SESSION

MOBILE COMPUTING + VIDEO PROCESSING + AD-HOC NETWORKS

Chair(s)

TBA

Hierarchical Bipartition Routing for Delivery Guarantee in Sparse Wireless Ad Hoc Sensor Networks With Obstacles

D. Gaußmann[1], **S. Hoffmann**[1], **E. Wanke**[1]
[1]Institute of Computer Science / Heinrich-Heine-Universität Düsseldorf, Germany

Abstract— *We introduce and evaluate a very simple landmark-based network partition technique called* Hierarchical Bipartition Routing *(HBR) to support greedy routing with delivery guarantee in wireless ad hoc sensor networks. The main assets of HBR are: 1) packet delivery guarantee, 2) unique virtual addressing, 3) no packet overhead in typical cases, 4) small routing tables, 5) no physical geographic coordinates necessary, 6) easy to combine with energy-aware approaches, and 7) simplicity. We evaluate the performance of HBR using realistic network topologies. We combine HBR with two energy-aware geographic greedy routing algorithms based on physical coordinates and virtual coordinates, respectively.*

1. Introduction and related work

In *geographic routing* protocols the decision to which neighbor the packet is sent is controlled by the position of the nodes and the distances between them. The position information of each node can be obtained either by devices such as GPS or Galileo (geographic coordinates) or by analyzing the network structure (virtual coordinates). Position awareness can often significantly improve the efficiency of routing. In [23] it is mentioned that protocols using position information for routing like MFR [29], COP [26], and GFG [14] are competitive alternatives to the classical routing protocols for wireless ad hoc networks, as for example DSR [16], AODV [24], and OLSR [8]).

In general, greedy routing algorithms do not guarantee packet delivery. A packet can be trapped in a local minimum where the algorithm will fail to find a next neighbor. The probability of reaching a so-called *dead-end* increases if the network is less dense or if the network contains obstacles where no nodes can be placed and/or connections are truncated by obstacles.

There are several attempts to obtain delivery guarantee for greedy routing algorithms. The authors of [4] propose *face routing*, which guarantees delivery in two-dimensional unit disk graphs (UDG). Face routing is applied to a planar sub-network obtained by considering the Gabriel Graph [4], [21], the Relative Neighborhood Graph [30], or the Morelia Graph [3]. In [27] a greedy-face-greedy (GFG) approach is considered, where greedy routing is based on COP as in [26] and face routing is similar to the one in [4]. Energy-aware routing is also proposed in LEARN [32], SPFSP [25], End-to-End (EtE) [10], and EEGR [34].

Landmark-based routing algorithms like VCap [5], JUMPS [2], GLIDER [11], VCost [9], and BVR [13] use virtual coordinates computed from the distances to specific nodes called *landmarks*, *anchors*, or *beacons*. In the first phase, a global and distributed election mechanism elects a set of nodes acting as landmarks. Then the landmarks flood the entire network or only parts of the network such that every node can compute its virtual coordinate depending on the distances to the landmarks. The virtual coordinates can then be used to route a message greedily through the network. Packet delivery is also not guaranteed if different nodes have the same virtual coordinates.

There are also several attempts to obtain delivery guarantee for landmark-based greedy algorithms. Most of them are based on a tree coordinate system like LTP [6] and ABVCap [22]. An energy efficient approach is introduced in HECTOR [23]. This protocol mixes the use of the tree-based coordinate system of [6] and the landmark-based coordinate system of [5] and [9].

An alternative way for delivery guarantee can be obtained by hierarchical addressing, see for example [31]. Tsuchiya solves this problem by allowing nodes to self-configure their addresses. The protocol uses a hierarchical set of landmark nodes that periodically send scoped route discovery messages. A node's address is the concatenation of its closest landmark at each level of the hierarchy. The overhead of route setup can be reduced to $O(n \log n)$ and nodes only hold state for their immediate neighbors and their next hop to each landmark. However, this requires a protocol that creates and maintains this hierarchy of landmarks and appropriately tunes the landmark scopes. Recent proposals adopting this approach have been fairly complex [19] in contrast to our design goal of configuration simplicity. See [15] for an overview of hierarchical protocols.

A second alternative to obtain delivery guarantee is clustering which is proposed by various researchers as for example in [7] and [18]. A closely related approach is the construction of connected dominating sets as routing backbones [33].

In this paper, we introduce and evaluate a very simple landmark-based network partition technique called *Hierarchical Bipartition Routing* (HBR) to support routing with delivery guarantee in wireless ad hoc sensor networks. It is a simple routing protocol that can easily be combined with any other greedy routing algorithm to obtain delivery guarantee. The efficiency of HBR increases if the network is

sparse and contains obstacles. These are exactly the networks where greedy algorithms will fail with high probability. The hierarchical bipartition of the network is performed on a landmark-based data structure setup in a pre-processing phase.

The space necessary to store the routing information at a node u is approximately $\log_2 n \cdot \log_2 \deg(u)$ bits in typical cases, where n is the total number of sensor nodes in the network and $\deg(u)$ is the number of neighbors of u. This is in general not larger than the size necessary to store the IDs of all neighbors of u. The amount of work to setup the complete data structure is in typical cases proportional to flooding the entire network $\log_2 n$ times. The sizes of the virtual addresses are then only a few bits larger than the sizes of the IDs.

We evaluate the performance of HBR using realistic network topologies. We combine HBR with two energy-aware geographic greedy routing algorithms based on physical coordinates and virtual coordinates, respectively. Our simulations show that the difference between using HBR and a weighted shortest path to escape a dead-end is only a few percent in typical cases.

To keep the advantages of greedy routing, we suggest to use HBR for finding the way out of a dead-end. This is especially very interesting if the network is sparse or contains obstacles. In these cases, the probability to reach a dead-end is very high. Our experimental evaluations even show the following: If more than 50% of the routes reach at least one dead-end, then the performance of stand-alone HBR is in general better than the performance of geographic greedy routing with a shortest path dead-end handling.

The main assets of HBR are already mentioned in the abstract. The drawbacks are: 1) works only for static network structures, 2) relative time-consuming and energy-consuming set-up phase, 3) an unbalanced partitioning can theoretically produce larger address lengths, and 4) unbounded theoretical worst-case stretch factor.

We do not introduce and evaluate HBR to be "yet another routing protocol better than others", but to be one of the more basic routing techniques that is worth to be classified and analyzed for networks whose sensor nodes are distributed in realistic environments.

2. Hierarchical Bipartition Routing

A *network* is modeled as an *undirected graph* $G = (V, E)$, where V is the set of *nodes* (*sensor nodes*) and $E \subseteq \{\{u, v\} \mid u, v \in V,\ u \neq v\}$ is the set of undirected *edges* (*connections*) between sensor nodes. A connection $e = \{u, v\}$ may have a positive *weight* $\omega(e) \in \mathbb{R}_+$, also denoted by $\omega(u, v)$ or $\omega(v, u)$. In general, the weight represents the amount of energy necessary to reach the neighbor node. A *path* $p = u_1, \ldots, u_k$ is a non-empty sequence of nodes $u_i \in V$, $1 \leq i \leq k$, such that $\{u_i, u_{i+1}\} \in E$, $1 \leq i < k$, is a connection between u_i and u_{i+1}. The *weight* of p is

$$\omega(p) = \sum_{i=1}^{k-1} \omega(u_i, u_{i+1}).$$

Path p is a shortest path between u and v, if there is no path p' between u and v with $\omega(p') < \omega(p)$.

The *ω-distance* $d^\omega(u, v)$ between two nodes is the weight of a shortest path between u and v.

The *hop distance* $d^h(u, v)$ between two nodes is the weight of a shortest path between u and v for the case that all connections $\{u, v\}$ have weight $\omega(u, v) = 1$.

To analyze the performance of HBR, we assume that every node $u \in V$ has a physical geographic position $(u_x, u_y) \in \mathbb{R}^2$ in the plane defined by a two-dimensional real vector. These positions are only used by the geographic greedy routing protocols, and not by the hierarchical bipartition technique. The *euclidean distance* between two nodes u and v is

$$d^e(u, v) = \sqrt{(u_x - v_x)^2 + (u_y - v_y)^2}.$$

Note that the ω-distance $d^\omega(u, v)$ and the hop distance $d^h(u, v)$ are defined by the network structure, whereas the euclidean distance is defined by the physical geographic positions of the nodes.

A network $G = (V, E)$ is *connected* if there is a path between every pair of nodes. Network $G' = (V', E')$ is a *sub-network* of G if $V \subseteq V$ and $E' \subseteq E$, it is an *induced sub-network* of G if $V' \subseteq V$ and $E' = \{\{u, v\} \mid \{u, v\} \in E,\ u, v \in V'\}$.

The nodes are assumed to be static. Each node is assumed to have a unique ID which is mainly used to break ties. The unique ID is also necessary to specify the target node of the packet, if we do not want to give every node a unique virtual address.

2.1 Initialization

In the first phase, an arbitrary node w and two *landmark nodes* x_0, x_1 are selected. Landmark node x_0 is one of the nodes that has maximum ω-distance to w, and landmark node x_1 is one of the nodes that has maximum ω-distance to x_0. Every node u with

$$d^\omega(u, x_0) \leq d^\omega(u, x_1) \text{ gets virtual address 0,}$$

and every node u with

$$d^\omega(u, x_0) > d^\omega(u, x_1) \text{ gets virtual address 1.}$$

Let G_0 be the network induced by the nodes with virtual address 0 and let G_1 be the network induced by the nodes with virtual address 1.

Lemma 2.1: If network G is connected, then G_0 and G_1 are connected.

Proof: Let v be a neighbor of a node u on a shortest path between u and x_0, then $\omega(u, v) + d^\omega(v, x_0) = d^\omega(u, x_0)$. If the virtual address of u is 0, then $d^\omega(u, x_0) \leq d^\omega(u, x_1)$.

Since $d^\omega(u,x_1) \leq \omega(u,v) + d^\omega(w,x_1)$, we get $d^\omega(v,x_0) \leq d^\omega(v,x_1)$. That is, all nodes v on a shortest path between u and x_0 have virtual address 0 and thus G_0 is connected. An analogous argumentation shows that G_1 is connected. ■

Next two landmark nodes $x_{\alpha\cdot 0}$, $x_{\alpha\cdot 1}$ from the connected sub-network G_α are selected for every virtual address $\alpha \in \{0,1\}$. Here $\alpha\cdot 0$ and $\alpha\cdot 1$ is the extension of α by symbol 0 or 1, respectively. Landmark node $x_{\alpha\cdot 0}$ is one of the nodes of network G_α that has maximum ω-distance to x_α in sub-network G_α, and landmark node $x_{\alpha\cdot 1}$ is one of the nodes of network G_α that has maximum ω-distance to $x_{\alpha\cdot 0}$ in sub-network G_α. A node u of G_α whose ω-distance to $x_{\alpha\cdot 0}$ is less than or equal to the ω-distance between u and $x_{\alpha\cdot 1}$ gets virtual address $\alpha\cdot 0$. It gets virtual address $\alpha\cdot 1$, if the distance between u and $x_{\alpha\cdot 0}$ is greater than the distance between u and $x_{\alpha\cdot 1}$.

The bipartition of every G_α into two further sub-networks $G_{\alpha\cdot 0}$ and $G_{\alpha\cdot 1}$ can be continued with the new virtual addresses α until all the created sub-networks consist of only one single node. In this case, the nodes of the network are uniquely identified by their virtual addresses. An inductive application of Lemma 2.1 shows that all the sub-networks G_α are connected.

2.2 Distributed address computation

Once the network is deployed, an arbitrarily selected node w starts flooding the network. The message carries a weight initialized to zero. The weight is increased by every forwarding node. If node u sends a message to a neighbor v_i then u increases the weight by $\omega(u,v_i)$. If a node receives more than one message, it will store and forward only the one with the smaller weight. If a node does not receive a new message for a while, the current weight represents the ω-distance to w. To be sure that the flooding is finished, the node has to wait for a time longer than the time required to propagate a message through the network.

The election of the landmark nodes for the bipartition of the network can be done by the following simple protocol. Assume we want to determine a unique node u with maximum ω-distance to some other node w. Then all nodes with a maximum ω-distance to w in its two-hop neighborhood (in case of a tie the nodes with maximum IDs) start sending a message back to w. (The route back to w can be stored during the update process of the ω-distance to w.) Node w receives all these messages and can select the node u with maximum distance to w. It has to wait for a while such that no further messages will arrive. Then it sends a message back to the winner, the node u with maximum distance to w.

2.3 Routing protocol

Assume a packet should be sent from a source node s to a target node t. If the virtual address of s starts with symbol 0 and the virtual address of t starts with symbol 1, then s is in sub-network G_0 and t is in sub-network G_1. In this case, the packet is sent step by step to a neighbor whose distance to landmark node x_1 is minimum until it reaches a node in G_1. Then it is routed within the connected sub-network G_1 using the second symbol of the virtual addresses, and so on.

More generally, let $\alpha\cdot d_u\cdot\alpha_u$ be the virtual address of the current node u and $\alpha\cdot d_t\cdot\alpha_t$ be the virtual address of the destination t such that $\alpha,\alpha_u,\alpha_t \in \{0,1\}^\star$, $d_u,d_t \in \{0,1\}$, and $d_u \neq d_t$. That is, the symbols left to d_u and left to d_t are equal in both virtual addresses. If $d_u = 0$ and $d_t = 1$, the packet is sent greedily towards landmark node $x_{\alpha\cdot 1}$, if $d_u = 1$ and $d_t = 0$, the packet is sent greedily towards landmark node $x_{\alpha\cdot 0}$. The packet does not leave the connected sub-network G_α. An inductive argumentation proves that HBR guarantees delivery.

Corollary 2.2: A packet sent with HBR always reaches its destination.

2.4 Address size

The sizes $|\alpha|$ of the virtual addresses α depend on the number of partitions necessary to obtain single node graphs. In this case, the virtual addresses are unique and can be used for the identification of the nodes. If the weights of all edges are equal, then a worst case for the address length is a complete network, that is, a network where every node is connected to all other nodes. Then the number of bipartitions and thus the address length is $n-1$. To avoid these worst-case situations, we can stop the bipartition process when all nodes of a sub-network G_α have hop distance ≤ 1 to landmark node x_α. This will considerably reduce the address size. Since now the nodes do not have unique virtual addresses, it is necessary to include the target ID in the packets.

The routing protocol can be extended as follows: Assume the packet reaches a node u that has virtual target address α but is not the target node. If a neighbor v of u is the target node, the packet can be sent to node v. Otherwise, the packet can be sent to x_α and from x_α to the target node. This is always possible, because all nodes with virtual address α are neighbors of x_α. The decision to stop the bipartition is very easy to implement, because the nodes of G_α that are candidates for flooding only have to check their list of neighbors.

If the weights of the connections are not all equal, but depend on the distances of the connections, then it is in general not necessary to abort the bipartition process. In practical cases, the different lengths of the connections lead to a partition of the network into two almost equally sized sub-networks. In this case, the virtual addresses are unique and it will not be necessary to use the original IDs.

2.5 Storage size

A node has to store its ID, its virtual address, a routing table, and temporarily during the initialization phase some ω-distances to landmark nodes and some source IDs. For the partition of sub-network G_α into $G_{\alpha\cdot 0}$ and $G_{\alpha\cdot 1}$, we only

need the ω-distances to $x_{\alpha\cdot 0}$ and $x_{\alpha\cdot 1}$. When the new virtual addresses $\alpha\cdot 0$ and $\alpha\cdot 1$ are assigned, it is no longer necessary to store these ω-distances, and the IDs of $x_{\alpha\cdot 0}$ and $x_{\alpha\cdot 1}$. For the routing of a packet, it is sufficient to know for every position i, $1 \le i \le |\alpha|$, the neighbor to which the packet has to be sent to if the own address α and the target address of the packet are equal at the first $i-1$ positions and differ at position i. If a node u with virtual address α has $\deg(u)$ neighbors, then the size of the additional routing information is only $|\alpha| \cdot \log_2 \deg(u)$ bits.

2.6 Worst-case behavior

From a theoretical point of view, the weight of a path routed by HBR can be arbitrarily larger than a shortest path between the source and target node. Figure 1 shows a simple example. The virtual addresses of the black and white nodes start with symbol 0 and 1, respectively. A shortest path between the source node s and the target node t has weight $2 \cdot a$. HBR routing will send the packet from s to v and then via x_1 to t. The weight of this path is $m \cdot a$, where m can be arbitrary large. However, this is not a typical case for randomly generated networks.

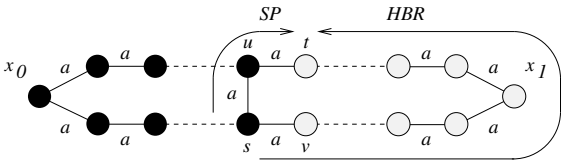


Fig. 1: An unrealistic worst case for the stretch factor of HBR

In the worst case, the size of the virtual addresses α can reach the number n of nodes. This is for example the case if the network is complete and all edges have the same weight, or if the network is a path and the connections have exponentially increasing weights $1, 2, 4, 8, 16, 32, \ldots$. However, if the number of nodes reachable with increasing distance is for most nodes approximately the same, the size of the virtual addresses is nearly $\log_2 n$. This is also confirmed by our experimental evaluations of randomly generated networks.

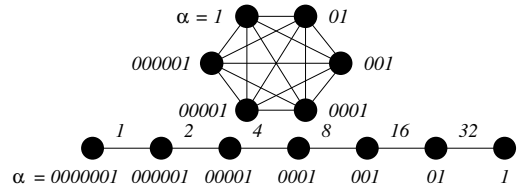


Fig. 2: Two unrealistic worst cases for the size of the virtual addresses

3. Two greedy routing protocols

We mainly want to use HBR to guarantee delivery for greedy routing protocols. For this reason, we combine HBR with geographic greedy routing based on physical and virtual coordinates.

3.1 Physical coordinates

The simplest energy-aware geographic greedy routing protocol sends the packet to a neighbor for which the ratio of cost over progress is minimum. This *cost over progress* (COP) power-aware framework is introduced in [20]. If we consider physical coordinates, the packet for target node t is sent from node u to a neighbor v of u for which

$$d^e(v,t) < d^e(u,t) \quad \text{and} \quad \frac{\omega(u,v)}{d^e(u,t) - d^e(v,t)} \quad \text{is minimum.}$$

If $\omega(u,v) = a$, $a > 0$, for all connections $\{u,v\}$, we obtain the simple geographic greedy routing that selects a neighbor closest to the destination [12]. For $\omega(u,v) = d^e(u,v)$, the routing is similar to compass routing [17]. If the angle between (u,v) and (u,t) is β and the distance to the target node tends to infinity, the ratio of cost over progress tends to $\frac{1}{\cos(\beta)}$. If $\omega(u,v)$ is defined by the commonly used energy function $a + b \cdot d^e(u,v)^c$, $a, b > 0$, $c \ge 2$, then an optimal routing tries to use equidistant steps towards the target node t. The best progress (also called the *characteristic distance*) is

$$d^\star = \sqrt[c]{\frac{a}{b \cdot (c-1)}},$$

see also [28]. The ratio of cost over progress has its minimum at the position with distance $d^\star$ from u in the direction to the destination.

3.2 Virtual coordinates

Our second greedy routing protocol is based on virtual coordinates which we will define by four landmark nodes denoted by A, B, C, and D. These four landmark nodes are selected similarly as in VCap (virtual coordinate assignment protocol) from [5]. The first landmark node A is one of the nodes with maximum ω-distance to an arbitrary node w. The second landmark node B is one of the nodes with maximum ω-distance to A. The third landmark node C is one of the nodes for which

$$d^\omega(C,A) + d^\omega(C,B) - 2 \cdot |d^\omega(C,A) - d^\omega(C,B)|$$

is maximum. And finally, the fourth landmark node D is one of the nodes for which

$$d^\omega(D,C) - |d^\omega(D,A) - d^\omega(D,B)|$$

is maximum. Since we are considering energy efficient routing, we use the ω-distances instead of hop distances [5] for the computation of the virtual addresses.

In case of a tie, the node with larger ID is chosen. The four landmark nodes A, B, C, D define for every node u a 4-tuple

$$(d^\omega(u,A), d^\omega(u,B), d^\omega(u,C), d^\omega(u,D)).$$

Our landmark-based routing protocol sends the packet to a neighbor v for which the ratio cost over progress is minimum. The progress $d'(u,t) - d'(v,t)$ is defined by distance function

$$d'(u,v) = \sqrt{\begin{matrix}(d^{\omega}(u,A) - d^{\omega}(v,A))^2 + (d^{\omega}(u,B) - d^{\omega}(v,B))^2 + \\ (d^{\omega}(u,C) - d^{\omega}(v,C))^2 + (d^{\omega}(u,D) - d^{\omega}(v,D))^2\end{matrix}}.$$

3.3 Dead-end handling

Both greedy routing protocols can reach a so-called *dead-end*, i.e., a node u that has no neighbor v closer to the destination than u. If a dead-end is reached, the packet is either sent along a shortest path or by HBR towards the destination node. In both cases the weight function ω is applied. The packet is sent hop by hop until a node is reached whose distance to the destination is less than the distance from the last dead-end node to the destination. Then the original greedy routing is continued.

It is obvious that a shortest-path routing is not possible in practice. We use shortest-path routing only to get a comparison with HBR under the assumption that following a shortest path is a good idea to get out of a dead-end.

The two geographic routing variants based on physical coordinates are denoted by GEO$^{\text{SP}}$ and GEO$^{\text{HBR}}$, the two variants based on virtual coordinates are denoted by LMR$^{\text{SP}}$ or LMR$^{\text{HBR}}$, depending on whether the dead-end problem is cleared with the help of a shortest path or by HBR, respectively.

4. Analysis

The analysis of HBR is done by randomly generated networks and randomly selected source and target nodes. The test environment and the obtained evaluation results are explained in the next subsections.

4.1 Experimental environment

Our networks have a size of $1000m \times 1000m$. The radio range is fixed at $50m$, the node density δ varies between $0.5 \cdot 10^{-3}$ and $9.2 \cdot 10^{-3}$ nodes per m^2, which corresponds to an average node degree between 4 and 72. If one of the randomly created networks is disconnected, we use the largest connected component, if its size is at least $\frac{2}{3}$ of the size of the complete network.

Networks with holes or obstacles are created with the help of black/white-masks. If the randomly selected position of a node hits a white-entry of the mask, the node is omitted. We do not try to find another position for this node. The masks we use for our evaluations are shown in Figures 3, 4, and 5.

The lakes-mask of Figure 3 (latitude $51.19°$, longitude $6.37°$) represents wet areas where sensors are lost during the dispersion process. The streets-mask of Figure 4 (latitude $40.70°$, longitude $-73.93°$) represents an area where the sensor nodes are assumed to be dispersed by vehicles driving along streets. The buildings-mask of Figure 5 (latitude $52.50°$, longitude $13.35°$) represents an example of a metropolitan area. Here we additionally remove all connections between sensors that can not see each other, because there is a building in between.

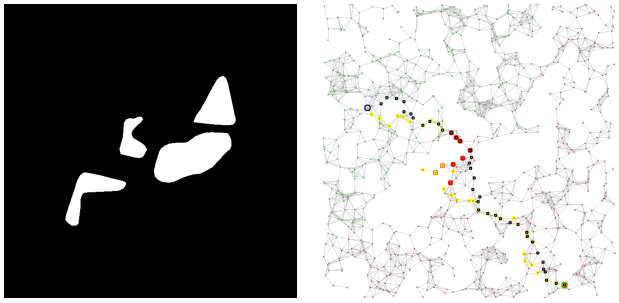


Fig. 3: Left: Lakes-mask, lat. $51.19°$, lon. $6.37°$; Right: Routing in a network generated with lakes-mask and density $\delta = 1.0 \cdot 10^{-3}$

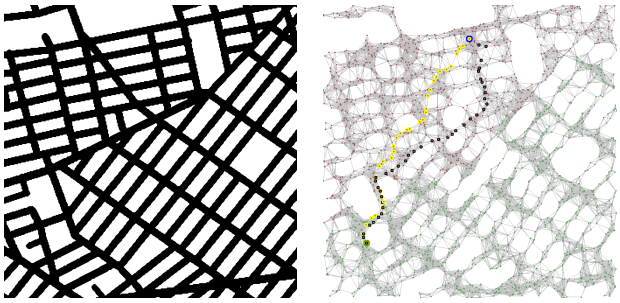


Fig. 4: Left: Streets-mask, lat. $40.70°$, lon. $-73.93°$; Right: routing in a network generated with streets-mask and density $\delta = 4.5 \cdot 10^{-3}$

We use the energy function $\omega(u,v) = 400 + d^e(u,v)^2$ for every connection $\{u,v\} \in E$, such that the characteristic distance is $d^\star = 20m$. Each of the Figures 3, 4, and 5 shows two routing paths to the right. The start node is encircled green, the destination node is encircled blue. The nodes traversed by HBR are colored black. The nodes traversed by GEO$^{\text{SP}}$ are colored yellow. The dead-end nodes of GEO$^{\text{SP}}$ are colored red. The small light green (light red) nodes have a virtual address starting with 0 (with 1, respectively).

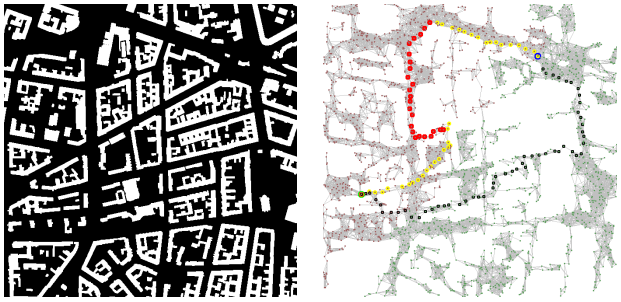


Fig. 5: Left: Buildings-mask, lat. $52.50°$, lon. $13.35°$; Right: Routing in a network generated with building-mask, only visible connections, and density $\delta = 4.5 \cdot 10^{-3}$

For every node density δ between $0.5 \cdot 10^{-3}$ and $9.2 \cdot 10^{-3}$ in steps defined by factor 1.2, we randomly create 1000 networks. For every network, we randomly selected 1000 source nodes and 1000 target nodes. Let $C(\text{HBR})$ and $C(\text{SP})$ be the sum of the costs to route from the 1000 source nodes to the corresponding 1000 target nodes in all 1000 networks using HBR and SP, respectively. The cost of a route is the sum of the weights of the used connections. The *overhead* of HBR is defined by

$$\frac{C(\text{HBR}) - C(\text{SP})}{C(\text{SP})}.$$

It is defined in the same way for the other routing protocols LMR^{SP}, LMR^{HBR}, GEO^{SP}, and GEO^{HBR}.

4.2 Evaluations

The Tables 1, 2, 3, and 4 show the average overhead in percent as a function of the node density δ. The overhead entries are colored continuously between green (0%) and red (50%). We think that it is more valuable to present the main results in a table than in a graphical illustration, because it is easy to create a graphical view from the values of the table, but not vice versa.

Table 1 considers networks where the sensor nodes are uniformly distributed. If the node density is greater than or equal to $1.0 \cdot 10^{-3}$, the two greedy routing algorithms GEO^{SP} and GEO^{HBR} on physical coordinates have less overhead than the greedy routing algorithms LMR^{SP} and LMR^{HBR} on virtual coordinates. For less dense networks ($\delta < 1.0 \cdot 10^{-3}$), HBR has even less overhead than GEO^{SP}, GEO^{HBR}, LMR^{SP}, and LMR^{HBR}. The difference between a shortest path dead-end handling and a HBR dead-end handling, i.e., the difference between GEO^{SP} and GEO^{HBR} and between LMR^{SP} and LMR^{HBR}, is only a few percent. This holds for greedy routing with physical coordinates as well as for greedy routing with virtual coordinates.

The average address lengths and the average number γ of routes that have at least one dead-end are shown in Table 1. These tables show that the sizes of the virtual addresses are only a few bits larger that the sizes of the IDs, which are at least $\lceil \log_2 n \rceil$.

HBR seems to be well suited for resolving the dead-end problem. The advantage of HBR is the very good performance in particular for sparse networks and for networks with obstacles. This show Tables 2, 3, and 4.

δ	HBR	LMR^{SP}	LMR^{HBR}	GEO^{SP}	GEO^{HBR}
0.5	2.94	20.06	21.20	26.01	27.29
0.6	5.83	22.10	24.63	27.17	30.17
0.7	11.52	22.11	26.69	26.84	32.54
0.9	16.87	17.15	21.35	21.92	28.22
1.0	17.92	12.10	14.32	15.60	19.63
1.2	17.96	9.59	10.55	10.91	12.76
1.5	18.02	8.27	8.64	7.98	8.71
1.8	18.05	7.38	7.51	6.27	6.49
2.1	18.09	6.76	6.80	5.20	5.25
2.6	18.21	6.23	6.24	4.49	4.50
3.1	18.33	5.78	5.78	3.94	3.94
3.7	18.59	5.39	5.39	3.47	3.47
4.5	18.82	5.02	5.02	3.02	3.02
5.3	19.08	4.87	4.87	2.61	2.61
6.4	19.32	4.94	4.95	2.26	2.26
7.7	19.58	5.15	5.15	1.95	1.95
9.2	19.82	5.29	5.29	1.69	1.69

δ	n	$\lvert\alpha\rvert$	$\lvert\text{ID}\rvert$	γ LMR	γ GEO
0.5	357	12.01	9	60.77	84.57
0.6	474	12.68	9	62.87	83.94
0.7	661	13.42	10	60.21	80.49
0.9	845	13.85	10	46.24	70.31
1.0	1030	14.10	11	28.25	53.14
1.2	1242	14.31	11	15.11	32.90
1.5	1491	14.51	11	7.32	16.58
1.8	1790	14.77	11	3.03	6.60
2.1	2149	15.07	12	1.09	2.00
2.6	2579	15.28	12	0.32	0.50
3.1	3095	15.59	12	0.08	0.08
3.7	3715	15.92	12	0.06	0.01
4.5	4458	16.16	13	0.04	0.00
5.3	5349	16.43	13	0.06	0.00
6.4	6419	16.76	13	0.09	0.00
7.7	7703	17.10	13	0.12	0.00
9.2	9244	17.29	14	0.15	0.00

Table 1: Top: Average overhead in percent as a function of density δ; Bottom: The average number n of nodes, the average address length $|\alpha|$, the size of the IDs ($= \lceil \log_2 n \rceil$), and the average number γ of routes that had at least one dead-end in percent as a function of density δ

References

[1] *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 13-17 March 2005, Miami, FL, USA*. IEEE, 2005.

[2] F Benbadis, J.-J. Puig, M. Dias de Amorim, C. Chaudet, T. Friedman, and D. Simplot-Ryl. Jumps: Enhancing hop-count positioning in sensor networks using multiple coordinates. *CoRR*, abs/cs/0604105, 2006.

[3] P. Boone, E. Chávez, L. Gleitzky, E. Kranakis, J. Opatrny, G. Salazar, and J. Urrutia. Morelia test: Improving the efficiency of the gabriel test and face routing in ad-hoc networks. In R. Kralovic and O. Sýkora, editors, *SIROCCO*, volume 3104 of *Lecture Notes in Computer Science*, pages 23–34. Springer, 2004.

[4] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. *Wireless Networks*, 7(6):609–616, 2001.

[5] A. Caruso, S. Chessa, S. De, and A. Urpi. Gps free coordinate assignment and routing in wireless sensor networks. In *INFOCOM* [1], pages 150–160.

[6] E. Chávez, N. Mitton, and H. Tejeda. Routing in wireless networks with position trees. In E. Kranakis and J. Opatrny, editors, *ADHOC-NOW*, volume 4686 of *Lecture Notes in Computer Science*, pages 32–45. Springer, 2007.

[7] C. Chiang, H. Wu, W. Liu, and M. Gerla. Routing in clustered multihop, mobile wireless networks. In *IEEE Singapore International Conference on Networks*, pages 197–211, 1997.

[8] T. Clausen, P. Jacquet, A. Laouiti, P. Muhlethaler, A. Qayyum, and L. Viennot. Optimized link state routing protocol (OLSR). *RFC 3626*, 2003.

δ	HBR	LMR^{SP}	LMR^{HBR}	GEO^{SP}	GEO^{HBR}
0.5	1.76	17.64	18.31	25.00	25.53
0.6	2.96	19.49	20.67	26.29	27.45
0.7	6.26	21.32	24.02	27.37	30.40
0.9	12.21	19.99	24.61	26.70	32.67
1.0	16.89	15.31	19.62	22.31	28.98
1.2	18.13	12.49	15.43	17.47	22.64
1.5	18.22	10.80	12.65	13.82	17.35
1.8	18.27	9.94	11.19	11.32	13.73
2.1	18.30	9.35	10.20	9.73	11.49
2.6	18.38	8.72	9.32	8.48	9.73
3.1	18.45	8.19	8.60	7.50	8.41
3.7	18.50	7.87	8.20	6.82	7.49
4.5	18.61	7.48	7.74	6.24	6.74
5.3	18.72	7.54	7.80	5.78	6.19
6.4	18.88	7.36	7.62	5.37	5.72
7.7	18.97	7.81	8.13	5.04	5.34
9.2	19.10	8.24	8.60	4.76	5.02

Table 2: Average overhead in percent as a function of density δ with lakes-mask

δ	HBR	LMR^{SP}	LMR^{HBR}	GEO^{SP}	GEO^{HBR}
1.2	4.08	21.59	23.23	27.53	29.35
1.5	7.57	23.27	26.50	28.56	32.14
1.8	12.87	22.53	27.46	27.18	33.08
2.1	16.70	18.36	22.89	22.72	28.67
2.6	17.59	14.61	17.67	17.82	22.09
3.1	17.81	12.62	14.58	14.23	16.87
3.7	17.95	11.67	12.96	11.87	13.38
4.5	18.03	11.09	11.98	10.42	11.34
5.3	18.14	10.76	11.42	9.62	10.19
6.4	18.19	10.40	10.87	9.06	9.46
7.7	18.26	10.18	10.53	8.63	8.94
9.2	18.18	9.92	10.20	8.26	8.52

Table 3: Average overhead in percent as a function of density δ with streets-mask

δ	HBR	LMR^{SP}	LMR^{HBR}	GEO^{SP}	GEO^{HBR}
1.8	2.95	24.06	25.55	27.79	29.29
2.1	4.70	25.40	27.76	29.37	32.01
2.6	7.56	26.45	30.27	31.14	35.71
3.1	11.90	27.01	32.89	32.67	40.11
3.7	16.40	27.03	34.64	33.05	43.29
4.5	19.10	26.31	34.17	32.49	43.86
5.3	20.71	25.60	32.98	31.51	43.04
6.4	21.20	24.89	31.38	30.40	41.42
7.7	21.31	24.71	30.87	29.97	40.36
9.2	21.37	24.59	30.57	29.61	39.55

Table 4: Average overhead in percent as a function of density δ with buildings-mask and only visible connections

[9] E. H. Elhafsi, N. Mitton, and D. Simplot-Ryl. Cost over progress based energy efficient routing over virtual coordinates in wireless sensor networks. In *WOWMOM*, pages 1–6. IEEE, 2007.

[10] E.H. Elhafsi, N. Mitton, and D. Simplot-Ryl. End-to-end energy efficient geographic path discovery with guaranteed delivery in ad hoc andsensor networks. In *PIMRC*, pages 1–5. IEEE, 2008.

[11] Q. Fang, J. Gao, L.J. Guibas, V. de Silva, and L. Zhang. GLIDER: gradient landmark-based distributed routing for sensor networks. In *INFOCOM* [1], pages 339–350.

[12] G.G. Finn. Routing and addressing problems in large metropolitan scale internetworks. Technical Report ISU/RR-87-180, Internetworks,ISI, 1987.

[13] R. Fonseca, S. Ratnasamy, J. Zhao, C.T. Ee, D.E. Culler, S. Shenker, and I. Stoica. Beacon vector routing: Scalable point-to-point routing in wireless sensornets. In *NSDI*. USENIX, 2005.

[14] H. Frey and I. Stojmenovic. On delivery guarantees of face and combined greedy-face routing in ad hoc and sensor networks. In M. Gerla, C. Petrioli, and R. Ramjee, editors, *MOBICOM*, pages 390–401. ACM, 2006.

[15] K. Iwanicki and M. van Steen. On hierarchical routing in wireless sensor networks. In *IPSN*, pages 133–144. ACM, 2009.

[16] D.B. Johnson, D.A. Maltz, and J. Broch. DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In C.E. Perkins, editor, *Ad Hoc Networking*, pages 139–172. Addison-Wesley, 2001.

[17] E. Kranakis, H. Singh, and J. Urrutia. Compass routing on geometric networks. In *CCCG*, pages 51–54, 1999.

[18] P. Krishna, M. Chatterjee, N.H. Vaidya, and D.K. Pradhan. A cluster-based approach for routing in ad-hoc networks. In *Symposium on Mobile and Location-Independent Computing*, pages 1–10. USENIX, 1995.

[19] S. Kumar, C. Alaettinoglu, and D. Estrin. Scalable object-tracking through unattended techniques (scout). In *ICNP*, pages 253–262, 2000.

[20] J. Kuruvila, A. Nayak, and I. Stojmenovic. Progress and location based localized power aware routing for ad hoc and sensor wireless networks. *IJDSN*, 2(2):147–159, 2006.

[21] J. Li, L. Gewali, H. Selvaraj, and M. Venkatesan. Hybrid greedy/face routing for ad-hoc sensor network. In *DSD*, pages 574–578. IEEE, 2004.

[22] K. Liu and N.B. Abu-Ghazaleh. Stateless and guaranteed geometric routing on virtual coordinate systems. In *MASS*, pages 340–346. IEEE, 2008.

[23] N. Mitton, T. Razafindralambo, D. Simplot-Ryl, and I. Stojmenovic. Hector is an energy efficient tree-based optimized routing protocol for wireless networks. In *MSN*, pages 31–38. IEEE, 2008.

[24] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector (AODV) routing. *RFC 3561*, 2003.

[25] J.A. Sánchez and P.M. Ruiz. Exploiting local knowledge to enhance energy-efficient geographic routing. In J. Cao, I. Stojmenovic, X. Jia, and S.K. Das, editors, *MSN*, volume 4325 of *Lecture Notes in Computer Science*, pages 567–578. Springer, 2006.

[26] I. Stojmenovic. Localized network layer protocols in wireless sensor networks based on optimizing cost over progress ratio. *IEEE Network*, 20(1):21–27, 2006.

[27] I. Stojmenovic and S. Datta. Power and cost aware localized routing with guaranteed delivery in unit graph based ad hoc networks. *Wireless Communications and Mobile Computing*, 4(2):175–188, 2004.

[28] I. Stojmenovic and X. Lin. Power-aware localized routing in wireless networks. *IEEE Transactions Parallel Distrib. Syst.*, 12(11):1122–1133, 2001.

[29] H. Takagi and L. Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transactions on Communications*, 32(3):246–257, 1984.

[30] G.T. Toussaint. The relative neighbourhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268, 1980.

[31] P.F. Tsuchiya. The landmark hierarchy: a new hierarchy for routing in very large networks. In *SIGCOMM*, pages 35–42, 1988.

[32] Y. Wangy, W.-Z. Songz, W. Wang, X.-Y. Li, and T.A. Dahlbergy. LEARN: Localized energy aware restricted neighborhood routing for ad hoc networks. In *SECON*, pages 508–517. IEEE, 2006.

[33] J. Wu. Dominating-set-based routing in ad hoc wireless networks. In I. Stojmenovic, editor, *Handbook of Wireless Networks and Mobile Computing, chapter 20*, pages 425–450. John Wiley & Sons, 2002.

[34] H. Zhang and H. Shen. Eegr: Energy-efficient geographic routing in wireless sensor networks. In *ICPP*, pages 67–75. IEEE Computer Society, 2007.

Progressive Download Video Rate Traffic Shaping Using TCP Window and Deep Packet Inspection

Ran Dubin[1], **Ofer Hadar**[1], **Rony Ohayon**[2] **Noam Amram**[3]
[1]Communication Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2]School of Engineering, Bar-Ilan University, Ramat-Gan, Israel
[3]LiveU Ltd, 5 Hagavish St, Kfar Sava 44641, Israel

Abstract - *Progressive download (PD) is a video streaming method over HTTP. Although PD is the most common streaming method over the internet it is highly inefficient from the internet service provider (ISP) point of view. ISPs need to compete with increasing competition, declining profitability and increasing client demand for network bandwidth (BW). ISPs, therefore, depend on the ability to optimize their network traffic, where video streaming has become the number one task.*

As ISPs depend on deep packet inspection (DPI) systems in order to optimized and control their network, client/server shaping solution cannot be leaned on. Furthermore, such solutions are traditionally created by different buffer -based systems such as Leaky Bucket., but it is problematic to implement them on buffer -limited systems. Therefore, a highly efficient video traffic solution is needed.

This paper presents a buffer -free video streaming traffic shaping solution, based on TCP window size and scale modification which depends on the CBR video encoding rate and network conditions. Our solution can save up to 60% percent of bandwidth per connection under certain viewing habits conditions .Our simulation, which consisted of 3600 users over the time span of one hour, managed to achieve better network utilization by up to 25%.

Keywords: Traffic Shaping, Video Mobile Network optimization, Flash Video, Progressive Download over HTTP, Deep Packet Inspection

1 Introduction

Video streaming is constantly gaining popularity, with hundreds of millions of Internet users viewing video online. Video streaming is now responsible for the majority of Internet traffic, and is expected to keep growing over the next years, growth that is expected to be even more substantial in mobile networks. Service providers now face many new challenges when managing their networks bandwidth (BW).
The MEDIEVAL [18] project is addressing the huge demand for video traffic by specifying an enhanced architecture to advanced mobile networks to handle video services and optimizing the Quality of Experience (QoE) rather than the Quality of Service (QoS). As part of the optimization proposed in MEDIEVAL a packet dropping approach for UDP traffic based on content priority was proposed in [20] for Scalable Video Coding (SVC) [19] however the mechanism for optimization of TCP based traffic has to adopt a different approach as proposed in here.
While traffic consumption was once dominated by peer-to-peer (P2P) networks, P2P traffic is now declining rapidly due to the increasing popularity of streaming services like YouTube, Netflix and Hulu. Streaming services has gained a market share of over 35% from the total network BW [1], while YouTube, for example, has over 2 billion video views per week and is responsible for more than 17% of mobile network traffic.

The most popular technology to stream video over the internet is called PD. PD uses HTTP over TCP in order to stream video content to the client's side. In this method, the video is packed in a container and encapsulated over the HTTP connection. The PD streaming server typically sends packets in the highest data rate possible with a constant payload size using the advantages of TCP's rate control mechanism. This way, video is encoded in constant bit rate (CBR), a very simple and reliable method. Video delivered using this technique can be played as soon as the player receives a small amount of data. However, while this technique is suitable for the client side, it isn't optimal from the ISP point of view. The PD technique requires a short time peak BW at the beginning of the transmission, in order to fill out the buffer at the client's side and avoid unsmooth video decoding (avoiding jitter phenomena). The average peak time in low resolution streams is ten seconds, but HD streams peak can last throughout the transmission. We have observed that the peak rate ratio compared to the CBR encoding rate can be up to 5 times more.

The initial peak video transmission creates two main problems. The first is the high BW demand from the server side, which leads to inefficient BW usage. The second problem is related to the fact that most clients tends to stop the streaming before it ends, meaning that data accumulated in the client's buffer will not be used.

Since ISP suffers from heavy congestion and limited BW resources, the PD method is not efficient in the sense of BW utilization. The purpose of this work is to improve the PD algorithm by reducing the initial peak rate and increasing the network's BW utilization, while enabling the client to obtain smooth video playback.

Another approach to video streaming that is gaining popularity at present is the Adaptive Streaming protocols (ASP) [2]. ASP implements various methods that adjust the video quality according to changing network conditions while transmitting, thus ensuring the best possible viewing

experience. ASP over HTTP is similar to PD: the stream is split into small chunks (each about a 2-10 second protocol, depending on the specific protocol implementation), and each chunk can be encoded into different bit rate and quantization variable bit rate (VBR). Based on the client's evaluation of network conditions, the suitable chunk is requested from the server. ASP, in general, consists of two different stages of operation: in the first approach, the "buffering mode", the player accumulates video content until an excessive amount of content has been buffered. Then the client automatically switches to the second stage, called "steady state", in which he will ask for a new chunk only when the previous one has been successfully received according to the new available network conditions. Using this buffer control technique, the protocol adapts the BW requirements while preserving the optimal QoE.

The ASP method, however, has one drawback. It is not efficient in the sense of memory requirement on the server side, due to several different encoding rates for the same video stream. Our suggested algorithm is a solution that combines the two approaches - DP and ASP. The proposed combination uses the advantages of both methods while using a single encoding of the original stream without any video encoding modification on the content provider side. This algorithm uses the simplicity of the first algorithm (PD) and the adaptive BW consumption of the second.

Traffic shaping is based on the ability to adjust the streaming data rate to the video data rate. The proposed algorithm can be regarded as an online video rate traffic shaping algorithm based on adaptive control over TCP window size. Traffic control/smoothing is usually attributed to VBR traffic. However, CBR traffic over TCP (PD), as previously mentioned, has a transferring rate of up to five times higher than the encoded CBR video rate traffic. Therefore, from the ISP's point of view, there is a need for traffic optimization from one hand, and from the other maintain high QoE.

There is a great deal of previous research regarding video rate smoothing algorithm [3-8]. However, most did not consider the streaming protocol, and is either more focused on the application level or based on preliminary video data rate knowledge, without consideration of real time environment.
Online smoothing for the live video streaming algorithm must have low time complexity, and usually deals with small chunks of real time video streaming. Smoothed video streams can be sent in a piecewise-CBR fashion [3]. Another suggested online method is VBR real time protocol (RTP), video rate smoothing at the proxy side [4]. [4]On the one hand, this method has a huge advantage over others, because it does not dependant on the server side, which leads to better QoE. On the other hand, expensive memory resources are used to store the stream at the proxy server. Our proposed solution would work at the proxy side as well and optimized the traffic from the proxy to the client.

Offline smoothing is ideal for video on demand, many research have been done on this subject including, work-ahead smoothing [5], Enhanced Piecewise Constant Rate Transmission and Transport (e-PCRTT) [6].Additional approaches to server side smoothing are "frame smoothed" streaming on the server side [7], and using a geometrical algorithm to determine optimal Leaky Bucket (LB) parameter for CBR video streaming [8]. However, this proposed algorithm needs extra information about the encoding .
It should be mentioned, that all of the aforementioned solutions are not suitable for ISPs because they depend on the will of the content video provider (server side) to shape their traffic.

A common solution for handling heavy congestion networks is based on the ability to identify the different network protocols and enforce QoS policies. Identification is based on DPI and enforced with different buffer algorithm such as LB. LB is a metaphor for a bucket with a hole: the bucket is full of water that drains in constant rate, which allows control over transmission speed. However, this requires a sufficient amount of memory, and while LB is considered to be a very reliable solution, it is not suitable to limited memory systems. Implementing traffic shaping in platforms with limited resources (buffers) can lead to major performance degradation.

Therefore, TCP window size control can be a suitable solution for real time systems with a limited buffer. A related work for QoS improvement using TCP window control and channel occupancy in wireless media [9] reduced the packet loss ratio of CBR traffic up to 45 percent. Our proposed solution controls and modifies the TCP window based on the stream encoded data rate. Window modification enables replacing the LB mechanism with TCP streams without dropping any packets, while adjusting to the client network conditions.

2 Proposed video rate shaping

2.1 Overview of the proposed shaping mechanism

Our solution is based on reading the PD video header, extracting the video data rate, and modifying the TCP window throughout the connection depending on the network condition, client buffer redundancy and video encoding rate.

Fig. 1 explains the general scheme of the proposed solution that consists of connection matching, DPI and TCP Window modification scheduler. The "Out gate" represents the opposite network interface card (NIC), through which we send the incoming packet to the other side of the network.

Connection matching purposes aim to control and manage the incoming traffic using 5-tupples: TCP\UDP, source\destination IP's and source\destination ports. Using the information from the first packet of connection, we redirected the packets to the appropriate next stage. If it is a new TCP connection or an ongoing connection that the DPI module didn't finish handling, then the next stage is DPI. If it isn't TCP or if the DPI module knows it is not PD then we forward the connection without any modifications. If it is a PD stream, we will redirect it to the TCP window modifier module.

The DPI module investigates the first packets of the connection, identifies each PD connection, updates

Connection Matching and calls the video container parser function. Parser is needed in order to extract the CBR data rate from the video header container. Here, we focused on the Adobe FLV [10] container, while using an FFmpeg [11] open source library as the base code for our container parser.

A study in [12] reveals that online video is viewed in full in less than 50% of the cases. Considering a movie streamed using PD can be fully downloaded after a mere few seconds, under this assumption traffic is not optimized. Therefore, our algorithm optimizes the traffic where QoE is ensured and optimized unwatched traffic that can be wasted. Therefore, the TCP window modifier module is divided into two stages: buffering state and steady state. The buffering state ensures that the client will achieve a sufficient amount of redundant buffer, which we define as *Threshold* [sec]. When the client redundant buffer size is greater than *Threshold* , we switch to steady state mode. Steady state job is to restrict the client buffer and prevent unnecessary streaming traffic from being sent from the server to the client. The combination of both states enables the algorithm to adapt to changing network conditions and remain unaffected by short term network problems.

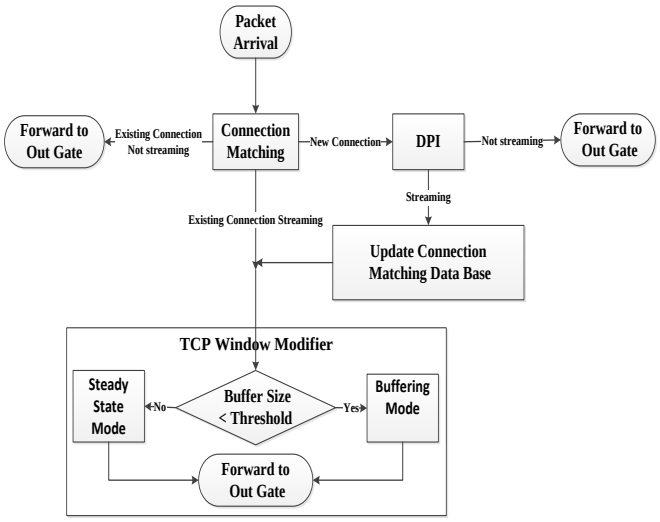


Figure 1: General Scheme of the proposed algorithm

2.2 TCP window size and scale modification

After parsing the container header using the DPI module, the TCP scale [13], (1) and TCP window [14] is to be updated. The proposed solution changes the SYN packet scale value and sets it to zero. Traditionally, the TCP window is calculated by (2). This is the proposed minimal window size (2) that should satisfy the minimum needs of our traffic window shaper.

$$Send.W = Segment.W << Send.W.Scale \quad (1)$$

$$W = \lceil RTT * DataRate \rceil \left[\frac{Bytes}{Sec} \right] \quad (2)$$

We gave discovered that (2) is insufficient because the server fills the window only if a full packet has enough space to enter. This is because PD stream servers use a fixed packet length size. Therefore, the optimized window size for streaming is (3.1). If (2) is used, then the stream is downloaded at a slower rate, which results in an unsmooth playing characteristic due to jitter effects. For example, given that W = 2500 and the Packet-Length = 1400 - because the window is too small for two packets, the streaming server will send a non -optimal window size with an actual size of 1400, which will cause an underflow of W_{uf} = 1100 Bytes. Another example: if W = 2900, then the W_{uf} = 100. This means we are W_{uf} short. In order to send a window that is suitable for our stream needs, it is required that compensate our window in order to send full window.

$$W_{\mathrm{mod}} = W \; Mod \; PacketLength \quad (3.0)$$

$$If\,(W_{\mathrm{mod}} == 0)\;\{W' = W + W_{\mathrm{mod}}\}$$
$$else\;\{W' = W + |W_{\mathrm{mod}} - PacletLeng\,th\,|\left[\frac{Bytes}{Sec} \right]\} \quad (3.1)$$

Therefore (4) underflow (W_{uf}) is the amount of data that was needed in order to send an optimize window. Overflow (W_{of}) is defined by the amount of redundant data that was sent in the current window (5) compared to (2).

$$UnderFlow = W_{\mathrm{mod}} \quad (4)$$

$$OverFlow = W' - W \quad (5)$$

From analyzing the previous examples, we can see that in the second one our optimized formula for streaming purposes created an increased W_{of}.

2.3 Modifier algorithm

Upon initialization of the algorithm, buffering mode is activated and the window size is set to W'. The client buffer is calculated with the consideration of the viewing progress over time. When the client buffer reaches a *Threshold* redundancy size, the steady state mode is activated. In order to control the client buffer we use two stages: soft shaping and aggressive shaping. In the first stage, TCP window size is modified with two kinds of windows: *W* and *W'*. The use of the small window - *W*, helps reduce and control the client buffer. However, in HD streaming the data rate is much higher, and aggressive temporary reduction of the window is needed (*tempW'* parameter). When control is achieved, adaptive increase of the client TCP window is necessary in order to preventing traffic peaks.

The following pseudo code illustrates the algorithm:

```
Buffering mode:
{
  WindowSize = W'
  If (Client Buffer Size > Threshold)
     {Go to Steady state mode}
  Else {Send modified packet and continue to
        Buffering mode}
}
Steady state mode:
{
  Gap = ⌈UnderFlow / OverFlow⌉
  Timer = T_size
  While ( (Timer > 0) && (Client Buffer Size > Threshold))
  {      //soft shaping stage
    Gap = Gap -1    // Gap minimum value is 1
    Send Gap modified packets with WindowSize = W' and
    one modify packet with WindowSize = W
  }
  If (Client Buffer Size < Threshold)
  {return to Buffering mode}

  While ( Client Buffer Size > Threshold)
  {    // aggressive shaping
    tempW' = W/2
    Send modified packet with window size of tempW'
  }

  If (Client Buffer Size < Threshold)
  {    //adaptive window size increases
    While (tempW' ≤ W')
    {
      NumOfPackets = 0

      While (Client Buffer Size < Threshold) &&
            (NumOfPackets < P_Gap)
      {
          NumOfPackets ++
          Send modified packet with window size = tempW'
      }
      If (Client Buffer Size > Threshold)
      {Return to Steady state mode}

      tempW' = tempW' + incFactor
    }// gradual window size increase. Every P_Gap packets
  }
  Return to Buffering mode
}
```

Algorithm 1**:** Window size modification traffic data rate shaping.

We now turn to explaining. Gap is the compensation between the overflow and underflow. If the gap is relatively small, less smooth traffic will be observed (resulting in a larger amount of small peaks). However, faster control over the client buffer is achieved. If the gap is larger, then the traffic will be smoother, but controlling the client buffer will be harder. Moreover, the system reaction time will be longer. If after *T_Size* seconds we cannot control the client buffer, we will switch to aggressive mode and adjust the window size to *tempW'*. Large *T_Size* slower system response time. In order to prevent unnecessary peaks in traffic, we adaptively increase the window size every *P_Gap* packets. We used a *Threshold* = 25 seconds of client buffer size. Influenced by adaptive streaming, the switch time period from buffering state to steady state is after the client accumulates 20-30 seconds of video redundancy (depending on the algorithm that is being used). We have tested the algorithm with different threshold sizes, and found that it is quicker to smooth the traffic when the client buffer size is small. The drawback of small threshold is that the traffic is less smooth. After hundreds of network testing with different ISPs, we found out that the optimized parameters best fitted from our standard deviation (STD) tests are: *T_Size* =3, *P_Gap* = 20 and *incFactor* = 10. However, the location of the implemented system can have different optimized values due to changing network conditions. Therefore, the system needs an initial optimization in order to a chive the desired results.

2.4 Implementation

The experiments were conducted on Ubuntu version 10.10, Intel I5-2300 2.8 GHz 8 GB RAM with 2 network interface cards, one for an internal network and the other for an external (WAN).The software was written in C/C++. In order to forward packets from one side to the other we used libpcap library [15]. The ADSL line capacity was 5Mbps link with download rate of 5.33 Mbps and upload rate of 0.7 Mbps. The following are the assumptions used to test our software:

Assumption 1 - The user will not skip to a later segment in the video. Skipping will close the current connection and open a new connection from the new segment point (assuming that the segment wasn't downloaded). This is the typical way PD works.

Assumption 2 - The user will not change the video resolution while watching (same explanation as assumption 1).

Due to space limitations, in this paper we will present only result with video resolution of 630 x 430.

3 Results

3.1 Results without the algorithm

In this case, where the stream was downloaded without shaping stream-1[16] was used. Fig. 2 shows that the ratio between the maximum download rate and the desired data rate is 7.

3.2 Shaping without buffer control

In this mode the algorithm only shape the traffic with the desired window (3) without steady state mode. In the

given example (Fig. 2), the stream is downloaded very fast because the download isn't optimal and the ISP suffers from overloading and high throughput on his line. It can be seen in Fig. 3 that we managed to minimize the maximum peak due to our TCP scaling and window modification. It is possible to decrease the maximum peak even more but in exchange to increasing initial play out delay. It can also be seen that we are above the desired data rate due to the overflow done by (3). Experiments have shown that using (2) will cause the stream to get stuck while downloading.

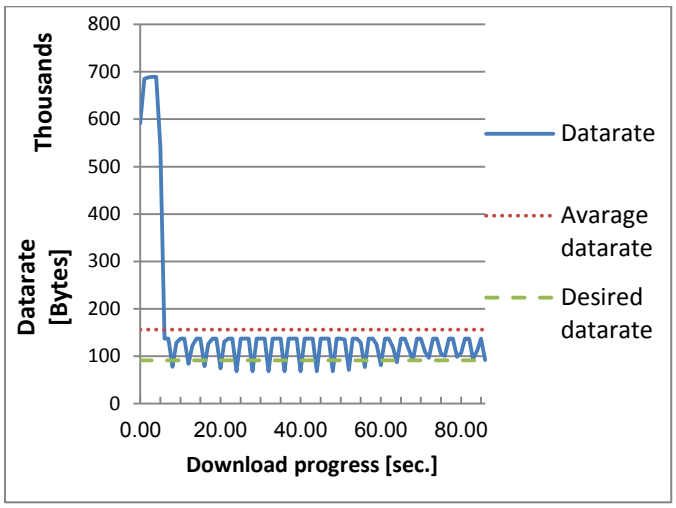


Figure 2: Stream-1 data rate without shaping. Desired datarate is the original video data rate

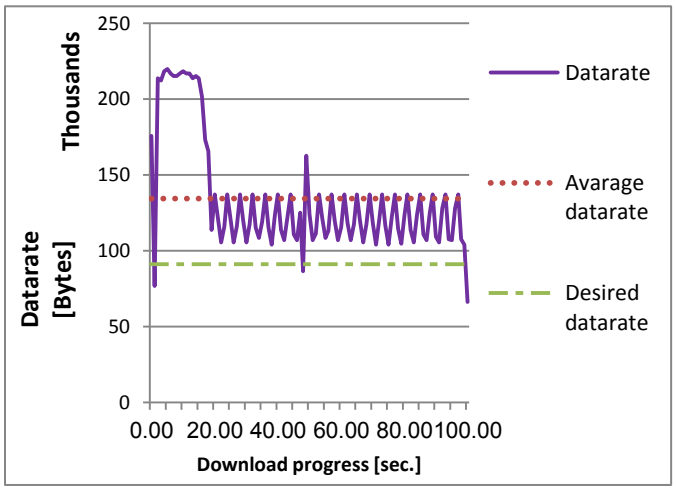


Figure 3: Stream-1 data rate with shaping without buffer control

Fig. 4 dash dot line shows that after 33 seconds we have already accumulated 25 seconds of redundancy stream buffered. In addition, when the stream is fully downloaded, the client has 42.5 seconds of redundancy. If we assume that most viewers will not watch the entire stream, this situation demonstrates the unnecessary download. We also observed, from other streaming sites and higher data rate networks, that PD can result in even faster download rates - we observed situations where the stream was downloaded fully in 10 seconds.

3.3 Shaping with buffer control

Fig. 5 presents the full algorithm, where at 32 seconds the algorithm switches to steady state mode. From Fig. 4 we can see that the maximum difference in the client buffer without the algorithm compared to shaping with buffer control is doubled. Furthermore, we can see that without shaping mode the client reached a 25 seconds buffer size in only 7.81 seconds. Table .1 sums up the statistical difference between the proposed methods.

Full video demonstration of the proposed solution with *Threshold* size of 5 [sec] can be seen here [17].

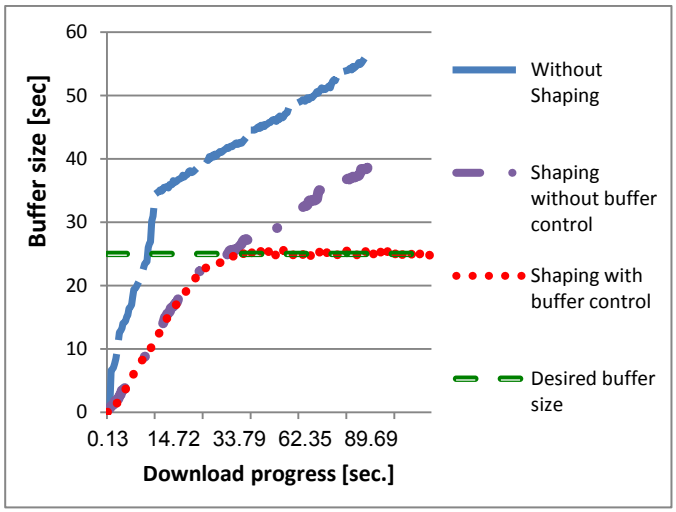


Figure 4: Stream-1 summery of buffer size comparison

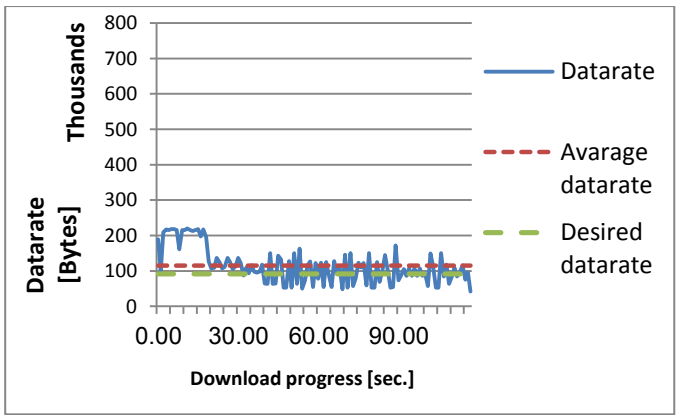


Figure 5: Stream-1 data rate shaping with buffer control

Table 1: Comparison of average data rate and standard deviation for different shaping modes for Stream-1.

Shaping method	Average data rate [Kbits/sec]	STD [Kbits/sec]
Without shaping	152.42	134.44
Shaping without buffer control.	131.28	36.7
Shaping with buffer control	112.34	48.98

4 Conclusions

In this work, a novel method and system for providing video data rate shaping for PD videos was introduced. The proposed solution is especially useful for mobile and 4G networks with existing or new DPI systems, as well for proxy servers. Since DPI systems have limited computational and memory resources, our proposed solution eliminates the buffer requirements and reduce the computational complexity.

5 Acknowledgment

We are grateful to Chen Hadad and Eyal Ron for their testing, evaluating and important inputs throughout this project. We would like to thank Tomer Margolin, Tal Vazana and Ofer Hermoni for their support and important feedback.

This work was supported by the EU MEDIEVAL research project. MEDIEVAL (MultiMEDia transport for mobIlE Video AppLications) [18] is a medium-scale focused research project (STREP) of the 7th Framework Program of the European Commission, addressing the core of the strategic objective "The Network of the Future".

6 References

[1] Allot Communication mobile trends report H2 2010.Available at: http://www.allot.com/MobileTrends_Report_H2_2010.html.

[2] S. Akhshabi, A.C. Begen, and C. Dovrolis, "An Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP," Proc. ACM Multimedia Systems Conf., ACM Press, 2011.

[3] J. Rexford, S. Sen, J. Dey,W. Feng, J. Kurose, J. Stankovic, and D. Towsley. "Online smoothing of live, variable bit-rate video". Proc. International Workshop on Network and Operating Systems Support for Digital Audio and Video, pages 249–257, May 1997.

[4] J. Rexford, S. Sen, and A. Basso, "A Smoothing Proxy Service for Variable-Bit-Rate Streaming Video," Proc. Global Internet Symp. Dec. 1999.

[5] J. Salehi, et al., "Supporting Stored Video: Reducing Rate Variability and End-to-End Resource Requirements through Optimal Smoothing, IEEE Transactions on Networking, Vol. 6, No. 4, August 1998, pp. 397 – 410.

[6] O. Hadar and S. Greenberg, "Statistical Multiplexing and Admission Control Policy for Smoothed Video Streams using e-PCRTT Algorithm", Proceedings of the International Conference on Information Technology: Coding and Computing, pp. 272 – 277, March 2000.

[7] A. Kashyap and B.Bing,"Efficient HD Video Streaming Over the Internet", Proceedings of the IEEE SoutheastCon pp.272-274 March 2010.

[8] Li, P., Lin, W., Rahardja, S., Lin, X., Yang, X., and Li, Z., "Geometrically determining the leaky bucket parametersfor video streaming over constant bit-rate channels," in [Proc. of ICASSP04], 20, 193–204 (February 2005).

[9] K.Ijiri, S.Ohzata, K.Kawashima, "window control for QoS improvement based on channel occupancy information over wireless LAN", Industrial Informatics (INDIN), 2010 8th IEEE International Conference ,2010.

[10] Adobe Inc. Adobe flash video file format speciation version 10.1. White paper, Adobe Systems Incorporated, August 2010.

[11] FFMPEG software library.Avilable: http://ffmpeg.org/

[12] Watching habits research can is available at: http://www.tubemogul.com/research/report/38-Brightcove-TubeMogul-Online-Video-and-the-Media-Industry-Q4-.

[13] V. Jacobson, R. Braden, and D. Borman."TCP extensions for high performance", 1992.

[14] V. Jacobsen and M. Karels, "Congestion avoidance and control," In Proc. ACM SIGCOMM, 1988.

[15] Libpcap software library and project .Available: http://www.tcpdump.org/.

[16] Avatar Trailer video can be watched at: http://www.youtube.com/watch?v=d1_JBMrrYw8.

[17] This Paper video example can be seen at: http://www.youtube.com/watch?v=B85L8KZODkQ&feature=mh_lolz&list=FLSHKdmosPXDR0gOk8vUeZzg.

[18] "Medieval Home Page." [Online]. Available: http://www.ict-medieval.eu.

[19] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC," IEEE Trans. Circuits Syst. Video Technology, vol. 17, no. 9, pp. 560–576, 2003.

[20] MEDIEVAL Project, Deliverable D5.1– "IP Transport Optimization: Initial Architecture", June 2011.

Algorithmic Support for Distributed IDS in MANETs

Paulo M. Mafra[1], **Joni da Silva Fraga**[1], **Altair O. Santin**[2]
[1]Automation and Systems Department, Federal University of Santa Catarina, Florianopolis, SC, Brazil
[2]Pontifical Catholic University of Parana (PUC-PR), Curitiba, PR, Brazil

Abstract— *The intrusion detection systems (IDS) are usually designed to work on local networks. However, with the development of mobile networks and their applications, it became necessary to develop new architectures for IDSs to act on these networks in order to detect problems and ensure the correct operation of data communications and its applications. This paper presents a distributed IDS model for mobile ad hoc networks that can identify and punish those network nodes that have malicious behavior. In this paper we describe the proposed model, making a comparison with major efforts in the literature on distributed intrusion detection systems for mobile ad hoc networks and present the results of tests obtained with an implementation of the proposed model.*

Keywords: Mobile Ad Hoc Networks; Intrusion Detection System; Distributed Algorithms;

1. Introduction

The last decade has witnessed a great evolution in communication technologies. Among these technologies are the Mobile Ad Hoc Networks, which form highly dynamic environments without the presence of concentrator units [1]. However, such technology is susceptible to a great variety of attacks by faulty or malicious units. The challenge that arises is to maintain the MANETs free from the activity of malicious or faulty nodes. In the face of the difficulty of avoiding the effects of malicious activities, mechanisms are necessary to at least minimize such effects. Intrusion detection systems (IDSs) can be used as one of these mechanisms [2]. The key question in these IDSs is to ensure that applications in mobile Ad Hoc environments can always evolve despite of failures, attacks by malicious entities or their own mobility. Works in this area deals mainly with problems applied to routing protocols in MANETs [2] and some works focus on problems of malicious behavior [2], [3], [4]. These works are limited to monitoring the communications in MANETs and the proposed IDS models usually have centralized elements and do not admit intrusions in their elements. There are no mechanisms to protect their own information.

The communication among entities in MANETs can be monitored by these entities in order to detect faulty or malicious behavior and to improve safety and reliability of these environments. In this paper we propose a secure and fully distributed IDS model for mobile Ad Hoc networks. We apply concepts of distributed systems and dependability. The use of these concepts allows, within certain limits, the construction of an IDS less subject to restrictions than those present in other IDS models. The algorithms presented in this paper define an IDS with functions performed by groups of components (fully distributed executions) and with mechanisms and techniques that are tolerant to malicious activities of their own components. In section 2 we describe the IDS organization. In section 3 we introduced the algorithmic base that supports our secure IDS. In section 4 we present an analysis of the costs involved in the communication of IDS nodes and some results of performed tests in a simulator. Following (section 5), related work are presented. We finally present our conclusions.

2. Architecture Description

Communication protocols in MANETs depend upon collaboration from equipment which form these network nodes. As such, we also assume this collaborative environment. However, we do not limit this collaboration. I.e. all the network nodes participate in the intrusion detection. The proposed IDS model is distributed and should assume a hierarchical stratification in order to attend the diverse IDS functions. The differentiation of functions and their distribution among nodes in this collaborative environment establish two classes of nodes: the "leader nodes", which perform higher level functions such as analysis; and the "collector nodes", which assume lower level functionalities such as sensors, collecting data for future analysis. We also assume that all the network nodes possess at least $2f+1$ neighbors, where f is the failure or intrusion limit that our algorithms should support[1].

2.1 Topology and Component Description

The hierarchical topology of the model introduces the idea of clusters, as well as in [2], [3]. However, in our model, we consider each cluster to have various "leaders". This various leaders form what we denominate as "leadership". The leaders are chosen by its connectivity and available energy, but if it is necessary to maintain the number of $2f+1$ leaders, any node could be chosen as a leader. These leaders

[1]This limit f will be used in our approach of a threshold for the occurrence of abnormalities which, once they are not surpassed, the IDS and its distributed functions will continue to supply the correct and expected behavior. At this limit, malicious and faulty nodes present in the cluster and further, departures or churn during a predefined period are accounted for at a given instant named *epoch_time*.

which form a leadership in a cluster use "secure channels" to exchange information and alerts among themselves. The collecting nodes send a summary of the data collected to the leaders also through "secure channels". The leadership which define the domains of a cluster are instituted based on $2f + 1$ leaders. Leadership and, consequently, the corresponding cluster no longer exist as of the moment in which a leadership has less than $2f + 1$ leaders.

The collecting nodes constitute the largest part of the cluster. In order to belong to a cluster, the collecting nodes need to possess secure channels with at least $f + 1$ leaders of the cluster. The data collected by the collectors always considers its neighborhood. Ie. the collecting nodes capture the data from each neighbor and store the information in them in a table. This information may be, for example, the quantity of packets received and sent by the neighboring node i. The leader nodes analyze the data sent by the collecting nodes related to the same cluster. Based on the analysis of this monitoring information, they make their decisions concerning malicious nodes. These decisions are in turn registered in local lists (ex. malicious nodes list) and later shared, compared, and synchronized with the remaining leader nodes which make up the same leadership which determines the existence of the corresponding cluster. Thus, it is not necessary consensus between the leaders of a cluster. Since the leadership has at least $2f + 1$ leaders, the same choice of $f+1$ leaders about the behavior of a given node is sufficient for this decision to be considered by the leadership.

2.2 Cryptographic Mechanisms

IDS communication, which occurs among (collector - leader) and (leader - leader) cluster nodes makes use of cryptography. Encrypting messages is done with the use of session keys and symmetric encryption. Key distribution involves the public keys (pair of asymmetric keys) for each participating node. The public key of this pair is signed by the committee that manages the system, generating a certificate for node's authentication. These asymmetric cryptography keys are one of the prerequisites for any participating node in the system and in the IDS. The leadership also authenticates its messages. These leadership authentications are founded in the threshold signature scheme (TSS) [5].

The public key of the leadership (E_l) will always been known by all the cluster nodes and is used in verifying leadership signatures. The corresponding private key (D_l), used to generate the leadership signatures, possesses guaranteed sanctity through threshold cryptography. In other words, the key D_l is not available in any moment in the system. Its use is through a K set of partial keys ($|K| = m, K = SK_1, SK_2, ..., SK_m$) derived from D_l using threshold cryptography. In this model, generating and verifying partial signatures is completely non-interactive, without needing messages exchanges to execute these operations. In the proposed distributed IDS, each one of these m partial keys is delivered to a specific leader from leadership. Any operation with D_l is only possible through the participation of at least $f + 1$ leaders and their partial keys.

2.3 The System Dynamics

Collectors send data monitoring summaries to leadership. Such data is sent upon concluding a time period called the transmission time (Tt). The leaders analyze the data sent from collectors at the end of each (Tt). We call epoch the periods in which each cluster "freezes" its composition. In other words, during one epoch it is assumed that the composition of the cluster does not change. At each epoch there are several data collections (several Tt's). The transmission time (Tt) occurs n times in each epoch ($Tt = epoch/n$).

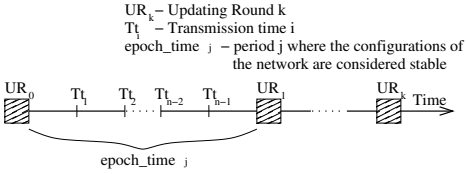


Fig. 1: Temporal relation among updating rounds

The changes which occur in the system during this period are not updated. I.e. possible system changes (within an epoch) such as faults, node entrances or departures, node exclusion, etc. are not taken into consideration in composing the cluster during that time. At the conclusion of each epoch, the cluster must synchronize itself. Thus, the updating round (UR) is initiated. These updating rounds, also present in [6], define a time period where cluster leaders exchange information in order to update their knowledge concerning the present state of the cluster. Therefore, in each synchronizing period (one updating round), the changes that occur in the cluster at last epoch are considered. Then, the composition for the new epoch is defined. During these updating rounds are also defined new roles for cluster nodes. The decisions taken during updating rounds will always depend upon the agreement of $f + 1$ leaders. The ratio of these periods of time is illustrated in Figure 1. At the end of the UR a new epoch is initiated.

3. Algorithmic Support for the IDS

3.1 Route Discovery and Update

In this model we adopted the DSR routing protocol [7], an on demand routing protocol widely used in MANETs with a local routing cache. We adapted the DSR algorithm to update the local routing cache, inserting $f+1$ disjoint routes to the leadership. This adapted algorithm (named RouteUpdate()) also updates the list of neighbors and the list of leaders from the node that is running it. All the existing routes from i to a leader l are obtained (through the Route Discovery mechanism in the DSR). Once the routing cache is updated, the list of neighbors from i is generated. Each neighbor node is extracted from this routes. If the node that is running this

algorithm is a leader, it attempts to get the maximum number of disjoint routes to other leaders as possible ($f+1$ at least). Otherwise, it gets just $f+1$ disjoint routes[2].

3.2 Data Analysis

Data analysis is performed in each updating round. This analysis takes into account available data from neighbors of a node and is implemented based on the system OctopusIIDS [8]. Each leader makes the analysis of each network node based on the data that it received from collectors. If the statistical analysis of $f+1$ nodes points node *i* as suspect, then it is considered suspect by the corresponding leader.

This analysis is done by all the leaders present in a leadership. The comparison (through the exchange of encrypted and authenticated messages) among the results obtained in analysis from these leaders is made concrete as well as in the leadership. The results obtained by these comparisons will be considered by the cluster as a whole. With $f+1$ leaders agreeing upon the analysis results, these results will be considered by the leadership. If there are at least $2f+1$ leaders, the cluster will always decide upon any analysis result, even in the presence of f malicious leaders.

3.3 Data Dissemination

Each leader will only be connected to a cluster if it possesses routes to at least $f+1$ leaders of that cluster, just as any node in the network. In order to the messages reach leadership, there must always be dissemination in leadership based on the correct leader. The algorithm 1 describes the steps of the dissemination protocol. In this algorithm, a node j disseminates a message msg_j in the $Leadership_j$ (lines 3-6 of algorithm 1) which corresponds to its knowledge about the leaders that form the cluster's leadership. Upon receiving the message msg_j, each leader in turn sends it to its respective leadership knowledge (lines 7-11). In this algorithm, at least one correct leader is reached with each resend, which returns to disseminate the message once again, using the recursion of the protocol $Disseminate()$ [9]. As each leader is connected to at least one correct leader and if the cluster leadership does not form disjointed graphs, then the majority of leaders will be reached through such dissemination.

3.4 Synchronizing periods: epoch times, and updating rounds

In order to deal with the dynamic aspects of the network and collecting data for the detection process, it was necessary to define times which determine the synchronization of the actions distributed throughout the system. As we work essentially with time periods, synchronizing the clocks is not necessary for the nodes to initiate synchronized operations.

[2]The collector nodes need to reach just one correct leader that will disseminate the message through the leadership.

Algorithm 1 Disseminate Protocol

1: {**On Initialization**}
2: $Received_k \leftarrow \{\}$;
3: {**On Disseminate**$(msg_j, Leadership_j)$ **at** *node j*}
4: **for all** $l_k \in Leadership_j$ **do**
% msg_j is send to leaders l_k known by node j
5: $send < DISSEMINATION, msg_j > to\ l_k$
6: **end for**
7: {**On Receive**$(< DISSEMINATION, msg_j >)$ **at** l_k}
8: **if** $(< DISSEMINATION, msg_j > \notin Received_k)$ **then**
9: $Received_k \leftarrow Received_k \cup \{< DISSEMINATION, msg_j >$ };
10: $Disseminate(msg_j, Leadership_k)$
11: $deliverDissemination(msg_j)$; % msg_j locally delivered in l_k
12: **end if**

Using their local clocks, the periods are controlled with their respective deadlines with timers which aid corresponding operational activation. The routine presented in Algorithm 2 is used to initiate a synchronized common activity among network nodes. It depends upon the course of the stipulated time and reception of at least $f+1$ corresponding sync messages.

The idea of this algorithm is to stipulate a period d and at the end of this period to start sending a $sync$ message to the leadership (lines 4-7 of algorithm 2). Upon receiving this message, the leader saves it and checks how many $sync$ messages he has received from different leaders. If the number of messages is greater than $f+1$ then this leader sends again a $sync$ message for the other leaders and collectors in order to force them to get in synchronizing period, if they still have not received $f+1$ messages (lines 8-13). After this last send, the sender may start participating in the updating round (UR) (line 14). At this time we also synchronize collector nodes, synchronizing the transmission times (Tts) (line 17).

Algorithm 2 Synchronizing() at node i

1: **Require** : $receive < sync_j >$ *or period d is elapsed in* N_i
2: **Init** :
3: $\delta \leftarrow$ **time**() % starts a new period counting d
4: **upon** $((\textbf{time}() - \delta) \geq d)$ **do** % at the end of d
5: **if** $(i \in Leadership_i)$ **then** % if i am leader
6: $Disseminate(sync_i, Leadership_i)$ % send of messages *sync* by i
7: $\delta \leftarrow$ **time**()
8: **upon** $receive(sync_j)$ **do**
9: $Sync_i^d \leftarrow Sync_i^d \cup \{sync_j\}$ % sync messages received by i
10: **if**$(\mid Sync_i^d \mid \geq f+1) \wedge (i \in Leadership_i)$**then**
11: $Disseminate(sync_i, Leadership_i)$ % call leaders to synchronize
12: **for all** $k \in Collectors_i$ **do**
13: $send < sync_i > to\ k$ % call collectors to start synchronizing
14: $UR()$ % a new synchronized UR is started
15: $\delta \leftarrow$ **time**()
16: **else if**$(\mid Sync_i^d \mid \geq f+1) \wedge (i \in Collectors_i)$**then**
17: $Tt()$ % a new synchronized Tt is started

3.5 Transmission times

Algorithms for the activities corresponding to transmission time (Tt) periods and updating round (UR) periods were defined. In the $Tt()$ algorithm, a deadline in which each

collector node must send a summary of collected data during the last period (time between Tt -1 and Tt) to the leadership was established through the Disseminate() protocol (lines 7-11 of algorithm 3). Upon receiving the message, the leader node verifies if the message is not older and saves it (lines 12-14). If the number of received messages in this (Tt) were greater than or equal to $2f + 1$ (it is the minimum required) (line 15) and if the number of received messages were greater or equal to the number of collectors minus f or if the $timeout$ has been achieved (line 16), then the leader node starts analyzing each cluster node by checking in the received data if the node is suspect by most of its neighbors (lines 17-21). In such case, this suspected node is inserted into the list of suspects from that leader node (lines 19-20).

Algorithm 3 Tt($monitoring\_data_i, Tt_j$)

1: **Init** :
2: $\gamma \leftarrow \textbf{time}()$
3: $Tt_j \leftarrow 0$
4: $Collected\_data_l \leftarrow \emptyset$ % buffer for monitoring data
5: $suspects\_List_l \leftarrow \emptyset$
6: $timeout \leftarrow p/3$ % p is the value of one Tt

7: **upon** $((\textbf{time}() - \gamma) \geq p)$ **at** $node\ i$ **do** % p is the period between two Tts
8: **if** $(i \in Collectors_i)$ **then**
9: $Disseminate(< monitoring\_data_i, Tt_j >, Leadership_i)$
10: $\gamma \leftarrow \textbf{time}()$
11: **end if**

% leader l receives data from collector i
12: **upon** $receive(< monitoring\_data_i, Tt_k >)$ **at** $l : l \in Leadership_i)$ **do**
13: **if** $Tt_k = Tt_l$ **then**
14: $Collected\_data_l \leftarrow Collected\_data_l \cup \{< monitoring\_data_i, Tt_l >\}$
15: **if** $(|\ Collected\_data_l\ | \geq 2f + 1)$ **then**
16: **if** $(|\ Collected\_data_l\ | \geq |\ Collectors_l\ | - f) \vee ((time() - \gamma) \geq timeout)$ **then**
17: **for all** $node\ n \in Cluster(Leadership_l)$ **do**
18: $analysis_l^n \leftarrow Data\_Analysis(Collected\_data_l, id_n)$
19: **if** $analysis_l^n \in is\_suspect$ **then**
20: $suspects\_List_l \leftarrow suspects\_List_l \cup \{id_n\}$
21: **end for**
22: $Tt_l \leftarrow Tt_l + 1$
23: $Collected\_data_l \leftarrow \emptyset$
24: $timeout \leftarrow EstimatedTime(timeout)$
25: **end if**

In this system, for the purpose of estimating the time when the majority of messages sent by collectors arrive at leader nodes, we apply an adaptive timeout that can auto-adjust over time. This adaptive timeout is based on the timeout used in TCP protocol proposed by Jacobson [10]. In it, each sent message involves the sender estimating a time interval to receive of an acknowledgment from the destination. We call this algorithm $EstimatedTime()$. In this algorithm the observed error from last timeout was first calculated and the next value for the timeout then estimated.

3.6 Updating rounds

An updating round (UR) is defined as a period of time where the cluster nodes update their views about the cluster. This view is composed by a set of lists (malicious nodes, leaders, collectors, etc). During these URs, the information about suspected nodes is shared among cluster leaders. In an UR node entrances and departures are also processed. Following, we explain how UR() algorithm works.

In UR() algorithm the list of suspects generated from last epoch during (Tts) is sent to other leaders through the Disseminate() protocol. Upon receiving this list, the leader node checks if the message is not older. If the number of received messages is greater than or equal to $2f + 1$, then the leader node executes the function *identifies_Suspect()* in which an analysis of each node is performed searching for nodes reported to be suspicious by at least $f + 1$ leaders. These nodes are included in the malicious node's list of the leader. In the sequence, the leader node disseminates a message asking for the election of a new coordinator. These messages are saved and when the number of messages reaches $2f + 1$, a leader is elected as coordinator.

The coordinator then processes the node entrance and departure requests in the cluster, generates a new view of the cluster (with new nodes) and spreads this view to all the nodes in the cluster. Upon receiving the new view of the cluster, if the node is a leader, it checks the message validity. If the message is valid then the node updates its view starting a new epoch. If it is not valid, a message asking for new coordinator is spread throughout leadership again. If the node that received the new view is a collector, the message signature is verified and its list of neighbors is updated.

3.7 Identification, entrance, and departure of nodes

The node identification process in the network should be secure enough so that it cannot create multiple identities. Thus, it is possible to prevent attacks like the Sybil [11]. In our model, node identification is carried out with the use of certificates. A certifying authority (CA) considered to be known and reliable by the network nodes generates a certificate for the public key for each node when it joins the system. The role of the CA in our model is assumed by a certifying entity[3]. With this model, we may define the user and his/her equipment. There will not be users using multiple equipment on the network.

Algorithm 4 presents the steps involved to insert a new node within a cluster. Upon entrance, a new node i must broadcast an entrance request message (REQIN). A neighbor, upon receiving REQIN, disseminate it to the leadership. This message informs its identification (node id) and its credential (its public key in certificate form) (lines 2-9 of algorithm 4). The new node i, upon receiving the new view of the leadership verifies the signature and initializes its view of the cluster, assuming the role assigned to it by the

[3]This certifying entity will not necessarily need to be an official PKI. It may be a system management commission, an administrator, etc.

Algorithm 4 Entrance of node i into the cluster

1: $broadcast < REQIN, id_i, cred_i >$ **to** j % broadcast REQIN
2: **upon receive** $< REQIN, id_i, cred_i >$ **at** node j **do**
3: **if** $VerifiedSignature(cred_i, admin)$ **then** % node j verifies the credentials of i
4: **if** $(j \in Leadership_j) \wedge (i \notin black\_List_j)$ **then**
5: $inactive\_nodes_j \leftarrow inactive\_nodes_j \cup < REQIN, id_i, cred_i >$
6: $Disseminate(< REQIN, id_i, cred_i >, Leadership_j)$
7: **else if** $(j \in Collectors_j) \wedge (i \notin black\_List_j)$ **then**
8: $Disseminate(< REQIN, id_i, cred_i >, Leadership_j)$
9: **end if** % node i wait for a response of the leadership in next UR
10: **upon** $receive < view, tUR_l, id_c >$ **at** $node\ i$ **do**
11: **if** $VerifiedValidity(cert_i : cert_i \in view.certificates)$ **then**
12: **if** $i \in view.Lead$ **then**
13: $role_i = leader$
14: **else if** $i \in view.Col$ **then**
15: $role_i = collector$
16: **end if**
17: $Leadership_i \leftarrow view.Lead$
18: $Collectors_i \leftarrow view.Col$
19: $RouteUpdate()$
20: $CN_i \leftarrow listNeighbors_i$
21: $listNeighbors_i \leftarrow CN_i \cap (view.Col \cup view.Lead \setminus view.black\_List)$
22: $tUR_i \leftarrow tUR_l$
23: **end if**

leadership[4]. The list of neighbors will be formed by the intersection of the list of neighbors, the list of collectors and the list of leaders of the new view, excluding the nodes that were identified as malicious (lines 10-23).

Algorithm 5 Departure of node i

1: {**On Departure() at node i**}
2: $Disseminate(< REQOUT, id_i, cred_i, tUR_i >, Leadership_i)$
3: **upon receive** $< REQOUT, id_i, cred_i, tUR_i >$ **at** node l **do**
4: **if** $l \in Leadership_i$ **then**
5: $leaving\_nodes_l \leftarrow leaving\_nodes_l \cup \{< REQOUT, id_i, cred_i >\}$
% node i waits a response from leadership in next UR
6: **upon** $receive < view, tUR_l >$ **at** $node\ i$ **do**
7: **if** $VerifiedSignature(cert_i : cert_i \in view.certificates))$ **then**
8: $node\ i\ leaves\ the\ system$
9: **end if**

Algorithm 5 presents the steps that involve the departure of a node from the system. In this algorithm, the node that want to leave the cluster, disseminate a message REQOUT to the leadership. Upon receiving the message, each leader adds the message in a list of leaving nodes. In the next UR, when the node that want to quit receive the new view, it checks the signature and can then leave the system. The entrances and departures of the system are always considered during periods of updating rounds. Otherwise, the node will be considered as a malicious node.

4. Results and corresponding analysis

4.1 Communication Costs

The communication costs of the algorithms proposed in this model depend principally upon the size of the cluster and its leadership. As such, we calculate these costs in terms of the messages sent from each algorithm. Table 1 presents these costs for each algorithm, in which: n is the number of nodes; l is the number of leaders in the cluster; c represents the number of collectors in the cluster. We adopted the limit of $f = (l-1)/2$. In other words, the number of malicious or faulty nodes is half minus one of the cluster's leader nodes (it is the limit of anomalies that the IDS supports to operate properly). In this calculation, we also assume the costs of the cryptographic mechanisms, presented in [6].

Table 1: Communication costs

Entrance of a node in the network	$c + (l^2 + 10l + 1)/4$
Transmission time (Tt)	$cl/2 - c/2 + (l^2 + 2l + 1)/4$
Updating Round (UR)	$(l^3 + 2l^2 + l)/4$

In the entrance process of a node in the network (Algorithm 4), this node sends a request to his neighbors (c in the worst case). Upon receiving the request, each neighbor node resends the request to the other leaders of leadership using the protocol Disseminate(). The cost of this protocol is given by $D = (f + 1) * (f + 1)$ where $f = (l - 1)/2$. In this calculation, we arrive at $D = (l^2 + 2l + 1)/4$. Thus, the total cost for a node gets into the network is then $(c + D)$, or $c + (l^2 + 2l + 1)/4$. In calculating the costs for transmission time (Tt), each collecting node sends a message with monitoring data to $f + 1$ leaders and the first leader to receive the message executes the Disseminate() protocol. As such, we arrive at $c * (f + 1) + D$. Through substitution, we arrive at $cl/2 - c/2 + (l^2 + 2l + 1)/4$. In calculating an updating round, each leader disseminates, into the leadership, a message containing the result of its analysis concerning each cluster node $(l * D)$ or $((l^3 + 2l^2 + l)/4)$.

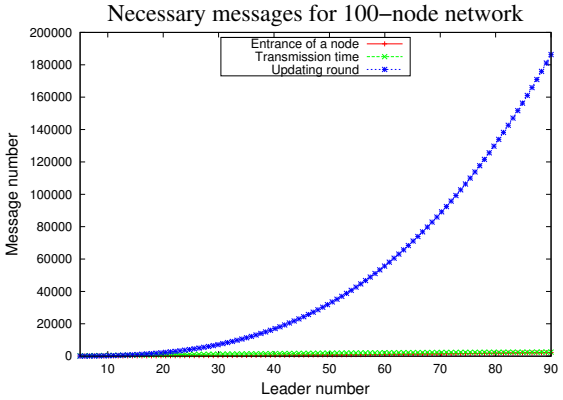


Fig. 2: Necessary messages by the number of leaders

In Figure 2, we present an example of the costs (in terms of messages) for the entrance operations of a network node, transmission time, and updating rounds in a 100-node network forming merely one cluster. With the use of communication cost information obtained, it is always possible to find more adequate values for the number of leaders in a cluster. In this example, the ratio between adequate leaders and collectors for a 100-node network

[4]The role of a new node is defined based on the need of leaders to maintain the cluster, node's connectivity and available energy.

should be approximately 20 leaders. From the graph, we note that as of 20 leaders, the ratio between the number of necessary messages for updating round operations (UR) grow exponentially. This same behavior is not noticed when the considered algorithms are transmission time (Tt) and entrance of a new node into the network. This algorithms are not as dependent on their costs in terms of messages as the updating round regarding the size assumed for the leadership.

4.2 Implementation and tests

The principle objective of the tests performed in this work was to verify the simulated limits the system would support through the Omnet++ version 4.1 simulator with a Mixim version 1.1 wireless network module. This tests were performed in a 300 x 400 meter rectangular environment. The period for transmission time was established at 60 seconds, and each epoch's time was set at 300 seconds. These amounts were obtained based on simulator tests, in such a way as to not saturate the system with messages generated during transmission times and updating rounds. The total simulation time was limited to 6010 seconds (20 URs). This total time was found to be sufficient to observe the proposed model's results. The node mobility rate in mps (meters per second) was defined at 2.0 mps. These values are similar to those employed in [12] where tests with MANETs were performed.

The malicious activity was defined considering that nodes with malicious behavior fulfill their functions in randomly routing messages. In other words, sometimes they transmit, other times they omit messages in routing. This behavior in our simulations follows a uniform distribution, in which 80% of the cases of the messages are discarded by the malicious nodes. In the other 20%, they participated in routing correctly. We chose this type of behavior as it is more difficult to detect than simply a node which discards all the messages it receives, or than nodes which retransmit only to similar (routing messages merely to a list of nodes which possess the same behavior pattern).

Table 2: Results of tests

Observed Feature	0% mal.	10% mal.	20% mal.	30% mal.	20% surp. f	30% surp. f
Loss of messages	6.52%	13.56%	20.63%	29.67%	29.69%	35.43%
Disseminated msg	100%	100%	100%	99.99%	99.68%	98.03%
Detection rate	100%	100%	100%	96.03%	95.08%	91.26%

In these tests we measure the rate of messages lost in communications (collector - leader) and (leader - leader). This showed us the network's behavior in terms of failures. Following that, we observe message delivery in leadership as to the use of the Disseminate() protocol. In other words, we observe whether a message is delivered within leadership when it arrives to at least one correct leader sent through its leadership knowledge of the cluster ($Leadership_j$ in Algorithm 1, lines 4-7) and after that, each leader in $Leadership_j$ fulfilling the algorithm resends the messages to its leadership knowledge (lines 8-11 of Algorithm 1). Finally, we verify the detection rate of the system's malicious nodes. In order to obtain data measured, three tests were carried out: the first 100 nodes were used (Table 2), the second and third began with 100, but then varying node entrances and departures rates (Tables 3 and 4).

In the first test, three distinct conditions were observed: without malicious nodes; 10%, 20%, and 30% of the malicious nodes respecting the $2f + 1$ leadership limit; and with 20% and 30% of the malicious nodes, but keeping the leadership set at 10% (surpassing the f limit). The results of this test are presented in Table 2. This test showed us that even with the occasioned loss of messages through network problems, node mobility, or malicious activity, the message dissemination rate among leaders remained very close to 100% (even when the f limit was surpassed). Consequently, the malicious activity detection rate was rather high. As such, even exceeding the limits supported by the presently proposed model, the system continues to work without the guarantee of message delivery.

Table 3: Results of tests without malicious nodes

Observed Feature	5% dep.	10% dep.	15% dep.	20% dep.
Loss of messages	5.0%	5.09%	14.53%	15.91%
Disseminated msg	100%	100%	93.66%	92.06%
Detection rate	100%	100%	97.13%	97%

In the second and third tests, two distinct conditions were observed, respectively, without malicious nodes and with 10% of the malicious nodes (therefore, respecting the $2f+1$ leader limit in leadership). In these tests, we adopted a fixed entrance of nodes at 10% to each updating round, but varying the departure in 5%, 10%, 15%, and 20%. With these tests, we observed the system's behavior with the node entrances and departures. The results of these tests are presented in Table 3 and Table 4.

Table 4: Results of tests with 10% of malicious nodes

Observed Feature	5% dep.	10% dep.	15% dep.	20% dep.
Loss of messages	11.94%	11.41%	20.46%	22.71%
Disseminated msg	100%	100%	93.37%	93.05%
Detection rate	100%	100%	97.07%	96.73%

Through these tests, we were able to observe that system message loss was greater with 10% malicious nodes than without any. However, the rate of messages spread throughout leadership was similar (with 10% maliciousness, and without any). We also observe that with a departure rate greater than its entrance rate, some messages were not widespread among leadership. This may be explained by the low network density over time. With such density, the nodes possess fewer connections. As a result, some messages may not arrive to any correct leader (the probability of this happening increases with reduced network density). In the detection rate factor, one observes behavior similar to message diffusion rates. In the same manner, if some messages are not spread throughout leadership, the detection system will end up erring. Finally, we judge the set of tests to be satisfactory and that it evidences the limits and effectiveness of our proposals.

5. Related Studies

Intrusion detection systems for MANETs were proposed in [2]. These systems use clusters to collaboratively detect intrusions. Each cluster possesses a leader which monitors all the traffic within its cluster. These studies have not used cryptography in message exchange, thus making it possible for various types of attacks to occur in the intrusion detection process. Nor have these systems assumed their leaders were alone in their clusters, with malicious behavior.

In another IDS [13], the node which detects suspect activity requests opinions from its neighbors concerning this suspect activity. After analyzing each neighbor's vote, the node makes a decision and informs it to the participating nodes who voted. However, this voting mechanism is vulnerable to message violation from and collusion with malicious nodes. In another study [14], a node hierarchy organizational model was developed on various levels, where the lowest level collects the data and the higher levels correlate the data sent to them. This study, to the contrary of the others cited here, permits the detection of several malicious nodes at the same time. However, the malicious nodes may only belong to the lower levels of the proposed hierarchy. In our proposal, any node may have malicious behavior, whether leaders or collectors. The only limitation in our model is that the number of malicious nodes cannot exceed f.

Studies concerning IDS for MANETs show that the majority of the systems proposed are capable of identifying few types of attacks or some routing protocol problems for these networks [4]. In our proposal we adopted a detection model based on anomalies. Thus we are able to identify and neutralize a large set of types of attacks and routing problems described in literature. Just as the architectures presented in [2], [14], our model also assumes a hierarchical stratification. In these models, the hierarchical topology introduces the idea of clusters. The majority of studies in literature do not deal with the entrance aspects, departure aspects, or node mobility within the network. In no related study were we able to find simulated test results or real environment test results. In [2] a time period was established for the network to reorganize itself, in which the leaders could be re-elected through a voting process. Merely some of the IDS presented ([15], [14]), indicate the use of cryptographic mechanisms to secure properties such as authenticity, confidentiality, and the integrity of messages exchanged between the IDS nodes.

The greatest contribution of this study, separating it from others present in literature, is the use of distributed systems concepts and dependability concepts applied to an IDS model for MANETs. The use of these concepts permitted - within certain limits - the development of a model less subject to restrictions. The proposed system is able to deal with various faulty or malicious nodes without there being interference in the network's normal behavior. IDSs normally developed for MANETs, in the literature do not have mechanisms to protect their own information and do not tolerate intrusions into its various components. The approach introduced in this paper aims to build an IDS infrastructure that tolerates malicious actions. Beyond this, our system is able to identify a large number of different attacks or variations of known attacks.

6. Conclusions and future study

In this paper, we presented our efforts to develop an IDS model for dynamic environments together with distributed algorithms that support this model. This proposal is centered on a hierarchical malicious behavior detection model for MANETs. This model follows the concepts of dynamic distributed systems, permitting the presence of various non-malicious entities. We presented the complexity of the proposed algorithms in terms of messages. The proposed model permits the correct functioning of the network while the faulty or malicious node limit is not exceeded. However, these tests showed us that even with the f limit exceeded, the system continues to function. In such a case, there is no guarantee that our algorithms always work correctly.

References

[1] D. Djenouri, L. Khelladi, and A. Badache, "A survey of security issues in mobile ad hoc and sensor networks," vol. 7, 2005, pp. 2–28.

[2] E. Ahmed, K. Samad, and W. Mahmood, "Cluster-based intrusion detection (cbid) architecture for mobile ad hoc networks," Australia, May 2006, pp. 1–11.

[3] L. Bononi and C. Tacconi, "Intrusion detection for secure clustering and routing in mobile multi-hop wireless networks," *Int. J. Inf. Secur.*, vol. 6, no. 6, pp. 379–392, 2007.

[4] A. Nadeem and M. Howarth, "Protection of manets from a range of attacks using an intrusion detection and prevention system," *Special issue on Mobile Computing technologies of Telecommunication System Journal*, pp. 1–12, 2011.

[5] V. Shoup, "Practical threshold signatures." Springer-Verlag, 1999, pp. 207–220.

[6] F. C. Pereira, J. da Silva Fraga, and R. F. Custódio, "Self-adaptable and intrusion tolerant certificate authority for mobile ad hoc networks," in *AINA)*, 2008, pp. 705–712.

[7] D. B. Johnson, D. A. Maltz, and J. Broch, *DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks*. Addison-Wesley, 2001.

[8] P. M. Mafra, V. Moll, J. da Silva Fraga, and A. O. Santin, "Octopus-iids: An anomaly based intelligent intrusion detection system," in *ISCC*, Italy, 2010, pp. 405–410.

[9] T. Chandra and S. Toueg, "Unreliable failure detectors for reliable distributed systems," *J. ACM*, vol. 43 2, pp. 225–267, 1996.

[10] V. Jacobson, "Congestion avoidance and control," *SIGCOMM Comput. Commun. Rev.*, vol. 18, no. 4, pp. 314–329, Aug. 1988. [Online]. Available: http://doi.acm.org/10.1145/52325.52356

[11] A. Vora, M. Nesterenko, S. Tixeuil, and S. Delaët, "Universe detectors for sybil defense in ad hoc wireless networks," *CoRR*, vol. abs/0805.0087, 2008.

[12] J.-H. Böse, "Atomic transaction processing in mobile ad-hoc networks," Master's thesis, Freie Universität Berlin, 2009.

[13] S. A. Razak, S. M. Furnell, N. L. Clarke, and P. J. Brooke, "Friend-assisted intrusion detection and response mechanisms for mobile ad hoc networks," *Ad Hoc Netw.*, vol. 6, no. 7, pp. 1151–1167, 2008.

[14] D. Sterne and G. Lawler, "A dynamic intrusion detection hierarchy for manets," in *Sarnoff Symposium, 2010. SARNOFF '10. IEEE*, april 2010, pp. 1 –8.

[15] Y. Zhang, W. Lee, and Y. Huang, "Intrusion detection techniques for mobile wireless networks," *ACM/Kluwer Wireless Networks Journal*, vol. 9, no. 5, September 2003.

Network Platform for Location-Based Network Applications in MANETs

Hiroaki Higaki
Department of Robotics and Mechatronics
Tokyo Denki University
+81-3-5284-5606
Senju-Asahi 5, Adachi, Tokyo 1208551 Japan
Email: hig@higlab.net

***Abstract*—Until now, network applications are developed based on the TCP/IP network model which is intrinsically based on stationary wired networks. Here, the socket interface is the most widely used API (application program interface). However, due to widely available mobile wireless computers, location-based network services are required to be developed, e.g. location-based advertisement and query for sensor data. In order to support development such applications, novel API and network platform are required. This paper proposes an API for location-based network applications by which the destination of transmission is specified by its location and a network platform supporting such applications on ad-hoc networks adopting a location-based routing protocol.**

I. Introduction

Due to widely available wireless communications and mobile computing technologies, wireless multihop networking becomes to be expected. MANETs (Mobile Ad-Hoc Networks) provides higher availability supported by location-transparent connectivity based on wireless multihop communication with ad-hoc routing in spite of higher mobility of wireless nodes. In addition, node identification is also important to realize communication among nodes. As various mobile wireless nodes gets popular, various network services are developed. Especially, according to location acquisition technologies, location-based network applications and services becomes attractive. Here, data messages are required to be transmitted to wireless nodes specified by their locations. However, currently available network platforms do not well support such location-based applications since they are designed for conventional wired networks with stationary nodes. This paper proposes a novel network platform dedicated to location-based applications in MANETs. By combination of location-based ad-hoc routing and application program interface (API) for location-based network applications by which destinations of data messages are specified by their locations, not only performance of application development but also execution performance of location-based applications in MANETs are expected to be improved.

II. Related Works

In the conventional TCP/IP Internet technology based on wired networks, nodes, i.e. computers connected to the Internet such as servers and clients (terminals), are identified by their IP addresses intrinsically independent of their geographical locations. For data message transmissions, their destinations are specified by the IP addresses and their routing protocols such as OSPF, RIP and BGP are designed to best fit to treat the IP addresses. The most widely used application program interface to TCP/IP Internet communication is the socket interface [3]. In order to describe data message transmission between nodes in an application program, a socket is created by an primitive socket() which returns a socket identifier socket_id. Sending of a data message is described by a send() primitive whose parameters are the socket identifier, an IP address of a destination node and a transmitted message; i.e., send(socket_id, destination_address, message). On the other hand, receipt of a data message is described by a receive() primitive whose parameters are socket identifier, an IP address of a source node and a received message; i.e., receive(socket_id, source_address, message).

In wireless multihop networks such as MANETs, wireless sensor networks and mesh networks, data messages are transmitted by using wireless multihop transmissions. Data messages are transmitted from a source wireless node to a destination one along a wireless multihop transmission route consisting of a sequence of intermediate wireless nodes which forward the data messages. For wireless multihop transmissions, various ad-hoc routing protocols such as AODV, DSR, TORA and OLSR have been designed [10]. In these protocols, each wireless node is specified and identified by its identifier. Hence, source and destination nodes are required to be specified by their identifier, e.g. their IP addresses, and the socket interface can be used in description of application programs.

Not only fundamental but also the most important property of mobile wireless nodes in MANETs is their locations which are changed by their mobility. Recently, location-based network services are widely discussed and are expected to provide their customers novel value. For example, location-based advertisement of shops and stores are expected to achieve higher effectiveness against its cost since explicit and implicit (potential) uses and customers are tend to be geographically maldistributed. Hence, data messages containing their sales information are required to be destined to some specific locations (areas or ranges). In sensor networks, gathering sensing data achieved within a specific area are often required. Hence, a query message is transmitted from a user terminal to wireless sensor nodes in the specified area. As shown in these examples, messages are required to be transmitted to destination wireless nodes specified not by their identifier but by their geographical locations. That is, GEOCAST [8] is

required to be available in MANETs supporting location-based services.

Different from the above mentioned ad-hoc routing protocols which are based on flooding of route request control messages or continuous exchange of control messages to maintain routing tables based on up-to-date network topology, location-based ad-hoc routing protocols have been developed and improved. GEDIR [6] and Compass [5] are greedy location-based ad-hoc routing which may achieve shortest wireless multihop transmission route though they might cause dead-ends. Face [1] and GPSR [4] are guaranteed delivery ad-hoc routing which are dead-end-free but may suffer longer transmission delay due to detour wireless multihop transmission route. In these location-based ad-hoc routing protocols, it is required for each intermediate wireless node to achieve the location of the destination wireless node for determination of its next-hop wireless node. In other words, these protocols can be applied to transmit data messages to wireless nodes in a specified geographical area.

III. Proposal

It is difficult for combination of conventional ad-hoc routing protocols based on flooding of route request control messages and the socket application program interface designed for stable networks, e.g. wired networks, composed of stationary nodes to support location-based services in mobile wireless networks. Although a set of destination wireless nodes are required to be specified by an area where they are included, the socket application interface requires to specify the destination nodes by their IP addresses. Hence, a translator (or a resolver) to achieve an IP address from location of an area is required. For mobile ad-hoc networks, various location services such as HRLI [7], DREAM [2] and ABLA [9] have been proposed. Most of these location services are mainly designed to achieve location information of mobile wireless nodes by their identifiers such as IP addresses and the reverse translation is not explicitly supported. Since HRLI and some others are fully or partially centralized services, i.e. one or multiple location servers provide the translation service, it may be possible to provide the reverse translation. However, since DREAM and ABLA provides the translation in fully distributed manner, it is difficult to provide the reverse translation with reasonable communication overhead. After achieving a set of IP addresses of the destination mobile wireless nodes in the specified area, data messages are required to be transmitted through the send() primitive invoked multiple times, i.e. transmitted to the destination nodes one by one in unicast. It requires redundant transmission of copies of the data messages since most of the intermediate mobile wireless nodes are shared by all the destination nodes located in the specified area. Moreover, in order to route a copy of data message to each destination node, a route detection procedure based on flooding is required in case of no cash entry in a source and/or intermediate nodes.

For avoidance of the communication overhead for flooding, introduction of the location-based ad-hoc routing protocols is effective. However, in an application program, location information of each destination mobile wireless node is required to be treated explicitly. That is, both location to identifier translation for invocation of the send() primitive and identifier to location translation for routing are required, which is obviously redundant.

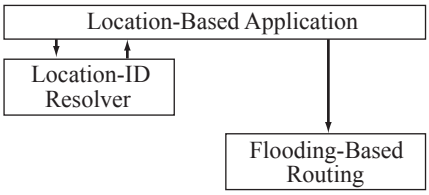


Fig. 1. Naive Platform Architecture in Conventional Method.

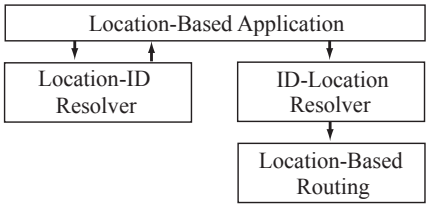


Fig. 2. Redundant Address Resolution in Conventional Method.

In order to solve this problem, this paper proposes a network platform for location-based network applications in MANETs. Here, data messages are routed by location-based ad-hoc routing protocols and a set of destination mobile wireless nodes are specified by an area; i.e. a location information.

A. Platform Architecture

In order to avoid the redundant translation between location information and identifier of destination mobile wireless nodes for addressing and routing to support location-based services, this paper proposes a wireless network platform in which location information is used in both addressing and routing and no translators (resolvers) are introduced. In an application program of a source wireless node of data messages, the destination is specified by an area determined its location information; i.e. longitude and latitude. The specified location information is directly transferred to the location-based routing for data messages without any translation or resolution. Data messages are routed by using the specified location.

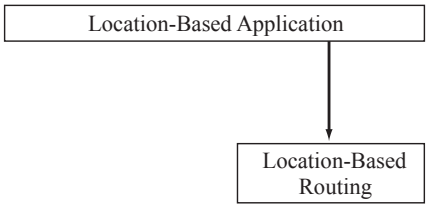


Fig. 3. Proposal of Platform Architecture for Location-Based Services.

As discussed in the next subsection, the destination is specified not by a point of location but by an area (or a region). Thus, there may be one wireless mobile node or

multiple wireless mobile nodes, or even no wireless mobile nodes in the area. Anyway, the data messages are required to be transmitted to the area; however, most of the location-based ad-hoc routing protocols require to be provided the location information of the destination location which is a point strictly though the destination is specified by an area in an application program. This problem is easily solved by introduction of the representative point of the area; e.g. the center of gravity of the area as shown in Figure 4. Now, without modification, the data messages are expected to reach one of the mobile wireless nodes in the destination area if it exists by using the location-based ad-hoc routing protocols.

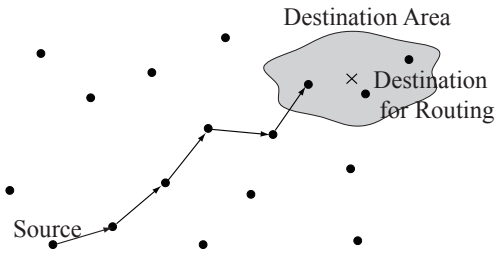


Fig. 4. Data Message Transmission to Representative Point in Area.

B. Application Program Interface

Since the proposed platform architecture requires no address resolution for data message transmissions in location-based services, required application program interface consists of only send() and receive() primitives. However, as discussed in the previous subsection, since the destination is specified not by identifiers of destination mobile wireless nodes but by location information of an area including the destination nodes, the area may contain multiple mobile wireless nodes. Hence, a location-based application program is required to specify that the data message is transmitted to ALL the mobile wireless nodes in the area or ANY one of them. That is, request for broadcast (ALL) or anycast (ANY) should be described in an application program.

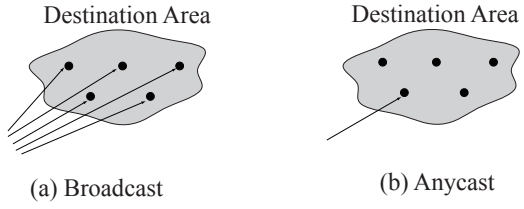


Fig. 5. Broadcast and Anycast.

There are various location-based services and they are classified into the following categories: oneway services and query-reply (client-server) services. In the former, only oneway transmissions of data messages are required as shown in Figure 6(a). Advertisement of some location-based information to nodes in the specified area is the representative service. On the other hand in the latter in Figure 6(b), a query message (or copies of the query messages) are transmitted to mobile wireless nodes in the specified area same as in the oneway transmissions. However, for transmissions of a reply message (or multiple reply messages), its destination should be specified by the identifier of the source node of the query message. Thus, an application program in the receiver mobile wireless node requires to achieve the identifier of the source node and to specify the destination of its reply message by the identifier. According to the considerations, we design

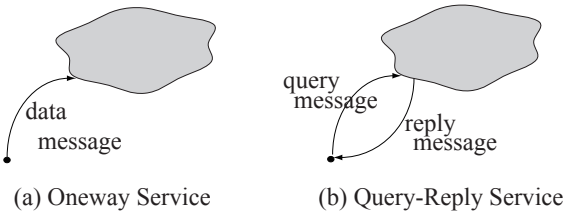


Fig. 6. Typical Communication in Location-Based Services.

the following application program interface; i.e. send() and receive() primitives.

[Send Primitive]

Send (destination_area, destination_ID, Delivery_Type, message) where destination_area is location information of destination, destination_ID is destination ID for reply message to which NULL is assigned in a oneway data message and a query message and Delivery_Type for multicast (ALL) or anycast (ANY).

- destination_area: location information of destination.
- destination_ID: destination ID of reply message.
- Delivery_Type: multicast (ALL) or anycast (ANY).

[Receive Primitive]

Receive (source_location, source_ID, message) where source_location is location information of source node and source_ID is source ID for reply message.

- source_location: location information of source.
- source_ID: source ID for reply message.

In order to available these primitives to application programs, the platform (library functions) embeds the current location information and identifier of the source node into the header in a data message.

C. Implementation

Data message transmissions specified by our proposed application program interface is implemented in the network platform based on the location-based ad-hoc routing.

For oneway data message transmission, the destination_area is specified but the destination_ID is NULL. If Delivery_Type is ANY, data message is transmitted to one of the mobile wireless nodes in the destination area. As discussed in subsection 3.1, data messages are routed as if the representative point in the destination area is the destination point though there may be no mobile wireless nodes at that point. When the data message is routed to one of the wireless nodes in the area, transmission procedure terminates. On the other hand, if Delivery_Type is ALL, copy of data message are propagated in the destination area by restricted flooding.

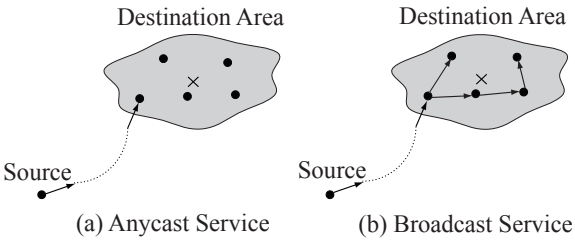


Fig. 7. Oneway Services.

For query-reply data message transmission, a query message is transmitted same as oneway data message transmission. However, a reply message is required to be transmitted differently from the query message. Since the destination of the reply message is specified not by its location but by its identifier, destination_ID in the send() primitive should be specified by the source node ID of the query message achieved through the receive() primitive. However, the underlying wireless multihop network is based on location-based routing protocol, destination_area is also required to be specified. Here, to reach the reply message to the destination node, the destination_area should include the current location of the destination node. In addition, independently of Delivery_Type, if destination_ID is specified, the data message is transmitted to all mobile wireless nodes in the specified area same as in the case that Delivery_Type is ALL. In mobile wireless nodes receiving this reply message, according to the result of comparison between its own identifier and the identifier specified in the message header, it only receives the message only in the case that these identifiers are the same.

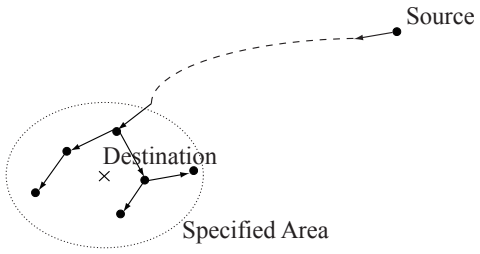


Fig. 8. Reply Transmission in Query-Reply Services.

IV. Evaluation

In order to evaluate the effect of the introduction of the proposed location-based network application platform, the authors develop two simple application programs. One is for location-based advertisement applications and the other is for location-based query-reply applications. In the former, a data message is distributed to all the mobile wireless nodes in the dedicated area which is different from the widely available multicast services where destination nodes are determined by a multicast address. In the latter, a query message is transmitted to a server node which is in the dedicated area, i.e., the server node is specified not by its address but by its location, and a reply message is transmitted to a sender client node. In both programs in the conventional method, translation from location information to an IP address is required and should be explicitly described in programs as a function call get_address(location_information). The function returns a list of IP addresses of mobile wireless nodes which is in the specified area. In the advertisement application programs, a data message is required to be transmitted to all the nodes by repetition of a message sending function call which implement unicast transmission of a data message. In the query-reply application programs, a query message is also required to be transmitted to all the nodes in the area since it is impossible to detect an IP address of a server node. On the other hand, in our proposed location-based wireless network platform, only one message sending primitive is required to be called since a data message is transmitted to all the wireless nodes in the specified area due to the underlying location-based platform including the proposed API and the location-based ad-hoc routing protocols.

Table I shows the results of the evaluation. In both advertisement and query-reply application programs, our proposed API and platform reduces the numbers of lines in the programs, i.e., it is expected that shorter development period and higher quality are provided by our proposed method than the conventional one.

TABLE I
Numbers of Lines in Sample Application Programs

	Proposed	Conventional	Reduction
Advertisement	57	115	50.4%
Query-Reply	73	231	78.4%

V. Conclusion

This paper has proposed a wireless multihop network platform and an application program interface for location-based services based on location-based ad-hoc routing protocol. Here, destination of data messages are directly specified by the destination location information in application programs. The application program interface and the network platform supports both oneway transmissions as for advertisement and query-reply transmissions as for sensor data retrieval. Since both of them are location-based, no address resolution mechanism is needed and avoidance of loss of performance is expected.

References

[1] Bose, P., Morin, P., Stojmenovic, I. and Urrutia, J., "Routing with Guaranteed Delivery in Ad Hoc Wireless Networks," Wireless Networks, vol. 7, pp. 609–616 (2001).
[2] Basagni, S., Chlamtac, I. and Syrotiuk, V.R., "A Distance Routing Effect Algorithm for Mobility (DREAM)," Proceedings of the 4th ACM International Conference on Mobile Computing and Networking, pp. 76–84 (1998).
[3] Comer, D.E., "Internetworking with TCP/IP," Prentice Hall (1991).

[4] Karp, B. and Kung, H.T., "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," Proceedings of the 6th ACM International Conference on Mobile Computing and Networking, pp. 243–254 (2000).
[5] Kranakis, E., Singh, H. and Urrutia, J., "Compass Routing on Geometric Networks," Proceedings of the 11th Canadian Conference on Computational Geometry, pp. 51–54 (1999).
[6] Lin, X. and Stojmenovic, I., "Geographic Distance Routing in Ad Hoc Wireless Networks," Technical Report in University Ottawa, TR-98-10 (1998).
[7] Nakagawa, H., Ohta, T., Ishida, K. and Kakuda, Y., "A Hybrid Routing with Location Information for Mobile Ad Hoc Networks," Proceedings of the 8th IEEE International Symposium on Autonomous Decentralized Systems, pp. 129–136 (2007).
[8] Navas, J.C. and Imielinski, T., "GeoCast – Geographic Addressing and Routing," Proceedings of the 3rd ACM International Conference on Mobile Computing and Networking, pp. 66–76 (1997).
[9] Oneda, R. and Higaki, H., "Lower Overhead Location Advertisement in Mobile Wireless Multihop Networks," Proceedings of the 22nd International Conference on Parallel and Distributed Computing and Systems, pp. 81–87 (2010).
[10] Perkins, C.E., "Ad Hoc Networking," Addison-Wesley (2000).

Ad-Hoc Cross Layered Energy based on-demand Routing Protocol for MANETs

K.Muthumayil[1], V.Rajamani[2], S.Manikandan[3]
[1]PSNA College of Engg. & Tech.,Dindigul, Tamilnadu, India
[2]Indra Ganesan College of Engg., Trichy, Tamilnadu, India
[3]R.M.D Engg.College, Gummudipoondi, Chennai, Tamilnadu, India

Abstract— **The design of efficient routing protocols for Ad hoc networks is a complex issue. These networks need efficient algorithms to determine ad hoc connectivity and routing. MANET aims not only to provide correct and efficient routes between pair of nodes but also to provide energy efficient route to maximize the life time of ad hoc mobile networks. In this paper, a dynamic energy conscious routing algorithm ACER where cross layer interaction is provided to utilize the energy related information from physical and MAC layers. This algorithm avoids the nodes which are having low residual energy. By maximizing the lifetime of mobile nodes routing algorithm selects a best path from the viewpoint of high residual energy path as part of route stability. The RTS/CTS transmission is a crucial step towards saving the energy of mobile nodes. In this scheme, the RTS/CTS transmission occurs after route discovery and route reply process but before the data transmission. BER at the end of a multi-hop route is calculated using the interference and noise level from physical layer. Routing is made after these information are collected from MAC layer. In this protocol, transmitter power is calculated after the initial transmissions based on the receiver power. The protocol is implemented for achieving quality of service (QoS) in terms of average energy consumption, packet delivery ratio, end-to-end delay and throughput.**

Index Terms— **Routing protocols, Ad hoc networks, Cross layer design, Quality of Service**

1. Introduction

Mobile Ad Hoc Network (MANET) is collection of multi-hop wireless mobile nodes that communicate with each other without centralized control or established infrastructure. In MANET each node communicates with other nodes directly or indirectly through intermediate nodes. Thus, all nodes in a MANET basically function as mobile routers participating in some routing protocol required for deciding and maintaining the routes. Routing[1],[2],[3],[4],[5] is one of the key issues in MANETs due to their highly dynamic and distributed nature. The routing protocols of MANETs are divided into two categories as table-driven and on-demand. In table-driven routing protocols, each node attempts to maintain consistent, up-to-date routing information to every other node in the network. Many routing protocols including Destination-Sequenced Distance Vector (DSDV) [1] and Fisheye State Routing (FSR) protocol belong to this category. In on-demand routing, routes are created as and when required. Route discovery and route maintenance are two main procedures: The route discovery process involves sending route-request packets from a source to its neighbor nodes, which then forward the request to their neighbors, and so on. Once the route-request reaches the destination node, it responds by unicasting a route-reply packet back to the source node via the neighbor from which it first received the route-request. When the route-request reaches an intermediate node that has a sufficiently up-to-date route, it stops forwarding and sends a route-reply message back to the source. Once the route is established, the route maintenance process is invoked until the destination becomes inaccessible along the route. Note that each node learns the routing path as time passes not only as a source or an intermediate node but also as an overhearing neighbor node. In contrast to table-driven routing protocols, on-demand routing protocols don't maintain all up-to-date routes. Dynamic Source Routing (DSR) [3] and Ad-Hoc On-Demand Distance Vector (AODV) [2],[4],[5] are popular on-demand routing protocols.

In addition to simply establishing correct and efficient routes between pair of nodes, one important goal of a routing protocol is to maximize the lifetime of ad hoc mobile networks. The residual battery energy of mobile nodes is a simple indication of energy stability and can be used to extend network lifetime [6],[7],[8],[9],[10]. This information has to be taken from the physical and medium access control layers of data link layer since these layers are responsible layers to compute the power consumption and residual energy computation. Many MAC layer protocol [11],[12]has been discussed previously.

The proposed Ad-Hoc Cross layered Energy based Routing (ACER) protocol prefers the wireless link requiring minimum transmitter power, but at the same time avoids the node with low residual energy. Hence, routes requiring required minimum transmitter power and high residual energy are preferred. The power related information are acquired from both physical and MAC layers. MAC layer functions are modified to provide the RTS/CTS packets after the route is discovered. In this protocol, the bit error rate at the end of a multi-hop route is calculated using the signal-to-noise ratio.

This paper is further organized into three sections. In section2, the energy efficient algorithms have been discussed. Section3 contains the proposed scheme that includes basic assumptions, four phases of the newly developed protocol such as route discovery, route reply, energy conservation and route repair. In section 4, the performance analysis has been given. Performance analysis is done using ns-2 simulator on the following parameters: average energy consumption, packet delivery ratio, delay and throughput.

2. Energy efficient routing algorithms

Rekha Patil and A.Damodaram [13] developed a routing protocol based on MAC information. The discovery mechanism in this algorithm uses battery capacity of a node as a routing metric. This approach is based on intermediate nodes calculating cost based on battery capacity. The intermediate node judges its ability to forward the RREQ packets or drop it. That is it integrates the routing decision of network layer with battery capacity estimation of MAC layer. Ivan Stojmenovic and Xu Lin [14] developed a new power cost-metric based on the combination of both node's life time and distance based power metrics. This provides basis for power, cost and power-cost localized routing algorithms where nodes make routing decisions solely based on the location of their neighbors and destination. The power aware routing algorithm attempts to minimize the total power needed to route a packet between source and destination. The cost-aware routing algorithm is aimed at extending the battery's worst-case lifetime at each node. The combined power-cost routing algorithm attempts to minimize the total power needed and to avoid nodes with a short remaining battery life time.

Power-aware Source Routing (PSR) discussed by Morteza Maleki, Karthik Dantu, and Massoud Pedram [15] is to extend the useful service life of a MANET. This is highly desirable in wireless ad hoc network since death of certain nodes leads to a possibility of network partitions, rendering other live nodes unreachable. This algorithm assumes that all nodes start with a finite amount of battery capacity and that the energy dissipation per bit of data and control packet transmission or reception is known and presents a new source-initiated (on-demand) routing protocol for mobile ad hoc networks that increases the network lifetime. Multicast Multi-path Power Efficient Routing by S.Gunasekaran and K.Duraiswamy [16] addresses the problem of power awareness routing to increase lifetime of total network. Since nodes in mobile ad hoc network can move randomly, the topology may change arbitrarily and frequently at unpredictable times. Transmission and reception parameters may also impact the topology. Therefore it is very difficult to find and maintain an optimal power aware route. A scheme has been proposed to maximize the network lifetime and minimizes the power consumption during the source to destination route establishment. This scheme is aimed to provide efficient power aware routing considering real and non real time data transfer.

Chansu Yu, Ben Lee and Hee Yong Youn [17] overviews on energy efficient routing approaches such as transmission power control approach, load distribution approach, sleep/power-down mode approach. For transmission power optimization, Flow Augmentation Routing (FAR) , Online Max-Min Routing (OMM), Power aware Localized Routing (PLR) protocols and minimum energy routing(MER) were discussed. For load Distribution Approach Localized Energy-Aware Routing (LEAR) and Conditional Max-Min Battery Capacity Routing (CMMBR) protocols were discussed. For sleep/power-down mode approach, SPAN protocol and the Geographic Adaptive Fidelity (GAF) protocol employ the master-slave architecture and put slave nodes in low power states to save energy. Unlike SPAN and GAF, Prototype Embedded Network (PEN) protocol saves more energy when the devices put into sleep state according to the need.

Many energy-aware routing protocols such as MREP, MLRP, HEAP [18] protocols were designed by considering node's remaining energy and/or link transmission power to prolong network lifetime through balancing energy draining among nodes. Here the subset of network nodes required to involve in a route searching process (measured in node's remaining energy) or the degree (measured in transmission power) to which intermediate nodes are required to participate in searching for such a path. V. Kanakaris, D. Ndzi and D. Azzi [19] overview on various routing protocols such as AODV, DSDV, DSR and TORA for their power consumption against mobility and concluded that TORA routing protocols needs more performance when compared to other protocols.

To cater various challenges in QoS routing in Mobile Ad hoc Networks, a Node Disjoint Multipath Routing considering Link and Node Stability (NDMLNR) protocol has been proposed by Shuchita Upadhayaya and Charu Gandhi [20]. In this scheme, the metric used to select the paths takes into account the stability of the nodes and the corresponding links.

3. Proposed scheme

In communication-related tasks, energy consumption depends on the communication mode of a node. A node may either in a mode of transmit, receive or idle. Transmission consumes more energy than the other two modes. Here directional antennas are used for better power consumption. It consumes power in a single direction between a sender-receiver pair. Here the transmission power of the sender is adjusted based on the receiver power for every link in the MANET. Once the receiver receives data it calculates

minimum receiving power of it. This power information is sent back to the sender. Then the sender alters its transmission power. It means that the sender have to send the further data using this power information. The nodes which have required minimum transmission power and the nodes with high remaining battery power is considered for stability. In MANET, there is high power consumption for sending RTS and CTS signals. Here the RTS/CTS handshaking happens after the route discovery but before the data transmission. For sending the RTS signal, the node has to wait for 5ms and then data transmission occurs. The proposed scheme ACER routing protocol is developed by using AODV as the base.

Basic Assumptions

All the nodes in the given area have same transmit power and each node selects a threshold energy level (E_{th}) and each node must maintain the value in their routing table to select the nodes during route discovery. The residual battery energy value can be obtained to the network layer where it is stored in the routing tables to make routing decisions based on the battery energy. When the residual energy is less than threshold energy value, that node is avoided in the route selection by the destination. On receiving the RREQ the intermediate nodes calculates the received signal (r(t)) strength which holds the following relationship for two-ray propagation model:

$$P_R = P_T G_T G_R \left(\frac{h_t h_r}{d^2} \right) 2 \quad (1)$$

Where P_T and P_R are transmitter power and receiver power respectively, λ is the carrier wavelength, d is the distance between the sender and the receiver and G_T G_R are the unity gain of the transmitting and receiving antenna respectively. Hence the node calculates the path loss using

$$Path\ loss = P_T - P_R \quad (2)$$

The main impact of physical layer affecting wireless ad hoc networks as perceived by the receiver is the degradation of the received signal strength due to free space loss. For every link, we have to calculate the SNR based on the received signal. The received signal r(t) can be written as

$$r(t) = f(t)\, m(t) + A(t) \quad (3)$$

Where m(t) is attenuated version of the transmitted signal, f(t) is the fading process and A(t) is an AWGN process.
The bit energy can be written as

$$Eb = P_r / Rb \quad (4)$$

The receiver sensitivity, the minimum received power necessary for a signal to be correctly detected is, P_{Rmin} as from (1). The receiver strength is the only one parameter which decides the correct reception of signals. The sender uses this P_{Rmin} for further transmissions.

The total amount of energy consumed per transmitted packet is written as

$$E_t = P_T * L/R_b \quad (5)$$

where

E_t = transmitted energy
P_T = transmitter power
L = packet length
R_b = data rate or bandwidth

The total amount of energy consumed per received packet is written as

$$E_r = P_{R\,min} * L/R_b \quad (6)$$

In general, the fading process f(t) can be written as

$$f(t) = a(t)\, e^{j\theta(t)} \quad (7)$$

a(t) = fading amplitude process
θ(t) =fading phase process
Assume X_N, here N is the number of nodes and the communication is taken place from X_i to $X_j$and i,j⊂N
Denote *a* is the fading amplitude, then the instantaneous link SNR can be written as

$$SNR_{x_i, x_j} = \frac{E[a^2] E_b}{A(t)}$$

$$= \frac{E[a^2] P_r}{A(t) R_b}$$

$$SNR_{x_i, x_j} = \frac{E[a^2] P_t G_t G_r (h_t h_r)^2}{A(t) (d^2)^2 R_b} \quad (8)$$

Then the bit error rate (BER) is calculated based on the SNR of the receiving node and modulation scheme. Assume that the channel fading is not presented in the case of strong LOS (the transmission is held over an AWGN channel), the link BER can be written as

$$BER_{x_i, x_j} = Q(\sqrt{2SNR_{link}})$$

$$= Q\sqrt{2E_b / A(t)}$$

$$= Q\sqrt{2P_r / A(t) R_b}$$

$$BER_{x_i, x_j} = Q\sqrt{2P_t G_t G_r (h_t h_r)^2 / A(t) R_b (d^2)^2} \quad (9)$$

At the end of n number of links, the route BER can be written as

$$BER^n_{route} = 1 - (1 - BER_l)^n \quad (10)$$

The node then calculates the residual energy E_{res} using the following parameters:

E_I – Initial energy taken by the node
E_t – Energy consumed in transmitting packets
E_r – Energy consumed in receiving packets
E_i – Energy consumption in IDLE state.

$$E_{res} = E_T - (E_t + E_r + E_i) \quad (11)$$

The node then piggybacks this residual energy along with the required minimum receiving power in the RREQ packet.

Using this method every node will forward the RREQ till it reaches the destination. The selection of best route is based on the remaining battery energy and transmission power of all intermediate nodes. The route with maximum remaining energy and minimum receiving power is selected by destination node.

Our proposed Ad-Hoc Cross Layered Energy based on-demand Routing Protocol consists of four phases:

A. Route discovery
B. Route reply
C. Route repair

Initially when a source node wants a route, "Route Discovery" phase is invoked. After source sends the RREQs, the destination receives the RREQ from all paths and route selection is done based on the high residual energy, required minimum receiving power and it replies with RREP through the best path selected. Then "Energy conservation" is invoked from MAC layer. If a route is broken, "Route Repair" is invoked to repair and establish a new route based on the criteria.

3.1. Route Discovery

In AODV, the proposed routing protocol modifies the route discovery procedure for balanced energy consumption. When the source node wants to send data packets to a destination node D and does not have a route to D, it initiates route discovery by broadcasting a route request RREQ (broadcast message) to its neighbors. RREQ packet length is about 40 bytes. Once a source node initiates the route discovery process it includes its required minimum receiving power and residual battery energy in the route request message just enough to reach to its neighbor nodes. In AODV, when a node receives a route-request message, it checks its residual energy (E_r) with threshold value (E_{th}).When E_r is higher than E_{th}, it appends its identifier and residual energy in the message and forwards it toward the destination. Thus, an intermediate node always relay messages if the corresponding route is selected. However, in proposed algorithm, a node determines whether to forward the route-request message or not depending on its residual battery energy (E_{res}). When E_{res} is higher than a threshold value (E_{th}), the node forwards the route-request message including its battery level; otherwise, it drops the message and refuses to participate in relaying packets. The node also calculates the required minimum receiving power by formula (1) and computes the signal strength r(t), SNR,BER and includes it's receiving power with the RREQ packet. Therefore, destination node will receive a route-request message only when all intermediate nodes along a route have good battery levels, and nodes with low battery levels can conserve their battery power.

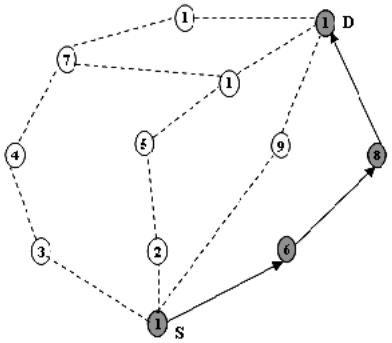


Fig 1. Route discovery process

In AODV protocol, when the route-request reaches an intermediate node that has a sufficiently active route, it stops forwarding and sends a route-reply message back to the source. In figure 1, the source node 1 sends route request broadcast message the path 1-6-8-12 has been selected as a stable path since it has nodes which have high residual energy and minimum transmission energy. Here the shortest path is 1-9-12 but it has less stability because of residual power and transmission power.

The primary objective of this Ad-Hoc Cross Layered Energy based on-demand Routing Protocol is to maintain a connected topology using the minimal receiving power and residual energy of the node based on their battery power so as to minimize the receiving power and maximize the lifetime of the node. Energy efficient routing protocols based on receiving power control find the best route that optimizes the total receiving power between a source-destination pair.

3.2. Route Reply

In AODV protocol, once the source node wants a data packet to be sent it initiates the route discovery process and waits for route reply and when it gets route then it transmits the data. However in this proposed protocol, the route selection is concerned with the remaining battery power (residual energy) and required minimum receiving power.
Among the various route request messages from the intermediate nodes, the destination node selects the best path having high residual energy and minimum and optimized receiving power as included in the route request messages. This path is more energy efficient as it involves optimized receiving power and nodes with maximum lifetime stability. If the route reply (RREP) packet has not come back to the source before the expiration of T_{SIFS}, the source will retransmit the same RREQ packet as broadcast packet. When the same intermediate nodes receive this same RREQ packet, first it checks its table to search for the particular RREP. It it has RREP, it is sent back to the source and the intermediate nodes simply drop because of same RREQ identifier.

3.3. Route Repair

After the destination selects the best route based on the residual battery energy and transmission energy, the source uses to transmit data packets through selected path. When a link break occurs in active route during the

transmission, route repair procedure is invoked. In this protocol, Local Repair (LR) message is sent to the source node by the upstream node of the broken link. Then the upstream node attempts to find the alternate path based on the residual energy and transmission power mentioned above.

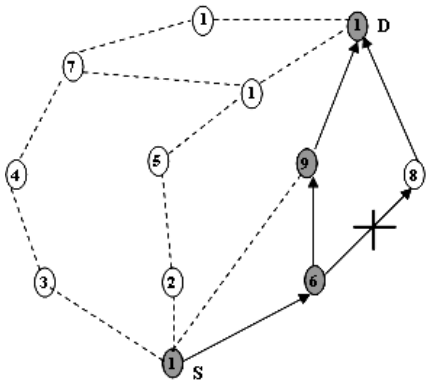


Fig 2 : Route Repair process

In figure 2, suppose the link between node 6 and node 8 is broken, the upstream node of the broken path 6 sends Local Repair (LR) back to the source. Also it attempts to find an alternate path. In this example node 6 finds the alternate path 1-6-9-12 since it has next high stability nodes based on the residual energy and minimum receiving power than all other paths.

4. Performance Evaluation

Simulation Environment

Simulation study has been carried out to show the performance of our proposed protocol. Simulation used here is NS2 (Network Simulator).Simulation results have been compared with various existing protocols like AODV and DSDV in terms of quality of service(QoS) parameters such as energy consumption and throughput.

TABLE III

SIMULATION PARAMETERS

Parameter	Value
Test Area	1500m X 1500m
Channel type	Wireless channel
Radio propagation	Two Ray Ground
Antenna type	Omni directional
Interface queue type	Drop tail with priority queue
Max.ifq length	50
Transmission range	250m
Number of nodes	20
Node separation	Half of radio range, vertically and horizontally
Transmission Bandwidth	1Mbps
MAC	IEEE 802.11 with RTS/CTS
Mobility Model	Random waypoint
Mobility Speed	0 – 10m/s
Mobility Pause Time	30s
Traffic Type	CBR, UDP
Packet size	512 bytes
Initial energy	100 Joules

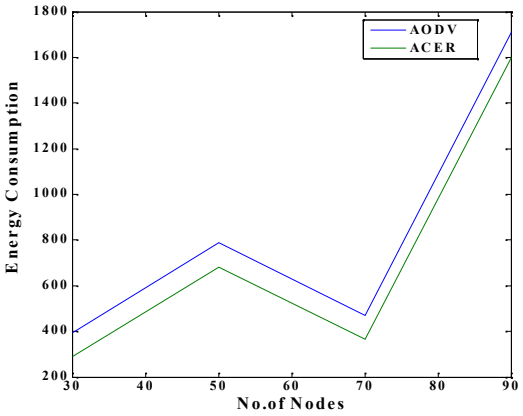


Figure 3: Energy Consumption Vs Number of Nodes (Fixed Topology)

Energy consumption of proposed protocol ACER and simulated protocol AODV is increased and decreased when the number of nodes is got increased and it is depicted in figure 1 to 4. Both topologies experiences that the proposed protocol ACER consumes less energy than AODV because the processing of packets and number of nodes at the same time load on many number of nodes consumes more energy for AODV is very large then ACER. Since ACER experiences the signal-to-noise ratio and computes bit error rate, it reduces it's energy consumption very low than AODV.

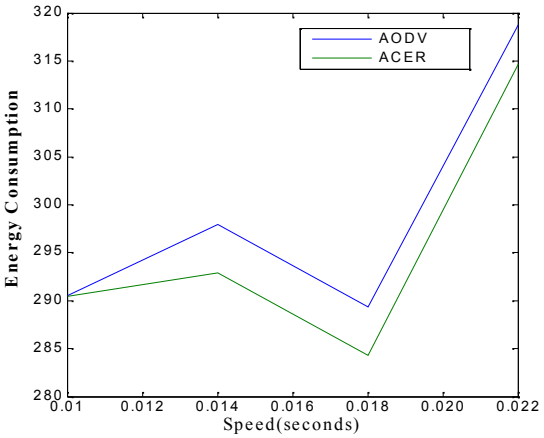


Figure 4: Energy Consumption Vs Speed (Fixed Topology)

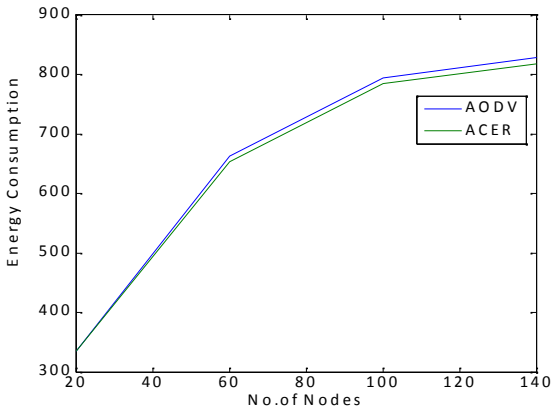


Figure 5: Energy Consumption Vs Number of Nodes (Random Topology)

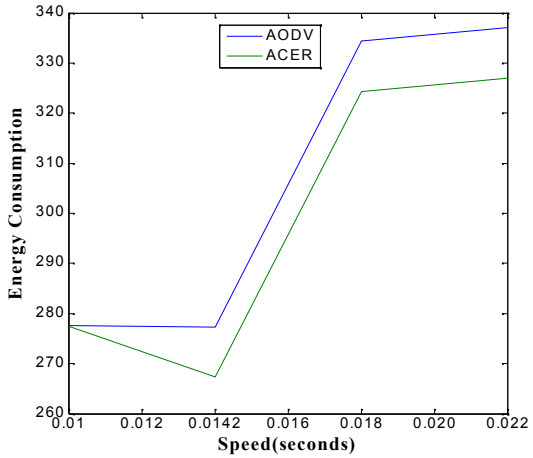


Figure 6: Energy Consumption Vs Number of Nodes (Random Topology)

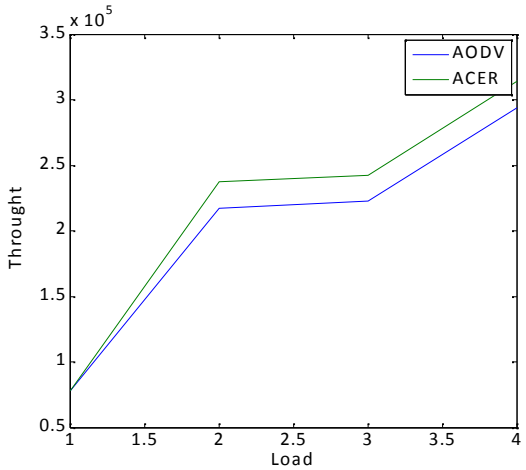


Figure 7: Load Vs Throughput (Fixed Topology)

Figure 5 to 8 depict the comparison of throughput against load and packet sized in fixed topology and random topology. Throughput is the average rate of successful delivery of packets in a communication channel.. It is the total successful transmissions within the time period from simulation starts and ends. In this scheme, the proposed protocol ACER achieves better throughput while the load in the path is increasing than AODV protocol. AODV makes no such stable route like ACER. AODV experiences heavy packet loss due to inconvenience in stable route. AODV suffers from heavy load. If there is high load in the path, AODV achieves only 85% successful delivery of packets. But our proposed protocol ACER achieves 97% of successful packet delivery eventhough there is high load.

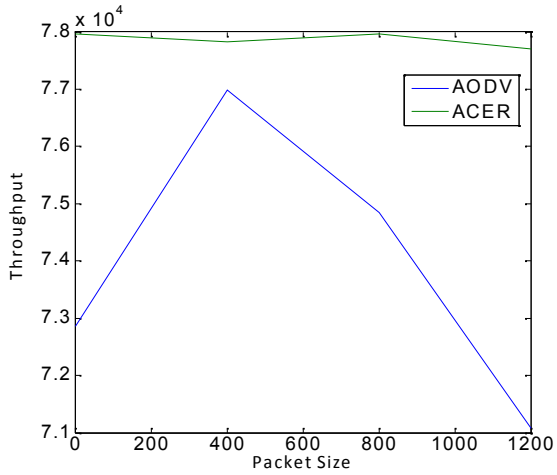


Figure 8: Packet Size Vs Throughput (Fixed Topology)

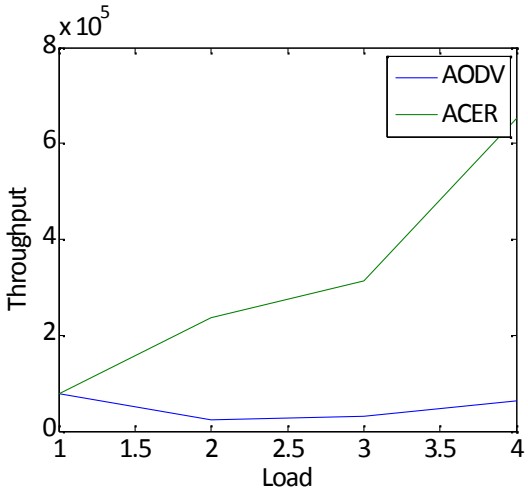


Figure 9: Load Vs Throughput (Random Topology)

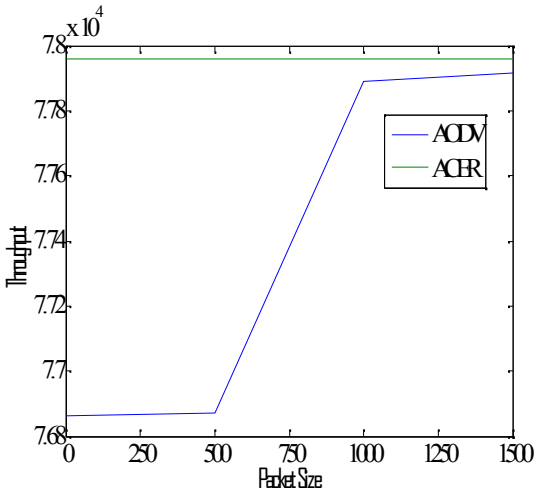


Figure 10: Packet Size Vs Throughput (Random Topology)

5. Conclusion

In MANET, there were no such power conscious algorithm which rectifies all the problems of ad hoc network. In our proposed algorithm, we show that the remaining energy of the node is used to calculate the stable path. The residual energy is calculated from initial energy and spent energy at each stage. The power of RTS/CTS mechanisms are reduced to a desired level in accordance with minimum receiving power of the intermediate nodes. The receiving energy is kept to a minimum level using the receiver's sensitivity. In this protocol, RTS/CTS transmission occurs after the route discovery and route reply to reserve the selected path so that the energy of the ad hoc node can be saved. RTS/CTS transmission consumes low energy since this transmission occurs only in the selected path. The MAC layer and physical layer functions are crucial to make a dynamic routing protocol and perform QoS parameters comparison. Here average energy consumption, packet delivery ratio, end-to-end delay and throughput are plotted against time. In future, these parameters are to be addressed against high mobility using many mobility models. Our proposed ACER performs better than AODV and DSDV in all these parameters when the energy of mobile nodes depends on the transmission of RTS/CTS after route discovery and route reply also depends on minimum receiving energy based on the receiver sensitivity.

References

[1]. C. E. Perkins and P. Bhagwat, .Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers,.in SIGCOMM '94: Proc. Conference on communications architectures,protocols and applications. New York, USA: ACM Press, 1994, pp.234.244.

[2]. C. Perkins, .Ad-hoc on-demand distance vector routing,. in MILCOM '97, Panel on Ad Hoc Networks, 1997.

[3]. D. Johnson and D. Maltz, .Dynamic source routing in ad hoc wireless networks,. Mobile Computing (ed. T. Imielinski and H. Korth), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[4]. C. E. Perkins and E. M. Belding-Royer, .Ad-hoc on-demand distance vector routing, 2nd Workshop on Mobile Computing Systems and Applications (WMCSA '99), 1999, New Orleans, USA, 1999, pp.90.100.

[5]. C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina. Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks. IEEE Personal Communications, 8(1):16–28, 2001.

[6]. Kamrok Lee, Jae Young Choi, Wook Hyun Kwon, Hong Seong Park "An Energy-efficient Contention-based MAC Protocol for Wireless Ad Hoc Networks" Networking, IEEE/ACM Transactions, Volume: 12 ,Issue:3 ,pp 493 - 506 ,ISSN: 1063-6692, June 2004.

[7]. Sylwia Romaszko and Chris Blondia "Cross Layer PHY-MAC Protocol forWireless Static and Mobile Ad Hoc Networks" EURASIP Journal on Advances in Signal Processing Volume 2009, Article ID 278041, 13 pages, doi:10.1155/2009/278041.

[8]. Thriveni J, Anita Kanavalli, K R Venugopal, L M Patnaik "Probabilistic Mean Energy Flooding to Increase the Survivability of MANET" Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol II , IMECS 2008, 19-21 March, 2008, Hong Kong

[9]. Natarajan Meghanathan, "Exploring the Stability-Energy Consumption-Delay-Network Lifetime Tradeoff of Mobile Ad Hoc Network Routing Protocols" Journal Of Networks, Vol. 3, No. 2, February 2008.

[10]. Geunhwi Lim Kwangwook Shin Seunghak Lee _ H. Yoon Joong Soo Ma "Link Stability and Route Lifetime in Ad-hoc Wireless Networks" Proceedings of the International Conference on Parallel Processing Workshops (ICPPW'02).

[11]. Sunil Kumar , Vineet S. Raghavan , Jing Deng "Medium Access Control protocols for ad hoc wireless networks: a survey" Elsevier doi:10.1016/j.adhoc.2004.10.001 2004.

[12]. Hsiao-Hwa Chen, Zhengying Fan, and Jie Li "Autonomous Power Control MAC Protocol for Mobile Ad Hoc Networks" EURASIP Journal onWireless Communications and Networking Volume 2006, Article ID 36040, Pages 1–10 , DOI 10.1155/WCN/2006/36040

[13]. Rekha Patil , DrA.Damodaram "Cost Based Power Aware Cross Layer Routing Protocol For Manet" IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.12, December 2008.

[14]. Ivan Stojmenovic, Xu Lin " Power-Aware Localized Routing in Wireless Networks" IEEE Transactions on Parallel and Distributed Systems, Vol.12, No.10, October 2001.

[15]. Morteza Maleki, Karthik Dantu, and Massoud Pedram "Power-aware Source Routing Protocol for Mobile Ad Hoc Networks" ISLPED'02, August 12-14, 2002, Monterey, California, USA.Copyright 2002 ACM 1-58113-475-4/02/0008.

[16]. S.Gunasekaran and Dr.K.Duraiswamy " Multicast Multi-path Power Efficient Routing in Mobile AdHoc networks" IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.4, April 2010.

[17]. Chansu Yu, Ben Lee and Hee Yong Youn "Energy efficient routing protocols for mobile ad hoc networks" Wirel. Commun. Mob. Comput. 2003; 3:959–973 (DOI: 10.1002/wcm.119)

[18]. Baoxian Zhang · Hussein T. Mouftah "Energy-aware on-demand routing protocols for wireless ad hoc networks" Wireless Netw (2006) 12:481–494 DOI 10.1007/s11276-006-6547-9.

[19]. V. Kanakaris, D. Ndzi and D. Azzi "Ad-hoc Networks Energy Consumption: A review of the Ad-Hoc Routing Protocols" Journal of Engineering Science and Technology Review 3 (1) (2010) 162-167.

[20]. Dr. Shuchita Upadhayaya and Charu Gandhi " Node Disjoint Multipath Routing Considering Link and Node Stability protocol: A characteristic Evaluation" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 2, January 2010.

Secure routing protocol for Mobile Ad Hoc Network using IPsec

A. El Hajjar[1-2], A.Lasebae[2], and D.K.Saini[2]
[1]Faculty of Engineering and Information Sciences, Middlesex University, London, United Kingdom
[2]Faculty of Computer and information Technology, Sohar University, Sohar, Sultanate of Oman.

Abstract - *Mobile Ad hoc network (MANET) is a wireless network that does not need any fixed infrastructure. It is based on dynamically formed mobile nodes that establish routes between each other and interconnect to transmit packets on a shared wireless channel. Designing routing protocols for mobile ad hoc networks and implementing new or existing security schemes on these networks has seen some extensive researches and a lot of experiments has been done in this topic. Our work propose a newly addition to the IPsec security suite named RSP (Random Secured Packets) with the use of AODV Routing protocol.*

Keywords: Routing, Security, Wireless, MANET, IPsec, Architecture.

1 Introduction

Mobile Ad hoc network (MANET) is a wireless network technology that does not need any fixed infrastructure. It is based on dynamically formed mobile nodes that establish routes between each other and interconnect to transmit packets on a shared wireless channel. Mobile Ad Hoc can be used in many scenarios, most importantly in disaster scenarios where a fast deployment of a network is a must in order for disaster relief personnel to coordinate their efforts through communication or in a battlefield scenario where soldiers will be able to communicate even with the absence of any control room. Easiness and fast implementation of a mobile ad hoc network whenever it is needed is one of its key successes. MANETs are envisioned to become key components in the 4G architecture, as they will offer multimedia services to mobile users in areas with no pre-existing communications infrastructure exists. [1]

In AD hoc network, nodes do not start out familiar with the topology of the network they are joining; instead they have to discover it. The idea is that each node that is willing to join the network will have to make itself visible by announcing its presence and thus this node, like any other node in MANETs will be capable of behaving as a host and as a router using either direct links or multi hop wireless links. Thus nodes in its range will be able to receive the announcement and they will learn about its presence. Since this is the basic behind Ad Hoc network, the importance of routing protocols for Ad Hoc network became the major study for any related studies on Ad Hoc networks.

However, designing a new routing protocol for Mobile Ad Hoc Networks is not an easy task. Since no fixed infrastructure is available for ad hoc networks, the topology keeps on changing on a continuous basis when new nodes are joining or leaving the network. Routing in Ad Hoc mobile networks is challenging mainly because of node mobility. The more rapid the rate of movement, the greater is the fraction of bad routes and undelivered packets."[2] Extensive research underwent to solve the complexity of routing protocols in Mobile Ad hoc network and several protocols where designed for this purpose. Most of these researches on MANETs revolved around three major routing protocols, all of them "On Demand protocols" such as AODV and DSR since using a "Table driven protocol" is not very realistic. The type of devices that usually join a MANET network such as portable devices and small devices that have limited power, limited battery life and limited storage capacity makes it very hard for this type of devices to store information about all devices in the network and update its table every specific period by broadcasting a message, as this will consume devices batteries dramatically and causes overhead in the network.

Routing primary security service is authorization. Typically, a router needs to make two types of authorization decisions authentication and integrity. Techniques like digital signatures and message authentication codes are used to provide these services. Denial of service attacks in a wireless network although of course it would be desirable, it does not seem to be feasible to prevent denial-of-service attacks in a network that uses wireless technology (where an attacker can focus on the physical layer without bothering to study the routing protocol).

Eavesdropping and masquerading are not very difficult. Node security is another major concern as mobile nodes can fall into hostile control. There have been widely reported cases of theft of cellular nodes, so MANET nodes would not be any safe. The node could be compromised and thus would act as a hostile node. Easy theft might also lead to node tampering. Tampered node might disrupt network operations or release critical information. The limited powers in the mobile nodes can lead to a simple denial of service attack where the attacker could create additional transmissions or expensive computations. Lack of fixed topology requires the routing protocols to be highly sophisticated and securing it presents a challenge in the presence of hostile nodes.

Until not long time ago, and since noticing the importance of designing a routing protocol for Mobile Ad Hoc networks, the security part of the routing protocol were completely ignored other than few proposed solutions. Those solutions either did not achieve the results they were designed for or created other challenges on the network such as large overheads on mobile ad hoc networks or in our opinion contradicted the whole idea of MANETs by ensuring that a centralized key exchange and distribution occurring between nodes.

Our main goal is to find a secure enough routing protocol that can be implemented on AODV without degrading the quality of service of MANETs and without contradicting the main idea of MANET, the decentralized architecture. Our proposed solution is to add a protocol to compute a certain pseudo random number which will be applied on certain packets that will be secured using one the IPsec protocols to the IPsec suite and implement it on AODV routing protocol without degrading the level of quality of service and with keeping the concept of decentralized architecture for MANET networks.

The rest of the paper is organized as follows. In section II, we briefly review the DSR and AODV protocols and we review other proposed solutions for securing routing protocol for Mobile Ad Hoc Networks. In section III we present a detailed description of the AODV protocol and the IPsec architecture and in section IV we present our proposed solution on how to secure routing for MANET using IPsec and in section V we present any further work that still need to be done.

2 Previous studies

The authors in [2] presented a good comparison of the performance of DSR and AODV and came out with the following results: DSR and AODV both use on-demand route discovery, but with different routing mechanics. In particular, DSR uses source routing and route caches, and does not depend on any periodic or timer-based activities. DSR exploits caching aggressively and maintains multiple routes per destination. AODV, on the other hand, uses routing tables, one route per destination, and destination sequence numbers, a mechanism to prevent loops and to determine freshness of routes. The general observation from the simulation is that for application-oriented metrics such as delay and throughput, DSR outperforms AODV in less "stressful" situations (i.e., smaller number of nodes and lower load and/or mobility). AODV, however, outperforms DSR in more stressful situations, with widening performance gaps with increasing stress (e.g., more load, higher mobility). DSR, however, consistently generates less routing load than AODV. The poor delay and throughput performances of DSR are mainly attributed to aggressive use of caching and lack of any mechanism to expired stale routes or determine the freshness of routes when multiple choices are available. Aggressive caching, however, seems to help DSR at low loads and also keeps its routing load down. On the other hand, AODV's routing loads can be reduced considerably by source routing the request and reply packets in the route discovery process. Since AODV keeps track of actively used routes, multiple actively used destinations also can be searched using a single route discovery flood to control routing load.

The authors in [3] concluded in their comparison of DSDV, DSR and AODV that when they compared the routing protocols based on generated parameters, DSR performed better. When they compared the routing protocols based on received packets vs. number of nodes, DSR performed better up to 10 nodes, AODV performed better for more than 10 nodes. When they compared based on total dropped packets Vs. No of nodes, DSR performed better. When they compared based on packet delivery ratio vs. No of nodes, AODV performed better and finally when they compared based on average End to End delay vs. number of nodes, AODV performed better up to 10 nodes, DSR performed better for more than 10 nodes.

Most of the secure routing protocols proposed on different papers are based on sharing a cryptographic symmetric key cryptography or on some sort of security association. Although this might be possible we do not believe that this the solution for securing Ad Hoc Network as this type of solution actually conflict with the whole idea of Ad Hoc network, such as an infrastructure-less network that is decentralized and that any node in the range can join the network and participate. The computational powers of the nodes also make the use of PKI during normal operations highly infeasible.

Other research papers study the routing protocols of MANETs and how to secure them which usually involve one of the two security mechanisms on the routing protocol, either securing the exchange of routing information or securing the data packet forwarded between nodes.

Secure Routing Protocol (SRP) is one proposed security solution that only secures the route exchange without addressing the protection of Data transmission which will need Secure Message Transmission Protocol (SMT) to cover it. In this proposal, SRP protocol based on SA (Security Association) between source and destination which is implemented by using public key and the two nodes can

negotiate a shared secret key. This way, it is ensured that the packets received from a node are actually from this actual node and not a malicious node thus the received packets is authenticated against the sender's id. This security solution is good, but only securing the routing exchange leave the message uncovered, and if protected using SMT, it will create a large overhead on the network. In the paper titled "Enhanced DSR for MANET with Improved Secured Route Discovery and QoS" the authors in [4] proposed the use of SRP on DSR routing to enable securing the route discovery in presence of adversarial node. The authors continue to presume that the route discovery is entrusted to SRP when integrated in DSR and thus the goal of SMT, if implemented on addition to S-DSR is to ensure secure data forwarding after discovery. The authors continue by saying that the proposed inclusion of SMT functionality in S-DSR will disperse data into multiple packets and reconstruct the packets at the destination. This will increase the processing overheads, but will provide secured data transmission.

Ariadne, SEAD are two other security solution proposed on DSR both designed by the same team, SEAD can only be used with any suitable authentication and key distribution scheme which is still not a straightforward matter and thus does not actually contribute to a solution. On the other hand, Ariadne also based on DSR provides an authentication mechanism using TESLA but also requires clock synchronization which is an unrealistic requirement for Ad Hoc. TESLA also incurs a delay since it requires that packets are delayed by the longest RTT in the network in both request and responses phases).

The authors of [5] proposed a security solution by developing ARAN (Authenticated Routing for Ad Hoc Networks). Although this authentication overcame the delays required by TESLA protocol, the authors conclude that the cost of ARAN is larger routing packets, which result in a higher routing load and higher latency in route discovery because of the cryptographic computation that must occur.

The authors of [6] proposed a secure routing for supporting Ad Hoc Extreme Emergency Infrastructures using IPsec for different routing protocols such as AODV, OLSR DYMO. The authors concluded that using the hybrid mode of AH and ESP towards securing MANETs will guarantee authentication and confidentiality. They carry on by discussing that the transport mode of the IPsec protocol was used in order to avoid high processing power overhead. We believe that due to the high mobility rate of MANET devices, using ESP and AH options at the same will generate large overhead and will create large delay.

3 IPsec and AODV

The Ad Hoc On-demand Distance Vector Routing (AODV) protocol is a reactive unicast routing protocol for mobile ad hoc networks. As a reactive routing protocol, AODV only needs to maintain the routing information about the active paths [7].

The AODV algorithm enables dynamic, self-starting, multi-hop routing between participating mobile nodes wishing to establish and maintain an ad hoc network. AODV allows mobile nodes to obtain routes quickly for new destinations, and does not require nodes to maintain routes to destinations that are not in active communication. AODV allows mobile nodes to respond to link breakages and changes in network topology in a timely manner [2].

In AODV, routing information is maintained in routing tables at nodes. Every mobile node keeps a next-hop routing table, which contains the destinations to which it currently has a route. A routing table entry expires if it has not been used or reactivated for a pre-specified expiration time. Moreover, AODV adopts the destination sequence number technique used by DSDV in an on-demand way.

It shares DSR's on-demand characteristics hence discovers routes whenever it is needed via a similar route discovery process. However, AODV adopts traditional routing tables; one entry per destination which is in contrast to DSR that maintains multiple route cache entries for each destination. The initial design of AODV is undertaken after the experience with DSDV routing algorithm. Like DSDV, AODV provides loop free routes while repairing link breakages but unlike DSDV, it doesn't require global periodic routing advertisements. AODV also has other significant features. Whenever a route is available from source to destination, it does not add any overhead to the packets.

However, route discovery process is only initiated when routes are not used and/or they expired and consequently discarded. This strategy reduces the effects of stale routes as well as the need for route maintenance for unused routes. Another distinguishing feature of AODV is the ability to provide unicast, multicast and broadcast communication. AODV uses a broadcast route discovery algorithm and then the unicast route reply massage.

IPsec (Internet Protocol Security) is an IP layer security mechanism suite that can be used to protect the entire path between two entities by authenticating and encrypting each IP packet of a communication session. IPsec is not a single protocol. Instead, IPsec provides a set of security algorithms plus a general where two communicating devices can choose the algorithm that suits it. The principle feature of IPsec is to provide the security in various scenarios by encryption and/or authentication to all traffic at IP level. IPsec allows the two entities to negotiate and select the required protection mechanism, such as "authentication only" or "authentication and encryption", select proper cryptographic transform to use for the chosen protection, and exchange the keys required for those transforms. [9]

Several services can be provided by IPsec based on the different algorithms such as Access Control, Connectionless integrity, Data origin authentication, Rejection of replayed packets, Confidentiality (encryption) and Limited traffic flow confidentiality.

The architecture IPsec is composed of the base protocol that implements Encapsulating Security Payload (ESP) and Authentication Header (AH) by processing the headers and interacting with the security policies database (SPD) and the security association database (SAD) to determine the security level afforded. ESP is the part of IPsec architecture that covers the use of ESP protocol to encrypt the packet. The authentication header (AH) is the part responsible of the format of the packet when IPsec is implemented on AH authentication option or on the ESP authentication option. AH and ESP do not actually provide any security protection but they must be used along with other cryptographic mechanism to provide such as Key Management. Key management is one important element in IPsec suite that is responsible of exchanging and distributing keys. The last element in the IPsec suite is the Domain of Interpretation protocol that contains values needed for all the IPsec protocols to work together. The three protocols that can be used in an IPsec implementation are ESP Encapsulating Security Payload and Encrypting and/or authenticating data, AH, Authentication Header provides a packet authentication service and IKE, Internet Key Exchange, Negotiates connection parameters, including keys, for the other two.

IPsec can also assure that a router advertisement and neighbor comes from an authorized router, a redirect message comes from the router to which the initial packet was sent and routing update is not forged.

4 Proposed solution

In our proposed potential solution for securing AODV routing protocol using IPsec to authenticate nodes when they join Mobile Ad Hoc Network and when they establish a new route or for data transmission over an established route.

When deciding to use tunnel mode in IPsec in comparison with Tunnel mode we based our decision on several points: if using transport mode, this would mean that transport layer header such as TCP Header will follow the ESP header immediately. Security Architecture allows transport mode to be used by end hosts but the security gateway can only use this mode when it is acting as an end node. When it is acting as intermediary, it cannot be used. However in Tunnel mode, the entire original IP packet is covered and not only the transport header will be protected by the IPsec. The initial IP packet is encapsulated by another IP packet that includes an IPsec header (ESP or AH). In a routing scenario basis, the two hosts, each in a remote location communicating through different nodes. All the forwarding nodes on its route simply take the IP packet without knowing what are the contents and the destination node will simply removes the outer header and IPsec header and decrypts the contents which also include the inner IP header.

Like the authors of [6], we are opting to use tunnel mode as Transport mode if used in MANET will add a great overhead on the network. The difference in our proposed solution is that what the authors in [6] propose is to use both ESP and AH options at the same time.

What we propose, as shown in figure 1 below, is to use ESP and AH on the routing discovery phase as the first provide confidentiality and data origin authentication and the second provide connectionless integrity and authentication for the IP datagrams since all the MANETs communication is encapsulated in an IP header. For data transmission phase, we add a new protocol to the IPsec security suite called RSP (random Secured packet) that implements a pseudo random generator that computes a random number of how many packets will be secured using both AH and ESP on the datagram based on how many packets the source is sending to destination. This way and although the overload on the network will still be quite large, it will be radically reduced in comparison with when applying both AH and ESP at the same time without using RSP.

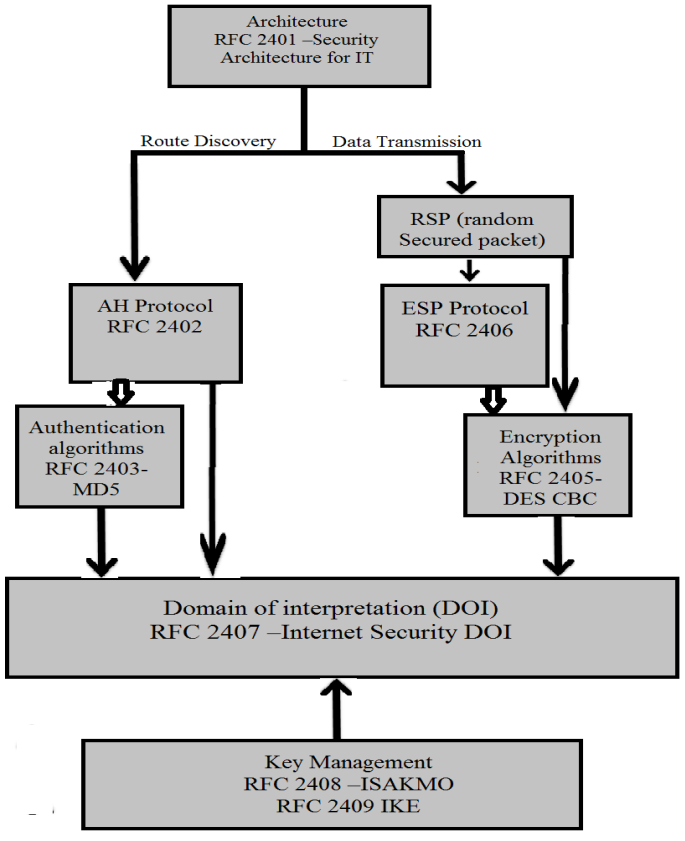


Fig 1: Newly proposed addition to the IPsec protocol suite

RSP, by reading the packets being received, and analyzing the message that is being transmitted chooses based on the number of packets and based on the type of communication, the number of packets that needed to be secured using both ESP and AH and thus applies the pseudo random generator to generate which packets will be secured. Making it randomly chosen make it a lot harder for anyone trying to listen to the transmission to guess which packets will be secured and which packets will not. Also by not having the same order of packets that will be secured using ESP and AH, makes it extremely hard if not impossible to guess the message being transmitted only from the packets that were sent unsecured by both ESP and AH.

5 Future work

We are currently finalizing the test-bed which will be used to test the newly developed addition the IPsec. Extensive complex simulation when carried by our test-bed will provide a more in depth performance analysis of the MANETs when secured using our newly adapted IPsec suite with AODV.

6 References

[1] Ronald Beaubrun and Badji Molo, Using DSR for routing Multimedia traffic in MANETs, International journal of Networks & communications (IJCNC), vol2, No1, January 2010.

[2] Perkins,C.E., Royer,E.M., Das,S.R., and Marina,M.K. (2001). Performance Comparison of Two on-Demand Routing Protocols for Ad Hoc Networks. IEEE Personal Communications, 16-28.

[3] V.K.Talsande and K.D.Kulat, (2011) performance comparison of DSDV, DSR, AODV protocol with IEEE 802.11 MAC for chain topology for Mobile Ad Hoc Network using NS-2, IJCA Special issue on "2nd National conference-Computing, communication and sensor Network" CCSN, 2011.

[4] Anil Rawat1, Prakash Dattatraya Vyavahare and Ashwani Kumar Ramani,, Enhanced DSR for MANET with Improved Secured Route Discovery and QoS , Department of Electronics and Telecommunication Engineering, Shri G. S. Institute of Technology and Science, India International Journal of Network Security, Vol.5, No.2, PP.158–166, Sept. 2007.

[5] Sanzgiri, K.; LaFlamme, D.; Dahill, B.; Levine, B.N.; Shields, C.; Belding-Royer, E.M., Dept. of Computer. Science., Univ. of California, Santa Barbara, CA, USA , Authenticated Routing for Ad Hoc Networks, IEEE journal, 2005.

[6] Emmanouil A. PANAOUSIS, Tipu Arvind RAMREKHA and Christos POLITIS, secure routing for supporting Ad Hoc Extreme Emergency Infrastructures, Future Network Summit, 2010.

[7] Park, and S. Corson, Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification. IETF Internet draft, 1997.

[8] Madjid Nakhjiri and Masha Nakhjiri, AAA and Network Security for Mobile Access, Motorola Labs, Johm Wiley & Sons, UK.

[9] K.A. Nusrath, et-al, MANIPSEC- IPSEC in mobile Ad Hoc Networks, Warangal, India

Application And Research On Mobile E-Business Payment Client Based On Cross-Platform

Zhang Zhong-ge, Zheng Nian-bin, Zhou Ze-fen, Peng Ge-gang, SHEN Qing
Talkweb Information System CO. LTD, China，410205

Abstract - *With the globalization of the information technology revolution and the development of intelligent mobile phone, the development of mobile e-business obtains more and more support. Mobile payment is the cornerstone and motive power of mobile e-business. Compared to other channels, such as website, SMS, and WAP, mobile phone client has become the most important channel because it has many advantages on portability, user experience and convenience. In view of the diversity of operating system for intelligent mobile phone, this paper presents a way on mobile client solutions based on cross-platform for mobile e-business and emphatically introduces the system's architecture and key issues. Actual operation situation shows that such a system has a good safety and practicality, and may become a powerful payment tool for mobile e-business applications.*

Keywords: Mobile e-business, Cross-Platform, Mobile phone client, Security

1 Introduction

Now the mobile communication technology and Internet technology has gradually become the two big pillars for the information industry. As the combination of both the industries, mobile e-business rapid develops [1]. Mobile e-business is the result of the rapid development of network communication technology. Based on wireless network technology, mobile e-business truly realizes the operation of cross-platform. With the rapid development of Internet technology, E-business has gradually become a new model for business activities. It fully supports Internet services. Users can use a smart phone or a PDA anytime, anywhere to search, select and purchase goods and services, and make use of electronic means to achieve payment [2].

Mobile payment is the core of mobile e-business. Mobile client has great significance for changes of payment means, expansion of payment scope, improvement of payment security and improvement of customer satisfaction [3, 4].However, the diversity of operating system and the large difference for intelligent mobile phone bring about high threshold for terminal application development of mobile payment, long development cycle and large difference in user experience. All of these have a direct impact on promotion deployment of mobile payment client and impede the development of mobile e-business for intelligent mobile phone. This paper puts forward a way on mobile client solutions based on cross-platform for mobile e-business and emphatically introduces the system's architecture and key issues to be resolved.

† This research was supported by the Ministry of Science and Technology of China , and the Ministry of industry of information technology of China (Project Name: Office information system based on homemade CPU/OS for Hunan province government No: 2102ZX01045-004-005-003)

2 Architecture

As shown in Fig1, the mobile client system is divided into mobile payment client, payment center system and version management system.

- **Mobile payment client:** It is the core system of the whole system structure. It is mainly responsible for user interface display for payment function and performs simple logic judgment on service. The structure is divided into application layer, application framework layer, the runtime library and operating system. Among them, the applications layer implements the functions of interface display and Interaction. These Interaction functions include user login, user registration, user payment, account management and other functions. Application framework layer provides framework support for application. The UI manager is responsible for the user interface layout management and subject management. Event manager encapsulates the events of bottom key and touch screen, which provides a unified treatment mechanism for different mobile phone manufacturers. The business process controller offers support for business flow and scene switches. Resource management is responsible for resource creation, resource recycling and resource release. Network management carries out the management of communication protocols, connection establishment, and data access and message analysis. System runtime layer mainly supplies support for program library and runtime library. Program library include local data access interface, local database access interface, wireless message interface, the underlying communication interface and multimedia interface. Runtime library offers a runtime environment. Operating system layer provides implement guarantee for the core function of application. These mainly include security, memory management, process management and network protocol stack.

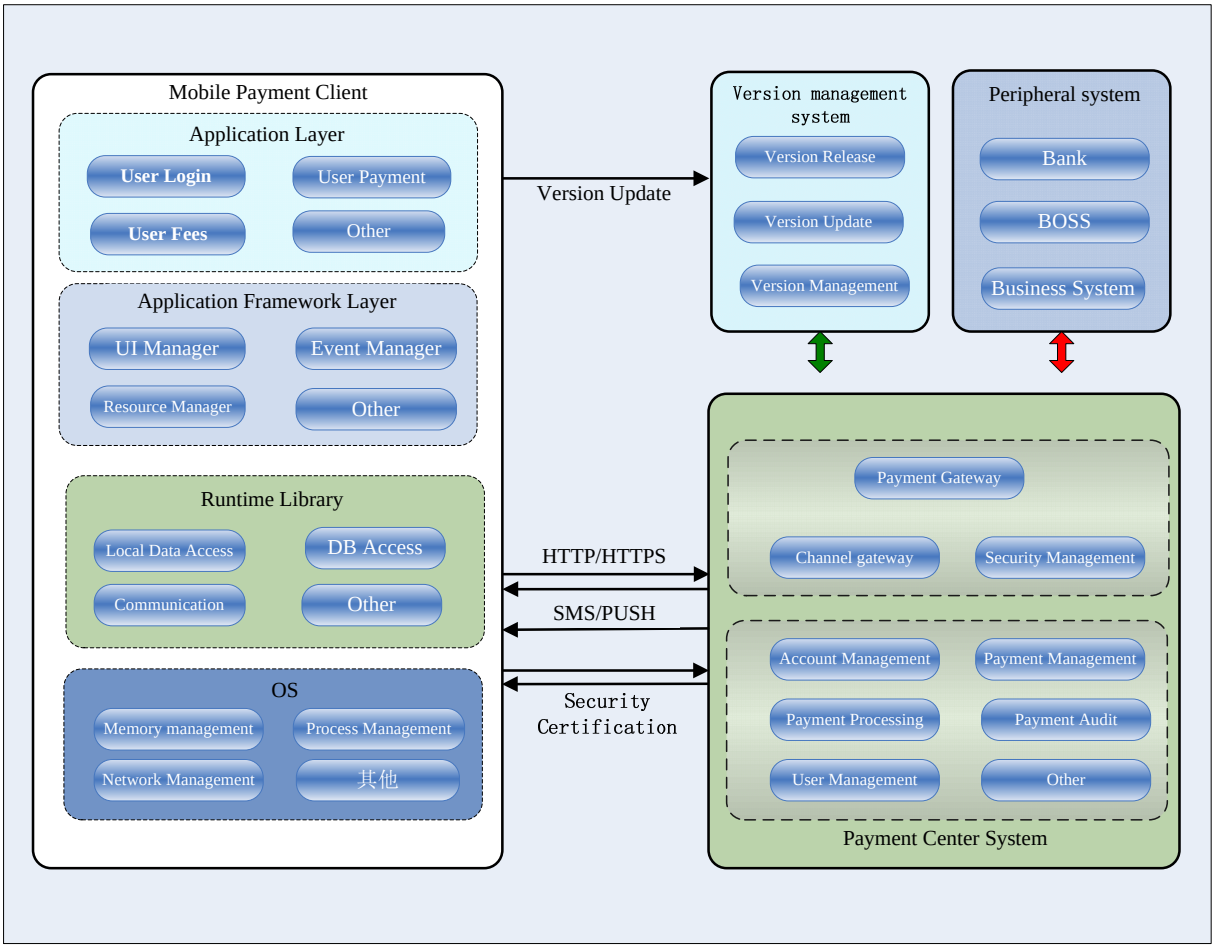


Fig1 the Architecture of Mobile E-Business Payment Client

Payment center system: Payment center system, which can be divided into front end and back end, provides service interface specification and implements the required business functions. Front end is in charge of the access of various channels and the three-party peripheral system, and performs protocol conversion between different systems. The other task is to carry out security monitoring, flow control, deployment management and transaction start-stop control, and provides uniform data view. Back end achieves the main business logic and data processing, and mainly includes account module, payment processing module, payment management module, payment and settlement module, marketing management module and processing module.

Version management system: It is an information management and storage platform for version and plug-in. Version management server responds to the request for client version and plug-in download. Version management system is not related to the service logic processing and is mainly responsible for the version release, version update and version management.

3 Research on Key Technologies

There are some key technology issues to be solved during the construction process of mobile electronic payment client system based on cross-platform. We presents our own solutions combined with the characteristics of mobile payment after we fully research traditional solutions. These research and solutions for these key technologies have played a crucial role in the mobile client system's construction. The following discussion will focus on cross-platform scheme and security scheme.

3.1 Cross-platform Scheme

The popularity of intelligent terminals effectively promoted the development of mobile e-business. However, the diversity of mobile phone's operating brings out high threshold for development, long development cycle and too large terminal adapter workload. All of these directly affect development and promotion of terminal application. Thus this also restricts the rapid development of mobile e-business on the other hand. How to realize cross-platform development has become an important issue need to be solved in the development of mobile e-business payment client.

3.1.1 Cross-platform Scheme Base on VM

At present, middleware and code switching is main cross-platform development scheme.

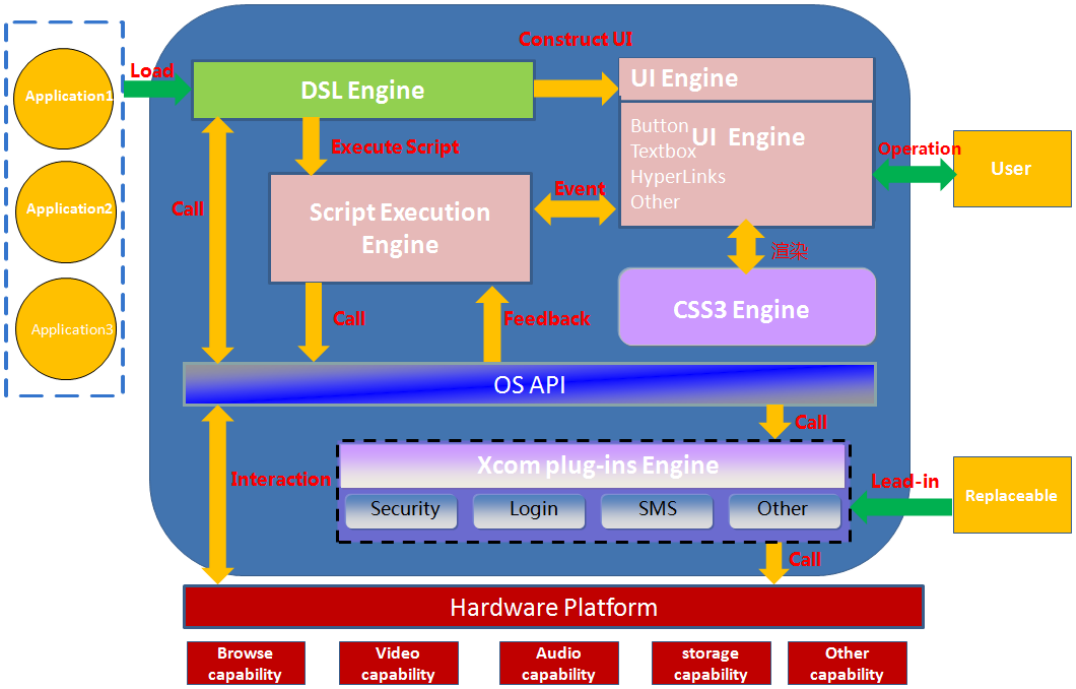


Fig2 the Architecture of Core engine

Middleware mode reduces the repetitive workload for developer by providing cross-platform SDK package. Nokia QT is the primary representative of the mode. After Nokia acquired Trolltech Company, it began to develop Qt to realize its own target. Now Qt could cover many platforms such as Symbian, windows mobile, windows, Linux and mac OS. Qt is essentially a set of cross-platform SDKs. Early, developers need to use C + + language to perform the development of Qt. Then Qt launched Qt Jambi so that developers can also use Java development. The regret is that the official has no plan to support Android. It is also difficult to say whether Android will be considered in the future. Although the third part open source project has implement code transplant in part, all of UI components belongs to Qt's own and original control could not be used. This brings out a great discount of user experience. The main drawback is that Qt uses its own control to meet code transplant demand but loses user experience of each platform [5].

In code switching mode, mobile phone developers adopt a unified language to develop such as JaveMe. Then source code is converted to each platform such as windows mobile by using a conversion engine. At last the resulted code is compiled and debug. AlcheMo is the primary representative for the mode. Currently AlcheMo supports iPhone，Android and windows mobile platform, but does not support symbian. AlcheMo introduces a bridging mechanism which can make developers for mobile phone application directly use native API to enhance platform's function. The main drawback of the mode is that separate debugging is required and the success rate of code switching is not high which leads it often need to modified by hand although developers just write code once[6].

In order to overcome the shortcomings of middleware mode and code switching mode, we put forwards cross-platform scheme based on virtual machine. It is different from existing J2ME JVM scheme that the scheme implement VM for various target platforms to ensure JVM unified. In our scheme, a set of client core business engine are built to realize the UI analysis and drawing. Then target application packages are generated after code are compiled and packaged by using visual developing platform (our own IDE). The target application packages call the underlying system capabilities and functions, and offer UI interface analysis and drawing through system instructions interpreted by the corresponding client core business engine. In this way, interface control and application functions for different platform are achieved. The key of our scheme is that JVM are adopted and its executable file uses binary intermediate format.

3.1.2 Core engine Based on Cross-platform

As shown in Fig2, the key technology is core engine based on cross-platform. The core engine of mobile e-business payment client is made up of browser engine, script execution engine (JVM engine), UI engine, CSS3 engine and XCOM plug-in engine. The browser engine is mainly responsible for loading application and UI engine performs UI drawing. Interface rendering work is CSS3 engine's task. The XCOM plug-in engine is in charge of behavior management for plug-ins. This management includes plug-ins location, plug-ins query, plug-ins download and plug-ins installation.

Standard WAP package and enhanced WAP package are two kind of application which can be produced by development engine based on cross-platform.

For the loading of Standard WAP, application will directly call system browser component via local API module. When the system browser needs to call a plug-in capability, the specific capability will be used by XCOM engine.

As for the loading of the enhanced WAP, enhanced WAP application will be decomposed into the scripts and pages, which are respectively processed by script engine, UI engine and CSS3 engine, after application package in application repository is loaded by browse engine. Enhanced WAP application consists of enhanced WAP tag description, scripts, and resource files.

3.2 Security Scheme

Due to the special characteristics of mobile e-business, mobile e-payment security is particularly important. Security issue has become the most important factors for mobile e-business [4]. During the building of the system, we put forward practical solutions to solve the access control, access security, data storage security and availability.

In access control, we mainly use user name and password authentication because payment business quota has been controlled from business requirement. Access control involves password initialization, login password's authentication and payment password's authentication. Among these processes, password initialization and login password's authentication are the most important and they are as shown below.

The process of password initialization is described as shown in Fig3. The authentication platform returns server public key to the mobile client after mobile client sends the request for downloading public key to authentication platform. User inputs initial login password and payment password, then mobile client generates random number and encrypts login PIN with server public key. In addition, mobile client randomly generates transmission key, and then the transmission key is encrypted in server public key. Then mobile client sends the request for user registration to the authentication platform. The request includes the login PIN cipher text and the payment PIN cipher text. The login PIN cipher text, which is encrypted by the server public key, contains 128 byte. The payment PIN cipher text, which is encrypted by the transmission key, contains 8 bytes. After receiving registration request, authentication platform invokes encryption machine interface to convert login Pin's cipher text and payment PIN's cipher text into PIN's cipher text encrypted by PVK encryption. If PIN cipher text needs to be transferred between different network elements, ZPK key shared between encryption machines is used for the encryption. Back-end authentication platform returns the initialization result to the mobile client after it performs the initialization of login PIN and payment PIN.

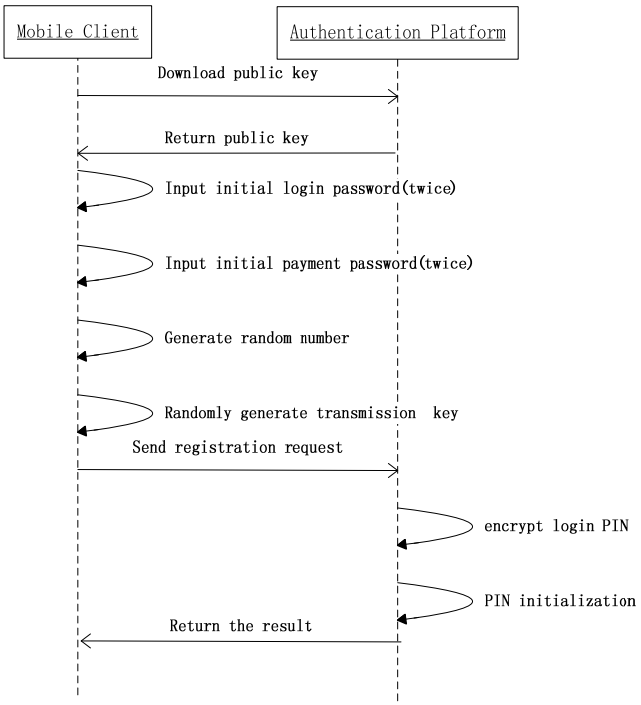


Fig3 the process of password initialization

The process of login password's authentication is described as shown in Fig4. The authentication platform returns server public key to the mobile client when it receives the request for downloading public key from mobile client. User inputs login PIN. Then mobile client generates random number and encrypts login PIN with server public key and sends the request for user login to the authentication platform. The request contains login PIN's cipher text. After receiving the login request, the authentication platform invokes encryption machine interface to convert login Pin's cipher text into PIN's cipher text encrypted by PVK encryption. If PIN cipher text needs to be transferred between different network elements, ZPK key shared between encryption machines is used for the encryption. The authentication platform returns the verification result to the mobile client after it checks the cipher text to be verified.

We also take other aspects in security solution into account. In communications security, we use HTTPS security protocol to communicate with backend server. In data storage security, we present strict requirements for sensitive data stored in the mobile terminal. Password in the client appears only in the memory and should be deleted immediately after is verified. In client availability, some measures also are taken to provide security protection. These measures include exception handling for call, power interruption and network interruption during transaction process.

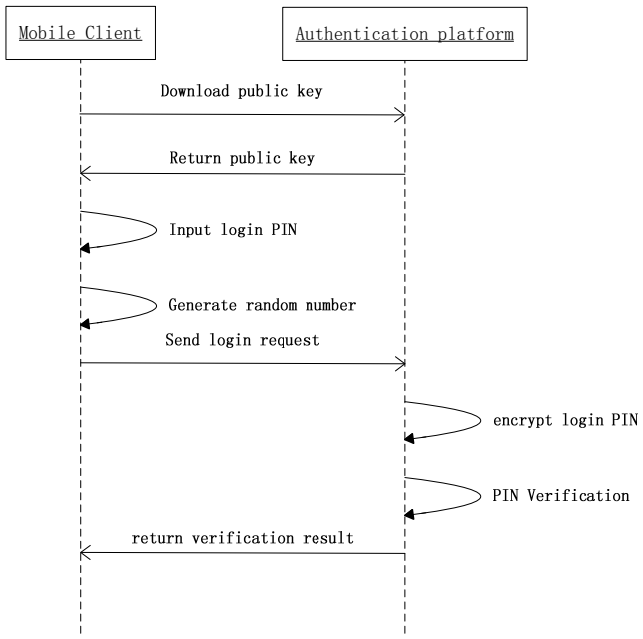


Fig4 the process of login password's authentication

4 Conclusion

This paper presents mobile client solutions based on the cross-platform for mobile e-business. The key issues to be resolved in the process of building the system are discussed based on the introduction of the architecture of the system. These crucial issues mainly include cross-platform scheme and security scheme. Actual operating condition states that the mobile e-business client based on the cross-platform has a good security and practicality, and could be an excellent tool for mobile payment applications.

5 ACKNOWLEDGMENT

We would like to thank deeply to the Ministry of Science and Technology of China , and the Ministry of industry of information technology of China for their support on this project.

References

[1] Hao Yu,Cheng Zhu,Hongming Cai, et al. "Role-Centric RESTful Services Description and Composition for E-Business Applications".2009 IEEE International Conference on e-Business Engineering(ICEBE '09),volume(1):103-110

[2] Li Cuizhi,Yue Yunkang ."A study on key technologies in the development of mobile e-commerce".2011 International Conference on E-Business and E-Government (ICEE), volume(1):1-4

[3] Lao Guoling,Liu Hanbing."Study of Mobile Payment Business Model Based on Third-Party Mobile Payment Service Provider", 2011 International Conference on Management and Service Science (MASS),volume(1):1-4

[4] Dizaj, M.V.A.,Moghaddam, R.A.,Momenebellah, S. ; Momenebellah, S. ,"New mobile payment protocol: Mobile pay center protocol 2 (MPCP2) by using new key agreement protocol: VAM" ,2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim), volume(1):12-18

[5] Akinkuolie, B.B.,Lin Chia-Feng,Shyan-Ming Yuan."A Cross-Platform Mobile Learning System Using QT SDK Framework".2011 Fifth International Conference on Genetic and Evolutionary Computing (ICGEC), volume(1):163-167

[6] Almangoush, R.,Aneiba, A. ,"A model for mobile voice message service using JAVA ME",2010 2nd International Conference on Computer Technology and Development (ICCTD), volume(1):42-55

A Multi-level Contextualization Framework for Authentication of Mobile Payment Applications

Deveeshree Nayak
KIIT University, Bhubaneswar, India
deveeshree@gmail.com

Srini Ramaswamy
Industrial Software Systems
ABB India Corporate Research Center
srini@ieee.org

Abstract: In this paper, the emerging need for a user-driven contextualized approach towards security authentication for today's emergent mobile e-commerce applications is outlined. The main advantages of contextualizing user-driven authentication is that it can be almost impregnable by malicious users who attempt to get personal details through various intrusion techniques or by plain theft of a mobile device. Drawing on existing work on context modeling, in this paper we develop a 4-layer framework that can support the provisioning of trustworthy authentication measures for mobility applications to garner widespread public adoption. If appropriately evolved, such technology might allow banking institutions in growing markets to leapfrog credit-card based payment schemes and rapidly reach an untapped 'banking market' segment of mobile users in new and emergent economies.

I. Introduction

In most modern day applications, there exists a dire need for a user-driven contextualized approach towards security authentication. Such needs are often evolved through contexts, wherein a context refers to the information which can be used to characterize the situation of an entity under consideration. An entity may be a person, location, object, etc. which are considered to be relevant for the behavior of an application. These contexts then need to be applied to address security authentication challenges through the use of policies. To establish contexts, researchers have applied context modeling, which is an attempt to describe not just the entities involved in a system but also their relationships. Such context modeling can be accomplished through either bottom-up or top-down techniques. In real-world applications (ex. mobile wallet) that have crept into the daily lives of millions of people, such contextualization must necessarily include some regionalization and localization for the needs of usable security authentication. To determine such regional and application-specific contexts, there must be a delicate and well-defined balance between the complexity of the authentication measure and the relative simplicity and ease of use by any individual. Thus the main scientific contribution of this paper is the development of a compelling case for personal and application-specific contextualization and a proposed that can accomplish this successfully. Such context-aware techniques have the potential to support the development of pervasive, user-centered computing applications that are flexible, adaptable, and capable of acting autonomously on behalf of their users.

II. Context Modeling Approaches

Current approaches to context modeling can be broadly classified into the three main categories: These include [1]:

1. *Object-role based context modeling:* these approaches are called fact-based context modeling approaches, have evolved from data base modeling techniques, and they attempt to create formal models of context to support automated processing for queries and reasoning. It also supports software engineering tasks such as analysis and design. Tools such as CML [2-4], based on OML [5], have been developed for this need. Such tools provide a database-style management of context information and typically offer interfaces for applications to query context information or receive notifications on context changes.
2. *Spatial context modeling*: Spatial context modeling is relevant for developing context-aware applications that are primarily location-based, for example mobile information systems. Such information may include the position of entities as well as its spatial relation to other related entities. This may be an area, range or distances to other entities. Typical queries for such a context management platform will include support for queries that help determine, position, range, nearest neighbor, geometric / symbolic coordinates, etc. [6]. Tools for special context modeling have been in the market by several vendors [8, 9, 10]. However, a major drawback of this approach will be the effort that it takes to gather and organize location data, as well as to keep it current.
3. *Ontology-based context modeling:* In this approach researchers attempt to support applicative needs that require a thorough representation of knowledge. Ontological representations attempt to describe complex context data that cannot be otherwise described by simple languages, by allowing for formal semantics based data descriptions, thereby making it available for relationship consistency checks. While it is highly expressive and could also support interoperability, it can be too complex and inappropriate for certain kinds of applications (such as 'thin' mobile-based payment systems). However, in [6], the authors have evolved a multi-tiered approach for ontology based contextual modeling, based on a 5-tiered ontology presented in [7]; namely, physical reality, observable reality, object world, social reality and cognitive agents.

III. Contextualization for Security Authentication

Given the above brief survey on contextualization for development of pervasive applications, specifically for the need to develop personalized and regionally / culturally relevant security authentication measures we foresee the following challenges: *(i)* First for each application and cultural / regionalization need, one needs to define a suitable context model for such systems to act as the primary basis for the policy generation. *(ii)* Second, the security authentication policies themselves have to be derived from this contextual model. *(iii)* Third, such a policy must also be enabled along with appropriate context management, i.e., the efficient management of context information and feasible context representations in order to allow reasoning.

The design of such societally pervasive systems thus cannot be highly application-centric – which has been the case thus far, and appropriate focus need to be paid to personally identifiable information (PII) and the management of such PII information from a contextual perspective. Moreover, while such context information must be rich enough to evaluate whether there is a chance for the user to understand the authentication measure with the application environment and policy situations that are prevalent locally, it should support decision making when adaptation to the context is necessary due to changes in the regional context that will additionally also require reasoning capabilities. Hence a framework for such systems must support multi-level context development, whereby higher level context information may be utilized to support emerging needs such as appropriate consistency verification and support in-depth reasoning about unforeseen, complex situations.

IV. Authentication Measures for Mobile Payments in a Regional Context

Mobile payments add a degree of flexibility for the end customer while reducing office / retail space costs for a vendor and hence are increasingly viewed as a significant business opportunity by all kinds of vendors. Typically, as shown in Figure 1, mobile payment mechanisms can be classified as one of the below three broad categories.

1. Account-based schemes: Account-based schemes are based on the principle that a customer will 'open' or maintain an account, using which they will be billed. Such schemes may be either direct cash based or token based (which map to some cash value) and typically are not small transactions. They can be further classified into phone based (some carriers in India allow 'deposit' of money / credit to facilitate such account-based schemes, card-based or point-of-sale (using a merchant). An interesting twist in this scheme that is popular in the Indian market segment is COD (cash-on-delivery), where in the customer has established a degree of trust with the vendor and hence the valuable purchased is shipped and payment received on delivery.
2. Mobile Wallet: Mobile wallet is a functionality that resides in a mobile device and supports secure interactions to digitally transact using the wireless backbone. They can help facilitate mobile payments, mobile commerce, manage mobile identify and engage in banking / financial transactions.
3. Micro payments: In a micropayment the user and seller each establish an account with a third-party service provider who monitors, collects and distributes micropayments, which are categorized as a small sum of money in exchange for something made available online. Due to the small nature of the transaction micropayments typically accumulate until they are collected as a single, larger payment.

While mobile payments add immense degree of user flexibility, security of these transaction mechanisms and ease of use of these interaction exchanges is of immense interest to both customer and the financial institutions. Security, trust and privacy are therefore critical for organizations to cultivate an effective, mutually rewarding relationship with the customer. Additionally, typically such services also involve other significant barriers to entry such as *(i)* High costs: For example if one wants to use services such as IBM's context service [12], which integrates different context information required by the client application from the different sources. However, it is often left to the customer to explore further integration to provide information spanning multiple instances. *(ii)* Low payout rates - operators also see high costs in running and supporting transactional payments which results in payout rates to the merchant being as low as 30% (usually this is around 50%) *(iii)* Low follow-on sales - once the payment message has been sent and the goods received there is little else the consumer can do. It is difficult for them to remember where something was purchased or how to buy it again. In such a nebulous business environment, it is easy for security authentication measures to not receive the necessary attention. For example, the SMS/USSD encryption ends in the radio interface, thereafter the message is often via plaintext.

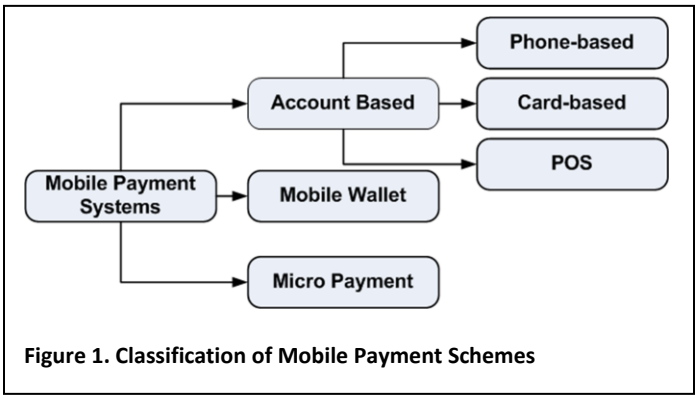


Figure 1. Classification of Mobile Payment Schemes

As a case study we will discuss the contextualization of security authentication for mobile payment applications motivated through an Indian context. Further we propose a context based framework for the development of such applications. Mobile wallet generally refers to payment services operated under financial regulations and performed either from, or via, a mobile device. Financial institutions, credit card companies, internet companies like Google, telecommunication companies, etc. now-a-day shave accepted using mobile payments as an alternative payment method to cash, checks or credit cards. Using this facility, a consumer can use a mobile phone to pay for a wide range of services such as music, videos, ringtones, online game subscription or items, transportation fares (bus, subway or train), parking meters and other such services. All over the world mobile payments have begun to grow and are being adopted in different ways. The combined market for all types of mobile payments is expected to reach more than double today's value and reach $600B globally by 2013. Additionally

contactless NFC (near-field communication) transactions are also rapidly evolving; money transfers through such mechanisms is expected to exceed $300B globally by 2013. In this context, the security of mobile payments which still remain a key customer concern needs to be addressed effectively. The availability of a more efficient trusted and secure means to enable issuers to provision wallets or other means of transaction mechanisms over the air to mobiles will undoubtedly pay huge dividends in the market place for organizations.

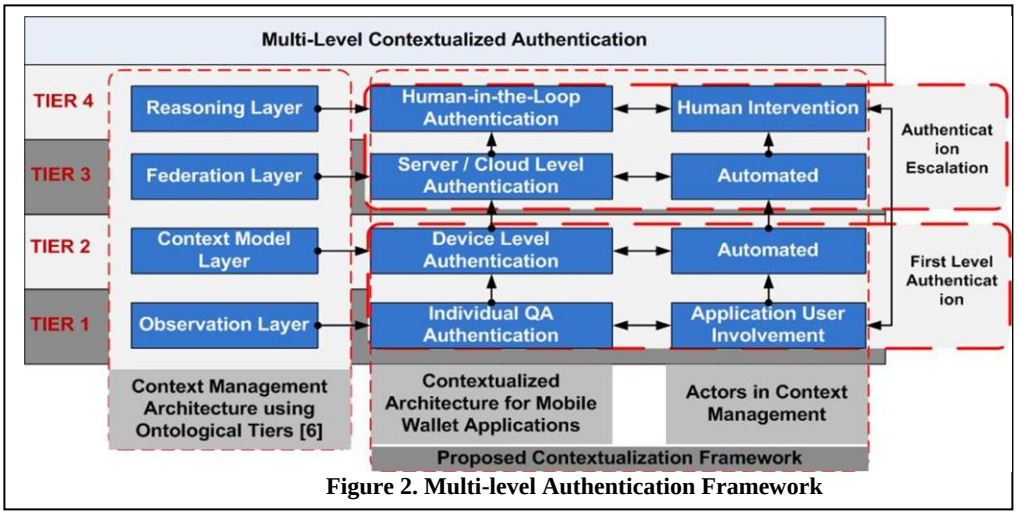


Figure 2. Multi-level Authentication Framework

In markets such as India, issues such as poor reliability and slower network speeds, etc. assumes greater significance. Furthermore, unlike many western countries, in several countries such as India, where there are deep layers of entrenched administrative bureaucracy, there are also issues with respect to using only personally identifying information (PII - such as name, DOB, father / mother's name, etc.) for authentication. For example, several tidbits of commonly used PII data for authentication are often widely available across multiple stakeholders – from schools, colleges, doctor's offices, other government service offices, etc. Hence there needs to be augment such authentication measures along with more 'private' measures of authentication. For example, many would very much remember who their favorite historical figures are in some ranked order, their first crush, their favorite heroes, heroines, or what was their most unforgettable day (not essentially birth day), etc.; roles / people / events, etc., that tend to resonate with a lot of the masses and thus be more personal and retained internally in their memories. Other such measures may be things such as favorite food, favorite movie, etc.

V. A Contextualized Model for Authentication for Mobile Wallet Applications

In this section we extend and adapt from the tired model developed in [6] to present a multi-level authentication framework for contextualized authentication of mobile wallet and related applications. Items in physical procession, especially in a mobile context, in certain regions of the world are more susceptible to theft and subsequent jail breaking to gain more access to the device. The 4-tier approach (shown in Figure 2), to user authentication presented in this section blends the three factors (i.e. *(i)* ***Knowledge***: something you know, *(ii)* ***Possession***: something you have, and *(iii)* ***Being***: something you are), for stronger user authentication in a multi-level mobility application authentication framework. As widely known, all these three methods are not fail-safe by themselves in every regional context, but combined within the 4-tier contextualization, they can provide a very resilient authentication framework for mobile applications. Such a mapping is presented in Figure 3.

Since the knowledge factor – for example, using information such as passwords, is acknowledged as a weak mechanism. Hence in our proposed framework we differentiate two types of PII knowledge – residual, imprinted and private PII knowledge of a mobile user from publically attainable PII information which is more widely available and needed for gaining essential societal services. In many regions, cultures and societies such private PII knowledge will not be widely known in many context beyond the user themselves (ex. date of their first date, name of their first crush, etc.). In our approach, we recommend utilizing such private PII knowledge for the first (Question-Answer based Authentication) QA tier, since we deem such private and personal knowledge to be of increasing value for user identification. While they may sound the same, there are important differences between authentication and identification. While authentication refers to the process of confirming or denying a person's claimed identity while identification refers to the process of establishing a subject's identity. Hence we propose to use such residual, imprinted and private knowledge at Tier-1 for rapid user identification. In certain contexts, these

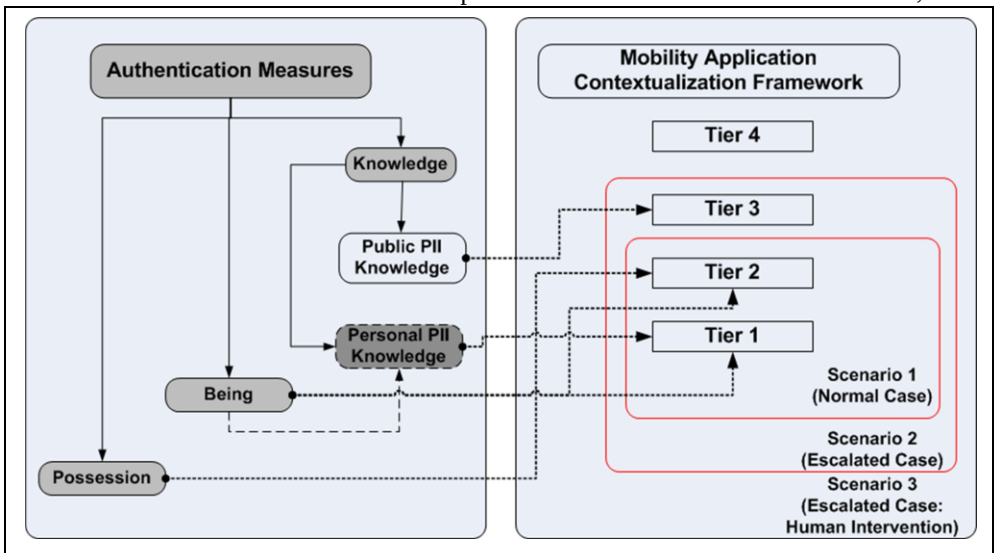


Figure 3. Authentication Measures Mapped to Mobile Contextualization Framework

may also be used in tier-3 at the server level for additional authentication. Such knowledge is different from PII information that a user might be required to enter in any other stored form that is not immediately under the user's control, etc. ex. for receiving a service – medical form information such as date of birth, blood group, or name, address, parents' name, previous addresses, etc. that may be needed in some university / college applications, etc. As shown in Figure 2, Tier (Level) 1 and Tier 2 based authentications therefore become a part of every interaction between the user and the mobile device. In all such interaction schemes, temporal and special activity profiling algorithms can be used to validate user accesses and actions.

In the proposed framework, at the Tier 1 level question-answer (QA) based authentication schemes are employed using private PII information, which can be effectively contextualized from a regional / cultural perspective. In this level, the system will not store any pre-structured information either in the device or the server and will effectively serve as the first line of authentication – for example using questions that are personally contextualized. For added complexity, reverse QA schemes can also be employed (where the user chooses the correct questions for a particular answer – information for which only the user knows the correct context).

Since this involves human interaction, use cases of appropriate scope and depth must be developed to capture the regional / localized contexts of this interaction. Tier 2 (Level 2) will involve device-level authentication measures, where in password-based, PIN-based, biometric-based, and multi-touch based authentication schemes can be effectively used to support the necessary authentication needs. Combined with the Tier-1 identification measures, this two-level scheme forms the first line of intrusion defense for any mobile device; especially in case of theft, which is rampant in developing countries. The mapping of strong authentication measures to the proposed tiered contextualization framework is shown in Figure 3. Figure 4 presents three different use-case scenarios that are specific to the mobile payment context for which appropriate context mapping models have to be developed.

Tier-3 (Level 3) authentication methods will be outside of the 'local (device and individual) system and will be effectuated at the server-level, and may include a combination of techniques that are automated. This can involve Capthas and other similar schemes that may include additional publically available PII (generic information stored at the server – during registration process, etc.) and a more general set of QAs that are not so deeply contextualized. However, the need for Tier-3 authentication will not be triggered until there are some alerts

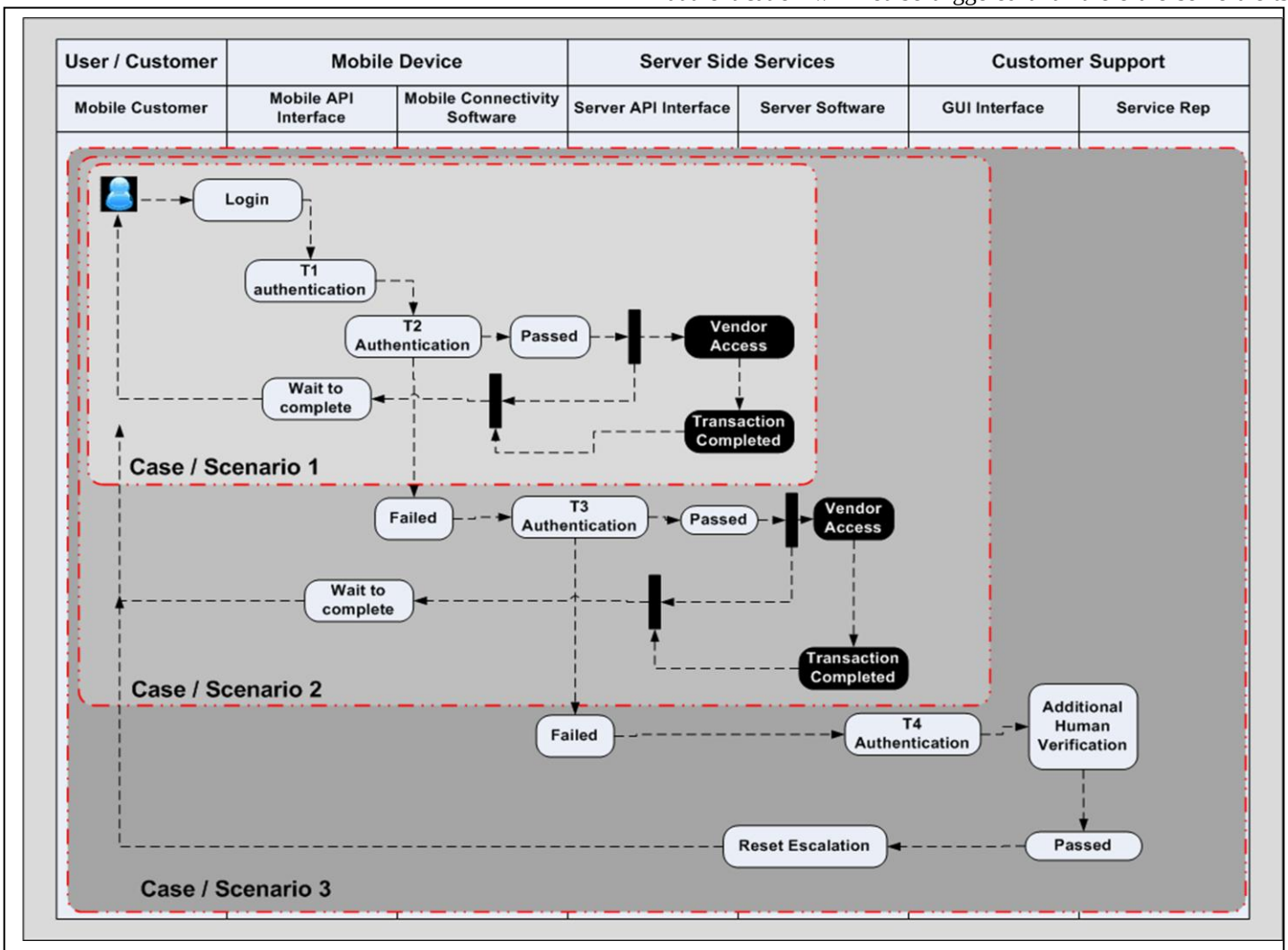


Figure 4. Normal and Escalated Use Case Scenarios

that are raised due to deviations from normal behavior patterns mined through low cost special-temporal algorithms at the device level. Upon unsatisfactory authentication at Tier-3 the system escalates the authentication need to involve a human in the loop who may be required to call an alternate number (home phone), or other kinds of schemes. Cumulatively, the approach presented in this paper, principally to distinguish public and private PII information to derive regional / localized contexts for user identification at the device level, can provide a scalable, light-weight, yet strong authentication measure for mobility applications.

VI. Conclusions

In this paper, we have proposed the need for contextualized, multi-level, multi-factor authentication mechanisms for user authentication in emerging mobile applications. We have proposed a notional framework for such authentication that accommodates deep personalization of authentication measures (possibly from regional and cultural perspectives). The tiered framework draws from existing literature on contextualization and adapts it to the emerging needs for mobile-based payment services.

However, it is clear that significant future work is needed to address the issue of personalization and contextualization and it assumes enormous business significance as the world rapidly advances to deploy mobile services for the common man. In countries such as India or similar developing countries, where some of the normally used QA based PII-driven authentication information are more easily attainable and where theft of personal mobility devices is significant, such approaches to addresses security and authentication is of dire significance to assure customers, and thereby build trust, to transition to using such systems. If simplicity of use, and user anxiety and apprehension in using these methods are adequately and appropriately addressed, such mobility-based payment systems carry significant potential for success as they will be incredibly useful for people in rural towns and villages, where there is already a very strong proliferation of mobility devices. Such technology might then allow for millions of users to buy into such services, and as a result banking institutions in such growing economies, may even leapfrog 'credit-card' based payment schemes, making them obsolete.

VII. References

1. C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, D. Riboni, "A Survey of Context Modelling and Reasoning Techniques", *Pervasive and Mobile Computing*, Elsevier, 2010
2. K. Henricksen, J. Indulska, A. Rakotonirainy, "Modeling context information in pervasive computing systems", *1st Intnl.Conf. on Pervasive Computing*, vol. 2414 of LNCS, Springer, 2002.
3. K. Henricksen, J. Indulska, "Modelling and using imperfect context information",*1st Workshop on Context Modeling and Reasoning (CoMoRea), PerCom'04 Workshop*, IEEE Comp. Society, 2004.
4. K. Henricksen, J. Indulska, "Developing context-aware pervasive computing applications: Models and approach", *Pervasive and Mobile Computing* 2 (1) (2006) 37–64.
5. T. A. Halpin, "Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design", *Morgan Kaufman*, San Francisco, 2001
6. C. Becker, D. Nicklas, "Where do spatial context-models end and where do ontologies start? A proposal of a combined approach", J. Indulska, D. D. Roure (eds.), *Proceedings of the First International Workshop on Advanced Context Modelling, Reasoning and Management, in conjunction with UbiComp 2004*, Nottingham, England: University of Southhampton, 2004.
7. A. U. Frank, "Ontology for Spatio-Temporal Databases". M. Koubarakis et al. (Ed): *Spatio-Temporal Databases— The CHOROCHRONOS Approach. LNCS*, Springer 2003
8. A. Leonhardi, K. Rothermel, "Architecture of a Large-scale Location Service", *Proceedings of the 22nd Conference on Distributed Computing Systems ICDCS*, Short Paper, 2002
9. ESRI, *Arclocation Solutions*, http://www.esri.com/library/brochures/pdfs/arclocation-sols.pdf
10. Intergraph, *Intelliwhere*, http://imgs.intergraph.com/intelliwhere/
11. Webraska, *Smartzone*, http://www.webraska.com/Products/
12. H. Lei, D. Sow, J. Davis II, G. Banavar, M. Ebling, "The design and applications of a context service", *ACM SIGMOBILE Mobile Computing and Communications Review*, volume 6(4), 2002.

Detection Techniques in MANET

Asma Ahmed[1], **S. Razak**[2], **A. Hanan**[2], **Izzeldin Osman**[3]
[1]Faculty of Computer Science and Information System, Universiti Technologi Malaysia, Johor, Malaysia
[2]Department of Computer Science Universiti Technologi Malaysia, Johor, Malaysia
[3]Faculty of Computer Science, Sudan University Science and Technology, Khartoum, Sudan

Abstract - *In the recent years, the security issues on Mobile ad hoc network (MANET) have become one of the primary concerns. Because of the mobility nature, MANET is more vulnerable to be attacked than wired network. Attack prevention measures, such as authentication and encryption, can be used as the first line of defense for reducing the possibilities of attacks. However, encryption and authentication cannot defend against compromised mobile nodes which carry the private keys. Intrusion detection techniques (IDS) the second mechanism to detect and response the attack which successfully penetrated the prevention mechanisms. The main categories of these techniques are anomaly-based detection, signature-based detection and specification-based detection. In this paper the different detection techniques- anomaly detection, signature detection and specification-based detection- are classified, as well as comparative discussion of these different techniques.*

Keywords: MANET, Anomaly Detection, Signature Detection, Specification-based Detection.

1 Introduction

The Internet and computer networks are exposed to an increasing number of security threats. Mobile ad hoc network (MANET) is vulnerable to security attacks due to its features of open medium, dynamic changing topology, cooperative algorithms and lack of cartelized monitoring. The flexibility provided by the open broadcast medium and cooperativeness of the mobile devices introduces new security risks. Security services, such as authentication services and access controls, can enhance the security of ad hoc networks. Nevertheless, these preventive mechanisms alone cannot deter all possible attacks (e.g., insider attackers possessing the key). Therefore, it is necessary to have other security mechanisms to deal with misbehaving insider nodes that possess the valid key and access rights. As result, intrusion detection is an indispensable part of security for MANET.

There are several techniques can be applied for the detection of attacks against routing protocols in MANET. The main categories of these techniques are: anomaly-based detection, misuse-based detection and specification-based detection [1]. These techniques apply to each of the routing protocols such as Ad hoc on-demand distance vector (AODV)[27], Dynamic source routing (DSR)[26], Optimized link state routing (OLSR)[10], and potentially other infrastructure protocols used in MANET.

The aim of this paper is to classify current techniques of Intrusion Detection System (IDS) and comparison between these detection techniques.

The paper is organized as follows Section 2 present the concept of prevention mechanisms extends with the limitation of these mechanisms. Section 3 provides an overview of IDS MANET as well as discusses the different detection techniques; these are anomaly-based, signature-based and specification-based. Comparison discussion between these detection techniques is presented in Section 4. Section 5 concludes the paper.

2 Prevention mechanisms

Prevention mechanism is used to secure network against external attacks, where it can be achieved by authenticating users and nodes [2][3][4], and by securing routing protocols used to create routes between nodes[5][6][7]. By signed the routing messages by each node, a large number of attacks can be prevented or eliminated, example of such attacks are:

-Spoofing attacks: Attacks in which nodes send routing messages pretending to be a different node.

-Modifying attacks: Attacks in which nodes modify routing messages in transit with the intention of misleading other nodes.

Modifying the routing protocols to require node authentication is a viable approach but has some limitations: firstly, it increases the overhead since it increases the size of routing messages and the amount of processing needed to process each routing message. Further, every node needs to verify the authenticity of the incoming routing messages.
However, these techniques in general they are designed for a set of known attacks. However, encryption and authentication cannot defend against compromised mobile nodes which carry the private keys. For this reason, there is a need of second mechanism to detect and response the attack that successfully penetrated the prevention mechanisms.

3 Intrusion Detection in MANET

Intrusion detection techniques (IDS) are a valuable technology to protect target systems and networks against malicious activities. Detection and response mechanisms are used to secure network against internal attacks. This can be achieved using intrusion detection systems [8][9]. IDS should be able to detect the malicious activities of attackers who successfully penetrated the prevention mechanisms.
Many intrusion detection systems have been proposed in traditional wired networks, where all traffic must go through switches, routers, or gateways. Hence, IDS can be added to and implemented in these devices easily [12][13]. On the other hand, MANET does not have such devices. Moreover, the medium is wide open, so both legitimate and malicious users can access it. Furthermore, there is no clear separation between normal and abnormal activities in a mobile environment. Since nodes can move arbitrarily, false routing information could be from a compromised node or a node that has outdated information. Thus, the current IDS techniques on wired networks cannot be applied directly to MANET.
Some assumptions are made in order for the intrusion detection systems to work in MANET [14]. The first assumption is that user and program activities are observable. The second assumption, which is more important, is that normal and intrusive activities must have distinct behaviors, as intrusion detection must capture and analyze system activity to determine if the system is under attack.
Several techniques can be applied for the detection of attacks against routing protocols. These techniques can be divided into the main categories of: anomaly-based detection, signature-based detection and specification-based detection [1]. These techniques apply to each of the routing protocols such as AODV, DSR, OLSR, and potentially other infrastructure protocols used in MANET (e.g. multicast, session management).

3.1 Anomaly-Based Detection

An Anomaly-based intrusion detection System, is a system for detecting computer intrusions by monitoring system activity and classifying it as either normal or anomalous. This technique tries to detect attacks by looking at activities that vary from the normal expected behavior. Detection techniques can be classified into three main categories [25]:

- Statistical-based.
- Knowledge-based.
- Machine learning-based.

Defining normal behavior major challenge in anomaly detection. Normal behavior can change over time and intrusion detection systems must be kept up to date. False positive – the normal activities that are detect as anomalies by IDS – can be high in anomaly-based detection. On the other hand, anomaly-based detection is capable for detecting previously unknown attacks. This is very important in an environment where new attacks and new vulnerabilities of system are announced constantly.

3.2 Misuse-Based Detection

Misuse-based or signature-based intrusion detection compares known attack signatures with current system activities. Generally misuse-based intrusion detection preferred by commercial IDSs since it is efficient and has a low positive rate.
Misuse detection provides very good detection results for specified, well-known attacks. However, they are not capable of detecting new, unfamiliar intrusions, even if they are built as minimum variants of already known attacks. The system is only as strong as its signature database, and this needs frequent updating for new attacks.

3.3 Specification-Based Detection

The specification-based intrusion detection technique [15] is usually based on building finite state

machines that reflect the expected behavior of the node. The implementation of this idea can be done by monitoring execution such program or protocol respect to the expected behavior. Specification-based intrusion detection technique is introduced as promising alternative that combines the strengths of anomaly-based and misuse-based detection techniques, providing detection of known and unknown attacks with lower false positive rate. It can detect new attacks that do not follow the system specifications. Moreover, it does not trigger false alarm when the program or protocol has unusual but legitimate behavior, since it uses the legitimate specifications of the program or protocol [28]. In specification-based a detector needs to monitor a node very closely and maintain information about the messages sent and received by the node. The detector then also needs to perform similar calculations as performed by a node executing the routing protocol. Therefore the complexity of the detector is typically similar to the complexity of executing the routing protocol itself. This increases the detector complexity, and the data that needs to be stored on the detector to a level that may not be acceptable. Another problem with specification-based detection is danger from misinterpretation of the protocol when the protocol is modeled in detail. This will lead to false alarms because the alarm may be due to a misinterpretation of the protocol in the detector finite state machine. To avoid this problem it has been proposed to simplify detectors by only modeling key characteristics of the protocol and not necessarily every detail. This decreases complexity and simplifies the detector but leaves open the possibility that an attack exploiting the portions of the protocol behavior that are not modeled by the detector will go undetected.

4 Analysis

IDSs on MANET use a variety of intrusion detection methods. The previous discussion illustrates the challenges associated with each of intrusion detection techniques. Anomaly-based detection systems implemented in MANET. Unfortunately, mobility of MANET increases the rate of false positives in these systems. The main benefit of anomaly-based detection techniques is their potential to detect previously unseen intrusion events. Misuse-based detection provides very good detection results for specified, well-known attacks. However, they are not capable of detecting new, unfamiliar intrusions, even if they are built as minimum variants of already known attacks. Updating of attack signatures is an important problem for this approach. Specification-based detection can detect routing attacks against routing protocols with low rate of false positives. However, it cannot detect some kind of attacks, such as DoS attacks. Table1 illustrate the characteristics of each technique.

Table1: Anomaly, Misuse and specification detection

Method	Characteristics
Anomaly-based	- capable for detecting previously unknown attacks - High false positive
Misuse-based	- very good detection results for specified, well-known attacks. - not capable of detecting new attacks.
Specification-based	- detection of known and unknown attacks. - low false positive. - needs to monitor a node very closely. - complsexity of the detector. - Cannot detect DoS attacks

Two key aspects concern the evaluation, and thus the comparison, of the performance of alternative intrusion detection approaches: these are the efficiency of the detection process, and the cost involved in the operation.

5 Discussion and Summary

Mobile ad hoc networks are attractive technology for many applications. However, this flexibility introduced new security risks. Since prevention techniques are never enough, intrusion detection systems (IDSs) are generally used to complement other security mechanisms. Intrusion detection for MANET is complex and difficult task mainly due to the dynamic nature of MANETs, their highly constrained nodes, and lack of central monitoring points. These different characteristics of MANET make conventional IDSs ineffective and inefficient for this new environment.

Recently, many researchers have been working on developing new IDSs for MANET. New approaches need to be developed or else existing approaches need to be adapted for MANETs. This paper, briefly explored the various intrusion detection methods suggested by the authors and also analyzed some challenges and problems of each method in MANET. There is an utmost need of a general foundation for all intrusion detection and supporting activities that can able to adapt dynamic network conditions. The requirement of the system like high security and low bandwidth should be satisfied by the IDS. Also the IDS that are able to detect known and unknown attacks should be considered.

References

[1] A. Mishra, K. Nadkarni, and A. Patcha. "Intrusion Detection in Wireless Ad Hoc Networks". IEEE Wireless Communications, Vol. 11, Issue 1, pp. 48-60, February 2004.

[2] Douceur, J. R. (2002). The Sybil Attack. In IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems. London, UK: Springer-Verlag. ISBN 3540441794.

[3] Capkun, S., Buttyan, L. and Hubaux, J.-P. (2003). Self-organized public-key management for mobile ad hoc networks. IEEE Transactions on Mobile Computing. ISSN 1536-1233.

[4] Yi, S. and Kravets, R. (2003). MOCA: Mobile Certificate Authority for Wireless Ad Hoc Networks. In 2nd Annual PKI Research Workshop Program (PKI 03.

[5] Xu, Y. and Xie, X. (2008). Security analysis of routing protocol for MANET based on extended Rubin logic. Sanya, China.

[6] Kim, J. and Tsudik, G. (2009). SRDP: Secure route discovery for dynamic source routing in MANETs. Ad Hoc Networks. ISSN 15708705.

[7] Yu, M., Zhou, M. and Su, W. (2009). A secure routing protocol against byzantine attacks for MANETs in adversarial environments. IEEE Transactions on Vehicular Technology. ISSN 00189545.

[8] Rao, R. and Kesidis, G. (2003). Detecting malicious packet dropping using statistically regular traffic patterns in multihop wireless networks that are not bandwidth limited. Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE. 5,2957-2961.

[9] Subhadrabandhu, D., Sarkar, S. and Anjum, F. (2004). E_cacy of misuse detection in ad hoc networks. Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on, 97-107. doi: 10.1109/SAHCN.2004.1381907.

[10] T. Clausen, P. Jaquet, et.al. "Optimized link state routing protocol". Internet Draft, draft-ietfmanet-olsr 06.txt, work in progress, 2001.

[11] G. Vigna, S. Gwalani, et al., "An Intrusion Detection Tool for AODV-based Ad hoc Wireless Networks," in Proceedings of the Annual Computer Security Applications Conference (ACSAC), Tucson, AZ, December, , pp. 16–27 2004.

[12] Y. F. Jou, F. Gong, et al.. "Design and Implementation of a Scalable Intrusion Detection System for the Protection of Networks Infrastructure". Proceedings of DARPA Information Survivability Conference and Exposition, Vol. 2, pp. 69-83, January 2000.

[13] E. Y. K. Chan et al., "IDR: An Intrusion Detection Router for Defending against Distributed Denial-of-Service (DDoS) Attacks". Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04), pp. 581-586, May 2004.

[14] Y. Zhang, W. Lee, and Y. Huang. "Intrusion Detection Techniques for Mobile Wireless Networks". ACM/Kluwer Wireless Networks Journal (ACM WINET), Vol. 9, No. 5, September 2003.

[15] C.Y. Tseng, P. Balasubramanyam, et al., "A Specification-Based Intrusion Detection System For AODV," in Workshop on Security in Ad Hoc and Sensor Networks (SASN)'03, 2003.

[16] C.K.Toh, "Ad Hoc Mobile Wireless Networks: Protocols and Systems," Prentice Hall Publications, 2002.

[17] S. Lee, B. Han, and M. Shin, "Robust routing in wireless ad hoc networks," in ICPP Workshops, pp.73, 2002.

[18] Y. A. Huang and W. Lee, "Attack analysis and detection for ad hoc routing protocols," in The 7th International Symposium on Recent Advances in Intrusion Detection (RAID'04), pp. 125-145, French Riviera, Sept. 2004.

[19] X. Wang, T. Lin and J. Wong, "Feature selection in intrusion detection system over mobile ad-hoc network," Technical Report, Computer Science, Iowa State University, 2005.

[20] J. Binkley and W. Trost. "Authenticated ad hoc routing at the link layer for mobile systems". Wireless Networks, 7(2): 139–145, 2001.
[21] Y. C. Hu, A. Perrig, and D. B. Johnson, "Ariadne: A secure on-demand routing protocol for ad hoc networks," in Eighth Annual International Confer-ence on Mobile Computing and Networking (Mobi- Com 2002), pp. 12-23, Sept. 2002.
[22] K. Sanzgiri, D. LaFlamme, B. Dahill, B. N. Levine, C. Shields, and E. M. B. Royer, "Authenticated routing for ad hoc networks," IEEE Journal on Selected Areas in Communications, vol. 23, no. 3, pp. 598-610,Mar. 2005.
[23] Y. C. Hu and A. Perrig, "A survey of secure wireless ad hoc routing," IEEE Security & Privacy Magazine, vol. 2, no. 3, pp. 28-39, May/June 2004.
[24] H. Deng, W. Li, and D. P. Agrawal, "Routing security in ad hoc networks," IEEE Communications Magazine, vol. 40, no. 10, pp. 70-75, Oct. 2002.
[25] Lazarevic A, Kumar V, Srivastava J. Intrusion detection: a survey, Managing cyber threats: issues, approaches, and challenges. Springer Verlag; 2005. p. 330.
[26] D.B. Johnson, D.A. Maltz, et.al. "The dynamic Source Routing Protocol for Mobile Ad hoc Networks (DSR)". Internet Draft, draft-ietf-manet-dsr-07.txt, work in progress, 2002.
[27] C.E Perkins, E. Belding-Royer. "Ad hoc On-demand Distance Vector (AODV)", Request For Comments (RFC) 3561, 2003.
[28] uppuluri P, Sekar R. "Experience with Specification-Based Intrusion Detection. In Proc of the 4th Int Symp on Recent Adv in Intrusion Detect LNCS 2212: 172-189. 2001.

The Worst Path Estimation for Real Time Communications over Ad hoc Networks

J. Bokri[1], **S. Ouni**[1], **and F. Kamoun**[1]
[1]Cristal Laboratory, ENSI, Manouba, Tunisia

Abstract- *The hard real time guarantee is one of the important topics which have received a lot of attention recently. The most important requirement in such cases is the end-to-end delay. This delay depends on the path connecting the source and the destination nodes. For that reason, we are interested to estimate the worst path which will lead us to estimate the delay in the worst case by using a depth decision tree.*

Keywords: Hard real time, ad hoc network, path, worst case.

1 Introduction

Nowadays, there is a subject that gets the attention of a lot of researchers which is the transfer of the hard real time data over the Ad hoc networks.

The most important requirement to assure the hard real time transfer is the end-to-end delay which should not exceed the deadline. To guarantee this delay, we have to focus on the worst case. In fact, if the delay in the worst case is inferior to the deadline, we can be sure that the deadline can never be exceeded [1][2][3]. However, this verification is not an easy task for the mobile Ad hoc networks where the paths connecting the source and the destination nodes can suddenly change. For that, many researches are interested to predict the new paths [4][5]. In our case, we are focusing on estimating the path between the source and the destination for nodes having known and cyclic trajectories.

2 Heuristic worst path estimation

Between the source of real time data and its destination, we can find several paths. In our case, the worst path is the longest one between those two nodes. In that case, the route passes by the intermediate nodes which are the furthest from the straight line linking the positions of the source node and the destination node, as explained in Figure 1.

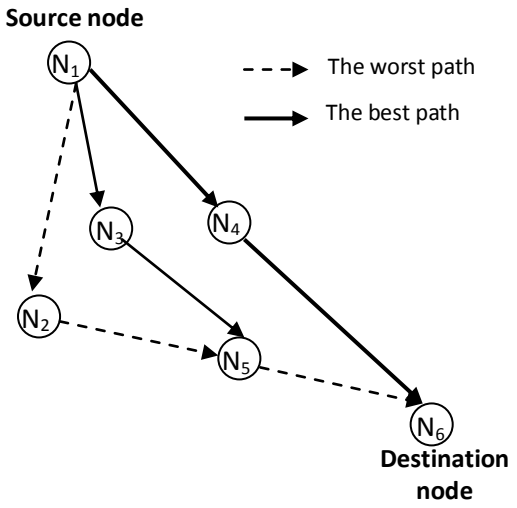


Figure 1: The worst path

In Figure 1, N_1 is the source node, N_6 is the destination node and N_2, N_3, N_4, N_5 are the intermediate nodes. The best path is (N_1, N_4, N_6) and the worst path is (N_1, N_2, N_5, N_6).

In order to explain the path estimation method, we will consider the following annotations:

$$\boldsymbol{V(N_i)} = \left\{ \boldsymbol{N_j(x_j, y_j)} / \boldsymbol{distance}\left(\boldsymbol{(x_i, y_i), (x_j, y_j)} \leq \boldsymbol{R} \right) \right\} \quad \textbf{(1)}$$

Where:

- $\boldsymbol{V(N_i)}$: The neighboring nodes of $\mathbf{N_i}$
- **R**: The range of the nodes (We assume that all the nodes have the same range).
- $\boldsymbol{D_{iD}}$: A straight line linking the current node N_i and the destination node.
- $\boldsymbol{D_{ij}}$: A straight line linking the current node N_i and its neighbor N_j
- $\propto_{\mathbf{ij}}$: angle between $\boldsymbol{D_{iD}}$ and $\mathbf{D}_{ij}$

In order to estimate the worst path, we will use a depth decision tree; It looks for the path which maximizes the distance and the number of intermediate nodes (hops) to the destination. Our heuristic (noted: **H**)

chooses the intermediate nodes with more important angle (noted: $\boldsymbol{\alpha_{ij}^*}$) to the straight line to destination. Moreover, it maximizes the number of intermediate nodes by choosing the nearest neighbors (having minimum $\mathbf{d_{ij}^*}$ and therefore maximum $\frac{\mathbf{1}}{\mathbf{d_{ij}^*}}$). Thus, $H(N_i)$ is given by the formula (2).

$$\mathbf{H(N_i)} = (\boldsymbol{\alpha_{ij}^*}, \frac{\mathbf{1}}{\mathbf{d_{ij}^*}}) \quad \textbf{(2)}$$

Where α_{ij}^* and d_{ij}^* are given respectively by the formulas (3) and (4).

$$\boldsymbol{\alpha_{ij}^*} = \mathbf{max_j}\left(\propto_{\mathbf{ij}}/\mathbf{N_j} \in \mathbf{V(N_i)}\right) \quad \textbf{(3)}$$

$$\mathbf{d_{ij}^*} = \mathbf{min_j}\ (\boldsymbol{distance}\ (\boldsymbol{Ni, Nj})/\mathbf{N_j} \in \mathbf{V(N_i)})) \quad \textbf{(4)}$$

The method of creating the depth decision tree which allows us to estimate the worst path is the following:

- For each node $\boldsymbol{N_j}$ belonging to $\mathbf{V(N_i)}$
  - Adding a new branch below root so that the neighboring node having a large value of $\mathbf{H(N_i)}$ must be on the left side and so on.
  - If $\boldsymbol{N_j}$ is not the destination node, then adding a new tree branches below $\boldsymbol{N_j}$ with considering $\boldsymbol{N_j}$ as a root and $\mathbf{H(N_j)}$ as attribute.

From the constructed decision tree, we can estimate the worst path by starting from the source node and looking for the closest branches to the left which lead us to the destination node. Every time we get stuck, we have to go back to the previous node and then look for a different branch again and so on.

Once we find all our branches that lead us to the destination, we can then calculate the maximum delay, in which the data will reach the destination.

3 The delay in the worst case

In order to determine the end-to-end delay, we should know the maximum delay for one hop which is the delay to transfer data between two neighboring nodes. However, in most of the TDMA-based Ad hoc networks, the maximum delay in one hop is the TDMA super-frame length. Thus, the end-to-end delay is equal to the number of intermediate nodes in the estimated path times the worst delay in one hop.

This way, the delay in the worst case (noted : $\boldsymbol{D_w(N_S, N_D)}$) to transmit data between the source node $\boldsymbol{N_S}$ and the destination node $\boldsymbol{N_D}$ is given by the formula (5).

$$\boldsymbol{D_w(N_S, N_D)} = \boldsymbol{NP} \times \boldsymbol{TF} \quad \textbf{(5)}$$

Where :

- **NP** is the total number of nodes belonging to the estimated path noted: $\boldsymbol{P_w(N_S, N_D)}$ between the source node $\boldsymbol{N_S}$ and the destination node $\boldsymbol{N_D}$.
- **TF** is the transmission frame in the TDMA-based Ad hoc networks.

4 Conclusions

In this paper, we have presented a new method to estimate the worst path between the source and the destination nodes in the worst case. For that, we used a depth decision tree. In future works, we will evaluate our approach based on simulations and mathematic demonstrations.

5 References

[1] S. Ouni, J. Bokri, and F. Kamoun, "DSR based routing algorithm with delay guarantee for Ad hoc networks," in JOURNAL OF NETWORKS, Academy Publisher, vol. 4 Issue : 5, pp. 359–369, July 2009.

[2] J. Bokri, S. Ouni and F. Kamoun, "End-to-end delay guarantee for TDMA-based Ad hoc networks with RT-DSR protocol, " in The Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2009), Sliema -Malta, October 11-16, 2009.

[3] Jihen Bokri, Sofiane Ouni, Farouk Kamoun, "A Novel Reservation Approach for TDMA-based Ad hoc Networks", in The Second International Conference on Communications and Networking (ComNet'2010), Tozeur, Tunisia, 4-7 November 2010.

[4] Hamid Mehdi, " MOBILITY PREDICTION WITH LLT ALGORITHM IN WIRELESS NETWORKS", in The International Conference on Information, Networking and Automation (ICINA), Puerto Rico, USA. October 21-23, 2010.

[5] Jianping Wang, " Exploiting Mobility Prediction for Dependable Service Composition in Wireless Mobile Ad Hoc Networks", IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 4, NO. 1, January-March 2011.

SESSION

SENSOR NETWORKS

Chair(s)

TBA

Load Balancing Algorithm for Wireless Sensor Networks

S. Wijedasa, S. Rizvi, and K. Ferens
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada

Abstract - *This paper presents an energy-efficient load-balancing algorithm for Wireless Sensor Networks. The algorithm specifies that only the nodes which have residual energy above a threshold can take part in routing, and that residual energy is to be updated at the highest opportunistic rate supported by the network. The algorithm was tested on and showed improvements to hierarchical data aggregation and basic directed diffusion; the proposed method updated residual energy during reinforced data transmission, the highest data rate in both protocols, by allowing neighbor nodes to promiscuously snoop traffic. Simulations show that the proposed method has lower residual energy variance (25%, 23% less) and longer network lifetime (10%, 23 % longer) than hierarchical data aggregation and basic directed diffusion methods, respectively. In addition, interest message flooding was reduced and network lifetime was increased by allowing only the nodes that have residual energy above a threshold to take part in interest propagation.*

Keywords: wireless sensor network; load balancing; network lifetime; energy consumption; energy efficient routing.

1 Introduction

In many WSN applications, such as environmental monitoring, a large number of energy conserving nodes are required to meet the requirements. For instance, in military applications, airborne vehicles sprinkle thousands of nodes over a battlefield for monitoring, surveillance, and reconnaissance purposes. Such nodes will not have the opportunity to replenish their power sources, since grid power (or other sources) may not be available. In addition, such nodes will be expendable due to the nature of their deployment. As such, these nodes must have low cost, small physical size, reduced functionality and short radio coverage. Most importantly, these nodes must consume low energy to extend their lifetime to a fullest possible extent.

One approach to meet the requirement of extending lifetime is to design energy efficient routing algorithms that have the objective to balance the work load among the nodes in the network. A load balancing algorithm aims to balance work load among each node in the network in a uniform or approximately equal manner. Ideally, when the workload of a node is the same as that of other nodes, then the residual energy of each node will decrease at the same rate with network operation. Over time, where there is no possibility to replenish power supplies, the network will cease to operate (die) when all nodes (die) reach zero energy at the same time. While the uniform balancing of tasks across all nodes in the network may not extend the lifetime of an individual greedy node to a fullest possible extent, such as in the case of a lazy node that does no work, load balancing in an approximately uniform manner will extend the lifetime of the network to the fullest possible extent. With uniform load balancing the network will survive the longest and the purpose of the WSN will be prolonged to its fullest extent. The needs of the many exceed that of the one. The nodes cooperate to achieve a greater system function.

The remaining parts of the paper are organized as follows: section 2 discusses related work and identifies the extensions and contributions of this work. Section 3 gives the details of the methods to improve load balancing among nodes in the network. Sections 4 and 6 discuss the simulation experiments performed to compare the proposed algorithms with other leading algorithms. Finally, conclusions and future work are given.

2 Related work

Many routing protocols have been proposed to efficiently manage the consumption of energy in WSNs. A large number of these protocols are extensions of the data centric protocol known as Directed Diffusion (DD) [1].

DD has four phases: interest propagation, exploratory data, reinforcement message, and reinforced data. A sink begins the process by expressing an interest for some data by flooding the network with a data interest packet. Each node that receives an interest packet establishes a gradient, from itself, directed towards the sending node. When a source node receives an interest for some data, for which its sensors support, the source node sends exploratory data packets to the sink along the gradients previously established by the interest propagation. The gradients act like wells for attracting data from a source to a sink. When the sink receives exploratory data packets, the sink responds by reinforcing the node from which it received the first exploratory data packet (This is the basic (simplest) DD behavior). The reinforced node will, in turn, reinforce the node from which it received the first exploratory data packet. In this way the least-delay path (gradient path) between the source and the sink is established. Finally, the source node sends reinforced data packets along the discovered least-delay reinforced data path to the sink.

This basic form of DD aims to minimize the delay in data packet transmissions. However, basic DD exhibits high energy consumption, high traffic, and unequal load distribution of the nodes. Most significantly, the nodes along the least-delay path will die prematurely because they will consume energy at a greater rate than other nodes.

Recently, much work has been done applying load balancing to WSNs [2] and, more specifically, directed diffusion. In [3], multiple paths are constructed between the sink and the sources. Data are spread and distributed along different paths according to several energy-related parameters, and, as such, the load across different nodes tends to balance. One of the problems of this and other multiple path approaches are that an insufficient amount of paths are generated due to the multiple distinct routes merging at some point in the network. While the merging of distinct routes can be seen as an advantage in terms of increasing the data aggregation capability, the nodes in the merged path will consume more energy than others; consequently, the network will be unbalanced and die prematurely. To overcome this problem, [4] proposed to force inactive nodes to take part in a routing path according to a local (greedy) algorithm. Similar to this work, the work in [5] proposed a multipath routing scheme, which creates several paths between source and sink. A node is selected to be on a path if it has high residual energy and large distance (low signal strength) from the sender. The rationale behind selecting distant nodes is to reduce the amount of hops that would otherwise result in if closer nodes were selected. However, in this algorithm, it is possible that distant nodes can be selected even though a more energy-efficient path to the source may exist. In this case, the nodes along the longer path would expend more energy, and thus, this would lead to premature network death. The work proposed in the present paper is similar to [5], but it differs by not choosing to select nodes based on their relative distance from the sender. This paper uses only residual energy to establish paths, since energy consumption is the primary parameter to preserve. Another approach taken to reduce the energy consumption of DD is to reduce the amount of interest packets and exploratory data packets (path discovery traffic) sent in the network. To reduce the path discovery traffic in the network, a possible method is to impose a sending restriction by creating a hierarchical structure, whereby in each layer, the parent nodes can only send traffic to their direct children and vice versa. Instead of flooding the network with interest packets, as is done in basic DD, a hierarchical structure limits the receiving nodes and thus reduces traffic; however, the trade-off is that the reliability is reduced as compared to flooding. One example is the Hierarchical Data Aggregation (HDA) scheme described in [6]. The HDA algorithm first forms a hierarchical layer structure during the interest propagation phase. During this phase each node will identify its corresponding parents and children. Hence, the structure limits communication between only parents and children of any two neighboring layers. In a network where sources were carefully placed at the last layer of the node hierarchy, HDA was able to achieve significant improvements compared to DD. However, simulations performed by the present work showed that HDA performs unsatisfactorily in terms of load balancing, as there was high variance in the distribution of residual energy at the time of first node death. Furthermore, the variance in the distribution of residual energy was even greater for realistic node arrangements where sources were situated at any layer of the structure.

The present paper extends the work of HDA by using latest residual energy information of the nodes at a higher rate instead of source count to establish parent-child relationships and thus routing paths. Furthermore, this paper implements the extension in a more realistic hierarchical structure where sources were randomly scattered throughout the network hierarchy at any layer of the structure. Through the mentioned extensions, this paper achieves a better degree of energy balancing among nodes and higher network life time as compared to the original HDA. Also similar extensions will be utilized to improve the basic form of DD as well.

3 Improved HDA scheme description

Following subsections describe the phases of the proposed scheme by highlighting the key differences and the similarities with the original version of HDA.

3.1 Interest propagation and hierarchy formation

The sink node initiates the process by periodically broadcasting an interest packet of a specific data type, the type of data in which it is interested. The sink is assigned as a "level 0" node in the structure. The sink includes a "level 1" indicator in the interest packet, which the receiving nodes use to establish their position in the structure. In other words, the children of the sink are positioned at "layer 1." The receiving nodes will add the sink as their parent and relay the interest packets towards their neighbors, informing them that their position in the structure is "level 2." In general, a receiving node establishes its level from the level indicator of the received packet ("level N"); adds the sending node to its parent list; and relays the interest packets to its neighbors, informing them of their position in the structure is "level N+1." This process continues until all nodes have been assigned a level in the structure, and each node will have discovered its parents. When a node receives its first interest, it determines its level to be the level indicated in the packet (say N), and records the neighbor sending this interest to be its parent. After receiving the initial interest, if another interest message containing the same level indicator (N) is received, the node adds the sender of such message to its parent list. If a received interest message contains a level indicator N+2, then the message must have been sent from the next highest level (N+1) node, and, therefore, the sender should be its child. Any other interests with different level numbers (i.e. neither N nor N+2) should be ignored. Thus each node will acquire knowledge about all its parents and

children. For more details see [6].

In addition to the level information and other parameters embedded within interest messages, this paper proposes that each sending node includes its residual energy in each packet it sends. Due to the broadcast nature of wireless transmission, each node can learn the residual energy of their neighboring nodes without any additional transmission, and without having the message addressed to them. Ultimately each parent node will learn the residual energy levels of its children and vise versa. The residual energy is sent in Reinforcement Message Packets (RMPs), but also in Reinforced Data Packets (RDPs). Including residual energy in RDPs is also another contribution of this paper since other work has not considered this advantage. Since the RDPs are the most frequent messages sent, every node will have the most frequently update of residual energy of other nodes. Furthermore, another contribution of this work is that a sink will send Interest Message Packets (IMPs) only to those nodes that have an energy level above an empirically determined threshold. This decision is to limit propagating an IMP only to those nodes with sufficient energy might seem to overly limit the reduction of an already limited set of recipient nodes (children of the parent) and potentially cause a termination of interest propagation. However, this would not be the case because, if a node has energy lower than a certain threshold, it should not be given the task to receive, process, and relay an IMP, since it will most likely deplete its energy, die, and cause the network to cease operation. The simulation shows that this approach has better load balancing and longer network lifetime. As an alternative to broadcasting, a multicasting approach could be adopted and will not affect the comparison purposes with the original version of HDA or with basic directed diffusion (where they broadcast to every neighbor) as the energy dissipation is handled in a similar manner. This paper denotes this extension as "INT Energy Piggyback with INT Threshold" as will be referred to in the Experimental Results Section.

3.2 Exploratory data propagation

If a node receives an interest message, and if the node supports that data type, i.e., the node is a source, then it will initiate the exploratory data sending process periodically. The exploratory messages will be propagated to the parent nodes only. The period of sending exploratory data will be extracted from the interest messages. The exploratory data sending frequency will be less than the interest propagation frequency. This paper models message propagation as a unicast transmission in order to align with DD for comparison purposes, even though a much better improvement can be achieved if multicast message transmission was adopted.

Another extension of this paper is that, at the exploratory stage of routing, since every node is aware of the current residual energy of their parents due to the inclusion of that information in interest packets, each node will adopt the energy threshold approach during each periodic transmission of exploratory data. In other words, a node (which includes a source) will unicast only to those of its parents which have residual energy above a given threshold, and thus balance the energy among its parents. This paper refers to this approach as "INT Energy Piggyback with exploratory data Threshold," as will be referred to later sections. Also, to compare with the original HDA, each node that receives an exploratory data message from a distinct source increments the source count thus establishing source node count in its node sub tree.

3.3 Reinforcement message propagation

Once the sink receives exploratory data corresponding to an interest that it sent previously, the sink will initiate Reinforcement Message Packet (RMP) propagation. In this phase every parent node includes either the source count details that they learnt previously, its residual energy, or both source count and residual energy in the reinforcement message packet. This is done so that their children can acquire this information and use it to decide which parent to send reinforced data, accordingly, later, in the reinforced data packet stage. The RMP frequency is the lowest among all message delivery frequencies, and, as such, the nodes will be updated with this information at the slowest possible rate, and will have old details about their parent nodes. The simulations show that this leads to unsatisfactory load balancing as the nodes are relying on old data. This is the exact approach adopted in old version of HDA. This extends that work by updating residual energy in the fastest possible manner, in the reinforced data stage, as discussed earlier. This paper denotes this process as "RMP Source Count Piggyback and RMP Energy Piggyback," respectively.

3.4 Reinforced data propagation

Once a corresponding source receives a RMP for a supported data type it will initiate the Reinforced Data Packet (RDP) propagation process. This is the fastest message propagation process among all message propagations. That is, the periodic rate of sending RDPs is the fastest compared to IMP, EDP, and RMP stages. The corresponding interval will be obtained from the RMPs.

This paper recognized that the residual energy was being updated at the slowest rate possible in the original HDA work. Since the RDP process is the fastest message propagation process, residual energy update would be greatly increased compared to the residual energy updated at the RMP rate. This together with the fact that radio transmission is inherently broadcast, this paper proposes to allow nodes to snoop the reinforced data packets to extract the residual energy of their neighbors, when their neighbors, acting as children, sending reinforced data to their parents.

Thus, in this phase every child node includes either the source count details that they learnt previously, its residual energy, or both source count and residual energy in the reinforced data packet. All neighboring nodes will "hear" and record this information. Children, of the children that send reinforced data packets to their parents, use this information to decide which parent to send subsequent reinforced data

according to the parent node which has the highest source count, to the parent node that has the highest energy, or a combination of both sourced count and highest residual energy. The receiving parent can aggregate the data received from its children nodes and relay to its parents and so on. Simulation shows that by updating the residual energy during the reinforced data packet stage, the variance of the load of the nodes in the network is much better balanced and the network lives longer.

4 Experimental results

A Java simulation was created to test the energy efficiency of the improvements made to the original version of HDA scheme. The results were compared with the original version. A *25m* by *25m* grid was created, and different numbers of nodes, ranging from *N = 360, 400, ..., 600* were placed in this area. The source to node ratio was maintained at 2.5%. Each node was given equivalent resources; same type and speed of processor; same amount of memory for buffering; radio range of *R = 4m*; and initial energy of $E=2x10^4$ Joules. During the operational state of the network, the energy consumption was set as follows: E_tx = 2 Joules for transmission and E_rx = 1 Joule for reception of each type of packet. The event triggering times for interests, exploratory data, reinforcements and reinforced data were 5 s, 10 s, 15 s and 1 s respectively.

4.1 Residual energy variance

The effect of the proposed modifications with respect to energy variance of the network was analyzed and compared with the original HDA protocol. Fig. 1 shows the variance of residual energy of the nodes in the network as a function of the number of reinforced data sent in the network. As expected the variance increases with increasing reinforced data transmission, since not all nodes can take part in the transmission of reinforced data, due to their physical location in the network. The basic HDA protocol updates source count (or residual energy) at each transmission of reinforcement message packets, which has the lowest periodic rate compared to IMP, EDP, and RMP phases. As shown in Fig. 1, the proposed method of updating residual energy at each reinforced data packet stage has a much lower variance than that of the original HDA protocol, because of the increased rate of updating residual energy. Plus, delivering packets only to the nodes having a energy level beyond a threshold level contributed to further lowering the variance. For example such a threshold was considered in IMP stage.

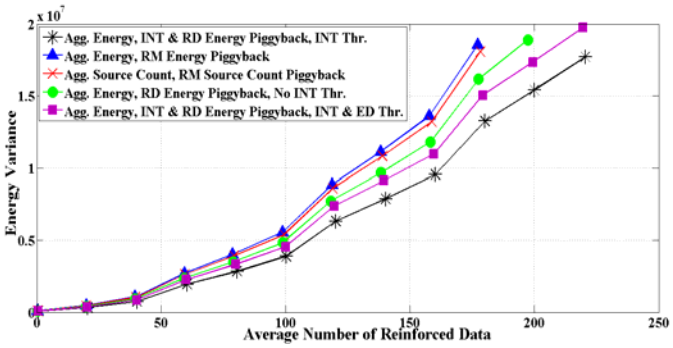


Fig. 1. Variance of residual energy vs average number of reinforced data events generated within the network.

4.2 Network life time

The lifetime of the network was determined by measuring the total number of reinforced data messages that the network nodes were able to deliver before they died, i.e., before the residual energy level of any given node fell below a certain threshold. The best performing variant of the original HDA protocol was found to be the "Agg. Source Count, RM Source Count Piggyback" method, which aggregated reinforced data packets based on the source count learned through the periodic transmission of reinforcement message packets. In other words, the parent node that contained the maximum number of sources in its sub tree was chosen to be the receiver and aggregator of reinforced data packets. The best performing variant of the proposed protocol was found to be the "Agg. Energy, INT and RD Energy Piggyback, INT Thr" method. This method selected the parent node to be the receiver and aggregator of reinforced data packets based on its residual energy learned through the IMP and RDP stages. The residual energy was piggybacked in interest message packets and reinforced data packets. In addition this method selected receivers of interest data packets based on their residual energy being higher than a threshold. As shown in Fig. 2 the proposed version was able to consistently deliver more reinforced data packets and, therefore, live longer, than the original HDA algorithm.

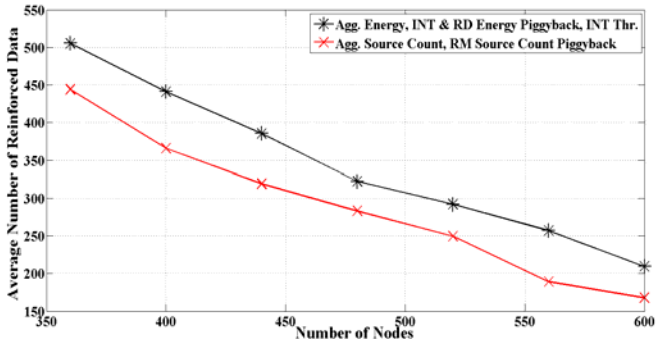


Fig. 2. Average number of reinforced data generated within the network vs. number of nodes with a source to node ratio of 2.5 %.

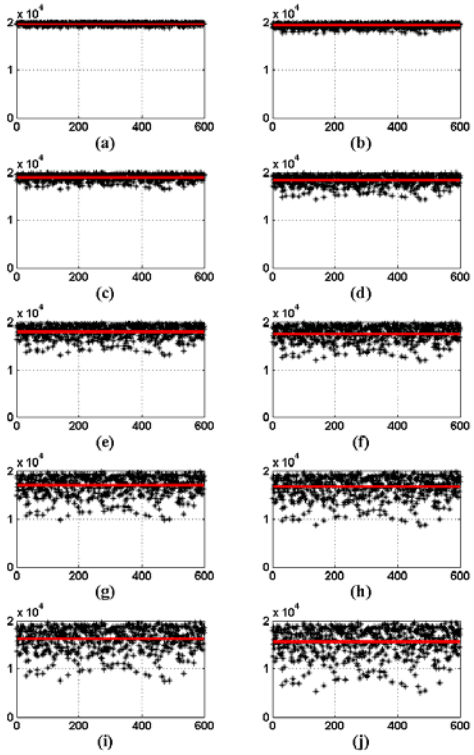


Fig. 3. Residual energy and mean residual energy presentation of each node at 20s intervals for improved version of HDA with INT & RDP energy piggiback and INT threshodling.

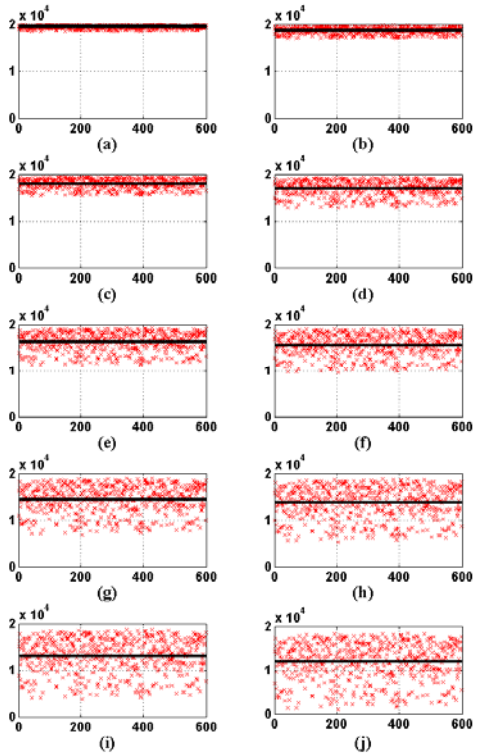


Fig. 4. Residual energy and mean residual energy presentation of each node at 20s intervals for original version of HDA with RM source count piggyback.

4.3 Energy balancing

The energy balancing performance of the two protocols was compared by measuring the residual energy distributions of each node for 600 nodes in the network, and for intervals of time ranging from 20s to 200s, incremented by 20s. As shown in Fig. 3 (proposed version) and Fig. 4 (original version), the proposed version shows consistent improvement over the original version of HDA. The average residual energy of the proposed protocol at each time 20s interval (graphs *a* to *j*, of Fig. 3) was higher than that of original HDA of the graphs *a* to *j* of Fig. 4. Furthermore the spread of energy is quite noticeably less in the proposed version than that of the original HDA for each time interval.

5 Improved directed diffusion

This paper's idea of updating residual energy at the reinforced data packet stage was also applied in the flat network of basic DD. The sink and the source were placed on the extreme ends of the network so that the maximum number of hops would be required. The simulations showed that by using the proposed residual energy based routing and higher rate of residual energy updates, the overall performance of the basic DD was improved. For comparison purposes, the same event timings, residual energy, and number of nodes were used as in HDA.

5.1 Interest propagation

Initially, the network enters the neighbor discovery process, and each node registers other nodes as its neighbor if they are within the radio range of 4m. The sink floods the network with an interest message packet. The interest message contains the event type the node is interested in and the interval for the next event. Whenever a node receives an interest message, it will add the sending node to its "gradient list." The gradient list is used to selectively route back exploratory data packets towards the sink. We simulate our network with only one interest sent out periodically.

5.2 Exploratory data propagation

Eventually the interest message reaches a node that can serve the interest. Corresponding node, which in this case identified as the "source" node will forward the requested data back to the "sink" along the gradients. This data is termed as exploratory data (ED) which is sent out periodically. In the proposed improvement of DD, the residual energy value of each node is piggybacked (included) in the exploratory data packet. When a node receives an

exploratory data packet, it will update its sender's list with the new residual energy value of the sender. In the basic directed diffusion, only the timestamps of the senders of the exploratory data packets are recorded and updated.

5.3 Reinforcement message propagation

When the exploratory data reaches the sink, it initiates the Reinforcement Message Packet (RMP) process. In the basic DD, the node unicasts the RMP to the node which passed the ED packet first. In the proposed improvement, the RM packet is sent to the node with the highest residual energy. This attempts to apply load balancing within the network. The RMPs are sent periodically with the lowest event frequency. When the sink receives the first exploratory data packet, it waits one ED packet interval to gain complete information about the residual energies of the sending nodes.

5.4 Reinforced data propagation

Once the source receives reinforcement message packet, it arranges for the reinforced data packet (RDP) to be sent back at a higher rate towards the sink. Reinforced data is sent at the highest rate within the network. The proposed improvement piggybacks the residual energy value of each node onto the RDPs. In this way the residual energy values are updated across the network at a higher rate compared to slower ED updates. As the simulations show, by updating the residual energy at a higher rate, the network lifetime increased. Reinforced data is relayed only to the node which reinforced the sending node (sent the reinforcement message packet). Eventually a shorter path is developed and the data is sent out by the source at a higher rate towards the sink.

6 Experimental results

A Java simulator was developed for both basic directed diffusion and the proposed improvement. The network was simulated for different number of nodes from N=350, 400, 450,..., 600. Initial residual energy for each node was equal to $2x10^4$ Joules; radio range = 4m.The energy consumption by a node for Transmission, E_tx=2 J and for Reception, E_rx=1. The event triggering times for interests, exploratory data, reinforcements and reinforced data were 5 s, 10 s, 15 s and 1 s respectively as in HDA.

6.1 Residual energy variance

As shown in Fig. 5, the proposed improvement has a lower variance (23%) in residual energy for large number of reinforced data packets than that of the basic directed diffusion.

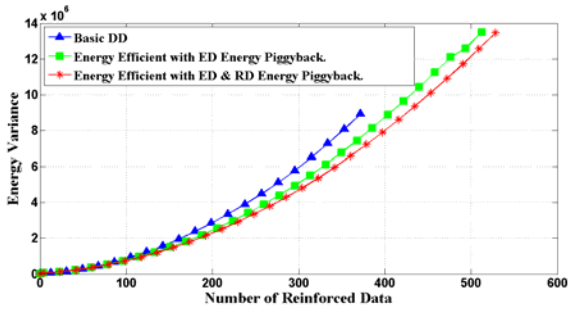


Fig. 5. Variance of residual energy vs average number of reinforced data events genenerated within the network.

6.2 Network life time

Similar to the section 4.2, we compare the network lifetime of both basic DD and improved DD schemes. Fig. 6 plots the node density of two networks versus the number of reinforced data passed by the nodes before any one node died in the network. Clearly the proposed improved version of DD (with piggy backing residual energy in exploratory data and reinforced data packets) delivers more reinforced data packets than the basic DD .The proposed improvement has a longer network lifetime (28 %) compared to basic DD.

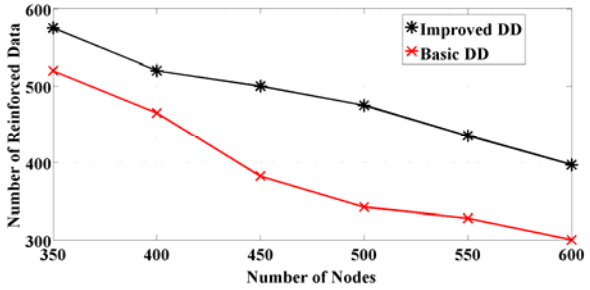


Fig. 6. Average number of reinforced data generated within the network vs number of nodes for basic and improved DD with ED & RD energy piggybacked.

6.3 Energy balancing

In four different phases, before any one node died, the network's residual energy distribution of each node was plotted and compared for both cases of basic DD and proposed improved version of DD with piggy backing residual energy in exploratory data and reinforced data packets. By comparing the Fig. 7d (basic DD) with Fig. 8d (improved version), it is clear that the proposed improvement was able to live longer by judiciously choosing nodes based on their residual energy and by gracefully lowering the average variance of residual energy near the "death floor" much better than basic DD.

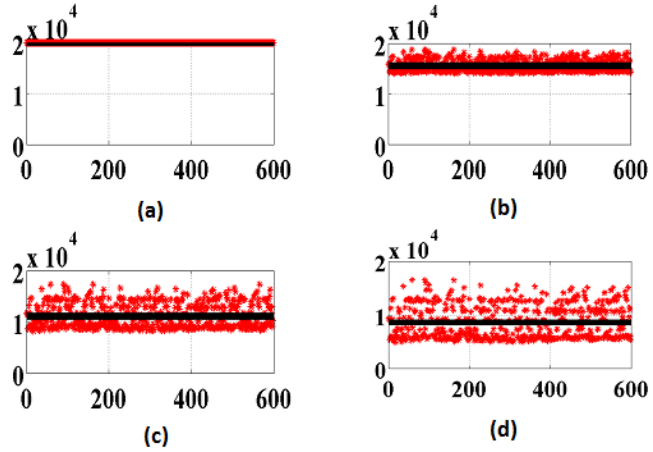


Fig. 7. Residual energy and mean residual energy presentation of each node for the basic DD.

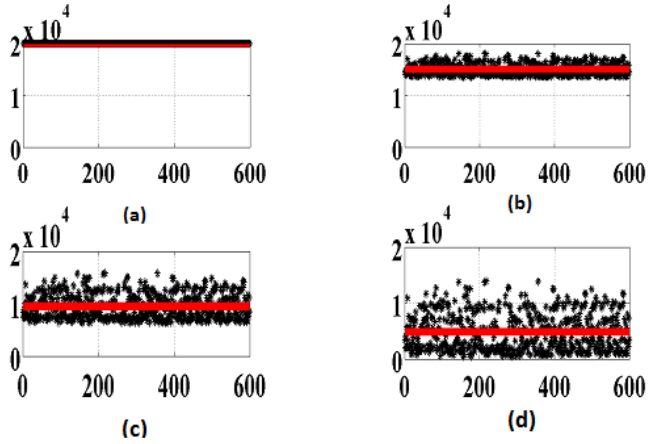


Fig. 8. Residual energy and mean residual energy presentation of each node for the improved DD with ED & RD energy piggybacking.

7 Conclusions and future work

This paper presented an energy efficient method of load balancing for wireless sensor networks. The proposed method updated residual energy during the reinforced data packet stage, the highest data propagation process in examples of flat (Directed Diffusion) and hierarchical networks Hierarchical Data Aggregation). The simulation, analysis, and comparison with hierarchical data aggregation and basic directed diffusion show that the proposed method has a lower variance in residual energy and a longer lifetime than both hierarchical data aggregation and basic directed diffusion. In addition, the interest message packet flooding was reduced and the network lifetime increased by limiting the nodes that could receive interest packets by their residual energy. The proposed method showed a 25% improvement in residual energy at the time of network death and an average improvement of 10% in extending the lifetime of the network as compared with hierarchical data aggregation. As compared with Directed Diffusion, the proposed method showed a 23% improvement in residual energy at the time of network death and a 28% improvement in extending the lifetime of the network. There was very little appreciable cost to piggybacking the residual energy values to existing packets.

Future work includes evaluating the same algorithm with many sinks and many sources to verify the algorithm's scalability assertion; in other words, since residual energy is updated opportunistically (piggybacked onto to already purposed data packets), the cost of updating residual energy is very low, amounting to the cost of transmitting a few bytes at each data transmission time. Also, future work includes an investigation of how bad paths can be self-healed by the network. In other words, even though a node may have higher residual energy than its neighbors, passing a packet through that node might lead to entrance to a bad path, and the network would need to recover from that. Finally, the cost of delay needs to be determined. In other words, choosing the highest residual energy node might lead to a path which has higher delay. The impact the algorithm has on delay needs to be quantified.

8 References

[1] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 2-16, Feb. 2003.

[2] S. Cui, K. Ferens, "Energy Efficient Clustering Algorithms for Wireless Sensor Networks," *International Conference on Wireless Networks*, Las Vegas, NV, USA, July 2011.

[3] N. Eghbali, A., M. Dehghan, "Load-balancing using multi-path directed diffusion in wireless sensor networks," *International Conference on Mobile Ad-hoc and Sensor Networks (MSN 2007)*, pp: 44-55, 12-14, 2007.

[4] S. Masoudi, A.M. Rahmani, A. Nasiri Eghbali, A. Khademzadeh, "GMPR: A Greedy multi-path routing algorithm for wireless sensor networks," *Second International Conference on Future Generation Communication and Networking (FGCN)*, pp: 25-30, 2008.

[5] R. Vidhyapriya , Dr.P.T Vanathi, "Energy Efficient Adaptive Multipath Routing for Wireless Sensor Network," *IAENG International Journal of Computer Science*, August 2007.

[6] Bin Zhou, Lek Heng Ngoh, Bu Sung Lee, Cheng Peng Fu, "HDA: A hierarchical data aggregation scheme for sensor networks, " *Computer Communications*, Volume 29, Issue 9, 31 May 2006, Pages 1292-1299, ISSN 0140-3664, 10.1016/j.comcom.2005.10.006.

TinyObf: Code Obfuscation Framework for Wireless Sensor Networks

Rafael Costa[1], Luci Pirmez[1], Davidson Boccardo[2], Luiz Fernando Rust[2] and Raphael Machado[2]
[1]Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[2]National Institute of Metrology, Quality and Technology, Rio de Janeiro, Rj, Brazil
{rafaelcosta,luci}@labnet.nce.ufrj.br, {drboccardo,lfrust,rcmachado}@inmetro.gov.br

Abstract - *The present paper proposes a code obfuscation framework to improve the security of WSNs. The obfuscation techniques employed in this framework aims at countering reverse engineering by compromising the disassembly stage of reverse engineering tools. We combine the obfuscation techniques with control flow redirection to protect sensitive data within the program code, making the disassemblers interpret sensitive data like program instructions. Using the proposed framework an attacker who wants to reverse a WSN program node in order to extract or modify sensitive data will spend much more time and effort.*

Keywords: Wireless Sensor Networks, Code Obfuscation and Security.

1 Introduction

Recent advances in micro-electromechanical systems and wireless communications technologies have enabled the building of low-cost and small-sized sensors capable of sensing, processing and communicating through wireless links. Wireless Sensor Networks (WSNs), composed of tens, hundreds and sometimes thousands of these small devices, are commonly used to monitor physical and environmental variables as temperature, humidity, and noise. WSNs are used for a wide range of applications, such as structural monitoring, natural resources mapping, tracking and monitoring of military targets, and smart environments control [1].

On the one hand, WSN offers new broad perspectives for various applications. On the other hand WSN poses a series of new challenges. In addition to challenges related to resource constraints, WSNs are subject to vulnerabilities associated with wireless communication and ad-hoc organization, both inherent characteristics of this type of network. Furthermore, in scenarios involving unprotected hostile outdoor areas, WSNs are prone to different types of attack that can compromise reliability, integrity and availability of this network [1]. In this paper we are interested in attacks that aim to compromise sensor nodes in order to violate its integrity or to extract confidential information of the code program.

A common approach to protect against node compromise attacks is code obfuscation. Programmers obfuscate their code with the purpose of making it difficult to discern and extract data from the code. It is becoming increasingly common to obfuscate code to protect intellectual property [2, 3]. However, the code may also be obfuscated to hide malicious behavior [4, 5]. A straightforward obfuscation technique is based on cryptography: the software developer encrypts sensitive data to maintain it unclear for an attacker. Nevertheless this sort of protection is based on the security of the cryptography key, which of course must be stored within the code. Once discovered the key, all sensitive data can be revealed. Furthermore, the processing overhead of encryption and decryption can be prohibitive since every access to the encrypted sensitive data demands their decryption and subsequent encryption after their use.

Another obfuscation technique is the instruction substitution of a program code performed at the assembly level producing a code semantically equivalent to the original code but syntactically different [6]. Depending on how the instructions are replaced and arranged within of the program code, the discovery of sensitive data using instruction substitution technique is harder than purely using cryptography. Moreover, the protection obtained by applying changes in the program code is not based on a single element (cryptography key) but on different kinds of substitutions that are performed along the program. So code obfuscation based on the instruction substitution can be used to protect sensitive data with a greater degree of resilience against automated reverse engineering tools and less overhead in terms of processing and memory usage than cryptography.

This paper proposes a code obfuscation framework for WSN that provides an efficient way to protect sensitive data stored at sensor nodes. The proposed framework combines obfuscation techniques that disrupt the disassembly generated by disassembler tools with control flow redirection to hide sensitive data within the program code. By deterring the disassembly stage, further reverse engineering stages may be questionable and any subsequent analyses circumspect. More specifically, we customized state-of-the-art code obfuscation techniques for MICAz motes to acquire locations in which sensitive data can be hidden and, consequently, protected from reverse engineering tools.

In this paper we use the proposed framework to protect sensitive data – cryptographic keys, watermark or serial number – of WSN nodes. Therefore the contributions of this work are: (i) a code obfuscation framework for WSN for countering the disassembly stage of reverse engineering without increasing considerably the use of sensor resources (presented in our experiments) and (ii) a mechanism to

protect sensitive data by combining the obfuscation techniques with control flow redirection, making possible to store sensitive data in program locations whose sensitive data bytes are disassembled like program instructions by reverse engineering tools . This paper is organized as follows. In Section 2 we give an overview of reverse engineering. Section 3 presents the TinyObf framework including the description of its logical architecture and the data protector mechanism with state-of-the-art code obfuscation techniques implemented and customized for the MICAz motes. Section 4 shows the experimental evaluation using application samples for the TinyOS. Section 5 discusses the related works and finally in Section 6 we present our concluding remarks.

2 Reverse Engineering

Reverse engineering of a program is the process that tries to reconstruct from an executable code a high-level language more readable for a human being. It is commonly divided in two stages: disassembly and decompilation. The disassembly stage translates an executable code into assembly language, which is the mnemonic notation of an executable machine code. The decompilation stage tries to construct high-level structures from the assembly language.

There are two main algorithms for disassembling: linear sweep and recursive traversal. Figure 1(a) shows the linear sweep algorithm that starts the disassembly process in the first byte of the text segment of the program and follows translating linearly until the end of the segment. The main weakness of this algorithm is that it is prone to disassembly errors when the text segment contains embedded data. Figure 1 (b) shows the recursive traversal algorithm that considers the control flow to disassembly a program. Thus, whenever the algorithm detects an instruction that can take more than one path, it tries to predict the next byte to be translated according to the control flow. The main weakness of this algorithm lies in how he tries to determine the next set of bytes that must be translated, because inaccuracies in determining the possible targets of an instruction can result in disassembly errors.

3 The TinyObf Framework

This section presents a generic code obfuscation framework that can improve security of WSNs. This framework uses obfuscation techniques that compromise the disassembly stage in order to protect sensitive data by making the reverse engineering process harder.

3.1 Logical Architecture

The logical architecture of the TinyObf framework, shown in Figure 2, consists of five components: Obfuscation Manager, Security Manager, Compiler, Obfuscator and Data Protection and two databases: Security Profiles and Obfuscation Rules.

The Obfuscation Manager is responsible for managing the operation of the other components. The Security Manager is the component that defines the security profile for the sensor node. A security profile specifies which obfuscation techniques will be applied and the obfuscation degree of the application (how many times each obfuscation technique will be applied). This component chooses the profile that has a higher security priority and a higher obfuscation degree among the security profiles of their application contained in the Security Profiles database. The compiler is the component responsible for compiling the source code. The Obfuscator component deals with the instruction substitution of a binary code according to the respective security profile. The Data Protection is responsible for hiding/protecting sensitive data in dead execution spots, i.e. spots in the code segment that never execute but whose bytes are, in fact, interpreted like program instructions by disassembler tools.

The Security Profiles and Obfuscation Rules databases are previously filled by a specialist at the pre-initial stage; the Security Profiles by the Application specialist and the Obfuscation Rules by the Security specialist. The Security Profile database contains for application the following fields: security profile identification, application name, obfuscation

```
global startAddr, endAddr;
proc DisasmLinear(addr)
begin
  while(startAddr <= addr < endAddr) do
    I:= decode instruction at address addr;
    addr += length(I);
  done
end

proc main()
begin
  startAddr := address of the first executable byte;
  endAddr := startAddr + text section size;
  DisasmLinear(ep);
end
```

(a)

```
global startAddr, endAddr;
proc DisasmRec(addr)
begin
  while(startAddr <= addr < endAddr) do
    if (addr has been visited already) return;
    I := decode instruction at address addr;
    mark addr as visited;
    if (I is a branch or function call)
      for each possible target t of I do
        DisasmRec(t);

      done
      return;
    else addr += length(I);
  done
end

proc main()
begin
  startAddr := program entry point;
  endAddr := startAddr + text section size;
  DisasmRec(startAddr);
end
```

(b)

Figure 1: Disassembly Algorithms: (a) linear sweep and (b) recursive traversal.

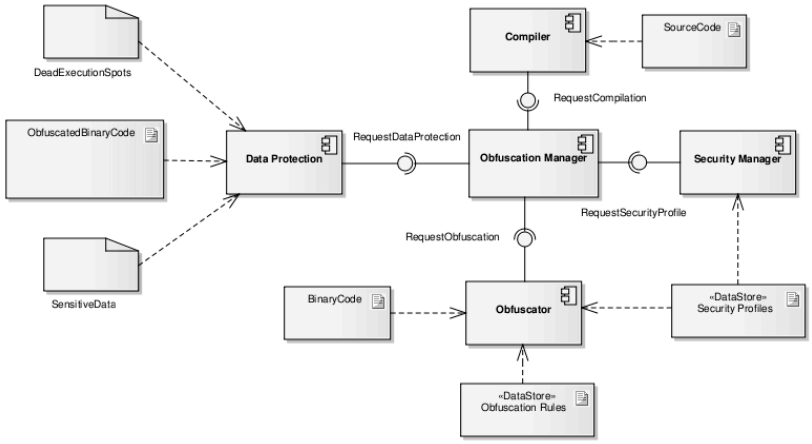


Figure 2: Logical Architecture of the code obfuscation framework.

techniques and degree of obfuscation. The Obfuscation Rules database stores the information used by the Obfuscator component to perform all the code obfuscation techniques in a specific hardware platform. Its fields are: name of obfuscation technique, set of instruction that will be replaced, set of instructions that mimic the same behavior of the instructions that will be replaced. This database is unique for each hardware platform and the security specialist must customize all obfuscation techniques to different instruction sets. In this paper we present the state-of-the-art code obfuscation techniques customized for MICAz motes. These techniques are: call obfuscation [7], return obfuscation [7], false return obfuscation [8] and jump obfuscation [3].

3.2 TinyObf Operation

The TinyObf framework has as input a source code, the platform, which the code will be compiled, and the sensitive data that should be protected. All the operations of the framework occurs offline, before the deployment of the program code. First, the Obfuscation Manager component requests to the Security Manager component the security profile that fulfils all the applications embedded in the source code of the sensor node. Next, the Obfuscation Manager requests to the Compiler component the compilation of the source code in a specific platform. Then, the Obfuscation Manager precedes the obfuscation and data protection procedures by calling the respective components: Obfuscator and Data Protection. The Obfuscator is responsible to apply the obfuscation techniques with their associated degree accordingly with the security profile for the binary code generated by the Compiler. It also deals with control flow redirection to create dead execution spots for later use by the Data Protection. Finally, the Obfuscation Manager requests to the Data Protection for hiding the sensitive data inside the dead execution spots in obfuscated binary code. After all, the obfuscated binary code can be installed in the sensor.

3.3 Sensitive Data Protection

We combine obfuscation techniques with control flow redirection to disrupt the disassembly stage and for protecting sensitive data. The disassembly stage deals directly with the binary code, which once disassembled, is not syntactically rich. Therefore the identification of parameters, procedure boundaries, procedure calls, and returns is done by making assumptions. Examples of such assumptions are: the sequence of instructions at a procedure entry (prologue) and at a procedure exit (epilogue), the parameter passing convention, and the conventions to make a procedure call. These assumptions are often referred, by researchers, as a 'standard compilation model.' However, these 'standards' are compiler specific; they are not industry standards. Even for a given compiler the standards may vary depending on the optimization scheme selected. When these assumptions are not reliable, such as for obfuscated or optimized code, the produced disassembly may be questionable and any subsequent program analyses circumspect.

We consider four obfuscations techniques to disrupt the assumptions adopted by disassembler tools: call obfuscation, ret obfuscation, false ret obfuscation and jump obfuscation. These techniques can change the control flow of a program making it possible to create dead execution spots while keeping the program behavior unchanged. However, the bytes contained in these spots will be translated as program instructions by disassembler tools due to the assumptions earlier described. We use these spots to embed sensitive data, which can be any data that the software developer wants to protect. Examples of these sensitive data are usually cryptographic keys, serial numbers or even watermark.

3.3.1 Call Obfuscation

According to the existing assumptions used by compilers, the next instruction to be executed after a function end is the instruction located after its respective function call instruction. Thus, the disassembler algorithms commonly disassembly the bytes stored after a CALL instruction.

This behavior can be used to obfuscate a program, redirecting the return of a function to another location. The semantics of the CALL instruction is based on stacking the subsequent instruction address (the function´s return address) and transfering the flow of execution to the address specified n the operand of the CALL instruction. One way to

accomplish call obfuscation can be obtained by replacing the CALL instruction to other set of instructions while keeping the same behavior.

For example considering the ATmega128L instruction set [9], commonly used for MICAz motes in a WSN, we can replace a CALL instruction by a combination of four LDI instructions, four PUSH instructions and a RET instruction. This combination first pushes the return address of the function and then pushes the address of the function to be executed. The instruction RET removes the address from stack and transfers the execution flow to the function that should be executed.

Figure 3 illustrates a non-obfuscated and an obfuscation function call. The stack used by ATmega128L stores 8-bit data coming from one of the general registers (R16 to R31). However, the memory address is 16 bits length, so it is necessary two LDI instructions and two PUSH instructions to load a memory address on the stack. In Figure 3(a) the 'function1' is called at address 0x02DD. When RET instruction is encountered at address 0x0310 the control flow goes back to the address 0x02DF. The Figure 3(b) shows an example of call obfuscation. This requires relocating the original instruction 'CPC R0, R3' at address 0x02DF to the address 0x0300, and then pushing this address and the address 0x0305 (address of the 'foo)' to the stack. This obfuscation technique also adds the RET instruction located at the address 0x02E5 to keep the original program semantics. When the RET instruction stored at address 0x02E5 is executed, the memory address of 'foo' is getting from stack and the control flow is transferred to it. The same occurs when the RET instruction of the 'foo' is executed, i.e., the return address 0x0300 is retrieved from stack and the control flow transferred to it. It is noteworthy that the memory interval between 0x02E5 to 0x0300 is a dead execution spot, which can be used to store sensitive data or to add junky bytes, in this case is used to store four bytes 0x0E940102 that can be a cryptographic key used by the application. However when disassembly the addresses 0x02E6 and 0x02E7 it will be translated as 'CALL 0x0102'.

3.3.2 Return Obfuscation

According to the existing assumptions used by compilers, when the disassembler encounters a RET instruction it believes that it reaches the function end. However it is possible to change (to obfuscate) the RET instruction for semantically equivalent instructions by directly manipulating the return address stored at the top of the stack. This behavior can obfuscate a program because the disassembler could not identify the proper end of a function.

For example, considering the ATmega128L instruction set, Figure 4 shows how we can replace a RET instruction by a combination of POP instructions, opaque predicate[1] [10], conditional branch and an indirect jump. First, two POP instructions are used to retrieve the return address of the 'foo' from the top of stack. Then, a simple opaque predicate, represented in our example with the compare instruction 'CP R30, R30', always evaluates to the true constant making the following conditional branch instruction 'BREQ 0x02' redirecting the control flow to the indirect jump instruction 'IJMP' at 0x0316 address. Finally, the 'IJMP' instruction redirects the control flow to the same target of the 'RET' instruction. The conditional branch instruction 'BREQ 0x02', in our example, creates two bytes of dead execution spots (0x0314 and 0x0315 addresses), which are used to protect the cryptographic key 0x0E940102, which is translated by the disassembler tool as CALL 0x0102.

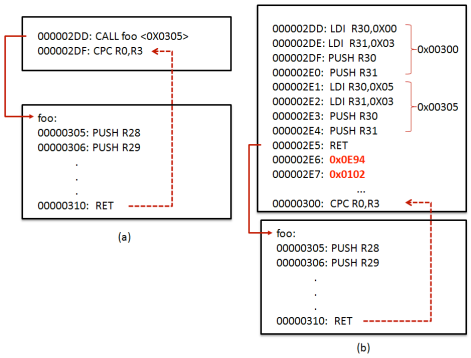


Figure 3: Example (a) non obfuscated and (b) obfuscated by call obfuscation.

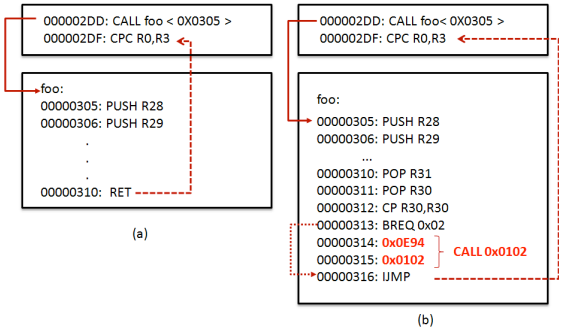


Figure 4: Example (a) non obfuscated and (b) obfuscated by return obfuscation.

3.3.3 False Return Obfuscation

This code obfuscation technique tries to compromise disassembler algorithms based on the RET instruction assumptions. In this case, when the disassembler encounters a RET instruction, it believes it reaches the end of the function and therefore the following instructions are not disassembled as instructions of the function. This behavior can be used to obfuscate a program forcing a premature end of a function using a false RET instruction, i.e., a RET instruction that does not characterize the end of the function. Thus, the disassembler will not consider the bytes stored in memory addresses subsequent of the false RET instruction. Furthermore, this technique makes a static program analysis tool believes that the next instruction is the instruction stored at the memory address located after function call.

Figure 5(a) shows an example of a non-obfuscated function foo. After the function execution, the instruction

[1] Predicate that always evaluate to either the constant true or the constant false regardless of the values of their inputs.

'CPC R0, R3' should be executed. In Figure 5(b) we have an example of how to use a false return obfuscation to force the premature end of 'foo'. First we use two instructions LDI and two instructions PUSH to stack the false return address 0x030E. Thus, a disassembler will believe that 'foo' ends at the 0x030B address. However instructions stored between 0x030C and 0x030D can be considered dead execution spots since the instructions stored at these addresses will never execute. In our example, false ret creates two bytes of dead execution spots (0x030C and 0x030D addresses), which are used to protect the cryptographic key 0x0E940102 which are translated as CALL 0x0102 by the disassemble tool.

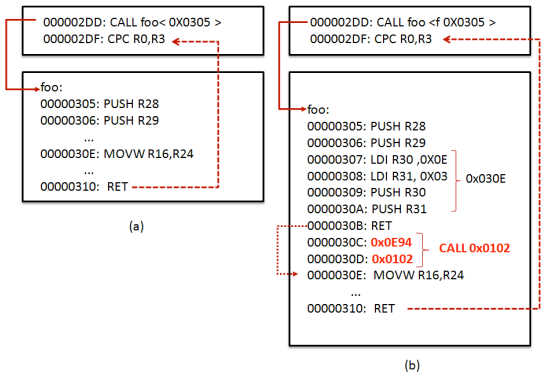


Figure 5 : Example (a) non obfuscated and (b) obfuscated by false return obfuscation.

3.3.4 Jump Obfuscation

This obfuscation technique is based on using a function called "branch function" that is capable of redirecting unconditional jump instructions of a program. It is used mainly for two purposes: (i) obscure the control flow of the program forcing an attacker to analyze an abstracted function (branch function) instead of a simple jump instruction and (ii) mislead the disassembly leading the disassembler to continue the disassembly process at the instruction following the CALL instruction.

Figure 6 shows an example of using a CALL instruction to the branch function 'f' that replaces the unconditional jump 'RJMP 0x0305'. The semantics of both instructions are the same, to redirect the control flow to the address 0x0305. In this example, two bytes of dead execution spots (0x02DF and 0x02E0 addresses) are created, which are used to store the cryptographic key 0x0E940102, which is translated as 'CALL 0x0102' by the disassembler tool.

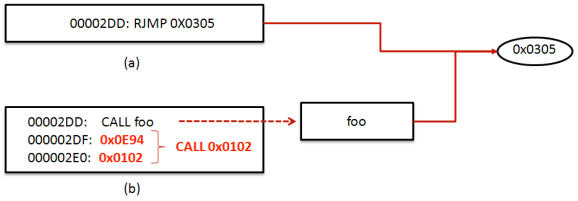


Figure 6: Example (a) non obfuscated and (b) obfuscated by jump obfuscation.

4 Experimental Evaluation

This section details the experiments with TinyObf in order to evaluate its efficiency in terms of the difficulty imposed for extracting sensitive data and its impact in terms of the resource consumption of the sensor.

The experiments were performed using the AvroraZ [11] to simulate a WSN. This network was composed by MICAz motes, which use 8-bit AVR microcontrollers (ATmega128) marketed by Atmel. The disassemblers chosen to evaluate the proposed framework were: avr-objdump [12], which uses linear sweep algorithm to perform the disassembly and IDA PRO [13], a commercial disassembler tool which basically uses transversal recursive algorithm.

The metric commonly used to measure the difficulty imposed by code obfuscation techniques to analyze a program was confusion factor CF (1) [3, 14, 15]. This metric calculates the number of units of the program (instructions, basic blocks or functions) that were incorrectly identified. The CF is defined by following equation: where A is the set of addresses of actual instructions, i.e. those that would be performed by the program, P is the set of address of instruction perceived by disassemblers and |A - P| is the set of addresses of instructions that are incorrectly identified by disassembler.

$$\mathbf{CF = |A - P| / |A|} \quad (1)$$

The impact of TinyObf in terms of resource consumption was measured using the following metrics: program size, processing cycles and energy consumption of CPU cycles (measured using AvroraZ). All of these metrics were calculated for the original and the obfuscated programs. We used three program samples of TinyOS 2.1.1 to perform the experiments: (i) RadioSenseToLeds, that broadcasts at 4Hz the value of a platform's default sensor, (ii) IDS (lymph node) and IDS (dendritic cells), both components of an Intrusion Detection System (IDS), program based in human immune system, and (iii) Delphos, program used to predict damage in wind turbines using Wireless Sensors and Actuators Networks (WSANs). The last application is a decentralized system where all decisions are processed inside the network. The objective of this system is to predict when a turbine will perform in a state of damage and act in his operation to prevent accidents, reduce maintenance costs and prevent delays that occur in power generation.

4.1 Results

In the following we show the results of our experiments considering the four metrics: confusion factor, program size, processing cycles and energy consumption. We also use the TinyObf framework to protect four bytes of data (sensitive data for each application). For all programs analyzed, the disassembler tools do not identify the functions boundaries neither the sensitive data, in our case, a cryptographic key 0x0E940102, translated as shown in the Section 3 as a call instruction to the address 0x0102 ('CALL 0x0102').

4.1.1 Confusion Factor

Figure 7 shows the CF for all applications using four obfuscation degrees. We can see that the CF increases as the obfuscation degree grows. So when we increase the degree of obfuscation, more difficult is to understand the program code.

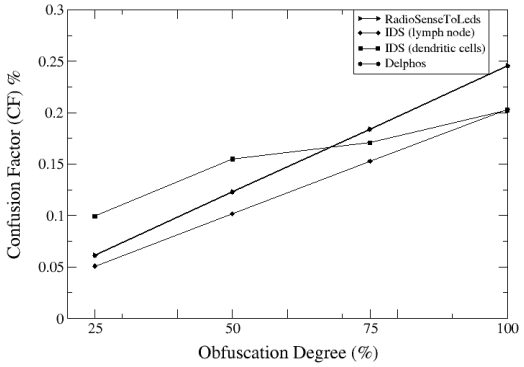


Figure 7 : Effect of code obfuscation on the program understanding.

4.1.2 Program Size

When we apply obfuscation techniques, some instructions can be changed or added to the program. Figure 8 shows the overhead in the program size after the obfuscation. Notice that as the obfuscation degree grows the difference between the original program (0% obfuscation degree) and the obfuscated programs is irrelevant even when the whole program is obfuscated. So the size overhead carried by obfuscation is considered worthless compared to the benefits of increasing the difficult to data discovery and program understanding.

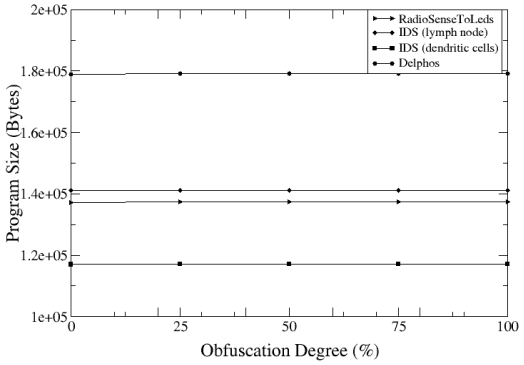


Figure 7: Effect of code obfuscation on the program size.

4.1.3 Processing Cycles

The changes performed by obfuscation techniques did not reflect negatively at time execution because few instructions were added to obfuscate a program then little more processor cycles is needed to do the same task. Figure 9 shows the increase in processing cycles of the obfuscated program compared with original program. Notice that as the obfuscation degree grows the difference in terms of processing cycles between original program (0% of obfuscation degree) and obfuscated program is irrelevant even when the whole program is obfuscated. So processing cycles overhead carried by the obfuscation techniques is considered worthless compared to the benefits of increasing the difficult to data discovery and program understanding.

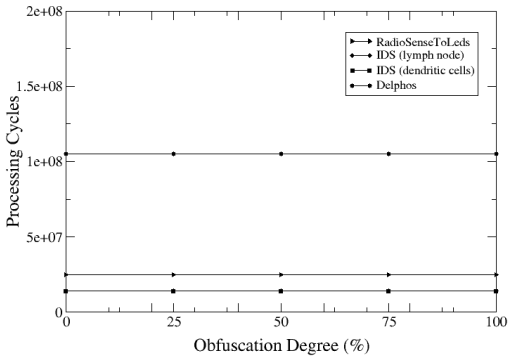


Figure 9 : Effect of code obfuscation on the processing cycles.

4.1.4 Energy Consumption

We also measure the energy consumption of changing or adding instructions in a program. Figure 10 shows the energy consumption of additional CPU cycles. Notice that as the obfuscation degree grows the difference of energy consumption between original program (0% of obfuscation degree) and obfuscated program are much less compared to the benefits of increasing the difficult to data discovery and program understanding.

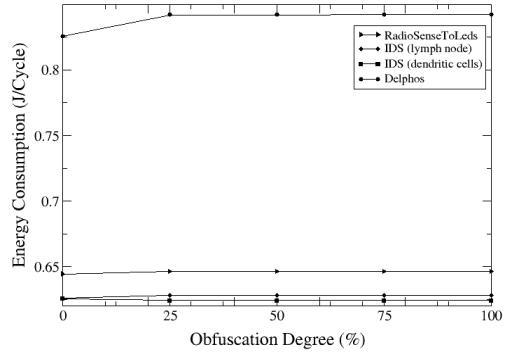


Figure 10 : Effect of code obfuscation on the energy consumption.

5 Related Works

There are plenty of studies proposing obfuscation techniques to protect intellectual property [2, 3, 6, 8, 16, 17]. Two of them are specific for WSN [16, 17]. The others [2, 3, 6, 8] disrupt disassemblers for x86 platform. The present work customizes these obfuscation techniques in order to use to other platforms, such as WSN.

The approach used by Gu [16] to improve security on WSNs is based on diversity. Each program deployed on WSN is different from one another. The diversity applied is given through obfuscation techniques that are capable to create different versions of the same program but semantically equivalent. This approach makes the exploitation of vulnerabilities difficult by adversary who wants to

compromise a WSN. This difficulty occurs because all the effort used to analyze one device is not helpful to compromise other devices because each program is different from one another. A great challenge of this is how to spread a different program for each sensor because this is a very costly task. WSNs usually consist of many nodes and deploy a different program for each sensor may take a long time because deploy is done individually for each sensor. The main difference between his proposal and the present proposal is the strategy to counter reverse engineering. The obfuscation techniques used by him tracks the decompilation stage. Notice that, since we use obfuscation techniques that disrupt the disassembly stage and it occurs before the decompilation one so any difficult imposed in the disassembly will propagate errors throw the decompilation stage. Both proposals technique are complementary and can be combined to obtain a higher degree of difficulty in understanding the program. We also combine the obfuscation techniques with control flow redirection in order to hide/protect sensitive data.

The method presented by [17] combines randomization and code obfuscation to protect executable code of WSN nodes. Randomization is defined as the composition of randomly arranged pieces of code while maintaining the program functionality. The work describes two obfuscation stages. First, it performs data obfuscation and second it obfuscates the functions that perform data obfuscation since the function are contained within of the program code. Even applying obfuscation in two different stages, the obfuscation techniques used by [17] are the same as in [16]. Thus the obfuscation used in this work is complementary.

6 Conclusions

In the present work we presented a framework that improves the security of software embedded in wireless sensor networks. Our contribution is twofold. On one hand, we describe a series of customized obfuscation methods that counter reverse engineering by increasing the resilience of the binary code against disassembly. On the other hand, we increase the protection from disclosure of sensitive data stored in WSN nodes. The effectiveness of our proposed solution is supported by our simulation results: while the confusion factor was increased, indicating an increasing of software protection, the overhead caused by the code transformations was negligible, be it in terms of energy consumption, execution time or size of code. The obfuscation techniques applied confuse the disassembler tools in terms of instruction translation and sensitive data disclosure.

7 References

[1] J. Yick et al., “Wireless sensor network survey,” In: Computer Networks. (2008).

[2] C. S. Collberg and C. Thomborson. Watermarking, tamper-proofing, and obfuscation tools for software protection. IEEE Transactions on Software Engineering, 28(8):735–746. (2002).

[3] C. Linn and S. Debray, “Obfuscation of executable code to improve resistance to static disassembly,” In: Proc. of the 10th ACM conference on Computer and communication security - CCS ’03. (2003).

[4] M. Christodorescu and S. Jha. Static analysis of executables to detect malicious patterns. In Proc. of the 12th USENIX Security Symposium, 2003.

[5] A. Lakhotia and P. K. Singh. Challenges in getting ’formal’ with viruses. Virus Bulletin, pp. 15–19, Sep. 2003.

[6] C. Collberg, C. Thomborson, and D. Low, “A Taxonomy of Obfuscating Transformations,” Technical Report 148, Department of Computer Science, The University of Auckland. (1997).

[7] A. Lakhotia, D. Boccardo, A. Singh, A. Manacero Jr, “Context-sensitive analysis without calling context,” In: Higher-Order and Symbolic Computation. Springer. 23(3): 275–313. (2010).

[8] J. Viega, M. Messier. “Secure Programming Cookbook for C and C++,” in O'Reilly Media publisher. (2003).

[9] ATmega128, available in http://www.atmel.com/dyn/products/ (Last accessed 01/2012).

[10] C. Collberg, C. Thomborson, and D. Low. Manufacturing cheap, resilient, and stealthy opaque constructs. In Proc. of the 25th ACM Symposium on Principle of Programming Languages, pp. 184-196, Jan. 1998.

[11] R. Alberola and D. Pesch, “AvroraZ: Extending Avrora with an IEEE 802.15.4 Compliant Radio Chip Model,” In: ACM Performance monitoring and measurement of heterogeneous wireless and wired networks. (2008).

[12] GNU Binutils, available in www.gnu.org/software/binutils/ (Last accessed 01/2012).

[13] IDA Pro, “Executive Summary”, available in http://www.hex-rays.com/ (Last accessed 12/2011).

[14] I. Popov, S. Debray, and G. Andrews. “Binary Obfuscation Using Signals,” In: Proc. Usenix Security '07. pp. 275-290. (2007).

[15] B. Lee, Y. Kim, and J. Kim, “binOb+: a framework for potent and stealthy binary obfuscation.” In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security. pp. 271-281. (2010).

[16] Q. Gu, “Efficient code diversification for network reprogramming in sensor networks,” In: Proceedings of the third ACM conference on Wireless network security. (2010).

[17] A. Alarifi and W. Du, “Diversify sensor nodes to improve resilience against node compromise,” In: Proceedings of the fourth ACM workshop on Security of ad-hoc and sensor networks - SASN ’06. (2006).

A Three-step Methodology for Stochastic Deployment of Wireless Sensor Networks

A. Ejnioui[1], C. E. Otero[1], I. Kostanic[2] and L. D. Otero[3]
[1]Information Technology, University of South Florida Lakeland, Lakeland, Florida, USA
[2]Electrical and Computer Engineering, Florida Institute of Technology, Melbourne, Florida, USA
[3]Engineering Systems, Florida Institute of Technology, Melbourne, Florida, USA

Abstract–*This paper presents a three-step methodology for stochastic deployment of wireless sensor networks. While the first step uses simulation to elaborate strategies for a given deployment scenario, the second step consists of applying vertical variance trimming techniques to reduce the number of non-redundant alternative strategies. The final step consists of formulating and solving the problem as a multi-attribute decision problem using grey number theory due to the uncertain character of simulation data. Decision-makers can use this methodology to determine the best deployment strategies based on mission specific goals. This methodology can be used to simplify the decision-making process and provide decision-makers the ability to consider all factors involved in the wireless sensor network deployment problem. The methodology can be easily customized to include numerous quality factors to further compare deployment strategies and identify the one that best meet applications requirements.*

Keywords: wireless sensor networks, stochastic deployments, statistical analysis, grey numbers, multi-attribute decision.

1 Introduction

Recent advances in micro electro-mechanical systems have led to the development of tiny low-power devices that are capable of sensing the world and communicating with each other. Such devices may be deployed in vast numbers over large geographical areas to form wireless sensor networks (WSN). WSNs provide the means for autonomous monitoring of physical events in areas where human presence is not desirable or impossible. Therefore, they are expected to facilitate many existing applications and bring into existence entirely new ones. Several proposed applications of WSNs include disaster relief, environmental control, military applications and border security [1].

In each application, the sensor nodes are deployed over the area of interest and tasked with sensing the environment and communicating with each other in multi-hop fashion to transmit the information back to a base station, also known as the information sink [2]. From the sink, the information is collected and typically relayed to a central location, across remote sites, where it is processed and analyzed.

For the most part, WSNs are highly application-dependent. This means that details such as node design, form-factor, processing algorithms, network protocols, network topology, and deployment scheme are customized for the proposed application. Among these, the deployment scheme is considered extremely important, since it directly influences parameters such as network complexity, connectivity, coverage, cost, and lifetime.

WSN deployments schemes are classified as deterministic or stochastic. Deterministic deployments typically result in optimal efficiency; however, due to the size and density required to provide appropriate network coverage in large geographical areas, careful positioning of the deployed nodes is impractical. Furthermore, several applications of WSNs are expected to operate in hostile environments [3]. This makes pre-defined deployment in some cases impossible; consequently, stochastic deployments become the only feasible alternative [1]. For these applications, sensor nodes may be dropped from a plane, delivered in an artillery shell, rocket or missile, or catapulted from a shipboard [2]. In these cases, the WSN has the utmost challenge of guaranteeing connectivity and proper area coverage upon deployment [4]. This requires detailed planning to find deployment strategies that meet application requirements in terms of network connectivity, coverage, cost, and lifetime.

This paper presents a decision-making methodology for stochastic deployment of WSNs. The methodology uses simulation, statistical analysis, and a utility-based multi-attribute decision process based on grey number theory to provide an innovative and unique approach that helps decision-makers determine goal-oriented deployment strategies from a set of alternatives. Furthermore, the methodology provides significant contribution to the current body of research by providing an extensible technique that takes into account important parameters, such as connectivity, coverage, cost, lifetime, involved in the deployment of WSNs.

2 Background Work

The deployment problem has been the topic of much research work; however, the majority of the methodologies concentrate on carefully positioning nodes to meet application

requirements [5, 6]. Insufficient work has been done to improve decision-making in stochastic deployment of WSNs. Furthermore, most of the current work provides methodologies that take into account one or two parameters at the expense of other network parameters. In [7], the authors present a methodology to maximize coverage and connectivity in randomly deployed WSN. In [8, 9], the authors present a methodology for decreasing node density (i.e., cost). In [10, 11], the authors point out the lack of research towards the WSN deployment problem and state that "While WSN design, architecture, protocols and performance have been extensively studied, only a few research efforts have studied the device deployment problem". Furthermore, the authors point out flaws in recent publications by stating, "Most of these works tackle the deployment problem only from a perspective of coverage and/or connectivity. The significance of deployment on lifetime is mostly overlooked".

These methodologies fail to provide decision-makers with holistic views that consider all parameters involved in the deployment decision-making process and allow them to make customized deployment decisions based on all application requirements.

3 Methodology

WSN are application-specific, therefore it is impractical to expect that the same solution can be used to address deployments in all environments. The proposed methodology is built on this fundamental assumption. For that reason, it requires decision-makers to execute the methodology under settings that provide appropriate characterization of application-specific requirements. Consequently, customized simulations specific to the deployment scenario at hand are required. Simulations of WSN deployments can be carried out using the popular ns2 simulator. Once simulation data are collected, the methodology uses the *Vertical Variance Trimming* (VVT) technique to eliminate statistically redundant deployment alternatives. Therefore, VVT provides a reduced number of deployment alternatives but equal characterization of the original deployment scenario. Finally, the reduced set of deployment alternatives is analyzed by formulating and solving a multi-attribute decision problem using grey numbers in order to rank deployment alternatives based on deployment goals. Figure 1 shows an overview of this methodology.

4 Vertical Variance Trimming

Vertical Variance Trimming (VVT) is a technique devised for reducing the number of deployment alternatives based on statistical analysis [12]. VVT works by determining the effects of deployment parameters (e.g., number of nodes, radio range, sensor range) on WSN network efficiency. Typical WSN deployment assumptions include: (1) higher network connectivity and area coverage is achieved by deploying higher number of nodes; and (2) higher radio and sensing range result in higher connectivity and area coverage. These are valid assumptions; however, the degree to which they are significant in a stochastic deployment scenario needs to be determined before making deployment decisions. To accomplish this and reduce deployment alternatives, VVT uses three fundamental techniques: *Single Factor Analysis of Variance* (ANOVA) [13], *Least Significance Difference* (LSD) [13], and *Critical Metric Value* (CMV).

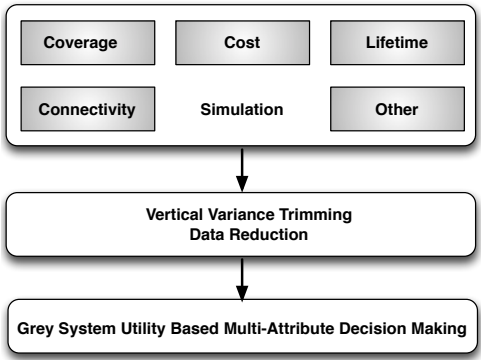


Figure 1. Methodology overview.

4.1 Single Factor ANOVA

Single factor ANOVA is a statistical procedure that tests the equality of two or more response means. In the deployment case, ANOVA can be used to determine the effects of deploying varying number of nodes at different radio and sensor range, on overall network connectivity and coverage. By using the F statistical test [13], the variance caused by natural error can be compared to the one caused by varying the deployment parameters. When these variances are similar, at least two deployment strategies are considered redundant and minimization of the number of deployment alternatives can be achieved. The test statistic F is computed using equation 1 [13],

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{SS_T/(a-1)}{SS_E/(N-1)} \qquad (1)$$

where N is the total number of observations and a the total number of treatments. A detailed explanation for computing SS_T and SS_E is presented in [13].

There are two ways that single factor ANOVA can be used in the WSN deployment problem. Table 1 displays identified use cases for single factor ANOVA.

Table 1. ANOVA Deployment Use Case.

Case	Number of Nodes	Radio Range	Sensor Range	Output
1	Variable	Fixed	Fixed	Connectivity
2	Variable	Fixed	Fixed	Coverage

In both cases, the effects of varying the number of deployed nodes on network connectivity and area coverage are determined by using a fixed set of values for radio range and sensor range. Once statistical differences have been identified using ANOVA, the LSD test is used to identify which deployment alternatives are different from one another.

4.2 LSD

The LSD test is used on pairs of deployment strategies to determine their statistical equality. Two deployment strategies are different if the difference in their mean response (i.e., connectivity, coverage) is greater than the least significant difference. This is obtained for balanced experimental data as specified in [13]. Otherwise, they are considered statistically similar. When this occurs, decision-makers can eliminate the redundant strategy that results in increased cost. Once all the deployment strategies have been analyzed using LSD, the CMV is used to further reduce the deployment alternative set to match specific application requirements.

4.3 CMV

After redundant deployment strategies have been removed, the Critical Metric Value (CMV) is selected to identify the minimum value for the decision metric that decision-makers are willing to accept for initial deployment. Deployment strategies resulting in values higher than the CMV are considered for initial deployment. Deployment strategies resulting in values below the CMV are considered for re-deployment to fill gaps and correct initial deployment. With this final step, VVT reduces the number of deployment alternatives significantly, which simplifies the application of the utility-based multi-attribute decision making process.

5 Multi-attribute Decision Making Using Grey Systems Theory

Multi-attribute decision making problems occur in situations where a finite set of alternatives need to be evaluated according to a number of criteria or attributes. The evaluation consists of selecting the best alternative or ranking the set of alternatives based on those attributes. The evaluation of various strategies to deploy wireless sensor networks can be approached by finding a set of criteria that provides the optimal benefit by minimizing cost factors. Deployment goals are customized based on specific application requirements. For example, multi-segment WSN require high connectivity and high coverage for the *Sensing & Relaying Segment (SRS)*, but high connectivity and low coverage for the *Relaying Segment (RS)* [8]. Other examples include WSN that use small autonomous vehicles after deployment to fill existing connectivity gaps in the network. In these applications, high area coverage with extended network lifetime is desired over high connectivity. In general, deployment strategies are composed of a fixed number of deployed nodes, fixed radio range, and fixed sensor range. For a given deployment scenario, there could be a number of deployment strategies, each providing different levels of connectivity, coverage, cost, and network lifetime. However, many decision problems present data that is imprecise or ambiguous leading to conflicting situations in which the evaluation of alternatives becomes difficult. This is the case when deploying WSNs. This information uncertainty has been modeled using fuzzy sets [14] or grey numbers [15]. While the former has been around for some time, only recently has interest been growing in the latter, since uncertainty can be modeled and manipulated in more flexible ways using grey number systems than fuzzy sets [15].

5.1 Grey Numbers and Grey Systems Theory

In practical applications, a grey number represents an indeterminate number that takes its possible value from an interval or a set of numbers. The symbol $\otimes$ denotes a grey number. The most basic types of grey numbers are [15]:

- Grey numbers with only a lower bound: $\otimes \in [\underline{a}, \infty]$ or $\otimes(\underline{a})$, where $\underline{a}$ is a fixed number representing the lower bound.
- Grey numbers with only an upper bound: $\otimes \in [-\infty, a]$ or $\otimes(\overline{a})$ where $\overline{a}$ is a fixed number representing the upper bound.
- Interval grey numbers: $\otimes \in [\underline{a}, \overline{a}]$ where $\underline{a}$ and $\overline{a}$ are the lower and upper bounds respectively.
- Continuous and discrete grey numbers: The former numbers can take any values within an interval while the latter can take only a finite number of potential values.
- Black and white numbers: When $\otimes \in [-\infty, \infty]$, that is when $\otimes$ has neither an upper nor lower bound, it is knows as a black number. On the other hand, when $\otimes \in [\underline{a}, \overline{a}]$ and $\underline{a} = \overline{a}$, it is knows as a white number.

After three decades of research, grey systems theory emerged as new discipline with contributions in [15]:

- Grey algebraic systems, grey equations, grey matrices, etc..
- Sequence operators and generation of grey sequences.
- System analysis based on grey incidence spaces and grey clustering.
- Grey prediction models.
- Decision making using grey target decision models.
- Optimization models using grey programming, grey game theory and grey control.

5.2 Strategy Selection

The first step in grey system decision making approach involves the selection of deployment strategies for a given deployment scenario. These strategies are developed during simulation of specific scenarios. The results are captured in a strategy vector as follows:

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \qquad (2)$$

where $i = 1, 2, \ldots, n$.

5.3 Scenario Attributes

When simulating a deployment scenario, a number of strategies are developed. Each deployment scenario can be characterized by the following attributes:

- *Nodes*: A deployment strategy consists of deploying a number of nodes scattered over a well-defined geographic area.

- *Radio range*: Each node has an antenna for transmitting and receiving data. This antenna has a well-known radio range.
- *Sensor range*: In addition to an antenna, a node can carry several sensors for monitoring specific parameters in the environment (e.g., temperature, noise, motion, etc.). These sensors have all well-defined sensing ranges.
- *Connectivity*: Each deployment strategy can be defined by the degree to which the deployed nodes are connected. This is known as the connectivity of the network resulting from the deployment strategy. In general, connectivity increases when higher number of nodes is deployed in a given strategy.
- *Coverage*: This is the area covered by the network of nodes in a deployment strategy.
- *Power*: The power consumed by the deployed nodes in a given strategy.

These attributes can be represented in the following vector:

$$A = [a_1\ a_2\ \cdots\ a_m] \qquad (3)$$

for $j = 1, 2, \ldots, m$.

5.4 The Deployment Scenario Matrix

For each deployment strategy, a value representing a benefit or cost is generally associated with each attribute. Because of the uncertainty of data in deploying WSNs, grey numbers are used to represent benefits or costs in the decision matrix. The overall assessment of a given deployment scenario based on the scenario attributes is captured using the following deployment scenario matrix:

$$\otimes D = \begin{bmatrix} \otimes d_{11} & \otimes d_{12} & \ldots \otimes d_{1m} \\ \otimes d_{21} & \otimes d_{22} & \ldots \otimes d_{2m} \\ \vdots & \vdots & \vdots \\ \otimes d_{n1} & \otimes d_{n2} & \ldots \otimes d_{nm} \end{bmatrix}$$

$$\otimes D = \begin{bmatrix} [l_{11}, u_{11}] & [l_{12}, u_{12}] & \ldots [l_{1m}, u_{1m}] \\ [l_{21}, u_{21}] & [l_{22}, u_{22}] & \ldots [l_{2m}, u_{2m}] \\ \vdots & \vdots & \vdots \\ [l_{n1}, u_{n1}] & [l_{n2}, u_{n2}] & \ldots [l_{nm}, u_{nm}] \end{bmatrix} \qquad (4)$$

where the rows represent strategies considered in the deployment scenario while the columns represent the attributes of that scenario. Note that the l_{ij} and u_{ij} represent respectively the lower and upper bounds of grey number d_{ij} for $i = 1, 2, .., n$ and $j = 1, 2, .., m$.

5.5 Goal Weights

In general, a deployment scenario will also be characterized by very specific goals. For example, the goals of a deployment scenario may consist of maximizing network lifetime, maximizing connectivity, minimizing cost and maximizing coverage in this listed order. The first goal entails minimizing power usage in the deployment scenario while the third goal entails minimizing the number of nodes deployed in the scenario. Optimization goals consist mostly of minimizing or maximizing one or more attributes associated with a deployment scenario. However, these goals may not have the same importance in some cases. As a result, a weight from 0 to 1 indicating the importance of a goal is given to each attribute in each scenario. In the absence of weights, all attributes are assumed to have equal importance. A weight vector is created where w_j represents the importance of each attribute as follows:

$$W = [w_1\ w_2\ \cdots\ w_m] \qquad (5)$$

5.6 Normalization of the Deployment Scenario Matrix

The scenario matrix can be normalized by using the core of the grey numbers in each column. If a grey number $\otimes \in [\underline{a}, \overline{a}]$ is continuous, then $\widehat{\otimes} = \frac{1}{2}(\underline{a} + \overline{a})$ is the core of $\otimes$ [14]. Grey numbers in the matrix can be normalized by using the sum of the cores in each matrix column as follows [15]:

$$\overline{l_{ij}} = \frac{l_{ij}}{\frac{1}{2}\left(\sum_{i=1}^{n} l_{ij} + \sum_{i=1}^{n} u_{ij}\right)} = \frac{2l_{ij}}{\sum_{i=1}^{n}(l_{ij} + u_{ij})} \qquad (6)$$

$$\overline{u_{ij}} = \frac{u_{ij}}{\frac{1}{2}\left(\sum_{i=1}^{n} l_{ij} + \sum_{i=1}^{n} u_{ij}\right)} = \frac{2u_{ij}}{\sum_{i=1}^{n}(l_{ij} + u_{ij})} \qquad (7)$$

for $i = 1, 2, .., n$ and $j = 1, 2, .., m$ where l_{ij} and u_{ij} are as defined in equation (4). The resulting normalized matrix is $\otimes\overline{D}$.

5.7 Weighting of the Normalized Scenario Matrix

The normalized scenario matrix can be weighted by multiplying the bounds of each grey numbers in the matrix by the weight of its attribute. Let $\otimes\overline{d_{ij}} = [\overline{l_{ij}}, \overline{u_{ij}}]$ be a grey number in the normalized matrix. Each grey number in the matrix is multiplied by its attribute weight as follows [15]:

$$\otimes\widehat{d_{ij}} = \otimes\overline{d_{ij}} \times w_j = [w_j\overline{l_{ij}}, w_j\overline{u_{ij}}] = [\widehat{l_{ij}}, \widehat{u_{ij}}] \qquad (8)$$

for $i = 1, 2, .., n$ and $j = 1, 2, .., m$. The resulting weighted normalized matrix is $\otimes\widehat{D}$.

5.8 Benefits and Costs in The Weighed Normalized Matrix

A simple weighted additive approach, similar to the COPRAS-G method, can be used to compute the benefits and costs of the attributes for each strategy in $\otimes\widehat{D}$ as follows [16, 17]:

$$P_i = \frac{1}{2}\sum_{j=1}^{k}\left(\widehat{l_{ij}} + \widehat{u_{ij}}\right) \qquad (9)$$

$$R_i = \frac{1}{2}\sum_{j=k+1}^{m}\left(\widehat{l_{ij}} + \widehat{u_{ij}}\right) \qquad (10)$$

assuming that the first k attributes are benefits while the remaining $(m–k)$ attributes are costs in $\otimes\widehat{D}$.

5.9 Relative Weight of Each Strategy

The importance of each strategy in the weighted normalized matrix can be calculated as follows [16, 17]:

$$Q_i = P_i + \frac{\sum_{i=1}^{n} R_i}{R_i \sum_{i=1}^{n} \frac{1}{R_i}} \qquad (11).$$

5.10 Utility of Each Strategy

The utility degree of each strategy can be calculated based on its relative weight as follows [16, 17]:

$$U_i = \frac{Q_i}{\max_{1 \le i \le n} Q_i} \qquad (12)$$

for i = 1, 2, .., n. The strategy with the highest utility degree is considered the best deployment strategy of the WSN under consideration given the m scenario attributes.

6 Case Study

Using simulation, the methodology is executed using a simplified deployment scenario consisting of rectangular deployment area measuring 500 m × 500 m; 50 to 100 nodes available for deployment with onboard radio capable of transmitting between 50 to 100 meters and sensors capable of covering between 30 to 60 meters. Cost is assumed to be directly proportional to the number of deployed nodes, and lifetime is related to transmission power. This power is simulated using the Log-Normal RF propagation model, whereby terrain obstructions are modeled as a zero-mean normally distributed random variable with standard deviation proportional to obstructions [18]. VVT was executed to reduce the number of deployment strategies based on the connectivity metric.

The case study identifies deployment strategies composed of specific number of nodes and radio range. Since there are 50, 60, …, n alternatives for the number-of-nodes factor and 40, 45, …, r alternatives for radio range; the initial number of deployment alternatives is $n \times r = 78$. Using VVT, the number of deployment alternatives was reduced to 51, resulting in 35% reduction of deployment strategies. In addition, a CMV value of 90% was used to further reduce the number of deployment strategies. The deployment alternatives for connectivity are summarized in Tables 2 and 3.

Table 2. Original set of deployment alternatives.

Nodes	Radio Ranges
50	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
60	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
70	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
80	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
90	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
100	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

Table 3. Trimmed set of deployment alternatives.

Nodes	Radio Ranges
50	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
60	50, 55, 60, 65, 70, 75, 85, 90, 95, 100
70	45, 55, 60, 65, 70, 75, 80, 85, 90, 95
80	40, 50, 55, 70, 75, 85, 90
90	40, 45, 55, 60 65, 70, 75, 80
100	50, 55, 65

Using the trimmed set of deployment alternatives, a deployment scenario in which network lifetime is of highest importance, followed by connectivity, cost, and coverage is created. This scenario is shown in Figure 2. This scenario shows 35 strategies where each strategy is characterized by the following attributes: number of deployed nodes, radio range of these nodes, range of the sensors on these nodes, connectivity of the network deployed by these nodes, coverage of this network, and its power usage. These attributes are represented by grey numbers derived from simulation experiments. The weights at the bottom of Figure 2 illustrate the priority goals of the deployment scenario. As the weight of each attribute shows, power usage has the highest weight, which means that network lifetime is of utmost importance. Next, connectivity, number of nodes, and coverage follow in importance based on their weights.

Strategy	Nodes		Radio Range (m)		Sensor Range (m)		Connectivity		Coverage		Power (mW)	
	l	u	l	u	l	u	l	u	l	u	l	u
7	50	55	100	105	30	35	109.82	122.15	44.61	59.36	209.74	237.91
12	60	65	95	100	30	35	108.22	118.70	51.04	134.78	183.07	209.74
13	60	65	100	105	30	35	116.71	123.29	51.04	134.78	209.74	237.91
19	70	75	95	100	30	35	114.55	119.77	56.72	184.11	183.07	209.74
23	80	85	90	95	30	35	113.08	117.70	61.67	246.20	157.90	183.07
33	50	55	100	105	35	40	109.82	122.15	56.14	99.97	209.74	237.91
38	60	65	95	100	35	40	108.22	118.70	62.40	138.08	183.07	209.74
39	60	65	100	105	35	40	114.55	123.29	62.40	138.08	209.74	237.91
45	70	75	95	100	35	40	114.55	119.77	67.93	187.14	183.07	209.74
49	80	85	90	95	35	40	113.08	117.70	72.71	248.97	157.90	183.07
59	50	55	100	105	40	45	109.82	122.15	66.17	100.41	209.74	237.91
64	60	65	95	100	40	45	108.22	118.70	72.17	138.15	183.07	209.74
65	60	65	100	105	40	45	116.71	123.29	72.17	138.15	209.74	237.91
71	70	75	95	100	40	45	114.55	119.77	77.43	186.85	183.07	209.74
75	80	85	90	95	40	45	113.08	117.70	81.95	248.31	157.90	183.07
85	50	55	100	105	45	50	109.82	122.15	74.93	97.72	209.74	237.91
90	60	65	95	100	45	50	108.22	118.70	80.56	135.00	183.07	209.74
91	60	65	100	105	45	50	116.71	123.29	80.56	135.00	209.74	237.91
97	70	75	95	100	45	50	114.55	119.77	85.45	183.23	183.07	209.74
101	80	85	90	95	45	50	113.08	117.70	89.61	244.22	157.90	183.07
111	50	55	100	105	50	55	109.82	122.15	82.63	91.91	209.74	237.91
116	60	65	95	100	50	55	108.22	118.70	87.80	128.62	183.07	209.74
117	60	65	100	105	50	55	116.71	123.29	87.80	128.62	209.74	237.91
123	70	75	95	100	50	55	114.55	119.77	92.23	176.29	183.07	209.74
127	80	85	90	95	50	55	113.08	117.70	95.92	236.72	157.90	183.07
137	50	55	100	105	55	60	109.82	122.15	89.51	82.97	209.74	237.91
142	60	65	95	100	55	60	108.22	118.70	94.12	119.01	183.07	209.74
143	60	65	100	105	55	60	116.71	123.29	94.12	119.01	209.74	237.91
149	70	75	95	100	55	60	114.55	119.77	97.98	166.02	183.07	209.74
153	80	85	90	95	55	60	113.08	117.70	101.11	225.78	157.90	183.07
160	50	55	100	105	60	65	109.82	122.15	95.80	70.91	209.74	237.91
165	60	65	95	100	60	65	108.22	118.70	99.73	106.19	183.07	209.74
166	60	65	100	105	60	65	116.71	123.29	99.73	106.19	209.74	237.91
172	70	75	95	100	60	65	114.55	119.77	102.93	152.42	183.07	209.74
176	80	85	90	95	60	65	113.08	117.70	105.39	211.42	157.90	183.07
Weight	0.30		0.10		0.10		0.40		0.20		0.50	

Figure 2. Deployment scenarios.

After normalization and weighting, the weighted normalized matrix appears in Figure 3. Among the attributes of this scenario, radio range, sensor range, connectivity and coverage are maximized whereas number of nodes and power usage are minimized. As such, the benefits of radio range, senor range, connectivity and coverage are computed using equation (9) while the costs of power usage are computed using equation (10). Next, the relative importance of all strategies and their utility degrees are computed using equation (11) and (12). Figure 4 shows the final results where strategies 176, 153 and 127 are the top three strategies that are highly desirable in the simulated scenario. Although strategy 176 requires a high

number of nodes, it shows a high range of nodes and sensors as well as a high degree of connectivity and coverage.

Strategy	Nodes		Radio Range (m)		Sensor Range (m)		Connectivity		Coverage		Power (mW)	
	l	u	l	u	l	u	l	u	l	u	l	u
7	0.0064	0.0071	0.0029	0.0030	0.0018	0.0021	0.0108	0.0120	0.0022	0.0029	0.0148	0.0168
12	0.0077	0.0084	0.0028	0.0029	0.0018	0.0021	0.0106	0.0117	0.0025	0.0067	0.0129	0.0148
13	0.0077	0.0084	0.0029	0.0030	0.0018	0.0021	0.0115	0.0121	0.0025	0.0067	0.0148	0.0168
19	0.0090	0.0097	0.0028	0.0029	0.0018	0.0021	0.0112	0.0118	0.0028	0.0091	0.0129	0.0148
23	0.0103	0.0110	0.0026	0.0028	0.0018	0.0021	0.0111	0.0116	0.0030	0.0122	0.0112	0.0129
33	0.0064	0.0071	0.0029	0.0030	0.0021	0.0024	0.0108	0.0120	0.0028	0.0049	0.0148	0.0168
38	0.0077	0.0084	0.0028	0.0029	0.0021	0.0024	0.0106	0.0117	0.0031	0.0068	0.0129	0.0148
39	0.0077	0.0084	0.0029	0.0030	0.0021	0.0024	0.0112	0.0121	0.0031	0.0068	0.0148	0.0168
45	0.0090	0.0097	0.0028	0.0029	0.0021	0.0024	0.0112	0.0118	0.0034	0.0092	0.0129	0.0148
49	0.0103	0.0110	0.0026	0.0028	0.0021	0.0024	0.0111	0.0116	0.0036	0.0123	0.0112	0.0129
59	0.0064	0.0071	0.0029	0.0030	0.0024	0.0027	0.0108	0.0120	0.0033	0.0050	0.0148	0.0168
64	0.0077	0.0084	0.0028	0.0029	0.0024	0.0027	0.0106	0.0117	0.0036	0.0068	0.0129	0.0148
65	0.0077	0.0084	0.0029	0.0030	0.0024	0.0027	0.0115	0.0121	0.0036	0.0068	0.0148	0.0168
71	0.0090	0.0097	0.0028	0.0029	0.0024	0.0027	0.0112	0.0118	0.0038	0.0092	0.0129	0.0148
75	0.0103	0.0110	0.0026	0.0028	0.0024	0.0027	0.0111	0.0116	0.0040	0.0123	0.0112	0.0129
85	0.0064	0.0071	0.0029	0.0030	0.0027	0.0030	0.0108	0.0120	0.0037	0.0048	0.0148	0.0168
90	0.0077	0.0084	0.0028	0.0029	0.0027	0.0030	0.0106	0.0117	0.0040	0.0067	0.0129	0.0148
91	0.0077	0.0084	0.0029	0.0030	0.0027	0.0030	0.0115	0.0121	0.0040	0.0067	0.0148	0.0168
97	0.0090	0.0097	0.0028	0.0029	0.0027	0.0030	0.0112	0.0118	0.0042	0.0091	0.0129	0.0148
101	0.0103	0.0110	0.0026	0.0028	0.0027	0.0030	0.0111	0.0116	0.0044	0.0121	0.0112	0.0129
111	0.0064	0.0071	0.0029	0.0030	0.0030	0.0033	0.0108	0.0120	0.0041	0.0045	0.0148	0.0168
116	0.0077	0.0084	0.0028	0.0029	0.0030	0.0033	0.0106	0.0117	0.0043	0.0064	0.0129	0.0148
117	0.0077	0.0084	0.0029	0.0030	0.0030	0.0033	0.0115	0.0121	0.0043	0.0064	0.0148	0.0168
123	0.0090	0.0097	0.0028	0.0029	0.0030	0.0033	0.0112	0.0118	0.0046	0.0087	0.0129	0.0148
127	0.0103	0.0110	0.0026	0.0028	0.0030	0.0033	0.0111	0.0116	0.0047	0.0117	0.0112	0.0129
137	0.0064	0.0071	0.0029	0.0030	0.0033	0.0036	0.0108	0.0120	0.0044	0.0041	0.0148	0.0168
142	0.0077	0.0084	0.0028	0.0029	0.0033	0.0036	0.0106	0.0117	0.0047	0.0059	0.0129	0.0148
143	0.0077	0.0084	0.0029	0.0030	0.0033	0.0036	0.0115	0.0121	0.0047	0.0059	0.0148	0.0168
149	0.0090	0.0097	0.0028	0.0029	0.0033	0.0036	0.0112	0.0118	0.0048	0.0082	0.0129	0.0148
153	0.0103	0.0110	0.0026	0.0028	0.0033	0.0036	0.0111	0.0116	0.0050	0.0112	0.0112	0.0129
160	0.0064	0.0071	0.0029	0.0030	0.0036	0.0039	0.0108	0.0120	0.0047	0.0035	0.0148	0.0168
165	0.0077	0.0084	0.0028	0.0029	0.0036	0.0039	0.0106	0.0117	0.0049	0.0052	0.0129	0.0148
166	0.0077	0.0084	0.0029	0.0030	0.0036	0.0039	0.0115	0.0121	0.0049	0.0052	0.0148	0.0168
172	0.0090	0.0097	0.0028	0.0029	0.0036	0.0039	0.0112	0.0118	0.0051	0.0075	0.0129	0.0148
176	0.0103	0.0110	0.0026	0.0028	0.0036	0.0039	0.0111	0.0116	0.0052	0.0104	0.0112	0.0129

Figure 3. Weighted normalized matrix.

Strategy	Relative Weight	Utility
7	0.0420	86.40%
12	0.0443	91.15%
13	0.0432	88.80%
19	0.0447	91.98%
23	0.0466	95.83%
33	0.0436	89.66%
38	0.0450	92.52%
39	0.0437	89.94%
45	0.0454	93.32%
49	0.0472	97.15%
59	0.0442	90.81%
64	0.0455	93.63%
65	0.0444	91.28%
71	0.0459	94.41%
75	0.0477	98.21%
85	0.0446	91.74%
90	0.0460	94.52%
91	0.0448	92.16%
97	0.0463	95.25%
101	0.0481	99.01%
111	0.0450	92.46%
116	0.0463	95.18%
117	0.0451	92.83%
123	0.0466	95.86%
127	0.0484	99.57%
137	0.0452	92.97%
142	0.0465	95.63%
143	0.0454	93.28%
149	0.0468	96.25%
153	0.0486	99.89%
160	0.0454	93.29%
165	0.0466	95.89%
166	0.0455	93.53%
172	0.0469	96.43%
176	0.0486	100.00%

Figure 4. Relative weights and utility degrees of all strategies.

With this information, the utility degree function derived from the weighted normalized matrix can be used to establish the goals for the deployment, rank the strategies and select the best ones desired for deployment.

7 Conclusion

A methodology for quantifying stochastic WSN deployments based on mission goals has been presented. This methodology uses simulation, data reduction through VVT, and a grey number-based decision making approach to rank alternative strategies in different deployment scenarios. This work presents a contribution to the current literature by providing an innovative approach to decision-making that considers all factors involved in the deployment of WSN. The approach can be easily customized to include numerous quality factors to further compare deployment strategies.

8 References

[1] H. Karl, A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, Wiley, 1st Edition, 2007.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey", *Computer Networks*, vol. 38, no. 4, pp. 393-422, 2002.

[3] P. Santi, "Topology Control in Wireless Ad Hoc and Sensor Networks", *ACM Computing Surveys*, vol. 37, no. 2, pp. 164-194, 2005.

[4] M. Younis and K. Akkaya, "Strategies and Techniques for Node Placement in Wireless Sensor Networks: A Survey", vol. 6, pp. 621-655, 2008.

[5] S. S. Dhillon and K. Chakrabarty, "Sensor Placement for Effective Coverage and Surveillance in Distributed Sensor Networks". *Proc. of IEEE Wireless Communications and Networking Conference*, New Orleans, LA, March, 2003.

[6] A. Efrat, S. Har-Peled, and J. S. B. Mitchell, "Approximation Algorithm for Two Optimal Location Problems in Sensor Networks", *Proc. of the 3rd International Conference on Broadband Communications, Networks and Systems*, Boston, Massachusetts, October, 2005.

[7] M. Ishizuka and M. Aida, "Performance Study on Node Placement in Sensor Networks", *Proc. of the 24th International Conference on Distributed Computing Systems Workshops* – W7:EC (Icdesw'04) –Volume 7, 2004.

[8] C. E. Otero, I. Kostanic, and L. D. Otero, "A Multi-hop, Multi-segment Architecture for Perimeter Security over Extended Geographical Regions using Wireless Sensor Networks", *Proc. of the 2008 IEEE Wireless Hive Network Conference*, Texas, 2008.

[9] S. Toumpis and L. Tassiulas, "Packetostatics: Deployment of Massively Dense Sensor Networks as an Electrostatic Problem", *Proc. of the 24th IEEE Conference on Computer Communications and Networking*, Miami, FL, March, 2005

[10] K. Xu, H. Hassanein, G Takahara, and W. Wang, "Relay Node Deployment Strategies in Heterogeneous Wireless Sensor Networks: A Single-Hop Communication Case", *Proc. of the IEEE Global Telecommunication Conference*, St. Louis, MO, November, 2005

[11] K. Xu, H. Hassanein, G. Takahara, and W. Wang, "Relay Node Deployment Strategies in Heterogeneous Wireless Sensor Networks: A Multiple-Hop Communication Case", *Proc. of the 2nd IEEE Conference on Sensor and Ad Hoc Communications and Networks*, Santa Clara, CA, September, 2005.

[12] C. E. Otero, I. Kostanic, L. D. Otero, S. L. Meredith, S.L. and M. Whitt, M. "Analysis of Stochastic WSN Deployments using Vertical Variance Trimming and the Analytical Hierarchy Process," *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, vol. 2,no. 1, pp. 169-190, 2011.

[13] D.. Montgomery, *Design and Analysis of Experiments*, Wiley, 5th Edition, 2001.

[14] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[15] S. Liu and Y. Lin, *Grey Systems: Theory and Applications*, Springer-Verlag, Berlin Heiderlberg, 2010.

[16] E. K. Zavadskas, A. Kaklauskas, Z. Turskis, and J. Tamosatitiene, "Multi-Attribute Decision-Making Model by Applying Grey Numbers," *Informatica*, vol. 20, no. 2, pp. 305-320, 2009.

[17] E. K. Zavadskas, Z. Turskis, J. Tamosaitiene, and V. Marina, "Selection of Construction Project Managers by Applying COPRAS-G Method," *International Conference on Reliability and Statistics in Transportation and Communication*, Riga, Latvia, 344-350, 2008.

[18] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 1996

Redundant Hop-count Mapping in a Random Wireless Sensor Network

Raymundo R. Hordijk and Oscar J.G. Somsen
SeWaCo, Netherlands Defense Academy, Den Helder, Noord-Holland, Netherlands.

Abstract - *Networks composed of a large number of tiny sensors, also known as "smart dust", can provide long range observation of phenomena that can only be detected at short range or contact. Mapping is necessary so that information from a particular sensor can be linked to a location within the network. Previously we determined that distances can be accurately measured by counting the number of hops or retransmissions that are necessary to transmit a message between two sensors and that these can also be used to determine angles and relative positions. In the present paper we investigate whether it is possible to extend the number of nodes to four and determine the angles more accurately. This works well, especially for very large or very small angles.*

Keywords: Networks mapping, tiny sensors, smart dust.

1 Introduction

Targets or events are often located by long range systems such as radar or sonar. However, some phenomena such as vibration or temperature can only be detected at short range. Detection of such phenomena require several sensors or even a network. We are interested in a network that would contain a large (~10^4) number of tiny sensors. Sensors of millimeter size have indeed been developed including energy generation and storage, sensory element and communication [1]. This type of network could be described as smart dust and has been proposed to secure military terrain [2] or e.g. in a vineyard [3].

Mapping is important in sensor networks. Information obtained from a sensor is of limited use if its location is unknown. In tiny, inexpensive sensors navigation aids such as GPS are not available [4]. Alternatively the communication aid can be used to determine distances [5] or angles [6]. A tiny sensor is likely to be equipped with a weak communication aid. It can only contact a limited number of sensors in the immediate neighborhood. In order to transmit a message across the network the neighboring sensor nodes need to retransmit it to their neighbors. This step has to be repeated a number of times until the message has reached its destination. The number of retransmissions or hops that is necessary to transmit a message between two sensor nodes may be used to estimate the distance between them.

Previously we have carried out simulations to investigate the effectiveness of hop-counting as a distance measure [2]. While the distance was typically overestimated the fluctuation of the hop-count (observed distance) between many simulated network realizations is small so that calibration is possible. When each node communicates with 5-10 neighbors, sufficiently large distances could be estimated with 1% accuracy or better in a homogeneous network. With three or more sensor nodes, the obtained distances can be used to determine angles. We proposed a protocol by which sensor nodes can use triangulation to determine their position within a wireless network [7]. While some bad cases occur e.g. when the three nodes are collinear. The angle measurements are much less sensitive to the aforementioned overestimate and accurate angles could indeed be obtained, even without calibration.

So far we only considered the mapping of three nodes in each network, in the present paper we consider a fourth node which will provide more information to estimate angles and relative positions. We investigate whether this redundance does indeed improve the accuracy of angle measurements. We present a number of simulations to determine the accuracy of these estimates and the possibility to use them to refine the position estimates within the network.

2 Hop distance and three-node simulations

The concept of hop-counting has been explained previously [2]. Briefly, two nodes are neighbors when they are within a hop-range from each other. A hop-route is a series of nodes where each is a neighbor to the next. The hop-count or hop-distance between two nodes is defined by the shortest possible hop-route that connects them. In an actual network this distance can be obtained by transmitting a message that includes the number of re-transmissions. A node may receive the message more than once but will re-transmit only the one with the lowest re-transmission number. This number is also its hop-distance from the origin (original transmitter).

In a computer simulation hop-distance between two node is done with a global search. This is most efficient with a procedure similar to the above. We first determine which nodes are neighbors to each other. Once a transmitting node has been selected we first determine the nodes that are located at a distance of one hop form this node, then the ones located at two hops, etc. until the network has been covered. This provides the hop-distance to all nodes within the network.

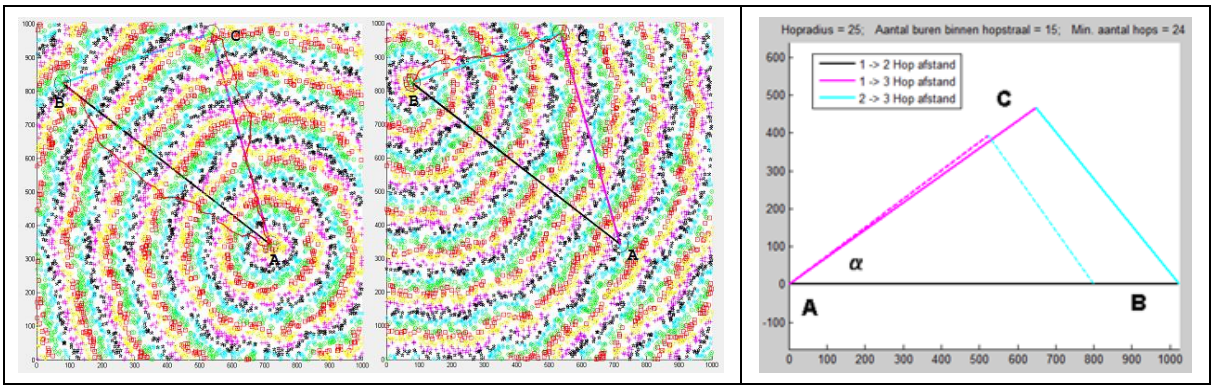


Figure 1: Example of a hop-count measurement with two transmissions.
Left: Transmission by nodes A and B. The colored rings illustrate the retransmission process (number of hops). Shortest hop routes are shown (red) for receiving nodes together with the straight line (black).
Right pane: Estimated positions. Dotted lines represent the 'real' distances between the reference nodes, and solid lines the 'hop' distances. The two triangles were superimposed with nodes A in the origin and nodes B on the positive x-axis and nodes C above this axis. Note that the result is a mirror image of the actual triangle (in the left pane)

Figure 1 illustrates an actual situation. In this case a network of 7600 nodes was simulated by randomly scattered in an square area of size 1000 x 1000 m. The hop-range (i.e. the range at which neighbors can contact each other) was set to 25 m so that on average a node is able to contact 15 neighbors. After initialization all nodes determine a random time at which they will transmit a message. The node with the earliest time actually transmits the message and is denoted "A". It arbitrarily places itself in the origin. All other nodes postpone their transmission. A collision may occur when a second node transmits a message before the message of "A" has travelled across the network. As shown the transmission starts by A sending the message to its neighboring nodes. These nodes forward the message to their neighbors and so on until the message has travelled across the network. The forwarded message includes the number of times that it has been forwarded and each receiving node can use this as an estimate of its distance to node A.

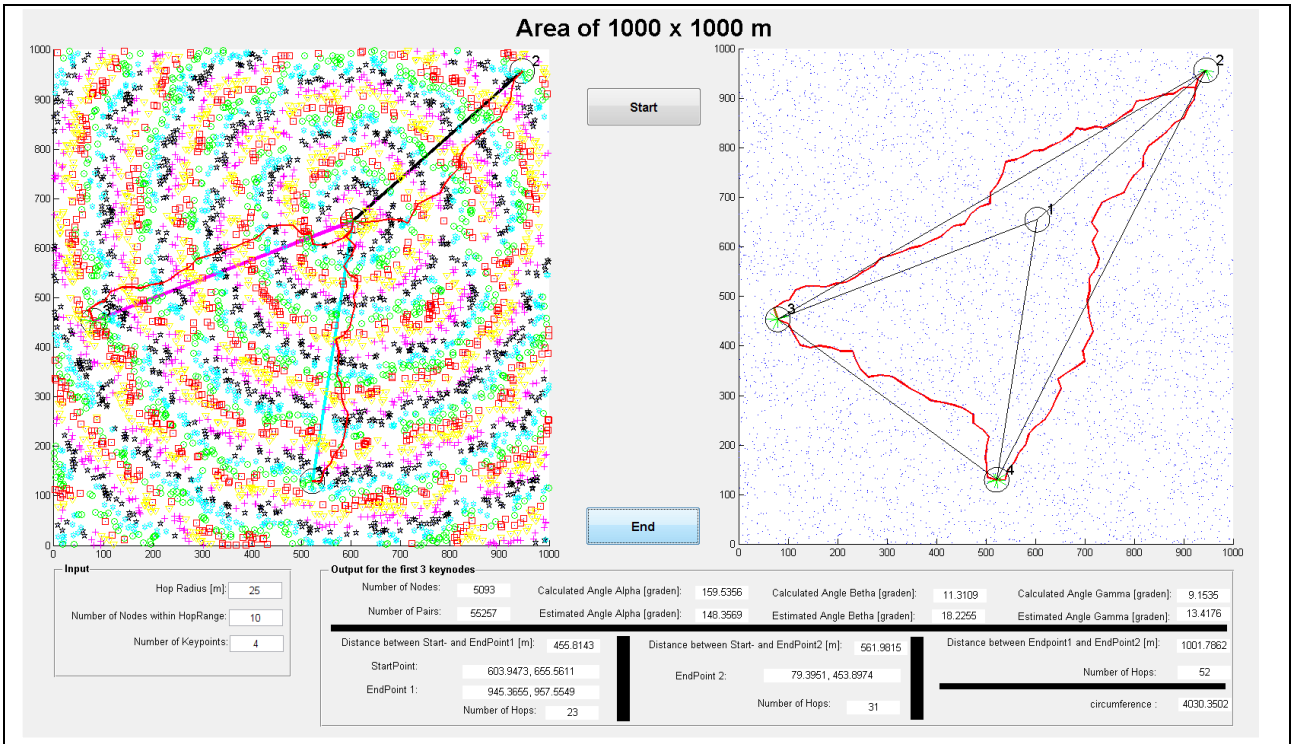
Once the first message has been received all nodes (except "A") determine a new random time at which they will transmit a message. Again the node with the earliest time is the actual transmitter and is denoted "B" it places itself (also arbitrarily) at the positive x-axis and includes its position estimate in the message. Again all other nodes postpone their transmission and use the number of times that this message has been forwarded as their distance with respect to node "B". As can be seen in figure 1, the shortest hop routes between transmitting and receiving nodes do not exactly follow the lines of sight and thus the distance derived from them overestimates the actual distance. The extent of this deviation depends on hop-range, node density and the length of the route.

As illustrated in figure 1, the distances observed in the above two transmissions can be used to estimate the shape of triangle ABC. In a simulation this can be compared to the triangle that can be determined from the actual positions of the nodes. We have studied this previously [7] and found a difference in distance and angle estimates. Since all three distances are overestimated the obtained triangle is larger than the actual triangle. However, since this overestimate is typically constant, the angles in triangle ABC form much better estimates of the actual angles especially when they are in a moderate range (between 20° and 120°). The orientation of the triangle cannot be determined. Also, two triangles can be constructed, which are each other's mirror image. However, based on the distances measured, triangle ABC provides the best possible map of the network.

3 Four-node simulations

An opportunity to improve the accuracy of the map occurs when more messages are transmitted. Figure 2 illustrates the extension of the process with a third transmission. The symbols (A, B, …) for the nodes have been replaced by numbers (1, 2, ..) so that it will be possible to continue to a arbitrary number. The node density was reduced to 5100 so that the hop range now includes an average of 10 neighbors (all other parameters were unchanged). The effect should be that the quality of the position estimates becomes less and the correction by including node 3 becomes more necessary and more visible.

As before, nodes 1 and 2 transmit a message and node 3 estimates its position relative to these nodes. Only the first transmission is shown explicitly in figure 2. The next step is that also node 3 transmits a message. With this information the distances between nodes 2, 3 and 4 are determined and used to estimate the position of node 4. Circles are drawn around the estimated positions of nodes 1, 2 and 3 and the intersections are determined. Three interactions are found near to each other (other intersections are far from this region and can be disregarded). These are the estimates of the position of node 4. Their average is determined and used as the best estimate.



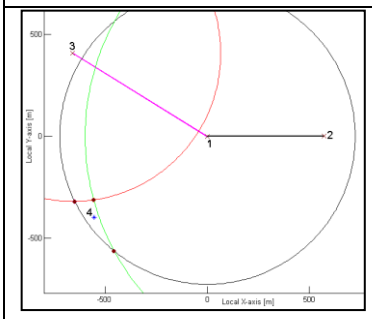


Figure 2: Hop-count measurement with three transmissions.
Topleft: User interface for the simulation. Shown is the transmission by node 1 that is received by nodes 2, 3 and 4. See figure 1 for an explanation of the symbols.
Topright: The shortest routes between nodes 2, 3 and 4. Blue dots represent the other nodes in the network.
Left: Position estimate for node 4: Circles are drawn around the estimated positions for nodes 1-3 (crosses) using the measured distances to node 4. Three intersections are found near to each other (dots). Their average is taken as the position estimate for node 4.

To investigate this we generated 50 random networks and analyzed the performance of position estimation using two- and three transmissions. The results are shown in figure 3. Since the positions of the nodes are determined relative to each other, the performance of the approach should not be determined by the positions themselves. Rather, figure 3 shows a comparison of the angle (2-1-4). Shown are the initially estimated angle (using only nodes 1 and 2 as a reference) and the corrected estimate (using also node 3). In the central region (angles between 40° and 100°) the initial estimates are best. The correction regularly improves this estimate, though not in all cases.

Larger initial errors occur in the extreme regions (angles below 20° or above 120°). In those cases large errors occur in quite a number of cases. The reason for this is that at those angles node 4 is more or less collinear with nodes 1 and 2 so that distance measurements are not suitable to determine its position. This is where the inclusion of a third reference node is most needed. In general this node will not be collinear with nodes 1 and 2. Therefore the position of node 3 is estimated with more accuracy and in turn node 3, in combination with node 1 or 2 should give a more accurate position estimate for node 4. This is indeed observed. In the extreme region, the correction with node 3 almost always leads to a more accurate estimate of angle 2-1-4. However, also in this case there are exceptions.

To get a better understanding why correction with node 3 does not always improve the position estimate of node 4, we have shown a few cases explicitly. In case I the angle 2-1-4 was initially estimated accurately at 140°. After correction the estimate was some 10° off. An explanation for this can be found in the corresponding diagram. In this case node 3 was even more collinear with nodes 1 and 2 than node 4 itself. Thus, the position estimate for node 3 is likely to be inaccurate. While node 3 is a good extra reference, in principle, to improve the position estimate for node 4, the uncertainty of its own position leads to a decreased accuracy.

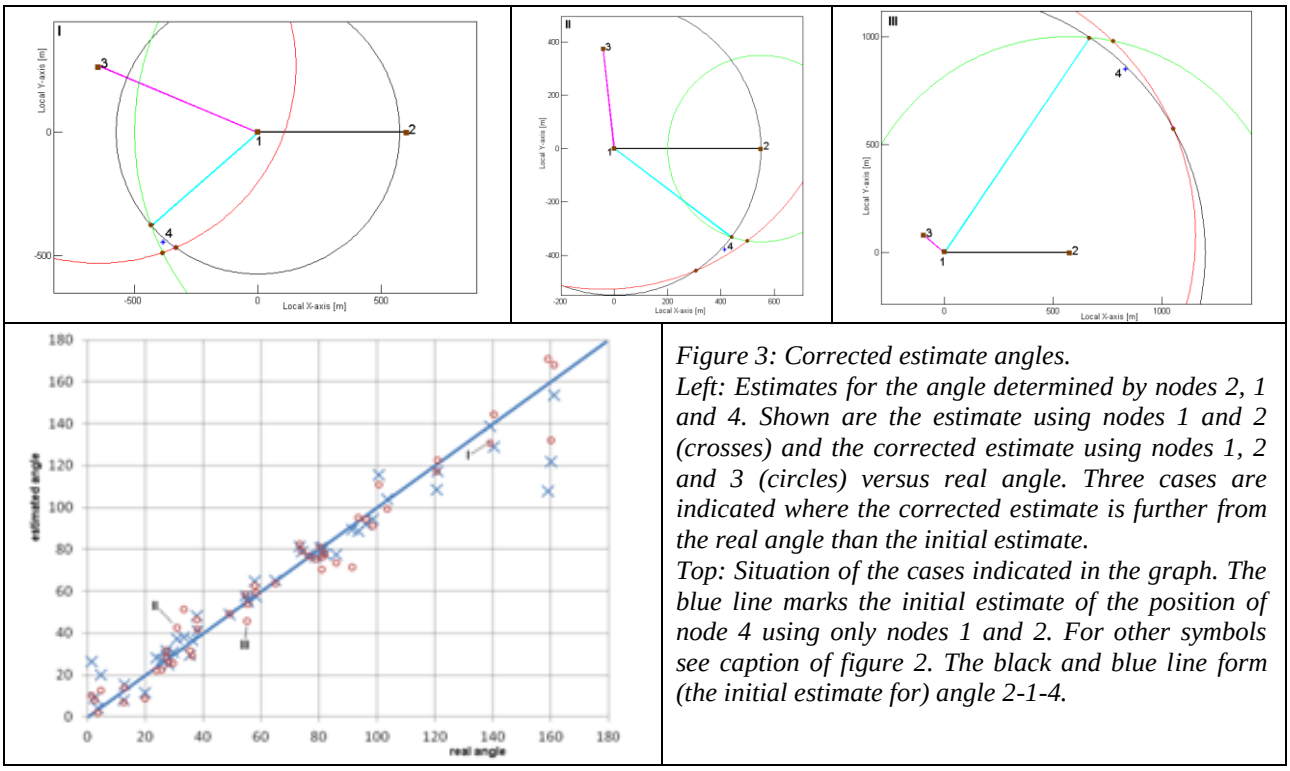


Figure 3: Corrected estimate angles.
Left: Estimates for the angle determined by nodes 2, 1 and 4. Shown are the estimate using nodes 1 and 2 (crosses) and the corrected estimate using nodes 1, 2 and 3 (circles) versus real angle. Three cases are indicated where the corrected estimate is further from the real angle than the initial estimate.
Top: Situation of the cases indicated in the graph. The blue line marks the initial estimate of the position of node 4 using only nodes 1 and 2. For other symbols see caption of figure 2. The black and blue line form (the initial estimate for) angle 2-1-4.

Case II is especially unfortunate. In this case angle 2-1-4 is not extreme and indeed the initial position estimate for node 4 was accurate. However, in this case node 3 is more or less collinear with nodes 1 and 4. Thus the position estimate of node 3 is probably accurate and nodes 2 and 3 provide a good basis to determine the position for node 4. But the position estimate determined with nodes 1 and 3 is much worse and reduces the accuracy of the final estimate. Yet another problem occurs in case III. In this case node 3 is very near to node 1. In that case the position estimate obtained by using these two nodes as reference is very inaccurate. While the other two estimates (using nodes 1 and 2 or nodes 3 and 2) are more accurate, the final result also shows a large error.

4 Discussion

We are investigating whether hop-distances, i.e. distance estimates obtained by counting the number of re-transmissions of a message as it passes through the network, can be used to reliably estimate positions of nodes within the network. Of course, only relative positions can be determined with this approach. The map that is finally obtained can be rotated or even mirrored with respect to the actual network. It is not possible to distinguish such cases using only distances. Also the size of the map is somewhat overestimated since hop-counting tends to overestimate the distance [2]. However, angles can be obtained with better accuracy [7] and provide information of the network structure that can be used for many purposes.

The type of mapping that we investigate is incremental. The first step occurs when a (random) node transmits a message. It will arbitrarily place itself in the origin. All receiving nodes then know their distance from the origin, but angles cannot yet be determined. When a second node (that has arbitrarily placed itself at the positive x-axis) transmits a message the receiving nodes know their distance from both nodes and can begin to estimate their position. There are still two possibilities. One above and one below the x-axis. When a third node (that has arbitrarily placed itself above the x-axis) transmits a message the receiving nodes can use their observed distance with respect to this node to decide between the aforementioned possibilities and arrive at a unique position estimate.

As of the third transmission it also becomes possible to refine the position estimates. This is what we have investigated in our present research. Our results indicate that the position estimate is indeed improved in many cases, though not in all. The improvement is most noticeable when the receiving node is more or less collinear with first two transmitting nodes. In that case these do not provide a suitable basis to determine it position using only distances. If the third transmitting node is not on the same line it helps to provide a much more suitable basis and can considerably improve the position estimate.

However, the enhancement does not occur in all cases. This might have been caused by inaccuracy of the distance measurements themselves, but this does not appear to be the case. In the cases that we have studied the position of the third transmitting node is usually the cause. Either it itself more or less collinear with the first two transmitting nodes so that its position estimate is inaccurate. Or it is in a position where it does not provide

a suitable basis to accurately estimate the position of the receiving node, even if the first two nodes did provide an suitable basis. These are the cases where the position estimate becomes less accurate by the correction.

The procedure for the position enhancement that we used here is still very coarse. We simply used the average of the three estimated positions independent of the above considerations. It should therefore be possible to improve the procedure. If the receiving node notices that the third transmitting node is not suitable to enhance the position estimate it might discard one or both of the two estimates and take the average of the remaining ones. An even better approach is possible if the accuracy of the three position estimates can also be estimated. In that case a weighted average of the three estimates could be obtained.

If the positions of the three transmitting nodes is less than optimal a receiving node may not yet be able to determine its position accurately. However, this should not be a major problem since the mapping process does not stop at this third transmission. At some later time a fourth transmission will occur and all receiving nodes will establish their distance from four transmitting nodes. If the first three transmitting nodes did not provide a suitable basis, the fourth one might. In fact, if the transmission occurs only for the mapping process the transmitting node need not be random. Each node could evaluate whether it provides a suitable addition to the basis and should therefore send a message or that it should better wait first for another, possibly more suitable, node to transmit a message. Since the size and shape of the network is not known at first, it is not possible to determine which node is the most suitable addition to the basis but some kind of selection is possible to make it more likely that it is. This selection may already be possible for the third and maybe even the second transmission.

The mapping process continues with each transmission. At some point a set of nodes spread all over the network should be available as a basis for position estimates. Some of the older nodes may be discarded from this set because they are less accurate, may no longer be accurate or simply because of limited data storage capacity. The above problems should no longer occur since this basis should always be suitable. With a sufficiently large basis set it may even be possible to estimate sensor positions in three-dimensions.

There are many advantages of using this mapping process with a changing set of basis nodes [8]. The system shows graceful degradation. Since any node can be part of the basis there are always sufficient basis nodes. If a node stops working it will automatically disappear from the basis. Also the basis set can be relatively large. This is especially useful if the network shows obstructions so that a transmission cannot always follow a straight path [9]. This will initially lead to false position estimates. However the basis of correctly positioned nodes grows until it spans the entire network and provides correct estimates everywhere. A main advantage is of course that there is no need for separate more expensive nodes to function as a positioning basis.

Of course there are also disadvantages. Each node has to perform many calculations to continually update and enhance its position estimate. Also separate positioning nodes may be equipped with accurate positioning devices such as GPS. The fact that this is not available in our network leaves the entire mapping process dependent on relatively inaccurate hop-distance measurements. This should be overcome by using indeed a relatively large basis set. However, the advantages of a flexible positioning basis set are large enough to continue researching this process.

5 References

[1] B. Warneke, M. Last, B. Liebowitz and K.S.J. Pister. Smart dust: "Communicating with a cubic-millimeter computer". COMPUTER 34:44-51, 2001.

[2] O.J.G. Somsen, R.R. Hordijk and Th.M. Hupkens. "Applicability of hop distance in random sensor networks". Wireless Personal Communications: DOI 10.1007/s11277-011-0376-6 (13 pp), 2011.

[3] A. Matese, S.F. Di Gennaro, A. Zaldei,, L. Genesio and F.P. Vaccari. "A wireless sensor network for precision viticulture": The NAV system. Computers and electronics in agriculture 69:51-58, 2009.

[4] S Čapkun, M Hamdi, and J-P Hubaux. "GPS-free positioning in mobile Ad-Hoc networks". Proceedings of the IEEE Hawaii Conference on System Sciences: 255-264, 2001.

[5] N Patwari, A Hero, AM Perkins, N Correal and B O'Dea. "Relative Location Estimation in Wireless Sensor Networks". IEEE Transactions on Signal Processing 51: 2137-2148, 2003.

[6] J. Bruck, J. Gao and A. Jiang. "Localization and Routing in Sensor Networks by Local Angle Information". ACM transactions on sensor networks 5, Article Number 7, 2009.

[7] R.R. Hordijk, Th.M. Hupkens, O.J.G. Somsen. "Towards hop-count mapping in a random wireless sensor network". In: Proceedings of the International Science and Technology Conference. Istanbul: 869-874, Dec 2011.

[8] RL Moses, D Krishnamurthy and RM Patterson. "A self localization method for wireless sensor networks". EURASIP Journal on Applied Signal Processing 4: 348-359, 2003.

[9] Y Kong, Y Kwoan and G Park. "Robust Localization over Obstructed Interferences for In building Wireless Applications". IEEE Transactions on consumer electronics 55: 105-111, 2009.

Polling-based Medium Access Control Scheme for Wireless Body Sensor Network

Tatiana Anonni Pazeto[1], Luis Fernando Refatti[2], Shusaburo Motoyama[3*]
[1]Informatics Department, Federal University of Mato Grosso (UFMT), Rondonópolis, Mato Grosso, Brazil
[2]Automation Department, Federal University of Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brazil
[3]Faculty of Campo Limpo Paulista (FACCAMP), Campinas, São Paulo, Brazil

Abstract - *The performance of polling-based MAC scheme for Wireless Body Sensor Networks (WBSN) is investigated in this paper. The objective is to study the suitability of the polling mechanism to gather the quasi real-time data from sensors placed in a human body. The main parameters used for study are packet loss, packet waiting time and buffer size at sensor nodes. To accomplish the objectives, a simulation platform is developed in C++ where the polling mechanism, the buffers and the sensor sources are all implemented. Due to the lack of sensor node models for WBSN applications, five sensor sources are developed based on the On/Off source model. The simulation results showed that the polling method is a very promising access scheme for the scenarios examined. The results also showed that source configurations can greatly affect the network performance.*

Keywords: Wireless Body sensor network; polling; performance; simulation; On/Off sources

1 Introduction

Wireless Sensor Networks (WSN) are composed of tiny electronic devices performing remote monitoring and have applications in different areas, such as, environment, plantation, human body, and others. In the near future these networks will be more present in various environments and scenarios.

With the great development of the microelectronic field, the sensors that are the nodes of WSNs became smaller, and some models can be referred to as nano-sensors. Due to the small size of the nodes and batteries, in addition to the the limited energy storage capacity, the sensors can be placed in locations of limited access, resulting in difficulties recharging or exchanging the batteries. Thus, the sensor nodes, apart from the operation with features designed, must mainly save energy.

Among the tasks performed by the sensors, one that more degrades the battery is the communication. Since the environment of transmission among the sensors is the air, if more than one node begins to transmit packets simultaneously, collisions will occur, and packets must be retransmitted. In the case of WSNs where the sensors are distributed over large geographic areas, these communications are complex because in addition to try to reduce the collisions, each node must discover a route and forward the packets to the next node. For wireless body sensor networks (WBSN), where the sensors are located in a restricted area (human body), the use of the centralized node or sink node is more convenient because it simplifies the communication, and it is able inclusive to avoid completely the collisions. Using the centralized or star configuration, the medium access control (MAC) scheme can be ordered, preventing that more than one node begin to transmit packets at the same time, avoiding packet collisions.

The main MAC scheme proposed in the literature for the WBSN is the standard 802.15.4 with beacon enabled star configuration which provides very low energy consumption [1]. However, since the scheme is not designed for WBSN applications some drawbacks have been pointed out [2] and recently many schemes of MAC protocols specifically for WBSN have been proposed [2-16]. Some proposals are variation of standard 802.5.4 [5], [8] and [11] and others are based on TDMA access technique [3], [4], [7], [10], [14], [15] and [16]. Each of the proposals explores some special features based on medical needs. For instance in [3-4] to deal with the light and heavy loads in the normal and urgent situations, a context aware MAC is proposed. To guarantee QoS of a WBSN in [12] it is proposed a MAC protocol based on random access technique. The proposal presented in [10], the heart beating is used for the purpose of clock synchronization. In [6] the beacon used for wake-up sensor nodes is used for battery charging, increasing the network life time.
In all of the above proposals, the nodes must be woken up periodically to synchronize the node clock with the centralized node clock using the beacon signal.

In this paper, the polling-based access scheme that avoids the need for periodical synchronization is examined. The polling access technique inspects each node in a predetermined sequence. At each inspection, if the sensor has data, it is transmitted - If not, the next node of the cycle is inspected, and so on, until the last node is completed. The

* This work was partially supported by FAPESP under grant No 2011/12463-0

cycle starts again once all nodes have been inspected. When a node is not inspected it can be in an "Off" state, thus saving energy. Another advantage of the polling access technique is that it allows for a different order of access to the sensors enabling QoS capability in WBSN.

This paper is divided into five sections. The second section is related to the operation of the polling access scheme for the wireless body sensor network. In the third section, five different sources developed and the scenarios to be analyzed are described. The simulation results obtained in different scenarios are presented in section four. Finally, the main conclusions are presented in section five.

2 WBSN and Scheduling

A WSN composed by biological sensors designed to monitor vital signs of a human body is usually called Wireless Body Sensor Network (WBSN). The WBSN - composed of many sensor nodes with processing, communications and limited energy capabilities - has the function of monitoring various activities of the human body, facilitating the attendance of patients who require remote medical attention.

Fig. 1 shows the configuration of a WBSN as proposed in [17]. As it can be seen in Fig. 1 the human body was divided into many regions: the head (region 1), thorax (region 2), two upper members (regions 3 and 4), abdomen (region 5) and two lower members (regions 6 and 7). Many sensors can be inserted in these regions.

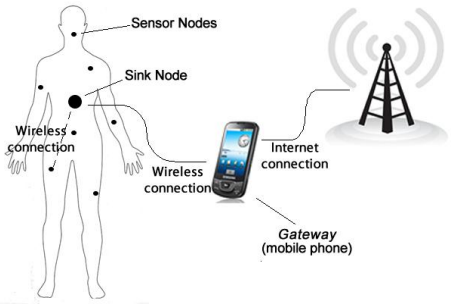


Figure 1. Environment of the WBSN studied.

Basically, there are two classes of MAC mechanisms: ordered and random access. In the former, a centralized node (or sink node) is used to organize the dispute for the output link. In the latter, each node transmits packets randomly to the physical medium and collisions may occur. For the WBSN, a centralized node is more convenient because collisions can be avoided, thus saving energy.

In this paper, an ordered MAC mechanism called polling is used. The idea is to use the similar time sharing system connecting the mainframe computer to the terminals in early times of computing. This scheme is chosen because it does not need frame synchronization as TDMA requires and still can keep almost a real-time treatment of messages for a reasonable amount of nodes. Fig. 2 shows the polling scheme proposed in this paper. The sink node, in normal operation, defines a cycle to attend the nodes. Based on this cycle, the sink node interrogates each sensor individually to check if there are packets to transmit. If there are, the sensor node receives a permission to start transmission while the others wait their turns. Thus, while a sensor node transmits packets, the others are performing their monitoring activities, awaiting their turns to transmit, and can store the generated packets in the buffer. After data transmission, a sensor node can await sleeping, thus saving energy.

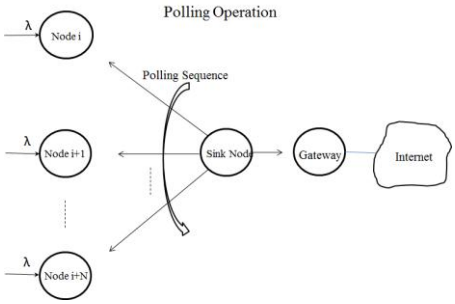


Figure 2. Polling Operation.

The communication protocol for the polling access scheme proposed in this paper can be simplified using the fact that the sensors are located near to the sink node. In normal operation the sink node broadcasts a packet carrying the node number to be investigated, i.e., it is sending an authorization to a sensor node to transmit the packets. This authorization packet has in the header enough bits for bit and frame synchronizations of a node. If a node has packets to transmit, it recognizes the node number and starts to transmit. After the transmission, the sensor node waits for acknowledgement in case of the need for retransmission. If a node doesn't have a packet to transmit, the transceiver can stay in an off state and only switches to on state if it has a packet to transmit. The sink node recognizes that a node is in an off state after the transmission of an authorization packet and waiting for a while. If the data packet from the polled node doesn't arrive, the sink node infers that the node doesn't have a packet to transmit and goes to another node in sequence to poll. Thus, in this protocol, the sink node does almost all the communication function, leaving the sensor node with only the packet transmission function.

The incoming packets to the sink node are queued at the output buffer to be relayed to the Internet or to a dedicated network where data are processed and stored as shown in Fig. 2.

3 Source Models and Description of the Scenarios for Analysis

For the analysis of the proposed polling-based MAC scheme, the sensor node modeling as the packet sources is

very important. Five different source models are proposed in this paper - all using the On/Off scheme for performance evaluation by simulation.

The first one is called Constant On/Off Source. This source generates one packet at each on interval, and the packet has a fixed size so that the intervals have an equal length obtained by dividing the packet size by the peak rate. In the off interval case, an average value is defined, and the off intervals are generated using negative exponential distribution.

The Constant On/Off Source operates in continuous monitoring mode, in which the sensors are always collecting the information and sending it to a sink node or a server. The sensors are processing all the time so they may have a shorter lifetime, but all the measurements are sent.

Four different sources based on the Constant On/Off Source that target energy saving have also developed. These sources have features that represent the behavior of sensors in the event-driven monitoring mode, where sensors send only relevant information to the event observer. The idea, for instance, is to define a sensor that is monitoring body temperature and send only those measurements that are above a certain value. The other criterion could be to transmit just the packets that are outside a certain range.

Based on these assumptions, a function that generates random values representing the measurement performed by the sensor was created. To select the packets that must be transmitted or discarded, the value generated is compared to a parameter supplied during the source configuration.

Table 1. FEATURES OF THE DEVELOPED SOURCES

Constant On/Off Source	Send all packets generated.
Threshold On/Off Source	Send only packets carrying information above a threshold.
Controlled Threshold On/Off Source	Send only packets containing information above a threshold or next packet when discarded packets reached a predefined number.
Out-range On/Off Source	Send only packets carrying information that are outside a certain range.
Controlled Out-range On/Off Source	Send packets satisfying Out-range On/Off Source criterion or next packet when discarded packets reached a predefined number.

To simulate sensors sending only measurements that are outside a certain range, two parameters should be informed to calculate that interval. These parameters are the average value and the percentage variation of this value. For example, in a sensor responsible for heart-beat monitoring, it is wished that only the measurements representing risks for a patient's life be sent. For instance, the normal heart-beat for a particular patient is 100 beats per minute, and it can vary between 80 and 120 per minute, then should be informed to the program the average value 100 and the percentage variation of 20%. Thus, in this case, only packets showing heart-beat measurements less than 80 or greater than 120 should be sent.

In the above-mentioned criteria, it is possible that there may be a hiatus where the nodes do not transmit any packet because no measurement satisfies the specified criteria for the transmission. Thus, to avoid a long silence of the sources, the discarded packets are counted and when this counting reaches a certain value the next packet is sent, regardless if the measurement satisfies the criteria established or not.

Tab. 1 shows the features of each developed source.

The following parameters shown in Tab. 2 are used for the generation of the five traffic sources.

Table 2. PARAMETERS FOR TRAFFIC GENERATION

Packet size	904 bits
Peak rate	39322 bits/sec
On Interval	22.989 msec
Off Interval	206.901 msec

The used packet size is the average packet sizes presented in the papers [18], [19] and [20]. The value of the peak rate was obtained in [21] and [22]. The On interval in Tab. 2 corresponds to the packet size divided by the peak rate. The Off interval is obtained considering that the sensors stay in the off state for 90% of the time [23]. From the sink node, the data are transmitted to a gateway, which can be a mobile device, as shown in Fig. 2. The sink node also contains the FIFO scheduler. The gateway forwards the information to the server. Different types of the sensor nodes, listed in Tab. 1, are placed in several parts of the body regions according to Tab. 3.

Table 3. SCENARIOS CONSIDERED

	Scenario 1	***Scenario 2***	***Scenario 3***
Region 1 (node 1)	*Constant Source*	*Constant Source*	*Constant Source*
Region 2 (node 2)	*Constant Source*	*Constant Source*	*Threshold Source*
Region 3 (node 3)	*Constant Source*	*Constant Source*	*Controlled Threshold Source*
Region 4 (node 4)	*Constant Source*	*Threshold Source*	*Controlled Threshold Source*
Region 5 (node 5)	*Constant Source*	*Controlled Threshold Source*	*Out-range Source*
Region 6 (node 6)	*Constant Source*	*Out-range Source*	*Controlled Out-range Source*
Region 7 (node 7)	*Constant Source*	*Controlled Out-range Source*	*Controlled Out-range Source*

Tab. 3 shows the three scenarios considered in the simulations, which aim to verify the performance of the

WBSN proposed. In Scenario 1, the Constant Source is used in many body parts to study the performance of the polling scheme to deal with heavier packet traffic, since the source constantly generates the packets. In the second scenario, mixed sources are configured having mostly Constant Sources. The third scenario is mixed with different types of proposed sources.

Each source was simulated with 10,000 packets, and the buffer sizes used are one, three, five or one thousand position(s). The choice of these values is justified by the fact that the nodes have little memory, and the last situation is equivalent to a fictitious buffer due to the limited energy of the nodes and can be considered infinite depending on the generation rate of packets. Since there are not frequent variations in measurements of physiological signals such as temperature and pressure, some losses can be accepted without impact to the system. A buffer of one-to-three positions was also used in [18].

At the sources using the controlled parameters, a signaling packet indicating the sensor node is active is sent every ten packets not transmitted.

The same output link of 250 Kbits/sec used in [21] is adopted. The data for statistical analysis are collected after discarding the first 2,000 packets to guarantee that statistical equilibrium is reached.

4 Analyses of Results

To analyze the behavior of the WBSN presented in Fig. 1, the three scenarios shown in Table III are studied. The aim is to analyze packet loss and queue time at sensor nodes using the polling medium access control. Moreover, the goal is also to verify the influence of the different sensor configurations in the WBSN performance.

Fig. 3 shows the simulation results for Scenario 1, according to the packet loss.

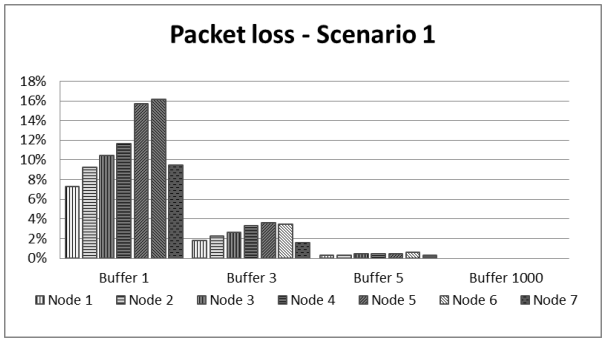


Figure 3. Packet loss in the first scenario.

As can be seen in Fig. 3, the packet losses in all nodes are very high - ranging from 7 % in node 1 to 16% in node 6 - for the buffer with one position because all the generated packets are transmitted. However, because of the increase in the buffer size for three positions, the packet losses become reasonable, reaching at most 4%. It shows that a buffer size with five positions has losses less than 1% and for one thousand positions no packet loss is observed. The packet loss using Constant Source types is not critical because the sources are sending packets constantly and any lost packet can be interpreted at the final server using an interpolation technique.

In Fig. 4 the average queuing time of packets is shown.

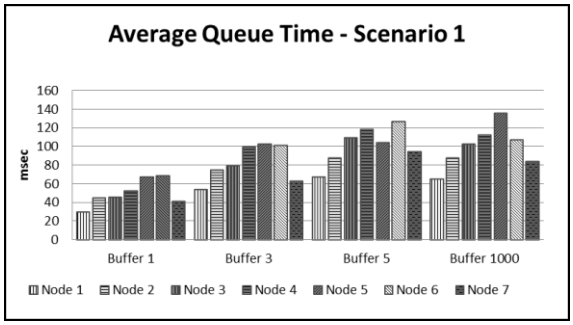


Figure 4. Waiting time of packets in first scenario

The minimum waiting time is about 30 msec in node 1 for one buffer size and the maximum is about 140 msec in node 5 for the one thousand buffer size as can be seen in Fig. 4. The waiting times are not long and are appropriate for quasi real-time processing.

In the simulations of all scenarios it has not considered the walk-time necessary, after the transmission of a packet, to move the inspection from one node to another node. This time in the case of WBSN is small and could be considered a constant value. The propagation time from a node to a sink node is also not considered.

Fig. 5 shows the simulation results for Scenario 2, according to the packet loss.

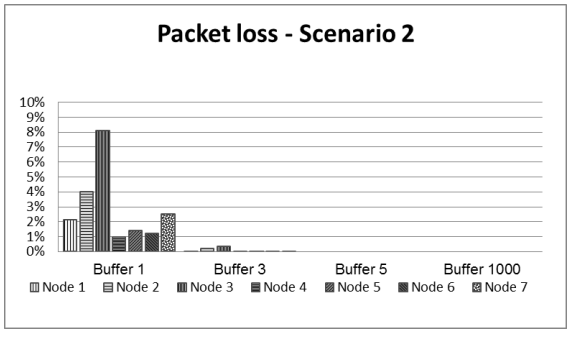


Figure 5. Packet loss in Scenario 2.

In the second scenario - where the traffic is generated in three sensors using the Constant Sources, in which all the generated packets are transmitted - it can be seen the nodes 1,

2 and 3 have more significant packet loss, reaching 8.14% at node 3. This fact is justified considering that all three sources have been configured with the parameters shown in Table II and, thus, the arrival times of packets will be similar. Since the service is cyclical, beginning with the first node, followed by the nodes 2, 3 and so on, node 3 has to wait for two nodes to be attended to in order for its turn to come, thus having much more loss, which is confirmed in the Fig. 5, with one buffer position. In other nodes, with one buffer position, the loss reached is at most 2.50% at node 7, which uses the Controlled Out-range Source for the generation of packets.

Comparing these values with the simulations using larger buffer sizes, the discarding has decreased considerably, and when the buffer size is set at five or one thousand positions, the packet loss has not happened in any of nodes.

Fig. 6 shows the percentage of packets not transmitted due to the restriction imposed at the sources.

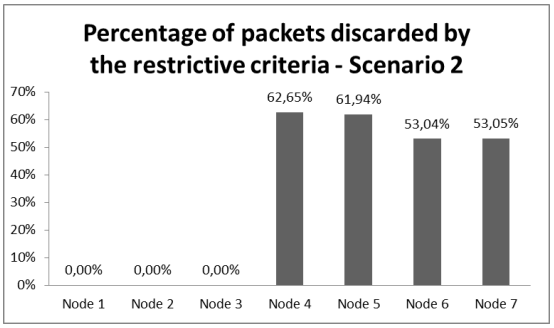


Figure 6. Percentage of packets discarded in Scenario 2 by the restrictive criteria.

Due to the use of Constant Sources in nodes 1, 2 and 3 there are no rejected packets in these nodes as can be seen in Fig. 6. In other nodes using sources with restrictive criteria, the rejection percentage exceeds 50% in all nodes, being the largest one in node 4 using the Threshold Source.

It can be pointed out that the figures presented in Fig. 6 are not affected by buffer sizes because the packet rejections are done using restrictive algorithm before the queuing in the node buffer.

It can also be concluded that the implementation of the restrictive algorithm is very important to save energy, considering the high energy consumption in a packet transmission.

Fig. 7 presents the simulation results of the queue time in the buffer for Scenario 2. The same influence of polling attendance of Fig. 4 is also observed - in this case with node 3 having more time to wait to transmit its packets. In other sources not using Constant Sources, the waiting times are smaller due to controlled packet generations.

When the buffer size is increased the waiting times are longer and have similar behavior, showing that the big buffer sizes are not necessary. A buffer with 3 positions is enough for low packet loss and capable to transmit almost all packets generated.

Fig. 8 shows the packet loss for Scenario 3 in function of the buffer sizes.

In Scenario 3, as the number of sensors with Constant Source is restricted to Sensor 1, the number of packets transmitted is reduced. Consequently, the packet loss is decreased.

Fig. 9 shows the average percentage of packets not sent due to the restrictive criteria applied at the sources for Scenario 3.

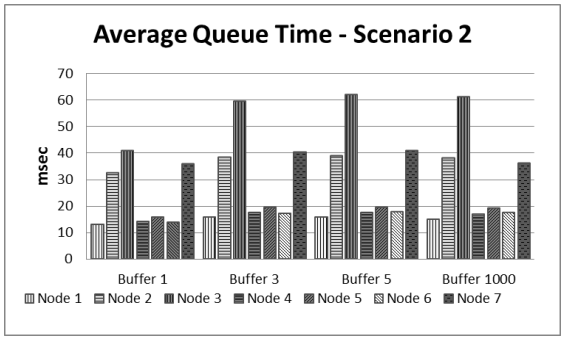


Figure 7. Waiting time of packets in second scenario.

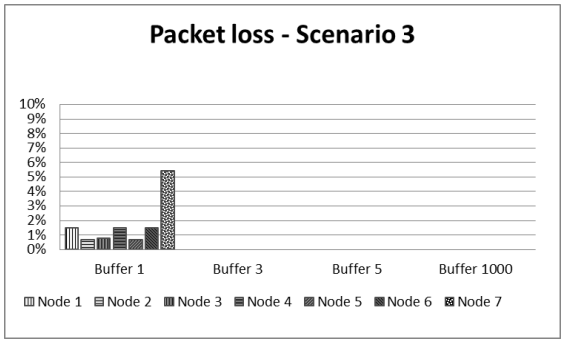


Figure 8. Packet loss in the second scenario.

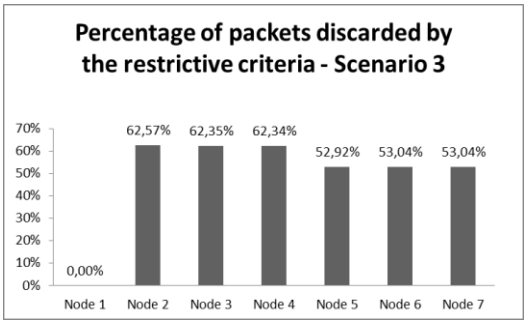


Figure 9. Percentage of packets discarded in Scenario 3 by the restrictive criteria.

Comparing Scenarios 2 and 3 in relation to the discarding of packets, nodes 1, 2 and 3 using Constant Sources have higher losses in Scenario 2, as is expected. Furthermore, since the cyclical service starts at node 1, this node becomes the most favored of the three nodes having lowest discarding. In the third scenario, since a diversity of sources is used - mainly those using the controlled parameters - the node that produces the highest discarding is node 7. This node is the last to be attended to in the cyclic polling so it has a longer waiting time to transmit and may even not transmit any packet in a given cycle, discarding all packets because they are inside of the predefined range or below of a threshold.

It can be seen in Fig. 9 that there is no rejected packet in node 1 due to the use of Constant Source as also observed in Fig. 6. However, the rejection percentages are high in other nodes using sources having restrictions to send packets as occurred in Scenario 2. The Threshold and Controlled Threshold Sources have higher rejection percentages ranging from 62.57% to 62.34% as can be seen in nodes 2, 3 and 4. Fig. 9 shows the Out-range and Controlled Out-range Sources also have significant losses reaching about 53% for all three sources as can be observed in nodes 5, 6 and 7.

Energy saving can also be observed in Scenario 3 by avoiding the packet transmission, although some energy is spent for the processing of restrictive algorithm.

In Fig. 10 the queue times for the third scenario are illustrated.

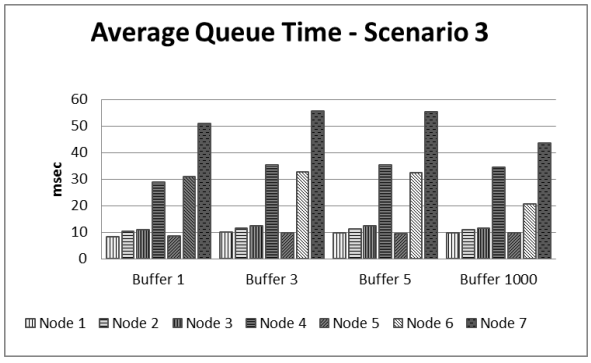


Figure 10. Waiting time of packets in third scenario.

By reducing the number of transmitted packets, the queue times are reduced in Scenario 3 compared to Scenarios 1 and 2, as expected. The reduction is significant in most of the nodes as can be verified in Fig. 10. In Scenarios 1 and 2, the average queuing times are 50.43 and 23.87 msec, respectively, while in Scenario 3 it is 21.21 msec, considering a buffer with one position. Moreover, the worst cases for Scenarios 1 and 2 occurred in node 5 (or 6) and in node 3, with queue times of 68 and 40.89 msec, respectively, while in Scenario 3, the longest queue time is 50.93 msec in node 7 using Controlled Out-range Source.

The use of diversity of the traffic sources is beneficial to the queue time and packet loss in most of the nodes. Moreover, the predominance of event-driven types of sources is also important. For the WBSN applications, a maximum of three buffer positions may attend the expectations of QoS for the packet discarding and queue time.

On the other hand, it is observed that for five and one thousand positions of buffer sizes, there is a similarity between Scenario 2 and 3 in the queue times because there are no significant variations in the values presented.

Since the polling mechanism has the function of traffic admission controller and by using a link of 250 Kbits/sec at the sink node there are no packets waiting at FIFO buffer for the three scenarios studied. The service time or system time is 3.5 msec in all analyzed scenarios.

5 Conclusions

In this paper the polling access scheme for Wireless Body Sensor Network (WBSN) was studied. The main technical advantage of the polling access mechanism is the non-necessity of frame synchronization and it has centralized control of sensors convenient for WBSN. The objective of the paper is to study the suitability of the polling mechanism to gather the almost real-time data from sensors placed in a human body. Thus, the main parameters used for study are packet loss and waiting time at the buffer of a sensor node. Since the sensor node for WBSN needs to save energy, the minimum buffer size needed to keep the packets before their transmissions was also examined. To accomplish the above objectives, a simulation platform was developed in C++ Builder where the polling mechanism, the buffers and sensor sources were all implemented. Since there is little sensor node models for WBSN applications in the literature, five sensor sources were proposed, all based on the On/Off model. One of the sensor sources developed was Constant On/Off Source which forwards the information continuously in the on interval and stays silent in off interval. The other sensor sources are event-driven, in which the information is transmitted only if it satisfies a certain condition. In addition, to facilitate the status management of the sensors, two other event-driven sources were developed in which a message is sent after a certain number of packets are not transmitted, regardless if the requirement for the transmission of data has been met or not.

The proposed human body environment for study consisted of seven sensors placed in different parts of body, forming a star topology, with the sink node at the network core. In this environment, three scenarios were proposed. The first scenario used a configuration with Constant Source in all nodes, while in the second scenario three Constant Sources are mixed with other types of sources. In the last Scenario five sources are mixed in different parts of body.

The simulation results for Scenario 1 in relation to the packet loss, considering only one position buffer at sensor node showed high losses ranging from 7% to 16%. The losses became reasonable for three or greater buffer sizes.

In Scenario 2 the packet losses for three Constant Sources are also high, reaching more than 8% in one of the nodes and about 2% in the most favored node of the polling scheme. However, these losses may not be critical for Constant Sources, because they are constantly sending the packets so that some lost information may be interpreted at the final server using some interpolation technique. But these sources must be used with care because of the high intensity of packet generation. For the other four sensor nodes in Scenario 2 the losses are smaller ranging from 1% to about 2%. But in these cases the losses may be critical because the sensors are already doing some kind of data selection. To overcome this situation, the results showed that for a buffer of three positions the losses are almost insignificant.

For Scenario 3, in a mixed situation of sources, the losses are more controlled reaching in the worst case about 5% for one position buffer and no loss in the case of buffer with three or more positions.

The simulation results for average queuing times at the sensors showed low waiting times for Scenarios 1, 2 and 3, ranging from 8 to 140 msec considering all buffer sizes. It was not considered the walk and propagation times in the simulations because they are small and constant values. It can be concluded that the polling access scheme is adequate for quasi real-time applications.

The simulation results also showed that at the sink node using the FIFO scheduler, no loss had occurred or no packet was waiting in the buffer because the polling access mechanism works as the admission controller and only one packet is processed each time.

The polling access scheme showed a very promising technique for WBSN applications but other scenarios will be investigated and compared to other access schemes in future works.

Acknowledgement

The authors would like to thank Allison Linn Davidson and Renato Motoyama for assistance in English revision.

6 References

[1] B. Latré, B. Braem I. Moerman, C. Blondia and P. Demeester, "A Survey on Wireless Body Area Networks", in Wireless Networks, Volume 17 Issue 1, January, 2011, Kluwer Academic Publishers Hingham, MA, USA

[2] B. Otal L. Alonso and C. Verikoukis, "Towards Energy Saving Wireless Body Sensor Networks in Health Care Systems" Proceedings of IEEE International Conference on Communications (ICC 2010), Second International Workshop on Medical Applications Networking (MAN 2010), Capetown, África dos Sul, 2010.

[3] Z. Yan and B. Liu, "A context aware MAC protocol for medical Wireless Body Area Network" Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International

[4] B. Liu, Z. Yan and C. W. Chen, "CA-MAC: A Hybrid context-aware MAC protocol for wireless body area networks" 13th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), 2011.

[5] Z. Xiuming, H. Song, H. Pei-Chi, A.K. Mok and C. Deji, "MBStar: A Real-time Communication Protocol for Wireless Body Area Networks" 23rd Euromicro Conference on Real-Time Systems (ECRTS), 2011.

[6] D. Layerle and A. Kwasinski, "A power efficient pulsed MAC protocol for Body Area Networks" IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2011.

[7] Y. Tselishchev, "Designing a Medium Access Control protocol for Body Area Networks" IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011.

[8] L.M. Borges, F. J. Velez, A.S. Lebres, "Performance Evaluation of the Schedule Channel Polling MAC Protocol applied to Health Monitoring in the Context of IEEE 802.15.4" 11th European Wireless Conference - Sustainable Wireless Technologies (European Wireless), 2011.

[9] S. Kutty and J.A. Laxminarayan, "Towards energy efficient protocols for wireless body area networks" International Conference on Industrial and Information Systems (ICIIS), 2010.

[10] L. Huaming and T. Jindong, "Heartbeat-Driven Medium-Access Control for Body Sensor Networks" IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 1, January 2010.

[11] K. A. Ali, J.H Sarker and H.T Mouftah, "Urgency-Based MAC Protocol for Wireless Sensor Body Area Networks" IEEE International Conference on Communications Workshops (ICC), 2010

[12] A. A. Khaled, H. S. Jahangir and T. H. Mouftah, "QoS-based MAC protocol for medical wireless body area sensor networks IEEE Symposium on Computers and Communications (ISCC), 2010

[13] X. Zhang,H. Jiang, X. Chen, L. Zhang, Z. Wang, "An Energy Efficient Implementation of On-Demand MAC Protocol in Medical Wireless Body Sensor Networks" IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009.

[14] S. Marinkovic, C. Spagnol and E. Popovici, "Energy-Efficient TDMA-Based MAC Protocol for Wireless Body Area Networks" Third International Conference on Sensor Technologies and Applications, 2009. SENSORCOMM '09.

[15] G. Fang, E. Dutkiewicz, "BodyMAC: Energy efficient TDMA-based MAC protocol for Wireless Body Area Networks" 9th International Symposium on Communications and Information Technology, 2009.

[16] S. S. Oliveira and S. Motoyama, "Applications Oriented Medium Access Control Protocols for Wireless Sensor Networks. IEEE Latin America Transactions, v. 7, Issue 5, 2009.

[17] B. A. Furouzan, "Data Communications and Networking". 4th Ed. The McGraw-Hill Companies. Nova York. 2007.

[18] H. Li and J. Tan, "An Ultra-low-Source Mediu Access Control Protocol for Body Sensor Networks". In: Proceedings of 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Shangai, 2005.

[19] R. Gravina, A. Guerrieri and A. Fortino, "Development of Body Sensor Networks Aplications using SPINE". In: IEEE International Conference on Systems, Man and Cybernetics. Singapura, 2008.

[20] B. Otal, L. Alonso and Ch. Verikoukis, "Novel QoS Scheduling and Energy-saving MAC protocol for Body Sensor Networks Optimization". In Proceedings of the Third International Conference on Body Area Networks (BodyNets08), Tempe, Arizona, 2008.

[21] S. Ullah, P. Khan, N. Ullah, S. Saleem, H. Higgins, K. S. Kwak, "A Review of Wireless Body Area Networks for Medical Applications". In: Int'l J. of Communications, Network and System Sciences. N.8, 2009.

[22] K. Chakrabarty and S. S. Iyengar, "Scalable infrastructure for distributed sensor networks". Springer-Verlag, 2005.

[23] V. Potdar, A. Sharif and E. Chang, "Wireless Sensor Networks: A Survey". International Conference on Advanced Information Networking and Applications Workshops. Bradford, 26-29 May 2009,pp.636-641.

Secure Network Coding in Unattended Wireless Sensor Networks

Faezeh Sadat Babamir
Department of
Computer Science
Shahid Beheshti University
Tehran, Iran
f.babamir@mail.sbu.ac.ir

Fattaneh Bayat Babolghani
Department of
Computer Science
Shahid Beheshti University
Tehran, Iran
f.bayat@mail.sbu.ac.ir

Kayhan Moharreri
Department of Computer
Science and Engineering
Ohio State University
Columbus, USA
moharrer@cse.osu.edu

Abstract—**Security issues in wireless sensor networks have been focused by extensive researches in recent years. Security concerns are particularly critical in disconnected or Unattended Wireless Sensors Networks (UWSNs). In this setting, the sink periodically collects sensed data and therefore the network will be left unattended most of the time. An adversary can take advantage of this behavior to modify or erase data. Thus, cryptographic techniques must be employed to ensure privacy and integrity of the information. In this paper, we consider network coding along with data sharing to provide confidentiality and integrity simultaneously. Moreover, every shared message will be signed and encrypted in efficient time to make the communication secure.**

Keywords-component; unattended wireless sensor networks, authenticity, confidentiality, integrity, time compelxity, modifies generalized laguerre functions, collocation method.

I. Introduction

In the past decades, wireless Sensor Networks (WSNs) attracted many researchers. A lot of them considered as important issues such as: routing, security, power awareness and data abstraction, but security is prior common assumption in most of these works. On the other hand, WSNs should collect small size and especially secure data in real-time manner. This is a crucial property as sensor nodes are small, low power with limited storage capacity. Therefore, classical algorithms may not be applicable, i.e. considering resource constrained sensors. These algorithms cannot guarantee the security of data. The aforementioned problem is even more critical in the new generation of WSNs referred to as Unattended or disconnected Wireless Sensor Networks (UWSNs) as sink periodically leaves and returns to the network.

The disconnected networks are established in critical or military environments. Hence, sink or collector is unable to gather data in real-time manner. Moreover, the network will be left unattended and will be periodically visited by the itinerant sink. This property provides some threats such as discovering and compromising sensor nodes by the adversary without detection of communication. The adversary, also, invisibly can perform to be intractable and unpredictable. A UWSN adversary may have different goals; some are curious and aim just to disclose data, while others aim to search data to replace them with forged message. The third type of adversary, known as the polluter, aims to inject invalid data to corrupt network called DoS attack or mislead the sink. In such setting, the main challenge is assurance about data integrity for long time.

In this research, we propose a scheme that encrypts shares and signs the generated data to provide confidentiality and integrity. We also leverage an efficient numerical solution for encryption; every sensor with unique identification encrypts shares, in which encryption is one-way without the knowledge of initial boundary conditions. Then a linear signing algorithm is applied to provide authentication and prevention of DoS attack. The signed generated data will be broadcasted to the neighbour sensors. Every neighbour uses network-encoding for received shares and homomorphic signs to remove previous signature and generate unique signature. This process decreases the size of total received shares.

Organization: Section *II* reviews the related work of UWSNs. Section *III* sketches our proposed algorithm including applied network coding, homomorphic and numerical encryption process. In section *IV* we have demonstrated our scheme is efficient. We have ended this paper with conclusion section.

II. Related Work

In UWSN setting, the adversary may have different goals. Reactive adversary is the adversary who starts compromising sensors after he identifies the target. To be more precise, such an adversary is inactive until it gets a signal that certain data must be erased, and then it wakes up and starts compromising up to l sensors per round. This is unlike the proactive adversary who can compromise sensors before identifying the target i.e. he essentially starts compromising sensors at round 1, *before* receiving any information about the target sensor and the target data collection round. He would choose and compromise different sensors in a geographic area even

before such signal is received. This powerful adversary who usually referred to as mobile adversary can even roam around the network and change from one set of compromised nodes to another, making such attacks more difficult to detect and prevent.

To defend against reactive adversary, many papers have been proposed encryption based schemes. Encryption can be employed to hide the collected information as well as the identity of the sensors that collect it. If the key of compromised node is not available, the reactive adversary is unable to distinguish the specific piece of collected data. However, proactive adversary can restore the keys of the other earlier compromised nodes to memorize encrypted data. These keys help adversary to encrypt some forged data and place them with the target data. Therefore, encryption is not enough to defend proactive adversary.

The faults of these solutions are discussed by Di Pietro et al. in [1]. They proposed super-encryption and re-encryption techniques to defend against mobile adversary. But they did not take into account the cost of time, memory and energy consumption overhead. In addition, the proposed solution have limitations because of dependency on symmetric (shared key) encryption. Symmetric setting prevents sensors to use data aggregation techniques. Another solution is asymmetric based scheme. Although it is more resource consuming than symmetric solution, the sensors can decrypt the ciphertext and perform data aggregation, eliminating redundancy to minimize memory and communication overhead. Therefore, in this scheme, extra efficiency through data aggregation will be obtained by more energy and memory consuming. In general, data aggregation is more considered than energy and memory consumption, since 1 bit transmitted may require the power equivalent to execute 800-1000 instructions [2].

D. Ma et al. in [3] proposed 2 approaches: First, FssAgg-BLS (a kind of signature) as ideal cryptographic tools for achieving data integrity and authentication for UWSNs in presence of active adversary. However, this public key encryption setting imposes extreme computational overheads on the network entities, which are intolerable for UWSNs applications. Second approach, FssAgg-Mac is based on symmetric key encryption, hash chains and Message Authentication Codes (MACs) which requires full symmetric key distribution and does not allow to be public verifiable. This makes it impractical for large distributed UWSNs. Later D. Ma et al. developed FssAgg-AR and FssAgg-BM in [4, 5] that more computational and storage efficient than FssAgg-BLS. However, all these schemes are still not efficient enough for UWSNs and are effective only against the reactive adversary that is relatively weak and easily to overcome.

III. Proposed Scheme

Ren et al. [6] prove that in order to achieve perfect secrecy, data sharing between neighbours is a suitable way. Therefore, in our scheme, sensor node collects data D and breaks it to equal shares $d_1, d_2..., d_n$. Using following process, the sensor sends signed encrypted d_is to the neighbours:

A. Share Generation, encoding, signing and broadcasting process

After sensor v_j collects data D, it proceeds following steps to achieve data integrity, confidentiality and also authenticity.

1) Shares D into equal $d_1,d_2,...,d_n$.
2) Using our numerical encrypting method(refer to section E), the sensor encrypts every d_i to Y_i .
3) Every Y_i will be signed by sensor v_j(known as δ_i).
4) Lastly, sensor v_j broadcasts every δ_i to the each neighbour.

Below we describe mathematically this algorithm. Set_j is the set of all neighbours of sensor j.

Alg. 1: Collecting and sending data(D, set_j)

{

Shares D into $d_1,d_2,..,d_n$.

Encodes d_i *by* $Y_i=f(d_i)$.

Sign Y_i *by* $\delta_i=Sig(Y_i)$

Obtain $pk_i=\{Y_i||\delta_i||t||TS||CNT\}$

Broadcast every pk_i *to every neighbour sensor belonging to* Set_j.

}

(t,TS,CNT will be defined in section B)

B. β-bounded moving(adapted from [7])

Every signedY_i should disperse enough to defend against mobile adversary.β is number of traversed hops up to now while CNT is number of hops which should be traversed. To determine β value, *DLE* variable is defined to determine the location entropy of data d_i. This concept makes trade-off between number of hops and energy communication. Moreover, more β consumes much energy communication but makes higher security against mobile adversary. *DLE* helps us to determine suitable value.

Finally, $pk_i=\{Y_i||\delta_i||t||TS||CNT\}$ is output of sensor v_j to another neighbour. More precisely, v_j in which Y_i, δ_i, t, CNT are encrypting vector of data share, signature of Y_i, sequence order of d_i and $CNT=\beta$ respectively. Also, TS is the time stamp of producing time. We define a tuple$UID=\{TS||t\}$, that can uniquely identify a share.

C. Network Coding

In this paper, we use two kinds of sensors that were called source sensor and forwarder sensor; source sensor should collect data and broadcast, while forwarder sensor receives the data from other sensors and then transforms

theses data packets into one packet and then forwards the resulting packet to the next hop. Furthermore, since communications consume more energy than computations, forwarding nodes encode received packets into one packet by using network coding solution. Clearly, network coding technique increases overall computation energy instead it significantly decreases communication consumption. Finally, the forwarding sensor signs the packet through homomorphic signature (refer to section *D*).

1) Basic setting

In this setting, we show the network with *G=(V,E)*. Source nodes and forwarding nodes are$S = \{s_1, s_2, \dots, s_P\} \subseteq V$ and $f = \{f_1, f_2, \dots, f_L\} \subseteq V$ respectively. The inputs of forwarding nodes are $Y_i, i \epsilon [1, n]$of pk_i and output packets are Z_j, $j \epsilon [1, p]$. Source nodes $S_l, l \epsilon [1, p]$propagate packets pk_i to the forwarding nodes. Each forwarding sensor, after receiving Y_i of pk_i from *n* incoming channels, computes following linear combination Z_j to transmit it to the *j*-the channel. The linear combination formula is:

$$Z_j = \sum_{i=1}^{n} (\alpha_i)(Y_i) \qquad (1)$$

In formula (1),$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ is encoding vector. The node randomly generatesα or α is pre-deployed, (depend on static network topology). It is proven that random coefficient optimises network performance with high probability because of independency of network topology.

2) Random linear network coding algorithm

In proposed scheme, every forwarding node receives some $Y_i i \epsilon [1, n]$and encodes them via network coding as explained in equation (1). Finally, it sends one packet containing an encoded vector of size *n*. For simplicity, we let pre-deployed encoding vector (α). Consider, Alg. 1, for encoding *n* packets. The final output is encoded vector of *Z*.

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1p} \\ \vdots & \dots & \cdots & \vdots \\ \alpha_{i1} & \alpha_{i2} & \cdots & \alpha_{ip} \\ \vdots & \cdots & \cdots & \vdots \\ \alpha_{j1} & \alpha_{j2} & \dots & \alpha_{jp} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_p \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p-1} \\ Z_p \end{pmatrix}$$

Figure1. Encoder Matrix

Our scheme is able to reconstruct thoroughly the primary data from all received packets. Moreover, by using aforementioned equation Y_is will be recovered in polynomial time (adapted [7]). In section *E*, we will propose a new numerical method to easily encrypt shares with *time efficiency*. This innovative solution is considerable either for sink or forwarding nodes, i.e. our scheme either on the node side or origin sink side runs efficiently.

D. Applied linear Homomorphic signature over $\mathcal{F}_2$

In this paper, we utilize Boneh et al. scheme which is inspired by Gentry, Peikert and Vaikuntanathan [7]defined linearly over binary field [8]. This signature is a short vector $\delta \epsilon Z^m$in $\Lambda_{2q}^{q.v}(\Delta)$, i.e. δ is in both $\Lambda_q^{\perp}(\Delta)$ and $\Lambda_2^{v}(\Delta)$ simoltaneusly. Mod 2 relates the signature to the message while mod q is designed to prove unforgeability of the scheme. This Δ is different for signing every packet.

The source sensor signs every Y_i using its identity based private key and then sends (Y_i, δ_i) to the forwarding neighbour node. Forwarding node receives Y_is along with their signatures.

Firstly, it checks the validity of signature. If it is not valid, forwarding sensor removes it as bogus data. Receiving enough valid data, forwarding sensor encodes and generates a homomorphic signature from share signatures without knowing the original messages (d_i) or the private key of source nodes. The detail of scheme is as follow:

1) Parameter setup phase

Following, we define parameters that used in [5] to describe applied signature. It is an *m*-dimensional lattice whose points are defined on $\mathbb{Z}^m$. Also, is a full-rank discrete subgroup of $\mathbb{R}^m$ and consist of vectors are generated by orthogonal to a certain "parity check matrix" $\Delta \epsilon \mathbb{Z}_q^{m \times n}$ modular integer *q*. The utilized lattices are defined:

$$\Lambda_q^{\perp}(\Delta) = \{e \epsilon \mathbb{Z}^m : \Delta . e = 0 \, mod q\} \qquad (2)$$

$$\Lambda_q^{u}(\Delta) = \{e \epsilon \mathbb{Z}^m : \Delta . e = u mod q\} \qquad (3)$$

$$\Lambda_q(\Delta) = \{e \epsilon \mathbb{Z}^m : \exists s \epsilon \mathbb{Z}_q^n with \Delta^t . s = e mod q\}$$

In formula (3), $\Lambda_q^{u}(\Delta)$is a coset of lattice $\Lambda_q^{\perp}(\Delta)$of formula (2) such that $\Lambda_q^{u}(\Delta) = \Lambda_q^{\perp}(\Delta) + t$ in which *t* holds in $\Delta . t = u mod q$.

2) Signature scheme

Firstly, we describe following functions that used in the Boneh et al. scheme:

- TrapGen(*q, n*): this algorithm receives an integer *q* and *n* holds in $m = [6n \, lgq]$. Also this algorithm outputs ($\Delta \epsilon \mathbb{Z}_q^{n \times m}, S \epsilon \mathbb{Z}^{m \times m}$), where Δ is statistically close to a uniform matrix in $\mathbb{Z}_q^{n \times m}$ and *S* is a basis for $\Lambda_q^{\perp}$.

- ExtBasis(*S, B*): let *m'* be an arbitrary dimension. This algorithm gets ($S, B = \Delta \| \Delta'$) where

$\Delta' \epsilon \mathbb{Z}_q^{n \times m'}$ and $S \epsilon \mathbb{Z}^{m \times m}$ be an arbitrary basis of $\Lambda^{\perp}(\Delta)$ for a rank n matrix $\Delta \epsilon \mathbb{Z}_q^{n \times m}$ that outputs a basis *T* of $\Lambda^{\perp}(B) \subset \mathbb{Z}^{(m+m') \times (m+m')}$.

- SamplePre(*Δ, T, u, δ*): this algorithm inputs matrix $\Delta \epsilon \mathbb{Z}_q^{n \times m}$, a basis *T* of $\Lambda_q^{\perp}(\Delta)$, a parameter δ and a vector $u \epsilon \mathbb{Z}^n$. Then outputs a sample which is statistically close to the distribution of $\mathfrak{D}_{\Lambda_q^u, \delta}$.

a) Signing algorithm

1- Choose a $id \xleftarrow{R} \{0,1\}^n$ randomly. If *id* has already been queried to the hash function *H,* then abort. (The simulation has failed).

*Setup(1^n; k):*On input of a security parameter *n* and a maximum data set size *k*, do the following:

1. Choose two primes *p, q = poly(n)*with $q \geq (nkp)^2$. Define *l :=[n/6 log q]*.

2. Set $\Lambda_1 := pZ^n$.

3. Use TrapGen(*q;l; n*) to generate a matrix $\Delta \epsilon F_q^{l \times n}$along with a short basis T_q of $\Lambda_q^{\perp}(\Delta)$. Define$\Lambda_2 = \Lambda_q^{\perp}(\Delta)$and $T:=p.T_q$. Note that T is a basis of$\Lambda_1 \cap \Lambda_2 = p\Lambda_2$.

4. Set $v := p.\sqrt{n.logq}.logn$.

5. Let $H: \{0,1\}^* \rightarrow F_q^l$be a hash function (modeled as a random oracle).

6. Output the public key $pkey = (\Lambda_1, \Lambda_2, v, k, H)$and the secret key *skey= T*.

The public key *pkey* defines the following system parameters:

- The message space is $F_p^n$and signatures are short vectors in Z^n.
- The set of admissible functions *F* is all F_p-linear functions on *k*-tuples of messages inF_p^n.
- For a function $f \epsilon F$defined by $f(m_1, \ldots, m_k) = \sum_{i=1}^{k} c_i m_i$, we encode *f* by interpreting the c_i as integers in (-p/2, p/2].

*Sign($skey, \tau, m, i$):*On input of a secret key *skey*, a tag$\tau \epsilon \{0,1\}^n$, a packet$pkey \epsilon F_p^n$and an index *i*, do:

1. Compute $\alpha_i = H(\tau || i) \in F_q^l$.

2. Compute $t \in Z^n$such that *t* mod *p* = *pkey* and$\Delta . t \bmod q = \alpha_i$.

3. Output $\sigma \leftarrow \mathrm{SamplePre}(\Lambda_1 \Lambda_2, \mathrm{T}, \mathrm{t}, \mathrm{v}) \in (\Lambda_1 \Lambda_2) + \mathrm{t}$.

*Verify(pk, τ, pk, σ, f):*On input of a public key *pkey*, a tag$\tau \in \{0,1\}^*$,a message $m \in F_p^n$, a signature$\sigma \in Z^n$and a function$f \in F$, do:

1. If all of the following conditions hold, output 1 (accept); otherwise output 0 (reject):

(a) $||\sigma|| \leq k.\frac{p}{2}.v\sqrt{n}$

(b) $\sigma \bmod p = pkey$.

Evaluate $(pkey, \tau, f, \vec{\sigma})$. On input a public key *pkey*, a tag $\tau \in \{0,1\}^*$, a function $f \in F$ encoded as$< f >= (c_1, \ldots, c_k) \in Z^k$and a tuple of signatures$(\sigma_1, \ldots, \sigma_k) \in Z^k$, output $\sigma = \sum_{i=1}^{k} c_i \sigma_i$.

After sink receives all signed encrypted shares, it verifies the homographic signature and decrypts them to reconstruct *D*.

In this signing algorithm, we apply linear signing and efficient encoding algorithms. More exactly, we firstly encrypt d_i into Y_i included in pk_i= *{Y_i||δ_i||t||TS||CNT }* by proposed numerical method. This encrypting solution prevents adversary to read data because our numerical encrypting solution Eq. (1) is a differential equation and insolvable without knowing boundary conditions. Boundary conditions are initial values of the Eq. (2) which is available for receivers. We discuss about our numerical technique in following section.

E. Numerical encryption process

In this section, we have an improvement in our last scheme that was mentioned in [38]. In fact, we used an Ordinary Differential Equation (which we denote as ODE) that has an exact solution for encrypting the data. This ODE can only be solved with its boundry conditions and we use that as the decryption part on the receivers' side; meanwhile senders have the exact solution that would be mentioned in Eq. (3) for encryption. The ODE would be well known, and the boundry conditions would be the secret key of recievers for decryption.

The receives decrypt and solve this equation by modified generalized Laguerre functions which are orthogonal functions. Collocation method is used in this approuch that reduces the solution of this problem to the solution of an algebraic equation. Moreover, in the graph of the $\| Res \|^2$, we show that the present solution is more accurate and faster in convergence for this problem.

Firstly, we introduce the well known equation from the problem of flow and diffusion of chemically reactive species over a nonlinearly stretching sheet. This equation is well known by the adversary and recievers. This non-linear ODE is [9-24, 35]:

$$\frac{d^3 f}{dd_i^3} + f\frac{d^2 f}{dd_i^2} + (\frac{df}{dd_i})^2 - id\frac{df}{dd_i} = 0 \qquad (1)$$

Subject to boundary conditions,

$$f(0) = 0, \qquad f'(0) = 1,$$
$$f'(d_i) = 0 \ as \ d_i \rightarrow \infty, \qquad (2)$$

Thses boundry conditions are kept secret on receivers.

In the Eq. (1), id is identification of every sender sensor, and in *Alg. 1*, data shares into $d_1, d_2, .., d_n$ that every $d_i$is entry of Eq. (1).

Here we note that the Eq.(1) subject to boundary conditions Eq.(2) has an exact solution in [11,17] that is used for enceypting the data:

$$f(d_i) = \frac{1}{\sqrt{1+id}}\left(1 - e^{-\sqrt{1+(id)d_i}}\right) \quad (3)$$

- Encryption:

Assuming that every d_i is l bits (also known as share length), then senders use a pseudorandom generator with the expantion factor of l to mask the predictibility of d_i. For every d_i we will generate a random number of the length s, where $s << l$ for the preformance efficnecy. Technically, that is defined $G:\{0,1\}^s \rightarrow \{0,1\}^l$; also, $r \in \{0,1\}^s$ is chosen randomly.

For every d_i, the sensors computes ciphertext as follows:

$$Y_i = f(G(r) \oplus d_i) \parallel r \quad (4)$$

Where f is only known to sender sensors, and it will be unpredictable for the adversary to guess d_is from the ciphertexts as they are randomly being masked by our pseudorandom generator. As a result of this encryption we can also apply Cipher Block Chaining (CBC) mode to this scheme where $f(G(r) \oplus d_i)$ will feed the next mask for d_{i+1} instead of generating a new $G(r)$. (assuming that the range of the function f can be adapted to size s). However, in practice, we found that CBC is biased towards the output of the function as f being used here is clearly not a psodurandom function.

- Decryption:

We introduce the method encrypting and solving this equation as follows. Also, different techniques have been used to obtain analytical and numerical solutions for this problem in [14, 22-24].

1) Modified Generalized Laguerre functions

The generalized Laguerre in polynomial manner are defined with the following recurrence formula[25-28]:

$$L_0^a(x) = 1, \quad L_1^a(x) = 1 + \alpha - x,$$
$$nL_n^a(x) = (2n - 1 + \alpha - x)L_{n-1}^a(x) - (n + \alpha - 1)L_{n-2}^a(x),$$

These are orthogonal polynomials for the weight function $w_\alpha(x) = x^\alpha e^{-x}$. We define Modified Generalized Laguerre Functions (which we denote MGLF) ϕ_j as follows[25]:

$$\phi_j(x) = \exp\left(\frac{-x}{2L}\right) L_j^1\left(\frac{x}{L}\right), \quad L > 0. \quad (5)$$

This system is an orthogonal basis [36, 37] with weight function $w(x) = \frac{x}{L}$ and orthogonality property [25]:

$$\langle \phi_m, \phi_n \rangle_{w_L} = \left(\frac{\Gamma(n+2)}{L^2 n!}\right) \delta_{nm}, \quad (6)$$

where δ_{nm} is the Kronecker function.

2) Function approximation with Laguerre functions

A function $f(x)$ defined over the interval $I = [0, \infty)$ can be expanded as:

$$f(x) = \sum_{i=0}^{+\infty} a_i \phi_i(x), \quad (7)$$

Where

$$a_i = \frac{\langle f, \phi_i \rangle_w}{\langle \phi_i, \phi_i \rangle_w}. \quad (8)$$

If the infinite series in Eq. (7) is truncated with N terms, then it can be written as [25].

$$f(x) \simeq \sum_{i=0}^{N-1} a_i \phi_i(x) = A^T \phi(x), \quad (9)$$

with

$$A = [a_0, a_1, a_2, \dots, a_{N-1}]^T, \quad (10)$$
$$\phi(x) = [\phi_0(x), \phi_1(x), \dots, \phi_{N-1}(x)]^T. \quad (11)$$

3) Modified generalized Laguerre functions collocation method

Let $w(x) = \frac{x}{L}$ and x_j, $j = 0, 1, \dots, N-1$, be the N MGLF-Radau points that are the colocation points. The relation between MGLF orthogonal systems and MGLF integrations is as follows [25, 29-32]:

$$\int_0^{+\infty} f(x)w(x)dx = \sum_{j=0}^{N-1} f_j(x)w_j + \left(\frac{\Gamma(N+2)}{(N)!(2N)!}\right) f^{2N}(\xi) e^{\xi}, \quad (12)$$

where $0 < \xi < \infty$ and

$$w_j = x_j \frac{\Gamma(N+2)}{\left(L(N+1)!\left[(N+1)\phi_{N+1}(x_j)\right]^2\right)}, \quad j = 0, 1, 2, \dots, N-1.$$

In particular, the second term on the right-hand side vanishes when $f(x)$ is a polynomial of degree at most $2N-1$ [25]. We define:

$$u(x) = \sum_{j=0}^{N-1} a_j \phi_j(x), \quad (13)$$

4) Solving the problem with modified generalized Laguerre functions

To apply modified generalized Laguerre collocation method to Eq. (1) with boundary conditions Eq. (2), at first we expand $f(d_i)$ as follows:

$$f(d_i) = \sum_{j=0}^{N-1} a_j \phi_j, \quad (14)$$

To find the unknown coefficients a_j's, we substitute the truncated series $f(d_i)$ into Eq. (1) and boundary conditions in Eq. (2). Also, we define Residual function of the form:

$$Res(d_i) = \sum_{j=0}^{N-1} a_j \phi_j'''(d_i) + \sum_{j=0}^{N-1} a_j \phi_j(d_i) \sum_{j=0}^{N-1} a_j \phi_j''(d_i) - \left(\sum_{j=0}^{N-1} a_j \phi_j'(d_i)\right)^2 - id \sum_{j=0}^{N-1} a_j \phi_j'(d_i) \quad (15)$$

$$\sum_{j=0}^{N-1} a_j \phi_j(0) = 0, \quad (16)$$

$$\sum_{j=0}^{N-1} a_j \phi_j'(0) = 1, \quad (17)$$

$$\sum_{j=0}^{N-1} a_j \phi_j(\infty) = 0. \quad (18)$$

By applying d_i in Eq. (15) with the N collocation points which are roots of functions L_n^a , we have N equations that generate a set of N non-linear equations; also, we have two boundary equations in Eq. (16,17). Now, all of these equations can be solved by Newton method for the unknown coefficients. We must mention Eq. (18) is always true; therefore, we do not need to apply this boundary condition.

All in all, we can find the polynomial approximation function of $f(d_i)$; therefore, every Y_i can be decrypted on the recivers.

The absolute error between MGLFMs solution and exact solution of the velocity profile $f(d_i)$ for $id = 0.6$ is shown in Figure 2. This graph shows the error is negligible; Also, we know that every d_i is corresponding to an integer value; therefore, the result of decrypting must be rounded to the nearest integer value. In addition, the graph of the $\| \text{Res} \|^2$ for MGLF at $id = 0.6\, b_2 = 0.1$is shown in Figure3. This graph illustrates the convergence rate of the method.

IV. Performance Analysis

This approach proposed above is secure against chosen-plaintext attacks (CPA-secure) because it uses random masking mechanism. So the same message encrypts to different values each time making it hard for the adversary to guess the real data flow underneath the masking. Also, we sign every encrypted shares that makes this scheme secure against chosen-ciphertext attacks(CCA-secure).

Besides, decryption is based on the modified generalized Laguerre. Modified generalized Laguerre functions are orthogonal functions that solved the system of non-linear differential equations governing the problem on the semi-infinite domain without truncating it to a finite domain. Modified generalized Laguerre functions were proposed to provide simple way to improve the convergence of the solution through collocation method by $N = 20$, $\alpha = 1$and $L = 0.99$. Figure 2 and Figure 3 show this approch is more accurate and faster convergence in this problem.

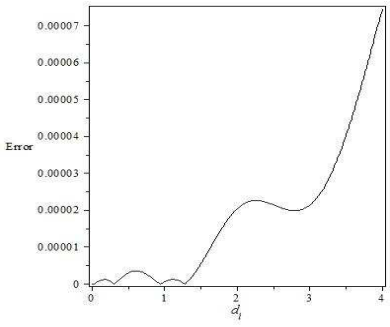


Figure 2. Graph of Error by MGLFMs solution for id=0.6

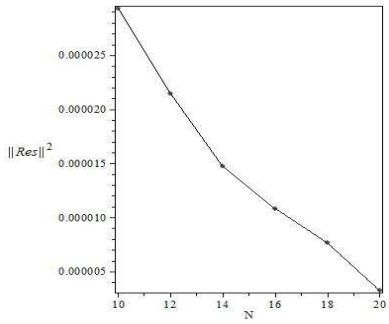


Figure 3. Graph of $\| \text{Res} \|^2 \| \text{Res} \|^2$ by MGLFs solution.

In addition, the code was writen in MAPLE 15 framework and the experiment was done on a Intel Core(TM) i3 CPU and 2.53 GHz processor. Under these conditions, the execution time of this solution was reported 0.96 seconds which is more efficient than other existing solutions. It is worth-mentioning that only the owner of the boundary conditions can solve this equetion in efficient time. To achieve a speed up, a reciever can make a look up table, and store all coeficients of the aproximation functions per each sensor *id*, after numerically calculating them once. Of course the trade off between memory and computation should be taken into account as the size of the table grows depending on the accurecy demand.

V. Conclusion

In this paper, we proposed an efficient scheme including special technique to defend against curious, and search-replace adversary as well as injection capable attacker. In fact, we shared and encrypted data using a numerical method (defence against curious, and search and replace adversary), and efficiently signed every unit of data to prevent injection attacks.

Moreover, based on an ODE and its boundary conditions, a new function for every sensor is released because every sensor has its special *id*. This equation is publically known but the calculated function is infeasible to obtain without knowing boundary condition. Hence, encrypted packet of any function provides no information about the original data. This technique is also firm against injection attack which is the most rampant attack in general unattended wireless sensor network. Furthermore, we can claim that, our model is applicable and scalable to real world applications and it is secure against statistical traffic analysis attacks since it blocks chosen-plaintext attacks and chosen-ciphertext attacks.

References

[1] Pietro,R. D. , Mancini, L. V.,Soriente,C.,Spognaedi, A.,Tsudik,G., "Playing hide-and-seek with a focused mobile adversary in unattended wireless sensor networks", Ad Hoc network, vol. 7, no. 8, pp. 1463-1475, 2009.

[2]Karlof,C.,Sastry,N., Wagner,D.,*"TinySec: A link layer security architecture for wireless sensor networks"*, in proceeding of the 2nd international conference on embedded networked sensor systems, New York, NY, USA: ACM, 2004, pp. 162-175.

[3]Ma,D.,Tsudik,G.,*"Extended abstract: Forward-secure sequential aggregate authentication"*, in proceeding of IEEE symposium on security and privacy (S&P), Oakland, CA, USA, May 2007, pp. 86-91.

[4]Ma,D.,*"Practical forward secure sequential aggregate signatures"*, in proceeding of the 3rd ACM symposium on information, computer and communications security (ASIACCS'08), ACM, NY, USA, 2008, pp. 341-352.

[5]Ma,D.,Tsudik,G.,*"A new approach to secure logging"*, ACM transaction on storage (TOS), vol. 5, no. 1, 2009, pp. 1-21.

[6]Ren, W., Zhao, J., Ren, Y.: network coding based dependable and efficient data survival in unattended wireless sensor networks. Journal of Communications.4, NO. 11,894-901 (2009)

[7] Gentry, C., Peikert, C., Vaikuntanathan, V.: trapdoors for hard lattices and new cryptography constructions: In STOC, ed. R. E. Ladner and C. Dwork, ACM, 197-206(2008).

[8]Boneh, D., Freeman, D. M.: linearly homomorphic signature over binary fields and new tools for lattice-based signatures. PKC'11, LNCS 6571. 1-16 (2011).

[9]Sakiadis, B. C.: Boundary-layer behaviour on continuous solid surfaces: I. boundary-layer equations for two-dimensional and axisymmetric Flow. AIChE J. 7, 26-28 (1961)

[10] Sakiadis, B. C.: Boundary-layer behaviour on continuous solid surfaces: II. boundary-layer equations for two-dimensional and axisymmetric flow. AIChE J. 7, 221-225 (1961)

[11] Crane, L. J.: Flow past a stretching plate. Z. Angew. Math. Phys. 21, 645-647 (1970)

[12]Andersson, H. I., Hansen, O. R., Holmedal, B.: Diffusion of a chemically reactive species from a stretching sheet. Int. J. Heat Mass Trans. 37, 659-664 (1994)

[13] Thakar, H. S., Chamkha, A. J., Nath, G.: Flow and mass transfer on a stretching sheet with a magnetic filed and chemically reactive species. Int. J. Eng. Sci. 38, 1303-1314 (2000)

[14] Raptis, A., Perdikis, C.: Viscous flow over a non-linearly stretching sheet in the presence of a chemical reaction and magnetic field. Int. J. Nonlinear. Mech. 41, 527-529 (2006)

[15] Rajagopal, K., Veena, P. H., Pravin, V. K.: Nonsimilar solutions for heat and mass transfer flow in an electrically conducting viscoelastic fluid over a stretching sheet saturated in a porous medium with suction/blowing. J. Porous Media. 11, 219-230 (2008)

[16] Bejan, A.: Convection heat transfer. Wiley-Interscience, New York, USA (1984)

[17] Akyildiz, F. T., Bellout, H., Vajravelu, K.: Diffusion of chemically reactive species in a porous medium over a stretching sheet. J. Math. Anal. Appl. 320, 322-339 (2006)

[18] Cortell, R.: MHD flow and mass transfer of an electrically conducting fluid of second grade in a porous medium over a stretching sheet with chemically reactive species. Chem. Eng. Process. 46, 721-728 (2007)

[19] Cortell, R.: Toward an understanding of the motion and mass transfer with chemically reactive species for two classes of viscoelastic fluid over a porous stretching sheet. Chim.Eng. Process. 46, 982-989 (2007)

[20] Prasad, K. V., Abel, M. S., Khan, S. K., Datti, P. S.: Non-darcy forced convective heat transfer in a viscoelastic fluid flow over a non-isothermal stretching sheet. J. Porous Media. 5, 41-47 (2002)

[21] Prasad, K. V., Abel, M. S., Datti, P. S.: Diffusion of chemically reactive species of a non-newtonian fluid immersed in a porous medium over a stretching sheet. Int. J. Non-Linear Mech. 38, 651-657 (2003)

[22] Ziabakhsh, Z., Domairry, G., Bararnia, H., Babazadeh, H.: Analytical solution of flow and diffusion of chemically reactive species over a nonlinearly stretching sheet immersed in a porous medium. J. Taiwan Inst. Chem. Eng. 41, 22-28 (2010)

[23] Kechil, S. A., Hashim, I.: Series solution of flow over nonlinearly stretching sheet with chemical reaction and magnetic fleld. Phy.Lett. A 372, 2258-2263 (2008)

[24] Dinarvand, S.: A reliable treatment of the homotopy analysis method for viscous flow over a non-linearly stretching sheet in presence of a chemical reaction and under influence of a magnetic fleld. Cent. Eur. J. Phys. 7, 114-122 (2009)

[25]Parand, K., Taghavi, A.: Rational scaled generalized Laguerre function collocation method for solving the Blasius equation. J. Comput. Appl. Math. 233, 980-989 (2009)

[26] Parand, K., Taghavi, A., Shahini, M.: Comparison between rational Chebyshev and modified generalized Laguerre functions Pseudospectral methods for solving Lane-emden and unsteady gas equentions. ActaPhysicaPolonica B. 40, 1749-1763 (2009)

[27] Coulaud, O., Funaro, D., Kavian, O.: Laguerre spectral approximation of elliptic problems in exterior domains. Comput.Method. Appl. Mech. Eng. 80, 451-458 (1990)

[28] Guo, B. Y., Shen, J., Xu, C. L.: Generalized Laguerre approximation and its applications to exterior problems. J. Comput. Math. 23, 113-130 (2005)

[29] Zhang, R., Wang, Z. Q., Guo, B. Y.: Mixed Fourier-Laguerre spectral and Pseudospectral methods for exterior problems using generalized Laguerre functions. J. Sci. Comput. 36, 263-283 (2008)

[30] Wang, Z. Q., Guo, B. Y., Wu, Y. N.: Pseudospectral method using generalized Laguerre functions for singular problems on unbounded domains. discret. contin. dyn. s. 11, 1019-1038 (2009)

[31] Iranzo, V., Falqus, A.: Some spectral approximations for differential equations in unbounded domains. Comput.Methods Appl. Mech. Engrg. 98, 105-126 (1992)

[32] Szeg, G.: Orthogonal polynomils. AMS, New York, (1939)

[33] Parand, K., Dehghan, M., Taghavi, A.: Modified generalized Laguerre function Tau method for solving laminar viscous flow: The Blasius equation. Int. J. Numer. Meth.Heat Fluid Flow. 20, 728-743 (2010)

[34] Parand, K., Shahini, M., Dehghan, M.: Rational Legendre Pseudospectral approach for solving nonlinear difierential equations of Lane-Emden type. J. Comput. Phys. 228, 8830-8840 (2009)

[35] Rajagopal, K., Tao, L.: Mechanics of mixture. World Scientific, Singapore, (1995)

[36] Gasper, G. Stempak, K., Trembels, W.: Fractional integration for Laguerre expansions. J. Math. Appl. Anal. 67, 67-75 (1995)

[37] Taseli, H.: On the exact solution of the Schrodinger equation with quartic anharmonicity. Int. J. Quantom. Chem. 63, 63-71 (1996)

[38] Babamir, F., S., Bayat Babolghani, F.: Linearly Time Efficiency in Unattended Wireless Sensor Networks, Real-Time Systems, Architecture, Scheduling, and Application, Dr. Seyed Morteza Babamir (Ed.), ISBN: 978-953-51-0510-7 (2012)

An Indoor Positioning Mechanism using Finite Memory Structure Filtering in Wireless Sensor Networks

Pyung Soo Kim[1], Eung Hyuk Lee,[2] and Mun Suck Jang[2]
[1]Department of Electronic Engineering, Korea Polytechnic University &
Center for Embedded Computer Systems, University of California, Irvine, Irvine, CA 92697, USA
[2]Department of Electronic Engineering, Korea Polytechnic University, 429-793, KOREA

Abstract— *This paper proposes an alternative indoor positioning mechanism in wireless sensor networks (WSNs). The proposed mechanism gives the filtered estimates for moving target's position and velocity in real-time, while removing undesired noisy effects and preserving desired moving locations. The well known finite memory structure (FMS) filter is adopted for the filtering. The filtered estimates for moving target's position and velocity have good inherent properties. Especially, it will be shown that a constant acceleration does not induce the error in filtered estimates for moving target's position and velocity. From discussions about the choice of window length and normalized noise covariance, it is shown that they can make the performance of the proposed mechanism as good as possible. Via extensive computer simulations, the performance of the proposed FMS filtering based mechanism is shown to outperform the existing infinite memory structure (IMS) filtering based mechanism for both zero acceleration and nonzero acceleration.*

Keywords: Indoor positioning system, Wireless sensor networks, Finite memory structure (FMS) filter, Infinite memory structure (FMS) filter.

1. Introduction

Recently, indoor positioning systems for wireless sensor networks (WSNs) have become very popular and thus been used successfully in a variety of scenarios, such as location detection and tracking of products stored in a warehouse, people within buildings such as hospitals and nursing homes. Because of indoor channel characteristics, accurate estimation mechanism is required for WSN positioning system. In an outdoor environment, the Global Positioning System (GPS) has been used in many outdoor applications for localizing people, cars, as well as other objects. However, GPS lacks the same level of efficiency when used within indoor environments because of obstacles that can weaken the signal of the GPS. Therefore, to track the positions for indoor systems in the WSN, several mechanisms are developed by measuring the distance or range value between the target and anchor sensor as shown in [1]-[4].

However, the estimated position can be corrupted by a couple of noises, the measurement noise and the system noise. The measurement noise is caused by inaccuracies in the tracking entity and the system noise is caused by turbulence or human error and other environmental factors. These noises occur due to multiple factors like environmental intrusion, inaccuracies in sensor measurements, turbulence affecting the target's movement, and human inability to navigate in a perfectly straight line. These noises must be filtered out in order to draw a true path of a moving target. Thus, several approaches introduce Kalman filter into WSN systems [5]-[7]. The Kalman filter has been well known as a recursive linear filtering model used to filter random inaccuracies in measurements to predict the most likely position and velocity of a moving target on real-time position coordinate. However, the Kalman filter has an infinite memory structure (IMS) that utilizes all past information accomplished by equaling weighting and has a recursive formulation. Thus, the Kalman filter tends to accumulate the filtering error as time goes. In addition, the Kalman filter has known to be sensitive and show even divergence phenomenon for temporary modeling uncertainties and round-off errors [8]-[12].

Therefore, in the current paper, an alternative indoor positioning mechanism is proposed in WSNs. The proposed mechanism gives the filtered estimates for position and velocity of moving target in real-time, while removing undesired noisy effects and preserving desired moving locations. For the filtering, the proposed mechanism adopts the well known finite memory structure (FMS) filter that utilizes only finite information on the most recent window [9][10]. The filtered estimates for position and velocity of moving target have good inherent properties such as unbiasedness, efficiency, time-invariance, deadbeat, and robustness due to the FMS.

Through discussions about the choice of window length and normalized noise covariance, it is shown that they can be considered as useful design parameters to make filtering performance of the proposed mechanism as good as possible. Finally, computer simulations shows that the performance of the proposed FMS filtering based mechanism can outperform the existing IMS filtering based mechanism.

The paper is organized as follows. In Section 2, an alternative indoor positioning mechanism using the FMS filtering is proposed for WSNs. In Section 3, a discussion about the choice of window length and covariance ratio is

shown. In Section 4, numerical simulations are performed to evaluate the performance of the proposed mechanism. Finally, conclusions are made in Section 5.

2. An Indoor Positioning Mechanism using FMS Filtering in WSN

To describe the random nature of moving target's successive locations in wireless sensor networks (WSNs), the following discrete-time 6th-order state-space model with sampling time T is introduced as shown in [5]-[7]:

$$\begin{aligned}\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} w(t) \\ \omega(t) \end{bmatrix}, \\ \begin{bmatrix} z(t) \\ \eta(t) \end{bmatrix} &= \begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} v(t) \\ \nu(t) \end{bmatrix} \quad (1)\end{aligned}$$

where

$$\begin{aligned} A &= B = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \\ C &= D = \begin{bmatrix} 1\ 0\ 0 \end{bmatrix}, \end{aligned}$$

and state vectors $x(t)$ and $y(t)$ along the *X* and *Y* directions are defined by:

$$x(t) \triangleq \begin{bmatrix} x_p(t) \\ x_v(t) \\ x_a(t) \end{bmatrix}, \quad y(t) \triangleq \begin{bmatrix} y_p(t) \\ y_v(t) \\ y_a(t) \end{bmatrix}.$$

The state variable $x_p(t)$ and $y_p(t)$ represents the moving target's random *positions*, and $x_v(t)$ and $y_v(t)$ represents the corresponding *velocities* at time t. The corresponding *accelerations* are defined by $x_a(t)$ and $y_a(t)$. These state vectors $x(t)$ and $y(t)$ cannot be observed directly, due to the random disturbances by fading and shadowing. To take these effects into account filtered estimates, the observation vector $[z(t)\ \eta(t)]^T$ is modeled by independent additive errors $[v(t)\ \nu(t)]^T$. The system noise vector $[w(t)\ \omega(t)]^T$ and the observation noise vector $[v(t)\ \nu(t)]^T$ are zero-mean white Gaussian and mutually uncorrelated.

As shown in (1), the state vectors $x(t)$ and $y(t)$ can be processed independently. Thus, in this paper, the moving target for the *X* direction is considered only. The moving target's successive locations for the *X* direction in the WSN can be represented by the discrete-time 3rd-order state-space model with sampling time T:

$$\begin{aligned} x(t+1) &= Ax(t) + w(t), \\ z(t) &= Cx(t) + v(t) \quad (2)\end{aligned}$$

where

$$\begin{aligned} x(t) &= \begin{bmatrix} x_p(t) \\ x_v(t) \\ x_a(t) \end{bmatrix} \triangleq \begin{bmatrix} x_m(t) \\ x_a(t) \end{bmatrix}, \quad w(t) \triangleq \begin{bmatrix} w_m(t) \\ w_a(t) \end{bmatrix}, \\ A &= \begin{bmatrix} \bar{A} & E \\ 0 & 1 \end{bmatrix}, \quad C = [\bar{C}\ \ 0], \end{aligned}$$

with

$$\bar{A} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \ E = \begin{bmatrix} T^2/2 \\ T \end{bmatrix}, \ \bar{C} = [1\ \ 0]$$

and the variances of $w(t)$ and $v(t)$ are Q and r, respectively.

The main task of the proposed indoor positioning mechanism is the filtering of moving target's random position in real-time, removing undesired noisy effects, while preserving desired position. For the filtering, the well known finite memory structure (FMS) filter in [9][10] is adopted. For the state-space model (2), the filtered estimate $\hat{x}(t)$ for moving target's position, velocity and acceleration processes linealy the only finite observations on the most recent window $[t-M, t]$ as the following simple form:

$$\hat{x}(t) = \begin{bmatrix} \hat{x}_m(t) \\ \hat{x}_a(t) \end{bmatrix} = HZ(t) = \begin{bmatrix} H_m \\ H_a \end{bmatrix} Z(t) \quad (3)$$

where the gain matrix H and the finite observations $Z(t)$ are represented by

$$\begin{aligned} H &\triangleq [h(M)\ h(M-1) \cdots h(0)], \quad (4) \\ Z(t) &\triangleq [z^T(t-M)\ z^T(t-M+1) \cdots z^T(t)]^T. \quad (5)\end{aligned}$$

The algorithm for filter gain coefficients $h(\cdot)$ in (4) is obtained from the following algorithm as shown in [9][10]:

$$h(j) = \Omega^{-1}(M)\Phi(j)C^T/r, \qquad 0 \le j \le M, \quad (6)$$

where

$$\begin{aligned} \Phi(l+1) &= \Phi(l)[I + A^{-T}\Omega(M-l-1)A^{-1}Q]^{-1}A^{-T}, \\ \Omega(l+1) &= [I + A^{-T}\Omega(l)A^{-1}Q]^{-1}A^{-T}\Omega(l)A^{-1} \\ &\quad + C^TC/r, \end{aligned}$$

with $\Phi(0) = I$, $\Omega(0) = C^TC/r$, and $0 \le l \le M-1$. H_m and H_a are the first 2 rows and the last 1 row of H, respectively. Note that $Z(t)$ in (5) can be represented in the following regression form:

$$Z(t) = \Gamma x_m(t-M) + \Lambda X_a(t) + \Theta W_m(t) + V(t) \quad (7)$$

where $X_a(t)$, $W_m(t)$, $V(t)$ have the same form as (5) and Γ, Λ, Θ are defined as follows:

$$\Gamma = \begin{bmatrix} \bar{C} \\ \bar{C}\bar{A} \\ \bar{C}\bar{A}^2 \\ \vdots \\ \bar{C}\bar{A}^M \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \bar{C}E & 0 & \cdots & 0 & 0 \\ \bar{C}\bar{A}E & \bar{C}E & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{C}\bar{A}^{M-1}E & \bar{C}\bar{A}^{M-2}E & \cdots & \bar{C}E & 0 \end{bmatrix},$$

$$\Theta = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \bar{C} & 0 & \cdots & 0 & 0 \\ \bar{C}\bar{A} & \bar{C} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{C}\bar{A}^{M-1} & \bar{C}\bar{A}^{M-2} & \cdots & \bar{C} & 0 \end{bmatrix}.$$

Ultimately, filtered estimates $\hat{x}_m(t)$ and $\hat{x}_a(t)$ for moving target's position, velocity and acceleration are obtained from (3) as follows:

$$\begin{aligned} \hat{x}_m(t) &= \begin{bmatrix} \hat{x}_p(t) \\ \hat{x}_v(t) \end{bmatrix} = H_m Z(t), \\ \hat{x}_a(t) &= H_a Z(t). \end{aligned} \quad (8)$$

Filtered estimates $\hat{x}_m(t)$ and $\hat{x}_a(t)$ have good inherent properties of unbiasedness, efficiency, time-invariance and deadbeat since the FMS filter used provides these properties. The IMS filter such as Kalman filter used in [5]-[7] does not have these properties unless the mean and covariance of the initial state is completely known. Among them, the remarkable one is the deadbeat property which filtered estimates $\hat{x}_m(t)$ and $\hat{x}_a(t)$ tracks the actual state vector $x_m(t)$ and $x_a(t)$ exactly in the absence of noises. The deadbeat property gives the following matrix equality as shown in [9][10]:

$$\begin{aligned} H \begin{bmatrix} C \\ CA \\ \vdots \\ CA^M \end{bmatrix} &= \begin{bmatrix} H_m \\ H_a \end{bmatrix} \begin{bmatrix} \Gamma & \bar{\Delta} \end{bmatrix} \\ &= A^M = \begin{bmatrix} \bar{A}^M & \sum_{l=0}^{M-1} \bar{A}^l E \\ 0 & I \end{bmatrix} \end{aligned}$$

where

$$\bar{\Delta} = \begin{bmatrix} 0 \\ \bar{C}E \\ \bar{C}\bar{A}E + \bar{C}E \\ \vdots \\ \sum_{l=0}^{M-1} \bar{C}\bar{A}^l E \end{bmatrix}.$$

Thus, the following identities are obtained:

$$\begin{aligned} H_m\Gamma &= \bar{A}^M, \quad H_m\bar{\Delta} = \sum_{l=0}^{M-1} \bar{A}^l E, \\ H_a\Gamma &= 0, \quad H_a\bar{\Delta} = 0, \end{aligned} \quad (9)$$

which will be used later.

As mentioned before, although the state space model (2) is developed with the consideration of random-walk acceleration, a constant acceleration can occur in actual situations. In this case, it will be shown in the following theorem that a constant acceleration on the observation window $[t-M, t]$ does not induce the error in filtered estimate $\hat{x}_m(t)$ for moving target's position and velocity.

Theorem 1: A constant acceleration on $[t-M, t]$ does not induce the error in filtered estimate $\hat{x}_m(t)$ for moving target's position and velocity.

Proof: When an acceleration is constant as $\bar{x}_a$ in the observation window $[t-M, t]$, the finite observations $Z(t)$ (7) can be represented in the following regression form:

$$\begin{aligned} &Z(t)\big|\{x_a(t) = \bar{x}_a \text{ for } [t-M, t]\} \\ &= \Gamma x_m(t-M) + \bar{\Delta}\bar{x}_a + \Theta W_m(t) + V(t). \end{aligned} \quad (10)$$

Then, the filtered estimate $\hat{x}_m(t)$ for moving target's position and velocity is derived from (3), (9) and (10) as

$$\begin{aligned} \hat{x}_m(t) &= H_m\big[\Gamma x_m(t-M) + \bar{\Delta}\bar{x}_a + \Theta W_m(t) + V(t)\big] \\ &= H_m\Gamma x_m(t-M) + H_m\bar{\Delta}\bar{x}_a \\ &\quad + H_m\big[\Theta W_m(t) + V(t)\big] \\ &= \bar{A}^M x_m(t-M) + \big[\sum_{l=0}^{M-1} A_m^l E\big]\bar{x}_a \\ &\quad + H_m\big[\Theta W_m(t) + V(t)\big]. \end{aligned} \quad (11)$$

The actual state vector $x_m(t)$ for moving target's position and velocity at current time t can be represented on the observation window $[t-M, t]$ as follow:

$$\begin{aligned} &x_m(t)\big|\{x_a(t) = \bar{x}_a \text{ for } [t-M, t]\} = \bar{A}^M x_m(t-M) \\ &+\big[\sum_{l=0}^{M-1} \bar{A}^l E\big]\bar{x}_a + \bar{\Theta} W_m(t) \end{aligned} \quad (12)$$

where $\bar{\Theta} = \big[\bar{A}^{M-1}\ \bar{A}^{M-2} \cdots I\ 0\big]$. Thus, using (11) and (12), the estimation error of filtered estimate $\hat{x}_m(t)$ is as follows:

$$\hat{x}_m(t) - x_m(t) = H_m\big[\Theta W_m(t) + V(t)\big] - \bar{\Theta} W_m(t)$$

which does not include the acceleration term. This completes the proof. ■

The Theorem 1 means that a constant acceleration does not degrade the moving target's tracking performance in the proposed indoor positioning mechanism. Therefore, it can be stated that the proposed FMS filtering based mechanism performs well for the moving target with constant acceleration as well as random-walk acceleration. This is a remarkable result of the current paper and cannot be obtained from the existing IMS filtering based mechanism in [5]-[7].

In addition, as mentioned previously, the proposed FMS filtering based mechanism has the deadbeat property, which means the fast tracking ability of the proposed mechanism. Furthermore, due to the FMS structure and the batch formulation, the proposed mechanism might be robust to temporary modeling uncertainties and to round-off errors, while the IMS filtering based mechanism might be sensitive for these situations.

3. Choice of Window Length and Normalized Noise Covariance

The important issue here is how to choose an appropriate window length M and normalized noise covariance Q/r to make the filtering performance as good as possible. They affect differently the performance of the proposed FMS filtering based mechanism for the WSN.

The noise suppression of the proposed mechanism might be closely related to the window length M. The proposed mechanism can have greater noise suppression as the window length M increases, which improves the filtering performance of the proposed mechanism. However, in case of too large window length M, the real-time application is somewhat difficult due to the computational load. This illustrates the proposed mechanism's compromise between the noise suppression and the computational load. Since M is an integer, fine adjustment of the properties with M is difficult. Moreover, it is difficult to determine the window length is systematic ways. In applications, one way to determine the window length is to take the appropriate value that can provides enough noise suppression.

The tracking ability of the proposed mechanism might be closely related with the normalized noise covariance Q/r when the window length M is determined. When the window length is fixed, the tracking ability of a filter increases and the noise-suppressing ability decreases as Q/r increases, and vice versa. Thus, Q/r is also a useful parameter in the adjustment of the tracking and noise-suppressing properties of the finite memory filtering based mechanism.

Therefore, it can be stated from above discussions that both the window length M and the normalized noise covariance Q/r can be considered as useful parameters to make the performance of the proposed mechanism as good as possible.

4. Computer Simulations

The performance of the proposed FMS filtering based indoor positioning mechanism is evaluated via extensive computer simulations and compared with the existing IMS filtering based mechanism in [5]-[7].

Computer simulations are performed for two cases. The first case considers 'zero acceleration' which the target moves with constant velocity. The second case considers 'nonzero acceleration' which the target moves with both constant and random-walk acceleration as shown in Fig. 1. The sampling period is taken as $T = 5$. The observation noise variance is $r = 40^2$ and the state noise variance is $Q = 0.173^2 I$. For the proposed mechanism, the window length is taken as $M = 20$. To make a clearer comparison, simulations of 30 runs are performed using different system and observation noises and each single simulation run lasts $500T$.

Fig. 2 shows the mean of root-squared estimation errors of the moving target's position for all simulations. The proposed FMS filtering based mechanism is comparable with the IMS filtering based mechanism in case of zero acceleration. On the other hand, the proposed mechanism is better than the IMS filtering based mechanism in case of nonzero acceleration for all simulations. Fig. 3 and 4 show the root-squared estimation error for mobile position and velocity using one sample simulation. It can be seen that the proposed FMS filtering based mechanism is remarkably better than the IMS filtering based mechanism on the interval where the acceleration varies abruptly. These results might come from the fast estimation ability due to deadbeat property and FMS structure of the proposed indoor positioning mechanism.

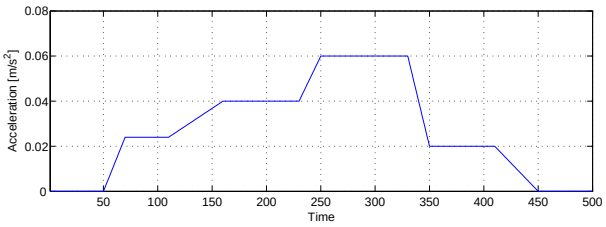


Fig. 1: Nonzero acceleration

5. Concluding Remarks

In this paper, the alternative indoor positioning mechanism using FMS filtering has been proposed for WSNs. The proposed mechanism gives the filtered estimates for moving target's position and velocity in real-time, while removing undesired noisy effects and preserving desired locations. The filtered estimates for moving target's position and velocity have good inherent properties. Especially, it has shown that a constant acceleration does not induce the error in filtered estimates for moving target's position and velocity. In addi-

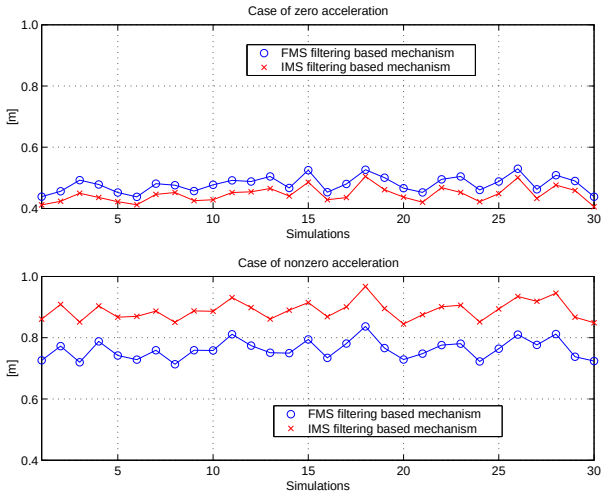


Fig. 2: Mean of root-squared estimation error of moving target's position for all simulations

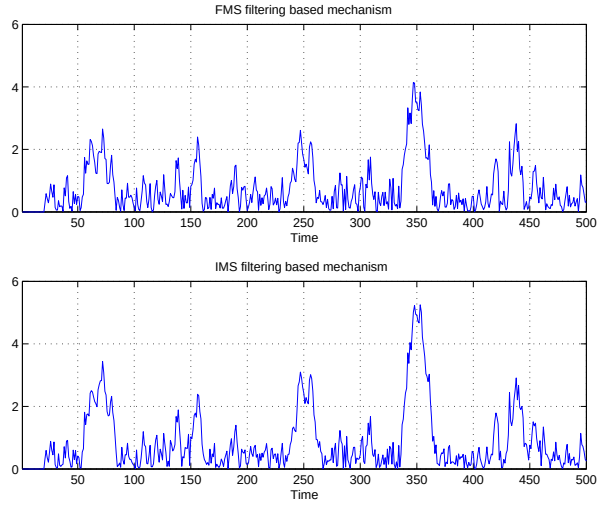


Fig. 3: Root-squared estimation error of moving target's position in case of nonzero acceleration

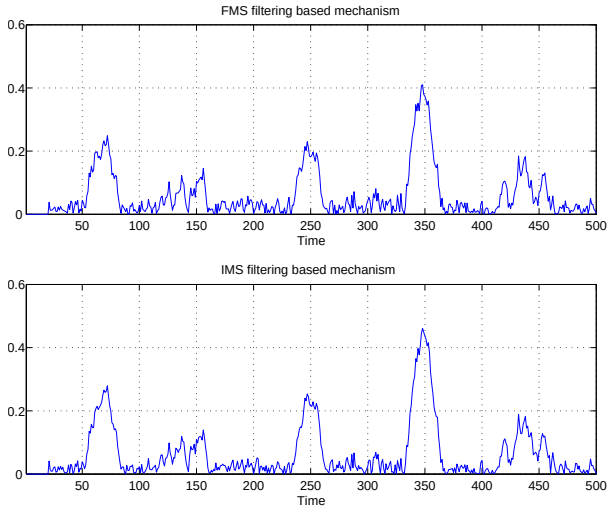


Fig. 4: Root-squared estimation error of moving target's velocity in case of nonzero acceleration

tion, it has shown that they can make the performance of the proposed mechanism as good as possible from discussions about the choice of window length and normalized noise covariance. Finally, through extensive computer simulations, the performance of the proposed FMS filtering based mechanism has shown to outperform the existing IMS filtering based mechanism for both zero acceleration and nonzero acceleration.

Acknowledgment

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the CITRC(Convergence Information Technology Research Center) support program (NIPA-2012-H0401-12-1007) supervised by the NIPA(National IT Industry Promotion Agency).

References

[1] N. Patwari, J. Ash, S. Kyperountas, A. H. III, R. Moses, and N. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, p. 54 69, 2005.

[2] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 6, p. 1067 1080, 2007.

[3] S. Bhatti and J. Xu, "Survey of target tracking protocols using wireless sensor network," in *Proc. 5th International Conference on Wireless and Mobile Communications*, 2009, pp. 110–115.

[4] I. Amundson and X. Koutsoukos, "A survey on localization for mobile wireless sensor networks," *Lecture Notes In Computer Science*, vol. 5801, p. 235 254, 2009.

[5] A. Ribeiro, I. Schizas, and S. R. G. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Systems Magazine*, vol. 30, no. 2, p. 66 86, 2010.

[6] J. Yi and L. Zhou, "Enhanced location algorithm with received-signal-strength using fading kalman filter in wireless sensor networks," in *Proc. 2011 International Conference on Computational Problem-Solving(ICCP)*, 2011, pp. 458 – 461.

[7] Y. Chen, D. Ni, L. Zhang, and C. Deng, "Realizing mobile node tracking in wireless sensor network based on kalman filter," in *Proc. 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN),*, 2011, pp. 299–302.

[8] R. J. Fitzgerald, "Divergence of the Kalman filter," *IEEE Trans. Automat. Contr.*, vol. 16, no. 6, pp. 736–747, 1971.

[9] W. H. Kwon, P. S. Kim, and S. H. Han, "A receding horizon unbiased FIR filter for discrete-time state space models," *Automatica*, vol. 38, no. 3, pp. 545–551, 2002.

[10] P. S. Kim, "An alternative FIR filter for state estimation in discrete-time systems," *Digital Signal Processing*, vol. 20, no. 3, pp. 935–943, 2010.

[11] Y. S. Shmaliy and O. Ibarra-Manzano, "Linear optimal FIR estimation of discrete time-invariant state-space models," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3086–3096, 2010.

[12] ——, "Time-variant linear optimal finite impulse response estimator for discrete state-space models," *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 2, pp. 95–104, 2012.

Multiple-Sensor Fusion Target Tracking using *ClusterTrack* Algorithm

S.M.R. Farshchi
ASR Centre for Embedded Systems Research
Sadjad Institute of Technology
Iran, Mashhad.
Mr.farshchi@sadjad.ac.ir

Hossein Sharifi Noghabi
School of Engineering and Design
Sadjad Institute of Technology
Iran, Mashhad.
hsharifi219@sadjad.ac.ir

Abstract— **Despite the minimal information provided by a binary proximity sensor, a network of these sensors can provide significant target tracking performance. This article deals with the performance examination of such a network for tracking multiple targets. We began with geometric arguments that address the problem of counting the number of distinct targets, given a snapshot of the sensor readings. Then necessary and sufficient criteria provided for an accurate target count in a one-dimensional setting, moreover, a greedy algorithm defined that determines the minimum number of targets that is consistent with the sensor readings. While these combinatorial arguments bring out the difficulty of target counting based on sensor readings at a given time, they leave open the possibility of accurate counting and tracking by exploiting the evolution of the sensor readings over time. To this end, we develop a particle filtering algorithm based on a cost function that penalizes changes in velocity. Finally, an extensive set of simulations, as well as experiments with passive infrared sensors, are reported.**

I. INTRODUCTION

Wireless sensor networks are set of tiny equipment's that have sensing processing and communicating competences together and provides lots of applications [1]. Amongst them, target tracking is a unique and important application that requires a cooperative processing to obtain robust results [2]. We assay the problem of tracking targets using an energy-efficient target tracking of binary sensors. There are two kinds of binary sensing models: the ideal binary sensing model and imperfect model. In the ideal model, each node can detect the target whenever it appears in the node's sensing range R (Figure 1(a)). Actually, detection ranges often vary regarding the environmental situations, such as the positioning of the target and the sensor. These factors make target detection near the boundary of the sensing range less predictable. Mentioned fact leads to an imperfect model in which the target is always detected within an inner disk of radius R_{in} but it is detected only with certain nonzero probability in an annulus between the inner disk and an outer disk of radius R_{out}. No detection will happen out of outer disk (Fig. 1(b)).

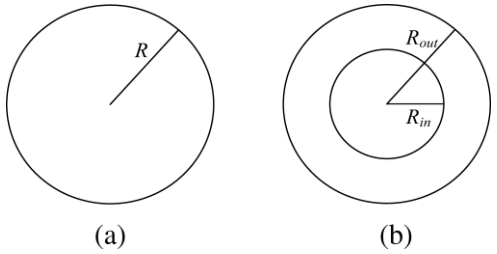


Fig. 1. Binary sensing models: (a) ideal, and (b) imperfect.

Each sensor produces a single bit of output, which is 1 when one or more targets are in its sensing range and 0 otherwise. These sensors are not able to distinguish individual targets, deciding how many distinct targets are in the range, or provide any location-specific information. Despite the minimal information provided by an individual sensor, a collaborative network has been shown in prior work [3] to yield respectable performance when tracking a single target: the resolution with which the target can be localized is inversely proportional to ρR^{d-1}, where ρ is the sensor density, R is the sensing range, and d is the dimension of the space. we study the problem of multiple targets tracking with binary sensors, without a priori knowledge of the number of targets.

Our focus is on the efficacy of collaborative tracking. Thus we assume that all of the sensor readings are available at a centralized processor, which can then estimate the targets' locations and trajectories. Distributed implementations of our algorithms, in which neighbors collaborate to estimate segments of trajectories, are possible, but not considered here.

In this approach we present an innovative distributed, energy-efficient, and fault tolerant target tracking algorithm using binary sensor networks that is able to track a target in both ideal and imperfect binary sensing models. Each awake node can estimate target's trajectory, location and speed locally, in cooperation with its neighbors. Likewise, our algorithm does not need synchronization for all networks and can track the target in real time under various paces and

moving in desultory directions. The algorithm in fault tolerant when a node fails to senses a target inside its range or lost packets because of collision.

The rest of the article is organized as follows: Section 2 literature review. Section 3 discusses our model and parameters briefly and then discusses the problem of target counting based on a snapshot of the sensor readings. Section 4, describes our particle filtering algorithm. Section 5 provides simulation results, while Section 6 describes our experimental setup and results. We end with Section 7, conclusion.

II. Related Works

The problem of tracking multiple targets using sensor networks has been explored in many references [4]. Owing to its simplicity and minimal requirements, the specific use of binary sensors for tracking applications has also drawn considerable attention of late. However, most of the work related to binary sensing has been applied to tracking a single target [6]. The tracking techniques employed in the large-scale deployment in [7] can be loosely interpreted in terms of a binary sensing model, even though a variety of sensing modalities and a variety of targets were considered in [8] contained a distributed tracking method for a binary sensor network, but assumed perfect knowledge about the number of targets and their identities, unlike our approach.

In our work, we investigate both target counting and tracking. Prior work on counting targets includes [9] but it assumed more detailed sensing capabilities than simple binary model. The classical framework for tracking is based on Kalman filtering, with a linear model for the sensor observations corrupted by Gaussian noise; for example, [10] investigated the use of Kalman filtering for distributed tracking. In recent years, the use of particle filters, which can handle more general observation models, has become popular [11]. However, most prior work about using particle filters for tracking in sensor networks [12] has assumed a richer sensing model than the binary model we consider. Exceptions are the prior work in [13] on the use of particle filters for tracking a single target using binary sensing, and also the preliminary results from our conference publication [5]. In this article, we build on [13] providing new analytical design criteria that assist in the efficient and reliable operation of our particle filter algorithm, and present a more detailed simulation-based analysis to evaluate the performance of the algorithm. In addition, we include simulation results and new theoretical proofs for two dimensions [14] only considered a one-dimensional setting).

III. Procedure for Paper Submission

A. Network Model and Assumptions

There are N nodes that randomly uniformly distributed a delimited two-dimensional region for monitoring. Each node has a unique ID and surely all the nodes together can cover entire region. To simplify, it is assumed that sensors have the same sensing range R under ideal sensing or the same radii R_{in} and R_{out} under the imperfect sensing. Nonetheless, this approach can work even when ranges vary from sensor to sensor. Each node begets one bit of information only when there is a change in target's status. Otherwise we get no information about features of the target. While there is no detection node remains silent to save energy and bandwidth. When it generates a new bit of information, in will be sent to the neighbors, nodes whose sensing ranges have intersect with node's coverage range.

Another assumption is that sensors are immobile. For example, Unattended Ground Sensors (UGS) used in the military and security applications to make this case worthy of study. We also assume that each node knows its own location. This assumption can be satisfied by using some low-power GPS or localization methods [10]. Since in this algorithm each sensor estimates target's position concerning to its location and range, in order to report these outputs, they do not need to be acquainted of their position in planer region. To estimate velocity, node requires position of neighbors adjacent to each other which can be organized via triangulation. Hence, it is not necessary for nodes to know geographical location of each node. In order to estimate trajectory, the location of neighbors specified by triangulation and then locating each node's geographical situation on the basis of some of the neighbors having GPS is a proper procedure [13] that can be done at the network deployment stage, but it is not mentioned here. Sensors exchange their location information through communication at the network deployment stage. Each sensor has its own local timer and can time stamp sent or received messages. Additionally, we assume that the target moves with velocity that is low relative to the node's sensing frequency. Consequently, time of discovery of the change in the target's presence within the node's sensing range differs little from the time at which the target moves within or out of this range under the ideal binary sensing model.

B. Snapshot-Based Inference

Our investigation began with asking under what circumstances an algorithm can reliably determine the number of distinct targets in the field, given a snapshot of the sensor readings. In order to develop fundamental geometric insights, we restrict attention in this section to an idealized model in which each sensor's coverage area is a circular disk of radius R: each sensor detects a target without fail if it falls within this disk, and does not produce false positives or negatives. While we develop our basic ideas and theorems in one dimension, we comment on their relevance and extensions to higher dimensions as appropriate.

C. Binary Sensing

Some *spatial separations* amongst the targets are clearly a necessary precondition for accurately disambiguating among different targets, but what does that mean, and how much separation is enough? For instance, is the following simple condition adequate: *each* target moves sufficiently (arbitrarily) far from the remaining targets at some point during the motion. Let us call this the condition of *individual*

separation. Unfortunately, as the following simple result shows, only this factor is not enough to count the number of targets accurately.

Even arbitrarily large individual separation is not sufficient to reliably count a set of targets using binary sensors. We give a construction in one dimension establishing the claim. Imagine a group of *m* targets moving at uniform speed along a straight line *L*. Initially, all targets are together and appear as one target to the sensor field. Now let target 1 speed up and move away from the rest of the group. Once it moves sufficiently far to the right, we can infer that there are at least two targets. Next, target 1 stops and waits until the rest of the group meets up with it, and then they all resume their motion. Then target 2 separates from the rest of the group and repeats the action of target 1, and so on. One can easily see that, in this scenario, *every* target achieves large individual separation from the rest, and yet no binary sensing-based algorithm can ever decide whether there are two targets or *m* targets, for an arbitrary value of *m*.

D.The Geometry of Target Counting

We begin with some geometric preliminaries. Suppose we have *N* binary proximity sensors deployed along a line. Each sensor's range is then an interval of length 2*R*. We use the notation C_i to denote the interval covered by sensor *i* (that is, sensor *i* outputs 1 if and only if a target falls in C_i). We assume that the domain of interest is covered by the union of the $\{C_i\}$, thus, there are no gaps in coverage. Any positioning of targets along the line leads to a vector of binary outputs from the sensors. In particular, we have contiguous groups of "on-sensors" separated by groups of "off-sensors." Geometrically, the on-sensors inform us about the intervals on the line where the targets might be, and the off-sensors tell us about the regions where there are no targets. If we let *I* be the set of sensors whose binary output is 1 and *Z* be the set of sensors whose output is 0, then all the targets must lie in the region *F*, which we call the *feasible target space*:

$$F = \bigcup_{i \in I} C_i - \bigcup_{j \in Z} C_j$$

The region *F* is a subset of the line, whose connected components are unions of portions of the sensing ranges of the on-sensors. An example is shown in Figure 2.

The feasible target space has an interesting geometric structure. While each on or off sensor contributes exactly 1 bit, the *information* content of the off sensors seems richer, especially in localizing the targets: the 1 bit only tells us that there is at least one target *somewhere* in the sensor's range, the 0 bit assures us that there is no target anywhere in the sensor's range. This observation leads to the following geometric property of the region *F*.

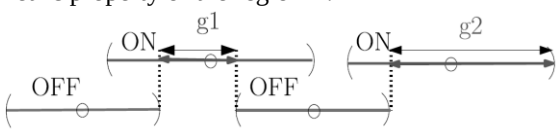


Fig. 2. A sample illustration for the feasible target space (*F*). Here, *g*1 and *g*2 represent the contributions of the "ON" sensors to *F*.

LEMMA 1. *Any two connected components of the feasible target space F are separated by at least distance* 2*R*.

PROOF. Choose a point *x* that is between two connected components of *F*. Since *x* must lie in the range of some sensor, and $x \_\in F$, that sensor must have binary output 0.

A sensor with binary output zero eliminates length 2*R* of the line for possible locations of the targets, and so the "gap" containing the point *x* must be at least as wide as 2*R*.

IV. TRACKING ALGORITHM

To demonstrate our basic idea, we use the example in Figure 4, which shows a target moving through an area covered by three nodes. At first, the target is outside of the sensing ranges of the nodes. Later, it falls in the sensing range of node N_x at time t_1, and then sensing ranges of N_y at time t_2 and N_z at time t_3. At last, it leaves sensing ranges of nodes, in that sequence, at times t_4, t_5, t_6, respectively.

The initial idea of the tracking algorithm under the ideal model was mentioned in [13, 14] and it can be mentioned briefly as follows. At the time of arrangement, first each node initializes its neighbor's status to "0" on its own list. whenever a sensor received "1" from a adjacent node it updates the status of that neighbor to value "1" on the list. At the moment at which the node senses a change in target's existence within its range, it identifies the arc of its sensing range border circle that the target is crossing. The target location is estimated as the middle point of the corresponding arc and broadcasted to neighbors. We can use two different local times when the target tracked in different locations to estimate its velocity. A weighted line fitting method is used to find a line, approximating a fragment of the target trajectory, that best fits the estimated target locations.

A. Initialization and Information Update

First, each node generates a list of its neighbors. Each record of the list saves these information including: neighbor node identifier, intersection points of the sensing circles of the node and its neighbor, an angle corresponding to the arc defined by these intersection points, and one-bit information produced by the neighbor, initialized to "0". As soon as receiving value "1" from a adjacent sensor, the sensor alters the status of the related neighbor in the list.

B. Location Estimate

To determine the location, we use different combinations of angles related to intersection points. As an example in Figure 3, If the neighbors both outcome bit "1", the corresponding central angles are combined by "&" operation that returns the intersection of these two angles. As shown in Figure 3(a), the common angle of $\angle 1o3$ and $\angle 2o4$ is $\angle 2o3$, so the node N_y estimates the target location as the

middle point of arc "23" when it senses that the target just

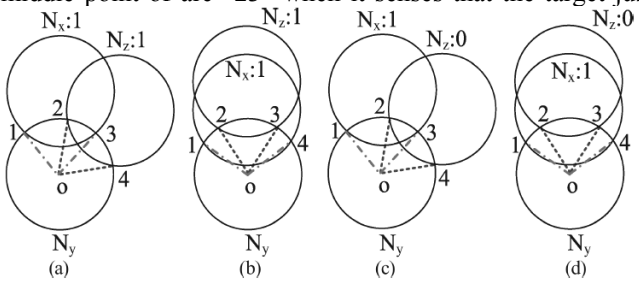


Fig. 3. Instances of angle combinations

moved within its sensing range. One special instance is shown in Figure 3(b), where the common angle is just one of the two angles.

C. Tracking Multiple Targets: The CLUSTERTRACK Algorithm

We call our proposed scheme CLUSTERTRACK. The method is specifically designed to prevent a subset of targets from monopolizing all of the available particles. To this end, instead of looking for clusters at the end, we monitor their formation throughout the tracking process, and limit the number of particles per cluster. We still retain K particles at each time instant. However, instead of picking the K best particles, we pick the K best particles subject to the constraint that the number of particles per cluster does not exceed a threshold H. A cluster is defined as a group of particles that are "similar," where similarity between two particles is measured in terms of a distance metric to be specified. Thus we scan the set of particles in increasing order of cost functions as before, but we retain a particle only if the number of similar particles retained thus far is less than the threshold H. This procedure enhances the likelihood that the particle filter catches all of the targets. In order to ensure that we do not end up scanning the entire sequence of particles at each instant, we can also put a limit L ($L > K$) on the number of particles that we consider. In this case, we stop the search for particles when either K particles have been retained, or L of them have been scanned, whichever happens first. The actual number of particles retained at time t is denoted by K_t, where $K_t \leq K$.

At the final time instant, we take the best particle from each of the clusters obtained, and designate it as our estimate of the trajectory followed by one of the targets. An alternative would be to choose a "consensus path" (e.g., based on a median filter at each time instant) for each cluster.

The pseudo code description for the CLUSTERTRACK at a particular time instant t is given in Algorithm 1. Cluster $_j$ represents the jth cluster, count $_j$ denotes the number of particles retained in Cluster $_j$, N_C is the number of clusters, H is the maximum number of particles to be retained from a particular cluster, and L is the maximum number of particles to be inspected in order to find the surviving particles at time t. We work under the assumption of smooth target trajectories (i.e., the targets do not have abrupt velocity changes), and hence pick a cost function that penalizes changes in velocity.

Let $\mathbf{P} = (\hat{x}[1], \ldots, \hat{x}[t])$ denote a particle. The instantaneous estimate of this particle's velocity vector at any time $n \in [1, t\text{-}1]$ is the increment in position $\hat{x}[n+1] - \hat{x}[n]$. The instantaneous contribution to the cost, in moving from time n to $n+1$, is taken to be the norm of the change in velocity

$$\begin{aligned} c[n] &= ||(\hat{x}[n+1] - \hat{x}[n]) - (\hat{x}[n] - \hat{x}[n-1])|| \\ &= ||\hat{x}[n+1] + \hat{x}[n-1] - 2\hat{x}[n]|| \end{aligned}$$

where $\|\cdot\|$ denotes Euclidean norm. Assuming that rapid accelerations are unlikely in smooth paths, the cost $c[n]$ should be inversely related to the probability that a target moves from the location $\hat{x}[n]$ at time n to $\hat{x}[n+1]$ at time $(n+1)$, given that it had moved from $\hat{x}[n-1]$ to $\hat{x}[n]$ between time instants $(n\text{-}1)$ and n. The net cost function associated with the particle $\mathbf{P}$ is simply taken to be the sum of the incremental costs:

$$\sum_{n=2}^{t-1} c[n].$$

V. TARGET MODELING

Since a moving target has dynamics time variant, a number of models can be established to describe a target's motion. Even though more models can give better overall estimates, it is less efficient since more time will be required to yield the estimate. In the sense of acceleration of a target, the

constant velocity (uniform) and acceleration (maneuvering)modes are most commonly considered to build models. In this project, these two models are also used. Linear accelerations are normally quite small and thus can be reasonably covered by a process noise in a nearly constant velocity model, i.e. the constant velocity motion plus a zero-mean noise with an appropriate covariance representing the small acceleration [8]. Alternatively, this mode can be described as a constant velocity model with no process noise. On the other hand, the acceleration mode has the acceleration increment during the sampling time, and this

should be included in the state space model. When the state space equation is given by:

$$x(k) = Fx(k-1) + Gw(k-1) \tag{1}$$

x is the state vector of a target defined as:

$$x = [\xi \quad \dot{\xi} \quad \ddot{\xi} \quad \eta \quad \dot{\eta} \quad \ddot{\eta}]' \tag{2}$$

where ξ and η denote longitudinal and lateral position respectively. In (3.1), w signifies process noise, which is zero-mean, white, and Gaussian with covariance $Q(k)$.

The state transition matrices and the noise gain matrices for each mode can be written in the following forms:

$$F_1 = \begin{bmatrix} 1 & T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad G_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{3}$$

and

$$F_2 = \begin{bmatrix} 1 & T & 0.5T^2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & 0.5T^2 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

$$G_2 = \begin{bmatrix} 0.5T^2 & 0 \\ T & 0 \\ 1 & 0 \\ 0 & 0.5T^2 \\ 0 & T \\ 0 & 1 \end{bmatrix}$$

where subscripts 1 and 2 in equations 3 and4 denote mode 1 (constant velocity motion) and mode 2 (acceleration motion) and T is the sampling time. Process noise covariance is simplified under the assumption that the process noise variance in each coordinate is equal and constant. In this case,

$$Q(k) = \sigma_{w_\xi}^2 = \sigma_{w_\eta}^2 = q$$

and q can be chosen by using the following inequality:

$$0.5\Delta a_{\max} \le \sqrt{q} \le \Delta a_{\max}$$

where $\Delta a_{\max}$ is the maximum acceleration increment over a sampling time.

The measurement model can be written as:

$$z(k) = Hx(k) + v(k)$$

$$H_1 = H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The measurement matrix implies that only the position of the target is measured from each sensor.

The mode sequence is assumed to be a first order Markov chain with transition probabilities:

$$\pi_{ij} = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$$

These transition probabilities imply that a target keeps its current mode with high probability rather than changes its mode. It is obvious that the results of our tracking algorithm become more accurate with an increasing number of neighbors reporting the status of the target. To evaluate the role of the adjacent nodes in tracking a target, following simulation configuration was used: 800 nodes distributed over an area with 800 units by 800 units. Each node's sensing range, *R (Rout* for imperfect model), differed from 40 to 150 units. To evaluate impact of sensor spatial density, first we assume fix range for sensing range, 40m, then, increase it to 150. If we diminish the area from 800m by 800m to 213.3m by 213.3m, in both of them number of neighbors changes but in the second also the spatial density of nodes per square meter increases. In terms of simulation for this concern placement of each node in relation to other nodes, remains constant and velocity of the target is adjusted proportionally to the sensing range, and stay constant if calculated in sensing range units.

All the paths limited inside the square area with length of 800− $R_{\max}$ located in the middle of the simulated region in order to exclude the boundary effect, where $R_{\max}$ is the maximum range (150 units) in the simulation. For the random turn trajectory, the length of each linear piece of the trajectory is random but proportional to the sensing range. As in [17], we set $R_{\min} = 0.9 \times R_{out}$ under the imperfect binary sensing model.

A. Tracking with Ideal Sensing

We considered five targets, and generated trajectories over 20 time instants for each of them. In keeping with our assumption of smooth target trajectories (i.e., no abrupt velocity changes), we picked the velocity of a particular target, at each instant, randomly within ±20 % of some mean value (using a uniform distribution). The model applies, for instance, if we consider the motion of vehicles on a freeway, over a reasonably short time window. The parameter ρR was taken to be 1 (i.e., the separation between consecutive sensors was equal to the sensing radius, so that the coverage areas for two adjacent sensors had 50% overlap).

With (roughly) constant velocity motion, as long as the velocities of two targets are not equal, they are guaranteed to separate out at some point of time. We therefore simulated two types of scenarios: (a) targets starting out well separated, getting close to each other, and then separating out again; (b) targets starting in close proximity to begin with, and then separating out. We found that our algorithm performed fairly well in both settings. Sample plots are shown in Figures 4(a) and 4(b), each corresponding to a single simulation run. We see that the algorithm succeeded in catching and tracking all targets. We note that the performance of the algorithm varied across simulation runs, and, over multiple runs, the algorithm generated between five and seven trajectories, with five of the trajectories almost invariably providing good approximations of the true paths. For example, the results from a simulation resulting in seven estimated trajectories are shown in Figure 4(c), where the additional spurious estimates are marked by the special characters. Note that we

Got spurious estimates of both types, low-cost smooth estimates (the estimate marked by "o"), and also high-cost estimates with sharp transitions (the estimate marked by "*"). In general, the emergence of low-cost spurious estimates was governed by the nature of the true trajectories: if the true trajectories allow smooth transitions from one to another, low-cost spurious estimates can arise. On the other hand, the

high-cost spurious estimates were seen to emerge only in (a subset of) those cases when the algorithm had to be rerun, because the trajectories generated in the first go could not satisfy the lower bounds on the target count.

The accuracy of location estimation resulting from our algorithm with fewer number of neighbors (sensing range of 40 units which means a node has 5 neighbors) is the same with outcomes of algorithm (1) with high number of neighbors (sensing range 150 units which means a node has 87 neighbors) under ideal binary sensing model. This can be proved by two ways, distributing enough sensors for other methods which is high and unnecessary for our approach. First we can use all distributed nodes to gain better accuracy and make our algorithm more fault tolerant. Because sensors fail with different reasons but our algorithm will still provide adequate accuracy even with failing of some of redundant nodes for location estimation. On the other hand our algorithm can works with some scheduling methods which means turning off some of the adjacent sensors to store energy. Our tracking algorithm achieves the same location estimate accuracy as algorithm (1) even when almost 95% of adjacent nodes are turned off.

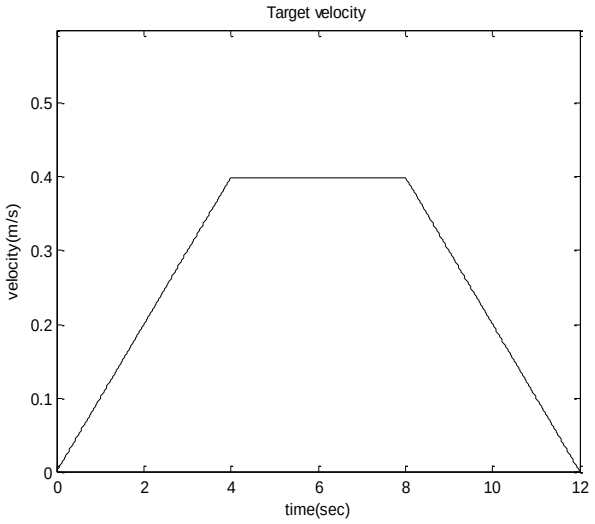
We also analyzed the number of messages exchanged and their corresponding energy cost. Let the target move from point O_1 to point O_2 with velocity $v(t)$ over time dt as shown in Figure 7. Area R_0 contains sensors that will broadcast bit "0" when the target moves from O_1 to O_2, and equal size area R_1 contains sensors that would transmit bit "1". Hence, the total number of messages generated by the target moving from O_1 to O_2 will be $2A\rho$, where A is the size of area R_0, ρ is the sensor density per unit square. A can be computed from Eq. (2), yielding $A = (2\alpha+\sin(2\alpha))R_2 \approx 4R_2\alpha$. Thus, the total number of messages generated is $4R\rho v(t)dt$. If over time t_r, the target moves distance D, then the number of messages produced is:

$$\int_{t=0}^{t_r} 4R\rho v(t)dt = 4R\rho D.$$

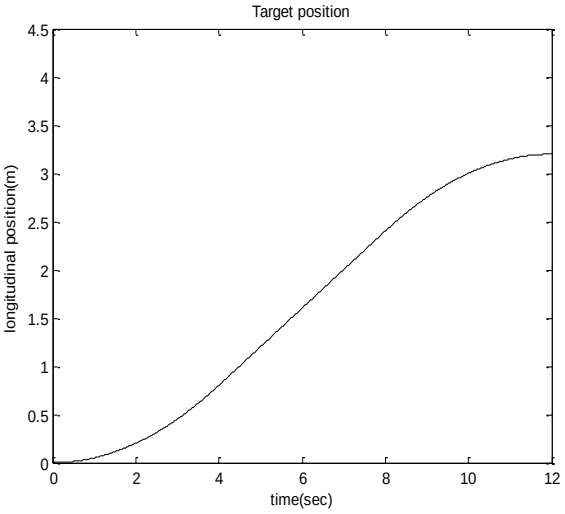
To verify this analysis, we performed two groups of simulations over an area of size 800 by 800 covered by 800 sensor nodes (the same density in each group). We set the sensing range R, 40 units in one group and 150 units in another. In both simulations, the target moves along a random trajectory (the same for each group) with a constant velocity and over the same distance. We ran the simulation 20 times for each group changing the topology of the network in each run. The exchanged messages were counted. The total number of messages exchanged is 474 when R = 40 units and 1759 when R = 150 units.

The ratio is 1759/474 = 3.71 which is close to the ratio of sensing ranges 150/40 = 3.75.

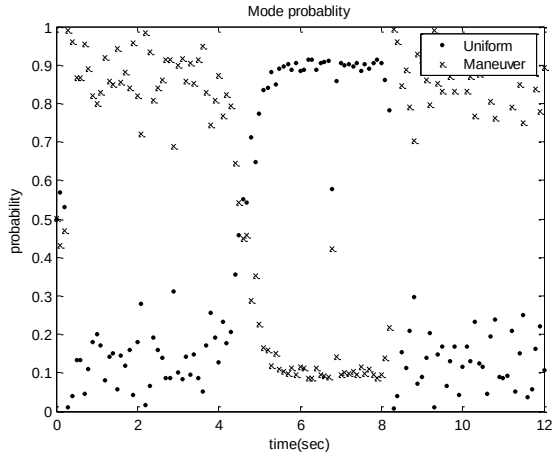
We used a small testbed with five PIR sensors placed uniformly along a line; see Figure 5. Each sensor sent a measurement to the base station when it changed state, and the base station was interfaced to a PC through a serial port. The data got time stamped at the PC, so that each of the final set of measurements included: value, position (mapped from node ID), and time. For the ground truth regarding target trajectories, the (human) targets were provided with separate sensor nodes (equipped with localization engines) with buttons, which they pressed as they passed by a set of known locations on the way. While each sensor in our experimental set up sent a measurement when it changed state, our problem formulation in Section 4 is based on the assumption that all sensors send their measurements at regular time instants. To apply our algorithm, therefore, we sampled the collected data at regular time instants, and assumed that the reading of a particular sensor at any time was the same as the one after its last toggle. Another implementation issue we faced was that, even when a target was detected as it entered the field of a sensor; the sensor output became 0.



(a)



(b)



(c)

Fig. 4. Example scenarios to depict the performance of Mode Probability in Longitudinally Moving Target Tracking (without measurement noise)

immediately after the detection, and kept toggling between 0 and 1 as the target moved toward the sensor. A probable reason for this is that the modules we used are meant for triggering a relay that resets after a certain amount of time, with the aim of minimizing false alarms, at the cost of some missed detections. To deal with this issue, we simply decided to neglect every 1 →0 transition that was

VI. CONCLUSIONS

Target tracking is one of the most important applications of sensor networks that can be done by collaboration between sensors. In this approach authors proposed a new algorithm for binary sensing models.

Reducing the target tracking error is one of the important criteria in developing novel system. To achieve this aim, the sensor fusion technique and the Interacting randomized algorithm are applied. Since the Kalman filter based on a single state space model has a defect in the case that a target changes its mode, the proposed algorithm using more than two different models is inevitable. Even though sensor fusion and proposed algorithm are totally different techniques, these can cooperate to provide the optimal estimates. By comparing the simulation using a single sensor and the proposed algorithm. The error reduction is greater when three sensor data are fused. The advantage in using the proposed algorithm is not only error reduction but also mode prediction.

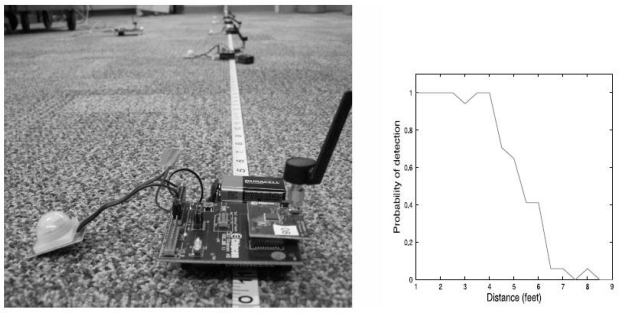


Fig. 5. Experimental setup and sensor characterization: The figure on the left shows the experimental setup, with the sensor modules placed uniformly along a line. The plot on the right shows the probability of target detection versus distance for a particular sensor module.

The introduced algorithm which is real-time distributed target tracking scheme without time synchronization for both the ideal and imperfect binary sensing models energy efficient and fault tolerant. In simulation phase, the accuracy of new algorithm tested under ideal model and outcome showed great precision. The analysis also demonstrated that for the assumption simulated, the application of sensors that hardly sense the target by the algorithm served to enhance the accuracy of localization minimized the estimation error for 50% in comparison with utilizing only sensors that do sense the target so significantly. Results of comprehensive simulations of this algorithm performed under different conditions and scenarios also verified that the presented algorithm overcomes other algorithms in terms of its accuracy of measuring the target location, velocity and trajectory applying the binary sensor networks.

REFERENCES

[1] Jaspreet Singh, Rajesh Kumar, et al., Tracking Multiple Targets Using Binary Proximity Sensors, IPSN'07, Cambridge, Massachusetts, USA.

[2] Jaehwi Jang, Master Thesis on Multiple-Sensor Fusion for Single Target Tracking Using Interacting Multiple Model (IMM) Algorithm, Seoul National University, Korea, 2001.

[3] J. H. Reed, K. J. Krizman, B. D. Woerner, and T. S. Rappaport, "An overview of the challenges and progress in meeting the E-911 requirement for location service," *IEEE Commun. Magazine*, pp. 30–37, Apr. 1998.

[4] J.Werb and C. Lanzl, "Designing a positioning system for finding things and people indoors," *IEEE Spectrum*, vol. 35, pp. 71–78, Sept. 1998.

[5] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *IEEE Pers. Commun.*, vol. 4, pp. 42–47, Oct. 1997.

[6] J. M. Rabaey, M. J. Ammer, J. L. da Silva, Jr., D. Patel, and S. Roundy, "Picorodio supports *ad hoc* ultra-low power wireless networking," *IEEE Comput.*, vol. 33, pp. 42–48, July 2000.

[7] R. Fleming and C. Kushner, "Low-power, miniature, distributed position location and communication devices using ultra-wideband, nonsinusoidal communication technology," Aetherwire Inc., Semi-Annual Tech. Rep., ARPA Contract J-FBI-94-058, July 1995.

[8] R. L. Moses, D. Krishnamurthy, and R. Patterson, "An auto-calibration method for unattended ground sensors," in *Proc. ICASSP*, vol. 3, May 2002, pp. 2941–2944.

[9] D. D. McCrady, L. Doyle, H. Forstrom, T. Dempsy, and M. Martorana, "Mobile ranging with low accuracy clocks," *IEEE Trans. Microwave Theory Tech.,* vol. 48, pp. 951–957, June 2000.

[10] A. Savvides, H. Park, and M. B. Srivastava, "The bits and flops of the n-hop multilateration primitive for node localization problems," in *Proc . Int. Workshop Sensor Nets. Appl.,* Sept. 2002, pp. 112–121.

[11] S. ˇ Capkun, M. Hamdi, and J.-P. Hubaux, "GPS-free positioning in mobile *ad-hoc* networks," in *Proc. 34th IEEE Hawaii Int. Conf. Syst. Sci.,* Jan. 2001.

[12] J. Albowicz, A. Chen, and L. Zhang, "Recursive position estimation in sensor networks," in *Proc. IEEE Int. Conf. Network Protocols,* Nov. 2001, pp. 35–41.

[13] C. Savarese, J. M. Rabaey, and J. Beutel, "Locationing in distributed *ad-hoc* wireless sensor networks," in *Proc. ICASSP,* May 2001, pp. 2037–2040.

[14] R. Nagpal, H. Shrobe, and J. Bachrach, "Organizing a global coordinate system from local information on an *ad hoc* sensor network," in *Proc. 2nd Int. Workshop Inform. Proc. Sensor Networks,* Apr. 2003.

[15] L. Doherty,K. S. J. pister, and L. E. Ghaoui, "Convex position estimation in wireless sensor networks," in *Proc. IEEE INFOCOM,* vol. 3, 2001, pp. 1655–1663.

[16] P.-C. Chen, "A nonline-of-sight error mitigation algorithm in location estimation," in *Proc. IEEE Wireless Commun. Networking Conf.,* Sept. 1999, pp. 316–320.

[17] M. A. Spirito, "On the accuracy of cellular mobile station location estimation," *IEEE Trans. Veh. Technol.,* vol. 50, pp. 674–685, May 2001.

Study on the Smart Application for Wireless Sensor Networks based Ubiquitous Livestock Farm System

Jeonghwan Hwang[1], **Hyun Yoe**[2]
[12]Department of Information and Communiaction Engineeringe, Sunchon National University, Suncheon, Jeollanam-do, Republic of Korea

Abstract - *Smart phone and its applications are currently bringing about significant changes in our lives and it is expected that applying such technology in the area of agriculture could increase the value added and productivity of agriculture with its various uses. This paper proposed smart application for monitoring and managing livestock farm in real-time anytime, anywhere. In the ubiquitous livestock farm system, wireless sensor networks environment sensor and CCTV are installed at livestock farm for collecting and monitoring livestock farm environment and video information on livestock breeding such as illumination, humidity, temperature and gas. The livestock farm environment and video information collected in such way can be used by user to monitor the livestock farm in real-time through the use of smart application of smart phone to control the livestock farm facilities anytime, anywhere. This smart application can provide user convenience and increase productivity by allowing users to control their livestock farm facilities.*

Keywords: Wireless Sensor Networks, Livestock farm, Smart Application, Ubiqiutous

1 Introduction

Mobile phone that had been primarily be used to make and receive calls has changed into application-centered mobile Internet device, which led to the paradigm shift in the mobile telecommunications industry[1][2].

Accordingly, ensuring high-quality applications is becoming a new element of competitiveness. Applications are the core element of smart phone and the applications for personal entertainment and everyday life purposes are forming the majority thus far. However, applications for business purposes are expected to be developed according to the advancement and development of devices and mobile standardization technology [3][4].

The percentage of smart phone users in Korea has surpassed 40% currently [5], and the introduction of smart phone is bringing about not only significant changes in the IT industry as well as our everyday lives but also many studies on various usages of smart phone[6][7].

Applying smart phone and its applications in the area of agriculture that is labor intensive where the application of IT technology has been relatively lacking can increase the value added and productivity of farming as it can be utilized in various ways.

It has become inevitable for the livestock industry of Korean agricultural industries, in particular, to face direct competition with the livestock industries of advanced nations due to the recent increase in feed price and the signing of FTA(Free Trade Agreement). In addition, many livestock breeders are experiencing difficulties from the increase in production cost such as fee, raw & subsidiary material and energy costs, in addition to the increase their death rate from various diseases [8][9].

Accordingly, the purpose of this paper is to proposed a smart application for monitoring and controlling the livestock farm environment and facilities by using smart phone in order to solve various issues the Korean livestock industry is currently experiencing.

In the WSN(Wireless Sensor Networks) based ubiquitous livestock farm system, WSN environment sensor and CCTV that collect information on livestock breeding environment such as the illumination, humidity, temperature and CO2 inside/outside livestock farm are installed to collect and monitor livestock farm environment and video information.

The livestock farm environment and video information collected in such way can be used by the user to monitor the livestock farm anytime, anywhere through the use of smart application of smart phone and the user is able to control the livestock farm facilities, thereby providing user convenience and resulting in productivity increase.

The composition of this paper is as follows. Chapter 2 explains the structure of the WSN based ubiquitous livestock farm system for smart application, as well as the process of service provided. Chapter 3 explains the implementation and operation of the proposed smart application. Lastly, Chapter 4 concludes the paper through conclusion

[2] Corresponding Author

2 Design of the Ubiquitous Livestock Farm System

2.1 System Structure

As shown in the Figure 1, the WSN based ubiquitous livestock farm system consists of physical layer that consists of sensor, CCTV and livestock farm facilities; and the middle layer that maintains the optimal livestock breeding condition by supporting the communication between the physical layer and the application layer and converting the livestock farm information into database to provide monitoring and control services; and the application layer in which various interfaces that support the livestock farm environment monitoring and facility control services exist

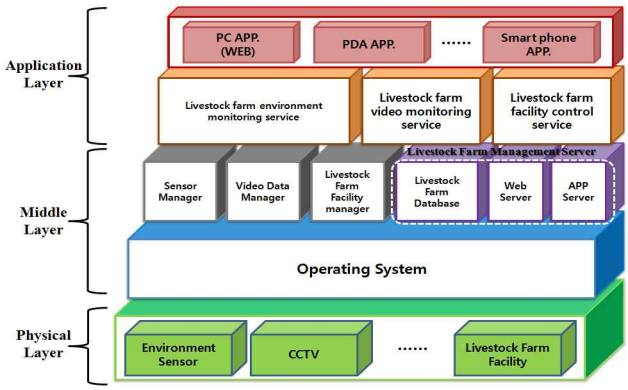


Figure 1. Ubiquitous Livestock Farm System Structure

The physical layer consists of environmental sensors for collecting livestock farm environment data, CCTVs for collecting livestock farm video data and livestock farm facilities for creating optimal livestock breeding environment. The environmental sensors are installed in livestock farm to collect livestock breeding environment data such as illumination, humidity, temperature and gas, and the sensor nodes, upon autonomously creating a network, wirelessly collects physical data obtained from the sensor nodes and measure environment changes. CCTVs are installed both inside and outside of livestock farm and the indoor CCTVs are installed to collect livestock farm and livestock video data and the outdoor CCTVs are installed to prevent theft, fire, etc. The livestock farm facilities refer to the devices that control the environmental elements that affect livestock growth such as illumination, temperature, humidity and gas, and they include lighting device, humidifier, air conditioner, ventilation fan, etc.

The middle layer consists of sensor manager for managing environment data collected from the physical layer sensors, video data manager for managing video data collected from CCTVs, livestock farm facility manager for managing livestock farm facilities, livestock farm database that stores livestock farm data and livestock farm management server for monitoring livestock farm and controlling its facilities. The sensor manager stores into livestock farm database the livestock farm environment data collected from the physical layer environmental sensors by formatting them into storable format, converting them into units according to measurement elements and using update inquiry for the processed data. The livestock farm facility manager, upon receiving control signal, operates/manages livestock farm facilities, and plays the role of storing the status of livestock farm facilities into livestock farm database. The video data manager provides stream data to the web. The livestock farm database stores in respective tables the livestock farm environment data such as illumination, temperature, humidity and gas collected from the sensors installed inside and outside of livestock farm, in addition to storing the livestock farm video data collected from CCTVs, livestock farm facility status and control data and the environment standard values for auto-control and status notification. The livestock farm management server is located between the user and the livestock farm database, and periodically tests and notifies the user the environment data stored in the livestock farm database, and compares them with the environment standard values stores in the livestock farm facility control table to control the facilities.

The application layer consists of applicant services that support various platforms such as laptop, web, PDA and smart phone, and it provides users with livestock farm environment monitoring service, livestock farm video monitoring service and livestock farm facility control service.

2.2 Service Process

The WSN based ubiquitous livestock farm system provides with 'livestock farm environment monitoring service', enabling the observation of internal/external environmental information of livestock farm, 'livestock farm video monitoring service', providing with livestock farm video in real time, 'livestock farm facilities control service', enabling the automatic control and manual control of livestock farm facilities by producer based on the environmental standard values, and 'danger alarm service', giving notification of dangerous situation at the livestock farm.

The livestock farm environment monitoring service shows the livestock farm environmental data, collected at the environmental sensors measuring the environmental elements, such as luminosity, temperature, humidity and CO2, to producer through GUI so that producers can identify the environment changes of internal and external of the livestock farm. Figure 2 shows the operation process of livestock farm environment monitoring service.

The detail of this service is that it collects livestock farm internal/external environmental information giving impacts to livestock growth such as luminosity, temperature, humidity and CO2 from the environmental sensors installed at inside/outside of livestock farm and transmits the information to sensor manager periodically. The sensor manager will

analyze the received data and extract each sensing value. Their formats will be changed and they will be saved in each table of livestock farm database. The livestock farm management server transmits livestock farm internal/external environmental information saved in the livestock farm database to producer and the producer can monitor the environmental information of livestock farm through this information.

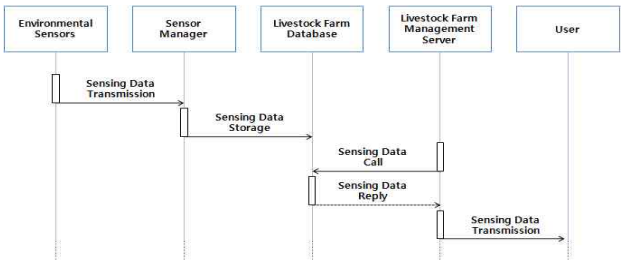


Figure 2. Environment Monitoring Service Process

The livestock farm video monitoring service provides producer/consumer with video of livestock farm /livestock-individuals through CCTV installed in the livestock farm. The CCTV sends the livestock farm video to video data manager and the video data manager provides with this information by web through Internet. Users can confirm the livestock farm video information through Internet. Figure 3 shows the operation process of livestock farm video monitoring service.

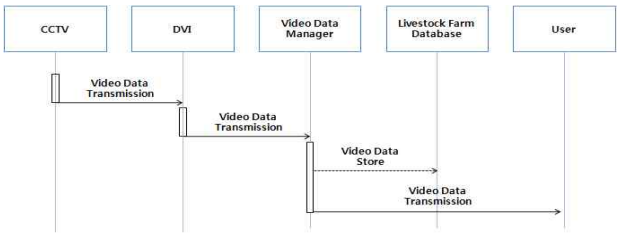


Figure 3. Video Monitoring Service Process

The livestock farm facility control service enables the livestock farm management server automatically control the livestock farm facilities, or, the producer manually control the livestock farm facilities based on the collected information at the CCTV and environmental sensors installed at inside/outside of livestock farm.

Figure 4 shows the operation process of livestock farm facilities automatic control service. The automatic control service saves the information collected from livestock farm at livestock farm database. The livestock farm management server calls up the information and compares it with the environmental standard values saved in the livestock farm use database. If it is more than or short of standard value, it will confirm whether the livestock farm facilities are operating as saved in the livestock farm database. Then it will send the control signal to livestock farm facility manager and control the livestock farm facilities. When livestock farm facilities operate, the livestock farm facilities status information is saved in the livestock farm database and it will be notified to user.

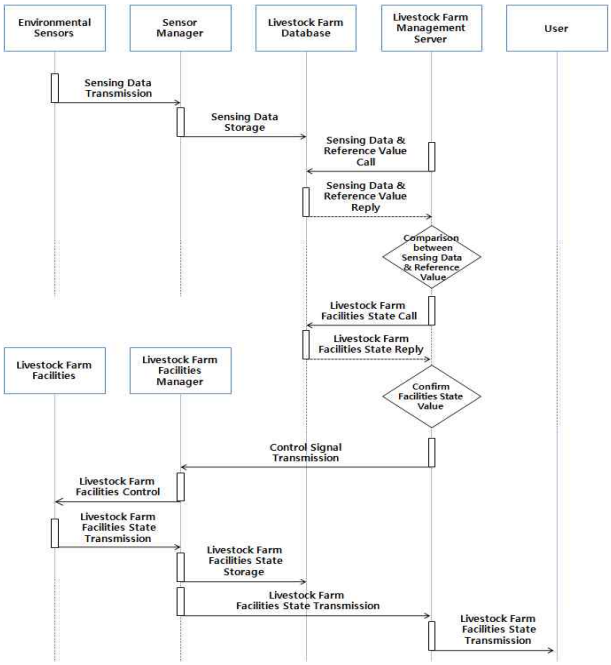


Figure 4. Facilities Automatic Control Service Process

Figure 5 shows the operation process of livestock farm facilities manual control service.

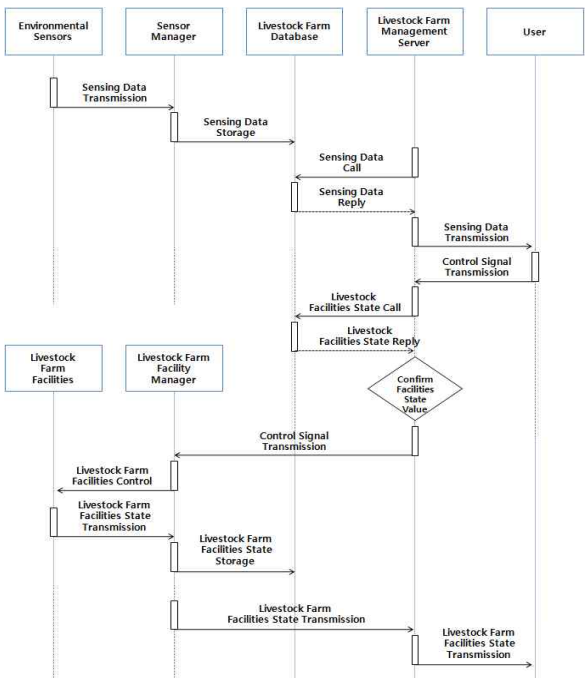


Figure 5. Facilities Manual Control Service Process

The manual control service saves the information collected from livestock farm in the livestock farm database and the livestock farm management server sends the information to the user in real time. If the user wants to control the livestock farm at this time, the user will send the livestock farm facilities control signal to livestock farm management server through GUI. The livestock farm management server will check whether the livestock farm facilities are operating through livestock farm database and send the control signal to livestock farm facilities manager to control the livestock farm facilities.

The danger alarm service tells the weather change and livestock farm status change to farmers in real time and takes emergency measure to prevent danger in advance. The data sensed at the environmental sensor is sent to the sensor manager. The sensor manager extracts the sensing values from received data and saves them in the livestock farm database. The saved sensing values will be periodically monitored by livestock farm management server. If it would be more than or less than the standard value, it will be notified to the element where the event had occurred. Figure shows the operation process of danger alarm service.

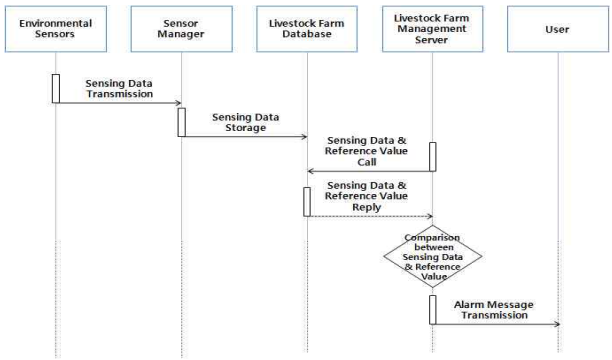


Figure 6. Danger Alarm Service Process

3 Implementation and Result

3.1 System Test-bed

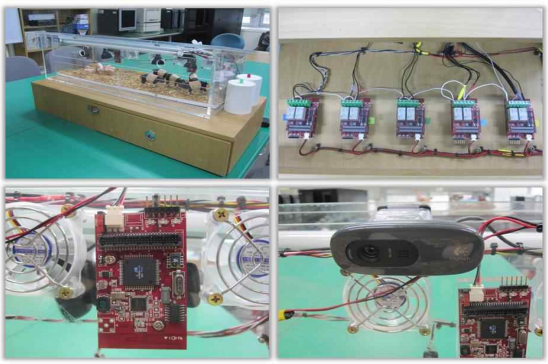


Figure 7. Ubiquitous Livestock Farm System Model

For the purpose of verifying the smart application, a virtual livestock farm model was produced in which environment sensor and web cam were installed for the livestock farm environment and video monitoring of the livestock farm.

In addition, environmental control facilities such as light, heating fan and ventilation fan were installed in the livestock farm to maintain optimal livestock breeding environment, and for the purpose of system control and monitoring test, ubiquitous livestock farm system application was implemented as shown in the Figure 8.

Figure 8. Web GUI (Test-bed)

The proposed smart application was implemented and operation in the virtual livestock farm and the result showed that applying the proposed smart application in an actual livestock farm would also yield the same result of being able to maintain optimal livestock breeding environment.

Figure 9. iOS based Smart Application (Testbed)

3.2 System Field Application

The effectiveness of the proposed smart application was confirmed through the result of test conducted by applying the system to a livestock farm model, and the system was applied and implemented at actual livestock farm.

The environmental sensor was installed, as shown in the Figure 7, in order to collect environment information such as the livestock farm temperature, humidity, illumination and CO2, and the installed sensor nodes transmit the livestock farm environment data through the WSN sensor gateway inside the livestock farm.

Figure 10. Installation Site

In addition, CCTV was installed, as shown in the Figure 8, for the purpose of 24-hour video surveillance of inside the livestock farm. The camera is used to identify the causes of any thefts or accidents or to check the current livestock farm condition in real-time by surveying and recording the inside of livestock farm 24 hours. The captured video is transmitted to the livestock farm management server and stored in the database upon being classified according to the livestock farm ID and camera number.

Figure 11. Environmental Sensor and CO2 Sensor

The livestock farm facilities that include lighting, humidifier, heating fan, air conditioning unit and ventilation fan were installed to control the livestock farm environment that could affect the livestock growth such as illumination, temperature, humidity and CO2, and according environment control devices were installed at the livestock farm as shown in the Figure. These are used to control the environment control devices to create an ideal livestock breeding environment inside the livestock farm.

Figure 12. DVI and CCTV

The livestock farm facilities that include lighting, humidifier, heating fan, air conditioning unit and ventilation fan were installed to control the livestock farm environment that could affect the livestock growth such as illumination, temperature, humidity and CO2, and according environment control devices were installed at the livestock farm as shown in the Figure. These are used to control the environment control devices to create an ideal livestock breeding environment inside the livestock farm.

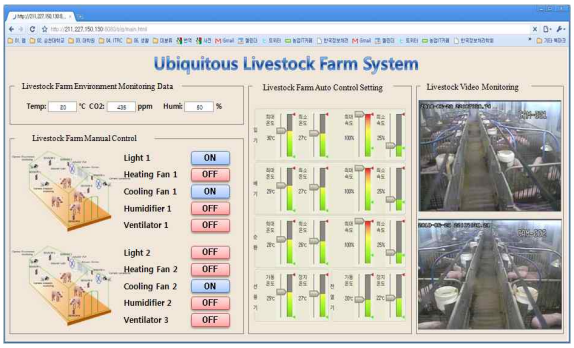


Figure 13. Ubiquitous Livestock Farm System Web GUI

3.3 Smart Application

Figure 14. iOS based Smart Application (iPad2)

As for the iOS-based smart application development system, Xcode 4.1.x IDE that operates based on MAC OS X 10.7.x was used to develop an application for the latest update of iOS 5.x. The Figure 14 shows the developed iOS-based

smart application operation, and (a) indicates the sensing value sensed through the sensors installed inside/outside the stall, and (b) controls or shows the livestock farm facility equipment. In addition, (c) is the video collected through CCTV and for controlling the CCTV, and (d) is for entering the standard value for the automatic control of livestock farm.

The application development environment for developing Android OS-based smart application, it operates in JDK 1.6 version of Window XP Service Pack3 OS, and Eclipse 3.6 (Helios) is used as the basic tool for Android development. As for the Android OS, the most commonly used version Android SDK 3.0 (Honeycomb) was used. The Figure is an operation screen of the developed Android OS-based smart application and it has the same function as iOS-based smart application.

Figure 15. Android OS based Smart Application (Galaxy Tab)

The smart application developed, as described above, was applied to an actual livestock farm, and it was able to collect the livestock farm environment and video information through the sensor and video surveillance camera and constantly monitor and control the livestock farm condition through its user intuitive GUI.

4 Conclusions

For the purpose of ensuring systematic and scientific livestock farming technology, this paper proposed smart application for wireless sensor networks based ubiquitous livestock farm system.

The WSN based ubiquitous livestock farm system for smart application consists of the physical layer for collecting pig farm information as well as environment control, and the application layer that consists of interface that supports service, and the middle layer that maintains livestock breeding environment in optimal condition by supporting the communication between the physical and application layers, converting the livestock and pig farm information into database and providing monitoring and control services.

For the purpose of verifying the proposed smart application, environment sensor and CCTV were installed at the livestock farm model to test the proposed smart application, and the proposed smart application was implemented at an actual livestock farm. The implement result showed that it was able to constantly monitor and control the livestock breeding environment of the livestock farm by using smart application, through which it was able to enhance livestock production efficiency and productivity and provide user conveniences.

Furthermore, it is expected to contribute to the reduction of labor force and the production of high-quality stock farm products, as well as ensure the competitiveness of the livestock industry.

5 Acknowledgment

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation

6 References

[1] A. Nusca, “Smartphone vs. Feature Phone Arms Race Heats up; Which Did you Buy?”, ZDNet, Aug. 2009.

[2] S Verstockt, D Decoo, D Van Nieuwenhuyse, F De Pauw, R Van De Walle, “Assistive Smartphone for People with Special Needs ; The Personal Social Assistanct”, 2009 2nd Conference on Human System Interactions, May. 2009.

[3] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, “Diversity in Smartphone Usage”, Proceedings of the 8th International conference on Mobile systems applications and services MobiSys 10, June. 2010.

[4] M. H Goadrich and M. P. Roger, “Smart Smartphone Development; iOS Versus Android”, Proceedings of the 42nd ACM technical symposium on Computer science education, March. 2011.

[5] Junbong Beak, Beomsuk Hong, Myoungho Choi, “Smart Phone 10 million era, Beginning of Mobile Business Bigbang”, KT Economic Management Institute, Digieco Reports, 8. April. 2011.

[6] D.H. Kim, C. Ryu, J.H. Lee, S.J. Kim, “Mobile Software Platform Trends for Smartphone”, Electronic Communications Trend Analysis, Korea Electronics and Telecommunications Research Institute, Vol. 25, No. 3, 2010.

[7] Young-Bong Kim, “Presence and Future of the SmartPhone and Security Apps”, Journal of Korea Contents Association, Vol. 8, No. 3, p42-p74, 2010.

[8] Jeong-hwan Lee, “[Livestock industry research series 11] What is a threat to South Korea's livestock industry?”, Focus attention GSnJ No.55, 2008.

[9] Yong-hl Yoo, Doo-hwan Kim, “The current state of automation in pig house establishment and prospection”, Korea society for livestock housing and environment, p29-p47(19), 2006.

Time Synchronization in Wireless Sensor Network: Techniques, Issues and Recommendations

Mubashir Saeed, Syed Osama Hassan, Sohail Jabbar
Bahria University Wireless Research Center, Bahria University Islamabad, Pakistan.

ABSTRACT

Wireless sensor networks comprise of multiple nodes of sensors that are organized and distributed in such a manner that they form a corporate network. Due to recent technology advancements, WSNs have become more feasible to be manufactured economically and technically and are widely being deployed for computation, communication and sensing purposes. The potential requirement of WSNs in worldwide applications is attracting their diverse usage hence have become most demanding in the last decade. Since WSNs comprise of nodes with limited energy, these nodes keep on operating until they are exhausted. To get the desired services précised and accurate, clock synchronization between the nodes is a fundamental building block to these sensor networks. This paper contains about all the issues that are the reasons to fluctuate the time of the network, enables the reader to grasp the concept of fluctuations, the drift and the offset and the classes of synchronization. Moreover few techniques are also suggested in the paper to synchronize the whole network.

KEWORDS: Wireless Sensor Network, load balancing, cluster routing, energy efficient routing protocol, static network, multi-hop routing

1- INTRODUCTION

Time synchronization means that all the sensor nodes of the network have a common notion of time. To achieve the goal of successful and efficient communication, it is necessary that the entire sensor network should be synchronized. Providing a common reference clock to the network is really challenging to achieve. Most of the sensor networks employ a sleep-clock with a relatively low frequency, to enable/disable a sensor for saving energy. All other clocks are termed as the individual local clocks for each node. Even if synchronized earlier, these local clocks drift over time due to a number of reasons.

$$\text{Clock synchronization} = \text{Synch of Clock drift} + \text{Synch of Clock offset}$$

The difference of frequencies of clock tapping is termed as clock drift and the difference of time between two clocks is said to be the clock offset.

There are multiple reasons for the clock fluctuations. Clock fluctuates as the time passes due to the available batteries, temperature and such other similar physical variations. This causes the change in supply voltage. Time synchronization is the issue of MAC layer. Media Access Control layer is used to provide addressing and control mechanisms for channel access enabling the several network nodes to communicate. The control mechanism provided by the MAC layer is said to be MAP (Multiple Access Protocol). Contention based technique CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) is the most common protocol and is widely being used. This technique works as; First of all, the whole traffic channel is checked. If it is found busy, then waited and tried later, if found free, then transmitted.

Ideal clock rate of the network is equal to 1. This rate varies when error occurs in the synchronized time of the nodes. So these nodes are required to be synchronized by some synchronization technique. In order to guarantee the accuracy of the executed synchronization, it must be maintained from following two aspects.

1. Offset Compensation
2. Drift Compensation

There are many protocols used to remove the error and variation of time in a network and synchronize them. Some protocols can only use offset compensation to synchronize the time. Most of the synchronization schemes perform drift compensation based on a short-term linear clock model. However, from a long-term perspective, an unwanted perturbation may occur in the clock model. To perform the offset compensation, a synchronization scheme should work based on an accurate estimation of the offset between sensor nodes via local time message exchanges. To achieve ideal synchronization, accuracy of the estimated offset is really critical. The existing synchronization techniques solve this problem by estimating and then by removing the delays in the process of transmission of the message.

2. TIME SYNCHRONIZATION IN WIRELESS SENSOR NETWORK

Lots of reasons have been sorted out in the available literature that explores the issues of time synchronization in sensor network. Upcoming paragraphs discuss this all in detail.

2.1 Causes of Time Synchronization Error

Non-determinism is the main disease of the precise network synchronization. Moreover, the wireless network uncertainties of message delay are larger as compared to wired networks. In order to understand the concept behind the sources of these errors, it is a better approach to fragmentize and segment the latency source of a message. After decomposing, we come up with the following ten times consumed during information communication.

1. Send time
2. Access time
3. Transmission Time
4. Propagation time
5. Reception time
6. Receive Time
7. Interrupt handling time
8. Encoding time
9. Decoding time
10. Byte alignment time

2.2 Factors Affecting Time Synchronization in WSN

Some other factors which influence the time synchronization in WSN are as follows;

1. Temperature
2. Phase noise
3. Frequency noise
4. Asymmetric Delay
5. Clock glitches

Sensor nodes should be sensitive to energy requirements instead of handling these factors.

2.3 Time synchronization schemes evaluation criteria

A nearly ideal technique fulfills all the enlisted evaluation criteria and meets the desired application requirements efficiently. These are;

1. Energy Efficiency
2. Scalability
3. Precision
4. Robustness
5. Lifetime
6. Scope
7. Cost and Size
8. Immediacy

2.4 Clock Models

Time intervals are measured by digital clocks which are consisted of timers '*h'* (also referred to as local clocks). These are used to counts the steps as a reading *h(t)* at the real time *t.* The increment to the counter is done by an oscillator of frequency *f.* At real time *t,* the rate *f* is given by the first derivative of the reading *h(t)* as *h(t): f(t)=dh(t)/dt.*

If the earlier mentioned fluctuations caused by 2.2 (also the Table 1) are arbitrary, the information from the clock reading is null. The fluctuations are at different rates and each rate leads to its own clock model. Available clock models (Figure 1) are classified into three categories and are explained below.

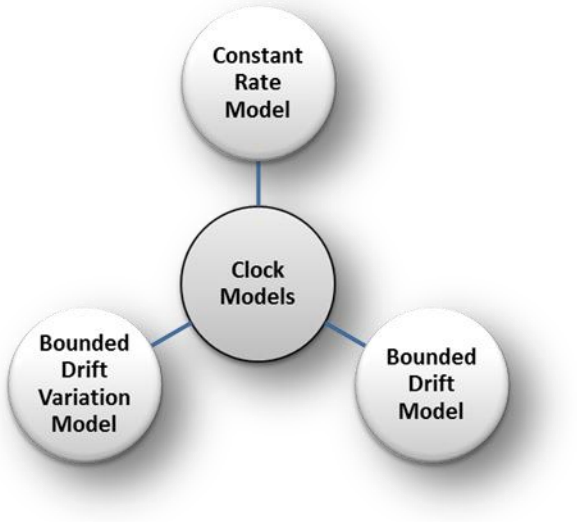


Figure 1: Clock Models

2.4.1 Constant Rate Model

In constant rate model, fluctuation rate is supposed to be constant. If the precision that is required is lesser than the rate fluctuation, this model is quite feasible.

2.4.2 Bounded Drift Model

Since the rate of the deviation from the standard rate 1 is supposed to be bounded, we can say this clock drift as $\rho(t)= f(t)-1=dh(t)/dt-1$, and denote the corresponding bound with $-\rho max \le \rho(t) \le \rho max$ *for all t.* The reasonable assumption is $\rho i(t)>-1$ for all the times t. This means that the running backward ($\rho i(t) < -1$) of clock and its stopping ($\rho i(t) = -1$) is not possible. So if the events *a*, *b* with *ta* < *tb* are occurred at the node *Ni*, whose clock drift ρi is bounded according the previous equation that is $-\rho max \le \rho(t) \le \rho max$ *for all t,* then the *Ni* node can calculate the lower and upper bounds $\Delta li[a,b]$, $\Delta ui\ [a,b]$ on the difference of the real time $\Delta[a,b] := tb-ta$ as the following.

$$\Delta^{l}_{i}\,[a,b] := \frac{h^{i}(tb)-h^{i}(ta)}{1+\rho max}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$

$$\Delta^{u}_{i}[a,b] := \frac{h^{i}(tb) - h^{i}(ta)}{1-\rho max} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Since bounds on the oscillator's rate are given by the vendor of the hardware, this model is sufficiently reasonable. There are usually non expensive oscillators on the nodes of the sensors so we can have ρmax belonging to [10 ppm, 100 ppm]. Note that in this model, the drift can jump arbitrarily within the bounds specified in $-\rho max \leq \rho(t) \leq \rho max$ for all t. Variation of the drift is limited by the upcoming model described below.

2.4.3 Bounded Drift Variation Model

In this model, the variation $\vartheta(t) = d\rho(t)/dt$ of the drift of the clock is supposed to be bounded.

$$-\vartheta max \leq \vartheta(t) \leq \vartheta max \quad \textit{for all } t$$

The assumptions that are made in bounded drift variation model are feasible only if this drift is being influenced by slowly varying conditions like the voltage of battery or the temperature. This implies to make the drift compensation possible. The nodes can calculate the bounds of its drift for the time of future and can also estimate its current drift.

2.5 Classes of Synchronization

Typical classes of time synchronization are depicted in (Figure 2) and are explained in detail below.

2.5.1 Internal or External

Internal synchronization requires no pre-determined master time and all nodes of the network are synchronized without any external source of time.

Whereas with respect to an external provided time, all the clocks of network are synchronized. This is called

Table 1: Various Reasons Due To Which Delays Occur and Hence Synchronization Errors Are Produced

Time Consumed	Deterministic	Magnitude	Reasons	Dependence	Taken by
Send time	X	0-100ms	Time for the generation of packet and its delivery to MAC layer.	Loads on processor	Sender
Access time	X	10-500ms	Packet wants to get access to the wireless channel.	Access to wireless link	Sender
Transmission time	√	10-20ms	Time to transmit packet bit by bit over the physical channel.	Baud rate and packet length	Sender
Propagation time	√	For the distances from 50-350 meters, <=1microsec	Time taken by the packet to travel from source to the receiver.	Distance b/w the Sx and Rx	Packet for propagation
Reception time	√	10-20ms	It is a delay to receive packet.	Baud rate, packet length and congestion	Receiver
Receive time	√	0-100ms	Time to process to process the incoming data packet.	Speed of processing and the variable data load	Receiver
Interrupt handling time	X	5microsec-30microsec	Time taken by the microcontroller to handle and execute an interrupt	Enabling and disabling of the interrupts and the execution of interrupt service routine	Both the sender and receiver
Encoding time	√	100-20microsec	Time required to convert the message into an EM wave	On the encoding techniques and the packet length	Sender
Decoding time	√	100-200microsec	Time taken for converting the EM wave to a message	On the decoding technique and the length of the packet	Receiver
Byte alignment time	√	0-400microsec	Time used to align the bytes for synchronization	Depends upon the radio speeds and the offsets of the bits	Both the sender and receiver

external synchronization and is done by taking the NTP as a reference. If the reference time is the time of an own network node, it creates no difference. Consistency within the network and outside provided time is required for external synchronization.
In our daily life, we usually face external synchronization while using wrist watches, cellular phones, computers and such other daily use home appliances. We always synchronize these times to a legal reference time to work proficiently.

2.5.2 Continuous or On-Demand

If the time of the entire network is always maintained, then the network is using continuous synchronization process. As a precautionary measure, network is continuously being checked to eradicate the variations.

On-Demand Synchronization:

1. Event Triggered on-demand synchronization
2. Time Triggered on-demand synchronization

Event triggered on-demand synchronization is required to be executed only after the event has occurred.
Time triggered on-demand synchronization is required to be executed if we need data observed by multiple sensors for specific time duration.

2.5.3 Extent or Ambit: All nodes or Subnets

Extent or Scope or Ambit of synchronization tells us about the nodes of the network that are unsynchronized or are required to be synchronized. All nodes or Subnet synchronization depends on the type of application being used. Event-Triggered time synchronization can be bounded to the collocated subset of nodes because they are the observers of event in question.

2.5.4 Rate Synchronization or Offset Synchronization

When the nodes measure identical time interval lengths, they are said to possess rate synchronization. If a mobile object crosses a network, the time it takes to cross the network can only be measured by rate synchronization. It is calculated by observing the appearing and disappearing time of the object.
To combine time stamps from different nodes, we use offset synchronization. In offset synchronization, the nodes of the network measure identical points in the time, say at time T. That is why all the software clocks display the time to be T when we look at their scope.

2.5.5 Time Scale Transformation or Clock Synchronization

To synchronize the network, we can basically do two fundamental tasks. One is to make all the clocks to show the common time at any instant, by synchronizing them. Other is to transform the scales of time that is to convert the local time of any node to local time of any other node.
Aforementioned both methods have their own advantages and a few drawbacks in either one but any how the data collected by using any of the respective methods may look like to be grasped from a single node. If we use clock synchronization, global flooding of messages is required and coordination is anticipated through the entire network, while in the transformation of time scales, the communication across the network is not required but only neighbor needs to bother.

2.5.6 Time Instant or Time Intervals

The information regarding time can be given by using either of the two; time instant or the time interval. The time instant T can be T=6 or any values and time interval T_i can be between two time instants like T_1= [5.5, 6.5].

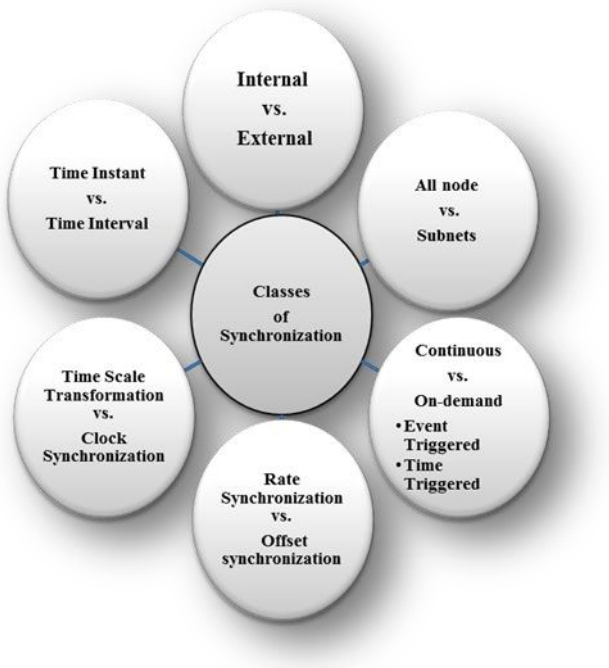


Figure 2: Different classes of synchronization and their sub types (if any)

In both approaches, by including the quality we can improve our required information of time. The use of time intervals with a guarantee of being correct is very useful for the sensor networks. Though this method is not being used widely but has a number of advantages over time instants. These are;

i. The bounds on the individual sensor local time with a guarantee of being correct allows to get the bounds with a guarantee of being correct from the data fusion of the sensors about any occurred event.
ii. Sensing, actuating and communicating become successful for several nodes if the time is re-determined.
iii. To get optimal and unambiguous local time, the bounds are combined for this local time and are made single.

2.6 Synchronization Techniques

Synchronization techniques are not to be mixed with the synchronization protocols. These techniques are the main methods followed by the respective protocols. Five different techniques are discussed below and are also depicted in (Figure 4).

2.6.1 Taking one sample

Consider two nodes N_s and N_r wants to communicate. Synchronizing the nodes means to establish some relationship between the local clocks of the nodes C_r and C_s.

2.6.1.1 Unidirectional Synchronization

Suppose N_s wants to communicate with N_r containing the local time stamp C_s^a. The node N_r receives the message at its local time C_r^b. The node N_r cannot determine the delay with which message is transferred. When N_s sends the message, the local time is $C_s^a > C_r^b$ and when the message is received the local time is $C_r^b > C_s^a$.
If we already know the former values of the delay (d) then $d_{min} \leq d \leq d_{max}$ and the estimation is given by
$(C_s^a \approx C_r^b - \frac{1}{2}(d_{min} + d_{max}))$ which reduces the worst-synchronization error. (Figure 3)

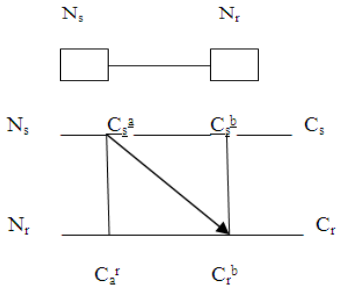


Figure 3: Uni-Directional synchronization

2.6.1.2 Round-trip synchronization

In this technique the node N_r inquires for time stamp C_s^b by sending a message to node N_s. The node N_r measures the round-trip time $D = C_r^c - C_r^a$, so by this process we can find the end to end distance of the time space between sending a query and getting a reply.
Similarly, without knowing the former values of delay the receiver node knows that it bounds between 0 and D. The estimation $C_r^c \approx C_r^b - \frac{D}{2}$ reduces the worst-synchronization error. (Figure 5)
The round-trip synchronization mechanism is also known as probabilistic time synchronization and this procedure reduces synchronization error given as follows; after getting the reply, the node N_r looks for the worst-case error synchronization $\frac{D}{2} - d_{min}$ which is below the threshold. If it is not below the threshold, it sends the new query to N_s until and unless both sending and receiving gets the desired synchronization error. More messages will be sent between the nodes if the threshold is small and this is the main disadvantage of this technique.

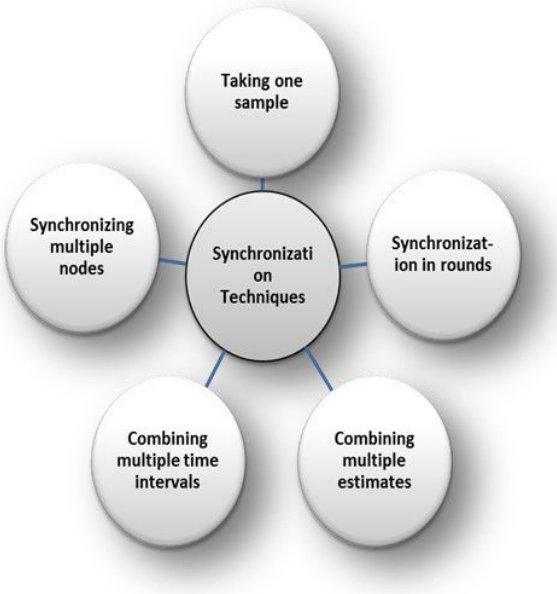


Figure 4: Synchronization Techniques

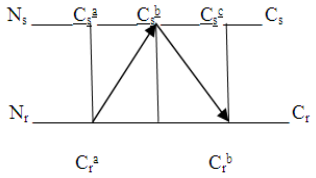


Figure 5: Round trip synchronization

2.6.1.3 Reference Broadcasting

In this approach we add another node N_i which can also be known as beacon node (Figure 6). In this technique the node N_i transmit a message to N_r, the delay d (to N_r) and d'(to N_s) are more or less equal. The node N_s sends a message to N_r with time stamp and so that N_r can measure the time interval $D = C_r^b - C_r^a$ and also can estimate $C_s^b \approx C_s^a + D$.

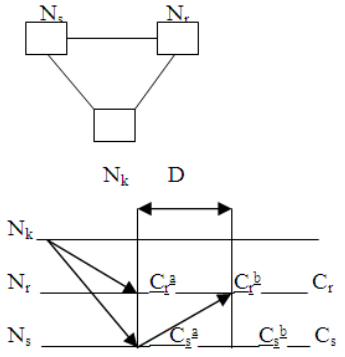


Figure 6: Synchronization using reference broadcasting

This approach became famous in the synchronization world of WSN, as RBS. The main advantage of RBS is that a broadcast message is received at the same time and thus synchronization error is smaller. The main disadvantage of reference broadcasting is that it requires a beacon node and physical broadcasts.

2.6.2 Synchronization in rounds

In a network of only two nodes, there is a need of refreshing the clocks periodically because clocks run at different speed. The length of a round-trip depends upon error budget and the quantity of relative drifts involving both clocks. The length of round-trip is denoted by τ_{round} and assuming that the length of round-trip consists of first period denoted as τ_{sample}. In the first period it takes either one or more samples from the techniques described above and in the second period no action is made by the nodes. The maximum interval of a round-trip τ_{round} has to satisfy the following relation.

$$\tau_{round} \leq \frac{Etotal - Esample}{\hat{p}\ maximum}$$

Where, E_{total} = total error, E_{sample} = error after taking samples and $\mathcal{P}_{maximum}$ = maximal drift rate.

The above relation shows that if E_{sample} and $\hat{p}_{maximum}$ are small the round-trip will be longer. In some cases E_{total} can be smaller than E_{sample} so to make it $E_{sample} < E_{total}$ we have to take multiple samples and it increment the length of first period τ_{sample}.

2.6.3 Combining multiple time estimates

Each circle stands for a single approximation of local time C_r^a of node N_r's at some occurrence, which happens at local time C_s^a of node N_s. The interpolation is the line 'through the middle of this cloud'.(Figure 7)

2.6.3.1 Linear Regression

The most widely being used technique; a linear relation is $C_r = \alpha + \beta$. C_s is hypothesized and the coefficients α and β are determined by the actual samples and reducing the double difference between the fixed C_r's. The large number of samples needs large amount of memory which increases the quality of regression [1].

The coefficient β can be represented as an approximation of C_r relative to C_s. Linear regression as a result completely recompenses for clock drifts. If the drift is changeable, the hypothesized linear affiliation between C_r and C_s does not explain the reality well. In such conditions the counted number of samples must be small.

2.6.3.2 Phase-locked Loops

An additional technique for dealing with continuous stream of samples is based on the belief of phase-locked loops (PLL). PLL takes the authority of the slope and interpolation by means of proportional integral (PI). The only benefit of using phase-locked loop approach is that it desires a smaller amount of memory as compared to the linear regression.

2.6.4 Combining multiple time intervals

The approach of 'taking one sample' is used to develop the inferior and superior boundaries on the local time of far-off node. The following figure (Figure 8) shows a series of inferior and superior boundaries on the local times C_r of far-off node N_r on the y-axis and the resultant local times C_s of N_s on the x-axis [1]. Here we have two clouds; created with the help of inferior bound and superior bound samples.

2.6.4.1 Convex-hull

This method fills in the two clouds one by one. One bend is drawn above the inferior bounds and the other bend is drawn below the superior bound samples. Convex-hull avoids the standard values and takes samples with maximum and minimum sample errors under consideration.

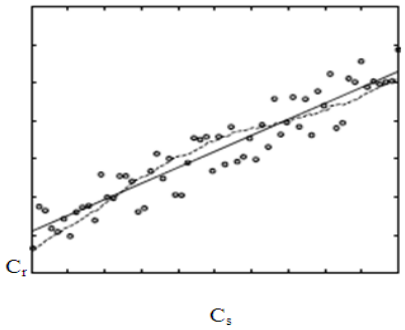


Figure 7: Every circle stands for a single approximation of node N_r's local time C_r^a at some event a, which occurs in N_s's at s local time C_s^a.

2.6.4.2 Synchronization of multiple nodes

Synchronization errors caused by delays are very hard to control, so we have to introduce multi-hop synchronization, which allows the nodes to communicate directly.

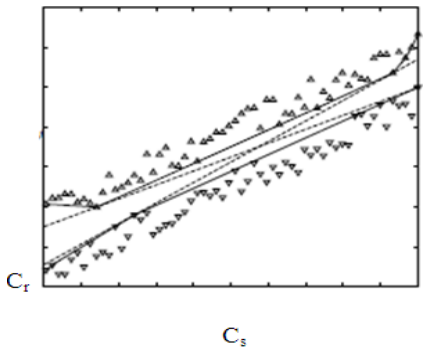


Figure 8: Series of inferior and superior boundaries on the local times Cr of far-off node Nr on the y-axis and the resultant local times Cs of Ns on the x-axis

2.6.5 Synchronization of multiple nodes

Synchronization errors caused by delays are very hard to control, so we have to introduce multi-hop synchronization, which allows the nodes to communicate directly.

2.6.5.1 Out-of-band synchronization

Synchronization is done by using out-of-band method such as GPS.

2.6.5.2 Clustering

All nodes in the clusters communicate with each other. Some nodes act as a gateway to interpret time stamp from one cluster to the other.

3.6.5.3 Tree construction

The solution for the synchronization problem is given by constructing a tree with only one master at the root [2, 3, 4, 5], for a multi-hop environment. The accuracy reduces as the distance of hop from the root increases.

2.6.5.4 Unstructured

In tree construction approach, synchronization is asymmetric and in unstructured, it is symmetric.

3. CONCLUSION

The attraction in wireless sensor networks has been increased because of their small size and rapid enhances, however one has to strive for acquiring the application goal. Ideal clock rate is impossible to maintain throughout due to the fluctuations reviewed in this paper. Fluctuations come from different sources that may be the environmental fluctuations that make the hardware faulty which in turn perterbates (varies) the time. Ten types of time that cause the delay in communication have been described, that further result in errors in the synchronized time of the network. The three clock models according to the source of errors have been explained. Moreover the classes of synchronization, the techniques used to synchronize time and the performance evaluation criteria have been told. Every scheme yields its own accuracy and consumes different magnitude of energy. These schemes can be graded according the energy efficiency, the required cost, the scope that the scheme has and the methods it uses to synchronize the whole network either internally or externally or makes synchronization in rounds. All the techniques can be executed to be post-facto or on demand (as a precautionary measure).

4. ACKNOWLEDGEMENT

We are very obliged to Mr. Muhammad Saeed Farooqui, Mrs. Kishwar Saeed, Mr. Tariq Mehmood and Mrs. Atia Tariq for their cooperation throughout this research work.

5. REFERENCES

[1]. Time Synchronization and Calibration in Wireless Sensor Networks Kay R¨omer, Philipp Blum, Lennart Meier ETH Zurich, Switzerland

[2]. Saurabh Ganeriwal, Ram Kumar, and Mani B. Srivastava. Timing-sync protocol for sensor networks. In First ACM Conference on Embedded Networked Sensor Systems (SenSys), November 2003.

[3]. Mihail L. Sichitiu and Chanchai Veerarittiphan. Simple, accurate time synchronization for wireless sensor networks In IEEE Wireless Communications and Networking Conference (WCNC'03), March 2003

[4]. Jana van Greunen and Jan Rabaey. Lightweight time synchronization for sensor networks. In 2nd ACM International Workshop on Wireless Sensor Networks and Applications, pages 11–19, September 2003.

[5]. Branislav Kusy and Miklos Maroti. Flooding time synchronization in wireless sensor networks, 2004.

Localization System for Multiple Mobile Objects in WSNs

Youngho KIM[1], Sehee WHANG[2], Sungjun KIM[2], Hyunchul KIM[2]and Sunshin An[2]
[1]Department of Telecommunication System Technology, Korea University, Seoul, South Korea
[2]Department of Electronics Engineering, Korea University, Seoul, South Korea

Abstract— *Monitoring application define an important class of applications used in wireless sensor networks. So it is very important regarding calculating and localization technologies. Localization service is a key-enable technology of diverse applications and widely exist in today's wireless sensor network. In generally, we are provided geographical location information by GPS(global position system). And how to find a geographical location method in unavailable to use GPS is what use fixed a location. However, this way has disadvantages, limited and cannot adjust mobility objects. In this paper, we propose the system for having mobility sensor node localization method , method localization using trilateration and rotation transform. We focus on using rotational transform, localization techniques and multi-objects.*

Keywords: Wiress senssor network; Localization; Mobile sensor network; trilateration; rotational transform

1. Introduction

In recent years technological advances have made the manufacturing of small and low-cost sensors economically and technically possible. These sensors can be used to measure ambient conditions in the environment surrounding them. Typically, wireless sensor networks (WSNs) contain hundreds or thousands of those sensors nodes. And users want to widely coverage area. Also, currently, most sensor node has mobility in WSNs. Due to the sensor features localization is the best suitable network to support wide area in such a scenario. This is one of the many reasons why we can not neglect the study of the collision effects and the noise influence. Many research centers worldwide (especially in Europe and USA) have focused their investigations on this kind of networks. One of the latest research lines in WSNs is called path discovery. There are many approaches which deal with this issue. However, due to the sensor constraints the design of the routing algorithm has to consider the Quality of Service (QoS) provided to the applications, in order to improve the related goals. In this sense, the use of rotating axis in WSNs offers an alternative way to relative position through the network. Typical applications like monitoring and activity recognition can be enhanced with this strategy. We present in this paper a new system for having mobility multi objects which introduces rotating axis to measure the relative position. The definition of a localization system. And time difference of arrival using acoustic has a widely application area and have been proposed to estimate the location of emitting sources. In theory, location can easily be calculated from TDOA and trilateration. However, this method is heavily affected by large measurement errors. So, I did as mentioned earlier, I used rotating transform before trilateration. It improve accuracy location data. The objective of this analysis is to determine the limitation of the accuracy of the general system caused by the errors in the range measurements, rather than data processing errors. Thus, we assume that:

- The data are taken simultaneously at each station
- In real, we don't know location information, however in simulation, we know distances trough positions.
- All of nodes only make negligible errors.

2. Related work

It is exactly infeasible to do localization without knowledge of the physical world[1]. According to the capacities of diverse hardware, we classify the measuring techniques into only three component (location, distance, angle).

2.1 Estimate Distance

To estimate distance , we usually use Radio Signal Strength(RSS), Time of Arrival(TOA) and Time Difference of Arrival(TDOA). There are many methods to calculate length between a node and the other nodes. However, in this paper , we mainly focus on RF(Radio Frequency) and acoustic of difference of signal arrival. These methods can get more accurate data[2][3].

2.1.1 Distance Measurement by using RF and acoustic TDOA

Nodes transmit signals prepare message including its ID, coordinates and surrounding temperature. These messages are sent by RF signals. And simultaneously send acoustic signals. When reference nodes are available, there is a category of measurements called TDOA. The transmitter sends a signal to a number of receivers at already known locations. Then, the receivers get the arrival time of the signal. The location of the transmitter is computed by the difference of the recorded arrival stamps. At this time, the time difference between the RF and acoustic signal is δ . and then we can indication of the following (1).

$$\delta T = \frac{d}{v_s} - \frac{d}{v_r f} \tag{1}$$

In generally, velocities of signal are 343 m/s for acoustic signals(v_s) and $3X10^8$ m/s for radio signals (v_{rf}).

2.1.2 Compensate Distance

Sound velocity depends upon temperature, barometric pressure, relative humidity, altitude, air composition, and so on. Therefore, to accurately estimate distances have to consider these factors[5]. However, it is not easy to reflect all elements. So we consider only temperature. Following equation is temperature in degree centigrade.

$$v_s(t) = 331.45\sqrt{1 + \frac{t}{273.16}} \quad (2)$$

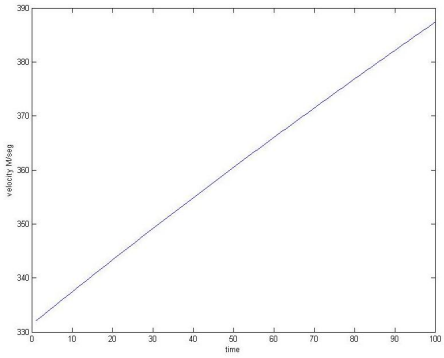


Fig. 1: Temperature in degree centigrade

2.2 Measurement location

Trilateration a method to determine the position of an object based on range measurements from three station located at known sites[1][2][3]. The relevant equations are nonlinear equations. and it is not easy to obtain an exact solution. A recursive formula for the solution of multi-tilateration systems can be found in. Figure 2. is a trilateration of example. At least, to estimate location we must know locations and distance of three nodes or more. A node unknown itself location calculates distance from three nodes known itself location.

2.3 Rotational transform

If the radio or radio signal coverage region can be described by determining which geometric areas that a node is in. The main idea in this paper is estimate rotational angel to rotate x-axis and y-axis. This angle can be estimate in a reference node and the other reference node. We can calculate new axis x' from (2) and angle of θ from (3).

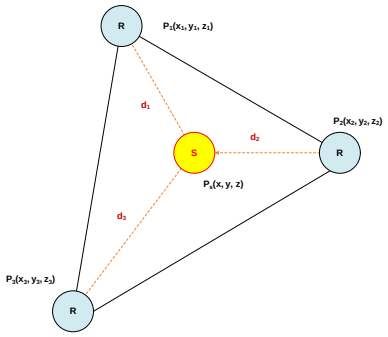


Fig. 2: Trilateration

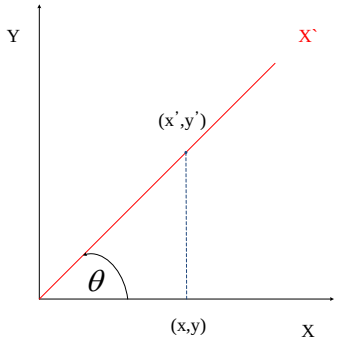


Fig. 3: Rotation transform x-axis to x'-axis

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & con\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
$$x' = xcos\theta - ysin\theta$$
$$y' = xsin\theta + ycos\theta \quad (3)$$

$$\theta = tan^{-1}\frac{a}{b} \quad (4)$$

3. Localization system

In this paper, we mainly focus on the localization system for mobile objects. And this system distributes operations to improve localization for large number of location of nodes, accuracy of location, easy to implementation system.

3.1 System Configuration

The proposed localization system can be separated by the role as above Figure 4. This system contains reference node, sink node, gateway. Furthermore, a user can monitor sensors through gateway. These nodes have same characteristic, homogeneous nodes.

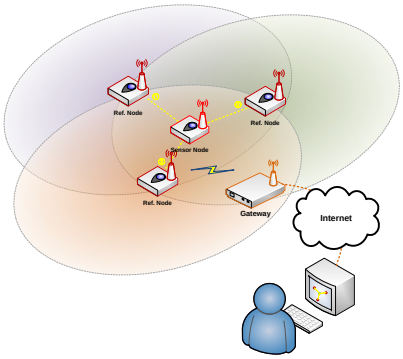


Fig. 4: Localization system configuration

3.1.1 Reference Node

Location of reference nodes is fixed. These nodes send information necessary to estimate location trough RF signal. These information contain temperature, reference node location, hop count, and so on. And if RF signal send to other nodes, acoustic signal simultaneously send to signal other nodes. Because, if two signals were not sent to other nodes, nodes could not calculate their position and distance. So, to accuracy get data, its have to send at the same time. Another important point is whether there are no necessary to reference nodes and reference nodes can change sensor nodes. Because sensor nodes have mobility. And then they are moving to get information.

3.1.2 Sensor Node

Sensor nodes have mobility. So They are moving to get information.Also they request order of neighbor nodes. Sensor nodes measure distance and time from three or more reference node. And then they calculate their position. In the next chapter I will give way to calculate position. At this time, its location will be compensated by its internal a angular sensor and a accelerating sensor. Sensor nodes can be divided into two roles. First thing are called first mover(Leader Node). This node have to get moving direction. Because if this node did not have direction, the others will move anywhere. So this node's role is very important. The other nodes are called following nodes.They move along a first mover node. Another important point is that if a sensor node doesn't receive three or less reference nodes, their role change sensor node to reference node. This function can improve accuracy, mobility and overlapping coverage region.

3.1.3 Gateway

Role of a gateway is very simple. It helps users' requests transmit and delivers the information about sensing values. Another main role is bridge. It allow to connect between user and nodes through existing networks.

3.2 System Initial Setup

Following steps are showing that system course of actions. These steps can be divided according to estimate location, switch reference node to sensor node, moving of nodes along the leader node.

3.2.1 Reference Nodes Setup

All of nodes are placed randomly in coverage region. And nodes are on standby. Received signal nodes from a criteria node send active message to neighbor node. They request query neighbor ID number owing to aware of neighbor node.

1) Received signal nodes from a criteria node send active message to neighbor node. They request query neighbor ID number owing to aware of neighbor node.
2) Received query messages nodes send their Id number and neighbor list to other nodes. At this point, if they have no list, they send null message.
3) Activated nodes save their neighbor node list in their list. If there are three or more the number of windows in their neighbor area, They are chosen reference node without doubt.
4) We can select reference nodes though recursive above steps.

3.2.2 Routing Table Setup between Reference Nodes among the reference nodes

1) When select reference nodes, reference nodes will recognize their neighbor nodes.
2) Each reference nodes update neighbor reference node list. And then we can map connective of interconnection for reference nodes.

Routing table for inter reference nodes will used when a leader node broadcasting forward direction.

3.3 System Operating

3.3.1 Estimate Reference Node Location

We divided two kinds of nodes, sensor node and reference node. At first, owing to reference nodes are fixed, we can find easily their location. we can obtain distance through difference signals velocities. And then draw the line X'-axis first reference node and having shortest distance reference node form first reference. This axis is new X-axis, called X'-axis. We can calculate difference x-axis to X'-axis. And then all of the reference node is rotated by θ. θ is angle of difference x-axis to X'-axis.

1) Draw new axis: We already know location 1st reference node from gateway. It is absolute values. And then line between 1st reference node and shortest reference node. This line is new X-axis. And orthogonal axis can be defined new Y-axis. Then we can find θ.
2) Find 'θ ': Finding θ is very simple. θ can be found by trigonometric function. For example, if shortest distance from 1st reference node's coordinates

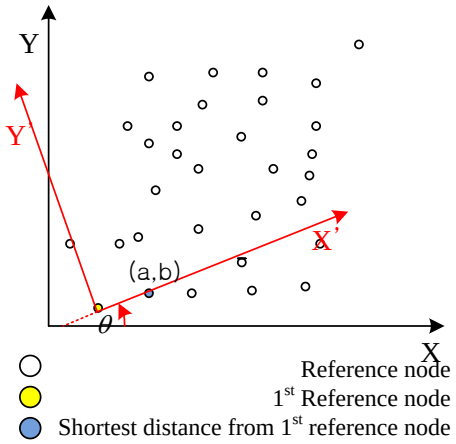


Fig. 5: Localization by using rotation axis

are$(\sqrt{3},1)$,$\theta$is π/6. The procedure is as follows (4).

$$\theta = tan^{-1}\frac{a}{b} = tan^{-1}\frac{1}{\sqrt{3}} = \frac{\pi}{6} \tag{5}$$

3) Rotate all coordinate: We find θ. Then we are able to rotate every coordinates. Each coordinates multiple -θthrough rotation transform. Then we can obtain relative positions. For example coordinate is$(\sqrt{3},1)$ converted to $(\sqrt{3},0)$. The procedure is as follows (5).

$$\begin{pmatrix} x^{'} \\ y^{'} \end{pmatrix} = \begin{pmatrix} cos(-\theta) & -sin(-\theta) \\ sin(-\theta) & con(-\theta) \end{pmatrix} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{3} \\ 0 \end{pmatrix} \tag{6}$$

3.4 Estimate Sensor Node Location

After reference node are setup, we can find location of sensor node from reference node.This mean use trilateration. In practice distance measurements inevitably contain errors, resulting in that the circles may not always intersect at a single point. This problem can be solved by a numerical solution to ac overdetermined linear system. Suppose an a sensor node and it is alive to obtain the distance $d_i^{'}$ to the i th reference node locating at (x_i,y_i),1≤i≤ n where n is the number of the neighbor node. Let $d_i^{'}$ be actual Euclidean distance from a sensor node to i th reference node,i.e.,

$$\begin{aligned} d_i &= \sqrt{(x_i - x_0)^2 - (y_i - y_0)^2} \\ d_1^2 &= (x_1 - x_0)^2 - (y_1 - y_0)^2 \\ d_1^2 &= (x_2 - x_0)^2 - (y_2 - y_0)^2 \\ d_1^2 &= (x_3 - x_0)^2 - (y_3 - y_0)^2 \\ &\vdots \\ d_n^2 &= (x_n - x_0)^2 - (y_n - y_0)^2 \end{aligned} \tag{7}$$

Subtracting the first equation from all of the rest equations gives

$$\begin{aligned} d_2^2 - d_1^2 &= x_2^2 - x_1^2 - 2(x_2 - x_1)x_0 + y_2^2 - y_1^2 - 2(y_2 - y_1)y_0 \\ d_3^2 - d_1^2 &= x_3^2 - x_1^2 - 2(x_3 - x_1)x_0 + y_3^2 - y_1^2 - 2(y_3 - y_1)y_0 \\ &\vdots \\ d_n^2 - d_1^2 &= x_n^2 - x_1^2 - 2(x_n - x_1)x_0 + y_n^2 - y_1^2 - 2(y_n - y_1)y_0 \end{aligned} \tag{8}$$

Rearranging terms, the above equations can be written in matrix from as

$$\begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_n - x_1 & y_n - y_1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_2^2 + y_2^2 - d_2^2 - x_1^2 + y_1^2 - d_1^2) \\ x_3^2 + y_3^2 - d_3^2 - x_1^2 + y_1^2 - d_1^2) \\ \vdots \\ x_n^2 + y_n^2 - d_n^2 - x_1^2 + y_1^2 - d_1^2) \end{bmatrix} \tag{9}$$

Then, this equation can be rewritten as

$$Hx = b \tag{10}$$

The least-squares solution of this equation is given by[1]

$$\hat{x} = (H^T H)^{-1} H^T b \tag{11}$$

1) Sensor nodes send query to neighbor reference nodes to collect information.
2) Received query message reference nodes send to sensor node including node ID, reference nodes count number.
3) Sensor nodes save received data in reference windows. And they have to decisions that reference nodes are selected by probability. And then we make a roulette by the probability.
4) Requested reference nodes give node ID, reference count number,temperature trough RF signal.Also they send acoustic signal at the same time.
5) A sensor node receives the two signals.Sensor nodes measurement distance by(7). These values store in the window. If windows buffers are full,it replace oldest thing.A sensor node calculates if it has three or more values in the windows buffer.
6) Reference nodes save got the counter values from a sensor node.
7) Recursive above the steps a to f.

The following figure 7 is showing phases that a new added sensor node choice the reference nodes by probability. Red marks mean that reference nodes are selected by a sensor node. At last phase, R_1 is elected. And then, R_2 is removed.

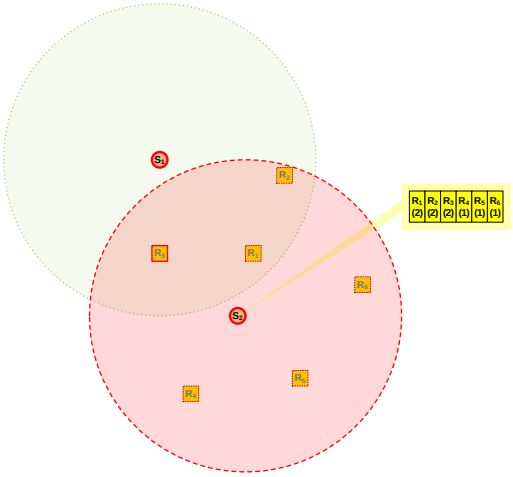


Fig. 6: Sensor Node Add

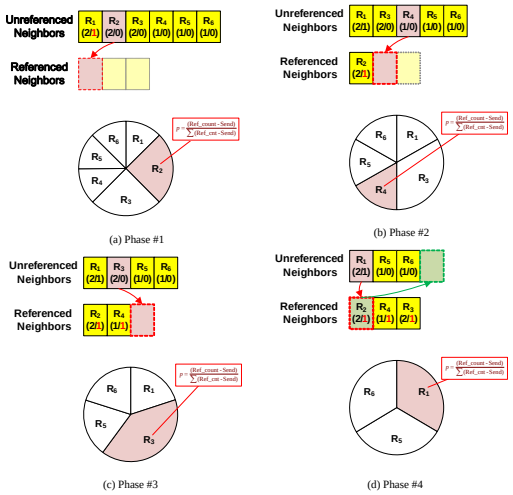


Fig. 7: New Sensor Node S_2 select reference node process

3.4.1 Group Control and Nodes Movement

A way how group control of nodes with mobility is that separate a leader node and following nodes. And then let following nodes follow a leader node.

3.4.2 Role Change Sensor Node to Reference node

In this paper, we consider that the number of sensors is limited. So we can't help changing node's role reference to sensor. Following figure 8 is showing role change.

1) A moving sensor node recognizes neighbor reference nodes.
2) If a sensor node move to signal reaches up to 70 80 percent, and then the node stop.And then a sensor node is measurement its distance and location.
3) A finished the setup sensor node changes the role to a reference node

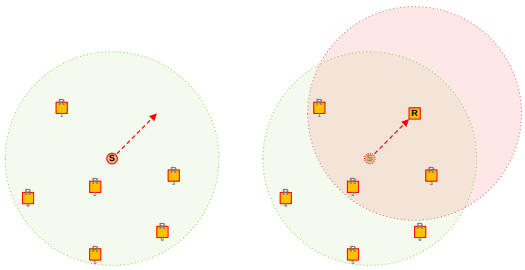


Fig. 8: Role Change A Sensor Node to A Reference Node

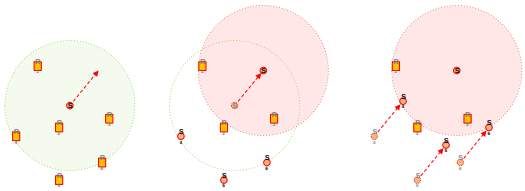


Fig. 9: Role Change A Sensor Node to A Reference Node

3.4.3 Role Change Reference node to Sensor Node(remove reference node)

If there is no reason remain. They change the role reference node to sensor node. A changed node have to move following a leader node.

1) Reference nodes perceive their position and there is no request nodes.
2) In limited time,if there is no more request, they change the role. and ask to remove node list adjacent reference nodes.
3) They start move following a leader node after a changed node receive confirm a message from adjacent node.

4. Localization algorithm

The function based algorithm gives authentic result. But these algorithm don't contain error of hardware. e.g. reflection, collision, receive error. It just indicate that Euclidean distance and rotation transform. Following algorithms use in MATLAB.

4.1 Euclidean Distance Algorithm

This algorithm can be calculate distance a node to reference node. These are put on the grid. But in the real system, They don't have any information about location . The detail algorithm is the following:

4.1.1 Input:

- Reference node initial position
- Sensor Position : (x_n, y_n)

We actually don't know real position. however we can calculate distance through TDOA. In this case, we define

real position to distance. And then we can estimate Euclidean distance.

```
   dimension(p1) = length(p1);
dimension(p2) = length(p2);

   if (dimension(p1)  = dimension(p2)
distance = -1;
disp('Invalid points - different dimensions');
return;
end

   total.distance = 0;
diff = p1 - p2;

   for i=1:length(p1)
total.distance = total.distance + diff(i)^2;
end

   distance = sqrt(total.distance);
```

4.1.2 Output:

All of output data store matrix form(distMatrix). A cell contain distance from a reference node.

4.2 Rotation Transform Algorithm

This algorithm use in calculate rad and rotation transform. It can be expressed through relative coordinates. We just know information about location in mobile nodes. It can be adjusted characteristic of even function and odd function owing to rotate -θ.

4.2.1 Input;

- Reference initial position
- Shortest distance from above node
- Position we want to convert

```
function pos =
Calculate2ndRelativePosition2(refnode,node1,nodei)
This function returns 2nd position of node

   a=EuclideanDistance(refnode,node1);
b=EuclideanDistance(node1,nodei);
c=EuclideanDistance(refnode,nodei);

   angle.rad=atan(node1(1,2)/node1(1,1));

   x=nodei(1,1);
y=nodei(1,2);

   x1 = ((x)*cos(angle.rad))+((y)*sin(angle.rad));
y2 = ((y)*cos(angle.rad))-((x)*sin(angle.rad));

   pos = [x1, y2];
```

4.2.2 output:

This algorithm also store date to matrix form. There are relative positions in data matrix.

5. Implementation and Result

To illustrate the effectiveness of the proposed method, experiments were simulated by MATLAB. How to obtain the distance a node to a node is replaced TDOA to define coordinates. We assume that distance can be measurement by TDOA. So there is no problem regarding this method.

5.1 Position of Fixed Nodes

All of nodes are randomly placed. This illustrate like real condition.

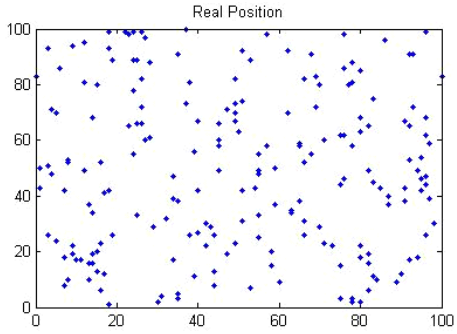


Fig. 10: Real Position

5.2 Relative location of nodes

These nodes rotated by $-\theta$. Our mainly purpose obtain location information. This can be obtained by rotation transform. And then we can easily estimate position in mobile networks. Following figure 11 shows relative location of nodes.

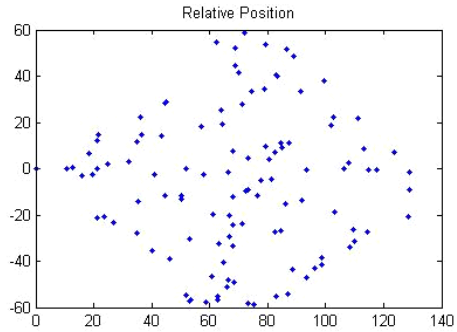


Fig. 11: Relative Position

5.3 After moving nodes

Following figures represent status after moving.

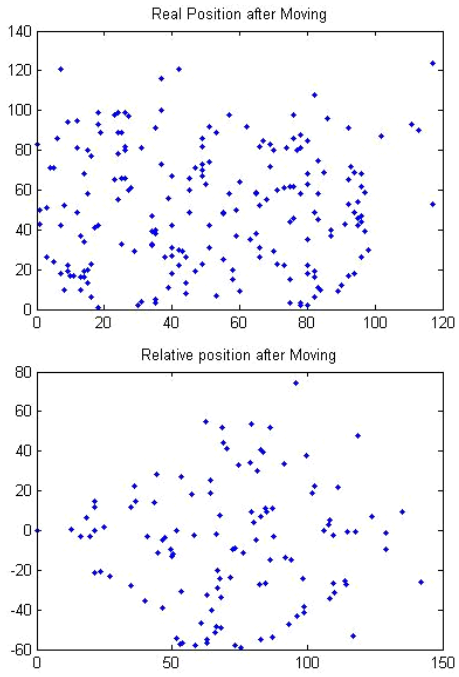


Fig. 12: After moving nodes

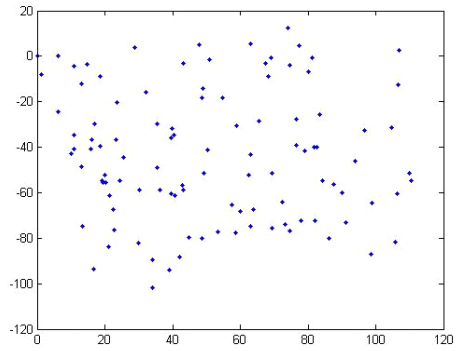


Fig. 13: Relate Position by Trilateration

5.4 Result

In this paper, I propose the localization system for multi mobile nodes using difference velocity RF signal and acoustic signals. This study has shown that attractive possibilities for making accurate position measurement by trilateration and TDOA. Figure 12 shows rotating relative position after moving. Significant simplification of the solution of the trilateration and rotating transform computation are achieved by recursion formula. Before estimate location through trilateration, I have used the rotation transform. It can be more easily and more accurate. Figure 13 sho that relative position using trilateration without rotation transform. This result obviously has error. Because angle of the negative partial has same result angle of the positive partial. Therefore, this system is especially useful in mobile sensor network. Currently, I keep going on test in the real field. This is the topic of the future.

References

[1] Yunhao,L.and Zhen,Y., Location, Localization, and Localizability: Location-Awareness Technology for Wireless Networks, Springer, 2010

[2] Hui, L. et al,âĂIJSurvey ofWireless Indoor PositioningTechniques and Systems.âĂİIEEE Trans. on System, Man, and Cybernetics, vol. 37, no. 6, pp. 1067-1080, Nov., 2007.

[3] Yanying, G. et al, âĂIJA Survey of Indoor Positioning Systems for Wireless Personal Networks.âĂİIEEE Comm. Survey / Tutorials, vol. 11, no. 1, pp. 13-32, 1stQuater, 2009

[4] D.A. Bohn, âĂIJEnvironmental Effects on the Speed of Sound.âĂİ83rd Convention of the Audio EngineeringSociety, New York, October, 1987

Implementing a Wireless Sensor Network Using MEMSIC'S Professional Kit

Stephen Ellis, Lorson Blair and Yenumula B Reddy
Department of Computer Science, Grambling state University, LA 71245, USA.

Abstract - *The research focuses on configuring MEMSICs professional kit to study the temperature, light, and humidity in a specific environment. The study involves the implementation of wireless sensor networks using Motes with specific requirements. We developed custom program application to target specific environmental elements specifically light and temperature (humidity, pressure, and other elements are not provided in this paper). By reconfiguring the database properties and base-station activity, we interact with the MoteView monitoring software and stored the results of incoming data. The experimental results are provided with appropriate tables and figures.*

Keywords: MEMSIC'S Professional, Wireless Sensor Network, MoteView, monitoring and collection of data.

1. Introduction

A Wireless Sensor Network (WSN) is a new technology that is utilized in data collection and monitoring fields. It is defined as a set of nodes that are organized into a cooperative network. In this project, we implemented a wireless sensor network using MEMSICS Professional Kit for Wireless Sensor Networks. The project is divided into two parts. Firstly, we aimed to achieve the network functionality. In achieving network functionality, we studied the operation of the motes (how they collect and transfer data, how they connect to each other and the base station etc) and software used to collect, retrieve, and view data. Further, we studied the interfaces used to program the motes. After initial network setup and collection of data, we developed a system to monitor light and temperature in four rooms in the TL James building, at Grambling State University. As a result, we programmed four of the motes with a custom program that collects and reports only light and temperature (humidity temperature) back to the base station. Visualization software (MoteView) provided by MEMSIC was used to view and analyze the data.

The references [1-3] provide the documentation for the MEMSIC professional kit. The sensor network concepts and some of the implementation methodologies are observed in [4-7]. Sensor networks are used flood control, agriculture, forest fire detection, temperature control, and parking identification. Today, sensors are every part of the instrument, war fields, and medical studies. The literature demonstrates the applications of sensors are unlimited.

1.1 Objectives

- To implement a WSN using MEMSIC motes - accomplished by first understanding and setting up a simple network, monitoring and analyzing data transfer and implementing a monitoring system that collects light and temperature (humidity temperature) data and the procedure of the data sent back to the base station.
- To develop a light and temperature monitoring system using MEMSIC Professional Kit for WSN - involves the reprogramming of the motes to get them to collect only light and temperature data (humidity temperature) and transfer the data to the base station. It requires completing three major tasks.
  - We modified the custom programs to collect and transmit only light and humidity temperature to the base station.
  - In the second step, the newly modified programs are compiled and downloaded to all the motes.
  - Finally, we used MoteView to monitor the system.

1.2 Restrictions

Implementation of modified data sampling application for a wireless sensor network leads a restriction on the client tier. MoteView, the software used to monitor the system, didn't allow any new applications to be added to its usable interface or the addition of a new relation in its database for data storage.

2. Hardware and software Requirements

2.1. Hardware

The implementation of the sensor network requires specific hardware supplied by MEMSICS. The MEMSICS Professional Kit consists of the following hardware:

- 6 Sensor Nodes
- 1 Base Station

- Processor/Radio Module
- 1 Data Acquisition Board
- 1 USB Programming Board

The specification of each hardware device is shown in the Table 1. The sensor nodes, the base station, USB programming board along with the additional hardware given below were used to set up of the network. Additional hardware includes:

- Batteries to power the sensor nodes.
- A laptop computer to connect the base station and used for data collection and analysis. The laptop was also was used to write the programs for the notes as well as upload these programs to the nodes. It should have the following minimum requirements:
  - 1 GB of free space in the destination drive.
  - 550 MB of free space in the C drive.
- USB extension cables to connect the devices to the laptop.

2.3 Software

The sensor networking kit includes several software packages that are needed to connect the motes, collect and analyze data, write and compile programs for the nodes, and upload these programs to the nodes. The list of software and their function (that are of importance to this project) are shown in Table 2.

MEMSIC provides additional software packages; however, only those listed above were used in our initial implementation and study of the WSN and the implementation of the light and temperature monitoring system.

To install the above software (with the exception of XMesh), the laptop needed to be programmed with Windows XP Professional Service Pack 2 (XP-Pro SP2).

3. Initial Network Setup/Data Collection

3.1 Topology Overview and Background

XMesh allows the motes to automatically connect in a mesh network topology. A mesh network is a multi-hop network in which all nodes (sensor motes) can communicate with each other to route data to and from a base station. Mesh networks provide multiple routes for data transfer and are highly fault tolerant. They also allow networks to expand over unlimited distances and allow nodes to converse energy. XMesh provides a TrueMesh networking service that is both self-organizing and self-healing. It can broadcast within a single area of coverage or arbitrarily between any two nodes or cluster. It offers quality of service (QOS) either by link level acknowledgement (best effort) or by end-to-end acknowledgement (guaranteed delivery). XMesh can also be configured in various power modes including high power (HP), low power (LP), and extended low power (ELP). Other features of XMesh include multiple transport services, health diagnostic, time synchronization, and over-the-air programming (OTAP) [3].

Each wireless sensor is programmed to measure humidity, humidity temperature, present temperature, pressure, light, vertical acceleration (y-axis), and horizontal acceleration (x-axis). These nodes have also been programmed to follow XMesh protocol.

3.2 Configuring the Client Tier

3.2.1 Primary Option

To set up the network, the necessary software had to be installed on the laptop computer to provide the interface to communicate with the nodes and collect and analyze data. MoteWorks (and all its components) and MoteView (and all its components) were successfully installed on the laptop with the Windows XP SP2 operating system.

3.2.2 Alternate Option

Using a preconfigured version of the OS Windows XP SP2 on a virtual machine such as VMware Player or Virtual Box provides a solution to combine the capability of the needed operating system Windows XP SP2 with newer versions of Windows or Linux simultaneously.

3.3 Network Deployment/Configuring the Mote Tier

Only four motes were used for the initial setup. The motes were powered up and placed at different locations. The location of the motes and the approximate distance between each and the base station is given in the Table 3.

4. Collecting Live Data

With the motes in their desired location, MoteView was opened, and the "Connect to WSN" button was clicked to collect live data. In the "Mode" option the operation, "Acquire Live Data", was selected from the "Select Operation Mode" options and "Local" was selected from the "Select Acquisition Mode" options. In the "Gateway" option/tab, the interface board MIB520, the serial port COM4, and the baud rate 56700 were all selected from the relevant drop down menus. Next, in the "Sensor Board" option, XMTS400 was selected as the sensor application. The "Done" button was then clicked and this initiated data collection.

Once live data collection was started, the behavior of the network was monitored. A snap shot of the network topology, with present temperature selected as the data being collected is shown in Figure 1. (N.B. Initial data collection can be stopped by simply clicking the "Stop XServe" button on the top MoteView interface. Data collection can be restarted by checking the "Live" checkbox, and clicking the "Start XServe" button.).

The different categories of data were being sent from the motes to the base station every 10 seconds. Data was collected over a 30 minute period. The default sampling rate for high power applications is 2 seconds, while the default for low-power applications is 3 minutes. The minimum sampling rate that can be applied to the motes is 300 milliseconds.

In addition to collecting data of environmental conditions, the motes also send health information about themselves. The health information includes data on how well the network is performing with respect to radio traffic, battery voltage, and parent's node Radio Signal Strength Indicator. Nodes also send health information for their neighbors.

5. Results

A sample set of results from the first deployment is shown in Table 4 and a sample set of results from our modified program in Table 5. Given the environment, the data collected over the time period was very consistent, predictable and valid. The data collected included humidity, humidity-temperature, current-temperature, pressure, light, horizontal acceleration, and vertical acceleration. All data were displayed in standard engineering units. This result can be viewed in MoteView. A snapshot of this interface is shown in figure 2. MoteView also gives us the ability to view charts and graphs of the data. You can select the data set you want to view the chart for. A snapshot of a chart for humidity temperature is shown in Figure 3.

6. Observations

The connections between the motes and the base station were made automatically by XMesh. Because the mesh network connections are influenced by environmental factors, the network topology changed approximately once every two minutes. In some instances, the topology change may lead the motes 7654 and 7653 connected to motes 7651 and 7651 respectively. This frequent change in network topology is XMesh's way of trying to find the most efficient route for data transfer. The frequency of the change is also due to the fact that the motes are relatively close to each other. Thus, either route connection would not cause a drop in efficiency.

Another observation made was that the transfer of data from mote 7653 to mote 7651 then to the base station was relatively efficient, with an average failure rate of 2.5%. Further, the transfer of data from motes directly connected to the base station had an average failure rate 1.5%.

Because all the motes were located in the same environment, the data collected over the 30 minute period was very consistent. Data collected were stored in a database. This data can then be retrieved later for analysis.

7. Data Transfer

Individual motes collect data from their respective environments and forward these data to the base station. The base station then transfers the data to the server, where it is stored in the database. The data can then be viewed at the client tier via MoteView. MoteView also provides tools to help with the analysis of data, drawing charts, graphs, and viewing the health status of the motes.

The motes also generate and forward health data packets. Health data allow us to monitor the health or state of the mesh network. Each mote sends its health packets towards the base station and also to its neighbor. In addition to its own health data packets, a mote also transfers a neighbor health packet. This allows the motes to keep track of the most optimal route for forwarding data packets towards the base station. All health data packets are logged to the database.

8. Sensor Data Restrictions

The Tables 6 and 7 give the specification for the humidity, temperature, and light sensors, the three main sensors of importance to our monitoring system.

Note: the voltage that the sensor operating is directly proportionate to the accuracy of the results.

9. Conclusions

We have completed the following objectives:

- Implementing a live wireless sensor network and the specified requirements for it to function successfully.
- Developed a custom program application to target specific environmental elements specifically as light and temperature.
- Reconfigured database properties and Base station activity to interact with the MoteView monitoring software and store results of incoming data packets.

Future Work:

- Implementation of a Clustering Algorithm and allow fusing of data from a cluster-head consequently cutting down network traffic.
- Implementing the event-based protocol to force motes to send data packets in the event of specific environmental changes.
- Implementation of Security Measures to identify Malicious Nodes and avoid them in the WSN.

Identifying algorithmic patterns to allow the motes that automatically adapt to the environmental changes and determine drastic events to report back to the base station.

Acknowledgement

The authors wish to express appreciation to Dr. Connie Walton, Provost, and Vice President of Academic Affairs, Grambling State University for their continuous support.

10. References

[1] "MoteWorks Getting Started Guide", Crossbow, April 2007.
[2] "MoteView Users Manual", Document Part Number: 7430-0008-05 Rev A, Crossbow, May 2007.
[3] "XMesh MoteConfig USER MANUAL". MEMSIC, Inc. Document Part Number: 7430-0112-02 Rev A.
[4] J. Stankovic, "Wireless Sensor Networks", *Handbook of Real-Time and Embedded Systems*, CRC, 2007.
[5] Martin Turon., "MOTE-VIEW: A sensor Network Monitoring and Management Tool", The Second IEEE Workshop on, pp 11-18 .
[6] J. Suh and M. Horton., "Current Hardware and software Technology for sensor Networks", First International Workshop on Networked Sensing Systems (INNS)", 2004.
[7] R.Szewczyk, J. Polastre, A. MainWaring, and D. Culler., "Lessons from a Sensor Network Expedition", first Europian Workshop on Wireless Sensor Networks (EWSN), 2004.

Tables

Table 1 Hardware Components and Specifications

Devices	Components and Description
Sensor Nodes	-IRIS Processor Radio Module - modules to enable the low-power wireless sensor networks measurement system. Available in 2.4 GHz. -MTS400 Basic Environmental Sensor Board – MTS400 multi-sensor board including temperature, humidity, barometric pressure, acceleration and ambient light sensing capabilities.
Base Station	-IRIS Processor/Radio Module – module functioning as a base station when connected to the USB PC interface. -MIB520 USB Programming Board – provides a USB Interface for data communications.
Processor/Radio Module	IRIS Reference Board
Data Acquisition Board	MDA300 Data Acquisition Board – high performance data acquisition board with up to 11 channels of 12-bit ADC analog input and onboard temperature and humidity sensors.
USB Programming Board	MIB520 USB Programming Board – provides a USB Interface for data communications.

Table 2 Software Packages used in the implementation.

Software	Function
XMesh	Multi hop networking protocol installed on each node.
TinyOS and MoteWorks	An event-driven OS for wireless sensor networks. It also provides tools for debugging.
NesC compiler	An extension of the C-language designed for TinyOS.
Cygwin	A Linux-like environment for Windows.
XSniffer	Network Monitoring Tool for the RF environment
MoteConfig	GUI environment for Mote Programming and Over the Air Programming (OTAP).
MoteView	An interface between a user and a deployed network of wireless sensors. Provides the tools to simplify deployment and monitoring.
Programmer's Notepad 2	A simple IDE for nesC code.

Table 3 Placement of nodes

		Distance from (in meters)				
Node ID	Location (Room #)	Base Station	7251	7252	7253	7254
Base Station	131		5	15	35	15
7251	131	5		10	15	20
7252	132	15	10		10	15
7253	Lobby	30	25	15		15
7254	Copy Room	15	10	10	15	

Table 4 Sample Set of Data

ID	Humidity (%)	Humidity-temperature (C)	Pressure-temperature (C)	Pressure (mba)	Light (Lux)	Horizontal acceleration (m/s^2)	Vertical dacceleration (m/s^2)	Time 10/24/2011 PM
7651	45.65	23.71	23.87	1006.3	308.89	3.332	-30.772	5:05:06
7652	48.79	24.89	24.91	1005.89	294.17	20.384	-24.5	5:05:06
7653	48.53	23.07	23.25	1005.69	514.3	0.196	-0.196	5:05:06
7654	43.97	24.32	24.78	1005.59	285	0	0	5:05:06

Table 5 A sample set of data from our modified light sensing application

Id	Time	parent	voltage [V]	lightc [lux]
7653	2/29/2012 15:25	0	2.5506	8.51
7652	2/29/2012 15:25	0	2.5249	10.35
7654	2/29/2012 15:25	0	2.5454	24.61
7651	2/29/2012 15:25	0	2.5351	11.27

Table 6 Humidity and temperature sensor specifications

Sensor Type	Sensirion SHT11	
Channels	Humidity	Temperature
Range	0 to 100%	-40°C to 80°C
Accuracy	± 3.5% RH (typical)	± 2°C
Operating Range	3.6 to 2.4 volts	
Interface	Digital interface	

Table7 Light sensor specifications

Sensor Type	Taos TSL2550
Channels	Light
Range	400 – 1000 nm
Operating Range	3.6 to 2.7 volts
Interface	Digital interface

Figures

Figure 1. Network Topology

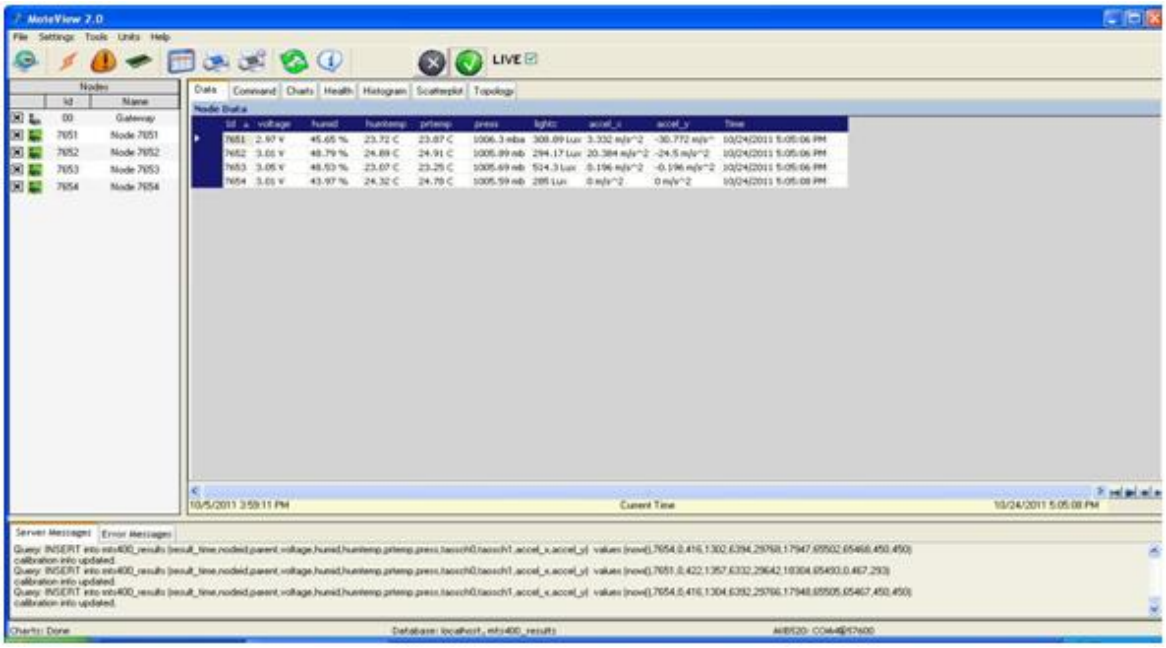


Figure 2. Tab in Mote-View (Data collected from initial data collection)

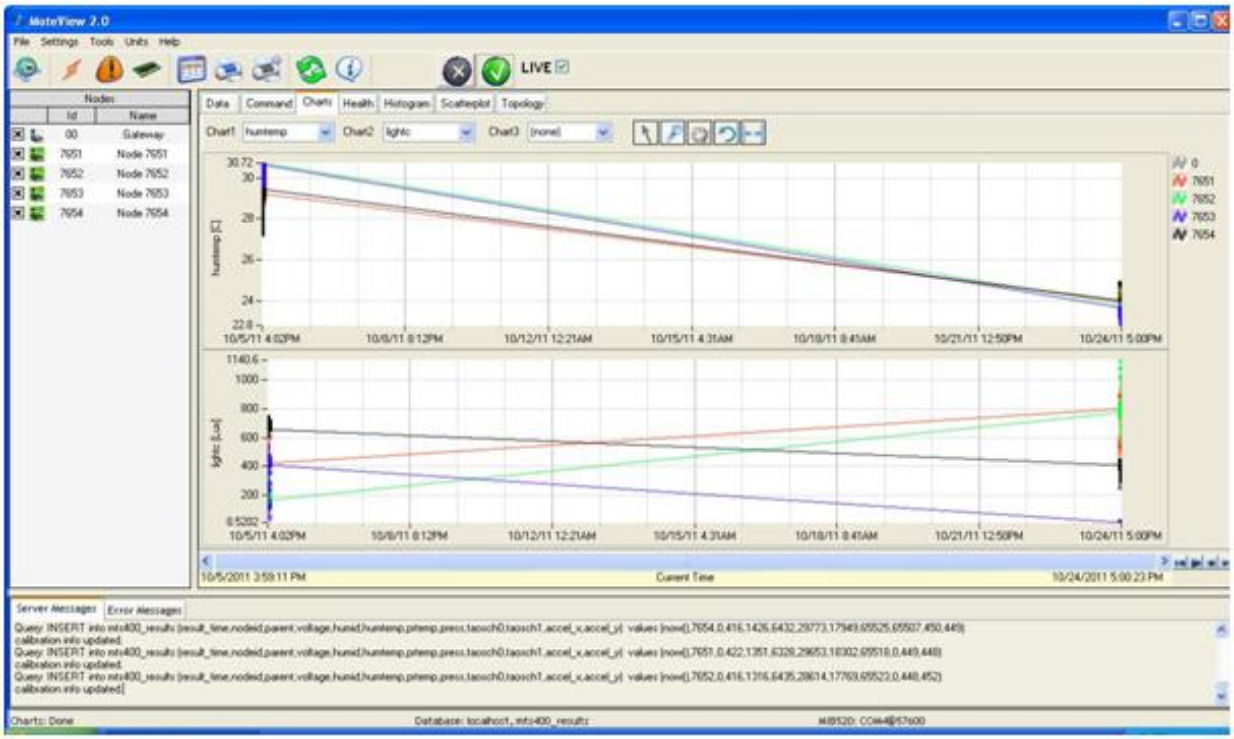


Figure 3. Mote-View chart for humidity temperature

Mobility of Data Collector Along Averages and Clustering Algorithm to Maximize the Lifetime in Wireless Sensor Networks

Abdullilah Alotaibi , Abdulaziz Almazyad,
Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia
Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia

Abstract

In this paper, we consider data gathering in wireless sensor network by utilizing mobility of data collector along averages and clustering algorithm . Basically, we have two classes of data, delay tolerant data and delay sensitive data. The mobile data collector moves along averages of the sensing field to collects data from the sensor nodes or cluster heads. Sensor nodes directly send data to their associated cluster heads in case of delay tolerant data, or packets are sent to a relaying node currently near the data collector in case of delay sensitive data. The data collector updates its trajectory after it knows the cluster head locations and moves toward them when they are away from its trajectory to collects data from them in a single hop.

We exploit the mobility of data collector to distribute the load over the network. We focus on data gathering with mobile data collector using clustering technique to maximize the lifetime of the network.

We also present a theoretical analysis to specify the impact of mobility of data collector along averages of the sensing field on the lifetime of network as compared to a network with data collector moving along perimeters of the sensing field and stationary data collector. Based on the simulation results we show that our scheme would maximize the lifetime of wireless sensor network. Our algorithm will be evaluated based on three basic metrics : Network Lifetime , Power Consumption and Average Packet Delivery rate.

Keywords: Data Collector; Wireless Sensor Network; Network Lifetime; Clustering; Mobility.

1. Introduction

Wireless sensor networks usually consist of a lot of sensor nodes, which are battery-powered tiny devices, a short-range wireless communication and a low capacity processor. These devices achieve three basic tasks: Monitor some surrounding environmental phenomenon, process and store the sensed data, and then transfer them through wireless communications to a sink node or data collector [1]. Increasing the lifetime of WSNs is crucial to enable their use in this wide range of applications. Because the sensor nodes are energy constraint, they send their data hop by hop using multi hop or other mechanism to a static or mobile sink or data collector .When any sensor node has data to transmit, if it is in the sink communication range it will send directly , otherwise it will use multi-hop to deliver its data to data collector. When multi hop transmission is used ,it will result an unbalanced energy load around the sink. Nodes near the sink or data collector are used as forwarding nodes to deliver data generated all over the network to the sink. Therefore, nodes around the sink deplete their energy much faster than other nodes [2]. Most of works use a single sink to improve the network lifetime such as in[3][4]. In the work presented in [5] the researcher uses a mobile data collector moves along boundary of the sensing field and collects data from the sensor nodes. In [6] the authors proved that using a mobile sink is more efficient than a static one and thus helps to increase the network lifetime. The authors utilize multi hop transmission to deliver data to the data collector at its current location. The scheme requires sensor nodes to be synchronized with the data collector to specify the current location. In [7] the authors utilize only one mobile sink that moves through a straight line while gathering data from the sensor nodes. This approach reduces the number of hops a packet has to travel in order to reach the sink.

In our work we present a distributed scheme that exploit mobility of data collector along averages and clustering algorithm to distribute the load over the network and there is no synchronization between the data collector and the sensor nodes. The sensing region in our work is divided into four equal squares and the data collector moves along the averages [1] of these squares. Any sensor node lies inside the data collector communication range (less than r m) is called a relaying node . We classify the data as presented in [5] into two types: delay sensitive data and delay tolerant data. In the case of delay sensitive data, relaying nodes receive data from other sensor nodes and send them to data collector or to a relaying node that currently near the data collector. Delay tolerant data are sent to a cluster head that the node belongs to and the cluster head waits for the data collector to comes and pick up

[1] Averages mean the Center of Four Squares.

data. The main contributions of this paper are described as follows:

1) We propose a new scheme for determining the movement of data collector based on the averages of the sensing field.

2) Our scheme finds good routing solution with a very low computational complexity and this scheme does not require a synchronization between sensor nodes and the data collector.

3) Our scheme combines the advantage of maximizing the lifetime of the network and minimizing the delay by maximizing the packet delivery success.

4)Our scheme has distributed routing where the sensor nodes make them routing decisions by themselves.

5)Our scheme exploits mobility of data collector moving along averages by increasing the number of relaying nodes, and clustering algorithm with cluster head rotation to distribute the load and it is also more scalable .

Our scheme is considered distributed scheme with no synchronization between data collector and sensors. It also combines data gathering with our clustering algorithm and mobility of data collector using good routing scheme to deliver data online and to maximize the network lifetime.

The rest of this paper is organized as follows. In Section II we describe our system model under study including the major assumptions. In Section III, we present our routing scheme. Section IV shows the experimental results. Finally, in Section V, we conclude the paper.

2. System Model and Problem Definition

Our wireless sensor network composed of N numbers of sensor nodes distributed randomly in the sensing field. This network is modeled as a graph G(N,E) where N is number of nodes and E is the set of all links(i,j) where j is the communication range of i. Node j is a neighbor to node i if node j is in node i communication range, and vice versa

2.1 Assumptions

Our scheme is based on the network model with the following assumptions:

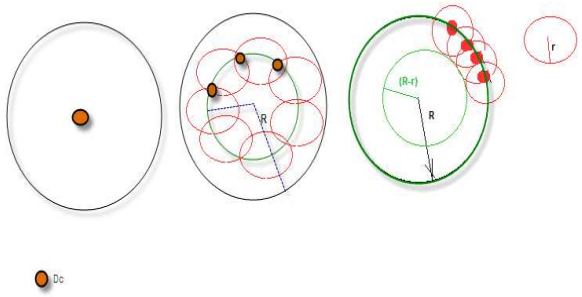
- The sensor nodes are distributed randomly and uniformly over the sensing field.
- The sensor nodes are stationary, After the deployment they still work in their location .
- The sensor nodes are homogeneous in that they have the same communication range and power level (50 joule).
- The sensor nodes are location aware after deployment.
- Data generation rate for all the sensor nodes are the same(150 packets per round).

The objective of our work is to maximize the lifetime of the network, which is the time until a specific percentage of the sensor nodes deplete their energy.

And to minimize energy consumption and maximize the packet delivery rate which is the average number of messages (150 packets per message) delivered successfully to data collector. Our simulation is applied to any specific percentage of sensor nodes.

2.2 Theoretical Analysis of Network Lifetime

We consider the sensing field as a circle with R radius, and the data collector with r radius in order to simplify the analysis. We also assume that the number of sensor nodes are N, and each sensor generates D packets per time unit. While the data collector moves along the perimeters of the sensing field, it will divide it into two partitions: One is covered by the data collector, and the other is uncovered as shown in Fig 1 (c) .We will detect and calculate the area of the two partitions. When the data collector moves along perimeter with radius r, the partition inside the field is a circle with radius R- r. So, the area of covered partition is equal to the area of the entire field subtracted by the uncovered partition (R-(R-r)).



(a) (b) (c)

Fig 1: Dc stationary in center (a), Dc moving along averages (b) Dc moving along perimeters (c).

A stationary data collector scenario

The lifetime of the network is determined by the lifetime of relaying nodes that is less than r m from the data collector. Since sensor nodes are uniformly distributed over the sensing field, we have $\frac{N\pi r^2}{\pi R^2}=\frac{Nr^2}{R^2}$ relaying nodes. With a perfect load balancing and each relaying node transmits $\frac{DNR^2}{Nr^2}=\frac{DR^2}{r^2}$ packets. The Lifetime of relaying nodes is $\frac{Einitr^2}{DR^2Etr}$ time unit.

A mobile data collector moving along the perimeter

The data collector moves along the perimeter of the sensing field so, we have $\frac{\pi(R^2-(R-r)^2N}{\pi R^2}=\frac{(2Rr-r^2)N}{R^2}$ relaying nodes. With a perfect load balancing and each relaying node transmits $\left[\frac{DNR^2}{(2Rr-r^2)N}=\frac{DR^2}{(2Rr-r^2)}\right]$ packets.

The lifetime of a relaying node is $\left[\frac{Einit(2Rr-r^2)}{DR^2Etr}\right]$

A mobile data collector moving along averages
The data collector moves along the averages of the sensing field so, the expected number of relaying nodes is: $\frac{\pi(R^2-(R-2r)^2N}{\pi R^2}=\frac{(4Rr-4r^2)N}{R^2}$
With a perfect load balancing and constant data generation rate, each relaying node will transmit $\frac{DNR^2}{(4Rr-4r^2)N}=\frac{DR^2}{(4Rr-4r^2)}$ packets .
the lifetime of a relaying node is $\left\lfloor\frac{Einit(4Rr-4r^2)}{DR^2Etr}\right\rfloor$ time units.
When we compare the network lifetime of our work with a mobile data collector moves along perimeter and stationary collector, we can see that our mobile scheme is longer than that of a network with a stationary data collector by a factor of 4R – 4r / r. And longer than that of a network with a mobile data collector moving on the perimeter of the sensing field by a factor of 4R – 4r / 2R – r . For the values of r = 100 m and R = 1000 m , the lifetime of our scheme is longer than the stationary collector with a factor of 36 times, whereas longer than the mobile data collector along perimeter with a factor of almost 2 times.
Stationary : Mobile moves along perimeter : Mobile moves along averages : 1 : 19 : 36

2.3 Cluster Head Election Process

The election of cluster heads can affect the performance of the entire network. A well- selected cluster heads cannot only minimize the energy consumption, but also maximize the lifetime of the network.
In our scheme, we classify the data as presented in [5] into two categories: Delay tolerant data and delay sensitive data. For delay tolerant data, we select cluster heads from all the sensor nodes except the relaying nodes that overlap with the trajectory of data collector. Cluster head send directly to the data collector in a single hop; and the sensor node that belongs to cluster head sends directly or through multi hop to its cluster head using Algorithm 2. Relaying nodes which have delay tolerant data send directly to the data collector if it is near, or wait for it. Sensor nodes which have delay sensitive data, will transmit their data directly to the data collector.
We classify cluster head election process into two stages: The first stage is to discover how many neighbors for each sensor node using Algorithm 4. For this, each sensor node broadcasts one hello message to its neighbors. From the number of messages received by a sensor node, each sensor node detects the number of neighbors being included. After that each sensor node also broadcasts another message containing the number of neighbors, and the remaining energy of the sensor node to its neighbors as well. The second stage is to select the cluster head which is the one that has the maximum value of adding the number of neighbors, and the remaining energy divided by 10. In the first case, while the initial energy for all the sensor nodes are the same, only the number of neighbors as a parameter will determine the cluster head. Initially, all the sensor nodes are initiated as cluster heads, but if a sensor node detects the presence of another cluster head from its neighbors (node has a bigger value of number of neighbors + energy /10); it becomes a participating node, and so, the sensor node that has a bigger value is the next hop. We should check all the neighboring sensor nodes before being a cluster head. The sensor node decides locally if it is a cluster head or not. In Fig 2, the sensor node number 2 with value equal to 8 is only a cluster head, and the sensor node number 3 with value 7 is the next hop for sensor node number 4 with value 6 and so on. If two sensor nodes have the same value they will become cluster heads, but each sensor node has only one cluster head.

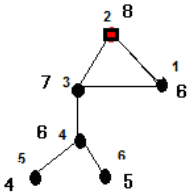


Fig 2: Cluster Head Election

Algorithm1 describes the process of electing cluster heads of the sensor nodes, assuming that each sensor node has only one cluster head that has the maximum value. Every sensor node will receive and transmit one message (150 packets) each round and can be a cluster head after the data collector completes a full round around the sensing field.

Algorithm1: **Cluster Head election Algorithm**:

```
Foreach sensor node Ni do
  If  node Ni didn't receive announcement
          message( nota relying node)
Then
     Foreach sensor node Nj do
       If node Nj is in range of node Ni
          Ni.numberOfNeighbours=
                Ni.numberOfNeighbours +1
          Ni.neighboursValues=Nj.energy/10+
                      Nj.numberOfNeighbours;
       End
     End
 End
End
```

```
Foreach sensor node Ni do
          Ni.Clusterhead = 1;
     Foreach sensor node Nj in neighbors of Ni    do
        If Nj (energy/10 + NumOfNeighbors) >
                Ni (energy/10 + NumOfNeighbors)
           Ni.Cluster head = 0;
          End
      End
End
```

2.4 Sensor Nodes Routing Schemes

When a sensor node has data, it uses two routing schemes to transmit its data to the data collector: One

is delay tolerant routing where the sensor nodes send data packets to their cluster heads in a single hop or more than one hop. This happens when the sensor nodes are away from the data collector trajectory. Cluster heads wait for the data collector to becomes near and picks up data. The sensor node which is connected to cluster head through multi hop utilizes the algorithm for multi hop transmission as in Algorithm 2.

Algorithm 2 : Using Multi Hop Transmission to Deliver Data to DC.

```
broadcastingNodes = set of {Relaying nodes around
                                                    DC}
For each sensor node Nk inside
                        set{ broadcastingNodes}
   For each sensor node Ni
    If ( N(i).nextHop  ~=0 )
      If ( (isInRange (broadcastingNodes (k).loc,
        N(i).loc,broadcastingNodes(k).range) =1))
              If (broadcastingNodes (k).numOfHops
                      + 1 < Nodes(i).numOfHops ||
               (broadcastingNodes(k).numOfHops + 1
                              = Nodes(i).numOfHops
                        &&
               Nodes(i).recievedAnnouncement=1&&
                                    isCloser(Nodes, i,
          Nodes(i).nextHop,broadcastingNodes(k))))
                    Nodes (i).numOfHops =
            broadcastingNodes (k).numOfHops + 1 ;
            Nodes(i).nextHop = broadcastingNodes
                                                (k).id;
             Nodes(i).recievedAnnouncement = 1;
              end
      end
    end
end
```

The cluster heads need to be located by the data collector to updates its trajectory accordingly. Whereas the other type is delay sensitive data, where data packets are sent directly to a relaying node that is currently located near the data collector. When a sensor node has delay sensitive data, it needs first to locate the data collector in order to send data to a relaying node that is currently located near the data collector. Relaying nodes are specified using Algorithm 3.

2.4.1 Delay-Tolerant Routing

In this routing, we have two cases: The first case is when the sensor nodes are away from the trajectory of data collector (more than r meters from DC or not relaying nodes). The other case is when the sensor nodes near the trajectory of data collector (less than r meters or relaying nodes) and have delay tolerant data.

In this routing, the data collector needs first to locate the cluster head locations. To locate the cluster heads, announcements are published vertically from cluster heads to south and north direction to extreme points outside the sensing field (sendingnode.x, network dimension+10) and (sendingnode.x , -10). Then the data collector sends queries using GPSR to east and west direction asking about the current cluster head locations (-10,sendingnode.y) and (network dimension+10,sendingnode.y).The replies are returned at the same path carrying the latest locations of cluster heads. As the data collector detects cluster head locations, it will update its trajectory accordingly and moves to them from the nearest point of its trajectory; pick up data from them in a single hop and come back to its original trajectory.

Any sensor node has delay tolerant data to transmit and away from the data collector trajectory, it will send data to its cluster head in a single hop if it is a neighbor or to the next hop that forward data to cluster head. Cluster heads wait for the data collector to becomes near and picks up data. A sensor node near the data collector trajectory is not included in the clusters; and sends its delay tolerant data directly to the data collector if it is near or waits until it becomes near. After the data collector completes a full round around the sensing field in anti clockwise direction; cluster head process is generated another time to select new cluster heads, so the topology of the network is changed. The announcement message has two fields: id which is the id of a cluster head, and location which is the current location of cluster head. Queries that are sent by the data collector have two fields: destination and location where destination is points outside the sensing field (network dimension+10, DC.y) and (-10, DC.y) and location is the current location of data collector. Replies have also two fields: Destination and locations where destination is the data collector which creates the query, and locations is the current locations of all the cluster heads. Each sensor node keeps a record of the next hop toward the cluster head or cluster head itself. This record (Ci) has two fields: The first field is id which is the id of the cluster head whose sensor node will send to, or it is the next hop. The second field is the next hop location which is a position of a sensor node neighbor or cluster head itself. The process of selecting cluster head node will construct a tree for all cluster head nodes, this tree is rooted at cluster head Ci and involves all the neighboring sensor nodes of Ci. We assume that each sensor node uses one cluster head.

Algorithm 3: The Process of Specifying the Relaying Nodes.

```
For each sensor node Ni do
If  Ni inside DC communication range then
      Nodes(i).numOfHops = 1;
      Nodes(i).nextHop = 0;
      Nodes(i).recievedAnnouncement = 1;
 else
      Nodes(i).numOfHops = NumberofNodes+1;
  End
```

2.4.2 Delay-Sensitive Routing

When any sensor node has delay sensitive data, it will use routing scheme that has two phases, the first phase is to locate the data collector and the second phase is to forward data directly to a relaying node that currently near the data collector.

To locate the mobile data collector, announcements of data collector are published vertically to south and north through a number of sensor nodes in the network toward an extreme points outside the sensing field carrying the current position of data collector. These announcements have three fields: location, time stamp and destination where location is the current location of the data collector, time stamp is the time associated with that location and destination is points outside the sensing field which are field (sendingnode.x, network dimension + 10) and (sendingnode.x,-10). Any sensor node has delay sensitive data to transmit, it sends queries horizontally to east and west directions asking about data collector location. The replies come back carrying the latest location of data collector. These queries have two fields: Sender and destination where sender is the sensor node that sends the query and destination is points outside the sensing field locations (-10, sending node.y) and (network dimension + 10, sendingnode.y). Replies have also three fields : destination, currentloc and time stamp where destination is a sensor node that creates the query, currentloc is the current location of data collector and time stamp is the time associated with that location. After the sensor node gets the latest location of data collector, it will transmit its data to a relaying node that currently inside data collector communication range using GPSR. When the data collector leaves to another location before sensitive data arrives, data packets will be sent to a relaying node that currently near the data collector using GPSR.

We use a similar approach that was used in the context of information discovery in large networks [8], to guarantee that the queries and replies reach sensor nodes which have the latest location of data collector or cluster heads. Announcements that are sent by the cluster heads or data collector traverse the sensing field vertically to south and north to extreme points. The queries and replies that are sent by a sensor node or data collector traverse the sensing field horizontally to east and west to extreme points. The intersection of announcement (column) and query (raw) is approved mathematically in the reference [9].

We use a property of the Greedy Perimeter Stateless Routing (GPSR)[10] to publish announcements, queries and replies. GPSR combines Greedy and Right-Hand rule methods on the connectivity graph. It starts with the Greedy mode where each sensor node forwards message to its neighbor node which is the nearest neighbor to the destination using Algorithm 5. If the message arrives to a local minimum sensor, GPSR switches to the Right-Hand Rule, this process continues until message arrives its destination.

By using GPSR we will publish the announcements of cluster heads or data collector vertically to south and north to destination points outside the sensing field and the queries and replies that are sent by the sensor node or data collector are published to east and west to destination points outside the sensing field.

While replies move along the same path for queries and carry the most recent location of the data collector it has encountered, a sensor node on its way will exchange information about the location of the data collector with the record of that sensor node before being transmitted to the next hop. Each sensor node stores the current information it knows about the data collector location in a record DC which has two fields: Loc and time stamp. All the neighboring nodes of the transmitter will hear the announcement and update their records accordingly because all transmissions are made through a wireless medium.

For summarizing, In the beginning each cluster head sends vertically announcements containing the location of a cluster head in the form of (x,y) using GPSR to points outside the sensing field to north and south. After that the mobile data collector sends queries horizontally containing the current location of data collector and replies are returned at the same path with the locations of all cluster heads. When a data collector knows the locations of cluster heads, It will updates its trajectory to move to the cluster head that is away from its trajectory. When any sensor node has delay tolerant data and not a relaying node, it transmits to a cluster head that belongs to, otherwise send directly to the data collector. When any sensor node has delay sensitive data to transmit, it sends queries to get the latest location of data collector and replies are returned to a sending node with the latest location of data collector. Then, the node will send its data directly to the data collector.

Algorithm 4: Setting Neighbors for a Sensor Node

```
for each sensor node Nodesi do
   Nodes(i).numberOfNodeNeihbours = 0;
   For each sensor node Nodesj do
      if(isInRange(Nodes(i).loc, Nodes(j).loc, 50) ==
                                                  1) then
         Nodes(i).numberOfNodeNeihbours =

         Nodes(i).numberOfNodeNeihbours +1;
       end
   end
end
```

Algorithm 5: Applying GPSR to Find Nearest Neighbor

```
recievingNodeID = 0;
closerNeighbourFound = 0;
if(N(sendingNodeID).numberOfNodeNeihbours
                                              ~=0)
    neighboursIDs=
     N(sendingNodeID).nodeNeighbours;
     minDistance=
    distance(destinationLoc ,N(neighboursIDs(1))
                                             .loc , 0);
            nodeToSendTo = neighboursIDs(1);
        for each neighboursIDs i = 2
            nextNeighbourDistance =
                distance(destinationLoc ,
                N(neighboursIDs(i)).loc , 0);
        if( nextNeighbourDistance  < minDistance)
         inDistance = nextNeighbourDistance;
```

```
            nodeToSendTo = neighboursIDs(i);
        end
    end
    distanceBetweenSendingNodeAndDestination=
distance(destinationLoc ,
                N(sendingNodeID).loc , 0);
    closerNeighbourFound = 0;
    recievingNodeID = Nodes(nodeToSendTo).id;
end
end
```

2.5 Experimental Results

In this simulation we compare our proposed scheme with two other schemes :a stationary scheme that has a stationary data collector and a mobile scheme that has a mobile data collector moves along perimeters of the sensing field [5]. In stationary scheme, data collector is placed at the center of the sensing field and all the sensor nodes send to it using multi hop transmission with the shortest path routing in terms of hop. If a sensor has two paths with the same number of hops, the shortest distance is taken. In the mobile scheme, the data collector moves along the perimeters of the sensing filed to collect data from the sensor nodes.

Our scheme has the objective of maximizing the network lifetime and minimizing the delay by maximizing the packet delivery rate to the data collector. .We also show the difference between the network lifetime of our scheme, the mobile scheme network lifetime and the stationary scheme.

To compute the energy consumption we use the general energy consumption model that presented in [2] which can be as follows.

$$ETr(r, b) = b \times (Eelec + Eamp \times r^{\gamma}) \quad (1)$$

$$ERc(b) = b \times Eelec \quad (2)$$

where ETr (r, b) is the energy spent to send b bits over r m , ERc(b) is the energy spent to receive b bits, Eelec is the energy spent by the transmitter or receiver to send or receive one bit., Eamp is the energy spent by the transmission amplifier for one bit and γ is the path-loss exponent. r is the communication range of sensors.

Table 1: Simulation Parameters

Parameter	Value
Eelec	50 nJ/bit
Eamp	0.1 nJ/bit/m2
Packet length	512 bits
Γ path-loss exponent .	Γ = 2
Initial energy	50 j
Data types	50%Delay tolerant and 50% delay sensitive

We use in our simulation the network sizes of 200,400,600,800 and 1000 sensor nodes that are randomly and uniformly distributed in fields of 300*300,400*400,500*500,600*600,700*700 m^2 fields, respectively. we test 10 instances and take the average for each network size. We use the simulation parameters as in Table 1.The communication range of the sensor nodes and of the data collector is set to 50 m and the data collector is not resource constrained . In our simulation. every sensor node generates 150 packets/round.

Fig 3 shows the comparison of lifetime between our scheme and two other schemes. The lifetime of a network with our scheme (cluster) is at least 7 times longer lifetime than that of a network with a stationary data collector and also longer than the mobile scheme with data collector moving along perimeters with a factor of almost 2 times. This is due to the mobility of data collector and the increase in the number of relaying nodes in our scheme distributed around the data collector trajectory as in Table 2. It is also resulted from our clustering algorithm where the sensor nodes send their data in a single hop to data collector, and, in turn, the sensor node sends to its cluster head with shortest path. Fig 4 shows a comparison of the average energy consumed per bit between the three schemes for different number of nodes. Besides load balancing, data collector mobility and clustering algorithm bring a reduction in the overall consumed energy.

From Fig 5 , it is observed that the average number of sensitive messages (one message equal to 150 packets) delivered successfully to the data collector is similar in both our scheme (cluster) and mobile scheme because the delay sensitive data is sent directly to the data collector, but our scheme has a shorter path to data collector; so, it has a little bit more messages than the mobile. Fig 6 shows that the average number of tolerant messages delivered successfully to the data collector in our scheme (cluster) is improved with 150 % over the mobile scheme. This is a direct result of the increase in the number of relaying nodes which have delay tolerant data to send, and also it's a result of applying clustering with our algorithm where cluster heads transmit in a single hop. The number of relaying nodes with our scheme (cluster) and mobile scheme is proportional to the size of the network, this means that when a network size increase, the amount of data generated over the network is also increased, which leads to increase the number of relaying nodes.

Table 2: Number of Relaying Nodes in Each network size

Number of relaying	Number of nodes			
	400	600	800	1000
stationary	21	12	27	17
Mobile	177	215	250	281
Ours	256	318	386	419

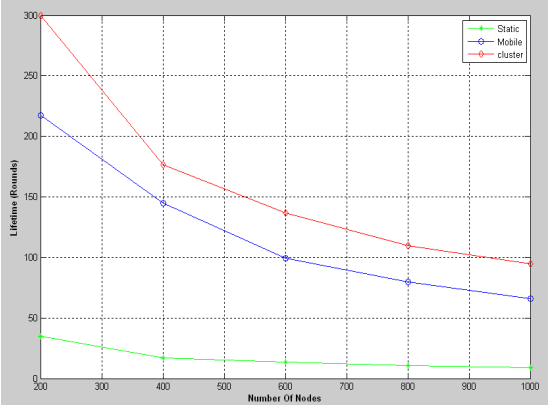


Fig 3: Lifetime comparison of three schemes.

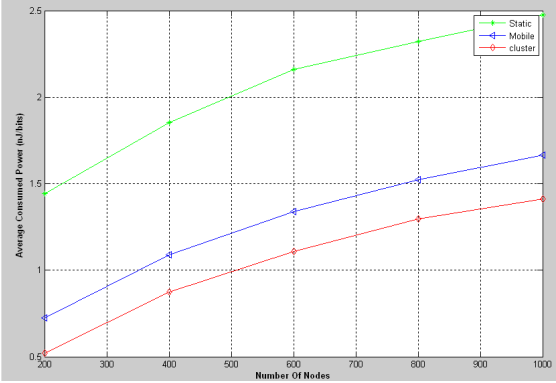


Fig 4: Energy consumption Comparison of three schemes.

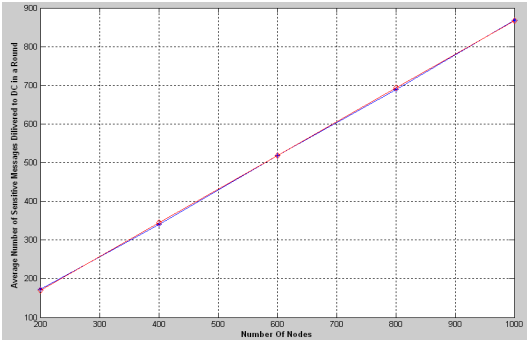


Fig 5 : Comparison of Average Number of Sensitive Messages delivered to Data collector(DC) Per round.

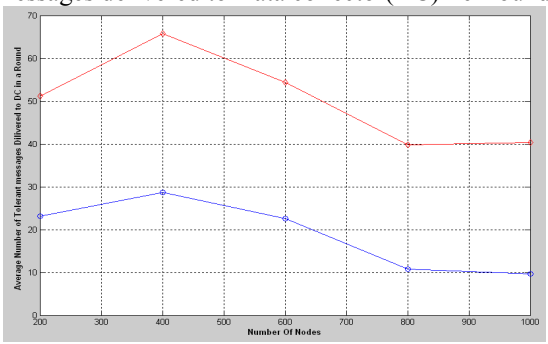


Fig 6 : Comparison of Average Number of Tolerant Messages delivered to Data Collector Per round.

3 Conclusion and Future Work

We propose a scheme for data collection using jointly mobility of data collector and clustering algorithm .With mobility we distribute the load around the network and using clustering we also distribute the load with rotation of cluster heads and finally the lifetime of network is maximized. There is no synchronization between sensor nodes and data collector and the computation complexity in our work is also simple. From simulation results we show that our scheme has the potential to maximize the lifetime of wireless sensor network better than the scheme with mobile data collector moving along boundary of sensing field. We are currently study the same scenario with some modifications to GPSR to decrease the transmissions to give a better results than previous and evaluate this scenario with more metrics.

4 References

[1] S. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, "Energy efficient schemes for wireless sensor networks with multiple mobile base stations," in Proc. The IEEE Global Telecommunications Conference (GLOBECOM), December 2003.

[2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy efficient communication protocol for wireless microsensor networks," in Proc. The 33rd Annual Hawaii International Conference on System Sciences, January 2000.

[3] B. Wang, D. Xie, C. Chen, J. Ma, and S. Cheng, "Employing mobile sink in event driven wireless sensor networks," IEEE Vehicular Technology Conference VTC, 2008.

[4] Z. M. Wang, S. Basagni, E. Melachrinoudis, and C. Petrioli, "Exploiting sink mobility for maximizing sensor networks lifetime," Proceedings of the 38th Annual Hawaii International Conference on System Sciences HICSS, 2005.

[5] Waleed Alsalih, Hossam Hassanein, and Selim Akl "Routing to a mobile data collector on a predefined trajectory" in the IEEE ICC 2009 proceedings.

[6] J. Luo and J.-P. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," in Proc. The 24th IEEE International Conference on Computer Communications (INFOCOM), 2005.

[7] A. Kansal, A.A. Somasundara, D.D. Jea, M.B. Srivastava, D.Estrin, "Intelligent fluid infrastructure for embedded networks, in: Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services – MobiSys'04, Boston, MA, USA, June 06–09, 2004, pp. 111–124.

[8] X. Liu, Q. Huang, and Y. Zhang, "Combs, needles, haystacks: balancing push and pull for discovery in large-scale sensor networks," in Proc. The 2nd International Conference on Embedded Networked Sensor Systems, 2004.

[9] Waleed Al-Salih, "Mobile data collectors in wireless sensor Networks ", A thesis submitted to the School of Computing in conformity with the requirements for the degree of Doctor of Philosophy, Queen's University Kingston, Ontario, Canada April, 2009.

[10] B. Karp and H.Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in Proc. The Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 2000.

A Geometric Tiling Algorithm for Wireless Sensor Networks

Adam P. MartinezÀ, Timothy NorfolkÀ, and Kathy J. Liszka ÿ

†The University of Akron
Department of Mathematics
norfolk@uakron.edu

ÿ *The University of Akron*
Department of Computer Science
liszka@uakron.edu

Abstract

The cover generation problem is relevant to the problem of creating large-scale wireless sensor networks with short-ranged sensor nodes that may not be capable of transmitting to the base station. Quickly and efficiently placing relay nodes allows the sensors to save on battery power and transmit information back to the base station via the relay nodes. Placing a minimal cover of relays is at least an NP-hard problem. We present a geometric tiling algorithm to construct an approximation to a minimal covering set in O(n) time.

Keywords: geometric tiling, minimal covering sets, wireless sensor networks

1. Introduction

There are a wide variety of polynomial time approximation schemes (PTAS) that can approximate solutions to the minimum geometric disk cover (MGDC) problem, but none in current literature can do so in O(*n*) runtime. We present an algorithm that computes an approximation to the MGDC problem with reasonable disk-to-point efficiency for many instances of the problem in linear runtime. The motivation for this research is to computer optimal designs for building wireless sensor networks (WSNs). Some WSN structure problems can be cast as MGDC problems. While for many applications long runtimes of current algorithms are not an issue, for time-sensitive problems or very large regions with large sets of sensor nodes (SNs), computing a covering set of relay nodes (RNs) can take unreasonably long.

Some applications can tolerate tens of hours computing an optimal networking solution that requires as few relays as possible. However, not all networking environments have the luxury of unlimited design and setup time. For time-sensitive applications, computing a fast and reasonably accurate solution to a covering set of a network can achieve a ñgood enoughò solution that will save lives. The network will be more costly, but it can start being built immediately. This kind of algorithm could be useful for providing real-time logistical and tactical information to moving front-line military units and ensuring that search-and-rescue teams have real-time information. Because these kinds of environments do not tolerate time delays, a less efficient network now is far more valuable than a more efficient network later.

This paper presents a geometric tiling algorithm for approximating a minimal covering set in the context of a two-tiered, single-hop WSN. This can alternately be described as an approximation scheme for the MGDC problem. The next section gives background on the problem along with related research. A formal description of the geometric tiling algorithm and an analysis of its performance is given in section 3. Analysis and experimental results are given in section 4. Conclusions and future work are given in the final section.

2. Background and Related Research

WSNs consist of a set of sensor nodes that collect information and wirelessly communicate with one another. There is either a Base Station (BS) that aggregates the information in the network or an outside access point to the network. There may or may not be RNs that act as network gateways for SNs within a small region around them. The presence of RNs determines if a WSN is *one-tiered* or *two-tiered* [1]. One-tiered WSNs have only sensor components. SNs are either deliberately placed or randomly dropped into a region, and an appropriate way to route data through the network must be found.

There has been significant research into algorithms for generating two-tiered covers of WSNs within the past decade. Two-tiered WSNs have not only a set of sensors in a region, but also have a set of relay nodes that act as network gateways for sensors in a small region around them. These algorithms usually assume a random distribution of SNs in a given region. A layer of relay nodes is placed such that the relay layer forms a cover over the sensors. The SN's sole purpose is to gather information and forward it to a local RN. For many applications the SNs are designed to be built as cheaply as possible, and thus do not have the battery power and design parameters to transmit data long distances. The RNs collect data from the SNs within a small region and relay the data either directly to or through one another back to a BS. Each RN in a single hop two-tier WSN has a direct connection to the BS. An example single-hop two-tiered network is shown in Figure 1.

Single-hop two-tier WSNs only require that the SNs transmit to the RNs, and do not require that the RNs be able to transmit to one another. The only requirement on the set of RNs is that it forms a covering set of the SNs. It is assumed that either the data will be consolidated at the RN for later collection or each RN has some capability of transmitting its information back to

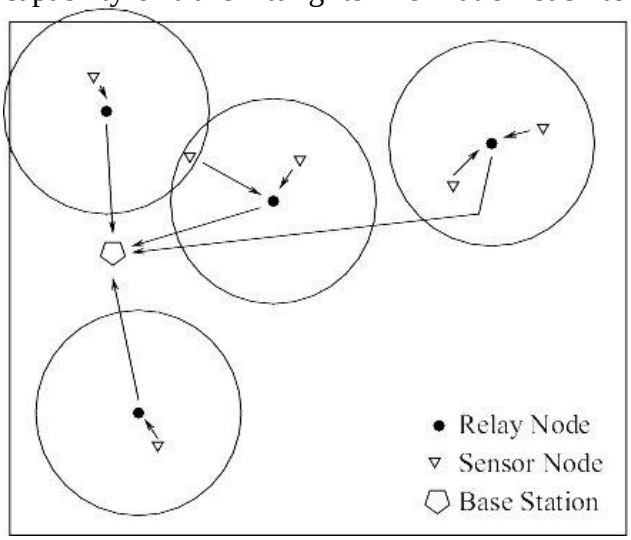


Fig. 1. A triangular grid of RNs

a BS. In the latter case, the RNs that are more distant from the BS will deplete their power more quickly than those further away [1].

Despite this drawback, this network architecture is still useful for networking environments where RNs have satellite uplinks, long range directional wireless communication or landline access. There is extensive study on this problem in terms of the minimum geometric disk cover (MGDC) and discrete unit disk cover (DUDC) problems.

2.1 Minimum Geometric Disk Cover

Given a region D containing a set P with n points, generate minimal covering set of unit disks C such that for each $p \in P$, $\exists\, c \in C$ such that $p \in c$.

2.2 Discrete Unit Disk Cover

Given a region D containing a set P with n points and a set of unit disks D, select a minimal covering set of unit disks $C \subseteq D$ such that for each $p \in P$, $\exists\, c \in C$ such that $p \in c$.

2.3 Comparison

The MGDC algorithm allows disks to be placed anywhere within the region, while the DUDC problem only allows disks to be placed in specific locations. Both the MGDC and DUDC problems have been proven to be NP-complete [2], but both also allow polynomial time approximation schemes. A PTAS generates a solution to an NP-Hard problem in polynomial time that is no more than some constant multiple of the optimal answer. A wide variety of PTAS have arisen for both of these problems. Many algorithms for DUDC have been proposed that generate solutions of no more than some constant multiple greater than one of the optimal solution in reasonable time [3, 4, 5, 6]. By comparison, algorithms for the MGDC problem generally require much longer runtimes, but can guarantee an arbitrary (1+ε) level of accuracy to the optimal solution of disks placed anywhere in the region. Depending on the accuracy required and the algorithm used, MGDC and DUDC PTAS can be as fast as $O(n^2)$ or slower than $O(n^{100})$. Some connected cover algorithms from the multiple-hop two-tiered problem, such as the 2CRNDC algorithm, are very similar to MGDC algorithms; only as a last step do they guarantee connectivity [7]. Some of the more recent work in this problem includes research by Liao and Hu [8], of which a modified algorithm is featured later in this work as a point of reference for the algorithm we present. Liao and Hu's algorithm build off of general set-based PTAS for approximating the MGDC [9].

3. Formulation of the Algorithm

We present an algorithm for generating a reasonably small unit disk cover of a set of points in $O(n)$ time [11]. The approach for this algorithm relies on the uniformity of a triangular grid. Consider the problem of finding the most efficient cover of a large but finite plane using

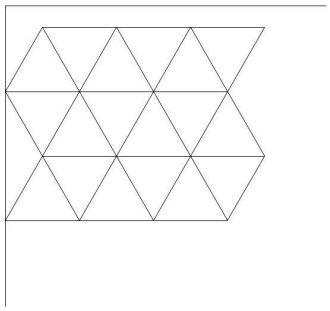


Fig. 2. A triangular grid of RNs.

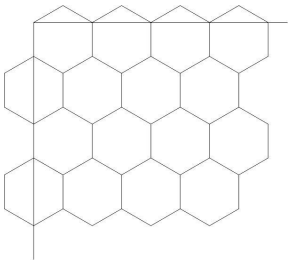


Fig. 3. A hexagonal cover of RNs.

disks of radius 1. Pompili et al. showed that the most efficient regular cover is a triangular grid of disks as in Figure 2, with a point-to-point transmission distance of ã3 [10]. However, our problem formulation does not require that we cover the entire region. We only need to provide a covering set for a set of n points in the region, representing SNs. We abstractly generate a cover of the region by overlaying a tessellation of hexagons of circumradius 1 with centers at each point on a triangular grid of edge length ã3. The hexagons in the region appear as in Figure 3. A hexagon of circumradius 1 is a regular hexagon inscribed in a circle of radius 1. Potential RN locations are only at the points on the triangular grid. Each RN only receives messages from SNs within the RN's corresponding hexagon. The algorithm iterates through the n SNs and adds the nearest point on the triangular grid to a solution set. By placing a unit disk at each point in the solution set, we produce an approximation to the minimal unit disk cover of the n points.

Problem Statement: given a region R containing a set P with n points, generate an

approximation C to the minimal covering set of unit disks such that for each $p \in P$, $\exists c \in C$ such that $p \in c$.

Table 1 provides the notations used. Given a region filled with n SNs, we approximate a covering set. For the purpose of this formulation we will assume we are given a square region containing the SNs. In practical problems, the region would be defined as the minimal square that contains the set of SNs. Label this square region R with side length s. Our objective is to approximate the minimal cover of the SNs using a triangular grid of RNs.

The RN-SN transmission range r forms a convenient non-dimensional scaling for this problem. We call the non-dimensionalized region R_n, with side length s_n. In this region, the RN-SN transmission range is 1. By scaling all distances involved by r, the algorithm generates a cover for any size region efficiently.

We abstractly tessellate the non-dimensionalized region R_n with hexagons of circumradius 1. The RNs at the centers of the hexagons form a triangular grid. Each SN in R will be mapped to points in D_n via the transformation

$$\left(x_(n,p), y_(n,p)\right) = \left(\frac{x_p}{r}, \frac{y_p}{r}\right), \qquad (1)$$

where (x_p, y_p) is the location os the SN p.

Given any square or rectangular region, we may orient the hexagons as in Figure 6, such that the hexagons fit neatly in the upper left corner of the region with minimal waste. We can then compute the coordinates of any RN. Tessellating the square region R_n this way requires no more than $\left\lceil \frac{s_n}{\sqrt{3}} \right\rceil + 1$ columns and $\left\lceil \frac{s_n}{1.5} \right\rceil + 1$ rows of hexagons of circumradius 1. Implementing the algorithm does not actually require the generation and storage in memory of the entire set of hexagons, merely the conceptual knowledge that we have overlaid it on the region.

We iterate through each SN and select the nearest hexagon in the grid. The regular nature of the tessellation makes finding the nearest RN a constant time arithmetic process. These RNs locations do not need to be pre-computed and then selected, as they can be computed on the fly using arithmetic and rounding.

Table 1. Variable definitions.

Variable	Description
R	The 2D region.
R_n	The non-dimensionalized 2D region.
s	The side length of R.
s_n	The side length of R_n.
r	RN ï SN transmission range.
C	An approximation to the minimal disk cover.
P	The number of sensor nodes.
n	The number of SNs.
x_p	The x-coordinate of sensor p.
y_p	The y-coordinate of sensor p.
$x_{(n,p)}$	The non-dimensionalized x-coordinate of sensor p.
$y_{(n,p)}$	The non-dimensionalized y-coordinate of sensor p.

The RNs, as they are set up in this problem, form a triangular grid. Each row on the triangular grid has the same y-coordinate. Every second row has an offset x-coordinate. We can immediately eliminate all but two potential candidates for the nearest RN to a SN by simply looking at the coordinates of the sensor. The sensor will fall between two rows of relays. Counting from the top, odd numbered rows will have a horizontal offset of ã3 / 2 from the even rows. Within each row, the x-coordinate of the SN will be closer to either the RN on its left or on its right. Each row then has a closest RN to the SN, and the closer of these two RNs is chosen to cover that SN.

We create a *selected relays* matrix M of booleans that stores whether or not the j^{th} RN in the i^{th} row of RNs must be selected to form a cover. This has the disadvantage of taking up a large block of memory by requiring a matrix of

$\left(\left\lceil \frac{s_n}{\sqrt{3}} \right\rceil + 1\right)\left(\left\lceil \frac{s_n}{\sqrt{1.5}} \right\rceil + 1\right)$ booleans, but avoids writing duplicate RNs to the solution set *C*. We then compute the location of the j^{th} RN in the i^{th} row for each true in *M*. These RNs make up *C*.

To map the locations of the RNs in R_n back to the region *R*, the corresponding location in *R* for an RN at x_n,y_n in R_n is

$$(x,y) = (rx_n, ry_n) \quad (2)$$

a set of points such that when we place an RN at each of these points, we have a cover of the SNs in *R*. Each SN will be within *r* units of distance of the nearest RN. This is an approximation to the minimal cover of disks of radius *r*.

Due to the properties of MGDC problem and the formulation of the triangular grid algorithm, there is no simple way to compare precisely how accurate of a solution the algorithm provides to the optimal solution. The MGDC problem is NP-complete, and so finding the optimal disk cover takes an unreasonable amount of time to compute for any random simulation. Additionally, the only mathematical bound the algorithm gives to the efficiency of its cover is that it is the most globally efficient layout of relays on the plane.

4. Analysis and Preliminary Experiments

The algorithm was run on a desktop machine using MATLAB R2011b. A set of points are generated using a uniform distribution in a square region of side length s = 10. This side length was chosen as it was large enough to see noticeable differences in solutions and computational times, as well as being small enough to compute in a reasonable time. The algorithm computes the locations of a triangular grid of unit disks that covers the region, then finds approximations to the minimal cover of unit disks on that grid.

Sensor densities of (n/s^2 = .1, .25, .5, .75, 1, 2) were considered to determine each the algorithm's response to networks of varying densities. An example of the output for n/s^2 = .5 is shown in Figure 4. Results of the simulation shows that the algorithm performed in O(n) time for simulations of all sizes of n. Memory usage was not a concern until the input data size grew to a very large n. Figure 5 shows the triangular grid algorithm across the trials we ran.

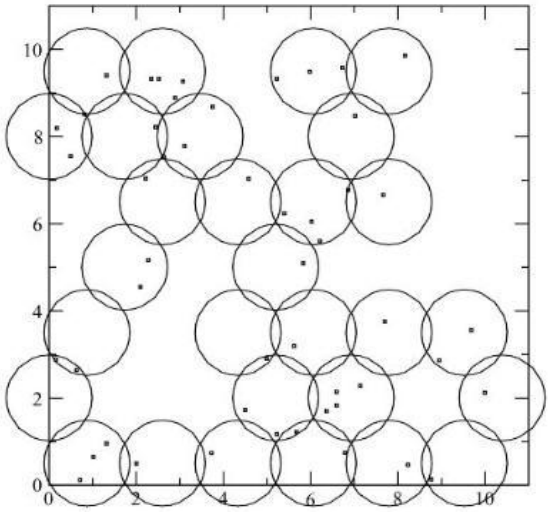


Figure 4. The triangular grid algorithm.

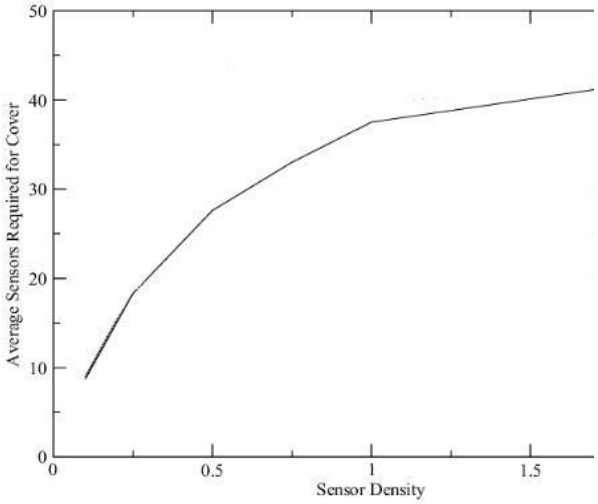


Figure 5. Average relays to cover sensor networks of various densities.

5. Conclusions and Future Work

We presented an O(n) triangular grid algorithm for approximating the minimum geometric disk cover of a set of points in a region. While the algorithm presented is not an epsilon-approximation, its speed and practical performance for generating approximations to an NP-Complete problem in linear time makes it suitable for some applications including guaranteeing coverage in dense or rapidly changing wireless sensor networks. For emergency situations and military applications, time is the primary issue for building an effective network, not cost. This algorithm provides a method for quickly generating covering sets of almost any size network.

The triangular grid algorithm could be improved. A better approximation could be found by culling relays from the grid using some simple enumerative techniques to identify unnecessary relays. Additionally, search techniques could be used to generate connected covers by finding unconnected spaces and connecting them with a shortest path of relays. While this would increase the runtime of the algorithm, it could provide fast solutions to the multiple-hop WSN problem. The runtime increase would likely be dominated by the runtime of the search algorithm. Common search algorithms, such as breadth-first and depth-first searches, are O(n^2). The algorithm would also generate an acceptable starting point for iterative methods for calculating a minimal cover.

Additionally, the algorithm deserves a comparison to an MGDC algorithm that is not restricted to a triangular grid. While the triangular grid algorithm provides adequate solutions on the grid, at this time we are looking at how the algorithm compares to a true MGDC algorithm in terms of RN-SN ratio.

6. References

[1] Xu, K., Hassanein, H.,Takahara, G. and Wang, Q., Relay node deployment strategies in heterogeneous wireless sensor networks. IEEE Transactions on Mobile Computing, 9(20:145-159, February 2010.

[2] Masuyama, S. and Ibaraki, T. and Hasegawa, T., The computational complexity of the m-center problems on the plane. *IEICE Transaction*s, 1981, E64-E, pp. 57-64.

[3] Clinescu, G. and Mndoiu, I. and Wan, P. and Zelikovsky, A., Selection forwarding neighbors in wireless ad hoc networks. *Mobile Networks and Applications*, 9:101-111, 2004.

[4] Carmi, P. and Katz, M. and Lev-Tov, N., Covering points by unit disks of fixes location. In Takeshi Tokuyaa, editor, *Algorithms and Computation*, volume 4835 of Lecture Notes in Computer Science, pp. 644-655. Springer Berlin/Heidelberg, 2007.

[5] Claude, F. and Dorrigiv, R. and Durocher, S. and Fraser, R. and Lpez-Ortiz, A. and Salinger, A., Practical Discrete Unit Disk Cover Using an Exact Line-Separable Algorithm, *Algorithms and Computation*, vo, 5868 of Lecture Notes in Computer Science, pp. 45-54, Springer Berlin/Heidelberg, 2009.

[6] Das, G. and Fraser, R. and LÃ²pez-Ortiz, A. and Nickerson, B., On the discrete unit disk cover problem. Naoki Katoh and Amit Kumar, editors, WALCOM: *Algorithms and Computation*, vol. 6552, Lecture Notes in Computer Science, pp. 146-157. Springer Berlin/Heidelberg, 2011.

[7] Hao, B. and Tang, H and Xue, G., Fault-tolerant relay node placement in wireless sensor networks: formulation and approximation. In *Workshop on High Performance Switching and Routing*, pp. 246-250, 2004.

[8] C. Liao and S. Hu, Polynomial time approximation schemes for minimum disk

cover problems., *Journal of Combinatorial Optimization*, 20:399-412, 2010.

[9] Nieberg, T. and Hurink, J. and Kern, W., Approximation schemes for wireless networks, *ACM Trans. Algorithms*, 4:49:1-49:17, August 2008.

[10] Pompili, D. and Melodia, T. and Akyildiz, I. F., Three-dimensional and to-dimensional deployment analysis for underwater acoustic sensor networks, Ad Hoc Networks, 7(4): 778-790, 2009.

[11] Martinez, A., A Geometric Tiling Algorithm for Approximating Minimal Covering Sets, a Master's thesis presented to the Graduate Faculty of The University of Akron, December 2011.

DISTRIBUTED SENSOR NETWORKS: AN APPROACH

Polzonetti Alberto, De Angelis Francesco, Marcantoni Fausto
School of Science and Technology
Camerino (ITALY)
name.surname@unicam.it

Abstract - *The emerging of global computing is changing the classical system model of computation - stand-alone and isolated - to an open, global and connected one, by showing the needs to suitably manage critical aspects of the concurrent and distributed computation in heterogeneous environment. In this paper, we present a regional project that aims to develop an open service platform in order to integrate heterogeneous sensing information in order to facilitate the administrative decision-making processes*

Keywords: *sensor network, distributed systems*

1 Introduction

During last years, several activities have taken place in order to monitor environmental changes with the aim of supporting the ordinary activities in the territory administration. This has brought the deployment of several sensing systems each one with own features and peculiarities. These systems work correctly in their specific application, but usually they are not reusable in different applications due to the lack of interoperability. Moreover, as sensor networks become more pervasive there emerges a need for interfacing applications to perform common operations and transformations on sensor data.

In this contribution, a regional project is showed, which aims to meet this issue. The project intends to study and experience a novel platform able to reorganize developed solutions (e.g. geographic information systems, sensor networks) in a new worldview that aims to improve and integrate the actual state of art in this sector.

The platform will employ monitoring systems and sensor networks to keep track of the status of the environment under a new point of view where all can be controlled and managed using a well-defined access way that facilitates the management of heterogeneous data provided by different sensing systems: the main goal is to obtain a wide open system with an elevate degree of interoperability among monitoring services.

2 Project Overview

Our project is attempting to analyze and deploy an open software platform able to support the realization, the maintenance and the evolution of a territorial control system: the interest is not only at research level, but it aims to apply this research in real case study involving civil protection organizations that will provide their already existing sensing systems.

The first phase of the project regards the study of this platform and its basic services, focusing on the possible problems and on the characterizing aspects that can simplify the management and the accessibility of environmental information useful for administrative decision making. In order to increase the interoperability, standard interfaces for web applications and services must be defined. This study will provide the individuation of basic services for applicative fields, such as the geographic one, the workflow one, the sensors one, the knowledge one.

The second phase regards the case study and the deployment of a prototype for the security and the monitoring of the territory. The basic idea is to experiment the platform to integrate existing sensing devices and web technologies for the gathering of territorial data to detect critical events.

3 The open special platform

The purpose of the former phase is to study the available methodologies and technologies, that can be used to support this open platform. This study must relate the main platform features with the actual specifications defined by the standardization committees, in particular standards regarding web technologies (e.g. web services [2]), communication protocols (e.g. ZigBee communication protocol suite [10]) and the interaction patterns (e.g. service interaction patterns).

In this phase, we need to define a service platform that integrates primary services, in a homogenous way, for the following application fields:

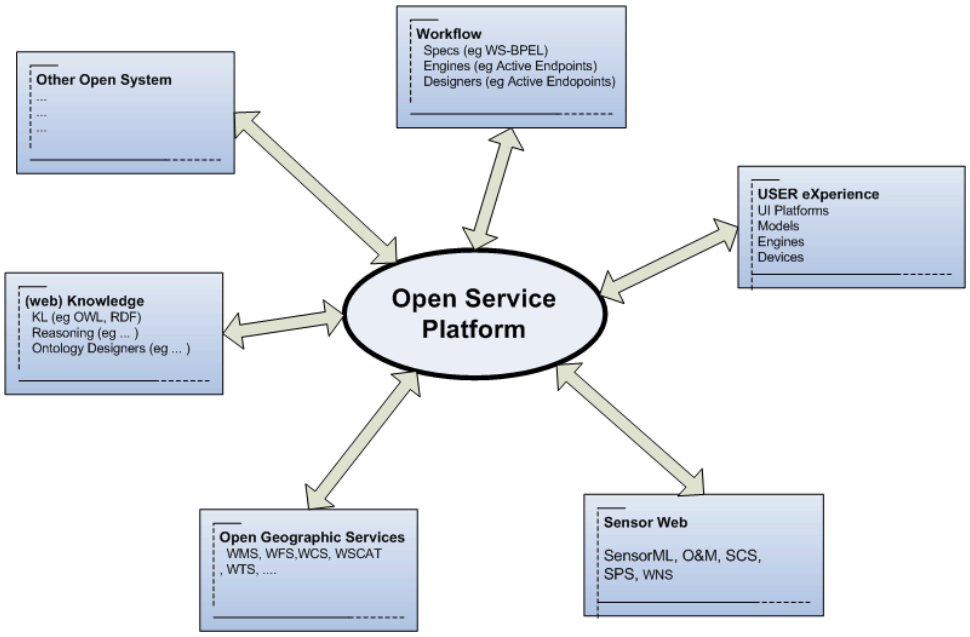


Figure 1 Open Service Platform

Geographic services: the aim of this services is to allow the access to geographic archives in order to obtain a virtual spaces modeling, a geographic data querying and a geographic-based data interpretation;

Services for the applicative cooperation based on synchronous and asynchronous collaboration profiles for the exchange of the informative flows between monitoring systems (e.g. workflow and Publish & Subscribe service)

Sensor services for the control sensor device and the management of distributed sensor information;

Knowledge services: that provide access to or operate over knowledge resources, like rule engines and automated reasoners;

Support services for the management of the technological infrastructure. These are centralized and demanded to specific operative points that also monitor the systems and the applications, control the security conditions and, in general, the operative and administrative applications needed for the correct operability of the informative system.

3.1 Geographic services

Nowadays national governments and international organizations are moving towards this direction and the growing adoption of infrastructures of standardized spatial data is a demonstration. INSPIRE European community aims to standardized territorial information and services simplifying the integration [3]:

"Art. 1 - … In order to bring about such integration, it is necessary to establish a measure of coordination between the users and providers of the information so that information and knowledge from different sectors can be combined"

"Art. 4 - The Infrastructure for Spatial Information in the European Community (Inspire) should assist policy-making in relation to policies and activities that may have a direct or indirect impact on the environment."

In this area, our interest is to simplify the representation of the territorial information, to share and to use them according to the specifications given by the Open Geospatial Consortium (OGC) [7].

What we have intention to do is a survey on the existing services to understand the current state of art and the needed improvements. Our approach aims to come to a point where the geographic information will not be anymore the object of the service, but they will be a "building block" for more "intelligent" systems.

3.2 Applicative cooperation

We define a workflow as "a block of activities correlated by different relationship typologies" and so we are interested in workflow management systems [4], that are able to describe the modeling, the definition, the execution and the monitoring of the business processes. The main aim is to identify a new

methodology to integrate workflow with different cooperation paradigms, like Publish & Subscribe, Tuple Space-based and so on. This is due to the heterogeneity of the involved system that can have limited resources (e.g. memory, energy, CPU). For this reason a more light cooperation way is needed in order to preserve the lifetime of this system and monitoring of critical events.

3.3 Knowledge domain

The increasing attention on semantic technologies brings a novel and efficient approach to share data among several data sources in a distributed networked environment. In this way, different applications can be integrated dynamically using a shared Knowledge based on ontology.

However, the main problem regarding Geographical Information Systems (GIS) [1] is the lack of integration among heterogeneous systems. An efficient solution is represented by the ontology-based GISs, a schema that brings a dynamic and open integration for different systems. Ontology solves this issue with a mechanism to specific in an explicit way the concepts to use in the applications: this helps especially the information sharing, because otology can specify the nature of the involved entities. What we propose is to create ontology and meta data based GISs, that are able to associate, extract and elaborate information expressed in different languages. In particular, ontology is a key conceptual tool for gathering, specifying, storing, maintaining, managing and representing explicit and implicit knowledge contained in a complex domain like GIS.

The advantages that we want to obtain using this approach are:

Physic and logic independence among system entities;

System scalability;

Improvement of information management;

Better Interoperability among GISs;

System customize based on user profiles;

3.4 Sensor services

Last years have seen a growing interest in the sensor networks, that represent a mature technology for the development of low-cost and low-consume distributed applications which are used to exchange environmental information within wide geographical areas. A sensor network is an autonomous collection of mobile nodes that communicate over wireless links. In general, a sensor is an autonomous unit that is able to elaborate a set of operations, to sample/interact with the environment and to cooperate with other devices. But these networks usually suffer from significant robustness issues due to limited resources, such as battery, memory and CPU.

Sensor networks have been widely used in the development of decision support systems [6], in particular, for environmental monitoring applications. However, this increasing deployment has brought to the development of different information management and storage mechanisms causing an impracticable cooperation among sensor devices provided by different vendors. For this reason, several initiatives have taken place in order to develop an intuitive and expressive approach for representing and integrating data provided by different sensor typologies, for example:

SINA: Sensor Information Networking Architecture [9];

SenseWeb [8]; IrisNet: Internet-scale resource-intensive sensor network services [5].

The main aim of this section is to develop a global and heterogeneous sensor network that allows information sharing using interoperable and standardized interfaces. In this direction, we consider the Sensor Web Enablement initiative proposed by OGC a possible approach for our problem.

4 Case Study: « Protezione Civile »

The second aim of our project is the deployment of a prototype for the security and the monitoring of the Le Marche region. Our experimentation will provide:

The classification and the specification of possible critical events and risk scenarios in order to identify a procedure to detect them;

The identification of the existing available monitoring systems;

The realization of prototypes for Sensor Web Services;

The verification of the prototype functionality;

The realization of a demonstrative application for the environmental monitoring.

The features that our system should have are: open-source code, tools to support a development community; interoperability with geographical standards and a customizable GUI.

4.1 The Scenario

The civil protection is a system that operates, in the regular scenario, to develop predictive analysis in order to monitor and prevent risks that can threaten the territory and, in the emergency scenario, to provide support for the emergency management. Civil protection system is a complex system, in which cooperate several institutions, operative structures and public and private organization.

In Italy, the "Protezione Civile" is an organization that deals with prevention, management and resolution o possible emergencies in a regional and national area. Its goal is to safeguard the lives, goods and buildings integrity from any calamities. In order to control the environment status, the "Protezione Civile" is supported by the RMIPR (Rete Metro Idro Pluviometrica), a regional network that monitors weather conditions: this net provides eleven typology of sensor distributed in the whole area, these sensors communicate each other by radio channels; our task is to extend this network with other private monitoring systems that are deployed in the region. This is the first step to realize an active monitoring system for the territory, that will adopt other than the use of the sensor networks and system already active, a new mobile system for video-acquisition, that should give the possibility to measure, recognize and rebuild in a 3D scenario, the environment according to the necessity.

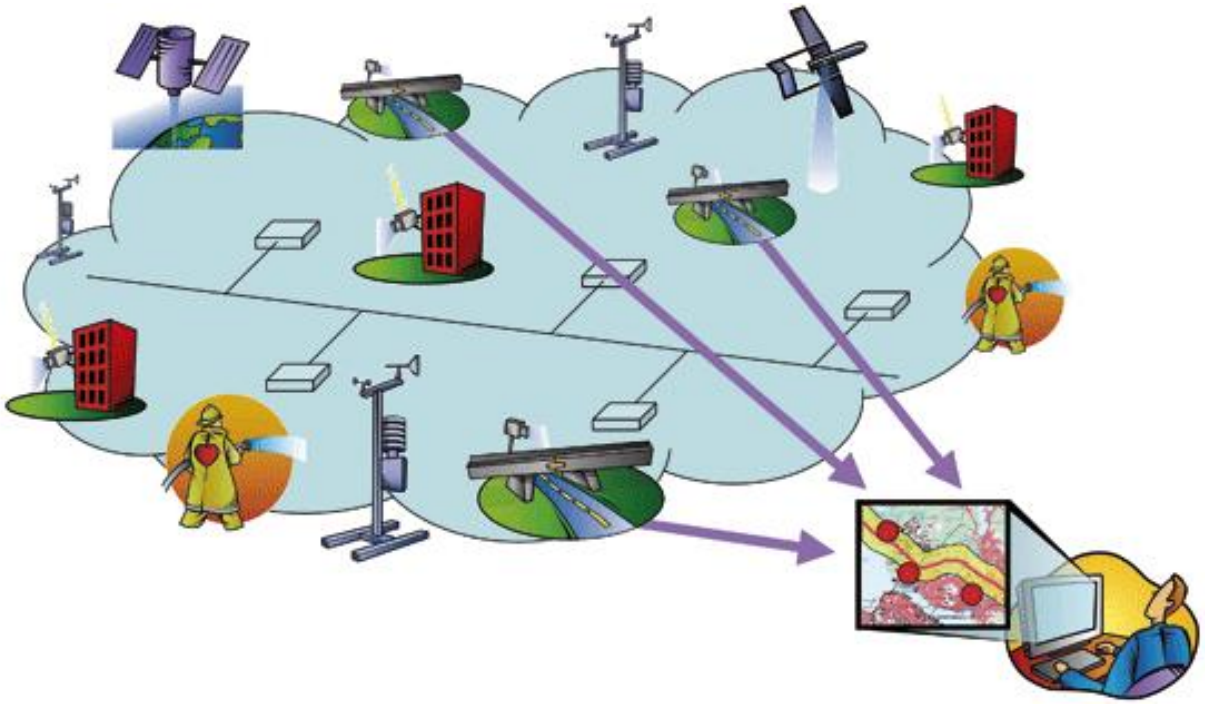


Figure 2 Example of Scenario

5 Conclusion

This work shows the introduction of project that can significantly support the development of monitoring applications. The development of intra- and inter-organizational applications represents one of the most involved problems of monitoring systems.

The development of the open service platform can enable public administrations to manage environmental information that can be used to support territorial decision-making process and to monitor for example outsourcing services (e.g. rubbish administration).

Solutions like the proposed one are not available, and in general there are not standardized and consolidated platforms able to integrate data from sensors, workflows and ontology in an efficient way; the only thing in this direction is the presence of some proprietary software created ad hoc for particular application scenarios, and not accessible to everybody. Anyways these kinds of software present several problems: they are static, extremely expensive and not interoperable.

6 References

[1] BERNHARDSEN, T., Geographic information systems: an introduction, John Wiley & Sons, New York (Usa), 2002

[2] BLOOMBERG, J., Principles of SOA, Application Development Trends Magazine, March 2003

[3] EUROPEAN PARLIAMENT AND THE COUNCIL, establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), Directive 2007/2/EC, March 2007

[4] Georgakopoulos, D., Hornick, M., Sheth, A., An overview of workflow management: From process modeling to workflow automation infrastructure, Springer Netherlands, April 1995

[5] GIBBONS, P.B., KARP, B., KE, Y., NATH, S., SESHAN, S., IrisNet: An Architecture for a World-Wide SensorWeb, IEEE Pervasive Computing, 2(4), October-December 2003

[6] JANKOWSKI, P., RICHARD, L., Integration of GIS-based suitability analysis and multi-criteria evaluation in a spatial decision support system for route selection, 1994

[7] REICHARDT, M., Sensor Web enablement, Open Geospatial Consortium (OGC) White Paper 05-063, 2005

[8] SANTANCHE, A., NATH, S., LIU, J., PRYANTHA, B., ZHAO, F., SenseWeb: Browsing the physical world in real time, IPSN 2006: Proceedings of the fifth international conference on Information processing in sensor networks, New York, (Usa), 2006

[9] SHEN, C., SRISATHAPORNHAT, C., JAIKAEO, C., Sensor Information Networking Architecture and Applications, IEEE Personal Communications (pages. 52-59), August 2001

[10] ZIGBEE ALLIANCE, INC., ZigBee Specification, White paper, 2006

SESSION

PROTOCOLS

Chair(s)

TBA

Adaptive Receiver Transmission Protocols for Receiver Blocking Problem in Wireless Multi-hop Networks

Kai-Ten Feng and Wei-Neng Lei
Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan
ktfeng@mail.nctu.edu.tw and roy2897.cm97g@nctu.edu.tw

***Abstract*—Due to the lack of a centralized coordinator for wireless resource allocation, the design of medium access control (MAC) protocols is considered crucial for throughput enhancement in the wireless ad-hoc networks. The receiver blocking problem, which has not been studied in most of the MAC protocol design, can lead to severe degradation on the throughput performance. In this paper, the diversified receiver transmission (DRT) is proposed to alleviate the receiver blocking problem without the adoption of additional control channels. The adaptive receiver transmission (ART) scheme is proposed to further enhance the system throughput with dynamic adjustment of the selected receivers. Simulations are performed to evaluate and compare the proposed two protocols with existing MAC schemes. It can be observed that the proposed ART protocol outperforms the other schemes by both alleviating the receiver blocking problem and enhancing the system throughput for the wireless multi-hop ad-hoc networks.**

Keywords: Wireless multi-hop networks, medium access control, receiver blocking problem.

I. Introduction

A wireless multi-hop network (WMN) [1] adopts wireless communication technologies to maintain connectivity and exchange messages between decentralized nodes in the multi-hop manners. This type of wireless networks are capable to perform self-creating, administering, and organizing the network connectivity. With the decentralized characteristics of the WMNs, feasible design of medium access control (MAC) protocol is considered important for performance enhancement. However, the connectivity between the network nodes are in general not guaranteed in the WMN, which incurs notorious exposed node and hidden node problems [2]. Some early attempts for resolving these problems in the literature [3], [4] suggested the usage of request-to-send (RTS) and clear-to-send (CTS) mechanisms, which were later adopted by the IEEE 802.11 standards. The well-adopted IEEE 802.11 MAC protocol suite [5], [6] can be employed in the WMNs since it has been specified to support decentralized operations called the ad-hoc mode.

However, it has been studied [7], [8] that the deployment of ad-hoc mode in the IEEE 802.11 network does not always result in feasible performance. Even though the hidden node and exposed node problems can be partially alleviated by adopting the distributed coordination function (DCF) in the IEEE 802.11-based protocols, an extended problem called receiver blocking or unreachability will be induced by the hidden node and exposed node problems thereafter. The receiver blocking problem occurs when the intended destination is located within the coverage of an on-going transmission pair. The destination node is not able to response to the corresponding RTS packet from the sender since the destination will be in the silent state caused by either the virtual carrier sensing (VCS) or the physical carrier sensing (PCS). In such case, the source node which is outside the range of this on-going transmission pair will confront a series of connection failure with its destination, which will result in the increase of unnecessary control overheads by initiating the RTS packets.

It has been investigated in several studies [9]–[12] regarding the severe performance degradation in ad-hoc networks. The dual-channel (DUCHA) [9] MAC protocol was proposed to alleviate the receiver blocking problem by adopting an additional channel for the transmission of control packets; while the data packet is transmitted in the data channel. However, each network node is required to install at least two transceivers in DUCHA scheme which is not always considered realistic due to hardware limitation and cost. On the other hand, the eMAC protocol [10] is proposed based on a multiple access collision avoidance (AMACA) protocol [11] to deal with the receiver blocking problem. The eMAC-table contains partial topology information of a network node and is periodically exchanged between the neighbor nodes. Therefore, a node can maintain and utilize the dual-hop neighborhood graph to determine the best strategy for the transmission of individual communication pause (ICP) packet which solves the ICP broadcast storm problem in the original AMACA protocol. Note that the ICP packet is utilized when a network node is notified to be silent for the duration of network allocation vector (NAV) and unfortunately becomes unreachable. Moreover, a enhanced IEEE 802.11 protocol that operates similar to the conventional DCF scheme is proposed in [12], which is called eDCF protocol in this paper for notational convenience. After sending the RTS packet to the intended receiver, the source node will set a timeout duration waiting for the CTS response. If the CTS packet has not been received after the timeout period, the eDCF scheme will provide an additional opportunity to select another receiver from its queue to deliver data packet since this channel within the coverage of the source node has already been reserved.

In this paper, the diversified receiver transmission (DRT)

[1]This work was in part funded by the Aiming for the Top University and Elite Research Center Development Plan, NSC 99-2628-E-009-005, the MediaTek research center at National Chiao Tung University, and the Telecommunication Laboratories at Chunghwa Telecom Co. Ltd, Taiwan.

mechanism is proposed to cope with the receiver blocking problem without adopting either additional control channels or transceivers. The DRT approach is proposed to provide additional opportunities for the transmission to multiple receivers and reduces the NAV duration in order to provide channel accessing opportunities for the other nodes in the network. However, the DRT scheme may suffer performance degradation under specific network scenarios which lead to the proposal of adaptive receiver transmission (ART) protocol in order to further enhance the network efficiency and channel utilization. The evaluation of the proposed schemes will be performed and compared with the conventional DCF, the eMAC, and the eDCF protocols via simulations. It will be shown that the receiver blocking problem can be effectively alleviated with the adoption of proposed DRT and ART schemes. The network throughput can consequently be enhanced.

II. Network Model and Problem Statement

Considering a set of nodes $\mathbf{N} = \{N_i \,|\, \forall\, i\}$ within a two-dimensional Euclidean plane, the locations of the set $\mathbf{N}$ are represented by the set $\mathbf{P} = \{P_{N_i} \,|\, P_{N_i} = (x_{N_i}, y_{N_i}), \forall i\}$. It is assumed that all the nodes are homogeneous and equipped with omnidirectional antennas under a single channel. The set of closed disks defining the transmission ranges of $\mathbf{N}$ is denoted as $\overline{\mathbf{D}} = \{\overline{D}(P_{N_i}, R) \,|\, \forall\, i\}$, where $\overline{D}(P_{N_i}, R) = \{x \,|\, \|x - P_{N_i}\| \leq R, \forall\, x \in \mathbb{R}^2\}$. It is noted that P_{N_i} is the center of the closed disk with R denoted as the radius of the transmission range for each N_i. Each node in the transmission range $\overline{D}(P_{N_i}, R)$ can communicate with N_i by utilizing the IEEE 802.11-based MAC features for channel allocations, including PCS, VCS, and binary exponential backoff (BEB) [13]. Moreover, the one-hop neighbor table for each N_i is defined as $\mathbf{T}_{N_i} = \{N_k \,|\, P_{N_k} \in \overline{D}(P_{N_i}, R), \forall\, k \neq i\}$. The receiver blocking problem associated with the receiver blocking group are defined as follows.

Definition 1 (Receiver Blocking Group). Given the set $\mathbf{S} \subseteq \mathbf{N}$ which includes all the transmitters and receivers, the receiver blocking group is defined as $\mathbf{B} = \bigcup_{N_i \in \mathbf{S}} \mathbf{T}_{N_i}$ since all the nodes in $\mathbf{B}$ are blocked either by the carrier sensing mechanisms or due to the on-going packet transmission.

Problem 1 (Receiver Blocking Problem). Let $\mathbf{B}$ be the receiver blocking group within the network. The receiver blocking problem occurs while a node $N_i \in (\mathbf{N} - \mathbf{B})$ intends to communicate with a node $N_j \in \mathbf{B}$. Due to the blocking nature of N_j, a large amount of useless connection-request packets will be issued by N_i, which leads to the degradation of network throughput.

Fig. 1 illustrates the schematic diagram for the receiver blocking problem with the network topology and the corresponding timing diagram. As shown in Fig. 1(a), it is considered that N_1 and N_2 constitute the on-going transmission pair as identified by the solid arrow. The receiver blocking problem happens if $N_A \in (\mathbf{N} - \mathbf{B})$ intends to initiate a communication link with $N_3 \in \mathbf{B}$, i.e. denoted by the dashed arrow. Based on Definition 1, the receiver blocking group is obtained as $\mathbf{B} = \{N_1, ..., N_9\}$, which lies within the light gray region as in Fig. 1(a). Referring to Fig. 1(b), N_A will attempt to communicate with N_3 by transmitting the RTS packet (i.e. RTS_1) after the successful channel contention. Based on the broadcast nature, N_B and N_C will also receive the RTS_1 packet and consequently set up their corresponding NAV timers in order to refrain from accessing the channel, i.e. $T_{NAV} = T_{CTS} + T_{Data} + T_{ACK} + 3T_{SIFS} + 3T_{prop}$. It is noted that the subscript in each timing parameter is utilized to denote its corresponding meaning, i.e. T_{CTS}, T_{Data}, T_{ACK}, T_{SIFS}, and T_{prop} indicate the time durations for the CTS packet, data packet, ACK packet, the short inter-frame space, and the propagation delay respectively. Moreover, T_{slot} and T_{DIFS} in Fig. 1(b) represent the slot time of conventional IEEE 802.11 standard and the time duration for the DCF inter-frame space respectively; while the parameter $T_{backoff}$ indicates the time interval for the current backoff window of a node.

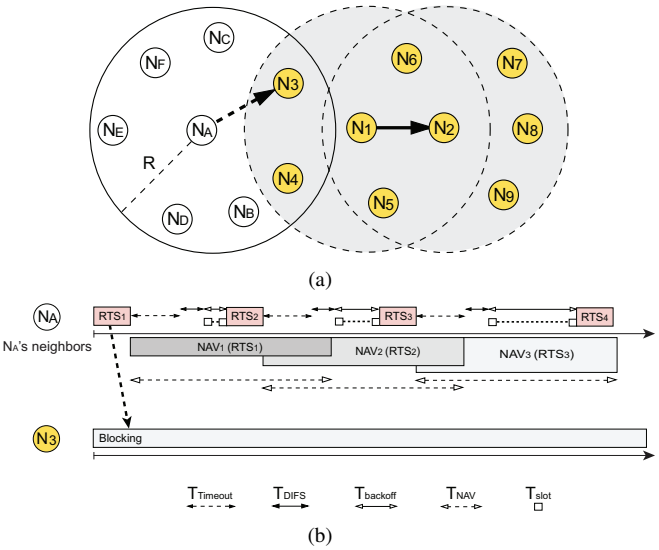


Fig. 1. The schematic diagram for the receiver blocking problem: (a) the network topology, and (b) the timing diagram.

However, N_3 will not respond to the RTS_1 packet with a corresponding CTS packet due to the PCS/VCS mechanisms. After a timeout $T_{timeout} = T_{CTS} + T_{SIFS} + T_{prop}$ for waiting the CTS packet, N_A will double its backoff window and re-initiate to communicate with N_3 by sending another RTS packet, i.e. the RTS_2 packet. In the meantime, N_B and N_C will refresh their corresponding NAV timers based on the newly issued RTS_2 packet as in Fig. 1(b). Consequently, N_A will result in a great amount of useless retries of sending RTS packets, which prohibit N_B and N_C from contending the channel and lead to the degradation of network throughput.

III. Proposed MAC Protocols

A. Diversified Receiver Transmission (DRT) Scheme

According to Definition 1, all nodes in the receiver blocking group $\mathbf{B}$ will not respond to the node $N_i \in (\mathbf{N} - \mathbf{B})$. Therefore, the transmission of the RTS packets from N_i will

fail in constructing the communication links to the nodes in $\mathbf{B}$. It is noticed that the unsuccessful reception of the CTS packets by N_i can be attributed to the factors as follows: (a) packet collisions; (b) error reception of the CTS packet from the receiver; and (c) the receiver located in the receiver blocking group $\mathbf{B}$. If the failure of acquiring the CTS packets is due to the factors (a) and (b), the conventional BEB method can be adopted to effectively resolve the drawbacks of the missing CTS packets by expanding the contention window and retransmitting the RTS packets. However, the BEB scheme will not suffice for alleviating factor (c), which will in general result in excessive and ineffective transmissions of the RTS packets.

One intuitive method to resolve factor (c) is to terminate the retransmission of the RTS packets since the RTS retries have no contribution in constructing the communication links with the node in $\mathbf{B}$ [10]. However, it requires node N_i to possess the information that the receiver is located within $\mathbf{B}$, which is considered inapplicable in realistic cases. The design concept of the proposed DRT technique is to increase the probability for selecting the destination that does not belong to the receiver blocking group $\mathbf{B}$. Instead of merely transmitting the RTS packet to its original intended receiver in $\mathbf{B}$, N_i will also attempt to utilize the same RTS packet for constructing the communication links with the other receivers which are not in the set $\mathbf{B}$, e.g. N_C as in Fig. 1(a). The policy of the DRT scheme is to utilize the designed RTS packet (called M-RTS) that will be specified and destined to more than one receiver, i.e. to the set $\mathbf{R}_M$ containing multiple receivers where M denotes the maximal number of receivers that will be specified within the M-RTS packet. In other words, additional receivers within the neighbor table $\mathbf{T}_{N_i}$ will be randomly chosen to accept the M-RTS packet other than the original targeted node that is located within the set $\mathbf{B}$, i.e. the value of M is designed to be always greater than one. In comparison with the original RTS packet, there is an additional CTS responding list in the M-RTS packet. This CTS responding list records the order of response for each receiver in $\mathbf{R}_M$, which ensures that the M receivers can arrange their CTS responses without collisions. Therefore, the probability for all N_i's receiving nodes to be blocked will be reduced from p_f to p_f^M, where $0 \leq p_f < 1$ is denoted as the probability of transmission failure. Consequently, the receiver blocking problem can be alleviated, which results in the enhancement of network throughput.

Fig. 2 shows the exemplified timing diagrams for the proposed DRT scheme. It is assumed that N_A wins the contention for channel access and is ready to transmit its data packets, where the maximal number of receivers within the multiple receiver set $\mathbf{R}_M$ is chosen as $M = 2$. First of all, the ideal case is considered where none of the selected nodes for $\mathbf{R}_M$ is located within the set $\mathbf{B}$, e.g. $\mathbf{R}_M = \{N_B, N_C\}$ as shown in Fig. 1(a). Based on the proposed DRT scheme, N_A will therefore transmit an M-RTS packet, i.e. the M-RTS$_A$ packet, which targets the two receivers N_B and N_C. Under the case with non-blocking receivers, N_B and N_C will sequentially

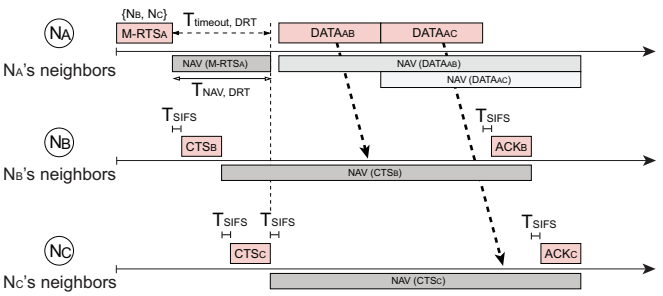


Fig. 2. The data delivery process of the proposed DRT mechanism.

feedback their CTS packets to N_A with the time difference of T_{SIFS}, where the order of the CTS feedbacks is specified within the M-RTS$_A$ packet. After the reception of the CTS$_B$ and CTS$_C$ packets, N_A will start the delivery of data packets to both N_B and N_C, respectively. Finally, the two receiver nodes will acknowledge the data packets by the corresponding ACK packets, i.e. ACK$_B$ and ACK$_C$.

On the other hand, the receiver blocking problem can happen when one of the selected nodes in $\mathbf{R}_M$ belongs to the set $\mathbf{B}$, e.g. $\mathbf{R}_M = \{N_B, N_3\}$. Similar to the explanation as in Fig. 2, N_A will initiate the M-RTS$_A$ packet that is addressed to both N_B and N_3. In this case, N_A will not receive the CTS packet from N_3 since N_3 is within the receiver blocking group $\mathbf{B}$. Therefore, the data packet towards N_B will be transmitted after the end of two CTS response time since the DRT protocol needs to wait for the required response time from the selected destinations. Afterwards, N_B will send the ACK packet if it successfully receives the data packet from node N_A. In the case that N_A does not receive any CTS feedbacks, N_A will re-initiate the contention process after a timeout period, which is M multiple of the original length defined in the conventional IEEE 802.11 protocol, i.e. $T_{timeout,DRT} = M \cdot T_{timeout}$.

Furthermore, consider the same case that N_A intends to transmit data packets to N_3 as shown in Fig. 1(a), it can be observed that N_A will continue to win the channel contention on those retrials to N_3 since all the other nodes will consistently be set at their NAV states. As shown in Fig. 1(b), the NAV timer assigned by N_A is longer enough to prevent the other competitors from contending the channel during its retransmission to the node in receiver blocking group $\mathbf{B}$. In order to provide the channel accessing opportunities for the competitors, the proposed DRT scheme reduces the NAV duration specified within the RTS packet to a shorter period of time which only protect until the end of current transmission of the CTS packet, i.e. a NAV duration of $T_{NAV,DRT} = T_{timeout,DRT}$ will be set within the RTS packet based on the DRT scheme. Therefore, the probability of channel contention can be increased under the occurrence of the receiver blocking problem.

B. Adaptive Receiver Transmission (ART) Scheme

As described in Subsection A, the DRT scheme may confront the inefficiency problem due to the requirement to allow a large amount of sequential feedbacks from the CTS packets.

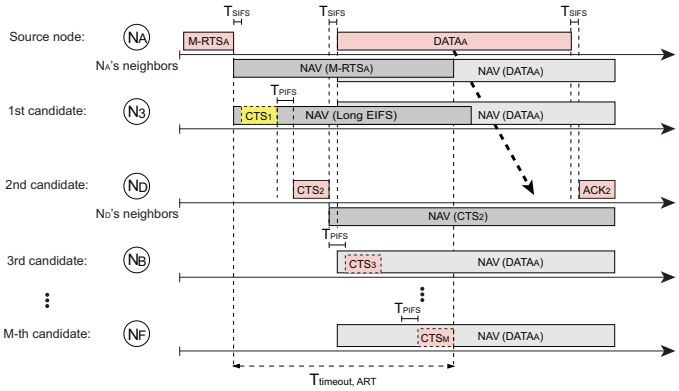


Fig. 3. The timing diagram for the ART protocol: If N_3 cannot correctly receive the M-RTS packet from N_A, it will set its NAV timer as $T_{LongEIFS}$ in order not to interfere the CTS reception of other nodes. After waiting for N_3's required CTS response time along with T_{PIFS}, N_D replies with a CTS packet and further triggers the data transmission. The other destination nodes will suspend their CTS feedbacks to N_A.

In order to alleviate the problem associated with the DRT scheme, the proposed ART protocol is designed to enhance the system throughput by conducting opportunistic CTS feedback. As shown in Fig. 3, N_A initiates the communication to the designated M receivers by broadcasting the M-RTS packet to its neighbors. Based on the order of receivers specified in the M-RTS packet from N_A, these M destinations are designed to potentially reply their corresponding CTS packets to N_A sequentially. One of the major design parameters in the ART scheme is that the inter-frame space between two CTS packets is modified from T_{SIFS} to $T_{SIFS}+T_{slot}$, which is coincidentally equal to the point coordination function (PCF) inter-frame space T_{PIFS}. Note that the adoption of $T_{PIFS} = T_{SIFS}+T_{slot}$ in the proposed ART scheme will not conflict with the original centralized PCF coordination since only ad-hoc operations are considered in the network. The reason to wait for additional T_{slot} within the T_{PIFS} is to allow the receivers to verify if they should continue transmitting their CTS packets. Since each receiver may not be able to hear the CTS feedbacks from other receivers to N_A, an elongated waiting time interval T_{PIFS} is required for each receiver to verify if there exists data transmission from N_A to its pervious receiver after a successful M-RTS/CTS negotiation. If a receiver does not hear the CTS transmission associated with the data packet from N_A to its previous receiver after time T_{PIFS}, the receiver will initiate the delivery of a CTS packet to N_A to request for data transmission. On the other hand, with successful M-RTS/CTS handshaking between N_A and the previous receiver, the data packet from N_A can therefore be transmitted after the short time duration T_{SIFS}. Consequently, by observing the on-going data transmission during the additional T_{slot} time interval, the remaining receivers will suspend their CTS feedbacks to N_A in order to prevent unnecessary channel reservation within their transmission ranges.

According to the mechanism as stated above, there is only one selected receiver that replies its CTS packet back to N_A, which consequently can reduce the waiting time for other data packets that are not destined to itself. Similar to the other non-destination neighbors, those unselected destinations must wait for the NAV period until the end of on-going communication. Note that if a node can correctly receives the M-RTS packet, it will set up its NAV timer for the time period as

$$T_{NAV,ART} = T_{SIFS} + (M-1)T_{PIFS} + M(T_{CTS} + T_{prop}) \quad (1)$$

Furthermore, N_A will re-initiate the contention process after $T_{timeout,ART} = T_{NAV,ART}$ if N_A does not receive any CTS feedbacks. After data packet has been designated to a specific receiver, the other non-selected receivers and non-destination neighbors will refresh their NAV period to become $T_{Data}+T_{SIFS}+T_{ACK}+2T_{prop}$ until the data has completed its transmission. Therefore, the channel can be completely reserved within the transmission range of a source node, and the channel reservation becomes more flexible if the source node fails to establish the link with its receivers in this round of transmission.

In certain situations, the receivers may receive scrambled signals that can not be decoded such that the M-RTS packet delivered from N_A will not be correctly received, e.g. the receiver N_3 as shown in Fig. 3. The reason is that these receivers are located in the receiver blocking group $\mathbf{B}$ where some neighbor nodes are simultaneously transmitting their packets. Therefore, in order not to interfere with either the CTS or ACK reception of other source nodes, N_3 is designed to wait for a longer NAV duration as long EIFS that can be obtained as $T_{LongEIFS} = T_{SIFS} + (M-1)T_{PIFS} + M(T_{CTS} + T_{prop}) + T_{DIFS}$, which is extended from the conventional $T_{EIFS} = T_{SIFS} + T_{CTS} + T_{prop} + T_{DIFS}$ in the IEEE 802.11 system. Even if N_3 can correctly receive the M-RTS packet from N_A, N_3 may not be able to reply its corresponding CTS packet since it can be NAVed by other on-going transmission in its neighborhood. N_3 will be requested to refresh its NAV timer for $T_{NAV,ART}$ similar to the other non-destination neighbors of N_A. Furthermore, consider a node, e.g., N_B, correctly receives the M-RTS packet from a source node N_A, and is notified to be one of the M receivers. During the time interval between the end of M-RTS transmission and before its CTS feedback, N_B may receive other M-RTS or CTS packets from its neighbors before N_B to broadcast its corresponding CTS packet to the original source node N_A. Under such situation, no matter if N_B will be informed to be the receiver from other source nodes, N_B will be requested to set its corresponding NAV timer according to the newly received M-RTS or CTS packet, which results in the termination of its original CTS feedback.

Referring to Fig. 3 as an example, it is assumed that N_A wins the contention for channel access and transmits its M-RTS packet to M destinations. All of N_A's neighbors will set their NAV period to be $T_{NAV,ART}$. Consider the case that N_3 is unable to receive the M-RTS packet correctly from N_A, N_3 will adjust its NAV timer as $T_{LongEIFS}$ in order not to interfere with the other transmissions in the network. After

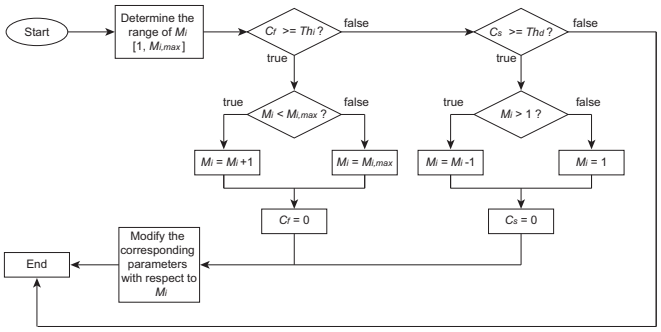


Fig. 4. The flow chart for dynamic adjustment of parameter M_i by adopting the ART protocol.

waiting for the time durations of both CTS_1 transmission and T_{PIFS}, N_D will reply with its CTS feedback, i.e. CTS_2, to N_A to request for data transmission. After observing the CTS_2 packet from N_D, the other nodes within the transmission range of N_D will set their NAV period to be $T_{Data} + T_{SIFS} + T_{ACK} + 2T_{prop}$ which is the same as that in the duration field of conventional CTS packet. As N_A has received its first CTS feedback from N_D, N_A will begin the data transmission to N_D after the time duration of T_{SIFS}. The CTS feedbacks from the other destinations, i.e. from the third to the M-th candidate, back to N_A will therefore be suspended.

1) Dynamic Adjustment of Parameter M in the ART Protocol: The maximal number of receivers M for each sending node should be determined in order to feasibly improve the network performance. The proposed ART scheme allows each node to maintain and dynamically adjust its own value of parameter M based on the real-time network environment. In order to further identify the dynamic behavior of parameter M, it will be modified as M_i where $i = 1$ to N with N denoting the total number of nodes in the network. Fig. 4 shows the algorithm for dynamically adjusting the parameter M_i at every node in the network. As a node wins the contention for channel access, e.g. node N_i, it will execute the algorithm to determine the value of M_i in this transmission round before broadcasting its M-RTS packet. First of all, the range of M_i for N_i will be determined for the dynamic adjustment algorithm as $[1,\ M_{i,max}]$, where the maximum value of this range $M_{i,max}$ can be obtained as

$$M_{i,max} = \min\{\omega_i, n_i\} \tag{2}$$

Note that the parameter n_i denotes the total number of neighbor nodes of N_i. The other parameter ω_i in (2) is utilized to limit the length of NAV timer of N_i's neighbor nodes not to exceed the best case of a successful data transmission. In other words, according to (1), the NAV duration for those unselected destination nodes after receiving the M-RTS packet from N_i will be constrained to be $T_{NAV,ART} = T_{SIFS} + (\omega_i - 1)T_{PIFS} + \omega_i(T_{CTS} + T_{prop}) \leq 3T_{SIFS} + T_{CTS} + T_{Data} + T_{ACK} + 3T_{prop}$. Consequently, the parameter ω_i will be selected as

$$\omega_i = \left\lfloor \frac{2T_{SIFS} + T_{PIFS} + T_{CTS} + T_{Data} + T_{ACK} + 3T_{prop}}{T_{PIFS} + T_{CTS} + T_{prop}} \right\rfloor \tag{3}$$

The main purpose of the minimization in (2) is to intuitively constrain the parameter ω_i derived from NAV duration not to exceed the total number of neighbors n_i of N_i.

As depicted in the flow chart shown in Fig. 4, the dynamic adjustment of parameter M_i will first verify with an increasing threshold Th_i to determine if the current M_i should be increased or not. The verification criterion is based on the number of successively transmission failures of the M-RTS packets from the previous rounds, which is denoted as C_f. If C_f is greater than Th_i, the adjustment algorithm considers this situation as potential occurrence of receiver blocking problem. In general, the probability of continuously M-RTS collisions will be small since the BEB mechanism can adequately avoid packet collision if there does not exist the receiver blocking problem. Therefore, the algorithm is designed to increase the current M_i value such that there will be additional receivers to assist the data delivery process from the source node. As shown in the left part of the flow chat in Fig. 4, the current M_i value will be verified whether it is less than the maximum value $M_{i,max}$. If the condition is true, the current value of M_i will be increased by one; otherwise, M_i is set equal to $M_{i,max}$. Consequently, the counter C_f will be reset to zero to initiate another accumulation of M-RTS transmission failures.

On the other hand, if C_f is less than the increasing threshold Th_i, the right part of the flow chart will be executed. In this case, the design consideration is to examines whether the current M_i value should be decremented if the number of continuously successful data transmission, indicated as C_s, is greater than the decreasing threshold Th_d. The reason is that larger value of M_i corresponds to excessive receivers are selected which can cause long delay of the corresponding CTS feedbacks. After the new M_i value is determined, the parameters associated with M_i will be adjusted accordingly such as M-RTS packet size, $T_{NAV,ART}$, and $T_{LongEIFS}$ for node N_i.

IV. Performance Evaluation

The performance of proposed DRT and ART schemes will be compared with existing protocols including the conventional IEEE 802.11a DCF protocol, the eMAC algorithm [10], and the eDCF protocol [12]. Most of the parameters are adopted from IEEE 802.11a standard; while the length of the M-RTS packet is equal to $20 + 6(M_i - 1)$ bytes. The network nodes are randomly distributed in a 180×180 square meters area. It is assumed that the active nodes always have data packets to deliver, and the packets size are considered to be the same. All the data packets generated from a source node are assumed to be transmitted to its network neighbors and a number of specific receivers are randomly selected. Both the MAC header and the control packets, i.e. M-RTS, CTS, and ACK packets, are transmitted at the rate of 6 Mbps; while

the payload part of a data packet is delivered at data rate of 24 Mbps. Moreover, the increasing threshold Th_i is chosen to be $Th_i = 3$, which indicates that the current M_i value of the source node is increased by one if the M-RTS packet is continuously failed by three times. This selection is considered reasonable in a normal node density of network layout since the BEB mechanism can partially alleviate the packet collision between the neighbor nodes. On the other hand, the decreasing threshold is set to $Th_d = 8$. In the following figures for performance comparison, each data point is averaged from 50 simulation runs where each simulation run is executed for 50 seconds.

Fig. 5 illustrates the performance comparison for average throughput under different number of nodes N with payload size $L = 3000$ bytes; while the throughput performance is compared under different payload sizes with $N = 60$ in Fig. 6. Note that the proposed ART protocol is evaluated under three cases as $M = 2$, 4, and with dynamic adjustment algorithm for M_i which is denoted as ART-DA scheme. As the total number of nodes in the network grows, one may observe from Fig. 5 that the throughput performance of all the schemes becomes worse since there can exist more packet collisions and additional interference from hidden nodes. The proposed ART-DA protocol can provide the highest throughput performance compared to the other schemes owing to its dynamic adjustment of selected receivers M_i. The throughput performance of eDCF protocol is similar to that of the DRT scheme since the eDCF protocol provides a second chance to deliver the data packet to another receiver if there is no CTS feedback from the original destination. Furthermore, as the payload size becomes larger as shown in Fig. 6, the throughput performance is increased in all the schemes since the source node is able to deliver additional information bytes after winning the channel contention. The proposed ART-DA protocol still outperform the other methods with the highest system throughput owing to its better channel utilization instead of constructing unnecessary connection attempts between the network nodes. Consequently, the simulation results show that the proposed ART protocols, especially the ART-DA scheme, can consistently outperform the other algorithms and effectively alleviate the receiver blocking problem.

Fig. 7 illustrates the comparison of control overhead versus number of nodes under $L = 3000$ bytes. Note that the control overhead is defined as the number of RTS/M-RTS packets over the number of CTS packets which implies the average required RTS/M-RTS packets for a protocol to acquire a CTS feedback from the selected receivers. In other words, as the control overhead is increased, the protocol will operates in a less efficient manner with worse channel utilization since it wastes excessive time in establishing the connection to obtain a CTS packet. As in Fig. 7, if the number of nodes is increased, additional control overhead for all the scheme can be observed which is attributed to the excessive packet collisions and retransmissions within the network. The conventional IEEE 802.11a DCF protocol results in the highest control overhead among all the schemes owing to its poor ability to handle

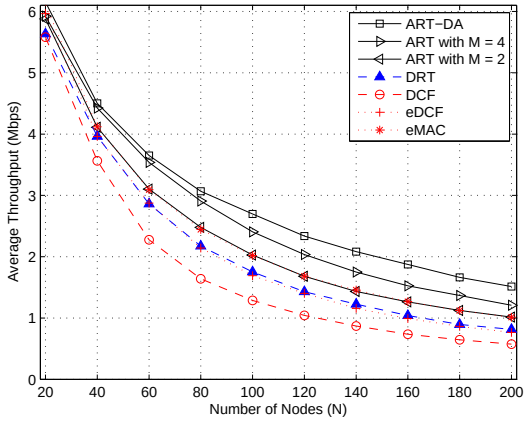


Fig. 5. Performance comparison: average throughput versus number of nodes.

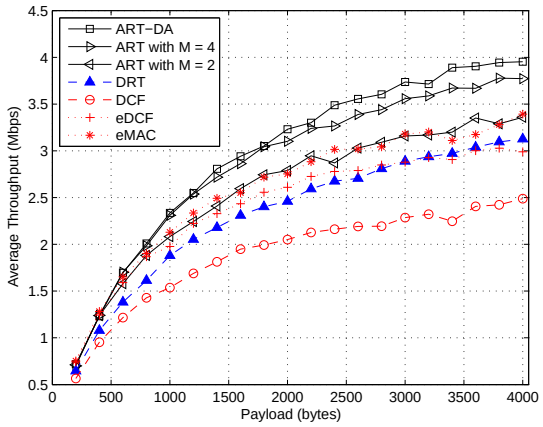


Fig. 6. Performance comparison: average throughput versus payload size.

the receiver blocking problem in the ad-hoc networks. Even though the throughput performance of the eDCF protocol is similar to that of the DRT scheme, excessive RTS packets are required by the eDCF protocol which is attributed to the second chance for delivering the data packet to another receiver that is not confirmed by the second receiver's CTS packet. It can be observed that the proposed ART-DA scheme can achieve reasonable lowered control overhead compared to other protocols. With less number of network node, the behaviors of the ART-DA protocol will be similar to the cases with smaller M values, e.g. $M = 2$; while the ART-DA scheme will behave similar to the situation with larger M under increased value of N. Therefore, as can be seen from Fig. 7, the total number of M-RTS packets of the ART-DA protocol will intersect the curves from smaller to larger M values as the number of nodes is augmented.

V. Conclusion

In this paper, the diversified receiver transmission (DRT) is proposed in order to alleviate the receiver blocking problem in the multi-hop ad-hoc networks. The adaptive receiver

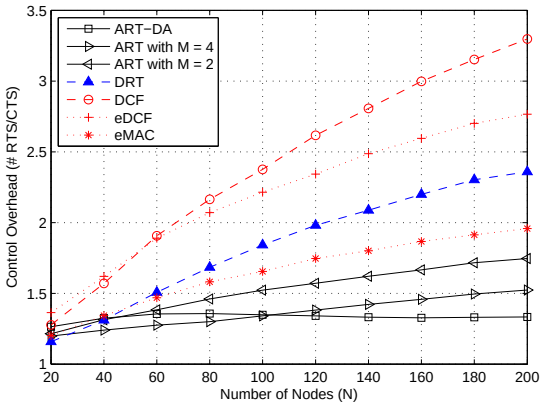


Fig. 7. Performance comparison: control overhead versus number of nodes.

transmission (ART) scheme is proposed to further improve the throughput performance with dynamic adjustment on the number of selected receivers. It is shown in the simulation results that the proposed ART scheme can effectively alleviate the receiver blocking problem, which consequently enhances the network throughput for wireless multi-hop ad-hoc networks.

REFERENCES

[1] C.S. Murthy and B.S. Mano, *Ad Hoc Wireless Networks: Architectures and Protocols*. Prentice Hall, 2004.

[2] F.A. Tobagi and L. Kleinrock, "Packet switching in radio channels:Part II-the hidden terminal problem in carrier sense multiple-access modes and the busy-tone solution," *IEEE Trans. Commun.*, vol. COM-23, no. 12, pp. 1417–1433, 1975.

[3] P. Karn, "MACA - A new channel access method for packet radio," in *Proc. ARRL/CRRL Amateur Radio 9th Computer Networking Conf.*, Sep. 1990, pp. 134–140.

[4] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in *Proc. ACM SIGCOMM'94*, Oct. 1994, pp. 212–225.

[5] IEEE 802.11 WG, *IEEE Std 802.11a-1999 (R2003): Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band*, IEEE Standards Association Std., 2003.

[6] IEEE 802.11 WG, *IEEE Std 802.11b-1999 (R2003): Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band*, IEEE Standards Association Std., 2003.

[7] S. Xu and T. Saadawi, "Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks?" *IEEE Commun. Mag.*, vol. 39, no. 6, pp. 130–137, Jun. 2001.

[8] K. Xu, M. Gerla, and S. Bae, "How Effective is the IEEE 802.11 RTS/CTS Handshake in AdHoc Networks?" in *Proc. IEEE GIOBECOM*, Nov. 2002, pp. 17–21.

[9] H. Zhai, J. Wang, and Y. Fang, "DUCHA: A New Dual-Channel MAC Protocol for Multihop Ad Hoc Networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3224–3233, Nov. 2006.

[10] K. Ghaboosi, M. Latva-aho, Y. Xiao, and Q. Zhang, "eMAC - A Medium Access Control Protocol for the Next Generation Ad Hoc Networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4476–4490, Oct. 2009.

[11] K. Ghaboosi and B.H. Khalaj, "AMACA - a new multiple access collision avoidance scheme for wireless LANS," in *Proc. IEEE PIMRC*, Sept. 2004, pp. 1932–1936.

[12] A. Chayabejara, S.M.S. Zabir, and N. Shiratori, "An enhancement of the IEEE 802.11 MAC for multihop ad hoc networks," in *Proc. IEEE VTC-Fall*, Oct. 2003, pp. 3020–3024.

[13] B.J. Kwak, N.O. Song, and L.E. Miller, "Performance Analysis of Exponential Backoff," *IEEE/ACM Trans. Netw.*, vol. 13, no. 2, pp. 343–355, Apr. 2005.

Study on the Sensor MAC Protocol for Environment Monitoring of Livestock Farm

Ho-Seok Jeong[1], **Hyun Yoe**[2]
[12]Dept. of Information and Communication Engineering, Sunchon National University, Suncheon, Jeollanam-do, Republic of Korea

Abstract - *Because the WSN has an extremely limited feature for storage and power supply capability, there is a problem that batteries of sensor nodes are frequently replaced and recharged. This paper suggests the event-driven CBMAC protocol for improving the battery consumption problem in such a sensor network. The suggested CBMAC exploits an event technique to reduce the traffic increase by proposing data collection at the certain interval when an event is arisen, and adjusts the sleep interval variably to lower consumption of the LPL. In order to test performance of the CBMAC, it applied the CBMAC into the livestock farm monitoring system to analyze breeding conditions of livestock and environmental information and to create an event table, and exploited MATLAB software to carry out simulations for evaluating the performance. As a result of the simulation, it could be found that the suggested CBMAC had improved the transmit/receive power of sensors by up to 40%, and the LPL consumption energy by about 25% compared to the B-MAC when events were arisen.*

Keywords: Wireless Sensor Networks, MAC, Protocol, Feedlot

1 Introduction

The ubiquitous computing collects a variety of information by using sensors without user's awareness, uses processors to make a decision, and makes carry out useful things for users through communication technologies [1][2].

The WSN(Wireless Sensor Networks) is a core field of the ubiquitous computing, which supports up to 60,000 nodes, and could secure longer communication distances by building a mesh network, therefore, it has been used for broader applications in the environment sector such as sensing of forest fire and flood, meteorological observation, precision agriculture etc., healthcare sector such as monitoring and tracking of patients, drug control etc., home automation, measurements of vehicle conditions, inventory control and so on [3][4].

Since sensor nodes consisting of such a WSN are operated with batteries, there is a drawback that requires to replace or recharge batteries frequently due to an energy problem, additionally it causes congestion of the network due to the increase of multiple packets created simultaneously from several regions when a lot of sensor nodes generate certain events [5][6].

A factor causing the most power consumption in the sensor network is Idle listening, which monitors channels regardless of the presence or absence of data transmission & receipt and consumes about 50% to 100% of the entire communication energy [7].

To reduce this energy consumption, many studies employed a method of keeping sensor nodes in the state of Sleep using MAC protocol. Important MAC protocol includes synchronous mode of S-MAC [8] and T-MAC [9], and asynchronous mode of B-MAC [10]. Synchronous mode of MAC shows high energy efficiency in distributed environment like Ad-Hoc network and is proper for streaming data service, but has a weakness of being not very useful in a small scale of network like WPAN environment. Synchronous mode of B-MAC can be materialized in a relatively simple way, but has a weakness that energy consumption increases due to the increase of overhearing when traffic increases.

To solve unnecessary data receipt involving existing MAC protocol, this paper proposes MAC protocol which uses event detection algorithm and improves energy consumption efficiency by reducing the increase of data traffic caused by data collection and reducing the frequency of LPL(Low Power Listening) occurrence through controlling Sleep interval variably when an event takes place.

This paper is organized as follows. Chapter 2 explains the proposed CBMAC protocol, and Chapter 3 describes the algorithm of CBMAC protocol. Chapter 4 evaluates performance of the proposed protocol through simulations, finally Chapter 5 finishes this paper with a conclusion.

2 CBMAC PROTOCOL

2.1 CBMAC (CONTEXT BASED MAC

CBMAC(Context Based MAC) analyzed requirements needed when raising livestock and climate characteristics in Korea, and configurated context information optimized for

[2] Corresponding Author

barn environment to generate an event through event-based data transmission scheme.

With regard to the system of monitoring barn environment, sensor nodes collect temperature and humidity data essential for livestock raising. When collecting data, temperature-humidity intervals suited for livestock rearing according to time take place. For example, recommended temperature for piglet rearing is from 10℃ to 20℃ with recommended daily temperature variation from 10℃ to −10℃, which shows a slow variation from higher to lower temperature over long time instead of sudden changes.

By setting boundary temperature (12 to 18℃ for piglets) as an event using these characteristics, CBMAC must not transmit data to nearby sensor nodes through the generation of event when the temperature reaches boundary temperature interval, which is recommended temperature for livestock rearing.

Nearby sensor node, which gets aware that data is not collected in the boundary interval, extends Sleep interval over two times compared to that of existing B-MAC and reduces data traffic and energy consumption.

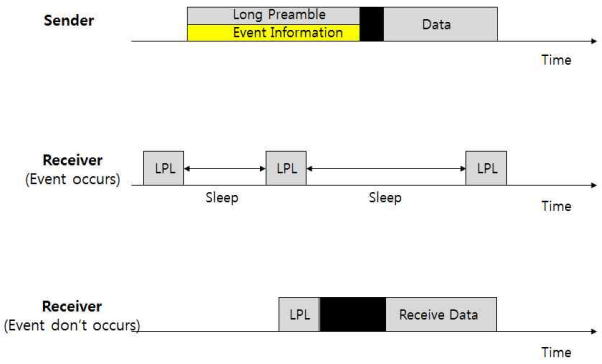


Figure 1. Data Transmission Method of the Proposed CBMAC

2.2 OVERHEARING PROBLEMS

When sensor nodes need to send data, the B-MAC transmits a preamble longer than the sleep interval before sending it to inform the data transmission to the neighboring sensor nodes. At this time, even sensor nodes disrelated with reception of the data should keep to receive the preamble till the end, just after checking the destination address contained in the data header following the preamble, it decides that itself is not one to receive and operates in the sleep mode.

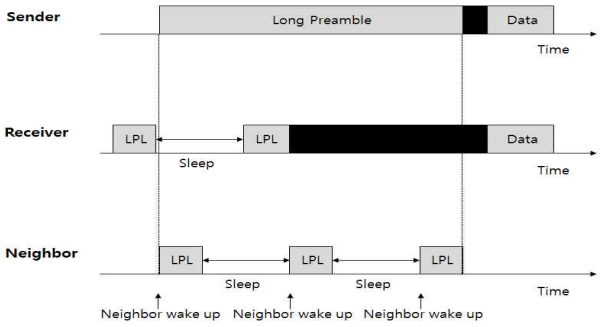


Figure 2. Overhearing cause by Sleep Interval Increase

Due to such an overhearing problem, in the CBMAC protocol, sensor nodes would send preamble signals with as much as about 2~3 times more stronger than the conventional B-MAC because of the extended sleep interval when an event is arisen if they want to inform data transmission to other nodes. At this time, there would be more overhearing than the B-MAC as the figure 2.

Therefore, in order to reduce such an energy consumption in the CBMAC, when it transmits a preamble, it sends DA, which is an address of the sensor node to be received, a preamble and a timer, which indicates time to complete data transmission, together, to make sensor nodes disrelated with the reception, which are awaken through the LPL during the preamble, check the DA and timer in the preamble, make them keep the sleep condition until the timer is out as the figure 3, so that it solve the overhearing problem.

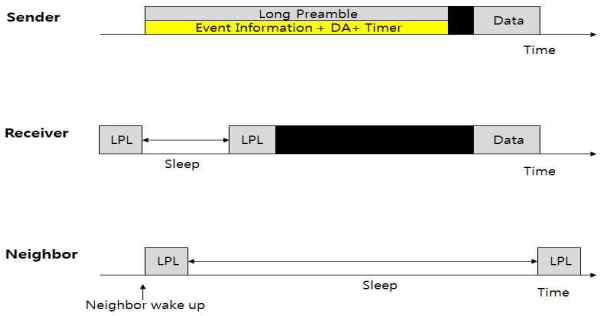


Figure 3. Overhearing Problem Solving

3 CBMAC ALGORITHM

CBMAC proposed by this paper designed algorithm based on event information as follows. As shown in figure 4, when an event doesn't take place to sensor node, sensor node performs LPL function of monitoring whether or not there are effective signals to the medium.

When there are effective signals to the channel, sensor checks that an event took place. When an event takes place, sensor substitutes event variable 1 and extends Sleep interval about twice. And according to doubled Sleep interval, the length of Timer is extended and Sleep interval is kept up to the remaining LPL point of time. When an event doesn't take place, sensor transmits Wake up Preamble and keeps itself in Sleep up to the remaining LPL point of time after transmitting data to nearby sensor nodes. When signals effective to channel are confirmed by LPL, sensor node is awakened through Wake up Preamble, and sensor node receiving Preamble checks whether or not to receive data. When it is effective, it receives data; and when it is not effective, it keeps itself in Sleep till Timer is terminated.

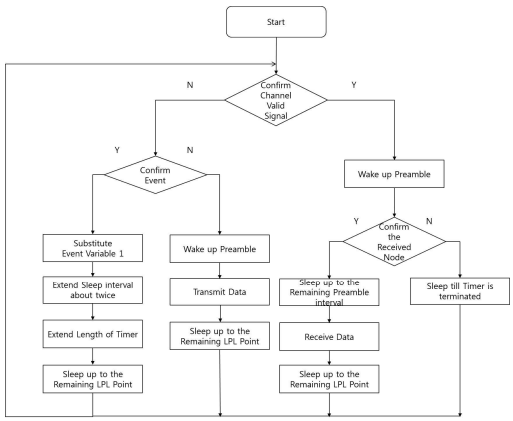


Figure 4. Control Algorithm when the Event has not occur

As shown in figure 5, when an event takes place to sensor node, sensor node checks whether there are effective signals to the medium; and checks whether an event took place when there are no effective signals.

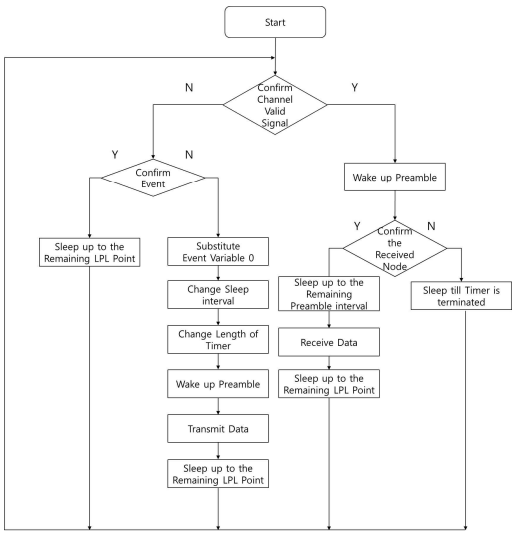


Figure 5. Control Algorithm when the Event occurs

When an event took place, sensor node checks whether the event continues and keeps itself in Sleep up to the remaining LPL interval; and when an event didn't take place, it substitutes event variable 0 and maintains Sleep interval and Timer length at normal interval. Then, sensor node transmits Wakeup Preamble to wake up other sensor nodes, and keeps itself in Sleep up to the remaining LPL point of time after sending Data

4 PERFORMANCE EVALUATION

4.1 SIMULATION ENVIRONMENTS

In order to analyze performance of energy consumption for the CBMAC proposed in this paper, it simulated the derived energy consumption model with the MATLAB. It used a total of 4 nodes in the experiment, and the network was constructed as a lattice form of 2×2 at intervals of 10m between sensor nodes considering that the length of a livestock room is 20m^2 in the livestock farm environment.

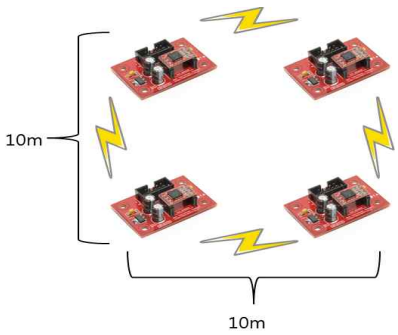


Figure 6. Network Topology

Referring to the energy consumption model, the transmit and receive energy of the sensor node is 19.5 and 21.8 mW, respectively, and the back-off timing is the same as the B-MAC transmission scheme to generate the contention window values randomly.

It compared with the B-MAC for verifying performance of the CBMAC, and it selected three kinds of temperatures / humidities, transmit/receive energy consumptions, data energy consumptions and energy consumptions of the LPL interval when an event is arisen for the performance evaluation factors.

4.2 SIMULATION RESULT

The figure shows a result of simulating the consumption rate of the energy to transmit/receive temperatures for the CBMAC and the conventional B-MAC based on the environmental data values collected through sensors at

intervals of 5℃. The simulation generated data 1,000 times for 9 sensor nodes to compare the B-MAC with the S-MAC focused on the probability of transmitting/receiving temperatures/humidities.

As shown in the figure 7, the B-MAC has constant energy consumption for data packets transmitted/received to keep the energy consumption constant regardless of the condition when data is generated, while the proposed CBMAC shows higher energy efficiency than the B-MAC because its amount of data transmitted is decreased up to 29% when the event of temperature is arisen between 10 and 30℃ that is adequate for breeding livestock in the livestock farm site.

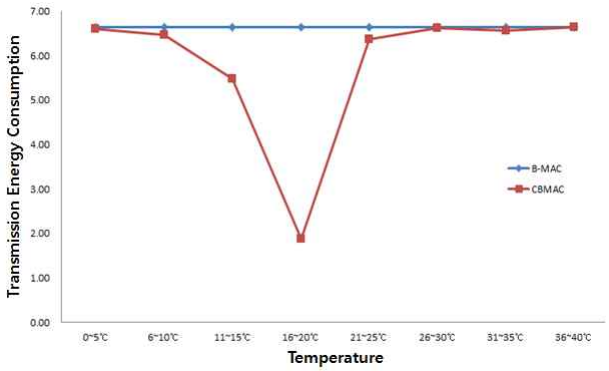


Figure 7. Compare the temperature transmission energy consumption of B-MAC and CBMAC

The figure 8 is the energy consumption rate when the humidity information is transmitted/received through sensors, which the humidity is also similar to the energy consumption rate in the B-MAC regardless of the condition. However, the CBMAC shows the energy efficiency above a maximum of 58% when an event is arisen. It was found that the energy efficiency is more increased than the B-MAC in general.

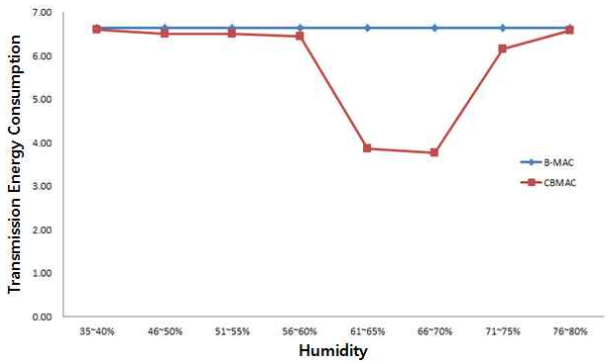


Figure 8. Compare the humidity transmission energy consumption of B-MAC and CBMAC

The figure 9 is the amount of energy consumption obtained by generating events based on the 16~20℃ that is adequate temperature for breeding fed pigs, which is the energy consumption rate obtained when events are arisen between 3~8, 15~28 o'clock by the hour. Comparing energy efficiencies between the CBMAC and the B-MAC, it was found that the former is about 40% higher than the latter.

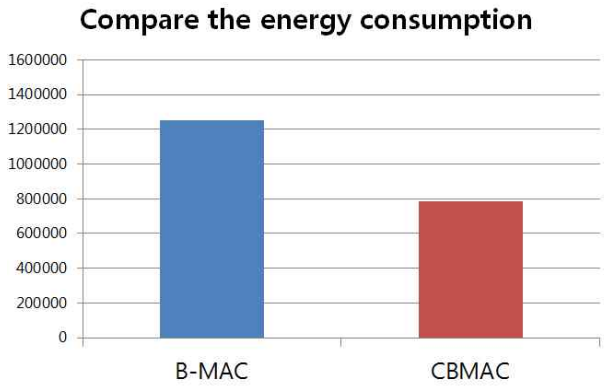


Figure 9. When an event occurs compare the energy consumption of B-MAC and CBMAC

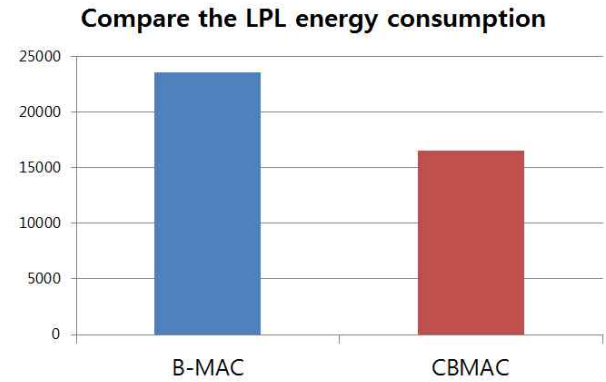


Figure 10. When an event occurs compare the LPL energy consumption of B-MAC and CBMAC

The figure 10 is the decrease rate of the LPL energy obtained by the increase of the sleep interval when events are arisen. It was found that the energy efficiency is about 25% higher than the B-MAC when events are arisen.

5 Conclusions

This paper proposed a method of reducing data attrition rate effectively by assigning an event suited for livestock farm based on B-MAC.

To reduce the number of packet transmission, sensor node in MAC protocol proposed by this paper analyzed livestock growth information and temperature characteristics in Korea and then generated an event table which lists cases when data transmission is unnecessary. When sensor detects a specific event on the event table in the process of collecting data, sensor collects data but doesn't transmit the data to other sensor nodes so that unnecessary data take place. When detecting an event, sensor node, capable of judging that there is no need for data transmission for a certain length of time, extends Sleep interval about twice and reduces LPL energy consumption capable of checking whether or not data took place.

In order to evaluate performance of the proposed CBMAC, the amount of energy consumption was calculated and a simulation was conducted with the calculated values by using the MATLAB. As a result of the simulation, it could be known that the rate of using energy to transmit/receive temperatures was 29% more efficient than the conventional B-MAC, and that it was also up to 58% more efficient for humidities depending on the event arisen. In addition, up to 40% of energy efficiency was taken place when certain events were arisen, and the rate of using the LPL was also up to 25% when events were arisen. Therefore, it could be seen that the proposed MAC protocol is more energy-efficient than the conventional B-MAC.

6 Acknowledgment

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation

7 References

[1] Jeong B.-M., “Foreign u-Farm Service Model Casebook”, Issues and Analysis Report of Korea National Information Society Agency, NCA V-RER-06005, Seoul, Korea, October 2006.

[2] Kwon O.-B., Kim J.-H., “A Basic Direction for Building Agricultural Radio Frequency Identification Logistics Information System”, Korea Rural Economics Institute, Seoul, Korea, December 2007.

[3] Shin Y.-S., “A Study on Informatization Model for Agriculture in Ubiquitous Era”, MKE Research Report, National IT Industry Promotion Agency, Seoul, Korea, 2006.

[4] Ian F. Akyildiz et al., “A survey on Sensor Networks,” IEEE Communications Magazine, Vol.40, No.8, 2002

[5] Chee-Yee Chong, Kumar, S.P. Booz Allen Hamilton, “Sensor networks: evolution, opportunities, and challenges”, Proc. IEEE, Vol.91 No.8, pp. 1247-1256, 2003

[6] Misic, J. Shafi, J. and Misic, V. B. "Avoiding the bottlenecks in the MAC layer in 802.15.4 low rate WPAN," in Proc. of ICPADS, PP. 363-367, 2005

[7] IEEE Std 802.15.4-2006, IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local metropolitan area networks - Specific requirements, Part 15.4: Wireless Medium Access Control(MAC) and Physical Layer(PHY) Specifications for Low-Rate Wireless Personal Area Networks(WPANs), 2006.

[8] W. Ye, J. Heidemann, and D. Estrin, “Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks,” IEEE/ACM Trans. Net., vol. 12, no. 3, June 2004, pp. 493–506.

[9] T. V. Dam and K. Langendoen, “An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks,” 1st ACM Conf. Embedded Networked Sensor Sys., Los Angeles, CA, Nov. 2003.

[10] Polastre, J., J. Hill, et al., “Versatile low power media access for wireless sensor networks”, In SenSys 2004 : Proceedings of the 2nd international conference on Embedded networked sensor systems, New York, NY, USA : ACM Press, (2004), pp.95-107.

Genetic Algorithm-Based Energy-Efficiency via Role Sharing Protocol for Wireless Sensor Networks

J.L. Liu
Department of Information Management, I-Shou University, Kaohsiung, Taiwan

Abstract - *This study proposes a genetic algorithm-based (GA-based) ERoS (Energy-efficiency via Role Sharing) protocol with capable of predicting the optimal probability of cluster heads to enhance the performance of cluster-based wireless sensor networks. The proposed GA-based ERoS protocol, termed ERoS-GA, which basically includes set-up and steady-state phases for each round in the protocol, and a preparation phase is added at the beginning of the first round to obtain the optimal probability of nodes being cluster heads from the base station via genetic algorithm computation. The preparation phase is performed only once before the set-up phase of the first round, and the processes of following set-up and steady-state phases in every round are the same as ERoS. Simulation results show that the proposed genetic-algorithm-based ERoS protocol can produce optimal energy consumption effectively for the wireless sensor networks, resulting in a prolongation for entire network lifetime.*

Keywords: Genetic Algorithm, Clustering Heads, Optimal Probability, ERoS-GA, Network Lifetime.

1 Introduction

Wireless sensor networks (WSNs) have been extensively applied in tactical combat situations, habitat monitoring, home security, and so on [1-3]. Considering that a wireless sensor network is composed of a large number of tiny sensor nodes with limited energy, an energy-efficient network protocol is one of the essentials in the WSN design. As reviewing the proposed protocols in the literature, cluster-based communication protocols can produce superior performance to achieve more balanced patterns of energy use in WSNs [4]. The well-known cluster-based communication protocol was Heinzelman *et al.'s* LEACH, low-energy adaptive clustering hierarchy [5, 6], which energy loads could be well amortized by periodically creating a small number of clusters based on a threshold function T(s) with a priori probability p (say, 5%), in the set-up phase. The technique uses cluster heads (CHs) to aggregate the sensed data from member nodes and forward the aggregated data to base station (BS). Simulation results in [5, 6] show that sensor nodes in the sensor field tend to dissipate the same level of energy over time since the CHs are dynamically rotated among nodes. However, LEACH uses a threshold function parameterized by a probability p that is needed to specify by user. The performance of sensor network is sensitive to the value of p. When p is large, many clusters are formed as a result of high energy consumption since many CHs could dissipate more energy in transmitting aggregated data to the BS. On the other hand, when p is small, only a few clusters are formed, which could increase energy dissipation of member nodes in transmitting sensed data to CHs. Accordingly, some researchers presented that the optimal value p_{opt} depends on parameters such as the total number of nodes distributed in the sensor field, the size of sensor field, the location of BS, and so on [7, 8]. Therefore, present work proposes a genetic algorithm-based (GA-based) energy-efficiency via role sharing protocol, termed ERoS-GA, to predict the optimal values of probability effectively for WSN applications.

1.1 ERoS: An energy-efficiency via role sharing protocol

LEACH as we know it is a stochastic cluster head selection algorithm shown in Fig. 1, which CHs are selected dynamically and periodically according to a threshold function in every round. The operation of LEACH is employed via several rounds, each round consisting of set-up and steady-state phases. In the set-up phase, the sensor field is divided into a small number of clusters. Each cluster includes a CH and several member nodes. Thereafter, in the steady-state phase, each member node transmits its collected data form surroundings to the closest CH, and then each CH receives and aggregates the data from its cluster members and forwards the aggregated data to the BS through a single-hop relay. It is clearly that each CH in LEACH is responsible for cluster management, as well as data aggregation and transmission. This places an excessive energy burden on the CH. Therefore, most protocols try to distribute this energy burden across the network by rotating the CH role between nodes chosen either randomly, or according to some residual-energy metric. However, the use of residual energy in CH selection still yields sub-optimal energy balance and network lifetimes. Based on this insight, we proposed a new protocol called ERoS (Energy-efficiency via Role Sharing) protocol in the previous work [9]. In the ERoS protocol, CHs are selected randomly based on a probability p in each round, yet achieves excellent energy balance by off-loading the data aggregation and transmission functions to other selected nodes. The role of CHs is just to form clusters. Data aggregation within each cluster is performed by an aggregation node, and data transmission to BS by a transmission node.

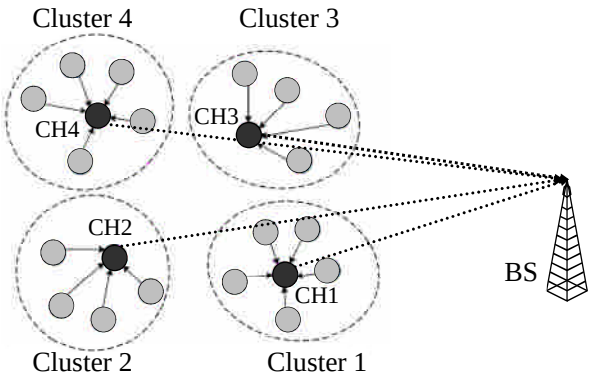


Figure 1. Clustering-based protocol with clusters in the wireless sensor network.

Each round in ERoS is divided into set-up and steady-state phases, just as in LEACH. The selection of CHs is determined purely using a probability value *p*. At the beginning of set-up phase, each sensor node picks a random number *t* in the interval [0, 1]. If $t < p$, the node advertises itself as a CH. Since CHs in ERoS are randomly selected, based only on a specified parameter *p*, several CHs could be located near each other, causing a local imbalance in energy consumption. To distribute the CHs more uniformly, a crowding distance check [10] is applied. After employing this crowding distance check, the surviving CHs advertise themselves to the other nodes in the sensor field via broadcast messages. Cluster formation proceeds with each of the non-CH nodes selecting the closest CH, that is, the CH whose broadcast signal appears the strongest. Picking the closest CH minimizes the energy required for member nodes to communicate with the CH. Each node sends a join-request message to its chosen CH, with its ID, geographical position, and a header. Cluster formation is complete when all nodes in the network have joined a cluster.

1.2 Selections of aggregation and transmission nodes

The energy required to transmit a wireless message over a distance *d* is proportional to d^{α} , where the value α depends on the distance between transmitter and receiver. The values $\alpha = 2$ and $\alpha = 4$ represent the free-space and multi-path fading models, respectively. Since data is aggregated in ERoS by the aggregation node (AN), the total data transmission energy used by nodes within the cluster is minimized when the AN is at the center of the cluster. Accordingly, the AN is chosen by the CH to be the node closest to the cluster center with residual energy higher than the average value within the cluster (see Fig. 2). The AN accepts data packets from member nodes and aggregates them to eliminate redundancy for reducing the size of data.

Each CH also selects the node with the highest residual energy level within the cluster to be the transmission node (TN), as shown in Fig. 2. The TN receives aggregated packets from the AN and forwards them to the BS. Since the TN has the highest residual energy in its cluster, it is the best candidate to transmit data packets to the BS located far from the cluster. Next, the CH sends a message with the IDs of AN and TN to its cluster members via a unique sub-area code. Finally, each CH creates a time division multiple access (TDMA) schedule and a unique spreading code, and transmits them to the members of its cluster. Clearly, the CH plays the role cluster administrator only in ERoS protocol.

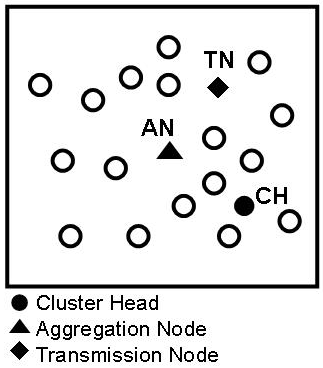


Figure 2. The distribution of CH, AN, and TN in the field.

After clusters have been formed, the steady-state phase begins, and the network starts its transmitting and receiving procedures. In this phase, all cluster members transmit their sensed data to the local AN. The time assigned for each data transmission slot depends on the number of nodes in the cluster. Except when transmitting, the member nodes remain in sleep mode to save energy. As described, ANs aggregate the received data packets and send them to the TN. At the end of the schedule, the TN receives aggregated data from the cluster AN, and forwards them to the BS.

2 Proposed ERoS-GA protocol

Our work introduces a genetic algorithm-based ERoS to determine the optimal value of *p* for various base station placements and sensor sizes. The GA-based optimization procedure is performed only once, before the set-up phase of the first round. The pseudo-code of the proposed protocol is described as follows.

Pseudo-code of the Proposed Protocol
BEGIN
1: Specify the probability (p_{set}), number of nodes (n);
2: $E_{init}(s)=E_0$, s=1,2, …, n;
I. PREPARATION PHASE
1: **if** ($E_{init}(s)>0$) **then**
2: $r \leftarrow$random(0,1);
3: **if** ($r < p_{set}$) **then** //p_{set} can set≥0.5
4: CCH{s}=TRUE; //node "s" be a candidate CH
5: **else**
6: CCH{s}=FALSE; //node "s" not be a candidate CH
7: **end if**
8: **end if**

9: SendToBS(ID_u, (x_u,y_u), CCH(u)) ← All nodes send messages to BS;
10: GAinBS(p_{opt}) ← Optimal probability is determined;
11: BC (p_{opt}) ← BS broadcasts a message back to all nodes;
12: **do** { //repeat for specified rounds

II. SET-UP PHASE

1: t←random(0,1);
2: **if** (E_{init}(s)>0) **then**
3: **if** ($t < p_{opt}$) **then**
4: CCH{s}=TRUE; //node "s" be a candidate of CH
5: **else**
6: CCH{s}=FALSE; //node "s" not be a candidate of CH
7: **end if**
8: **end if**
9: **if** (CCH{s}=TRUE) **then**
10: **if** (distance>distance threshold) **then**
11: CH{s}=TRUE; //crowding distance check
12: **else**
13: CH{s}= FALSE; //give up to be a CH;
14: **end if**
15: **end if**
16: **if** (CH{s}=TRUE) **then**
17: BC (ADV) ← broadcast an advertisement message;
18: Join(ID_i, (x_i,y_i), E(i)); //non-cluster head node "i" join into the closest CH
19: Cluster(c); //form a cluster c;
20: GC{c}← (x_c, y_c); //compute the geometric center
21: **do**{
22: AN{u}=TRUE; //node "u" be the aggregation node
23: } **while** (E(u)> $\overline{E}$(c) & min{dist(u,GC(c))})
24: **do**{
25: TN{v}=TRUE; //node "v" be the transmission node
26: } **while** (E(v)=max{E(c)})
27: **end if**

III. STEADY-STATE PHASE

1: **If** (AN(s)=TRUE) **then**
2: Receive(ID_i, DataPCK) //receive data from members;
3: Aggregate(ID_i, DataPCK) //aggregate received data;
4: TansToTN($ID_{AN,}$ DataPCK); //transmit received data;
5: **else**
6: **If** (MyTimeSlot=TRUE) **then**
7: TansToAN(ID_i, DataPCK); //transmit sensed data;
8: **else**
9: SleepMode(i)=TRUE; //node "i" at a sleep state
10: **end if**
11: **end if**
12: **If** (TN(s)=TRUE) **then**
13: Receive(ID_{AN}, DataPCK); //receive data from AN
14: **If** (MyTimeSlot=TRUE) **then**
15: TansToBS(ID_{TN}, DataPCK); //transmit data to BS;
16: **end if**
17: **else**
18: SleepMode(s)=TRUE; //node "s" at a sleep state
19: **end if**
20: } // one round is completed

END

3 Analysis of energy dissipation in ERoS

We use a first-order radio model [5] in analyzing ERoS protocol. The parameter values used in our simulation model are listed in Table I. According to the radio energy dissipation model illustrated in Fig. 3, the energy required by the transmit amplifier $E_{Tx}(l,d)$ for transmitting a l-bit message over a distance d between a transmitter and a receiver is given by

$$E_{Tx}(l,d)=\begin{cases} l\times E_{elec}+l\times\varepsilon_{fs}\times d^2 & if\ d\le d_0 \\ l\times E_{elec}+l\times\varepsilon_{mp}\times d^4 & if\ d\ge d_0 \end{cases} \quad (1)$$

where $d_0=\sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$ expresses the threshold distance, E_{elec} represents the energy consumption in the electronics circuit to transmit or receive the signals, and the terms of $\varepsilon_{fs}d^2$ and $\varepsilon_{mp}d^4$ represent amplifier energy consumption for a shorter and longer distance transmissions, respectively. To receive the l-bit message, the energy $E_{Rx}(l,d)$ required by the receiver is given by

$$E_{Rx}(l)=l\times E_{elec} \quad (2)$$

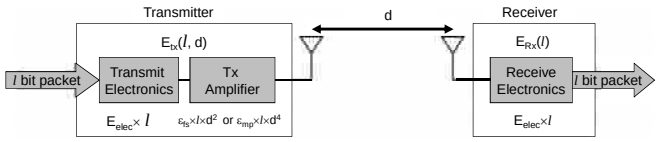


Figure 3. First-order radio model.

Table I. Parameters of the first-order radio model

Parameters	Values
Initial energy (E_0)	0.5 J
Transmitter Electronics (E_{elec})	50 nJ/bit
Receiver Electronics (E_{elec})	50 nJ/bit
Size of Data Packet (l)	2000 bits
Transmitter Amplifier (ε_{fs}) if $d\le d_0$	100 pJ/bit/m^2
Transmitter Amplifier (ε_{mp}) if $d\ge d_0$	0.0013 pJ/bit/m^4

We analyze the energy consumption under the first-order radio model for ERoS as follows. Let a total of n sensor nodes be distributed uniformly in the sensor field of size $M\times M$ (m^2), and be grouped into k clusters. The energy costs required to transmit/receive control messages are neglected in the following analyses of energy consumption, since data packets (l) are far larger than control messages (l_{ctrl}). The energy required per round for an AN to receive data packets from member nodes, and aggregate the received data and forward them over a distance d_{ANtoTN} to the TN is

$$E_{AN}(l,d) = l \times \left[E_{elec}\left(\frac{n}{k} - 1\right) + E_{DA}\frac{n}{k} + E_{elec} + \varepsilon_{fs} \times d^2_{ANtoTN} \right] \quad (3)$$

where E_{DA} represents the energy dissipation for aggregating data. In addition, the energy required per round for a TN to receive aggregated data packets from AN and forward them over a distance d_{toBS} to the BS is

$$E_{TN}(l,d) = \begin{cases} 2l \times E_{elec} + l \times \varepsilon_{fs} \times d^2_{toBS} & if\ d_{toBS} < d_0 \\ 2l \times E_{elec} + l \times \varepsilon_{mp} \times d^4_{toBS} & if\ d_{toBS} \geq d_0 \end{cases} \quad (4)$$

The energy dissipation for a member node, or a non-aggregation node, is

$$E_{non-AN}(l,d) = l \times E_{elec} + l \times \varepsilon_{fs} \times d^2_{toAN} \quad (5)$$

where d_{toAN} represents the distance between a cluster member and its AN. Since the nodes are assumed to be uniformly distributed in the sensor field, the expected value of squared distance from a member node with coordinate at (x, y), to its AN, which located approximately at the center of a cluster in ERoS protocol, is given by

$$E\left[d^2_{toAN}\right] = \frac{1}{A}\iint (x^2 + y^2)dxdy \quad (6)$$

Assuming the shape of clusters is a circle, (6) can be integrated as

$$E\left[d^2_{toAN}\right] = \frac{1}{2\pi}\frac{M^2}{k} \quad (7)$$

The expectation of d^2_{toAN} in (7) is the same as that of d^2_{toCH} in the work of Heinzelman *et al.* [6], since they assumed the CH to be at the center of cluster. Clearly, the assumption used in LEACH is incorrect, since the CHs are not located at the center of clusters in most cases. In the general case, the value of d^2_{toCH} should be the twice that of d^2_{toAN} [10]. Similarly, the expected value of the squared distance from the AN to TN, assuming the TN at (x', y'), also can be approximated as

$$\begin{aligned} E\left[d^2_{ANtoTN}\right] &= \frac{1}{A}\iint (x'^2 + y'^2)dx'dy' \\ &= \frac{1}{2\pi}\frac{M^2}{k} \end{aligned} \quad (8)$$

Clearly, the The expectation of d^2_{ANtoTN} in (8) is the same as that of d^2_{toAN}. Since the energy dissipation for a cluster ($E_{cluster}$) is the summation of E_{AN}, E_{TN} and $E_{non\text{-}AN}$, the total energy consumption for the entire sensor field can be computed by $k \times E_{cluster}$ and formulated as

$$\begin{aligned} E_{total} &= k \times \left(E_{AN} + E_{TN} + (\frac{n}{k} - 1)E_{non-AN} \right) \\ &\approx k \times \left(E_{AN} + E_{TN} + \frac{n}{k}E_{non-AN} \right) \end{aligned} \quad (9)$$

Therefore, the total energy dissipation for a round is given by

$$E_{Total} = \begin{cases} l \times \left[2(n+k)E_{elec} + nE_{DA} + k\varepsilon_{fs}E[d^2_{toBS}] + \varepsilon_{fs}\frac{(n+k)M^2}{2\pi k} \right] & if\ d_{toBS} < d_0 \\ l \times \left[2(n+k)E_{elec} + nE_{DA} + k\varepsilon_{mp}E[d^4_{toBS}] + \varepsilon_{fs}\frac{(n+k)M^2}{2\pi k} \right] & if\ d_{toBS} \geq d_0 \end{cases} \quad (10)$$

where $E[d_{toBS}]$ is the expectation of d_{toBS}. Equation (10) shows that the total energy dissipation is most significantly affected by the distance between TN and BS, and the size of the sensor field. We assume the coordinates of the BS to be (0.5M, 0.5M+B), the values of $E[d^2_{toBS}]$ and $E[d^4_{toBS}]$ can be obtain to be

$$E[d^2_{toBS}] = \frac{M^2}{6} + B^2; \quad E[d^4_{toBS}] = \frac{7M^4}{180} + \frac{2}{3}B^2M^2 + B^4 \quad (11)$$

The corrected equations for total dissipation in LEACH are presented in [10] as follows.

$$E_{Total} = \begin{cases} l \times \left[2nE_{elec} + nE_{DA} + k\varepsilon_{fs}E[d^2_{toBS}] + \varepsilon_{fs}\frac{nM^2}{\pi k} \right] & if\ d_{toBS} < d_0 \\ l \times \left[2nE_{elec} + nE_{DA} + k\varepsilon_{mp}E[d^4_{toBS}] + \varepsilon_{fs}\frac{nM^2}{\pi k} \right] & if\ d_{toBS} \geq d_0 \end{cases} \quad (12)$$

Comparing (10) and (12), we can see that although ERoS increases required power by $2lkE_{elec}$, this increase is small, since number of cluster k is small. However, it reduces the energy consumption roughly by $\frac{l\varepsilon_{fs}nM^2}{2\pi k}$ when $(n-k) \approx n$. Therefore, it is beneficial to assign the AN function to a node whose locates at or near the cluster center rather than the CH.

4 Analysis of optimal number of cluster in ERoS

From the mathematical expressions of (10), the total energy consumption E_{Total} is a function of the number of clusters k. Assuming that $(n+k) \approx n$, the analytical optimal solution for k can be obtained via setting the derivative of E_{Total} to k equal to zero. Therefore, the optimal number of clusters (k_{opt}) and probability for generating CHs (p_{opt}) can be formulated as

$$k_{opt}=\begin{cases}\sqrt{\frac{n}{2\pi}}\frac{M}{\sqrt{E[d_{toBS}^2]}} & \text{if } d_{toBS}<d_0\\ \sqrt{\frac{n}{2\pi}}\sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}\frac{M}{\sqrt{E[d_{toBS}^4]}} & \text{if } d_{toBS}\geq d_0\end{cases}\quad (13)$$

and

$$p_{opt}=\begin{cases}\sqrt{\frac{1}{2n\pi}}\frac{M}{\sqrt{E[d_{toBS}^2]}} & \text{if } d_{toBS}<d_0\\ \sqrt{\frac{1}{2n\pi}}\sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}\frac{M}{\sqrt{E[d_{toBS}^4]}} & \text{if } d_{toBS}\geq d_0\end{cases}\quad (14)$$

Clearly, the value of k_{opt} is approximated as $1/\sqrt{2}$ times the value of LEACH-GA presented in Ref. [10]. When the BS located at the centroid of sensor field, the values of $\sqrt{E[d_{toBS}^2]}$ is given by [7]

$$\sqrt{E\left[d_{toBS}^2\right]}=\frac{1}{A}\int\sqrt{x^2+y^2}\,dA=0.765\frac{M}{2}\quad (15)$$

and the form of p_{opt} can be simplified as

$$p_{opt}=\sqrt{\frac{1}{2n\pi}}\frac{2}{0.765}\quad (16)$$

Equation (16) states that the parameter p_{opt} is just function of the total number of sensor nodes only when the BS located at the center of sensor field. Namely, the value of probability at the center of sensor field is independent of the domain size.

5 Genetic algorithm-based clustering

At the beginning of preparation phase, each node initially determines whether or not it should be a candidate cluster head (CCH), using the following cluster head selection procedure. First, every sensor node selects a random number r from the interval [0, 1]. If r is smaller than p_{set}, based on a prescribed probability p_{set}, then the node is a CCH. The value of p_{set} can be a large value in our protocol, p_{set}=0.5 for example. Thereafter, each node sends its ID, location information, and whether or not it is a CCH to the BS. As the BS receives messages sent by all nodes, it performs GA operations to determine the optimal probability, $p_{opt}= k_{opt}/n$, by minimizing the total amount of energy consumption in each round. Therefore, the objective function used in the GA can be formulated as

$$f(\vec{x})=\sum_{c=1}^{k}\sum_{i=1}^{q}(E_{elec}+\varepsilon d^{\alpha}[i,CCH(c)])\times x_c+\sum_{c=1}^{k}(E_{elec}+E_{DA}+E_{elec}+\varepsilon d^{\alpha}[CCH(c),BS])\times x_c\quad (17)$$

where $\vec{x}=[x_1, x_2, \ldots, x_c, \ldots, x_k]$. The values of x_c are one for our binary-GA when it is a CCH, otherwise, it is zero. The parameters $\varepsilon=\varepsilon_{fs}$ and α=2 were used for $d<d_0$; while, $\varepsilon=\varepsilon_{mp}$ and α=4 were set for $d\geq d_0$. The symbol q represents the number of member nodes in a cluster. The optimal probability p_{opt} is determined by the $1/\sqrt{2}$ times the value of obtain from GA based on (14). Once the optimal probability p_{opt} is found, the BS broadcasts the value of p_{opt} to all nodes. The set-up and steady-state phases begin. The procedures of set-up and steady-state phase are the same as in LEACH.

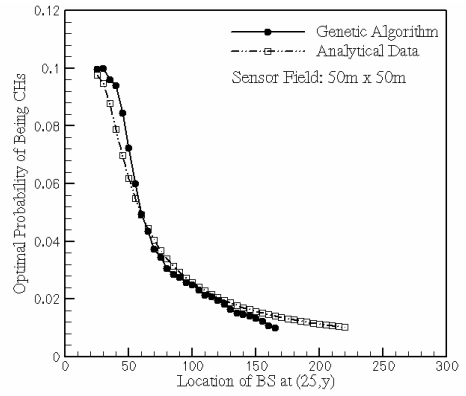
6 Simulation results

This work assumes that all sensor nodes are homogeneous and distributed uniformly over the sensor field with limited energy that the links between nodes are symmetric, and that messages from all nodes can reach the BS. The nodes are distributed randomly in a square of size $M\times M$. Two sizes of sensor field are studied for 50m×50m and 100m×100m domains. In this study, each simulation is repeated for 30 independent runs. In addition, control packet sizes for broadcasting packet and packet header were 50 bits long, and the energy dissipation for aggregating data was 10 nJ/bit/signal.

6.1 Comparison of optimal probability of CHs

The parameters ε_{fs} and ε_{mp} were specified as 100 pJ/bit/m^2 and 0.0013 pJ/bit/m^4, respectively. The total number of sensor nodes was 100. The GA simulation is repeated for 100 independent runs, and solutions are obtained from the average of the runs. Figure 4(a) shows the comparison of optimal probability obtained from model analysis and GA-based computation for a variety of locations of BS for the sensor field of 50m×50m. The comparison depicts that the distribution of present computed optimal probability using ERoS-GA quite agrees with the analytical formulas of (14). Moreover, the result shows that the optimal probability, p_{opt}, is clearly affected by the locations of BS. When the BS is located near the sensor field, the values of p_{opt} are large. On the contrary, the values of optimal probability decrease as the BS moves farther from the sensor field. Figure 4(b) displays the distributions of optimal probability obtained using model analysis and ERoS-GA for a variety of locations of BS for the sensor field of 100m×100m. The results obtained using ERoS-GA was comparable to that of analytical approaches. From Figs. 4(a) and (b), it is showed that the value of probability at the center of sensor fields is independent of the domain size, that is agree with the analytical solution of (16).

(a)



(b)

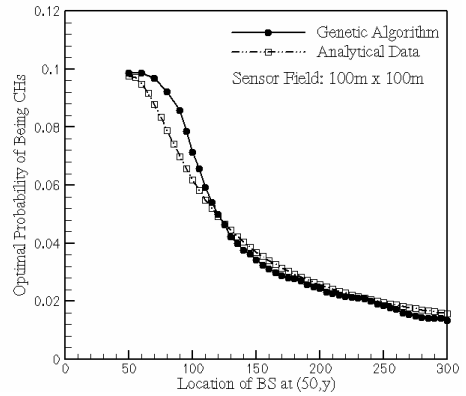


Figure 4. Comparison of optimal probability distributions between analytical analysis and ERoS-GA for the sensor fields with (a) 50m×50m and (b) 100m×100m.

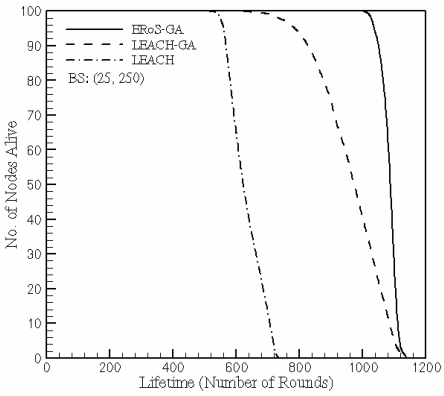
6.2 Comparison of network lifetime

The control packet sizes for broadcasting packet and packet header were 50 bits length for the present computations. Each simulation is also repeated for 30 independent runs, and solutions are obtained from the average of the runs. Table II lists the simulation results obtained using LEACH, LEACH-GA [10] and presented ERoS-GA protocols for BS located at different positions for the sensor field of 50m×50m. The initial energy for all nodes was 0.5(J). The number of rounds required when the number dead of nodes is 1%, 20%, 50%, and 100% are recorded during simulations. From our results, the values of p_{opt} clearly depend on the positions of BS. The value of optimal probability is the largest when the BS is at the center of sensor field, and it decreases when the BS moves outward. Moreover, the proposed ERoS-GA has better performance in most cases than that of LEACH and LEACH-GA in prolonging sensors' lifetime. Figures 5(a) and (b) show the comparisons of performance for BS located at two coordinates of (25, 250) and (25, 350), respectively. Our protocol clearly has excellent performance as compared with other protocols. When the location of BS is far from the sensor field, presented protocol prolongs the lifetime of network significantly since it uses ERoS protocol and with an optimal probability in forming clusters.

Table II. Comparison of network lifetimes (number of rounds) for sensor filed of 50M×50M

BS (25, y)	Protocol	Prob.	Nodes Dead			
			1%	20%	50%	100%
y=25 (center)	LEACH	0.05	1467	1618	1691	1850
	LEACH-GA	0.1307	1610	1732	1818	2040
	ERoS-GA	0.0998	1796	1822	1830	1843
y=50 (border)	LEACH	0.05	1438	1583	1661	1874
	LEACH-GA	0.0946	1512	1663	1717	2078
	ERoS-GA	0.0722	1711	1736	1746	1767
y=100	LEACH	0.05	1346	1473	1543	1787
	LEACH-GA	0.0334	1356	1482	1554	1815
	ERoS-GA	0.0249	1388	1404	1419	1441
y=150	LEACH	0.05	951	1027	1098	1298
	LEACH-GA	0.0181	927	1108	1205	1357
	ERoS-GA	0.0134	1240	1256	1272	1287
y=250	LEACH	0.05	540	576	616	718
	LEACH-GA	0.010	686	874	971	1106
	ERoS-GA	0.010	1059	1073	1085	1097
y=350	LEACH	0.05	220	247	283	360
	LEACH-GA	0.010	407	574	660	757
	ERoS-GA	0.010	748	769	780	797

(a)



(b)

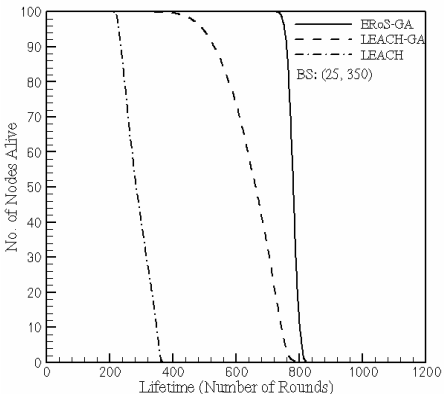


Figure 5. Performance comparisons of network lifetime for the BS located at points of (a) (25, 250) and (b) (25, 350) for sensor filed of 50m×50m.

Table III lists the simulation results obtained using LEACH, LEACH-GA and presented ERoS-GA protocols for BS located at different positions for the sensor field of 100m×100m. Generally, sensor nodes in the large sensor field tend to consume more energy per round for transmitting their sensed data to AN as compared with small one. In this case, the proposed ERoS-GA also showed better performance in most cases than that of LEACH and LEACH-GA in prolonging sensors' lifetime.

Table III. Comparison of network lifetimes (number of rounds) for sensor filed of 100M×100M

BS (50, y)	Protocol	Prob.	Nodes Dead			
			1%	20%	50%	100%
y=50 (center)	LEACH	0.05	718	992	1112	1326
	LEACH-GA	0.1394	1035	1304	1402	1693
	ERoS-GA	0.0986	1284	1351	1360	1379
y=100 (border)	LEACH	0.05	682	927	1049	1278
	LEACH-GA	0.1009	838	1042	1227	1557
	ERoS-GA	0.0714	1068	1187	1205	1264
y=150	LEACH	0.05	596	792	909	1131
	LEACH-GA	0.0483	604	790	900	1101
	ERoS-GA	0.0342	706	909	926	952
y=250	LEACH	0.05	410	503	585	724
	LEACH-GA	0.0261	395	525	611	745
	ERoS-GA	0.01845	535	707	717	726
y=350	LEACH	0.05	192	251	333	448
	LEACH-GA	0.0141	344	457	557	755
	ERoS-GA	0.0100	464	580	590	601

7 Conclusions

This paper proposed a GA-based ERoS protocol, termed ERoS-GA, to determine the optimal probability for cluster formation in WSNs. The LEACH or previous proposed ERoS protocol requires the user to specify a probability to determine whether a node becomes a CH or not. Considering that the probability value (p) for forming clusters in the protocols is difficult to obtain an optimum setting from available prior knowledge. Thus, we designed an additional preparation phase prior to the set-up phase of the first round in the ERoS-GA protocol to gather information about node status, IDs, and location and send it to the BS, which determines the optimal probability to use in the CH selection mechanism. Results showed that the distributions of optimum probability obtained using ERoS-GA was comparable to that of energy model analysis for BS located at different positions in the two sensor fields with 50m×50m and 100m×100m. Moreover, the proposed ERoS-GA method demonstrated good performance in prolonging sensors' lifetime when compared with LEACH and LEACH-GA, since the use of ERoS protocol and the optimal probability can yield optimal energy-efficient clustering.

8 Acknowledgment

The work was supported by National Science Council of Republic of China under Grant Number NSC 100-2221-E-214-040.

9 References

[1] Kemel Akkaya and Mohamed Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," Ad Hoc Networks, Vol. 3, pp. 325–349, May 2005.

[2] Ameer Ahmed Abbasi and Mohamed Younis, "A Survey on Clustering Algorithms for Wireless Sensor Networks," Computer Communications, Vol. 30, pp. 2826–2841, Oct. 2007.

[3] Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, and John Anderson, "Wireless Sensor Network for Habitat Monitoring," ACM WSNA'02, Atlanta, Georgia, pp. 88–97, Sept. 2002.

[4] Dharma. P. Agrawal and Qing-An Zeng, "Introduction to Wireless and Mobile Systems", Pacific Grove, Thomson Brooks/Cole, 2003.

[5] Wendi R. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan, "Energy-efficient Communication Protocol for Wireless Microsensor Networks," IEEE Proceedings of the 33rd Annual Hawaii International Conference on System Science, pp. 1–10, Jan. 2000.

[6] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan, "An Application-specific Protocol Architecture for Wireless Microsensor Networks," IEEE Transactions on Wireless Communications, Vol. 1, Issue 4, pp. 660–670, Oct. 2002.

[7] Seema Bandyopadhyay and Edward J. Coyle, "Minimizing Communication Costs in Hierarchically-clustered Networks of Wireless Sensors," Computer Networks, Vol. 44, No. 1, pp. 1–16, Jan. 2004.

[8] Kun Yang, Yuan-Ming Wu, and Hai-Bo Zhou, "Research of Optimal Energy Consumption Model in Wireless Sensor Network," 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, China, pp. V7-421–V7-424, Apr. 2010.

[9] Jenn-Long Liu and Chinya V. Ravishankar, "ERoS: Role Sharing for Improved Energy Efficiency in Cluster-Based Wireless Sensor Networks", The 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'11), pp. 61–67, Las Vegas, Nevada, July 2011.

[10] Jenn-Long Liu and Chinya V. Ravishankar, "LEACH-GA: Genetic Algorithm-Based Energy-Efficient Adaptive Clustering Protocol for Wireless Sensor Networks", International Journal on Machine Learning and Computing, Vol. 1, No. 1, pp. 79–85, Apr. 2011.

Small Group Relay Protocol using TDMA Contention Slot

Byoungchul Ahn[1], Sang-Ho Hwang[1], Chang-Hyeon Park[1] and Sang-Ho Moon[2],
[1]Dept. of Computer Engineering, Yeungnam University, Gyungsan, Korea
[2]Div. of Computer Engineering, Yeungnam College of Science & Technology, Daegu, Republic of Korea

Abstract - *WiFi and Bluetooth have become attractive wireless communications nowadays. But application area of Blutooth wireless connections are growing slowly. In the field of voice communications for small group, it is very difficult to apply the Bluetooth technology to the Ad-hoc communication although it has Scatternet feature. This paper presents a relay protocol to communicate among several users with Ad-hoc capability. The proposed protocol can relay data or voice to other nodes by multi-hop retransmission using TDMA wireless networks. This TDMA based approach is very efficient for the personal area networks and it easy to implement its function on FPGA. The maximum number of nodes is sixteen nodes for 64Kbps voice communications. The simulation results show that node connection set-up time is very long when the number of nodes is more than 10 nodes. The proposed network delivers very reliable voice packets since the packet loss is 7% when nodes move at the speed of* 6*m/sec.*

Keywords: WiFi, Bluetooth, TDMA, small group relay, Ad-hoc

1 Introduction

Wireless Ad-hoc networks have become in steep research topics and significant advancements in recent times due to their important advantages. One of advantages is easy to set-up a network fast in environments where there is no existing network. For this reason, it can be utilized in emergency rescue operations, military operations, and *etc*. There are, however, several technical challenges in ad-hoc networks. First, ad-hoc networks are characterized by high bit error rates and path breaks due to changing network topology. If the topology change does not occur frequently, people can tolerate packet losses up to 5% according to topology changes[1]. Second is transmission delay or processing delay. In real time transmission such as voice application, there are delay problems. This delay is the transmission delay and processing delay. Processing delay is caused by a codec or any other system processing, and the transmission delay is caused by delivery. The transmission delay is related to the number of hop in the Ad-hoc network. In general, the maximum end-to-end delay bound of voice packet is assumed to 285ms by ITU-T[2]. Thirdly, ad-hoc network have small data payload. The wireless network frame must include not only network information, but also a preamble for synchronization. Therefore, the efficiency of such networks in poor for small data payloads[3].

Considering three major challenges, this paper proposes a protocol to relay data or voice up to 16 nodes to apply for small group voice communications. The routing method is described to manage the network by joining new nodes or disconnecting nodes.

2 Related works

Studies for real-time speech on wireless Ad-hoc are conducted by Jarhl, Kwong and Venkat [4][5][6]. Frank Kargl *et al.* discuss voice transmission over Bluetooth[4]. They present a new routing protocol called Bluetooth Scatternet Routing(BSR). But they discuss the possibility, and have not implemented. Kwong *et al.* use multi-path routing protocol called MSDR in order to speech quality[5]. But the processing overheads are not solved. G. Venkat Raju *et al.* have proposed a Localized Distributed heuristic for Minimum number of Transmissions(LDMT)[6]. In order to reducing transmission delay, this algorithm minimizes voice retransmission.

Several researchers have studied the problem of capacity reduction in multi-hop wireless networks[11][12]. They observe that the performance degrades quickly as the number of hops increases due to using a single radio for transmitting and receiving packets. A good way to improve the capacity of wireless is to use more network interfaces or to use speech compression in the case of voice applications. Some researches use only one network interface due to cost. Another way to improve the capacity of wireless is to use schedule transmission slots in time[13][14][15] and to use multiple non-interrupting frequency channels[16][17][18].

This paper presents a method to communicate whole group by voice. The number of nodes is limited to 16 nodes, which is used for a small group. To relay the voice to other nodes, non-interrupting frequency hopping method is used to avoid collisions.

3 Protocol model

3.1 Routing Protocol

There are a lot of routing protocols in Ad-hoc network. However, most of these routing protocols are not appropriate to communicate voice data because they are designed to communicate data exchanges. A simple routing protocol is proposed for voice communication for a small group. The protocol does not support unicast transmission. Most packets are used to broadcast or multicast transmission

3.1.1 Joining

A node to connect to an already established network must check and try to synchronize the network by receiving a start cycle frame, which is transmitted from master node, and control frames, which are transmitted from slaves. If the received control frame is more than one, a node with the smallest hop count is chosen by the parent node. And to connect to the network, a new node sends "Join_Request_Message" to the parent node through the competition slot. If its parent node received the packet is not master, the node retransmits the packet to master through the control frame. If a master node is received "Join_Request_Message", it assigns an empty slot number for the new node, and transmits it to the new node. If all slots are used, the master sends "Join_Listen_Only_Message".

3.1.2 Path Set-up

The control frame has 96-bit routing information. It contains the sequence number of parent node and descendent nodes' in the network. If the parent node of a node receives the control frame, it stores the slot number of its descendent node in the routing table.

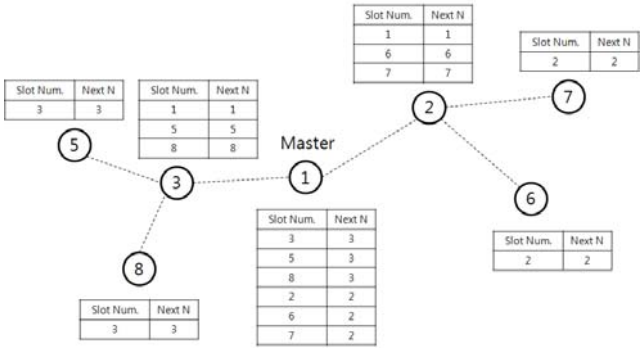


Figure 1. Example of a routing table

For example, if Node-3 at Figure 1 sends its parent the routing information that Node-5 and Node-8 are descendent nodes, the Node-3's parent node receives that packet and updates the "Next Node" in the routing table as number 2, which means Node-5 and Node-8 are connected to Node-3. Also it means that all packets that sent to Node-3, Node-5 and Node-8 are transmitted through Node-3. Node-3 will ignore all routing information of the control frame coming from Node-3's parent node, but the path of Node-3's parent node is stored at Node-3. The information of the parent node in routing table is used to determine the relay path to the master and to check if its parent is alive or dead. Other information in the routing table is used to find the direction toward its descendent nodes.

3.1.3 Path Management

The routing table managed by each node is consisted of Node_Number, Next_Node(link direction), TTL(time to live) and Sequence Number(use to prevent being updated by old routing information). When a node receives the control frame, TTL in routing table is renewed in order to check if. When a node does not receive any control frame, the TTL is reduced. If TTL becomes zero, the path is canceled. Due to frequent topology changes, there could be a possibility to receive other routing information. In this case, the routing table must be updated with new routing information which sequence number is bigger than old routing information.

For synchronize with the master node, all nodes should update their hop count using the control frame of their parent node in every cycle. The master node's hop count is 1 and the hop count of slaves is incremented by one when they join the network. When received parent node's control frame with zero hop(level), the node retransmits zero hop count to its descendant nodes and deletes the route to parent node. Nodes which want to join in the network should not choose the node transmitting control frame with zero hop count for parent node.

3.2 Frame structure

The "Start of Cycle" slot is transmitted by the master node. If neighbor nodes of the master receive the "Start of Cycle" frame, they send "Join Request Message" to the master to join the network during contention period. When the master receives it, the master assigns an empty slot number to participate in the network to that node. Each node which allocated its slot number always broadcasts its control frame in order to notice its connection to its neighbor nodes in the network. All new nodes which want to join to network must transmit "Join_Reqeust Message" to the master using contention period after "Start of Cycle". Payload slots after the control frame are used to data transmission and they are divided into 16 small slots.

3.2.1 Contention Peiod

When a new node wants to join the network, it must use contention period. It should send "Join_Reqeust Message" to the master using this period. If the master receives that message directly, it transmits an empty slot number through its own control frame. If other nodes except the master in the

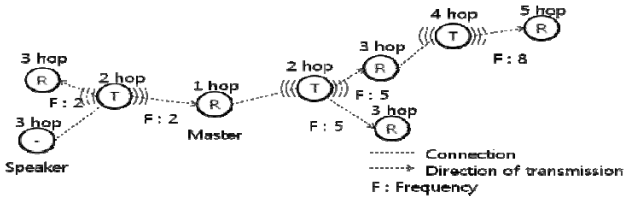
network receive this message, they must relay the message to the master using their own control frame. If there are collisions during the contention period because all nodes may use this period freely, they should send messages after random back-off time in order to reduce wireless collisions.

3.2.2 Start of cycle frame and Control frame

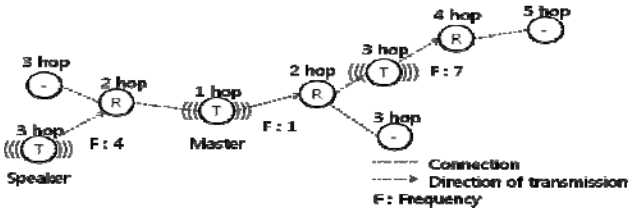
A new node assigned slot number by the master transmits its own synchronous signal named the control frame. All nodes in the network can receive this frame by the dedicated frequencies. Through this signal, new nodes recognize the established network can connect the network. This control frame includes routing information to create a routing table. Therefore, a new node to connect the network can create its own routing table. This frame also contains control command sent by nodes or the master node. Control commands are sent by broadcasting. To prevent loop, control command has sequence ID.

3.2.3 Data frame

Voice packets are transmitted using data slots. Data slots are consists of 16 small slots, and small slots are divided into two sections with eight channels. Nodes transmit and/or receive packets as selected order according to their hop count shown at Figure 2. If the hop count is even, it is receive-transmission mode. If the hop count is odd, it is transmission-receive mode.



a. Even numbered nodes transmit messages



b. Odd numbered nodes transmit messages

Figure 2. Packet Relay Method

Each node has own FHT(Frequency Hopping Table). All nodes should be aware of the FHT of the other nodes. By default, all nodes listen on the frequency of its parent node. If a node is assigned a channel and transmit data, each node sets the dedicated frequency to receive packets according own routing table.

At Figure 3, if node-8 transmits packets, depending on the routing table, node-3 receives packets using the frequency of node-8 and node-1 receives packets using the frequency of node-3, and other nodes receive packets using the frequency of parent node. If a node does not have a path in the routing table, it has to listen to the network using the frequency of the parent node.

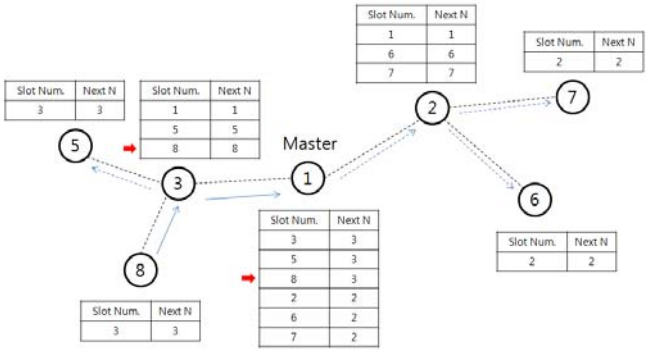


Figure 3. Path selection example

4 Simulation

The proposed protocol is simulated NS-2 network simulator and verified its operation [19]. Table 1 is parameters to simulate the protocol. The total nodes are 16 and mobility is human walking speed. The data bandwidth is 1Mbps and it is enough to sample 16Kbps and transmit voice data. The voice packets are generated by 20*msec*.

Table1. Simulation parameters

Field size	500m x 500m
Number of nodes	16
Mobility	0 – 10 m/s
Bandwidth	1Mbits/s
Application	CBR
Transmitted period	20 ms
Packet size	150 bits
RF Range	50m

When each node receives messages, it does not send ACK to the sender. Therefore a node resends “Join Request” again to the Master if it does not receive ”Join OK” from the Master node in a given time. If several nodes send “Join Request” at the same time, the setup time is increased by collision.

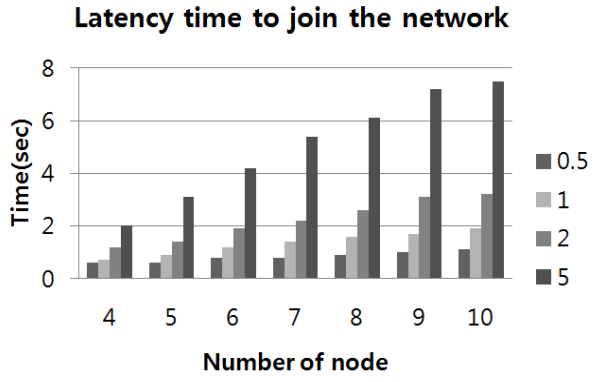


Figure 4. Average Latency time

Figure 4 shows the latency time to join the network. Longer retransmission time makes long latency time to join the network. If the network topology is shown in Figure 5 using sixteen nodes, the connection time is shown in Figure 6. The voice transmission delay of a given multi-hop connection is related to the number of hops. As the number of hops is increased, the delay is increased linearly.

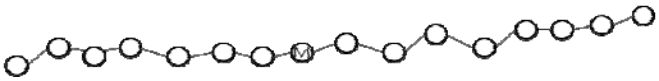


Figure 5. Network topology for the Join latency time estimation

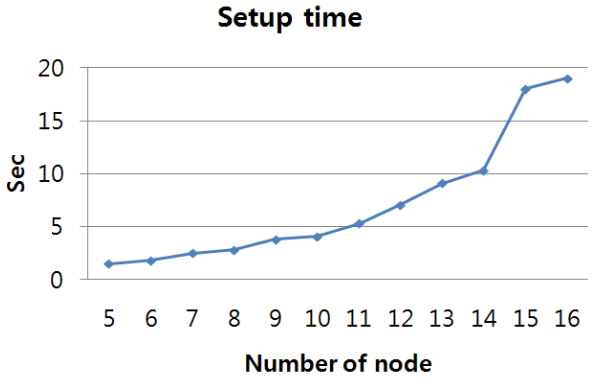


Figure 6. Connection time for Figure 5

Figure 7 shows the packet loss rate according to node mobility. To measure the packet loss, 100 netwotk topologies are generated by randomly, 1000 messages are transmitted at each topology and loss rates are measured for each transmission. Figure 7 shows average loss rates.

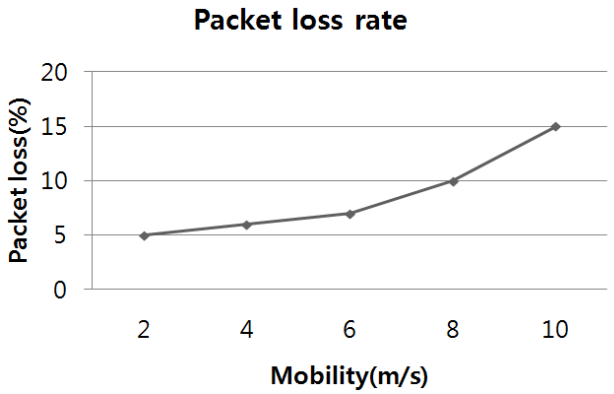


Figure 7. Packet loss rate for mobility of nodes

5 Conclusions

A new relay protocol is presented to use a small group voice communication. The simulation results show that node connection set-up time is very long when the number of nodes is more than 10 nodes. But the set-up time can be reduced by adjusting retransmission waiting time. Data loss rate is increased as the mobility speed is increased. When nodes move at speed of 6m/sec, the packet loss rate is 7 percent. This means that the proposed network delivers stable voice transmission for human sports activities. In the future, more experiments are required for fast topology changes. The proposed protocol is planned to implement on FPGA and verify its functions late 2012.

Please address any questions related to this paper to Prof. Ahn by Email (b.ahn@yu.ac.kr).

6 References

[1] N. Jayand and S. W. Christensen, “Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample-Interpolation Procedure”, IEEE Transactions on Communications, vol 29, no. 2, pp. 101-109, February 1981.

[2] ITU-T Recommendation G.114, “One-way transmission time”, May 2003.

[3] Chi-hsien Lin, Hui Dong, Upamanyu Madhow, and Allen Gersho, “Supporting Real-time Speech on Wireless Ad-hoc Networks: Inter-packet Redundancy, Path Diversity, and Multiple Description Coding” in Proceedings of ACM workshop on WMASH pp. 11-20, October 2004.

[4] Frank Kargl et al, "Bluetooth-based Ad-Hoc Networks for Voice Transmission", in Proceedings of 36th Annual Hawaii International Conference on System Sciences, January 2003.

[5] M. Kwong, S. Cherkaoui, R. Lefebvre, "Multiple description and multi-path routing for robust voice transmission over ad-hoc networks", in IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob, pp. 262-267, 2006.

[6] G. Venkat Raju, T. Bheemarjuna Reddy Shyamnath Gollakota, and C. Siva Ram Murthy, "A near optimal localized heuristic for voice multicasting over ad-hoc wireless networks", in Communications, 2007 ICC'07. IEEE International Conference on, pp. 1648–1653, June 2007.

[7] D.B. Johnson, D.A. Maltz, "Dynamic Source Routing in Ad-hoc Wireless Networks", in Computer Communications Review – Proceedings fo SIGCOMM' 96, Aug. 1996.

[8] S. Corson, J. Macker, "Mobile ad-hoc networking (MANET): Routing protocol performance issuse and evaluation considerations", IETF 1999.

[9] C. Perkins, E. Royer, "Ad-Hoc On-Demand Distance Vector Routing", Proceedings of 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 153-181, Feb. 1999.

[10] T. Camp, J.Boleng, V.Davies, "A survey of mobility models for ad-hoc network research" , Wireless Communications & Mobile Computing(WCMC): Special Issue on Mobile Ad-hoc Networking: Research, Trends, and Applications, Vol. 2, no. 5, pp. 483-502, 2002.

[11] P. Gupta and P. R. Kumar, "The capacity of wireless networks", IEEE Trans on Info Theory, Mar 2000.

[12] J. Li, C. Blake, D. S. J. De Couto, H. I. Lee, and R. Morris, "Capacity of ad-hoc wireless networks", in MOBICOM, 2001.

[13] T. -W. chen, J.T. Tsai, and M. Gerla, "QoS routing performance in multihop, multimedia, wireless networks", in Proceedings of IEEE ICUPC'97 1997.

[14] Yu-Ching Hus, Tzu-Chieh Tsai, Ying-Dar Lin, and Mario Gerla, "Bandwidth routing in multi-hop packet radio environment", in Proceedings of the 3rd International Mobile Computing Workshop, 1997.

[15] Chunhung Richard Lin, "On-demand QoS routing in multihop mobile networks", in Proc. IEEE INFOCOM, April 2001.

[16] A. Nasipuri and S. R. Das, "A multichannel CSMA MAC protocol for mobile multihop networks", in WCNC, 1999.

[17] Z. Tang and J. J. Garcia-Luna-A , "Hop-reservation multiple access (HRMA) for ad-hoc networks", in INFOCOM, 1999.

[18] Anastassios Michail and Anthony Ephremides, "Algorithms for routing session traffic in wireless ad-hoc networks with energy and bandwidth limitations" in Proceddings of 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2001.

[19]] ns-2: network simulator, http://www.isi.edu/nsnam/ns/

SESSION

SECURITY + MIMO

Chair(s)

TBA

A Quest for Security in Wireless Sensor Networks using Game Theory

Mehran Asadi, Christopher Zimmerman, and Afrand Agah
Department of Computer Science, West Chester University, West Chester, PA, USA

Abstract— *In this paper, we propose a protocol that uses game theory to prolong the battery life of a wireless sensor network and to address the presence of nodes that may act maliciously by not sending packets in order to save their own battery power. We accomplish this by helping the sensors optimize their decision making process about whether or not to forward any data packets they may receive. By using game theory, we attempt to find an optimum configuration that will extend node's battery life while still allowing nodes to forward an acceptable amount of packets through the network. Through simulation, we evaluate the proposed protocol based on packet throughput and conservation of battery power.*

1. Introduction

A Wireless Sensor Network (WSN) consists of wireless sensors, small devices that collect data readings such as light or temperature from an environment. The sensors then send the data to a base station, a central location for the data to congregate. Wireless sensor networks have potential to revolutionize the way in which the real world is monitored and controlled. Also, such networks impose a series of security challenges to network designers. Among these security problems, Denial of Service (DoS) attacks, defined as any event that diminishes or eliminates a network's capacity to perform its expected function, degrade networks' intended services to its users. One simple form of a DoS attack is vulnerability by arbitrarily neglecting to route some messages. A subverted or malicious node can still participate in lower-level protocols, and may even acknowledge reception of data to the sender, as it drops messages on a random or arbitrary basis. Such a node is neglectful. The dynamic source routing protocol is susceptible to this attack. Because the network caches routes, communications from a region may all use the same route to a destination, and a malicious node can degrade or block traffic from a region to a base station. Game theory is a field of study that attempts to model decision making which has been used in various fields such as economics, politics, biology [20]. Game theory has previously been applied to wireless sensor networks, but within the context of modeling multiple nodes in the network attempting to share a shared medium: their radio communication channels [12]. We use game theory in our project for the purpose of extending a sensor's battery life. We accomplish this by helping the sensors optimize their decision making process about whether or not to forward any data packets they may receive. On one hand, if a node decides to never forward any packets, it conserves its battery power, but no data flows through the network. However, if a node forwards every packet that it receives, that node demonstrates its reliability and traffic flows through the network but the node will run out of battery power much faster than if the node were to not forward any packets. By using game theory, we attempt to find an optimum configuration that will extend a node's battery life while still allowing the node to forward an acceptable amount of packets through the network.

This paper is organized as follows. Section II reports the related work. Section III formulates the game. Section IV evaluates the performance of the proposed protocol, and Section V concludes the paper.

2. Related Work

Numerous techniques have been proposed in recent years for estimating battery lifetime. In addition, a variety of strategies have been proposed to exploit battery characteristics for designing more battery friendly systems and communication protocols. Authors in [21] report on systematic experiments that conducted to quantify the impact of key wireless sensor network design and environmental parameters on battery performance.They evaluated the extent to which known electrochemical phenomena, such as rate-capacity characteristics, charge recovery and thermal effects, can play a role in governing the selection of key wireless sensor networks design parameters such as power levels, packet sizes. They have also analyzed the non-trivial implications of battery characteristics on wireless sensor networks power control strategies, and find that a battery-aware approach to power level selection leads to a 52% increase in battery efficiency. The most work in this area relies on simulation of generic battery models.There are a number of approaches of energy management in sensor networks, including topology management and network layer optimization. Authors in [8] empirically examine the gain of battery runtime due to the battery recovery effect, and found this effect significant and dependent on duration. They also proposed a more energy-efficient duty cycling scheme that is aware of battery recovery effect, and analyzed its performance with respect to the latency of data delivery. The benefit of behaving well is

not obvious in the case of a delay between granting a favor and repayment, which is when nodes of a wireless sensor network forward for each other [7].

3. Proposed Protocol

It is our interest to investigate how selfish behavior by individual players may affect the performance of the network as a whole. In a wireless sensor network, each node generates its own data and forwards traffic for others. Forwarding others traffic can consume a considerable amount of battery life.

3.1 Game Formulation

In this paper, we first aim to study the mathematical modeling of battery discharge behavior in a wireless sensor network. Each player tries to maximize its own benefit, which is the available battery of each individual node. However if a node forwards all incoming packets then over time the node would diminish its own energy reserves. Based on this, nodes have a tendency of not forwarding packets and acting selfishly to conserve energy. The point of this paper is to give incentives to those nodes that participate in the network activities by forwarding incoming packets. Solving this problem means finding a Nash equilibrium [19] for the whole network, whereas each node is pre-programmed with a set of rules, maximizing the payoff for the entire network. We assume that each node has a discrete representation for its remaining energy, and the incentive for each node is to have a better reputation, where each node can be positively or negatively affected by its reputation. Over time, nodes with low reputation can be isolated and labeled as selfish/malicious nodes, and at each node, there is a trade-off between saving energy resources and maintaining their reputation. A game is formulated as $G =< N, A, \{u_i\} >$where N is the set of players (decision makers), A_i is the action set of player i, $A = A_1 * A_2 * ... * A_n$ is the Cartesian product of the sets of actions available to each player, and $\{u_i\}$ is the set of utility functions that each player i wishes to maximize, where $u_i : A \rightarrow \Re$. Our proposed framework enforces cooperation among nodes and provides punishment for non-cooperative behavior. We assume that the rational users optimize their profits over time. The key to solve this problem is when nodes of a network use resources; they have to contribute to the network life in order to be entitled to use resources in the future. The base station keeps track of the behavior of other nodes, and as they contribute to common network operation, their reputation increases. We are interested in solving a game by predicting the strategy of each player, considering the information that the game offers and assuming that the players are rational.

3.2 Equilibrium

We formulate a model that captures a situation in which two bargainers have the opportunity to reach agreement on an outcome in some set X and perceive that if they fail to do so then the outcome will be some fixed event D. Here the set X is the set of feasible divisions of good reputation and D may be the event in which neither party receives any reputation. The set of Nash equilibria of a bargaining game of alternating offers is very large. One such equilibrium is that in which both players always proposes x^* and always accept a proposal x if and only if $x = x^*$. For any agreement x and period t, there is a Nash equilibrium for which the outcome is the acceptance of x in period t [20]. One such equilibrium is that in which through period $t-1$, each player demands the maximum reputation and rejects all proposals, and from period t on proposes x and accepts only x. The procedure we study is one in which the players alternate offers in periods of the game. The first move of the game occurs in period 0, when player 1 makes a proposal (forward my incoming packet), which player 2 then either accepts or rejects.

3.3 Payoff and Reputation

Each node i has a von Neumann-Morgenstern utility function [19] defined over the outcomes of the stage game G, as $u_i : A \rightarrow \Re$, where A is the space of action profiles. A's action profile space is listed below:

$$A = \begin{cases} Act_1 & \text{forward packet} \\ Act_2 & \text{don't forward packet} \end{cases}$$

Let G be played several times and let us award each node a payoff which is the sum of the payoffs it received in each period from playing G. Here,

$$u_i^t = \alpha r_i^t - \beta c_i^t$$

where r_i^t is the gain of node i's reputation, c_i^t is the cost of sending or forwarding a packet for the node as energy loss, and α and β are weight parameters. We assume that measurement data can be included in a single message that we call a packet. Packets all have the same size. The transmission cost for a single packet is a function of the transmission distance [22]. At time t, each node calculates the utility to be gained for each of the two actions available. For forwarding a packet, the utility is calculated as:

$$u_{A_1}^t = T * r_i^{t+1} - B * (c_s + c_r)$$

where r_i^{t+1} is the predicted gain of node i's reputation. For sending a packet, c_i^t is broken down into two constant values: c_s and c_r. c_s is the voltage cost to send a packet and c_r is the voltage cost to receive a packet. B is the weight parameter for cost. B represents the importance of being conservative about sending packets when a node has a low battery leave. T is the weight parameter for the gain component of the

Table 1: Parameters and Notations

Cost of forwarding packet at node i	c_i
History at node i	h_i
Rating of node i	ρ_i
Reputation at node i	r_i
Utility at node i	u_i
Weight Parameters	α_i, β_i

equation. T represents the number of units of time since node i has last forwarded a packet. T starts at 1 for each node i and increments every time any node i decides to not forward a packet. When a node sends a packet, T is reset back to 1. If a node has recently sent a packet, it may not be important to send another packet right away, which is why T starts at a low value. But as time passes without forwarding any packets, it is important that a node sends data through the network, which leads T to increase. The utility for not forwarding a packet is calculated as:

$$u^t_{A_2} = T * 0 - B * c_s$$

Since there is no gain in reputation when not sending a packet, the gain is 0. However, receiving a packet from another node still costs energy. After calculating the utility for each of these actions, the node will perform the action that yields the greater utility. The strategy for each node i at time t is:

$$s_i(h^t) = \begin{cases} Forward & \text{if } u^{t+1}_{A_1} > u^{t+1}_{A_2} \\ Do\ not\ forward & \text{otherwise} \end{cases}$$

In order to compute the values of a node's gain, we turn our attention to the work proposed in [15]. In this work the authors proposed the concept of subjective reputation, which reflects the reputation calculated directly from the subject's observation. In order to compute each node's reputation at time t, we use the following formula:

$$r^t_i = \sum_{k=1}^{t-1} \rho_i(k)$$

where $\rho_i(k)$ represents the ratings that the base station has given to node i, and $\rho_i \in [-1, 1]$.

3.4 Configurations

We use two major network configurations. The first configuration, named case1, consists of a network of wireless sensors which broadcast packets to any nodes within range. Since the nodes in our experiment are located within a small distance of each other, all nodes in the network are capable of broadcasting directly to every other node in the network. Whenever a node receives a packet from another node and forwards the packet, that packet is re-broadcast to every node within range. For networks of a large size, this generates a large amount of traffic. In an attempt to remedy this, another network configuration was developed. Case2 utilizes a neighbor system. Each node has a neighbor table that holds the IDs of several neighbors, which are determined by a handshaking process that occurs after the nodes boot up and send an initial voltage reading to the base station. The neighbor relationship is bi-directional. Whenever a node receives a packet, it checks the data. This gives the ID number of the node that just sent the packet. If the ID found in path of the packet is not found in the neighbor table of the receiving node, the node ignores the packet and no further action is taken. However, if the ID of the node that just forwarded the packet matches an ID in the neighbor table, then the node's number of packets received is incremented and the node will take the appropriate action with the packet. Since we did not have access to a large testing area where we could spread the nodes out further, this is an attempt to emulate a less dense, less traffic-heavy network than what is found in case1.

3.5 Malicious Node Detection

In our simulations, we introduce malicious nodes into the network to see how they affect the network and if there is a way to detect and neutralize such nodes. Malicious nodes randomly drop packets, reducing the throughput of the network. Malicious nodes also consume additional power when randomly deciding whether or not to drop packets. The base station keeps track of the reputation of each node in the network. Periodically, the base station will decide whether or not a node is acting malicious based on its throughput. The base station takes the current reputation of each node in the network and calculates the average, as well as the standard deviation. If a node's reputation is lower than the average minus the standard deviation, that node is deemed malicious. The base station sends a packet to that node ordering the node to turn its radio off and shut down.

4. Performance Evaluation

In the case1 scenarios, the simulation starts with an initial voltage reading from each sensor. Next, packet is broadcast once every 200 milliseconds for 300 seconds. Then a final voltage reading is sent to the base station. In the case2 scenarios, after the initial voltage reading, the neighbor handshaking phase takes place. After the neighbor handshaking process, each node broadcasts data once every 200 milliseconds for 300 seconds. Lastly, a final voltage reading is sent to the base station. During the simulation, if a node receives a packet it will forward it or apply the game theory strategy, depending on the scenario. After sending the packets, each node turns its radio off for 10 seconds to get rid of the traffic in the network. Then, every node turns their radio on and sends one final voltage packet to the base station. This gave us a clear start and end voltage for calculating voltage loss. For malicious node detection, the base station checks to see if any nodes are malicious

after 60 seconds into the simulation, and then once every 30 seconds after that. Any nodes that are deemed as malicious are turned off via radio.

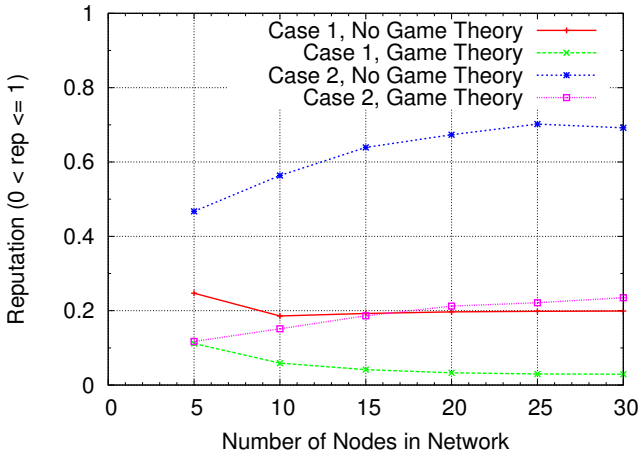


Fig. 1: Average network throughput for normal nodes.

As shown in Figure 1, reputations for simulations using game theory have a lower reputation, than simulations not using game theory, regardless of network size. By implementing game theory, the average throughput of the network drops, but this is to be expected since the sensors are dropping packets based on a set of rules in order to save power.

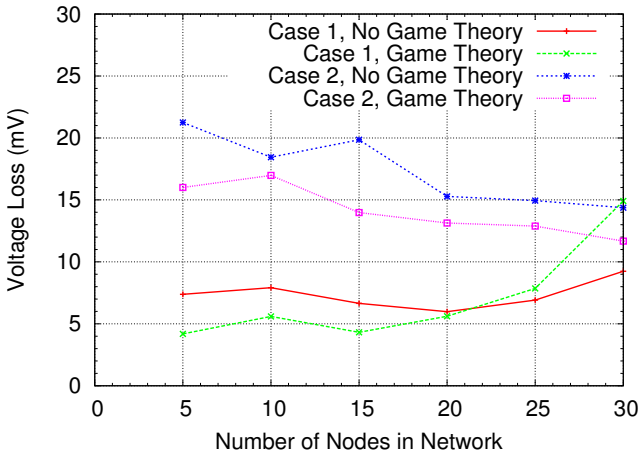


Fig. 2: Average network voltage loss for normal nodes.

Figure 2 shows that for case1, implementing game theory results in a lower voltage loss for smaller networks, but results in a greater voltage loss for larger networks. As the size of a network increases, so does the traffic. Since deciding whether or not to forward a packet by using a strategy also consumes power, there is a point where the frequency of deciding whether or not to forward an incoming packet is so high that the energy used for implementing the strategy is greater than the amount of energy the node tries to save by not forwarding packets. For case2, implementing game theory consistently results in a lower network voltage loss. The neighbor system helps reduce the amount of traffic in the network, which prevents the voltage loss that happens with larger networks in case1.

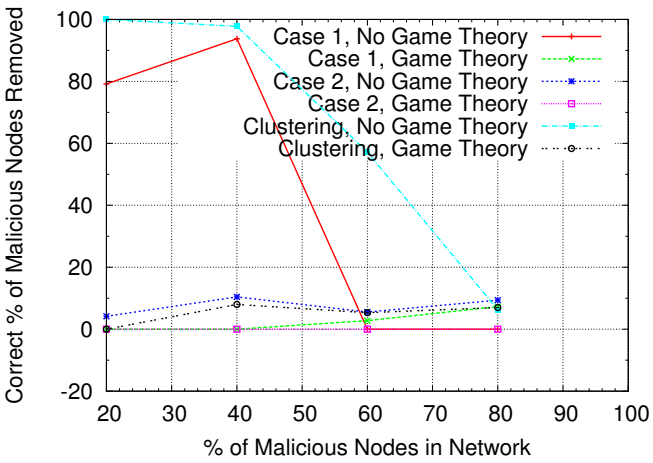


Fig. 3: Average percentage of malicious nodes correctly removed from network.

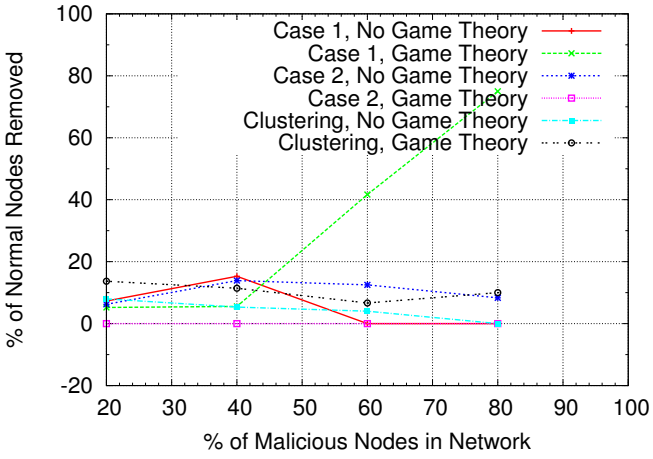


Fig. 4: Average percentage of normal nodes incorrectly removed from network.

As indicated by figures 3 and 4, our procedure for determining malicious nodes needs work. Aside from our first configuration that does not use game theory, a low percentage of malicious nodes are successfully removed from the network. Also, our procedure turns off some normal nodes in the network, which lowers the performance for the network.

As seen in Figure 5, average reputations for game theory cases are lower than non-game theory cases. The ineffectiveness of our procedure to detect and disable malicious nodes

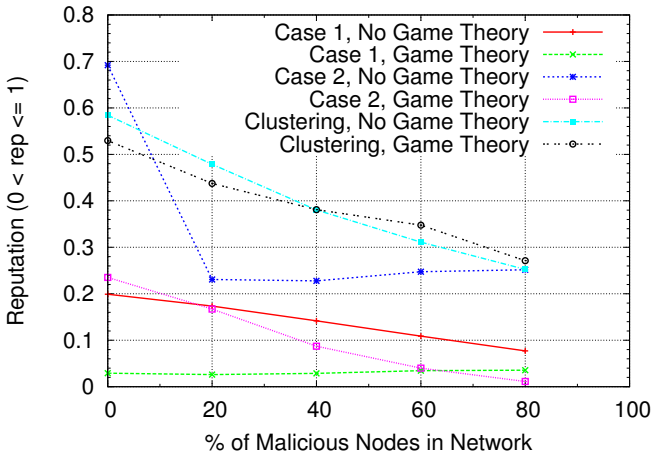


Fig. 5: Average network throughput for malicious nodes.

may have some influence in these results.

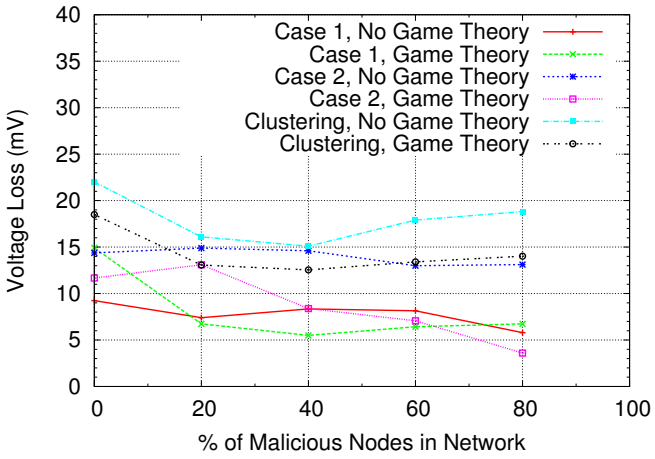


Fig. 6: Average network voltage loss for malicious nodes. Broadcast(case 1), hop-by-hop(case 2)

As seen in Figure 6, in most cases, voltage loss is lower with game theory implemented than if not, even with the presence of malicious nodes. Once again, the ineffectiveness of our procedure to detect and disable malicious nodes may have some influence here.

Due to how utility is modeled in our project, utility is bound to decrease. Therefore, a better utility is not defined by how quickly it can rise, but rather how slowly it can decrease. As seen in Figure 7, for our case1 scenarios, utility for networks implementing game theory have a lower utility than those which do not implement game theory. However, this is caused by the high amount of traffic in a larger network. As seen in Figure 8, for our case2 scenarios, utility for networks implementing game theory have a higher utility than those which do not implement game theory. Despite the large network size, the neighbor system reduces the amount

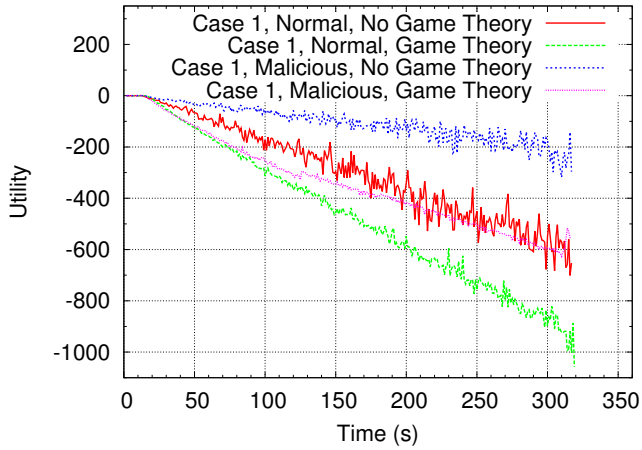


Fig. 7: Average network utility for Broadcast (case 1) scenarios (Network size - 30 nodes).

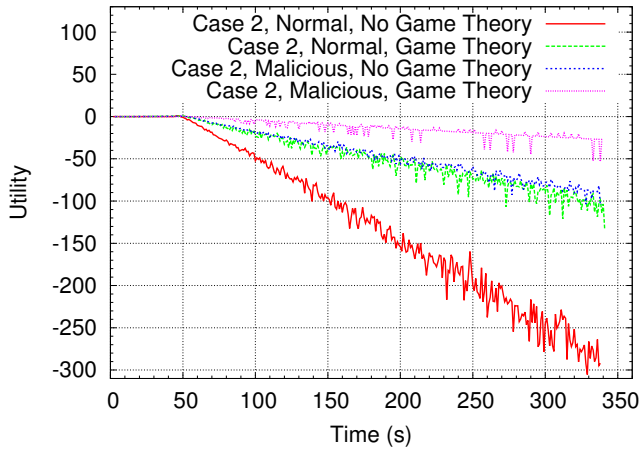


Fig. 8: Average network utility for hop-by-hop (case 2) scenarios (Network size - 30 nodes).

of traffic that flows through the network.

5. Conclusion and Future Work

Our results indicate that, under most cases, implementing game theory in a WSN is beneficial by helping reduce the amount of voltage consumption throughout the network. By adding a decision making process of when to send and not send packets, the sensors conserve energy while maintaining an acceptable amount of throughput. Further work includes experimenting with different strategies in order to save power, as well as improving our procedure of how to detect and neutralize malicious nodes. Other possible extensions of this project would be to experiment with packets of different priorities and implement coalitions of nodes.

Acknowledgements

This work is supported by the National Science Foundation under grant number 1054492.

References

[1] C. Zimmerman, A. Agah and M. Asadi, "Applying Economical Modeling to Wireless Sensor Networks for Maximizing the Battery Life," *The 26th Pennsylvania Computer and Information Science Educators (PACISE) Conference*, Poster Presentation, Shippensburg, PA, 2011.

[2] C. Zimmerman, A. Agah and M. Asadi, "Incorporating Economical Modeling to Extend Battery Life in Wireless Sensor Networks," *Graduate Research and Creative Projects Symposium*, Harrisburg, PA, 2011.

[3] A. Agah, S. K. Das and K. Basu, "Preventing DoS attack in Sensor and Actor Networks: A Game Theoretic Approach," *IEEE International Conference on Communications (ICC)*, 2005.

[4] A. Agah, S. K. Das and K. Basu, "Enforcing Security for Prevention of DoS Attack in Wireless Sensor Networks using Economical Modeling, " *Proceedings of the 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS)*, 2005.

[5] I. F. Akyldiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, 2002.

[6] S. Buchegger and J. L. Boudec, "Performance Analysis of the CONFIDANT Protocol Cooperation Of Nodes Fairness In Dynamic Ad-hoc NeTworks," *International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2002.

[7] S. Buchegger and J. L. Boudec, "Nodes bearing grudges: toward routing security fairness and robustness in mobile ad hoc networks," *Proceedings of the 10th Euronicro Workshop on parallel, Distributed and Network-based Processing*, 2002.

[8] C. chau, M. H. Wahab, F. Qin, Y. Wang and Y. Tang, " Battery Recovery Aware Sensor Networks," *the 7th International symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2009.

[9] C. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenge," *Proceedings of the IEEE, special issue on sensor networks and application*, vol. 91, no. 8, 2003.

[10] S. Eidenbenz, G. Resta and P. Santi, "The COMMIT protocol for truthful and cost-efficient routing in ad hoc networks with selfish nodes," *IEEE transactions on mobile computing*, vol.7, no.1, 2008.

[11] M. Felegyhazi, L. Buttyan and J. P. Hubaux, "Equilibrium Analysis of Packet Forwarding Strategies in Wireless Ad Hoc Networks the Static Case," *Proceedings of Personal Wireless Communications*, 2003.

[12] M. Felegyhazi and J.P. Hubaux, "Game Theory in Wireless Networks: A Tutorial," *EPFL - Switzerland, LCA-REPORT*, 2007.

[13] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," *In First IEEE International Workshop on Sensor Network Protocols and Applications*, SPNA, 2003.

[14] R. Machado and S. Tekinay, "A survey of game-theoretic approaches in wireless sensor networks," *Elsevier Computer Networks journal*, vol. 52, 2008.

[15] P. Michiardi, R. Molva, "Core: A Collaborative Reputation mechanism to enforce node cooperation in mobile ad hoc networks," *Communications and Multimedia Security Conference*, 2002.

[16] P. Michiardi, R. Molva, "Prevention of denial of service attack and selfishness in mobile ad hoc networks," Research Report RR-02-063, Institute Eurecom, France, 2002.

[17] P. Michiardi and R.Molva, "Game theoretic analysis of security in mobile ad hoc networks," *Institute Eurecom*, Research Report, France, 2002.

[18] P. Michiardi and R.Molva, "Simulation-based analysis of security exposures in mobile ad hoc networks," in *European Wireless 2002: Next Generation Wireless Networks: Technologies, Protocols, Services and Applications*, 2002.

[19] G. Owen, *Game Theory*, 3rd ed. New York, NY: Academic Press, 2001.

[20] M. Osborne and A. Rubinstein, *A Course In Game Theory*,The MIT Press, 1994.

[21] C. Park, K.Lahiri and A. Raghunathan, " Battery Discharge Characteristics of Wireless Sensor Nodes: An Experimental Analysis, " *IEEE SECON*, 2005.

[22] T. S. Rappaport, "Wireless Communications: Principles and Practice," *Prentice Hall*, 2002.

Capacity of MIMO Fading Channels

A. Riaz ul Hassnain Syed [1], **B.** Tariq Adnan[2], **and C.** Ammara Masood [1]
[1] *University of Engineering and Technology Peshawar, Abbottabad Campus PAKISTAN*
[2]University of New South Wales Australia.

Abstract - *Multiple antenna system are projected to contribute a key role in potential multimedia wireless communication systems in the future. Such systems are predicted to provide high channel capacities in a limited bandwidth, which enables the efficient usage of spectrum. Essentially to the realization of these channel capacities are the characteristics it displays especially when analyzing the ergodic capacities and the outage capacities. To characterize the channel, certain methods have been employed. This paper reviews the basic background of MIMO systems and discusses the channel capacity using uniform power allocation in terms of ergodic and outage capacities.*

Keywords: MIMO, Capacity, Fading channels, Ergodic capacities, outage capacities

1 Introduction

Various channel models have been presented over a period bits/sec, of the communication for each MIMO user can be improved. Therefore, this directly increases the network capacity and quality of service. This development enables MIMO systems to be viewed as an extension of the so-called radio propagation between two or more points. Radio propagation is an important aspect of any radio design or radio network planning. Therefore, a variety of analyses through process models have been identified. [1] One of the models is the MIMO configuration.

In the digital communication field, the recent introduction of MIMO - multiple input multiple output or can also known as volume-to-volume 'wireless link', has appeared as one of the most significant breakthrough in modern communications. [2] One of the main motivations for this advancement is the chance of resolving the bottleneck of traffic capacity in future Internet-intensive wireless networks.

To define MIMO systems simply, given an arbitrary wireless communication system, we consider a link for which the transmitting end as well as the receiving end is equipped with multiple antenna elements.[2] Saying this, it can clearly be seen that MIMO architectures are a form of receive and transmit diversity scheme. In addition to that, this form of configuration involving multiple antennas on the input and the output gives us the ability to dramatically improve capacity gains through increased spatial dimension. MIMO systems can provide significant increase in data throughput and link range without additional bandwidth or transmit power.

This basically means that the signals on the transmitting antennas on one end and the receiving antennas on another end are 'combined' in such a manner that the quality, with references to signal to noise ratio (SNR) and the data rates in bits/sec, of the communication for each MIMO user can be improved. Therefore this directly increases the network capacity and quality of service. This development enables MIMO systems to be viewed as an extension of the so-called 'smart antennas', where it uses antenna arrays along with signal processing algorithm to improve wireless transmissions. Multipath propagation, conventionally known to be a drawback of wireless communications, was made a turn around with the introduction of MIMO systems. This became a key feature of MIMO, where it uses multipath propagation properties to multiply transfer rates. A high rate signal is split into multiple lower rate streams. Then each of these stream is transmitted from a different transmit antenna in the same frequency channel. This is known as multipath, the signal taking different routes to reach the receiver. The different signals will then arrive at the receiver antenna with different multipath delay spread. Then the receiver can proceed to separate these different signals, creating parallel channels for free. Spatial multiplexing is a very powerful technique for increasing channel capacity at higher Signal to Noise Ratio (SNR). [3] The panorama for magnitude improvement in wireless communication performance at no cost of extra bandwidth is largely responsible for the success of MIMO. This has encouraged tremendous progress in diverse areas such as channel modeling, information theory and coding, signal processing, antenna design and multi-antenna-aware cellular design, fixed or mobile. [3] This paper covers the capacity of MIMO going through a channel experiencing Rayleigh fading. The first portion of this paper will be introducing the system model of MIMO systems. Follow up on that would be the MIMO signal model where certain analysis of the model with respect to the system model will be explained. Next section would see the analysis of the simulation results about the ergodic capacity and 99% outage capacities of MIMO fading channels.

2 Sytem Modal and Signal Model

2.1 System Model

The way it has always been traditionally for wireless communications is the usage of a single antenna for transmission and a single antenna for reception. Such systems are known as single input single output (SISO) systems. As the times began to change, significant progress has been made in developing systems that employ multiple antennas at the transmitter and the receiver in order to achieve better performances. [4]

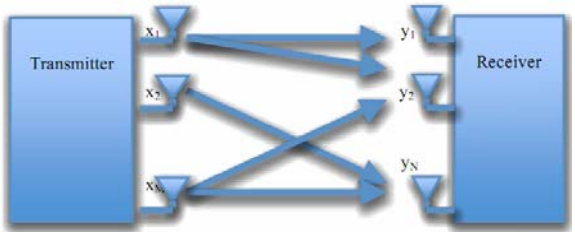


Fig. 1 Block diagram representing multiple transmitting antennas and multiple receiving antennas

Fig.1 depicts the system model of the MIMO systems. As seen, on the transmitter end we have M transmit antennas, represented by XM. On the receiver end, we have N receive antennas. Transmission from the transmitter's M antennas can go in any path, therefore forming multipath propagation before it is being received at the various receiving antennas. There are several benefits of using multiple antennas. Firstly, the link budget and spatial diversity improvements. Spatial diversity refers to the fact that the probability of having all antennas at bad locations is significantly lower as the number of antennas increases. Link budget improvement on the other hand, refers to the fact that the signals from the various antennas can be combined to form a signal stronger than any of the individual signals. For receive spatial diversity, signals received on multiple antennas are weighted and combined. Example of this is Maximum Ratio Combining (MRC), where the receiver co-phases the signals and sums them together, weighting each branch with a gain proportional to the amplitude on the received signal on that branch. Other diversity methods used at the receiver includes Selection Combining where we simply select the branch with the strongest SNR to be the output signal and Equal Gain Combining where the receiver co-phases the signals and sums them up. There are two types of transmit spatial diversity, open- loop and closed-loop. Open-loop transmit diversity involves transmitting signals from multiple antennas in some deterministic pattern, that does not depend on the channel. Open-loop techniques include cyclic delay diversity (CDD) and space-time block codes (STBC). Closed loop transmit diversity techniques, in contrast, require channel information to guide transmissions. An example is the Transmit Beam Forming (TxBF), where proper magnitude and phase weights computed from the channel estimation are applied again across antennas to aim the signal in a given desired direction. MIMO systems with spatial diversity achieve better performance. This basically means that it has a longer range for a given data rate, as compared to SISO systems at a given same location. The other advantage to exploit rich spatial dimensionality is via spatial multiplexing. This means transmitting and receiving multiple data streams from multiple antennas at the same time, and in the same frequency spectrum. This is possible because the signals received at different antennas are unique combinations of the transmitted data streams. Advanced digital signal processing algorithms can be used to recover the original data. Spatial multiplexing can be implemented in either open-loop or closed-loop. In open-loop spatial multiplexing, different streams are simply transmitted from different antennas. In closed-loop spatial multiplexing, every stream is transmitted from all of the antennas using weights computed from the channel estimation. MIMO systems with spatial multiplexing achieve higher peak data rates and increases spectrum efficiency.

2.2 Signal Model

In order to design efficient communication algorithms for MIMO systems and to understand the performance limits, it is important to understand the nature of MIMO channels. For a system with M transmit antennas and N receive antennas, assuming frequency-flat fading over the bandwidth of interest [5], we obtain the following:

$$\begin{bmatrix} y1 \\ \cdot \\ \cdot \\ \cdot \\ yN \end{bmatrix} = \begin{bmatrix} h1,1 & \cdots & h1,M \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ hN,1 & \cdots & hN,M \end{bmatrix} \begin{bmatrix} x1 \\ \cdot \\ \cdot \\ \cdot \\ xM \end{bmatrix} + \begin{bmatrix} n1 \\ \cdot \\ \cdot \\ \cdot \\ nN \end{bmatrix}$$

Fig. 2 Matrix Version

Or can also be denoted as:

$$y = Hx + w \quad (1)$$

We see that y denotes the N x 1 received signal vector and x is represented by the M× 1 transmitted signal vector. n is the N× 1 zero-mean complex Gaussian noise vector with independent, equal variance real and imaginary parts, thus giving us N (0, □^{2} IN) [6]. The matrix H is the N× M normalized channel matrix. Each element h(N, M) represents the complex gains between the Mth transmit and Nth receive antenna.

For a memory less 1x1 (SISO) system the capacity is given by:

$$C = \log_2\left(1 + \rho|h|\right) bits/\sec/Hz \quad (2)$$

Where h is the normalized complex gain of a fixed wireless channel or that of a particular realization of a random channel. In (2) and subsequently, ρ is the SNR at any receiver antenna. As we deploy more receiver antennas the statistics of capacity improve and with N receiver antennas we have a single-input multiple-output (SIMO) system with capacity given by:

$$C = \log_2\left(1+\rho\sum_{i=1}^{N}\left|h_i\right|^2\right) bits/\sec/Hz \qquad (3)$$

Where h*i* is the gain for these receiver antennas, *i*. Note the fundamental element of (3) is that increasing the value of N only results in a logarithmic increase in average capacity. Correspondingly, if we decide to opt for transmit diversity, in the common case where the transmitter does not have channel knowledge or state information, we have a MISO system with M transmit antennas and the capacity is given by:

$$C = \log_2\left(1+\frac{\rho}{M}\sum_{i=1}^{N}\left|h_i\right|^2\right) bits/\sec/Hz \qquad (4)$$

Where the normalization by M ensures a fixed total transmitter power and shows the absence of array gain in that case, as compared to the case in (3) where the channel energy can be combined coherently. Again, note that capacity has a logarithmic relationship with M. Now we consider the use of diversity at both transmitter and receiver giving rise to MIMO systems. For M transmitting antennas and N receiving antennas we have the now famous capacity equation:

$$C = \log_2\left(I_N+\frac{\rho}{M}HH^*\right) bits/\sec/Hz \qquad (5)$$

Where * means transpose-conjugate and H is the M x N channel matrix. Note that both (4) and (5) are based on M equal power (EP) uncorrelated sources.

Let $\Re$ denote the covariance matrix of x, then the capacity of the system described as based by (1) is given by

$$C = \log_2\left(I_N+H\Re H^*\right) bits/\sec/Hz \qquad (6)$$

Where tr($\Re$) $\leq \rho$ holds to provide a global power constraint. Note that for equal power uncorrelated sources, $\Re = \frac{\rho}{M} I_M$ and (6) becomes (5). This is optimal when H is unknown at the transmitter and the input distribution maximizing the mutual information is the Gaussian distribution. It is also demonstrated that the capacity in (5) grows linearly with min (M; N) rather than logarithmically. This result can be intuited as follows: the determinant operator yields a product of min (M;N) non-zero eigenvalues of its (channel-dependent) matrix argument, each eigenvalue characterizing the SNR over a so-called channel eigen-mode. An eigenmode corresponds to the transmission using a pair of right and left singular vectors of the channel matrix as transmit antenna and receive antenna weights respectively. Thanks to the properties of the log, the overall capacity is the sum of capacities of each of these modes, hence the effect of capacity multiplication. Clearly, this growth is dependent on properties

of the eigenvalues. If they decayed away rapidly then linear growth would not occur. However (for simple channels) the eigenvalues have a known limiting distribution and tend to be spaced out along the range of this distribution. Hence it is unlikely that most eigenvalues are very small and the linear growth is indeed achieved. With the capacity defined by (5) as a random variable, the issue arises as to how best to characterize it. Two simple summaries are commonly used: the mean or also known as the ergodic capacity given as

$$C = E\det\left[\log\left(I_N+\frac{\rho}{M}HH^*\right)\right] bits/\sec/Hz \qquad (7)$$

and also, the outage capacity, C□. Outage capacity measures are often denoted by capacity values supported (a), for example, 90% or 99% of the time, and indicating the system reliability. Therefore we have:

$$\text{Prob}(C\ (H) > C_\alpha) = \alpha \qquad (8)$$

A full description of the capacity would require the probability density function or equivalent. Some vigilance is required when interpreting the above equations. The important thing to consider is that our discussions concentrate on single user MIMO systems. Although so, many results can also apply to multi-user systems with receive diversity. [7]

Finally the linear capacity growth is only valid under certain channel conditions. It was originally derived for the independent and identically distributed flat Rayleigh fading channel and does not hold true for all cases. For instance, if large numbers of antennas are packed into small volumes then the gains in H may become highly correlated and the linear relationship will reach a state of little or no growth at all due to the effects of antenna correlation. On the other hand, other propagation effects not captured in (5) may provide a means to highlight the capacity gains of MIMO such as multipath delay spread. This can be shown in the case when the transmit channel is known but also in the case when it is unknown.

In general, the consequences of the channel modeling are critical. Environments can easily be chosen which give channels where the MIMO capacities do not increase linearly with the numbers of antennas. [1] However, most measurements and models available to date do give rise to channel capacities which are of the same order of magnitude as the promised theory. [7] Also the linear growth is usually a reasonable model for moderate numbers of antennas which are not extremely close-packed. Here the capacity is given by CEP in (5). This was derived and showed $\Re = \frac{\rho}{M} I_M$, that is, optimal for identical and independent Rayleigh fading channels. Also, this was derived starting from an equal power assumption.

3 Simulation Results and Analysis

In this section, we examine the ergodic capacities and 99% outage capacities of a MIMO Rayleigh fading channel with M transmit antennas and N receive antennas for SNR ranging from 0 dB to 40dB. We consider M=N=1; M=1, N=2; M=2, N=1 and M=N=2 respectively.

3.1 The ergodic Capacity comparison between the different number of M and N.

At first, we assume the channel information is known at transmitter (CSIT). So, it used the waterfilling solution. Then we can see the capacity is increased while the number of antennas is increased. (Fig. 3.) In the SISO case, where both M and N are equal to 1, the capacity at SNR equals to 5 dB is 1.8 bit/s/Hz while 3.6bit/s/Hz in 2× 2 MIMO case which is approximately twice as the SISO case. This feature keeps from low SNR to high SNR situation. Secondly, the capacity of asymmetrical antennas at transmitter and receiver is nearly the same at 2× 1 antennas case and 1× 2 antennas case.

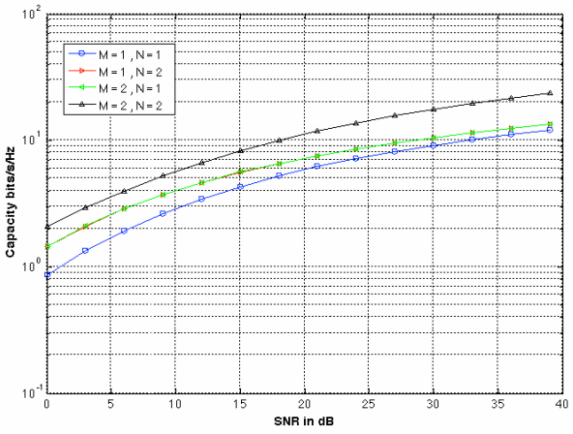


Fig. 3 Capacity comparison between the different number of M and N (Waterfilling allocation)

Then, we examined when the channel information is unknown at transmitter, so we used the equal power allocation. (Fig. 4.) In this case, the performance of symmetrical number of antennas is as same as the previous case, double the capacity when the antennas increased from 1× 1 to 2× 2. However, the capacity of 2× 1 model (MISO) is different from the 1× 2 (SIMO) one. The more receiver antennas achieve higher capacity compared to the more transmitter antennas' case. The performance is worse at receiver when the channel state information is unknown at transmitter, especially at low SNR. That is because at low SNR with Full CSI, waterfilling can provide a significant power gain.

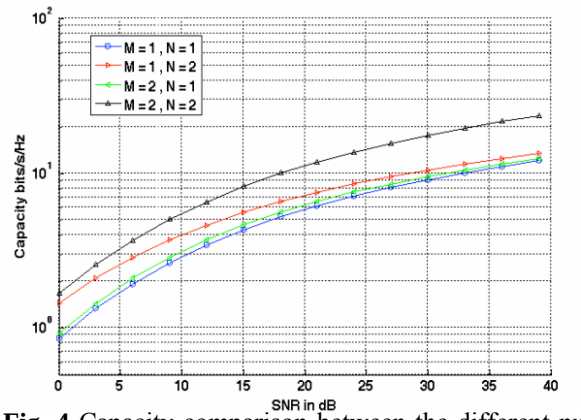


Fig. 4 Capacity comparison between the different number of M and N (Equal power allocation)

3.2 The 99% outage capacities.

The result is shown in Fig. 5.

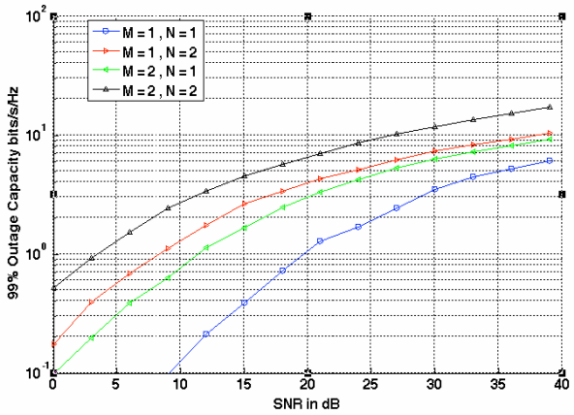


Fig. 5 99% outage capacity

The q% outage capacity Cout of a fading channel is the information rate that is guaranteed for (100-q)% of the channel realization. The 99% outage capacity is much better at 2× 2 (MIMO) than the case of SISO. That means in MIMO case, the communication is more reliable by the increased receive and transmit antennas. This is because transmitting and receiving data from different antennas can reduce the BER rapidly.

3.3 Ergodic capacity increases linearly in min(M,N) at high SNR.

Result is shown in Fig. 6.

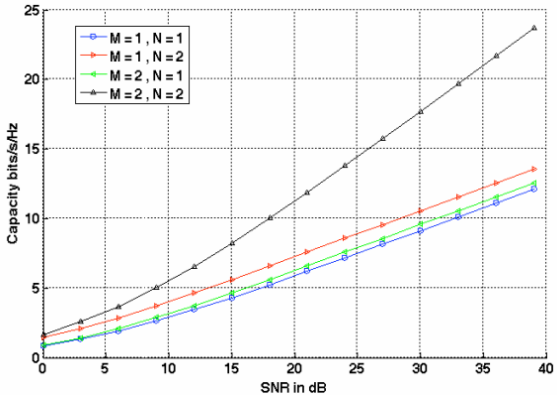


Fig. 6 The ergodic capacity between different M and N

This figure is in the case the channel state information is unknown at the transmitter. The ergodic capacity:

$$C = \varepsilon_H \left[\log\det\left(I_N + \frac{1}{\sigma^2}\frac{P}{M} I_M HH^H\right)\right] = \varepsilon_H \left[\sum_{i=1}^{m} \log\left(1 + \frac{P}{M}\lambda_i\right)\right], \quad (9)$$

where m=min (M,N) and $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of the Wishart matrix

$$W = \begin{cases} HH^H & N < M \\ H^H H & N \geq M \end{cases} \quad (10)$$

For large SNR, $C_{EP} = \min(M, N)\log P + \mathcal{N}$, the capacity grows linearly with min(M,N).

4. Conclusion

As a conclusion, this paper has covered the major reviews of MIMO systems and its potential use for further enhancement in future wireless networks. Information theory reveals that great capacity gains can be realized from employing MIMO systems. Whether we achieve this fully or at least partially in practice, it can be seen that this depends on a sensible design of transmit and receive signal processing algorithms. It is obvious that the success of MIMO algorithm has integrated into various commercial standards such as 3G, WLAN and beyond and this clearly will rely on a fine compromise between the maximization of data rates and diversity, example space time coding, solutions. This also includes the ability to adapt to the time changing nature of the wireless channel using some form of feedback. To this end more progress in modeling, not only the MIMO channel but its specific dynamics, will be required. As new and more specific channel models are being proposed it will useful to see how those can affect the performance trade-offs between existing transmissions algorithms and whether new algorithms, tailored to specific models, can be developed. Finally, upcoming trials and performance measurements in specific deployment conditions will be a key to evaluate precisely the overall benefits of MIMO systems in real-world wireless scenarios.

5. References

[1] Rafal Zubala, Herbert Kokoszkiewicz, B.W.M. Kuipers, L.M. Correia, "A Simple Approach to MIMO Channel Modelling"(Onlinepublishing)http://www.eurasip.org/Procs ipco/Eusipco2006/p apers/1568981957.pdf

[2] Gesbert, D., Shafi, M., Da-shan Shiu, Smith, P.J., Naguib, A. " From theory to practice: an overview of MIMO space-time coded wireless systems" Selected Areas in Communications, IEEE Journal, Volume 21, Issue 3, April 2003 Page(s):281 – 302

[3] MIMO – Wikipedia, the free encyclopedia (OnlineReference) http://en.wikipedia.org/wiki/Multiple input_multiple-output_communications

[4] J.M. Gilbert, Won-Joon Choi, Q.Sun "MIMO Technology for Advanced Wireless Local Area Networks", Atheros Communications, Inc. pp. 413-415.

[5] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, H. Vincent Poor "MIMO Wireless Communications" Hardback, Cambridge University Press, 2007.

[6] Shuo Pan, Salman Durrani, M.E. Bialkowski, "MIMO Capacity for Spatial Channel Model Scenarios",(Online Reference)

[7] Taesang Yoo, Andrea Goldsmith, "Capacity of fading MIMO channels with channel estimation error", Dept. of Electrical Engineering, Stanford University.

[8] H. Ramezani. (2006, Oct.). MIMO Rayleigh fading Channel Capacity. Mathworks. [Online]. Available:
[9] http://www.mathworks.com/matlabcentral/fileexchange/12491

An Efficient Security Aware Route Optimization Technique In Mobile IPV6 Communication Networks

Md. Junayed Islam[1], **Md. Waliullah**[1] **and Md. Moktadir Rahman**[2]
[1]School of Computing & Mathematical Sciences, University of Greenwich, London, United Kingdom
[2]School of Engineering & Mathematical Sciences, City University, London, United Kingdom

Abstract – *Privacy and Route optimization are important issues for reliable and efficient Mobile IPV6 communication networks. An efficient security based route optimization technique through a pseudo-tunnelling module has been proposed in this paper. Simulation results in OMNET++ show better performance than any other approaches having dependency on less number of cryptographic messages with absolute validation of nodes. Also higher throughputs, reduced data traffic loads, less WAIT time of data packets in home agent and congestion controlled Round Trip Time have proved the overall performance improvement of the entire communication network.*

Keywords: Mobile IPV6, Home Agent, Foreign Agent, Type2 Routing Header, Binding Update, Return Routability Procedure, UMU-PKIV6.

1 Introduction

In Mobile IPV6, the packets which will be sent from the mobile node towards their correspondent node must go through the home agent prior to being delivered. Such interception of packets by the home agent is known as dogleg routing that causes a non-optimised longer paths and an unexpected higher communication delays between mobile nodes and their correspondents [1].

Mobile IPv6 base specification includes a route optimization scheme called the Return Routability Procedure (RRP) which allows packets to be directly sent between a mobile node and its correspondent without involving the home agent. When route optimization is performed the mobile node's data traffic is not protected by IPsec, which leaves the communications vulnerable to eavesdropping on the visited network.

2 Proposed Return Routability Procedure

2.1 Validation of MN and HA

Definition of a new data structure named *HomeNetworkInfo* has been introduced in the *IPv6InterfaceEntry* [2], which is liable to contain all information of home network of the mobile node's [MN] interface.

2.2 Validation of CN and HA

The CN is verified by HA through the PKIV6 system during exchanging their neighbour solicitation messages and the HA generates a *Care-of keygen* token as below which is sent in CoT message towards CN, encrypted by the public key of CN.
Care-of keygen token:= First (64, HMAC_SHA1 (care-of-address | nonce |))

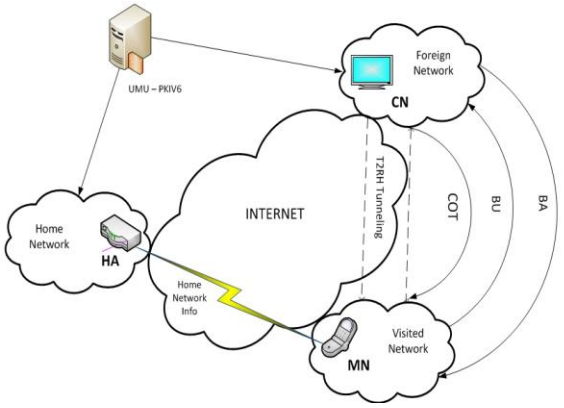


Figure 1. Proposed RRP

2.3 BU Authentication

The MN now generates a BU and sends to CN [Fig 1]. The content of BU message includes a *sequence number, nonce, CoA* and a *MAC of (Care-of keygen token, (CoA | Address of CN | BU))*.

2.4 Route Optimized Secured Tunneling

A *Type 2 Routing Header (T2RH)* based tunneling module has been developed by using OMNET++ [2] as part of this research. Once the Return Routability procedure has been completed as well as the mobile (MN) has performed the registration with correspondent node; subsequent communication can be carried out between mobile node (MN) and correspondent node (CN) by bypassing home agent (HA).

The HoTI, HoT and CoTI messages are eliminated in this proposed *Return Routability (RR) procedure*. The CN does not have to generate Home keygen token and Care-of keygen token as is required in Traditional Return Routability procedure. A CN, which can be a mobile device with low processing power, is saved of computation load.

3 Simulation Results & Evaluation

Table 1 shows less number of messages are used by Proposed Return Routability Procedure (*RRP)* comparatively with other approaches.

Table 1. Comparative Messages used in Security Solutions

Security Solution	**Messages**	**Dependency on PKI**
Traditional RRP	8	No
Improved Bombing Resistant Protocol	6	No
Enhanced cga based RO	7	Yes
Proposed RRP	3	Yes

In Figure 2, it shows number of data bytes sent and received over time by foreign network's access point are less than the home agent's one having comparatively less difference between them. This simulation here clearly shows that the the proposed mechanism of mipv6 routing mechanism reduces about **49.50%** of traffic from home agent by directing the packets directly from mobile node to correspondent node.

On the other hand, reduction of traffic in home agent reduces the probability of collision as well. So the simulation results in Figure 2 show increased throughput of the home agent which makes sure the traversal of data packets through a less congested path that leads to comparatively **30.76%** reduced packet dropping probability.

Packet dropping probability thus can be calculated as -

$$P(S) = \frac{L^s / S!}{\sum_{i=0}^{S} L^i / i!}$$

The heterogeneous communication such as network setup [Figure 1] in this research having mixture of such wired or wireless links solves several challenging problems in modeling and control like time varying traffic load in home agent, end-to-end congestion controlled round trip time (RTT) [Figure 3] and variety of application demands etc.

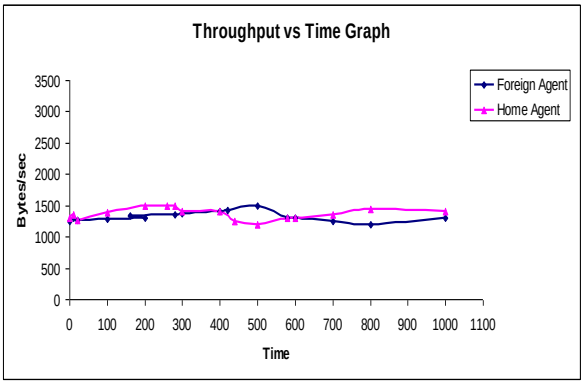


Figure 2. Throughput (bytes/s) vs Time (sec) Graph

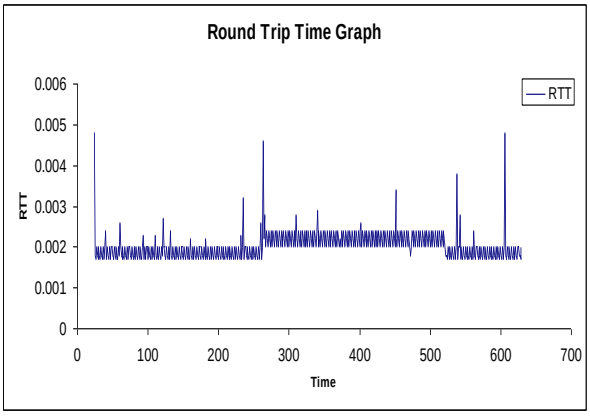


Figure 3. Round Trip Time (sec) Graph

4 Conclusion & Future Work

This paper has introduced several engineering directions that would improve the limitations of mobile ipv6 protocol. Simulation results show performance improvements by reducing traffic load from home agent, solves the dogleg routing problem in mipv6, increased throughput of the home agent, makes sure the traversal of data packets through a less congested path, reduced total *WAIT time* in queue for home agent.

As a future work this research work can be extended by introducing Multiple Care of Address (mCOA) Registration technique as well as an effective path selection method in mCOA for MIPv6.

5 References

[1] T. Narten, E. Nordmark, W. Simpson, and H. Soliman. Neighbor Discovery for IP version 6 (IPv6). RFC 4861 (Draft Standard), September 2007.

[2] Faqir Zarrar Yousaf, An Accurate and Extensible Mobile IPv6 (xMIPV6) Simulation Model for OMNeT++, Communication Networks Institute, Dortmund University of Technology.

SESSION

PERFORMANCE ANALYSIS, EVALUATION, AND MONITORING

Chair(s)

TBA

Performance Analysis of String Topology Underwater Acoustic Networks

Zaihan Jiang[1], Mengchu Zhou[2], and Lichuan Liu[3]
[1]Acoustic Division, US Naval Research Laboratory, Washington, DC, 20375, USA
[2]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA. Email: zhou@njit.edu
[3]Department of Electrical Engineering, Northern Illinois University, DeKalb IL, 60115, USA. Email: liu@niu.edu

Abstract - *The design of network topologies and protocols is critical to improving underwater acoustic network performance. A string topology is one essential component of underwater networks. This work analyzes its performance in an underwater environment. Theoretical results show that the number of nodes, traffic intensity, packet size, and probability of collision affect the end to end network throughput significantly while physical channel capacity fundamentally restricts it. This research gives one a theoretical direction to design future topology and network protocols for underwater acoustic networks. This work is supported by the Office of Naval Research.*

Keywords: Underwater Acoustic Network (UAN), end-to-end throughput, string topology, traffic density, channel capacity

1 Introduction

A string topology network has been widely applied to an underwater environment, e.g., in sea web experiments, a string topology has been deployed [1][2]. As shown in Fig. 1, in a string topology, nodes are sending message to one destination, usually a gateway node, with a few relay nodes in between to re-deliver messages.

An underwater environment is significantly different from its territorial counterpart [3][4][5][6][7][8][9]. From a network's view point, one has to consider the following essential issues in such environment:

1) Long propagation delay, due to the slow acoustic signal propagation speed in water medium, which affects link/network protocol design;
2) Very limited bandwidth;
3) Very limited energy;
4) Unstable link condition;
5) Transmission loss is frequency/environment dependent.

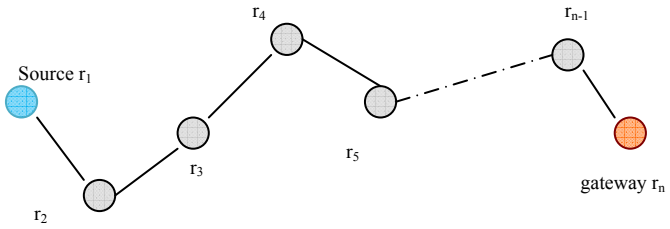


Fig. 1 A string topology underwater network

A string topology is the basic component in an underwater acoustic network. However, its performance is not well known in an underwater environment [10]. Why the end-to-end throughput is so small in underwater? Which factors fundamentally restrict the end-to-end throughput? What kind of protocols fit in it? How can the design of network topology be optimized? This work concentrates on the study of aforementioned questions and gives insight and answers to these important questions. The rest of the paper is organized as follows. Section 2 analyzes the end-to-end throughput theoretically and gives a close-form solution in both ideal and realistic cases; Section 3 gives insight analysis of the factors contributing to the throughput. Finally, conclusion is drawn and future research direction is given in Section 4.

2 Theoretical analysis

In this section, the end-to-end throughput is first investigated in an ideal case and then a realistic case given an underwater acoustic network of a string topology.

2.1 Ideal case

To simplify the analysis, we make the following assumptions of the topology:

1) There is only one sender who generates data packets; other nodes are relay nodes;
2) Neighboring nodes are equal distanced at *d*;

3) Nodes' transmission power is the same.

We also assume:

- The Medium Access Control (MAC) protocol is contention-based;
- The sender generates packets with the same size L;
- The packet generation at the source node follows a Poisson process with rate λ;
- Each node has unlimited buffer to save packets such that none will be dropped; and,
- The link is stable, i.e., only collision may cause the loss of a packet but not because of any other reasons. Each hop has the same link condition.

As shown in Fig. 1, let r_1 = source, r_n = Gateway node, and r_i $(1 < i < n)$ is an intermediate node. A packet is only generated from source and transmitted from r_i to $r_{i+}1$ where $1 \le i < n$. The successful transmission of a frame at r_i is expressed as:

$$P_i = P_r\{successful\ \ reception\ \ at\ \ r_{i+1} \mid frame\ \ transmitted\ \ by\ \ r_i\} \quad (1)$$

A packet is transmitted successfully only if it is received properly at the next node. T is the transmission time of a packet and can be expressed as:

$$T = \frac{L}{TR} \quad (2)$$

in which TR is the transmission rate. The probability that no traffic is generated by node r_{i+1} during a frame's reception vulnerability period, which is $2T$, is:

$$\frac{e^{-(2T)(\lambda_{i+1})}(2T\lambda_{i+1})^0}{0!} = e^{-2T\lambda_{i+1}} \quad (3)$$

Where $\lambda_0 = \lambda$ is packet generation rate of node r_0 while λ_i is the rate of node r_i, $1 \le i < n$. We also have:

$$\begin{aligned} \lambda_1 &= \lambda \\ \lambda_2 &= \lambda p_1 \\ \lambda_3 &= \lambda p_1 p_2 \\ &.... \\ \lambda_{n-1} &= \lambda(p_1 p_2 \cdots p_{n-2}) \\ \lambda_n &= \lambda(p_1 p_2 \cdots p_{n-1}) \end{aligned} \quad (4)$$

and:

$$p_i = e^{-2T(\lambda_i + \lambda_{i+1} + \lambda_{i+2})} \quad (5)$$

Combing the above equations together, we obtain n nonlinear equations with respect to n variables $\lambda_1, \lambda_2, \ldots,$ and λ_n. We have:

$$\begin{aligned} P_1 &= e^{-2\lambda T(1+p_1+p_1p_2)} \\ P_2 &= e^{-2\lambda T(p_1+p_1p_2+p_1p_2p_3)} \\ P_3 &= e^{-2\lambda T(p_1p_2+p_1p_2p_3+p_1p_2p_3p_4)} \\ P_4 &= e^{-2\lambda T(p_1p_2p_3+p_1p_2p_3p_4+p_1p_2p_3p_4p_5)} \\ &\vdots \\ P_{n-2} &= e^{-2\lambda T(p_1p_2\cdots p_{n-3}+p_1p_2\cdots p_{n-2}+p_1p_2\cdots p_{n-1})} \\ P_{n-1} &= e^{-2\lambda T p_1 p_2 \cdots p_{n-2}} \\ p_n &= 1 \end{aligned} \quad (6)$$

$P_n = 1$ due to that gateway node r_i uses RF signal such that acoustic signal from other nodes does not collision with its signal. Since $0 \le p_i \le 1$, we get:

$$\begin{aligned} \ln p_1 &= -2\lambda T(1 + p_1 + p_1p_2) \\ \ln p_2 &= -2\lambda T p_1(1 + p_2 + p_2p_3) \\ \ln p_3 &= -2\lambda T p_1p_2(1 + p_3 + p_3p_4) \\ &\vdots \\ \ln p_{n-2} &= -2\lambda T p_1p_2\ldots p_{n-2}(1 + p_{n-2} + p_{n-2}p_{n-1}) \\ \ln p_{n-1} &= -2\lambda T p_1p_2\ldots p_{n-2} \\ p_n &= 1 \end{aligned} \quad (7)$$

They can be expressed as:

$$p_2 = \frac{\frac{\ln(p_1 - 1 + 1)}{-2\lambda T} - 1}{p_1} - 1$$

$$p_i = \frac{\frac{\ln(p_{i-1} - 1 + 1)}{-2\lambda T p_1 p_2 \cdots p_{i-2}} - 1}{p_{i-1}} - 1, \qquad i = 3, \ldots n-1 \quad (8)$$

Subject to: $0 \le p_i \le 1$ (9)

From (9), we have:

$$\begin{aligned} p_1 + 1 &\le \frac{\ln p_1}{-2\lambda T} \le 2p_1 + 1 \\ (p_2 + 1)p_1 &\le \frac{\ln p_2}{-2\lambda T} \le (2p_2 + 1)p_1 \\ (p_3 + 1)p_1p_2 &\le \frac{\ln p_3}{-2\lambda T} \le (2p_3 + 1)p_1p_2 \\ (p_4 + 1)p_1p_2p_3 &\le \frac{\ln p_4}{-2\lambda T} \le (2p_4 + 1)p_1p_2p_3 \\ &\vdots \\ (p_{n-2} + 1)p_1p_2\cdots p_{n-2} &\le \frac{\ln p_{n-2}}{-2\lambda T} \le (2p_{n-2} + 1)p_1p_2\cdots p_{n-2} \end{aligned} \quad (10)$$

The effective throughput of the network is expressed as:

$$\eta = \lambda_{n-1} p_{n-1} L(1 - \partial) \quad (11)$$

Where α is the bit error rate. Hence, we have:

$$\eta = \lambda(p_1 p_2 \ldots p_{n-2}) p_{n-1} L(1-\partial) \tag{12}$$

Finally, by combining (8) ~ (12) together, we have:

$$\eta = \lambda(p_1(\frac{\frac{(p_1-1)[1-\frac{1}{6}(p_1-1)^2]}{-2\lambda T}}{p_1}-1)(\frac{\frac{(p_2-1)[1-\frac{1}{6}(p_2-1)^2]}{-2\lambda T p_1}}{p_2}-1)\ldots$$
$$(\frac{\frac{(p_{n-2}-1)[1-\frac{1}{6}(p_{n-2}-1)^2]}{-2\lambda T p_1 p_2 \ldots p_{n-2}}}{p_{n-1}}-1)L(1-\partial) \tag{13}$$

he title approximately 2.5 centimeters (1 inch) from the top of the first page and use 20 points type-font size in bold. Center the title (horizontally) on the page. Leave approximately 1 centimeter (0.4- inches) between the title and the name and address of yourself (and of your co-authors, if any.) Type name(s) and address(s) in 11 points and center them (horizontally) on the page. Note that authors are advised not to include their email addresses.

2.2 Realistic case

Due to the significant propagation delay, because of the slow propagation speed of acoustic signal in water medium, and system delay, which can also be un-ignorable [11], the real transmission time of a packet (we use $\tilde{T}$ to represent it) is much bigger in a real scenario than that in the ideal one. $\tilde{T}$ can be expressed as

$$\tilde{T} = T + \frac{d}{c_w} + S_d \tag{14}$$

where S_d is the system delay and c_w is the sound speed in water medium. The bigger transmission delay causes two effects: First, the probabilities in (8) become higher, which help the end-to-end throughput positively; second, the effective transmission rate is reduced to:

$$\tilde{TR} = TR \times \frac{T}{\tilde{T}} \tag{15}$$

which impacts the throughput negatively. Correspondently, the throughput can be expressed as:

$$\tilde{\eta} = \lambda(p_1(\frac{\frac{(p_1-1)[1-\frac{1}{6}(p_1-1)^2]}{-2\lambda \tilde{T}}}{p_1}-1)(\frac{\frac{(p_2-1)[1-\frac{1}{6}(p_2-1)^2]}{-2\lambda \tilde{T} p_1}}{p_2}-1)\ldots$$
$$(\frac{\frac{(p_{n-2}-1)[1-\frac{1}{6}(p_{n-2}-1)^2]}{-2\lambda \tilde{T} p_1 p_2 \ldots p_{n-2}}}{p_{n-1}}-1)L(1-\partial)\frac{T}{\tilde{T}} \tag{16}$$

Equation (16) demonstrates that the relay node placement, underwater sound speed and system delay all contribute the variation of the end-to-end throughput. Due to the scope of this paper, we will address this issue in detail in the future with open sea experiment data.

3 Performance analysis

Equation (13) and (16) shows the insight information about relationship between the end-to-end throughput and physical channel capacity, network topology design, and traffic load:

- The throughput is proportional to the transmission rate, and proportional to the channel bit error rate (BER) reversely. The higher the BER, the lower the throughput as illustrated in Figs. 2-3. In other words, the physical channel capacity restricts the network throughput fundamentally.
- The throughput is directly affected by the traffic generation rate from two aspects: the increase of traffic rate brings more data into the link; meanwhile, the possibility of collision among neighboring nodes increases, which contributes to the decrease of network throughput. In addition, the packet size impacts the throughput similarly. The results are shown in Fig. 4-5.
- As descripted in Fig. 6, the throughput is affected by the number of nodes in the string indirectly. The traffic stream becomes thinner with the addition of a node, since the probability of no collision in the node is usually less than 1, due to the inevitable collision that occurs within a node.
- (10) reveals that the probability of a successful transmission in the 1st node affect successful reception of following relay nodes; the one in the upper stream nodes of a string affects the one of lower stream nodes. Eventually, the end-to-end throughput is impacted.
- (16) shows that the distance of neighboring nodes, sound speed and system delay all impact the end-to-end throughput.

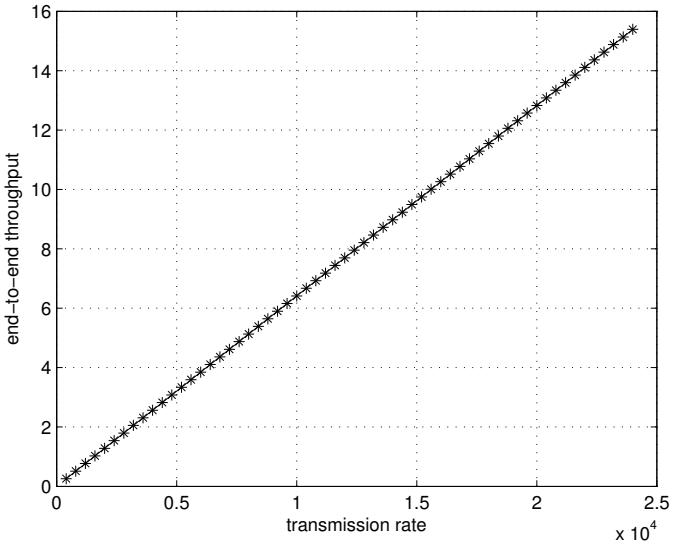


Fig. 2 Throughput vs. transmission rate (number of nodes = 6)

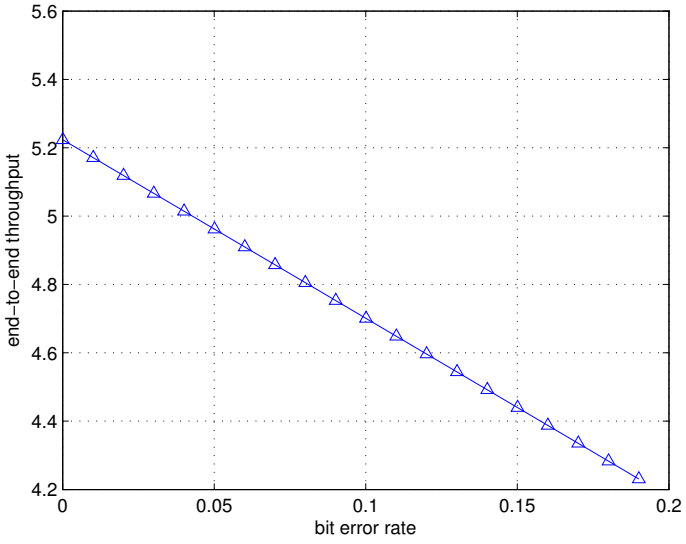


Fig. 3 Throughput vs. channel BER (number of nodes = 6)

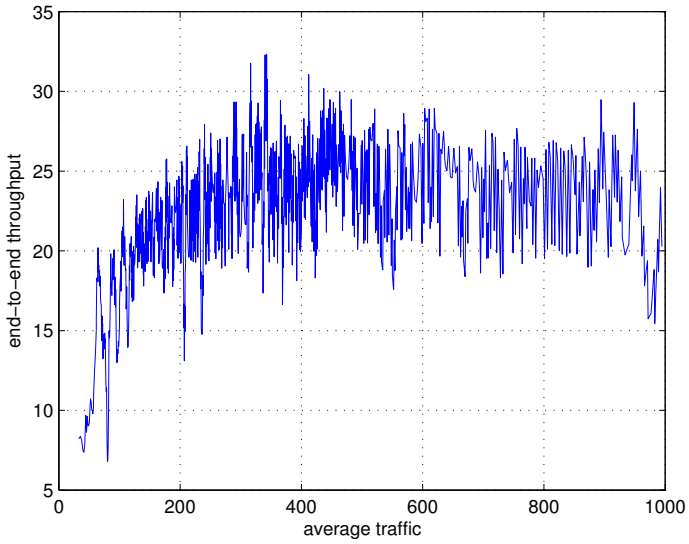


Fig. 4 Throughput vs. traffic load (number of nodes = 4)

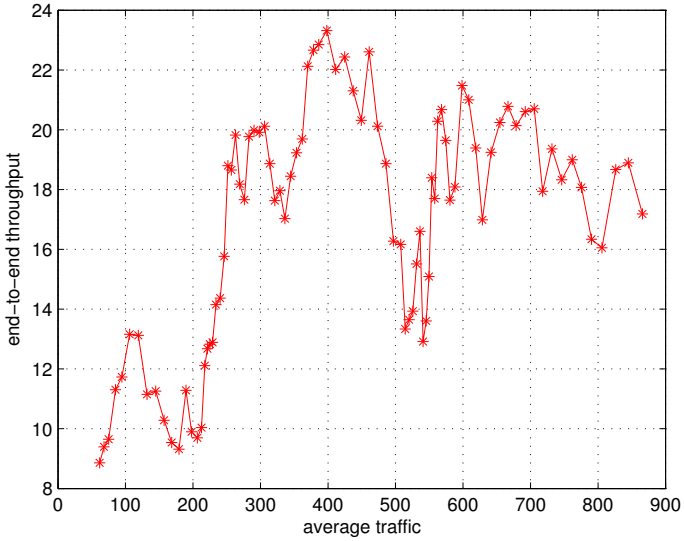


Fig. 5 Throughput vs. traffic load (number of nodes = 5)

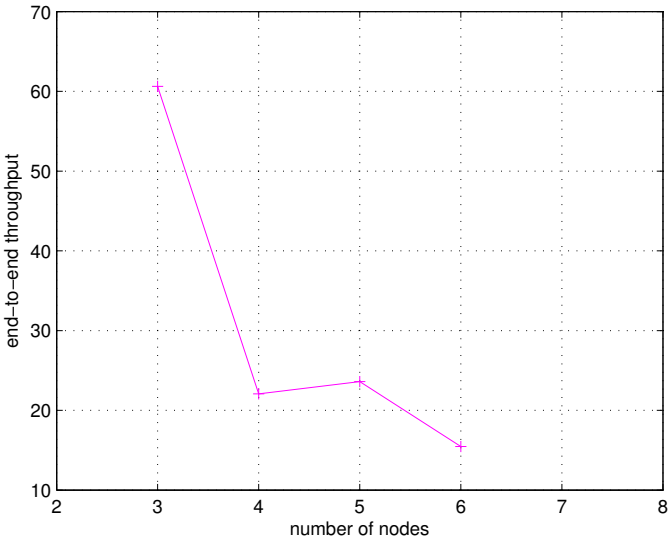


Fig. 6 The throughput vs. number of nodes in the string

4 Conclusions and future research

A string topology is one of the essential topologies in an underwater acoustic network. In this paper, we assume a contention-based MAC protocol and analyze the end-to-end network throughput of a string topology underwater acoustic network. Close-form formulae are given for both ideal and realistic cases. It is shown that the network end-to-end throughput is affected by traffic generation rate, packet size, network topology design (number of nodes, etc.), and probability of collision within a node; as well as physical channel capacity restricts the throughput fundamentally. We also find that the relay node placement, sound speed in water medium, and system delay also impact end-to-end throughput if the delay caused by propagation and system transmission is not ignorable.

In the analysis, we assume a contention-based MAC protocol and get poor end-to-end throughput. To achieve higher throughput, a non-contention-based MAC protocol is definitely needed in underwater environment. In the future, we will optimize the topology design and investigate a contention-avoided MAC for underwater acoustic networks. We also further investigate the network performance in realistic case and plan to compare that with real experiments' result in open sea environment.

5 References

[1] J. Rice, "SeaWeb Acoustic Communication and Navigation Networks", *Proceedings of the International Conference "Underwater Acoustic Measurements: Technologies & Results" Heraklion, Crete, Greece, 28th June – 1st July 2005.*

[2] J. Rice, "Undersea Networked Acoustic Communication and Navigation for Autonomous Min-Countermeasure Systems", *5th International Symposium on Technology and the Mine Problem, April, 2002.*

[3] I. F. Akyildiz, D. Pompili, T. Melodia, "Underwater acoustic sensor networks: research challenges", *Ad Hoc Networks (Elsevier), vol. 3, no. 3, pp. 257-279, March 2005.*

[4] H. Schmidt, "Autonomous underwater vehicle networks as integrated acoustic observation systems", *Acoustical Society of America Journal, Vol. 117, Iss. 4, pp. 2409 – 2410, April, 2005.*

[5] L. Berkhovskikh and Y. Lysanov, "Fundamentals of Ocean Acoustics". New York: Springer, 1982.

[6] A. Quazi andW. Konrad, "Underwater acoustic communications," *IEEE Commun. Mag., pp. 24–29, Mar. 1982.*

[7] R. Coates, "Underwater Acoustic Systems", New York: Wiley, 1989.

[8] J.A. Catipovic, "Performance Limitations in underwater acoustic telemetry", *IEEE J. Oceanic Eng., 15(3):205-216, Jul. 1990.*

[9] Z. Jiang, "Underwater Acoustic Networks – Issues and Solutions", *International Journal of Intelligent Control and Systems, Vol. 13, No. 3, Sept. 2008.*

[9] J. Gibson, G. Xie, Y. Xiao and H. Chen, "Analyzing the Performance of Multi-hop Underwater Acoustic Sensor Networks", In *Proc. MTS/IEEE Oceans 2007 Conference, Scotland, June 2007.*

[10] Y. Zhu, Z. Jiang, Z. Peng, J. Cui and M. Zuba, "Toward Practical MAC Design for Underwater Acoustic Networks", Submitted to *MASS'12.*

Spectrum Management in Cognitive Radio Networks: Modeling and Performance Evaluation

Md. Akbar Hossain, Nurul I Sarkar
School of Computing and Mathematical Sciences
Auckland University of Technology (AUT), Auckland, New Zealand

Abstract— *Nowadays, 'Cognitive Radio' (CR) is one of the most promising concepts which facilitate the flexible usage of radio spectrum and enhance the spectrum utilization by enabling unlicensed users to exploit the spectrum in opportunistic manner. However, the most important challenge is to share the licensed spectrum without interfering with transmission of other licensed users. Therefore, to alleviate the above problems a proactive spectrum management schemes has been designed to access the unoccupied spectrum opportunistically with minimum latency. The primary function of the management schemes is to characterize the available channels based on spectrum sensing of CR node and create a backup channel list for further use. In this paper, we model the licensed users' activity and create a scheme to build up available channel list and backup channel list. Our simulation results show that, around 65% enhancement of channel utilization can be achieved through channel management.*

Keywords: Cognitive Radio, Channel prediction, Spectrum Sensing, Spectrum Management, spectrum Mobility

1. Introduction

T HE increasing demand for new wireless services and applications, as well as the increasing demand for higher capacity wireless networks, the wireless networks become highly heterogeneous, with mobile devices consisting of multiple radio interfaces. In this context, it is essential to have updated information on radio environment to enhance the overall network performance. The outcomes of several investigations have shown that the lack of spectrum is not an issue, but the fact that radio resources are used inefficiently. Therefore, a promising functionality is required to be built into future terminals to have the cognitive capability to assist with the Dynamic Spectrum Allocation (DSA), which allows more efficient utilization of radio resources by changing the spectrum allocation on demand. A cognitive radio is a self-aware communication system that efficiently uses spectrum in an intelligent way [1]. The most significant characteristic of a cognitive radio is the capability to sense surrounding radio environment such as information about transmission frequency, bandwidth, power, modulation, etc. and make a decision to adapt the parameters for maintaining the quality of service. It autonomously coordinates the usage of spectrum in identifying unused radio spectrum on the basis of observing spectrum usage. Therefore, spectrum handoff occurs when a licensed user further utilizes this unused radio spectrum and find that CR nodes occupy the channel [2]. In order to avoid service termination, the CR user will perform link maintenance procedure to reconstruct the communication. In general, channel management procedure can be categorized into a) proactive spectrum management b) reactive spectrum management. In proactive scheme, CR nodes observe all channels to obtain the channel usage statistics, and generate a list of candidate and backup channel list for spectrum mobility while maintaining the current transmission [3]. Reactive spectrum management operates in on-demand manner, i.e. CR nodes perform spectrum mobility after detecting the link failure [4]. From system design, point of view reactive spectrum management is more suitable than the proactive scheme as it requires very complex algorithm for concurrent operation. On the other hand, proactive spectrum management poses very faster spectrum switching with respect to reactive spectrum scheme, resulting better QoS in on-going transmission. Moreover, selection of reactive and proactive spectrum management schemes is depends on sensing time. In order to capture the dynamic and random behavior of both licensed and unlicensed users, we focus on proactive spectrum management issues to enhance the network wide performance in terms of throughput and collision. Therefore, we proposed proactive dynamic channel selection algorithms to deal with spectrum mobility more robustly and effectively based on proactive channel prediction. Based on channel prediction, al the observed channel is graded and form two different channel lists. In this paper, we also present a CR node modeling in OPNET which utilize the channel prediction and spectrum management functionalities.

The rest of this paper is organized as follows. An introductory idea of the work is presented in details in section one. Then, section II describe the background study of spectrum mobility in cognitive radio networks. In section III spectrum mobility is being characterized through primary user traffic prediction. Proactive spectrum management scheme is introduced based on the channel prediction in section IV. Finally, we present our performance study and node design in section V which is followed by conclusion and future works.

2. Literature Review

In cognitive radio ad-hoc networks spectrum mobility is one of the main performance bottlenecks which include transmission delay, routing discovery as a consequence throughput degradation. This problem is somehow related to multichannel MAC problem in traditional mobile ad hoc networks despite of fixed channel assignment. Therefore, we need a novel spectrum management schemes where CR users will pause the current transmission and vacate the operating channel due to presence of license user as well as determine the available channel to re-establish the communication. In this regard author in [5] proposed two different kind of observation method called proactive method and on-demand or reactive method. In the proactive method, the CR user periodically observes all the channel usage statistics and determines the candidate set of channels for spectrum handoff. In contrast to proactive method, the candidate channels are searched with an on-demand manner in reactive method. Therefore, the instantaneous outcomes from wideband sensing will be used to determine the candidate channel list for spectrum handoff [6]-[4]. Such sense and react approach causes for frequent service disruption in communication and degrade the QoS due to higher handoff latency. While the proactive approach, the latency of spectrum handoff would be smaller even though it incurs a larger overhead due to periodic observation. In [7] a detailed proactive spectrum framework has been proposed assuming exponential and periodic traffic model where CR users utilize past channel histories to make prediction on future spectrum availability. In [8] author proposed MAC layer proactive sensing schemes to maximize the probability of channel opportunities and minimize the channel switching delay for spectrum mobility. The problem of spectrum mobility in cognitive radio network has been widely investigated in the last few years. L. Giupponi, in [9] proposed a fuzzy-based spectrum handoff to deal with the incompleteness, uncertainty and heterogeneity of a cognitive radio scenario and spectrum quality grading scheme is presented in [10] to enhance the QoS. In [11], Chang and Liu proposed a strategy that optimally determines which channel to probe and when to transmit in a single channel transmission. A sensing sequence is proposed in [5] to maximize the chances of finding and idle channel but it does not guarantee the minimum discovery delay. To minimize the delay, in [12]authors proposed a Bayesian learning method which could predict the underutilized radio spectrum proactively. Further enhancement is shown in [13] through the concept of building a backup and candidate channel list, unfortunately no algorithm or schemes has been provided. In most of the literature spectrum mobility is mainly concern to licensed band although CR nodes have the capability to use any portion of the spectrum not only from licensed band but also unlicensed band. Therefore authors in [14] first come up with an idea to build the backup channel list from unlicensed band in static manner and proposed a Markov channel model to evaluate the scheme. It is observed that, in these studies, the dropping probability and the number of handoff is reduced in case of the appearance of primary users. So far we consider about the spectrum sensing methods, but how we could share this sensing information among different CR nodes is a network architectural issue. Hence from the network architectural perspective, different spectrum mobility management schemes are presented in [15] and illustrate the basic comparison of centralized and distributed methods. In case of centralized architecture, spectrum mobility management is maintained by CR base station whereas in distributed ad-hoc networks each particular node is responsible to carry out the same task. Therefore, a reliable common control channel is required to exchange channel information is common control channel (CCC) or rendezvous problem. So far in the literature a reasonable amount of work has been done on CCC problem based on either dedicated global control channel [16], [17],[18],[19],[20],[21],[22],[23],[24] or network wide synchronization channel hopping sequence [25],[26],[27],[28]. However due to the dynamic nature of licensed users dedicated CCC is impractical and it suffers from CCC saturation and single point failure problem in high dense network. Like other distributed networks, network wide synchronization is not a feasible assumption for large distributed networks. In order to fully utilize the scare radio spectrum, several dynamic spectrum sharing schemes have been extensively studied [29],[30],[31],[32] from game theoretic view point for flexible and fair spectrum usages through analyzing the intelligent behaviors of network users.

The preliminary literature review reveals that most researchers have focused on spectrum selection processes in a static manner which failed to capture the dynamic behavior of radio environments. A study of spectrum mobility under dynamic radio environment is required to assist efficient design and deployment of such networks. In our proposal, proactive licensed users' traffic predictive model is used to predict the best available spectrum and classify them according to expected channel duration and SNR which will further use to build the candidate and backup channel list.

3. Problem Formulation

In this paper, a single hop cognitive radio based ad-hoc wireless LAN is considered which is composed of several primary users and CR users as shown in figure 1. Moreover a cognitive radio is assumed to search N licensed channel for spectrum opportunities. Each CR user is assumed to have equipped with two transceiver, one is for data communication and the additional one is for acquiring the control information including sensing. Although having an additional antenna may increase the size and price of CR node, eventually it overcome the problem of common control channel. A channel is modeled as widely used renewal

process alternating between ON and OFF states. An ON (OFF) period can be considered as a time period in which licensed users are present i.e. as a binary time series. Hence, one of the main tasks is to model the ON (OFF) period so that CR users can utilize any portion of OFF periods for their transmission. The channel usages model is depicts in figure 2. Suppose *i* is the channel index and X_t^i denote the number of channel *i* at time *t* such that:

$$X_t^i = \begin{cases} 1 & \text{if channel is ON (BUSY),} \\ 0 & \text{if channel is OFF (FREE).} \end{cases} \quad (1)$$

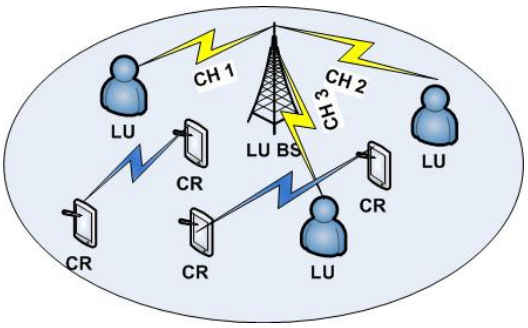


Fig. 1: Cognitive Radio Networks.

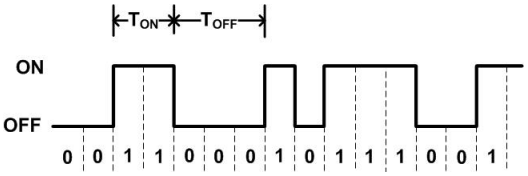


Fig. 2: Binary Channel Model.

For an alternating renewal process[33], let $f_{T_{ON}}(X)$ be the Pdf of the ON duration and $f_{T_{OFF}}(X)$ be the Pdf for the channel's OFF duration. Hence, the channel utilization μ is the expected fraction of time when the channel stays in its OFF state:

$$\mu = \frac{E[T_{OFF}]}{E[T_{ON}] + E[T_{OFF}]} \quad (2)$$

In equation 2, both ON and OFF periods are assumed to be independent and identically distributed (i.i.d). Since each licensed user arrival is independent, each transition follows the Poisson arrival process. Hence the length of ON and OFF period can be expressed using exponential distribution [34],[35] with pdf $f_X(t) = \lambda_X \times e^{-\lambda_X t}$for ON state and $f_Y(t) = \lambda_Y \times e^{-\lambda_Y t}$ for OFF state. Therefore, channel utilization μ in equation 2 can be written as:

$$\mu = \frac{\lambda_X]}{\lambda_X] + \lambda_Y]} \quad (3)$$

Where $E[T_{ON}^i = \frac{1}{\lambda_X}$ and $E[T_{OFF}^i = \frac{1}{\lambda_Y}$ are the rate parameter for exponential distribution. $E[T_{ON}^i]$ and $E[T_{OFF}^i]$ is the mean of distribution. Let $P_{ON}(t)$ be the probability of channel *i* in ON state at time *t* and $P_{OFF}(t)$ be the probability of channel *i* in OFF state at time *t*. The probabilities of $P_{ON}(t)$ and $P_{OFF}(t)$ can be calculated as:

$$P_{ON}(t) = \frac{\lambda_Y}{\lambda_X + \lambda_Y} - \frac{\lambda_Y}{\lambda_X + \lambda_Y} e^{-}(\lambda_X + \lambda_Y)t \quad (4)$$

$$P_{OFF}(t) = \frac{\lambda_X}{\lambda_X + \lambda_Y} + \frac{\lambda_Y}{\lambda_X + \lambda_Y} e^{-}(\lambda_X + \lambda_Y)t \quad (5)$$

Thus by adding equation 4 and equation 5, we can get

$$P_{ON}(t) + P_{OFF}(t) = 1 \quad (6)$$

4. Spectrum Management

The main focus of this section is to make an efficient channel selection and decision model which assists the CR user to spectrum mobility and enhance the spectrum utilization. The proposed spectrum management cognition cycle shows in figure 3 involves four major tasks such as spectrum sensing, spectrum analysis, spectrum classification and spectrum mobility. In the model, we also consider single hop network operation and ignored the route selection and route maintenance issues. Moreover, we consider two radio transceivers architecture which are sensing radio and data radio. The sensing radio is dedicated to spectrum monitoring includes particular radio environment and incumbent PU data base. The output of spectrum sensing process then feed into spectrum analysis process to characterize the spectrum hole information. All of this information is then processed by spectrum classifier and create available channel list based on channel duration (licensed spectrum ideal time) and quality of service. If there is no primary channel detected to be free then it will select unlicensed spectrum according to traditional CSMA/CA protocol. Due the dynamic nature of radio environment, the spectrum hole information or PU activities would be changed over time. Therefore, spectrum mobility is the processes which conveys this information from spectrum sensing process to spectrum classifier and relist the available channels. Further classification has been done on available channel list to create candidate channel list through fine scanning on the channels listed in available channel list. In the next step a back up channel list will form so that any CR node will select the channel which could maximize the channel utilization and finally transmit on that particular channel. In this report, we subdivided the model in two steps where CR users firstly perform proactive channel prediction to create available channel list. The usability status of all the channels is varying over time due to licensed user activities. Therefore, in the second step, CR user will update the available channel list that created in step one to adapt the radio environment dynamics.

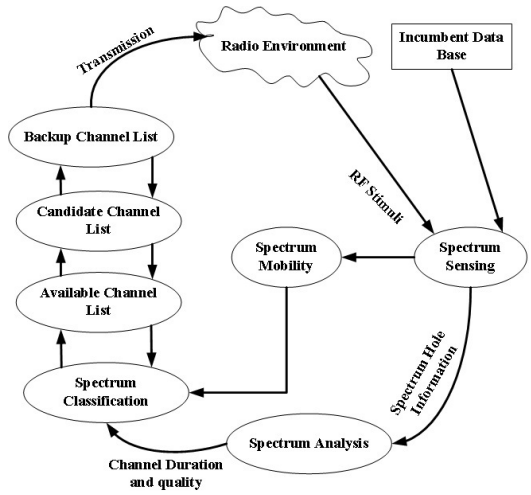


Fig. 3: Cognition Cycle of Spectrum management.

4.1 Step 1: Available Channel List

In our proposed model, an additional radio transceiver is assumed to monitor the radio environment continuously and create an available channel list to use for data transmission. Algorithm 1 shows the operational flow of step one where CR node will perform spectrum sensing to find unoccupied available channels through fast scanning. We also consider the geographical constrained imposed by regulatory authority of the particular place to protect per-defined incumbent licensed user using incumbent database. Moreover, CR user also uses the knowledge of previous scanning result to estimate the $P_{OFF}(t)$ of current channel selection. Here is $E[T^i_{MIN}$ the expected time required for minimum data transmission with the lowest packet size.

Algorithm 1 Available Channel List with Proactive Channel prediction

Load:Licensed User Database
K = Number of Licensed Users (Protected by Regulation)
for $i = 1$ to $N - K$ **do**
  Calculate $P_{OFF}(t)$
  if $E[T_{OFF}] \geq E[T_{MIN}]$ **then**
    $Avail_ch_list \leftarrow Channel(i)$
  else
    $Avail_ch_list \leftarrow Channel(unlicensed)$
  end if
end for

4.2 Step 2: Channel Classification and Update

The aim of this step is to classify the available channel list in two categories named as candidate channel and backup channel list. All the channels that are in available channel list can be treated as candidate channel as long as there is no licensed user operating. This candidate channel can become a backup channel through fine scanning. To be more precise all the candidate channels are scan at every 6 s for at least 30 s and outcome of the scanning result is the backup channel list. Any CR users before switching to the backup channel must be scan for at least 6 s to deal with any imperfect prediction of channel state in step one. Algorithms 2 illustrate the operational flow of channel classification as well as update scheme.

Algorithm 2 Candidate Channel List

Load:Available Channel List from Step 1
L = Number of Available Channel
M = Number of Licensed Channel in the System
U = Channel Utilization
C = Number of Candidate Channel
for $S = 1$ to L **do**
  Sense Channel (S)
  for $V = 1$ to M **do**
    if $P_{threshold}(S) \geq P_{threshold}(V)$ **then**
      $Candidate_{ch} \leftarrow 1$
    else
      $Candidate_{ch} \leftarrow 0$
    end if
  end for
end for
for $P = 0$ to $C - 1$ **do**
  $Backup_{ch} \leftarrow max(U(Candidate_{ch}), Backup_{ch}(P))$
end for

5. Performance Study

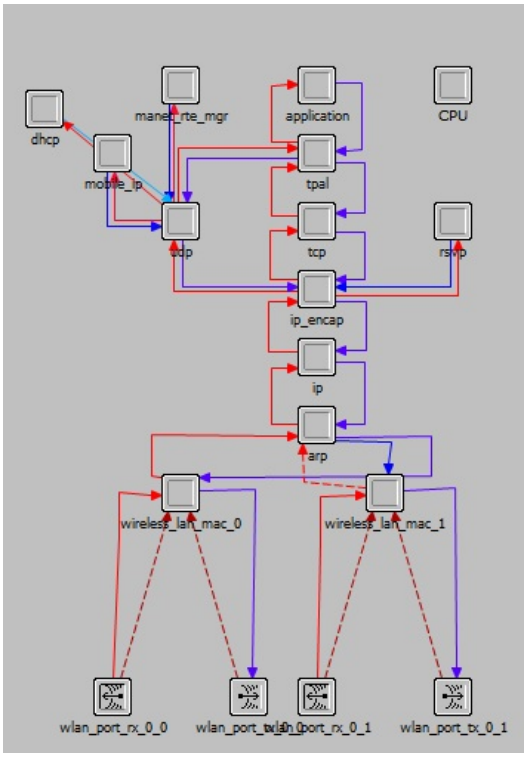
5.1 Simulation Setup

Here, we simulate an ad-hoc based cognitive radio IEEE 802.11g ad-hoc network consisting of multiple CR users operating at 2.4 GHz with eight license users. Each CR user is uniformly distributed over the network is assumed. We consider the case where a CR user tries to exchange packets with its neighbors. All the channels are assumed to have exponential distributed ON/OFF periods. To test our proposed algorithm we created a customized Cognitive radio wireless node where proactive predicted model is implemented in MATLAB and networking operation is done through OPNET 16.0. Our CR node is developed based on two wireless radio transceivers architecture which is shown in Figure 4(a). In the node model, the scanning radio is dedicated to listen the radio environment and creates a channel list that could be used for data transmission. The process model for scanning radio is a simple function $CR-Scan-data()$ which call the MATLAB library function and perform scanning and eventually updates the channel status table. This channel information is then sent to the data radio transceiver as an input to initiate the data transmission. In order to exchange the channel information of the neighboring nodes we adopt SYN-MAC[28] network initialization state protocol where it is assumed that all neighboring CR nodes are synchronized and has the channel set information of their neighboring nodes. We've designed our system having 5 different frequency channels and each User is assigned a particular frequency band which is designated as licensed user. At a particular time maximum five users can be present in the system. Therefore, the number of CR user is totally depending on how many empty slots is present. If the entire

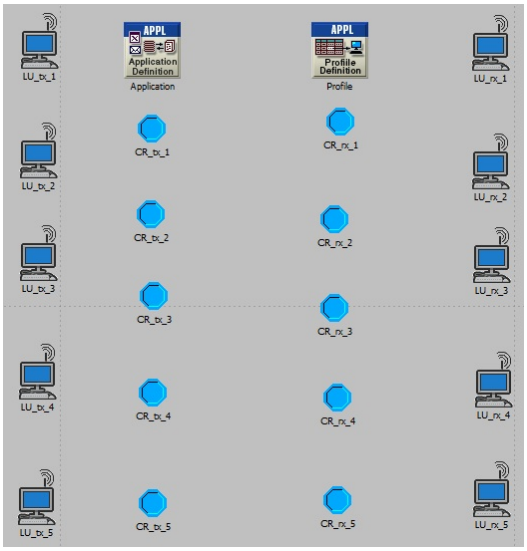
five licensed user are accessing the channels simultaneously then there is no room for CR user to operate in licensed user band. In that case CR user should go for unlicensed band which is out of scope or our current work. The network topology that has been used in order to measure the network performance is shown in Figure 4(b).

5.2 Simulation Results

In the first phase of our simulation we present the concept of cognitive radio networks along with proactive channel prediction. Figure 5 depicts the CR node performance in present of licensed user and shows that there is only one primary user operating that means we have four empty slots at 2,3,4,5 MHz band for the CR user to operate. In this case, with dedicated sensing transceiver, CR users can obtain perfect information of past and current channel status, and make accurate prediction of the future channel. In the first run we only allowed one CR user to enter in the system and find the available spectrum to start communication with CR receiver which is in 2 MHz In the second run another CR node enter the system and get the free channel at 3 MHz But in the third run we allow another licensed user who is owner of 3MHZ frequency band to initiate the data transmission. Therefore CR node 2 should vacate the 3 MHz frequency band immediately for the LU3 which is shown in left bottom side of figure 5. Right bottom side of figure 5 also shows the same cognitive radio concept for LU 2 and CR 1. The only exception is that, in this case CR 1 has to force to terminate or migrate to unlicensed band due to unavailability of licensed frequency band. In this part of research we didn't take consider the unlicensed frequency band as available channel list. Definitely, it could increase the system performance even though it might need more complex algorithm to mitigate CR users' coexistence with other unlicensed users coexistence with other unlicensed users. The next part of the simulation is to show how much spectrum utilization can be achieved when licensed user share the radio spectrum with other radio i.e. cognitive radio. The simulation scenario that has been used is shown in Figure 4(b) where we have 5 LUs, 5 CR nodes and low resolution video as LU application data. When only licensed users are using the network the overall utilization is 98.4kbps where as it could reach up to 286 kbps if we allowed the CR users to coexist with licensed users (figure 6). Hence, proactive predictive channel management schemes explore the cognitive radio concept for best utilization of unused spectrum while avoiding collision with the licensed users.



(a) Node Model



(b) Network Topology

Fig. 4: Cognitive Radio in OPNET.

6. Conclusion

In this paper we have presented proactive predictive channel management scheme in dynamic spectrum allocation systems. We have also proposed algorithms to construct the available channel list and backup channel list using the predictive traffic pattern of license users and maximize the

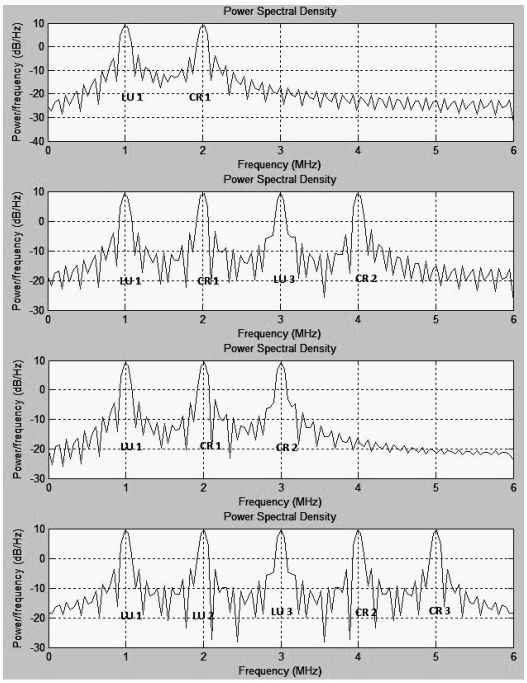


Fig. 5: Cognition Cycle of Spectrum management.

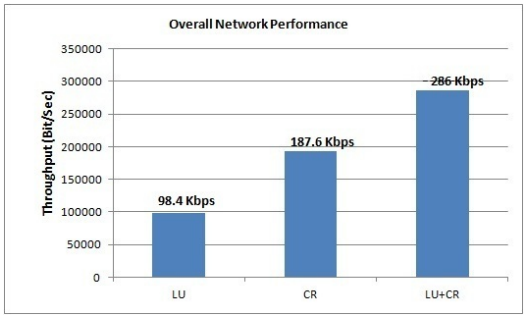


Fig. 6: Cognition Cycle of Spectrum management.

network utilization. To ensure the protection of incumbent licensed user, these channel lists is also being updated in dynamic manner. Our simulation result shows that performance enhancement and better network utilization is possible when CR users utilize the unused spectrum of licensed user in proactively.

7. Future Works

In the future, we would like to extend the proposed channel management scheme from single hop scenario to multi-hop scenarios. Future work may also involve co-existence among CR users for efficient channel selection algorithm.

References

[1] I. Mitola, J. and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13 –18, aug 1999.

[2] T. Weiss and F. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *Communications Magazine, IEEE*, vol. 42, no. 3, pp. S8 – 14, mar 2004.

[3] J. Gambini, O. Simeone, U. Spagnolini, Y. Bar-Ness, and Y. Kim, "Cognitive radio with secondary packet-by-packet vertical handover," in *Communications, 2008. ICC '08. IEEE International Conference on*, may 2008, pp. 1050 –1054.

[4] L.-C. Wang and C.-W. Wang, "Spectrum handoff for cognitive radio networks: Reactive-sensing or proactive-sensins?" in *Performance, Computing and Communications Conference, 2008. IPCCC 2008. IEEE International*, dec. 2008, pp. 343 –348.

[5] H. Kim and K. Shin, "Efficient discovery of spectrum opportunities with mac-layer sensing in cognitive radio networks," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 5, pp. 533 –545, may 2008.

[6] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz, "Cognitive radios for dynamic spectrum access - dynamic frequency hopping communities for efficient ieee 802.22 operation," *Communications Magazine, IEEE*, vol. 45, no. 5, pp. 80 –87, may 2007.

[7] L. Yang, L. Cao, and H. Zheng, "Proactive channel access in dynamic spectrum networks," *Physical Communication*, vol. 1, no. 2, pp. 103 – 111, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1874490708000268

[8] H. Kim and K. Shin, "Adaptive mac-layer sensing of spectrum availability in cognitive radio networks," *University of Michigan, Tech. Rep. CSE-TR-518-06*, 2006.

[9] L. Giupponi and A. Pérez-Neira, "Fuzzy-based spectrum handoff in cognitive radio networks," in *Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on*. IEEE, 2008, pp. 1–6.

[10] Y. Zhang, X. Chen, and L. Fu, "A demand-based spectrum allocation algorithm in cognitive radio networks," *Frontiers in Computer Education*, pp. 1011–1018, 2012.

[11] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 27–38.

[12] A. Motamedi and A. Bahai, "Mac protocol design for spectrum-agile wireless networks: Stochastic control approach," in *New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on*. Ieee, 2007, pp. 448–451.

[13] C. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional requirements for the 802.22 wran standard," *doc.: IEEE*, pp. 802–22, 2006.

[14] M. Kalil, H. Al-Mahdi, and A. Mitschele-Thiel, "Analysis of opportunistic spectrum access in cognitive ad hoc networks," *Analytical and Stochastic Modeling Techniques and Applications*, pp. 16–28, 2009.

[15] G. Salami, O. Durowoju, A. Attar, O. Holland, R. Tafazolli, and H. Aghvami, "A comparison between the centralized and distributed approaches for spectrum management," *Communications Surveys & Tutorials, IEEE*, no. 99, pp. 1–17, 2010.

[16] L. Le and E. Hossain, "A mac protocol for opportunistic spectrum access in cognitive radio networks," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*. IEEE, 2008, pp. 1426–1430.

[17] H. Su and X. Zhang, "Cross-layer based opportunistic mac protocols for qos provisionings over cognitive radio wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, pp. 118–129, 2008.

[18] ——, "Opportunistic mac protocols for cognitive radio based wireless networks," in *Information Sciences and Systems, 2007. CISS'07. 41st Annual Conference on*. IEEE, 2007, pp. 363–368.

[19] ——, "Cream-mac: An efficient cognitive radio-enabled multi-channel mac protocol for wireless networks," in *World of Wireless, Mobile and Multimedia Networks, 2008. WoWMoM 2008. 2008 International Symposium on a*. IEEE, 2008, pp. 1–8.

[20] L. Ma, X. Han, and C. Shen, "Dynamic open spectrum sharing mac protocol for wireless ad hoc networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*. Ieee, 2005, pp. 203–213.

[21] B. Hamdaoui and K. Shin, "Os-mac: An efficient mac protocol for

spectrum-agile wireless networks," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 8, pp. 915–930, 2008.

[22] J. Jia, Q. Zhang, and X. Shen, "Hc-mac: A hardware-constrained cognitive mac for efficient spectrum management," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, pp. 106–117, 2008.

[23] D. Raychaudhuri and X. Jing, "A spectrum etiquette protocol for efficient coordination of radio devices in unlicensed bands," in *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, vol. 1. IEEE, 2003, pp. 172–176.

[24] X. Jing and D. Raychaudhuri, "Spectrum co-existence of ieee 802.11 b and 802.16 a networks using the cscc etiquette protocol," in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*. IEEE, 2005, pp. 243–250.

[25] L. DaSilva and I. Guerreiro, "Sequence-based rendezvous for dynamic spectrum access," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*. Ieee, 2008, pp. 1–7.

[26] K. Bian, J. Park, and R. Chen, "A quorum-based framework for establishing control channels in dynamic spectrum access networks," in *Proceedings of the 15th annual international conference on Mobile computing and networking*. ACM, 2009, pp. 25–36.

[27] C. Cormio and K. Chowdhury, "Common control channel design for cognitive radio wireless ad hoc networks using adaptive frequency hopping," *Ad Hoc Networks*, vol. 8, no. 4, pp. 430–438, 2010.

[28] Y. Kondareddy and P. Agrawal, "Synchronized mac protocol for multi-hop cognitive radio networks," in *Communications, 2008. ICC'08. IEEE International Conference on*. Ieee, 2008, pp. 3198–3202.

[29] L. Cao and H. Zheng, "Distributed spectrum allocation via local bargaining," in *Proc. IEEE SECON*, vol. 5, 2005, pp. 119–127.

[30] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 3, pp. 517–528, 2007.

[31] J. Huang, R. Berry, and M. Honig, "Auction-based spectrum sharing," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 405–418, 2006.

[32] I. Malanchini, M. Cesana, and N. Gatti, "On spectrum selection games in cognitive radio networks," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*. IEEE, 2009, pp. 1–7.

[33] D. Cox, D. Cox, D. Cox, and D. Cox, *Renewal theory*. Methuen London, 1962, vol. 1.

[34] W. Lee and I. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 10, pp. 3845–3857, 2008.

[35] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *Selected Areas in Communications, IEEE Journal on*, vol. 4, no. 6, pp. 833–846, 1986.

Design and Analysis of an Independent, Layer 2, Open-Access WiFi Monitoring Infrastructure in the Wild

J. Milliken[1], A. Marshall[12]
[1]Department of Electrical and Electronic Engineering & Computer Science
Queens University Belfast, Belfast, Northern Ireland
{mmilliken02, a.marshall}@qub.ac.uk
[2]School of Computer Technology, Sunway University, Kuala Lumpur, Malaysia

Abstract - *While WiFi monitoring networks have been deployed in previous research, to date none have assessed live network data from an open access, public environment. In this paper we describe the construction of a replicable, independent WLAN monitoring system and address some of the challenges in analysing the resultant traffic. Analysis of traffic from the system demonstrates that basic traffic information from open-access networks varies over time (temporal inconsistency). The results also show that arbitrary selection of Request-Reply intervals can have a significant effect on Probe and Association frame exchange calculations, which can impact on the ability to detect flooding attacks.*

Keywords: WLAN, MAC, monitoring, analysis, flood.

1 Introduction

The increasing use of the WiFi protocol by the public has encouraged many innovations in social and economic domains. New business opportunities have developed around this rollout, with increasing numbers of cafes and fast food outlets offering WiFi as a selling point [1]. One result of this approach is that the importance of security and reliability within these networks becomes ever more significant [2].

While threats to user privacy and service availability in WiFi networks are well documented [2],[3], to date they have been addressed by providing additional (security) services for business users and, much more recently, home environments. Products which protect these systems are available [4], but none have been targeted specifically for protecting public WiFi infrastructure.

Although there has been some research in establishing techniques for detecting attacks against WiFi networks, none has yet produced a reliable solution to this problem. One of the reasons for this is the difficulty in obtaining real-world data to work with in security assessments [5]. The majority of data used for validation is contrived or unverified either through use of synthetic network traffic modelling [6], lab testbed approximations [7], or protected data [8]. This has led to a significant disparity between research conclusions and their practical application to real networks [5].

The creation of specific WiFi monitoring networks to support security research has been attempted in previous works [9], [10], [11]. The authors in [9], [10] established a single-use capture installation for WiFi traffic and analysed MAC layer traffic metrics. However these monitoring systems were deployed on campus networks and do not give sufficient information on the capture system itself to allow for duplication by other researchers.

In [11] the authors develop a large campus monitoring system and deal directly with the logistics of data protection concentrating on layers above the MAC, but no equipment details are given. Some non-academic environments have been examined [12], however the data tends to be from central RAID servers and cannot be made available for alternative research use or validation [8].

Thus while WiFi monitoring networks have been deployed in previous research endeavours, none have examined publically accessible live network data from public WiFi environments [13]. Furthermore none have given sufficient information to be able to replicate and verify the collection systems, and hence the data derived from them. Therefore it is impossible to determine the effect of the methodology used for data collection on the accuracy of the results drawn from the collected traffic.

1.1 Motivation

We have established that to date security in public, open-access WLAN networks has received limited attention, but is an area of growing importance as WiFi rollout expands. One of the major factors is the lack of freely available datasets in these network environments. This is a consequence of a lack of monitoring systems tailored for open-access networks coupled with a lack of detailed information on how to reproduce those systems already described in research.

A major challenge for any public monitoring system is preservation of users' personal data, for both legal and ethical reasons. Therefore the approach adopted here is to use only MAC and physical layer information which does not disclose identifiable personal data. The collection and analysis of MAC layer traffic maintains user confidentiality and privacy which can be a significant barrier for live network experiments.

A key challenge therefore is to use only MAC and physical layer information to gather sufficient information to allow subsequent analysis of the behaviour of the network and its users'. This approach also provides

interesting research relevant material as all information about connections, access, users and traffic usage are available without requiring more sophisticated higher level information which would disclose users' identities [14].

This motivated the authors to investigate and produce an open-access WLAN monitoring system that can be deployed alongside live networks with minimum network disruption to provide a source of live network data. There is a clear need for a well-documented, replicable system for use in WLAN research data gathering applications. This would ensure maximum confidence in the conclusions drawn from the data, as the results can be verified independently.

2 System Design Considerations

There are multiple factors to take into consideration when designing any network monitoring system [15]. These include cost, data quality, resource consumption, and physical and networking access to the equipment. Additional factors become influential once the system is intended for use in multiple locations or for other researchers to utilise. These include automation, sophistication, portability, ease of replication and assembly / disassembly effort required. For use in public environments yet more issues have to be addressed, such as size, privacy, concealment and network disruption.

In order to address all of these, sometimes competing, concerns, an independent network was chosen. An independent monitoring network is defined here as a system which is totally delinked and independent from the network which is to be monitored, and is designed to collect the same data from the wireless medium but with minimal network disruption. This is in contrast to an integrated system, which connects directly with the equipment which is to be monitored.

The two principle concerns for the authors were ensuring that the traffic collected is not affected by the act of monitoring and that minimal disturbance of the monitored network is possible. Guaranteeing a lack of network disruption in particular was seen as paramount for obtaining the agreement of businesses offering open-access services to the public. Equally important was the assurance that user privacy could be adequately maintained.

2.1 Structure

Constructing an independent system requires devices which can carry out all data collection, storage, networking, processing and reporting requirements in the most economical and space saving manner possible. The principal components of such a system are outlined below and in Figure 1. Specifics about the actual devices, code and data used are available from the authors[1].

Monitor Station. An Access Point (AP) can also act as a Monitoring Station (MS), by invoking monitor mode to collect all WLAN frames transmitted on a set frequency. Since monitors are passive devices they do not contribute to the network traffic. Thus they are much more suited to capture the activity of a network environment than monitors embedded within APs, which only see traffic for the device.

Attacker Station. APs can also act as an Attacker Station (AS), carrying out WLAN attacks as requested. This is required if there are to be packets injected into the network to test the response of the system or to generate capture files that are known to contain attacks.

Mini-PC. This device is used to facilitate outside communication with the installation, on-site processing and time synchronisation. The mini-PC requires small form factor, reasonable processor power and 2GB of RAM.

Network Attached Storage (NAS) Hard Drive. The capacity of the hard disk should be around 1TB and have dual disks to allow for RAID 1 recovery in the event of a network or system failure.

PoE (Power over Ethernet) Switch. The switch acts as the central connection between all other equipment. The MS / AS may be PoE powered as this reduces reliance on cables and power socket availability for device positioning.

Internet Connectivity. Internet connectivity is a key aspect of the design given it allows remote reporting of summary data as well as failure assessment and updates without requiring a physical person on site. A 3G dongle or any similar wireless interface is a portable option.

2.2 Software / Hardware Identification

The sophistication of the system is largely dependent on the software employed and thus simple, repeatable and portable solutions were sought out. Principal software used in the system is listed below to allow other researchers to copy the system or validate conclusions based on it. All devices are commercially available.

MS / AS. The MS devices in use are Ubiquiti Nanostation Loco APs with an inbuilt bi-directional antenna and the AS devices are Ubiquiti Picostation APs with an attached omni-directional antenna. These devices have an OpenWRT Backfire 10.03.1 build already available in the OpenWRT repositories. Both stations require SSH for secure communication and remote login, NTP for time synchronisation and NFS to allow mounting of the NAS equipment for data storage. Each MS also has TCPDUMP installed as the capture programme while an AS contains the OpenWRT version of the Aircrack-ng WiFi attack suite. The remainder of the scheduling, reporting, logging and traffic capture handling processes are done using BASH scripts.

Mini-PC. The Mini-PC in use is an Acer Aspire Revo R3700 running Ubuntu 10.4. Programs required for the device include an NTP-server, NFS-server and SSH as well as Python. The Mini-PC acts as a time synchronisation server while NFS is used to mount the NAS.

Python (>2.5) is employed as the analysis platform for data amassed on the system. Given all captures are being collected remotely this does not necessarily allow for

[1] They have not been provided in an open online form as the installation and data collection was agreed only for research purposes

physical collection of raw data. Bandwidth requirements also come into play. Thus a python program is utilised to summarise traffic details on a weekly basis and report the results to central online storage. This strikes a balance between data fidelity and processing power / bandwidth.

In depth details about the operation of the software or scripts in use are available from the authors. The equipment and software in use was selected with the purpose of being extensible to larger deployments if necessary. The backend equipment is capable of handling many more additional monitors and attackers as required, limited only by the number of network ports on the PoE switch.

The extensibility of this system is one of the primary drivers for the cost of the system, setting it apart from, for example, a home router with a USB Flash disk attached. The proposed system allows for central control, monitoring and logging of data from multiple locations as well as failure tolerance and on-site processing. These capabilities can be extended easily to additional devices, limited only by the number of Ethernet ports on the PoE switch, which is not the case with a more rudimentary system.

A flash, or solid state, drive may have faster read and write times and be less liable to errors than the RAID array proposed, but for the capacity chosen the price would be disproportionately high. Simply adding flash storage to an existing AP is deemed an unacceptable solution too, as invoking packet capture and storage processes place a large burden on an already resource constrained device, slowing down normal operations. This violates the principle of not influencing the network being monitored.

2.3 Data Collected

Data collected is restricted to 802.11 Layer 2 MAC frames as this alleviates many of the confidentiality and user privacy issues. In many cases alleviating these concerns for network owners and administrators can be the largest barrier to successful deployment of a system such as this. The MS captures are all truncated to a maximum frame size of 120bytes in order to allow for a sufficient amount of the MAC header to be dissected for all packets but all other payload data is obfuscated.

It is possible for entire packets to be collected however the authors note that there may be legal and data volume ramifications of this [11]. All collected data is stored locally with reports and summaries reported via internet connection. It is possible for raw data to be retrieved through a researcher on site or via FTP if deemed viable.

A further benefit of collecting data in this manner is the lack of disruption to the network owner / administrator. The network data is collected entirely without interaction with the monitored system, meaning that is can be installed without fear of corrupting or influencing the data being monitored. Conversely in an integrated system the monitor can impact on the monitored network. This can be as simple as slowing down network equipment due to extra load or more subtle, such as if the monitor is subverted as a means of an attacker accessing data or causing a DoS condition.

3 System Deployment

The equipment and layout described has been deployed successfully for medium term (roughly 3 week) remote data acquisition installations in multiple locations. One particular location, with two capture traces taken at different times, is discussed. A current trial with this deployment structure is presently deployed in a live network environment in Sunway Pyramid Shopping mall in conjunction with Sunway University, Malaysia.

The first traffic capture trace was taken at a Caravan park in Northern Ireland in August 2010. The WiFi network at the park is provided by JSR Technology [16], who established an 802.11g public access network with a common password required for access. The deployment location was within the attic of an enclosed washroom at the centre of the park within sight of the targeted Access Point Tower (APT) with 3 outdoor directional APs.

A second capture trace was taken at the same caravan park in February 2011 in order to ascertain how much the network packet characteristics can change over time. This time period was chosen as the inclement weather in Northern Ireland in this month leads to a drastic reduction in usage of the park, giving an insight into annual variation in traffic characteristics. The same equipment was used at the same location and with the same orientations as the previous installation in order to maintain compatibility of the results for each MS in both timeframes.

Device	Programs in Use
Mini-PC	NTP, NFS, SSH, Python, Bash, sendmail
Monitor Station (MS)	NTP, NFS, SSH, Bash, TCPDUMP
Attacker Station (AS)	NTP, NFS, SSH, Bash, Aircrack-ng

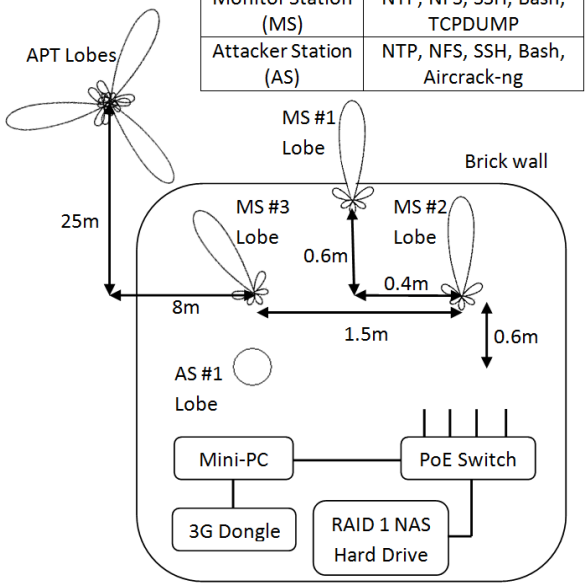


Figure 1. Layout of Equipment with Software

4 Analysis of Traffic

The data presented in the following section (see Table 1 and Table 2) is based on packet data and traffic summary details for the Caravan Park site in both timeframes

indicated for two of the monitors, MS #1 and MS #3. In Table 2 details for a single AP are given; this references all packets which were sent to or from the WiFi MAC address for that AP and all probe requests that were responded to by that MAC.

Table 1: MAC-Level Traffic Details for Two WLAN Packet Capture Periods and two Monitoring Stations (MS).

Per Capture	**Timeframe 1**		**Timeframe 2**	
Capture Start	11 Aug. 2010		22 Feb. 2011	
Capture length	21 (days)		21 (days)	
Per Monitor	***MS#1***	***MS#3***	***MS#1***	***MS#3***
Tot. # Packets	144.5M	126.7M	76.6M	30.4M
Capture Size	10.6Gb	9.4Gb	6.1Gb	1.9Gb
% Data Packets	26%	27%	25%	22%
% Mgmt. Packets	38%	35%	47%	65%
% Control Packets	36%	38%	28%	13%

Table 2: MAC-Level Traffic Details for two Monitoring Stations (MS) During the First Traffic Capture Period, Focusing on the Individual Traffic Associated with a Single Access Point.

Per Capture	**Timeframe 1**	
Capture Start Date	11 August 2010	
Capture length (Days)	21	
Per Monitor	***MS#1***	***MS#3***
# Packets	15.2M	15.2M
% Data Packets	39.5%	30.1%
% Management Packets	41.8%	65.5%
Av. # Beacon Frame (per hour)	11609	18222
Av.# Probe Frame Exchanges (per hour)	95.9	27.7
Av. # Association Frame Exchanges (per hour)	0.47	0.15

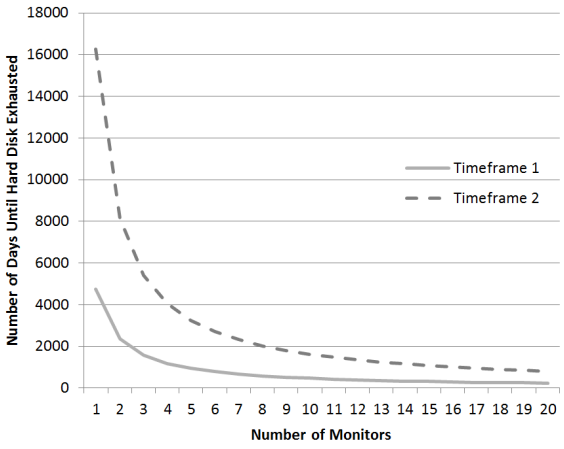


Figure 2. Hard Disk Exhaustion Rate for Monitoring System Based on a 1TB Hard Drive.

Using the information from Table 1 it is possible to estimate the likely length of operation based on number of monitoring stations and timeframe of operation, where the limiting factor is a 1TB hard-drive space. This is presented in Figure 2, which shows that the space available is suitable for extension of up to 20 monitors for Timeframes 1 and 2, giving operational lifetimes of approximately 200 and 800 days respectively.

While these values are within reasonable bounds of operating times for this installation, this is not certain to be the case on more highly utilised networks. In which case the difference (average 30%) between the two timeframe curves could have a significant impact on the estimation of required hard drive space.

4.1 Temporal Inconsistency in WLANs

Two factors regarding the traffic statistics were noteworthy from these results. The differences between:

A. Traffic details for the same MS over two separate timeframes (Table 1)
B. Traffic details for two MS's monitoring the same network (Table 2)

The disparity is evident from the values in Table 2 but less indicative in Table 1, possibly obfuscated by the size of the trace. In order to determine the difference over time and monitoring station more formally, two statistical tests were performed on the Beacon, Probe and Association frame datasets for MS#1 and MS#3; two sample Kolmogorov-Smirnov [17] and two sample Mann-Whitney [18]. Both of these are nonparametric tests for the equality of continuous, one-dimensional probability distributions.

For both tests between timeframes and monitor locations the tests rejected the hypothesis that the datasets are drawn from the same dataset. The inference then is that there is unlikely to be correlation between the traffic distributions of Beacon, Probe and Association frames in different timeframes.

4.2 Effect of Temporal Inconsistency on WLAN Attack Detection

A lack of correlation between traffic characteristics over time may have implications for attack detection where thresholds are employed. Example WLAN attacks where this is employed include Probe and Association Flood attacks [3]. The attack operates by sending a "disproportionately" high number of Probe or Association requests to an AP. This consumes AP resources and, either through resource exhaustion or over-work, the AP can hang, restart or shutdown, causing a Denial of Service condition.

Conventional attack detection methods rely on establishing a level that represents "proportionate" in the network. Any deviation above this level is deemed intrusive or suspicious behaviour and an alert is generated. It is accepted that these threshold values are not constant and must be established for each network, however in current IDSs' the typical approach to threshold selection is expert (i.e. human) knowledge. This is equivalent to a "best guess" scenario and is set only once when the monitoring system is installed.

With evidence of temporal inconsistency this calls into question the fidelity of a set threshold during network operation. It may be more appropriate for the Flood detection threshold to be dynamically established at intervals to maintain an optimum level of security. A further implication is that when investigating the performance of algorithm or detection system performance in live networks, the testing needs to occur at multiple times during operation in order to determine experimental accuracy in real world deployments.

4.3 Window Timeout In Independent WLAN Monitoring Systems

Another metric which can impact on threshold selection and influence Flood attack detection in independent networks is the window timeout for frames. Acquisition of data from a live network with no internal access to the networking devices under examination provides a set of challenges with regards to window length assumptions. This manifests itself in two issues; Request-Reply Intervals and estimation of the number of connected clients.

In the traffic details documented in Tables 1, 2 and 3 a frame window (size = any 10 consecutive frames) is established after the reception of a request or reply frame. If an additional or retry frame of the same type is received within this window it is re-established. If the window expires then the interval between the logged request and logged reply is the calculated interval.

These intervals can be used to determine how well the network is performing (or loaded) proportional to how quickly it can respond to requests. For Flood attack detection purposes they can also be used to assess how busy the AP is in processing requests. If the request and reply chain has been closed then the AP is no longer busy using resources to process that conversation.

4.4 Variation in Request – Reply Exchange

In an environment where an independent monitoring system is employed, internal AP parameters for the monitored network may not be known for a number of reasons:

A. There may be APs from different vendors with unique window lengths.

B. The windows may appear to be different for the monitor depending on its position relative to the monitored system (Hidden node).

Where AP access is available before installation often these parameters are not easily determined or set.

The lack of internal AP information in independent monitoring systems creates difficulties with calculation of correct request-reply intervals, complicated by the presence of excessive re-transmissions. This becomes an issue when tracking Probe and Association frame exchanges.

Given the unreliability of the wireless medium it is possible for packets to be lost or corrupted in transit, which can require legitimate re-transmissions from either client or AP. In the case of requests it is not evident whether the retransmissions are occurring faster than the responder can process or if the packets are lost. In the case of replies it is not evident whether the first reply was received correctly and no more action has been taken or whether the packets have been lost.

The window length determines how many packets the monitor should wait before discarding an unpaired request or accepting the close of a conversation with a reply. Table 3 shows the effect of varying this window length on the average percentage increase on frame exchange times against what the exchange time would be if no retries were accepted.

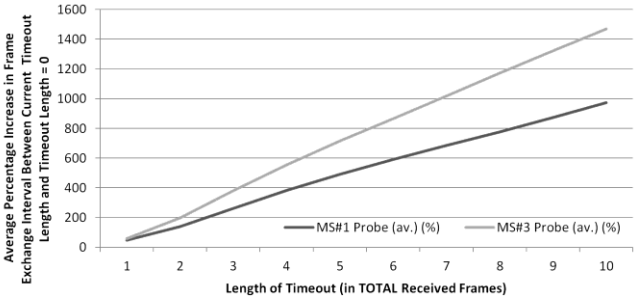


Figure 3. Effect of Request – Reply Window Lengths on Exchange Interval for Probe Frames (vs. No Retry values (%)) for Capture 1 MS #1 and MS#3

This retransmission chain can extend for a considerable amount of time, significantly increasing the time interval of the exchange. From Table 3 the disparity can influence the interval length by between 21% and 1468% depending on

Table 3: Effect of Request – Reply Window Lengths on Exchange Interval for Probe and Association Frames (vs. No Retry Values (%)) for Capture 1 MS #1 and MS#3.

Frame Type	Length of Timeout (in TOTAL Received Frames)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MS # 1										
Probe (average.) (%)	46	138	261	383	492	590	685	776	872	973
Assoc. (average.) (%)	23	34	39	45	53	53	54	55	56	56
MS # 3										
Probe (average.) (%)	58	196	379	554	714	866	1016	1172	1323	1468
Assoc. (average.) (%)	21	48	51	60	60	60	60	65	63	63

the frame type under consideration. Furthermore it may be noted that while these values are not identical across each MS or AP analysed, the trend is still observed in each. Figures 3 and 4 show that probe frames do not reach a levelling out value within the measurement parameters investigated here however association frames level out after a timeout window size of 5 frames.

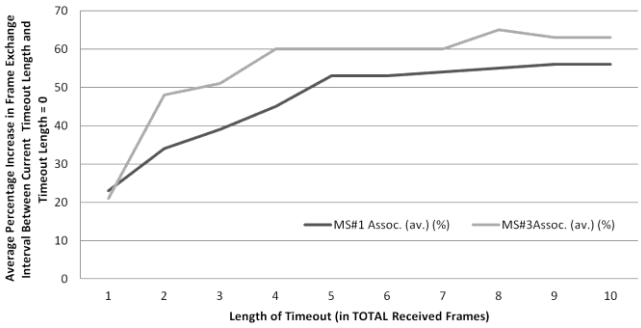


Figure 4. Effect of Request – Reply Window Lengths on Exchange Interval for Association Frames (vs. No Retry values (%)) for Capture 1 MS #1 and MS#3

4.5 Effect of Request-Reply Interval on WLAN Attack Detection

Selection of an appropriate window parameter helps maintain a link between the network traffic seen by the AP and the monitoring system itself. A window value too low can increase the amount of connection attempts seen and discard valid exchanges which were delayed by propagation. A window value too large can decrease the number of connections seen and absorb valid exchanges, having assumed subsequent exchanges are delayed by propagation.

Determination of an appropriate window impacts on detection performance in a similar fashion if the window is too large or too small. Too small and it is assumed that unanswered requests are discarded by the AP. Valid retries may then be counted as new attempts at connection, and could push alerts over any Probe Flood detection threshold unduly. Furthermore additional reply frames may be delinked from their request conversation, leading to suspicious activity alerted through any protocol non-adherence detection algorithms. Should the window be set too large then floods can be misread as valid request retry chains.

Consequently the effect of this variation in exchange length is that any WiFi Flood attack detection algorithm in independent monitoring systems which relies on traffic statistics is susceptible to changes in seemingly "best guess" or arbitrarily chosen values.

4.6 Connected Client Estimation

Estimation of the number of clients connected to the AP is another issue present in independent monitoring systems. Without direct access to the DHCP details of the APs it is not possible to categorically state how many clients were still being afforded memory resources on the network. The absence of data packets does not necessarily indicate that the AP has discarded the client association. Consequently the actual timeout period for association resource allocation reclaim is unknown.

Given that a client does not have to rigorously follow the 802.11 protocol and de-authenticate and disassociate before leaving the network, an assumption must be made. The procedure followed in the course of this work relies on Request To Send (RTS) and Clear To Send (CTS) exchanges, given they are a necessary preclusion to sending valid data frames from clients to APs in networks where it is employed.

Upon detection of a matched RTS-CTS pair the client in the exchange is deemed to have been connected during the hour of capture during which the exchange occurred. According to this system a count of unique active clients associated per hour is kept, with the effective maximum timeout interval being 2 hours and effective minimum timeout interval being 1 hour.

This approach leads to the estimation of connected clients shown in Table 4. Note that for Capture 2 MS #3 in Table 1 the percentage of data traffic is reasonably high (21%), however no clients are assumed in Table 4. Having investigated the data capture trace further it appears that this data frame set primarily contains QoS best effort and null data frames. This confirms the validity of the approach for counting clients via RTS-CTS exchanges rather than data packets as it gives artificially high results.

Table 4: Client Estimation for 2 WLAN Packet Capture Periods and for two Monitoring Stations (MS) During Each Capture.

Per Capture	**Timeframe 1**		**Timeframe 2**	
Per Monitor	***MS#1***	***MS#3***	***MS#1***	***MS#3***
Est. Av. # of clients per day	9.4	2.52	0.0	0
Total # of Unique AP-Client Exchanges	23	9	5	0

The disparity in the usage levels in both timeframes shows that the number of users can vary wildly over large timeframes, particularly where seasonal factors are a consideration, such as in a caravan park, this can have a large impact on connected clients. Thus the traffic on a network cannot be assumed to be the same at all times. To get a fuller picture of the characteristics of a monitored network it must be assessed over multiple timeframes, as one timeframe cannot be considered typical by default.

5 Conclusions

The continued use of WiFi networks to extend mobility for internet services in public environments presents opportunities for intruders and challenges for security research. Direct traffic monitoring within a live network can give researchers an up-to-date source of information on new

attacks, traffic behaviours and performance of detection algorithms.

This paper has described a monitoring installation which is designed for supervising and collecting WiFi traffic in public networks while maintaining user data privacy. It was constructed with readily obtainable materials so that the deployment and results can be easily replicated and verified. The installation takes the form of an independent network, which was determined, using a set of assessment criteria, as the most appropriate solution.

MAC-level summary information gathered from a deployment of the system at two different capture intervals has identified two factors affecting WiFi security in independent networks; temporal inconsistency in traffic observations and selection of Request-Reply window timeout parameters. Temporal inconsistency affects the reliability of attack detection algorithms based on thresholds, particularly Probe and Association Flood attacks, which may suffer if not tested over multiple timeframes, even for the same network.

It has been shown that for independent monitoring systems the lack of internal AP access requires that assumptions be made about Request-Reply intervals. This is also true for integrated systems in cases where multiple APs are used and general wireless measurement issues have impact, such as multipath or hidden nodes. In both systems the choice of timeout parameters for Request-Reply intervals can have a significant effect on frame exchange times, which can impact on traffic based WiFi attack detection approaches.

All of these factors; frame interval calculation and traffic temporal inconsistency, arise not only as a result of an independent monitoring system but also from a lack of investigation in current research into the challenges of network collection for public open-access networks. However once these challenges have been overcome, where a public, open-access WLAN infrastructure is to be investigated, independent networks have multiple characteristics which make them attractive solutions.

6 Acknowledgements

The authors would like to acknowledge Jason Rosborough of JSR for permission and assistance in monitoring the network resident within the caravan park. This work was funded by EPSRC (UK) under grant number EP/H004793/1

7 References

[1] JiWire, 2009. JiWire Mobile Audience Insights Report [online]. Available at http://www.jiwire.com/downloads/pdf/JiWire_MobileAudienceInsights_1H09.pdf [Accessed on 9 Sept 2011].

[2] CPP, 2010. UK Wireless Network Hijacking, A CPP White Paper [online]. Available at: http://blog.cpp.co.uk/files/uploads/cpp-research/UK_Wireless_Network_Hijacking_2010.pdf [Accessed on 9 Sept 2011].

[3] Tews, E., Beck, M., 2009. Practical attacks against WEP and WPA. In: Proceedings of the second ACM conference on Wireless network security (WiSec'09). New York, USA.

[4] Cheng, et al., 2006. Jigsaw: solving the puzzle of enterprise 802.11 analysis. In: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '06). New York, USA.

[5] Ibrahim, et al., 2008. Assessing the challenges of Intrusion Detection Systems. In: Proceedings of the 7th security conference. Las Vegas, USA.

[6] Stakhanova, N., Basu, S., Wong, J., 2008. On the Symbiosis of Specification-Based and Anomaly-Based Detection. Computers and Security, 29(2), pp.253-268.

[7] Chen, G., Yao, H., Wang, Z., 2010. An Intelligent WLAN Intrusion Prevention System Based on Signature Detection and Plan Recognition. In: Proceedings of Second International Conference on Future Networks. Sanya, China.

[8] Afanasyev, et al., 2010. Usage Patterns in an Urban WiFi Network. IEEE/ACM Transactions on Networking, 18(5), pp.1359-1372.

[9] Yeo, et al., 2005. An Accurate Technique for Measuring the Wireless Side of Wireless Networks. In: Proceedings of First International workshop on Wireless Traffic Measurements and Modeling. Berkley, USA.

[10] Kotz, D., Essien, K., 2005. Analysis of a Campus-Wide Wireless Network. Wireless Networks, 11(1-2), pp.115-133.

[11] Bratus, et al., 2009. Dartmouth Internet Security Testbed (DIST): Building A Campus-Wide Wireless Testbed. In: Proceedings of Second Workshop on Cyber Security Experimentation and Test. Montreal, Canada.

[12] Deshpande, U., McDonald, C., Kotz, D., 2008. Refocusing on 802.11 Wireless Measurement. In: Proceedings of 13th Passive and Active Measurement conference. Cleveland, USA.

[13] Tala, et al., 2011. Guidelines for the accurate design of empirical studies in wireless networks. In: Proceedings of the 8th International ICST Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TRIDENTCOM '11). Shanghai, China.

[14] Armknecht, F., 2007. Who Said That? Privacy at Link Layer. In: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM). Anchorage, USA.

[15] Mahajan, et al., 2006. Analyzing the MAC-level Behaviour of Wireless Networks in the Wild. In: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '06). New York, USA.

[16] JSR Technology, http://www.jsrtechnology.com/.

[17] Massey, F.J.,1951. The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association. 46(254), pp.68-78

[18] Halpern, M.,1960. Extension of the Wilcoxon-Mann-Whitney Test to Samples Censored at the Same Fixed Point. Journal of the American Statistical Association. 55(289), pp.125-138

Analyzing the Performance of Handoff in IEEE 802.16e Mobile WiMAX Networks

Sharmistha Khan[1], Ji Hyun Lee[2], and Golam R. Khan[1]

[1]Department of Electrical & Computer Engineering, Prairie View A&M University, Prairie-View, TX
[2]Department of Electrical Engineering, Tuskegee University, Tuskegee, AL

Abstract

It is well known that Mobile WiMAX is the latest technology that promises a broadband wireless access over long distance. In mobile WiMAX networks, it is essential to provide continuous network connectivity to satisfy high levels of mobile service quality or mobility. So, in order to support mobility in mobile WiMAX, it is necessary to provide handoff. Handoff is an essential process in wireless networks to guarantee continuous, effective, and resilient services during Mobile Station (MS) mobility. Although different kinds of handoff optimization schemes have already been adopted to support short interruption time, cost effectiveness, low handoff latency, etc., in order to receive excellent performance in the handoff process, we are looking for other methods. This paper focuses on the analysis of the performance of the handoff process in Mobile WiMAX. This paper aims to find out the factors/parameters of the WiMAX module that affect handoff performance the most, such as handoff duration of less than 50 ms and mobility speed up to 120km/hour. Simulation results show that some of the parameters of the WiMAX do not have any influence on handoff latency, and some others have a great impact towards achieving shorter handoff duration time. The results also show that the handoff times could vary for different speeds of the Mobile Station (MS).

Keywords: Handoff , Mobile WiMAX

I. INTRODUCTION

The WiMAX Forum developed the most modern wireless technology named WiMAX in early 2001, which is a telecommunications protocol that provides fixed and fully mobile Internet access. There are many positive aspects of this technology; among them, one of the most important is, the support of a large coverage area. WiMAX provides the support of wireless connectivity with a minimum range of 30 miles. WiMAX technology also offers high speed broadband access to mobile internet which transfers data, voice, and video. In WiMAX when a user uses a 20 MHz data rate the bandwidth for this data rate can be up to 70 Mbps. Whereas WiFi offers a short range of data transfer with a maximum bandwidth of 54 Mbps [1] [2].

The IEEE 802.16 standard forms the basis of WiMAX technology. The WiMAX Forum gradually improves the functionality and approves different generations for this standard. Usually these standards differ in two different forms generally known as “802.16d” or “Fixed WiMAX” and “802.16e” or “Mobile WiMAX”. The 802.16d-2004 standard has no support for mobility. Mobility support is when a user is in motion or in a vehicle and can access the wireless network easily. To solve the problem of mobility the IEEE 802.16e-2005 standard was published which had full support for mobility. Since this standard introduced the support for mobility it is known as “Mobile WiMAX” [2].

In Mobile WiMAX networks when a Mobile Node (MN) changes its location the MN moves the point of attachment to the network. In this situation it is essential to provide continuous network connectivity to satisfy high levels of mobile service quality. Here the major issue concerning implementation of Mobile WiMAX is providing effective handoff. Handoff is the process of changing a Mobile Station’s (MS’s) network connectivity from one Base Station (BS) to another BS. Providing the support for ongoing video call, or Voice over IP (VoIP) conversations for mobile users makes it necessary to make the handoff process be as fast as possible. Therefore, in order to decrease handoff interruption time and unnecessary call drops it is important to implement

different handoff optimization methods. Providing different Quality of Service (QoS) support such as high speed data transmission, low handoff latency, a reduced amount of packet loss in Mobile WiMAX network, and various handoff enhancement approaches have been proposed.

For the past few years of Mobile WiMAX technology many researchers have explained the handoff process in Mobile WiMAX along with different handoff techniques. Some proposed different effective algorithms for improving handoff methods; while others proposed vertical handoff schemes with the aim of reducing handoff signaling overhead on the wireless backbone and providing a low handoff delay to mobile nodes [3]. On the other hand, other researchers introduced Secure Internet Protocol (IPSec) based end-to-end securing solution for real-time services, which provides a secure and fast seamless handoff solution while preserving the QoS and security when moving between heterogeneous access networks [4]. Again, various fast handoff techniques have also been suggested in literature to meet QoS requirements. Hence, in general, we found that different researchers worked on different aspects but with the same general motto, to improve handoff performance in Mobile WiMAX. All of the handoff techniques implemented the process in such a way that it always supported short interruption time, low handoff latency, high speed, and was cost effective. Still, a great deal of research is being conducted in order to determine a more efficient handoff processes to reach the desired performance. If we want to improve the performance of the handoff methods we need to first analyze the present condition of the performance of the implemented handoff processes in Mobile WiMAX. The performance of the handoff processes are regulated by the parameters of the WiMAX module. Therefore, we need to adjust those parameters in order to increase handoff performance by achieving shorter handoff duration time. Not only the parameters but also the velocity of the MS may have some impact on the performance of the handoff process. Therefore, our main focus is to analyze the performance of the handoff methods in Mobile WiMAX by investigating different performance related parameters with varied MS velocity.

This paper is organized as follows. After the introduction, the handoff in Mobile WiMAX is described in Section II. Simulation results are presented and discussed in Section III. Based on the simulation results, conclusions are given in Section IV.

II. HANDOFF IN MOBILE WIMAX

Handoff is a process with the intention of changing the network access point of a mobile node without any data loss or disturbance of the current connection while a call is in progress. Therefore, in mobile WiMAX, the basic meaning of handoff process is to provide uninterrupted connectivity when a mobile station (MS) transfers from the air-interface of one base station (BS) to the air-interface of another BS. Telecommunication reasons for conducting handoff can vary. Here we have tried to mention some of them.

• When the cellular phone is moving away from the area covered by the serving BS, the phone gets outside the range of the serving BS. Therefore, in order to avoid call termination, the call needs to be transferred to the area covered by another BS [5].

• When the signal strength is not enough to maintain a proper call at the edge of a serving cell, the call needs to be transferred to another cell [6].

• When the capacity for connecting new calls of a cell in a BS becomes full or more traffic is pending, it is required to free-up some capacity in the first cell for other users who can only be connected on that cell. Therefore, the existing calls or the new calls from a phone that is located in an area that is overlapped by both cells can be transferred from the first cell to the second cell [5] [6].

• In non-CDMA networks several phones use different cells but same channel. As a channel is being used by several phones, disturbing co-channel interference comes from one to another. Therefore, in order to avoid the interference, the call is transferred to a different channel in the same cell or to a different channel in another cell [5].

• When the behavior of a MS changes in such a way that a fast moving MS suddenly stops. Therefore, to provide better capacity a large cell can be adjusted with a small cell. [6]

• In vertical handoff, a faster network is occasionally available. Therefore, the phone changes its network to that cheaper one [6].

Handoff is divided roughly into two broad categories: hard and soft handoffs with different variants based on the used technology. They are also known as “break before make” and “make before break” handoffs. Generally, in hard handoffs old connections are broken before new connections are

created; in soft handoffs, both existing and new connections are used during the handoff process. IEEE 802.16e defined the handoff process into three major classifications known as Hard Handoff (HHO), Macro Diversity Handoff (MDHO), and Fast Base Station Switching (FBSS), where HHO is mandatory in WiMAX systems, the other two types of handoff are optional [7].

The HO procedures can be divided into two major phases known as network topology acquisition parts and the actual HO process. The network topology acquisition phase can be divided into 3 sub procedures, which are network topology advertisement, MS scanning of neighbor BSs, and association procedure. These three sub-procedures are executed before HO through the backbone network. Again, the actual HO phase can be divided into some other sub phases which are cell selection, handoff decision and initiation, synchronizing with new downlink and obtain parameters, obtaining uplink parameters, ranging and uplink parameter adjustment, MS re-authorization, re-register, and termination with the serving BS [6]. All of these sub-phases are executed during HO.

III. SIMULATION AND RESULTS

The main goal of these simulations was to investigate the impact of different parameters of the WiMAX module during handoff as well as examine the properties of Mobile WiMAX. A very simple and basic scheme was designed in order to keep the simulation process simple and trouble-free. To implement handoff in Mobile WiMAX we have used the Network Simulator 2 (NS-2 Version 2.29), which forms a basis for the simulations. Additionally, in order to support Mobile WiMAX, we have to use two other particular modules known as WiMAX and Mobility module from the NIST project [8]. Again, the additional modules also lack some functionality to support the implementation of Mobile WiMAX. Therefore, we tried to get the results and measurement as accurate as possible from the simulation according to the simulator's perspective.

The simulation scenario consists of three 802.16e BSs (BS0, BS1, BS2) and one MS; where the BSs are aligned on a line and the MS travels through the coverage areas of these three BSs. The BSs are located in such a way that there will be some overlap within the coverage areas of the neighboring BSs. We have selected some constant values for some of the elements such as cell size, transmission power of BSs, and the route of the MS. Our main concern was to adjust the parameters of the WiMAX module to get faster handoff times in the NS-2. We simulated the whole scenario with speeds 10 meters/second to 40 meters/second with 1 meter/second steps. The first one will occur when the MS reaches the edge of the coverage area of BS0, it then needs to change its connection access point from BS0 to BS1 to gain continuous service. Again, the second handoff will occur when the MS reaches edge of the coverage area of the BS1, then to get continuous service it needs to once again change its connection access point from BS1 to BS2.

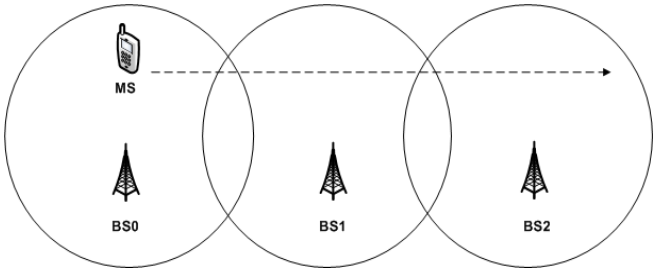


Figure 1: Simulation Scenario.

In simulation, our main goal was to find out the parameters that have the greatest impact on the performance of the handoff process in Mobile WiMAX. According to the scenario we have measured the time for the two handoffs.

Adjusted Parameters

In order to achieve less handoff time we have adjusted the list of parameters of the WiMAX module. When we changed the value of a parameter we ran the simulation and measured the handoff time for each adjustment. During the time we changed the values of some parameters from the example file of the WiMAX module that has predefined values. Then we adjusted the rest of the parameters that are offered by the MAC/ 802.16 in NS-2 and have some impact on the handoff time. Some of the parameters were set as their default value in the simulation code that we did not mentioned separately. When we completed the adjustment successfully, we found that some of the parameters did not have any impact on handoff latency, where some other parameters had a greater influence. We will discuss these parameters in brief description in the simulation result section.

Constant Parameters

As the implementation process of handoff in Mobile WiMAX is quite complex we want to keep the simulation process simpler. Therefore, we have used some constant value for some of the parameters that are related to the configuration of the BSs such as the BS coverage area, transmission power, and operating frequency. By keeping these parameters constant we could adjust some other parameters to achieve less handoff duration time for both handoffs. For each adjustment we used the same simulation scenario, where the number of BSs, location of the BSs, and the route of the MS were mentioned. The bit rate for transmitting data was also kept constant around 1.4 Mbps. From literature, we have found that to send a MPEG-1 file a data transmission bit rate of 1.5 Mbps is required [9]. Therefore, comparing this value with the used data bit rate of this simulation can be considered as a moderate bit rate. The packet size and the duration time for sending a packet were also kept constant, 1600 bytes and 15 ms respectively.

Handoff Times with Varied MS Velocity

To investigate the impact of the different velocities of the MS is one of our main concerns in this research. During the adjustment of the different parameters, we also varied the velocity of the MS. We have done the simulation by changing the speeds of the MS from 10 m/s to 40 m/s with a 1 m/s step. We chose the highest speed 40 m/s as it equals 144 km/h, which is well above to the standard limit (100 km/h) for a seamless handoff [7]. When we adjusted the parameters we varied the speed and investigated the results of the handoff times. The impacts of the different velocities of the MS on handoff times will be presented in the following Figures for both handoffs.

We have seen from Figure 2 and Figure 3 that the handoff times were varied for different MS speeds. We have found that for the MS speed of 10 m/s to 19 m/s, the handoff times varied consistently by a few milliseconds within the range of 35 ms to 45 ms, which is less than 50 ms (50 ms is the standard of the handoff latency). After 20 m/s MS speed, the handoff times increased inconsistently and went up to approximately 150 ms, this is much larger than the latency limit. But this result can differ for real time traffic because here the MS speeds were controlled by the simulator smoothly. On the other hand, due to the lack of the simulator, we could not implement some of the major features of Mobile WiMAX like QoS, authentication, and service flow. Therefore, the results can also vary for those missing features.

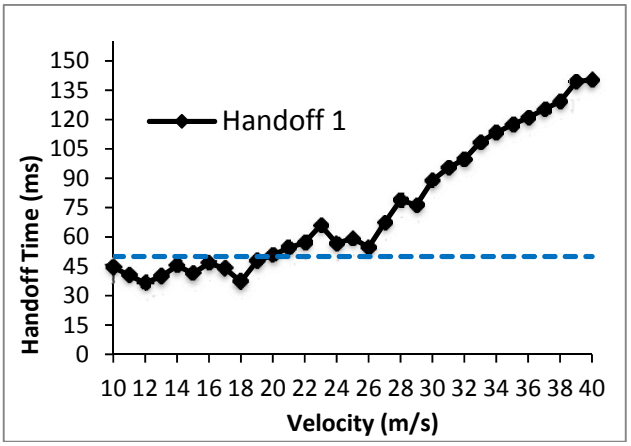


Figure 2: Simulated Handoff Times with Varied MS Speeds for First Handoff.

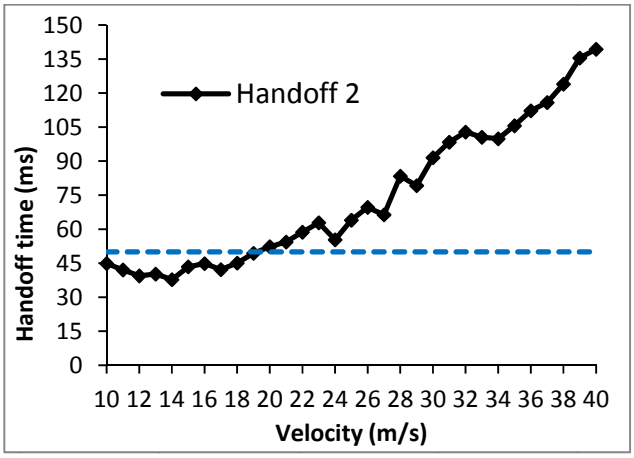


Figure 3: Simulated Handoff Times with Varied MS Speeds for Second Handoff.

IV. CONCLUSION

In this paper the performance of the handoff process in Mobile WiMAX network is investigated. Simulation was done with the NS-2 simulator. As NS-2 does not have all of the necessary components to support the implementation of Mobile WiMAX, two additional modules known as mobility and WiMAX from NIST are used in parallel with it. The parameters of the WiMAX module were adjusted and investigated to achieve shorter handoff duration time during both handoffs. After the adjustment of all the parameters, the

velocity of the MS was varied within the range from 10 m/s to 40 m/s. The main objective of this research was to find out the parameters that have the greatest impact on handoff performance as well as on handoff latency during the handoff in Mobile WiMAX. We have seen that some of the parameters have direct influence on the handoff latency where others do not.

V. REFERENCES

[1] Wikipedia, "IEEE 802.11," <url:http://en.wikipedia.org/wiki/IEEE_802.11>, retrieved December 2010.

[2] Wikipedia, "WiMAX," <url: http://en.wikipedia.org/wiki/WiMAX>, retrieved December 2010.

[3] Zhang, Y., "Vertical Hando_ between 802.11 and 802.16 Wireless Access Networks," Waterloo, Canada, pp. 3, 2008.

[4] Diab, B. W.,Tohme, S.,"End-to-End Security and Seamless Handover Solution for Real-Time Communications over 3G Networks," ACM, New York, pp.1, 2009.

[5] Wikipedia, "Handover," <url: http://en.wikipedia.org/wiki/Handover>, retrieved December 2010.

[6] Makelainen, A., "Analysis of Handoff Performance in Mobile WiMAX," pp.12-68, 2007.

[7] WiMAX Forum: Feb 2006. "Mobile WiMAX – Part I: A Technical Overview and Performance Evaluation," url: < http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part1_Overview_and_Performance.pdf> retrieved January 2011.

[8] NIST Information Technology Laboratory "Seamless and Mobility," < url: http://www.nist.gov/itl/antd/emntg/ssm_seamlessandsecure.cfm>, retrieved December 2010.

[9] ISO. "ISO/IEC 11172-1:1993 - Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 1: Systems".<url: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=19180> Retrieved 2011-01-10.

Evaluation of Video Transmission Quality of Service over Next Generation Networks

Obeten O. Ekabua
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(obeten.ekabua@nwu.ac.za)

Francis L. Lugayizi
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(francis.lugayizi@gmail.com)

Abstract: *The challenge in service quality for the performance of Next Generation Networks (NGNs) is a growing concern. The bandwidth limitation for multimedia applications in NGNs such as voice and video telephony along with the increasing number of applications on the Internet, service classification and efficient resource management have all become quite challenging tasks. In order to handle the different types of applications in the network to improve the service quality in video and voice transmission, there has to be efficient resource and traffic management, and using routing protocols is one way in which this can be done. We designed three network models that are configured with OSPF, EIGRP and one with both OPSF and EIGRP routing protocols, and then used the QoS parameters of their throughput, packet loss, convergence time, packet delay variation and end-to-end delay as our performance evaluation metrics. Our main source of network traffic was a typical video conferencing application. The results obtained during the simulation indicates that combining both EIGRP and OSPF is more reliable in providing Quality of Service than OSPF routing protocol when the main traffic used in the network is video, but when dealing with a standalone real time application network, EIGRP is better than OSPF.*

1 Introduction

With recent trends and technology advancement in the development of converged broadband next generation networks (NGNs) and advanced multimedia services, the potential has increased for delivering video services to end users "anywhere, anytime" using the World Wide Web. A wide variety of these services do exist this is mainly due to the availability of various systems to deliver the services. These systems are built on top of tools and applications that provide the necessary communications and computer-aided support (e.g., multimedia conferencing/streaming enablers, image analysis and visualization tools, immersive and collaborative virtual environments).

A Next Generation Network is an interesting innovation that mainly drives to reduce costs on the side of service providers while at the same time enhancing the capability of a given network to stay open to new services and applications. This innovation basically involves the transformation of public switched telephone networks (PSTN) which are circuit-based networks into packet-based networks that mainly depend on Internet protocol. Therefore, it is one of those innovations to change the telecommunication industry forever.

The development of NGNs has further led to yet another concept – convergence, this represents the shift from the traditional 'vertical silos' architecture i.e. a scenario where services were provided through different networks (mobile, fixed, IP) to a situation where communication services are accessed and used seamlessly across different networks and provided over different platforms in an interactive way [1]. The biggest driving force behind this has been the Internet and it has stayed liked that.

Converged NGNs deliver different types of traffic across heterogeneous end-user environments [2]. For example, video and audio streaming have special bandwidth, loss and delay requirements, in scenarios where data or a video fails to arrive in expected time, play out in a particular application may pause, this becomes annoying to the user. Therefore, in order to meet the requirements of a specific video or audio service traffic delivered over networks in conjunction with other commercial traffic, QoS mechanisms such as class-based traffic prioritization are necessary. The wide variety of video services imposes different Quality of Service (QoS) requirements on underlying networks. One aspect is delay tolerance, with service requirements ranging from strict real-time and delay-intolerant data transmission to delay-tolerant services.

2 Related work

In order to deliver multimedia across NGNs, there are various requirements that have to be met especially for specific real-time/interactive video service traffic delivered over networks in conjunction with other commercial traffic (e.g., voice calls, streaming multimedia, and Internet traffic). QoS mechanisms such as class-based traffic prioritization and effective IP routing protocols are necessary. The wide variety of multimedia and video services imposes different Quality of Service (QoS) requirements on underlying networks. One aspect is delay tolerance, with service requirements ranging from strict real-time and delay-intolerant data transmission to delay-tolerant services. In [3], [4], [5], [6], and [8] the authors categorize the importance of various QoS parameters for different interactive video services. Prioritization and resource allocation schemes for various types of video traffic delivered over various networks have been addressed in [3, 8]. Further studies have more specifically focused on evaluating support for the delivery of real time video services over high speed 3G/4G networks [9],[12, 10] and other types of broadband networks [11], with evaluation results showing generally reliable performance.

The research community has continued to carry out various investigations on new approaches, methods, strategies, techniques and tools for analyzing, designing, controlling and evaluation of future Next Generation Systems that support user interworking of multimedia applications and their mobility. Emphasis is now being given to QoS and its related aspects both in access and core networks in the presence of multimedia traffic, as [12] suggested. The real question above all is how can we evaluate QoS of service especially in an instance where it is impacted by an unexpected event? This may sound pretty an easy question to answer but that isn't the case. We do not differ so much from [12] , when he stated that "the only good reason to measure anything is to reduce uncertainty with respect to some course of action that must be decided". Hardy [12] proposed and presented a method where he analyzed QoS using two processes, mainly measurement and evaluation. In our study we focus more on evaluating the performance of two routing protocols based on selected metrics, we also evaluate these protocols but from a network routing point of view. Routing is part of network traffic engineering, the major objective of the latter is to ensure and improve network performance at the same time maintain the QoS requirements by optimizing network resources. One way this can be done is through the application of efficient IP routing protocols that will enhance the general performance of the network especially when it comes to QoS of the various multimedia applications.

3 Network Convergence versus Network Quality of Service

Due to the increase in the digital content on the Internet, more and more people, institutions and businesses are shifting their activities to the Internet through creating IP-based networks, applications and services. This is mainly because there has been an increase in access to high-speed broadband, availability of computing power and devices along with the communication media discussed later on in the work. All the above activities have been a major contribution to the utilization of the "convergence" term when discussing next generation networks.

Convergence is seen as a shift from traditional 'vertical silos' architecture i.e. a situation in which different services were provided through separate networks to a situation in which these services are accessed and used seamlessly across different networks and provided over platforms in an interactive way[4]. The different levels of convergence include Network convergence, Service convergence, Industry/market convergence, Legislative, institutional and regulatory convergence, Device convergence and Converged user experience.

Quality of Service as defined by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Recommendation is the "collective effect of service performance which determines the degree of satisfaction of a user of the service". As discussed earlier on, the next generation networks hosts a wide variety of service applications for multimedia transmission across the networks. All these applications and networks have their own desired level of service standards as far as the users of those applications and the service providers themselves are concerned. This therefore calls for a systematic way of realizing these objectives, and this is mainly achieved through putting in place mechanisms to control network traffic, allocate and manage network resources. Most researchers believe that QoS is strongly correlated to the main characteristic of a network – its heterogeneity [13].

4 Next Generation Network Challenges

Next generation networks are meant to support a wide variety of network traffic services and user mobility that is constraint free at the same time ensuring that there is a guaranteed Quality of Service for users at any given time and anywhere. However this doesn't come cheap, there are quite a number of challenges that need to be addressed. The challenges are:

4.1 Application Traffic Types

The diverse nature of network traffic comes with a recognized challenge to any network engineer in their quest to guarantee quality of service to the users of that network in terms of Quality of Experience. In multimedia traffic, there are mainly two types of network traffic - real-time/interactive and non-real time. Examples of real time/interactive traffic are video and audio while file downloading is an example of non-real time traffic. Network traffic from real time/interactive applications are very sensitive to delay, whereas those applications that are non-real time may not have strict delay requirements, although they do not accommodate errors at the same level as the real time/interactive applications. One of the major aims of any network is to sufficiently meet the quality demands of all classes of applications and therefore guarantee Quality of Service in these applications. Therefore the major requirement of a next generation network is the ability to support a wide variety of services and applications with their respective traffic features, for instance diversity in delay and requirements in bandwidth.

With the evolving diversity in network traffic types, QoS provision within next generation networks is a real challenge, thus many researchers are now apportioning more resources to this cause and this is also one of the major aims of this thesis, though the slight difference here is that this thesis focuses on real time/interactive QoS video transmission over next generation networks.

4.2 Traffic Characterization

Traffic characterization is a way of dealing with traffic flow by determining its impact on the general performance of the network. The biggest concern for network engineers here is controlling data burstiness and the network characterization itself. Data burstiness is part and parcel of the different IP Networks' patterns. Much as the area network traffic modeling and characterization continues to become a research area of concern, this thesis doesn't proceed to venture into this area.

4.3 Protocol Specific

Next Generation Networks are IP based which means that they mainly rely on core network protocols. Much as Internet protocol may guarantee more scalability than Asynchronous Transfer Mode ATM, it does come with various challenges. One of the main problems is the QoS provision [16]. The Internet Engineering Task Force (IETF) working group is further evaluating and analyzing the various ways that QoS on IP can effectively be implemented.

4.4 Costing

One of the objectives of NGNs is to minimize on communication costs but there is always going to be a challenge of higher bandwidth costs and transmission rates. This is yet another challenge that various researchers are putting their heads together to strike equilibrium of service costing without affecting the QoS and the users at large.

4.5 Network Capacity

The evolution of telecommunication networks was fuelled by the demand for higher capacity. The continued upward trend in technology especially in telecommunications has further increased the rate at which multimedia applications are being developed. All these applications have a desired amount of bandwidth to ensure efficiency in service. Therefore, Next Generation Networks need to be able to meet the bandwidth capacity of the various multimedia applications. If all these factors are to be considered (which is always the case), then network designers and programmers are faced with many challenges especially when it comes to traffic engineering and network dimensioning.

4.6 Mobility Management

Heterogeneous networks have to promote mobility. The issues of global mobility in the core IP network are being solved by the introduction of the Mobile IP protocol [17] but Mobile Internet protocols also come with their own challenges for example triangular routing, and duplication of IP fields ("IP within IP") Researches have emphasized that in mobility, the handovers from different cells should be properly handled to maintain the desired QoS [18].

4.7 Scalability

Scalability is the degree of adaptability to a particular force of demand, thus the challenging question network engineers have to ask themselves is - will the network be able to support an increase in the number of users at any given one time? This is entirely because the number of people using telecommunications has continued to increase thus the NGN should be able to meet the demand without affecting the Quality of Service.

4.8 Heterogeneous network Compatibility

A reliable next generation network should be compatible with different types of networks currently in use and must have provision for future developments. Diversity in a network comes with key issues that shouldn't miss mentioning, and these include: Mobility management within applications of these networks, compatibility with other networks, secure interworking within access networks. End-to-end QoS of service guarantee to users of these networks. Various research groups are addressing the compatibility issues for the networks to work together [17].

5 Evaluation of Video Quality of Service

One way of guaranteeing effective video quality of service in next generation networks is through network traffic routing. Routing is a process where routers in a network specify paths that the various packets should follow during transmission across the Internet [18]. Routers and Routing protocols are at the center of this process; the latter are the various rules that specify how the routers are going to communicate with each other by disseminating data. There are quite a number of routing protocols widely in use in the telecommunication industry, this work concentrates mainly on Open Shortest Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP) as one of the ways of evaluating video transmission QoS in Next Generation Networks; this is mainly because they are the pre-eminently considered protocols for video transmission in a network.

5.1 Protocols under focus

The protocols presented in this paper are OSPF and EIGRP, These are the protocols we used in our experiments (section 8), and they are evaluated based on the quantitative metrics of convergence time, packet delay variation, end to end delay, traffic sent, traffic received and packet loss. These metrics are the ones we used to further evaluate the performance of 3 different network environments from which results and conclusions were drawn.

I. Open Shortest Path First (OSPF)

OSPF is an Interior Gateway Protocol (IGP) which is one of the main protocols used in the Internet Protocol (IP)-based Internetworks. The routing protocol is a public (open standard) that is based on the link state. The various concepts and operations of the OSPF link state are fully described in Request for Comments (RFC) 1583.

II. Enhanced Interior Gateway Routing Protocol (EIGRP)

EIGRP is an interior gateway protocol developed by Cisco Systems and introduced with Software Release 9.21 and Cisco Internetworking Operating System (Cisco IOS) Software Release 10.0. It is a suitable protocol for a variety of network topologies and multimedia. As we shall later on find out in the simulation stage, EIGRP is one protocol that has quick convergence time. The protocol is considered to be one that lets routers exchange information more efficiently than other network protocols.

5.2 Experiment set up for Network Topology

Figure 1 is a depiction of the 4 subnets (Cape Town, Durban, Johannesburg and Mafikeng). This particular figure was duplicated thrice in order to implement the three network scenarios with different routing protocol configuration. The network topology was made up of different devices and different configuration utilities.

These included:

I. The four subnets, each subnet with a video Ethernet server, LAN with Ethernet workstations and a CS_7000 Cisco router;

II. The routers are connected with the PPP_DS3 Duplex Links;

III. Ethernet 10 Base T Duplex Links that connect the various Ethernet workstations in the LANS together;

IV. Failure Recovery Configuration Utility;

V. Application Configuration Utility;

VI. QoS Attribute Configuration and a Profile Configuration Utility.

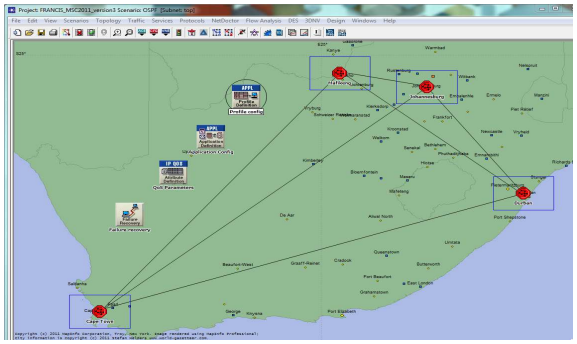


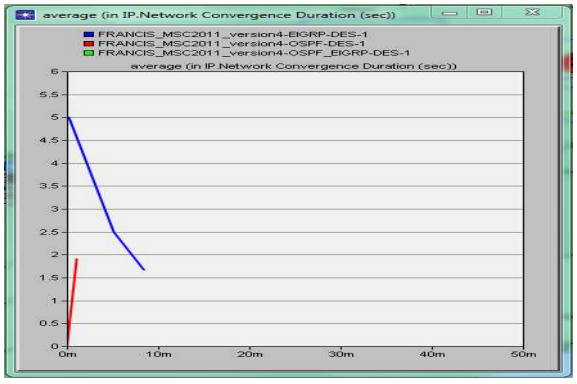
Figure 1: Network Topology with four subnets

5.3 Evaluation and experiments results

During the experiment conducted, we set up and simulated 3 different network scenarios i.e. OSPF scenario, EIGRP scenario and in the third scenario both OSPF and EIGRP are combined. A video conferencing application was configured to generate the main traffic of the network. The sections that follow present the results of the experiment.

5.3.1 Convergence time

Convergence is the time it takes all the various routers in a network to share specific information, of which this time should always be at a minimal.



Scenario Name	End to End Delay (ms)
OSPF_EIGRP	0.43171
OSPF	0.43194
EIGRP	0.63589

Figure 2: Convergence Duration

Analyzing figure 2, we observed that during convergence, the EIGRP network is faster than OSPF and OSPF_EIGRP networks. This was mainly because of the high rate at which EIGRP detected the rapid changes within the network and further communicated the changes with other neighboring routers until all the routers in the network were updated. Between OSPF and OSPF_EIGRP networks, OSPF_EIGRP network was slower than OSPF hence the same reason why OSPF had not yet reflected on the graph in figure 8-1. Table 1 depicts the convergence times of the networks.

Scenario No.	Scenario Name	Routing Protocol	Convergence Time(sec)
1st	EIGRP	EIGRP	0.5473
2nd	OSPF	OSPF	3.7552
3rd	OSPF_EIGRP	OSPF_EIGRP	5.0001

Table 1: Average value of convergence time

5.3.2 End to End delay (video conferencing)

End-to-end delay is the elapsed time for a packet to be passed from the sender through the network to the receiver and that the higher the delay between the sender and receiver, the more insensitive the feedback loop becomes, and therefore, the protocol becomes less sensitive to short term dynamic changes in the network.

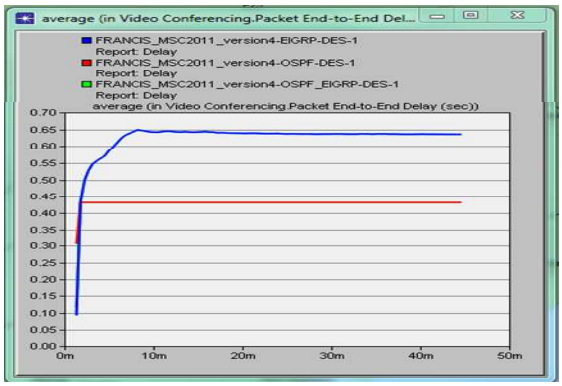


Figure 3: End to End delay

So, looking at figure 3 along with Table 2, the OSPF_EIGRP network had a lesser end-to-end delay compared to OSPF and EIGRP and it was network congestion that brought about this result. Basically end-to-end delay mainly depends on the speed of the network and the degree of network congestion. Table 8-3 below shows the average values of end-to-end delay of the different networks.

Table 2 Average of end to end delay

5.3.3 Packet Delay variation (video conferencing)

Packet delay variation refers to the difference in end-to-end delay of the packets. This variation sometimes brings about an effect known as jitter. This variation is measured by taking the difference in delay of the packets. Figure 4 illustrates video conferencing packet delay variation.

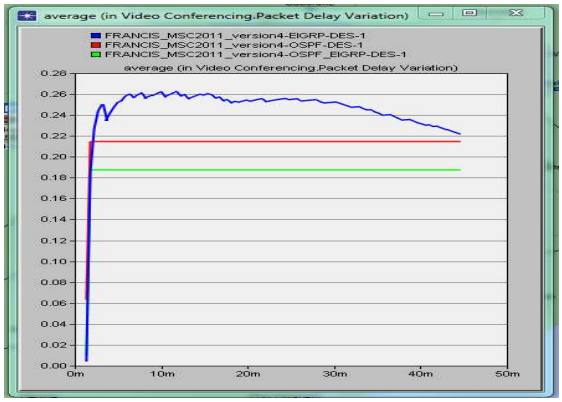


Figure 4: Video Conferencing Packet Delay variations

Table 3 below shows the results of packet delay variations of the different networks in our simulations. It was observed that the OSPF_EIGRP network records a less packet delay variation than OSPF and EIGRP networks; that is why OSPF and EIGRP individually had a high packet delay variation.

Scenario Name	Packet Delay variation (ms)
OSPF_EIGRP	0.18749
OSPF	0.21466
EIGRP	0.22186

Table 3 Packet Delay Variations (ms)

5.3.4 Video Conferencing Traffic Sent

A video conferencing application was our main source of traffic in our simulation experiments. It generated the traffic throughout the various scenarios. The video resolution of this application was set at a high resolution with 15 frames/secs as the frame inter - arrival time. The frame size of 128x240 pixels was used and the Type of service set at Best effort. The users of the network in the various areas had video streaming servers from which they were accessing the video from. Figure 8-4 illustrates the total traffic sent from the video conference application.

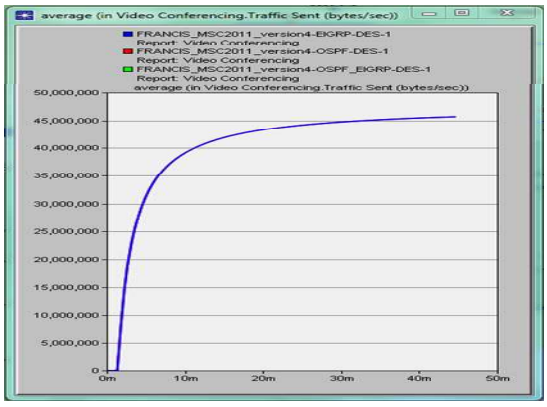


Figure 5: Video Traffic Sent

The Figure also shows a total traffic of approximately 45688129 bytes per seconds through the network.

5.3.4 Video Conferencing Traffic Received

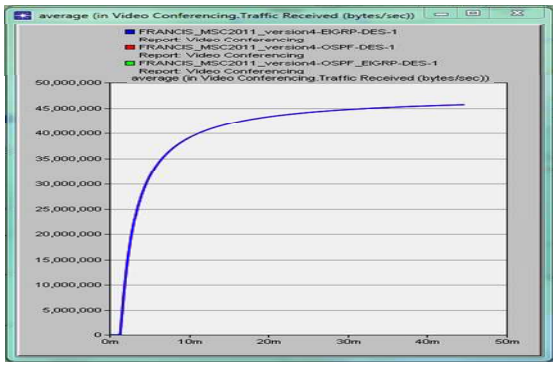


Figure 6: Video Traffic Received

The figure above illustrates a small percentage drop of packets because of the network congestion. Table 4 shows that EIGRP experiences the least packet loss in comparison with OSPF_EIGRP and OSPF networks that experience almost the same amount of packet loss.

Scenario Name	Sent(bytes/sec)	Received (bytes/sec)	Packet Loss
EIGRP_OSPF	2333436	233379	0.02%
EIGRP	2333436	233405	0.01%
OSPF	2333436	233265	0.07%

Table 4: Average value of sent and received (bytes/sec) for video.

6 Conclusion and Future work

In our study, it can be concluded that EIGRP has a much faster convergence time than OSPF and

OSPF_EIRGP networks and this was because EIGRP is a fast protocol when it comes to accessing a network's topology information updates compared to the other protocols which tend to struggle.

The results generally confirmed that combining both EIGRP and OSPF together especially in heterogeneous networks is bound to assure QoS in these networks since they host a multitude of time sensitive applications, but if a network developer chooses to focus on a single real time application network, then EIGRP is better routing protocol than OSPF in terms of guaranteeing the desired QoS of a network.

In future, we intend to explore and find out whether these protocols are bound to produce the same results in a different network environment such as a network based mainly on IPv6 due to the fact that this study was based on a network with IPv4 environment.

References

[1]. J. Sen, M. Sayyad and B. Hooli. Convergence and next generation networks. *Arxiv Preprint arXiv: 1012.2524* 2010.

[2]. P. Molinero-Fernández, N. McKeown and H. Zhang. Is IP going to take over the world (of communications)? *ACM SIGCOMM Computer Communication Review 33(1),* pp. 113-118. 2003.

[3]. M. Kalman, E. Steinbach and B. Girod. Adaptive media playout for low-delay video streaming over error-prone channels. *Circuits and Systems for Video Technology, IEEE Transactions on 14(6),* pp. 841-851. 2004.

[4]. J. Sen, M. Sayyad and B. Hooli. Convergence and next generation networks. *Arxiv Preprint arXiv: 1012.2524* 2010.

[5]. V. Raeisaenen and J. Wiley. *Implementing Service Quality in IP Networks* 2003.

[6]. C. Semeria. Supporting differentiated service classes: Queue scheduling disciplines. *Juniper Networks* 2001.

[7]. G. K. Wallace. The JPEG still picture compression standard. *Commun ACM 34(4),* pp. 30-44. 1991.

[8]. V. Bhaskaran and K. Konstantinides. *Image and Video Compression Standards: Algorithms and Architectures* 1997.

[9]. D. Wu, Y. T. Hou, W. Zhu, Y. Q. Zhang and J. M. Peha. Streaming video over the internet: Approaches and directions. *Circuits and Systems for Video Technology, IEEE Transactions on 11(3),* pp. 282-300. 2001.

[10]. H. Kanakia, P. P. Mishra and A. Reibman. An adaptive congestion control scheme for real-time packet video transport. Presented at ACM SIGCOMM Computer Communication Review. 2010,

[11]. L. Kleinrock. The latency/bandwidth tradeoff in gigabit networks. *Communications Magazine, IEEE 30(4),* pp. 36-40. 1992.

[12]. W. C. Hardy and J. Wiley. *QoS: Measurement and Evaluation of Telecommunications Quality of Service* 2001.

[13]. D. Goderis, S. Van den Bosch, Y. T'Joens, P. Georgatsos, D. Griffin, G. Pavlou, P. Trimintzios, G. Memenios, E. Mykoniati and C. Jacquenet. A service-centric IP quality of service architecture for next generation networks. Presented at Network Operations and Management Symposium, 2002. NOMS 2002. 2002 IEEE/IFIP, Florence, Italy. 2002,

[14]. J. Bannister, P. Mather and S. Coope. *Convergence Technologies for 3G Networks: IP, UMTS, EGPRS and ATM* 2004.

[15]. H. Zhou and Z. Zhang. Examining QoS guarantees for real-time CBR services in broadband wireless access networks. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2011* pp. 25-40. 2011.

[16] T. Walingo and F. Takawira. TCP over wireless with differentiated services. *Vehicular Technology, IEEE Transactions on 53(6),* pp. 1914-1926. 2004.

[17]. C. E. Perkins, S. R. Alpert and B. Woolf. *Mobile IP; Design Principles and Practices* 1997.

[18]. E. Crawley, R. Nair, B. Rajagopalan and H. Sandick. A framework for QoS-based routing in the internet. Presented at RFC. 1998,

SESSION

WIRELESS NETWORKS + ENERGY EFFICIENCY ISSUES

Chair(s)

TBA

A Fundamental Modeling of Wireless Mesh Network for Emergency Communications

Satoshi Takahashi
Graduate School of Systems and Information Engineering,
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, Japan 305-8573
Email: takahashi2007@e-activity.org

Tokuro Matsuo
Graduate School of Science and Engineering,
Yamagata University
4-3-16 Johnan, Yonezawa, Yamagata, Japan 992-8510
Email: matsuo@tokuro.net

Abstract—**This paper studies a wireless mesh network for information infrastructures in the huge disasters. It is very important to ensure the communication infrastructures when the huge disasters are occurred. We consider the special wireless mesh device, which it has solar panels to charge the electronic power. We model a simple situation of the wireless mesh network to achieve the new infrastructure, also we formulate some optimization problems to evaluate the wireless mesh networks.**

keywords : Infrastructure for Disaster, Wireless Mesh Network, Optimization Problem.

I. Introduction

The local governments have been taking countermeasures against natural disaster since the Great East Japan Earthquake occurred in 2011. It is a prime task that the local government prepares the infrastructures such as life lines (electronic and water), transportation infrastructures (railways and highways) and communication infrastructures (telecommunication and Internet). Particularly, securement of communication is discussed actively. When the natural disaster was occurred, many public agencies (e.g. the local government) communicate each other for gathering and sending information. However physical communication infrastructures does not use to be emergency communication infrastructure, since these infrastructures occurs disconnecting and loss of power. The wireless mesh network is thus focused on as an robust infrastructure[1], [2], [3]. These researches focus on decreasing throughput caused by going through some access point and proposes algorithms for solving the problem.

This paper discusses a situation of planning and creating the new infrastructure, which is getting electronic power and communication autonomy. Also We discuss some optimization problems for evaluation the planned infrastructure. It is important to plan the wireless mesh network as one of emergency infrastructures in the situation that the infrastructures does not get enough electronic power.

The rest of this paper consists of the following five parts. In Section 2, we shows detail of the infrastructure used by our research. Section 3 describes the wireless mesh network model and terms. Section 4 shows some optimization problems considering in the model, and discuss about the optimization problems. In Section 5 we conclude this paper and show the future work.

II. Infrastructure for the disaster

We describe the infrastructure device of emergency networks for the disaster. The goal is achievement of the robust network like Fig. 1. In this figure, there are two types of access point devices in the network. One is for business operator use access point, the other is for home or small group use access point. The network combines these access point to create the communication network. Each access point equips the self power generation system with the solar panel for countermeasure against loss of power by the natural disaster. Also the

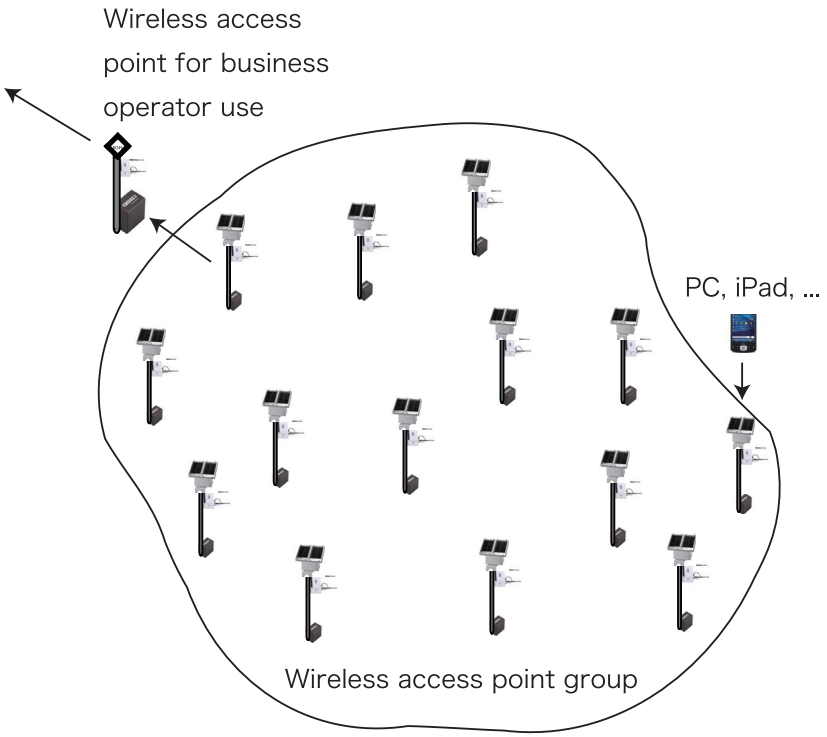


Fig. 1. Concept of infrastructure for disaster

access point equips some communication device like a smartphone or PDA for using emergency communication. The local government can gather many information from the network, since the end-user can access to the network by the access point.

Fig. 2 shows details of the access point. The access point consists of the solar panel, WiFi device, a battery and communication device. The reason of that we use two distinct type of access points is: if there is only small access point then we should have many access points to cover the whole of region, if there is only large access point then the creating cost is the problem since the device is very expensive.

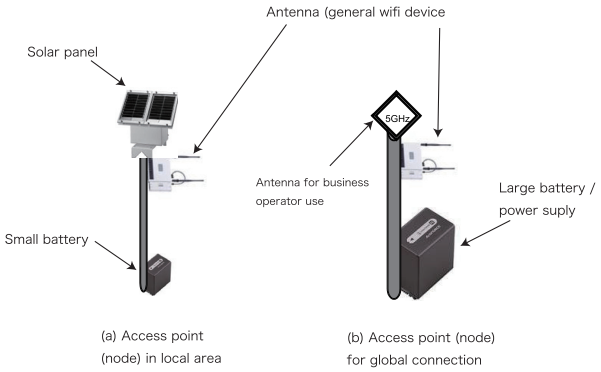


Fig. 2. Access point device

III. Preliminaries and Model

This section describes abstract of the wireless mesh network and definitions of terms for explore the model. Also we describe a fundamental communication model using by the wireless mesh network. In this research, we say that the communication device is a pair of the communication device and access point with solar panel. Let $V = \{v_1, ..., v_n\}$ be a set of communication devices. Each communication device v_i has a capacity of battery $c_i \in \mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$, a radius of communication $r_i \in \mathbb{R}_+$ and a location information (x_i, y_i). Fig. 4 shows concept of radius of communication. Also the efficiency of power production of solar panel is denoted by $e_i(t) \in \mathbb{R}_+$ with time t. The wireless mesh network is able to represent as the directed graph, denoted by $D = (V, A)$. Let $A = \{(v_i, v_j) \mid v_i, v_j \in V, v_j \in N(r_i)\}$ be a set of arcs of the directed graph D. $N(r_i)$ shows a set of neighborhood, denoted by $N(r_i) = \{v_j \in V \mid \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r_i\}$. The wireless mesh

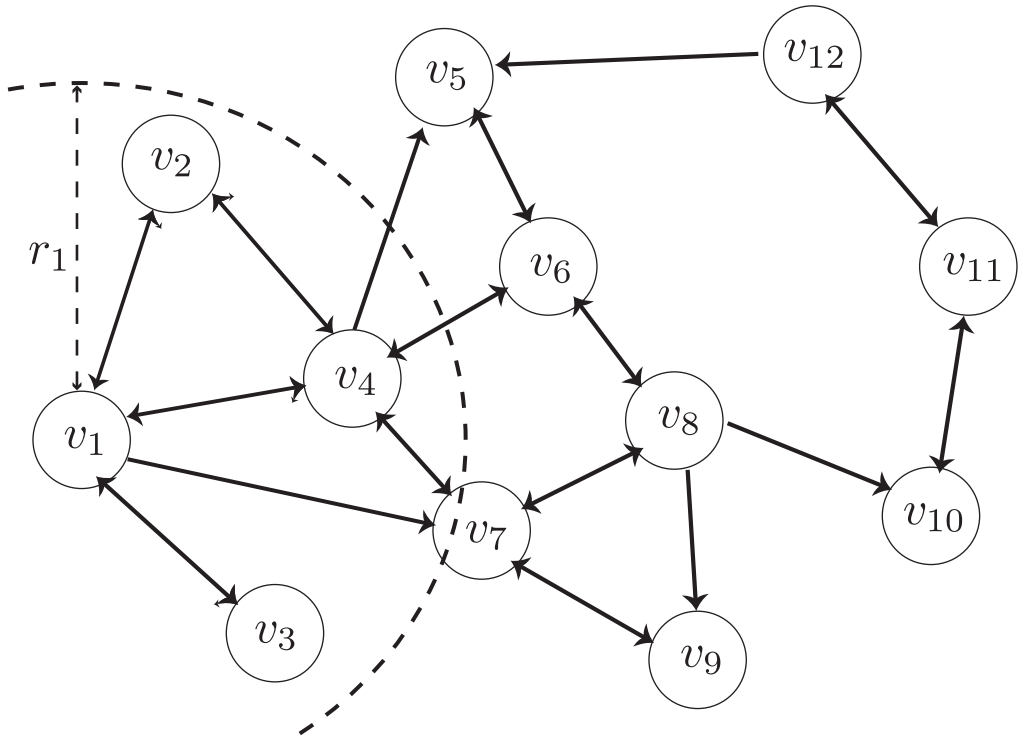


Fig. 3. Example of wireless mesh network

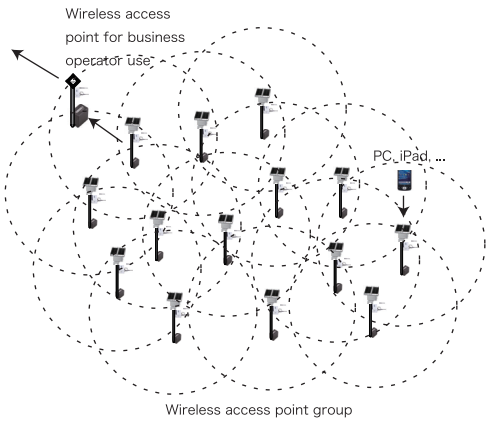


Fig. 4. Radius of communication

network can be denoted by $(D, \boldsymbol{c}, \boldsymbol{r})$.

The communication between each device is given a tuple of start vertex $s \in V$, terminal vertex $t \in V$ and communication traffic q_{st} as (s, t, q_{st}). To simplify the discussion, the communication traffic is defined by $q_{st} \in \mathbb{R}_+$. When the network communicate between s and t, the problem is a path of communication. If $t \in N(r_s)$, then the network can direct communication between s and t. Otherwise, $t \notin N(r_s)$, the communication has to go through some vertices. In this case the communication in a s-t path P_{st}. The communication spend the electronic power based on the communication traffic. When the network communicate between s and t, the power consumption $b_v(q_{st})$ is occurred in each vertex $v \in V[P_{st}]$ of P_{st}. Fig. 3 shows an example of the wireless mesh network. The constitution is as follows. we connect to some vertices in a radius of communication r_1 of the vertex v_1. Other vertices is operated as the same. Now, we consider the communication between v_1 to v_6 with traffic q_{16}. There are some v_1-v_6 paths, however, we decide a path $P_{16} = v_1, v_7, v_8, v_6$. Then the battery of each vertex in P_{16} is decreasing by $b_v(q_{16})$ for all $v \in V[P_{16}]$.

IV. Some Problems

This section formulates some problems occurred in the wireless mesh network for countermeasure against the disaster and discuss about it.

A. Network Flow Problems

To consist of the communication infrastructure, the problem is how robustly is the network in communication. We thus discuss how large traffic

do the network flow in. To know the flow in the network, we reduce this problem to the maximum flow problem. The maximum flow problem is to solve the maximum value of s-t flow, given the arc capacity, a start and a terminal vertex. Also the problem is one of primal combinatorial optimization problems[6]. Now we focus on only one pair of start and terminal vertex and formulate the problem based on above model. Considering network is $D = (V, A)$, where $A = \{(p_i, p_j) \mid p_i, p_j \in P, p_j \in N(r_i)\}$. We compute an arc capacity as follows. Suppose that $\partial^+ : A \to V$ is a map of tail of the arc and $\partial^- : A \to V$ is a map of head of the arc, the capacity $d : A \to \mathbb{R}_+$ of the arc $a \in A$ is defined by

$$d(a) = \min\{\lfloor c_{\partial^+(a)}/b_{\partial^+(a)}(1)\rfloor, \lfloor c_{\partial^-(a)}/b_{\partial^-(a)}(1)\rfloor\}.$$

Then $\lfloor c_{\partial^+(a)}/b_{\partial^+(a)}(1)\rfloor$, $\lfloor c_{\partial^-(a)}/b_{\partial^-(a)}(1)\rfloor$ represents how number of the communications with respect to the vertex $\partial^+(a)$ and $\partial^-(a)$. Also we denote the flow of the arc as $f : A \to \mathbb{Z}_+ = \{x \in \mathbb{Z} \mid x \geq 0\}$, then the maximum flow problem finds the flow such that to maximize $\sum_{a\in\delta^+(s)} f(a)$ with the capacity constraint $0 \leq f(a) \leq d(a),\ \forall a \in A$ and the flow conservation law

$$\sum_{a\in\delta^+(v)} f(a) - \sum_{a\in\delta^-(v)} f(a) = 0,\ \forall v \in V \setminus \{s, t\},$$

where $\delta^+(v) = \{a \in A \mid \partial^+(a) = v\}$ and $\delta^-(v) = \{a \in A \mid \partial^-(a) = v\}$.

To estimate the maximum s-t traffic, this is the criteria of decision making of how does plan the sending and gathering information. At the maximum flow problem with single pair of vertex, if each arc's capacity is integer, then there exists a maximum integer flow[4]. To use this property, we estimate the number of constant communication traffic. Also, the maximum flow has some polynomial time algorithms[5], [6], [4], it is easy to get the criteria of the network planning. The above discussion fix the pair of communication vertex, generally, the communication consists of multiple pairs of vertices. Also these pairs communicate different information each other. We thus consider the multi commodity flow problem. The multi commodity flow problem is that for some pairs of (s_i, t_j), the network flows the commodity flow with maximizing sum of the commodity flows in the network. For the multi commodity flow problem, there are some theoretical researches about complexity and properties of the solution[8], [9]. However the other research does not broom since the problem is hard to solve although this problem is similarly practical problem[7].

B. Path Packing Problem

Path packing problems are considered as the problem which estimates traffic capacity of the network. Path packing problems compute the number of embedded s-t paths on the network, given a network, vertex capacity (arc capacity) and start and terminal vertex, satisfies capacity constraint. The solution of the problem is one of criterion of not only traffic capacity but also network reliability, since the solution gives concrete communication paths. Suppose that $\mathcal{P}_{st}$ is a set of a s-t paths and $X_p \geq 0$ is a decision valuable representing the number of embedded path. Then the path packing problem is formulated as follows,

$$\begin{array}{ll} \max & \sum_{p\in\mathcal{P}_{st}} X_p \\ \text{s. t.} & \sum_{p\in\mathcal{P}_{st}} X_p e_v(q_{st}) \leq b_v,\ v \in p : p \in \mathcal{P}_{st}, \end{array}$$

where a communication traffic q_{st} of s-t pair is constant. The vertex capacitated path packing problem is known to be NP-hard[10], and [11] proposes effective approximation algorithm for in-tree packing problem which is upper class compared with path packing. We can consider the path packing problem with multiple pairs. In this case, the problem is more hard and the solution obtained by something approximation algorithms does not give good criteria.

C. Facility Location Problem

Next, we consider the access point location problem. This is the problem that given the number of access points and locations of communication devices, the goal maximizes the number of devices covered by choice access points. Let S be a set of access points candidates and $f : S \to \mathbb{R}_+$ be a function gives the number of covered devices. Then the access point location problem is formulated as follows given a number of access point u.

$$\max f(T) \text{ s. t.} |T| \leq u,\ T \subseteq S.$$

This formulation is an optimization problem of which the objective function is set function and the constraint is an uniform matroid constraint that is

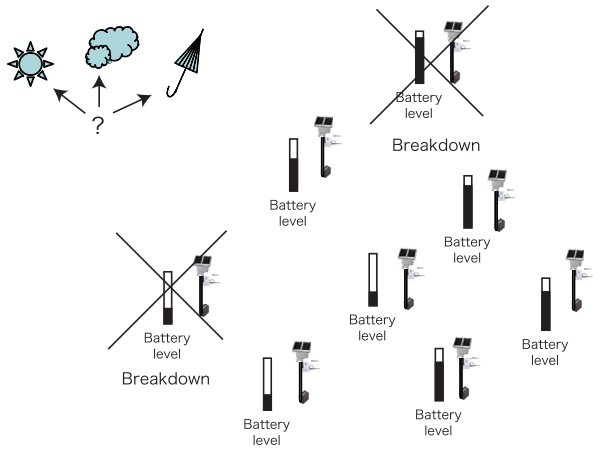


Fig. 5. Example of dynamic situation

the cardinality of element of the matroid is at most u. The optimization problem with matroid constraint has some theoretical researches, e.g. [7] and [12], [13].

D. Dynamic Conditions

The above optimization problems which compute the criterion for planning the network are the all static condition. However the realistic situation should consider the dynamical changing of battery usage by the solar panel and the time-line. It is difficult to formulate the dynamic situation as optimization problem. We thus plan the dynamic simulation for estimating the criterion of the network. The dynamic simulation is used by escape simulation in the natural disaster, and complex environment simulations. However since the simulation model widely changes by the planning network, we should classify the simulation conditions precisely. Fig. 5 show the example of dynamic condition.

V. Conclusion

This paper discusses the fundamental model and the optimization problems in the wireless mesh network with the solar power generation system as the emergency infrastructure. Considering the wireless mesh networks, we evaluate the networks by the criterion obtained by the situations and objectives. Also we should consider not only the static situation but also the dynamic situation when we prepare the wireless mesh network. The future work is that we analyze each optimization problem classifying the situations and environment.

References

[1] M. NOZAKI, K. GYODA, and M. KAWAI. Dynamic Segmentation Schemes for a Wireless Ad-Hoc Community Network, *Proceedings of IEEE MDMC'98(The 3rd International Symposium on Multi-Dimensional Mobile Communications)*, pp.202-206, 1998.

[2] H. Tsubouchi, Y. Fukuda, T. Ikenaga and Y. Oie. Performance Evaluation of Disjoint Path Routing for Multi-interface Multi-channel Wireless Mesh Network, *Proceedings of IEEE/IFIP WONS 2008 The Fifth Annual Conference on Wireless On demand Network Systems and Services*, pp.117-122, 2008.

[3] R. Baumann, S. Heimlicher, V. Lenders and M. May. Routing Packets into Wireless Mesh Networks, *Proceedings of IEEE Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'07)*, 2007.

[4] R. K. Afuja, and T. L. Magnanti, and J. B. Orlin, *Network Flows : Theory, Algorithms, and Applications*. Prentice Hall; United States ed, 1993.

[5] N. Katoh, and T. Ibaraki, Resource Allocation Problems. Handbook of Combinatorial Optimization (Vol. 2), D. Z. Du, and P. M. Pardalos (Eds), pp. 159–260, 1998.

[6] B. H. Korte, and J. Vygen. *Combinatorial optimization: theory and algorithms*, Springer, (2004).

[7] A. Schrijver, Combinatorial Optimization: Polyhedra and Efficiency (Algorithms and Combinatorics) . Springer varlag , 2003.

[8] H. Hirai. Half-integrality of node-capacitated multiflows and tree-shaped facility locations on trees, *RIMS-Preprint*, 1687, 2010.

[9] H. Hirai. The maximum multiflow problems with bounded fractionality, *Proceedings of The 42th ACM Symposium on Theory of Computing (STOC 2010)*, 2010.

[10] Y. Tanaka, S. Imahori, M. Sasaki and M. Yagiura. An LP-based heuristic algorithm for the node capacitated in-tree packing problem. *Computer & Operations Research*, Vol. 39, pp.637-646, 2012.

[11] Y. Tanaka, S. Imahori and M. Yagiura. Lagrangian-based Column Generation for The Node Capacitated In-Tree Packing Problem, *Journal of The Operations Research Society of Japan*, Vol. 54, No. 4, pp. 219–236, 2011.

[12] S. Iwata, A Fully Combinatorial Algorithm for Submodular Function Minimization. Journal of Combinatorial Theory, Series B, Vol. 84, pp. 203–212, 2002.

[13] S. Iwata, and J. B. Orlin, A Simple Combinatorial Algorithm for Submodular Function Minimization. Proc. of the twentieth annual ACM-SIAM symposium on Discrete Mathematics, pp. 1230–1237, 2009.

Energy-awareness Metrics for Multihop Wireless User-centric Routing

Antonio Junior[1], **Rute Sofia**[1], **and Antonio Costa**[2]
[1]SITI, University Lusofona, Lisbon, Portugal
[2]ALGORITMI, University of Minho, Braga, Portugal

Abstract— *This paper proposes and validates energy-awareness node-based ranking and energy-awareness successor-based ranking routing metrics focused on improving energy efficiency of multihop approaches in heterogeneous wireless environments. The validation is carried out through discrete event simulations based on real data set traces and controlled random topologies for the specific case of AODV.*

Keywords: Multihop routing; energy-efficiency; wireless user-centric networks.

1. Introduction

The recent advances in wireless technologies a such as *Wireless Fidelity (Wi-Fi)* is assisting the rise of *User-Centric Networking (UCNs)* architectures, i.e., architectures where nodes are often carried by regular Internet end-users [1], [2]. Examples of such environments can be a network formed on-the-fly after a disaster of some nature or even a municipality network where some nodes are based on end-user devices (through Internet access sharing).

Albeit being often spontaneously deployed, user-centric wireless environments rely on traditional multihop routing approaches. Multihop routing has been extensively analyzed and optimized in terms of resource management, but in terms of energy-efficiency, there is a lack of a thorough analysis. In other words, multihop routing is shortest-path based, but the metrics applied today normally relate to *Quality of Service (QoS)* aspects, or to hop count. In terms of energy-awareness, these protocols are lagging behind.

We highlight that there is considerable related work in the fields of energy-efficiency and energy-awareness for sensor networks. However, these are environments where nodes are homogeneous in terms of energy capability, and the architectures are often static. In contrast, in user-centric networks, nodes are expected to be heterogeneous in terms of energy resources, and the topology exhibits high variability as nodes tend to disappear and appear in the network, based on their carriers interests and behavior.

In previous work [3], [4] we have discussed the potential of current energy-aware routing approaches for wireless networks, and whether or not they may make sense when applied to routing in user-centric environments. We have also proposed concepts that could assist in making multihop routing more efficient in terms of energy-awareness that consider heterogeneous devices, without necessarily having to change operational aspects of the underlying algorithms, or protocols. Following such line of thought, this paper proposes and validates two routing metrics to improve network lifetime based on current multihop approaches. The first metric is based on a single (source node) energy-awareness perspective, while the second is based on the perspective introduced by the source node and potential successors. What we want to analyze is up to which point can a combination of the perspective of both a source and successor node improve network lifetime.

We evaluate the proposed metrics through discrete event simulations based on realistic assumptions based on the *Ad-Hoc On-demand Distance Vector Protocol (AODV)* [5], being the goal to analyze whether or not this approach can improve the overall network lifetime, without incurring a significant penalty.

The rest of this paper is organized as follows. Section 2 describes related work focused on multihop energy-efficiency. Section 3 goes over the discussion on energy awareness with single node vs. the two node association perspective in multihop routing. Section 4 describes our proposed metrics. In section 5, we present the performance evaluation while section 6 presents the evaluation results. Conclusions and future work are presented in section 7.

2. Related Work

A few approaches [6], [7] have surveyed multihop proposals focused on energy-efficiency, considering both the energy spent when nodes are engaged in active communication or passive communication, i.e., in idle mode. Such work has as underlying scenarios homogeneous environments, and several proposals combine different energy-aware metrics to maximize the network lifetime.

Relevant proposals [8], [9], [10] making multihop routing adaptive, have explored new metrics having in mind different types of optimization, e.g., reduction of energy spent across a path, considering the residual energy capacity of a node or avoiding nodes with low residual energy, on the global network.

Another relevant overview [11] has been provided by C. K. Toh, who discusses different routing properties to

consider in multihop routing. One of them is efficient utilization of battery capacity. In this work, the author also addresses the performance of power efficiency in ad-hoc mobile networks by analyzing four approaches which have as common goal to select an optimal path, being the optimum the minimization of the total power required on the network (across all nodes) and also the maximization of the lifetime of all nodes in the network.

The *Energy-efficient Unified Routing (EURo)* [12] develop an routing scheme that accommodates any combination of transmission power, interference and residual energy to optimize the energy efficiency of multihop wireless networks. The E^2R routing protocol [13] uses an opportunistic forwarding scheme to deliver control messages and data packets in a multihop wireless network to energy efficiency in multihop green wireless networks. Unlike other opportunistic routing protocols, it neither uses pre-selected static paths nor does it prepare forwarding candidates.

The authors of [14] outline some steps towards the definition of energy efficiency metrics for designing energy-aware wireless network. The approach is to estimate the optimal message size in terms of power consumption, to estimate the average amount of energy spent to transmit one bit and the relationship between traffic and power consumption. However, an energy-aware metric for multihop routing is not defined.

The *Maximum Residual Packet Capacity (MRPC)* protocol [15] comprises a node perspective parameter (battery power of the node) and a link perspective parameter (packet transmission energy in a link) across the link between nodes. MRPC identifies the capacity of a node not just by its residual battery energy, but also by the expected energy spent in reliably forwarding a packet over a specific link. However, such formulation is more adequate to capture scenarios where the link transmission cost depends on the physical distance between nodes and on the link error rates. Hence, the approach does not consider a energy-awareness as a primary resource of the network.

Zhang et. al. [16] combine node lifetime and link lifetime between every two adjacent nodes to select an optimal path. This route lifetime prediction algorithm explores the dynamic nature of mobile nodes in a route discovery period for predicting the lifetime of routes discovered, and then select the longest lifetime route for persistent data forwarding when making a route decision. The dynamic nature of mobile nodes mentioned, make use of the energy drain rate of nodes and the relative mobility estimation rate of adjacent nodes. However, this work does not consider the energy-awareness as a prime metric for the wireless link between two nodes, i.e. successor-based metric. Our work is in line with this, as we also attempt to explore combinations between a source and a successor node, to increase the network lifetime.

Finally, we highlight that *The Internet Engineering Task Force (IETF)* Working Group *Routing Over Low Power and Lossy Networks (ROLL)* is currently discussing multihop metrics tailored to energy-efficiency [17].

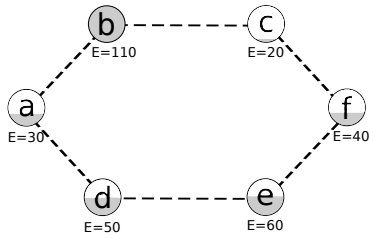


Fig. 1: Approaches of energy-aware metrics.

3. Energy Awareness in Multihop Routing, Single Node vs. the Two Node Association Perspective

In previous work [3], [4], we have considered energy-awareness based on a single node perspective. In other words, when building a shortest-path we take into consideration a node "energy-aware" cost.

During our research, we have realized that despite such approach providing a significant improvement, there are cases that could benefit, at a first glance, of a cost derived of the energy-awareness of two nodes, the father and the successor on different paths, in particular considering multihop on-demand approaches, where the information (node cost) will be provided from source to destination, but the path itself will only be available based on responses from the nodes involved. One of such cases is illustrated in Figure 1, where we consider two possible paths between source node a and destination f. A example of the energy-awareness cost provided by our metric is represented by E. As there are two potential paths, let us assume that based on shortest-path computation and on our metric, the best path would be the one providing a global higher level of E (the higher sum). Hence, the path $a-b-c-f$ (total cost of 200 units) would be preferred over path $a-d-e-f$ (total cost of 180). However, looking at the global topology we can see that path $a-b-c-f$ has an energy bottleneck (node c) and depending on traffic flowing, choosing this path may result in nodes having to recompute the path between a and f. If node b (as father node) had a perception that its neighbor node c (as successor node) would soon no longer be available, it could announce a lower value for its weight. Node c would do the same, and hence there seems to be some probability for path $a-d-e-f$ to be chosen over $a-b-c-f$. This is a simple example that considering the energy cost derived of the energy-awareness of an association between two nodes is advantageous, since it may give a perspective on the duration of a link between the nodes, based on the nodes energy capacity and drain rates.

The situations mentioned, of fluctuations in terms of energy-awareness concerning path computation in multipath

environments are expected to be more serious in UCNs, as this networks are expected to exhibit more variability in terms of node movement. User-centric routing has therefore to take into consideration some aspects which are intrinsic to the way that humans move and establish social contacts. Thus, the relation between a father and a successor node seems to be relevant to be considered also from an energy-awareness perspective. We highlight that there is a different between energy-awareness from a link perspective and energy-awareness derived from the association of two specific nodes. The former relates to the link quality itself; while the latter is simply a composition of the energy-awareness level of two associated nodes.

Hence, we address in the next section our proposed heuristics, which are based on the association of father and successor node.

4. Our Proposal

This section provides an overview on our heuristics. In previous work we have proposed and validated an energy-awareness metric which we named as *Enery-awareness Node Ranking (ENR)* [18][1]. In this section, we build on top of this node-based cost, but consider a father/successor approach.

Based on the notion that in UCNs nodes are heterogeneous in terms of energy capacity ENR explores the fact that nodes that have been in idle mode for the majority of their lifetime, and that still exhibit a good estimate for their future energy level are the most adequate candidates to constitute a shortest-path.

In ENR we estimate how much of its lifetime has node i been in idle mode, to then provide an estimate towards the node's future energy expenditure, as this will for sure impact the node's lifetime. Such periods are the ones that are expensive to i in terms of energy. Hence, we consider the total period in idle time, t_{idle} over the full lifetime expected for a specific node, which is given by the sum of the elapsed time period T with the estimated lifetime of the node, as provided in equation 1. The estimated lifetime $C(i)$ provided by Garcia-Luna-Aceves et. al. [19] have considered the ratio between residual energy and drain rate which can capture the heterogeneous energy capability of nodes.

$$ENR(i) = \frac{T - t_{idle}}{T \times C(i)} \quad (1)$$

ENR is therefore a node weight which provides a ranking in terms of the node robustness, from an energy perspective, and having as goal to optimize the network lifetime. The smaller $ENR(i)$ is, the more likelihood a node has to be part of a path.

Based on ENR, we consider in this work the *Energy-awareness Father-Son (EFS)* metric, which considers a composition of the ENRs of both a father and successor nodes (c.f. Figure 2), as specified in equation 2.

[1]Short version under submission.

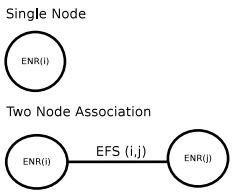


Fig. 2: Perspectives of the metrics.

$$EFS(i,j) = ENR(i) \times ENR(j) \quad (2)$$

EFS provides a ranking which we believe is useful to assist the routing algorithm to converge quickly in particular in multipath environments, as the selection on which successor to consider shall be made up from, by the father node. The goal is, similarly to ENR, to improve the network lifetime without disrupting the overall network operation. Hence, the smaller $EFS(i,j)$ is, the more likelihood a link has to become part of a path.

5. Performance Evaluation

This section provides a performance evaluation for EFS having as benchmark native AODV, as well as our previously proposed metric ENR. The evaluation has been performed by carrying out NS-2 (version 2.34) simulations. The scenarios are *Wireless Fidelity (Wi-Fi)* based. We have considered the NS-2 default physical layer, two-ray ground propagation model and DCF *(Distributed Coordination Function*) for MAC layer with 802.11g parameters.

For AODV we considered the native NS-2 module, here referenced as *AODV-native.* This module, *AODV-native,* considers hop-count as the metric to compute a shortest-path. Moreover, the original $C(i)$ has been developed to be applied to *Dynamic Source Routing (DSR)* protocol. The original specification of $C(i)$ therefore selects a best path based on a *min-max* approach, where the best path is the one that has the lowest bottleneck in terms of energy. Hence, we adapted the AODV to select the path in a min-max way as the original specification of the $C(i)$. We refer to this implementation as *AODV-minmax*. *AODV-ENR* and *AODV-EFS* represent AODV with our two metrics ENR and EFS, respectively. We describe next the topologies that have been considered in this evaluation.

5.1 Scenario 1 - Small Topologies

For this first scenario we have considered 25 nodes randomly distributed across a square with an area of 400m x 400m and 800m x 800m, respectively. We then considered a Poisson traffic model where each flow is based on *Variable Bit Rate (VBR)*, average packet size of 512 bytes, sending rate of 256 Kbps. Sources and destinations are randomly selected from the available nodes. Then, we consider 2 and 4 flows as a way to represent two different load levels. The

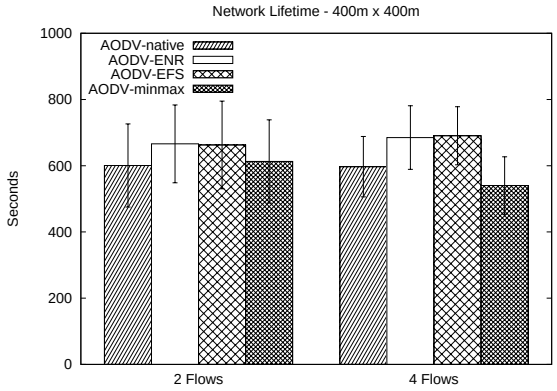


Fig. 3: Scenario-1, 400m x 400m.

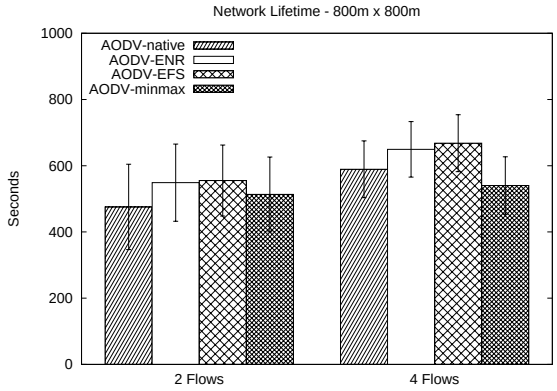


Fig. 4: Scenario-1, 800m x 800m.

simulation time has been set to 1000 seconds. In these set of experiments, all of the nodes are static, as what is relevant to us is to understand how the network behaves in terms of energy consumption. Hence, each node has been modeled to have different levels of energy parameters in order to represent heterogeneous Wi-Fi enable devices.

5.2 Scenario 2 - Traces-based Topology

For the design of this case, we have considered the settings that were obtained on traces available in the CRAWDAD project (*A Community Resource for Archiving Wireless Data At Dartmouth*) to be an example of a real environment to evaluate our proposed energy-awareness ranking metrics.

These traces have been collected based on GPS receivers, which logged data every 10 seconds. The NCSU scenario [20] comprises a human mobility data collected from 35 trace files, each corresponding to the perspective of a single node in one day, i.e., 24 hours. NCSU relies on a topology area of 2586.85 meters (X length) by 2347 meters (Y length) considering an uniform node speed of one meter per second (m/s). The authors have randomly selected 20 participants out of the students sharing a common interest, i.e., enrolled on the computer science department. Hence, more than one node contributed to different trace files, which there is no way to distinguish and to understand which node provided which trace.

For fair as possible, we have considered the same previews Poisson traffic model.

6. Evaluation Results

The results extracted intend to analyze benefits in terms of *network lifetime*. We define network lifetime as the time period since a topology becomes active, until a topology becomes disconnected, from the perspective of destination nodes. In other words, such time period is counted since the topology becomes active, until a destination cannot be reached by any of the available sources in the topology.

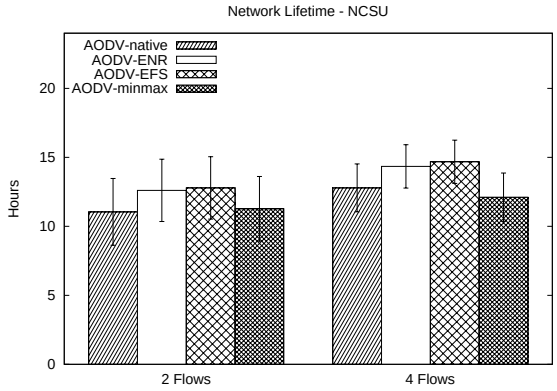


Fig. 5: Scenario-2.

Even though we analyze benefits in terms of network lifetime, we also want to understand the impact of the metrics in the overall network performance. For that, we consider three additional aspects: (i) *average end-to-end delay*, the time a packet takes between source and destination, comprising propagation and queuing delay. The end-to-end delay is computed per destination and then averaged across all destinations; (ii) *throughput*, the average number of bytes reaching destination nodes, measured in Kbps. The results presented correspond to the average throughput in the network, which is computed first per destination and then averaged across all destinations in the network; (iii) *average packet loss*, the percentage of packets that does not reach the destination. Average packet loss corresponds to the number of packets dropped between source and destination, averaged across all of the destinations.

To generate statistical sound results we relied on Akaroa2 [21]. All results have been computed within a 95% confidence interval.

6.1 Network Lifetime

To better analyze if our metrics behave coherently across different scenarios, based on the parameters described, all

Table 1: Network lifetime improvement.

AODV	400m x 400m		800m x 800m		NCSU	
	2flows	4flows	2flows	4flows	2flows	4flows
ENR	10.9%	14.7%	15.4%	10.2%	14.2%	12.1%
EFS	10.3%	15.6%	16.7%	13.4%	15.8%	14.7%
minmax	2.0%	-9.5%	7.9%	-8.3%	2.0%	-5.2%

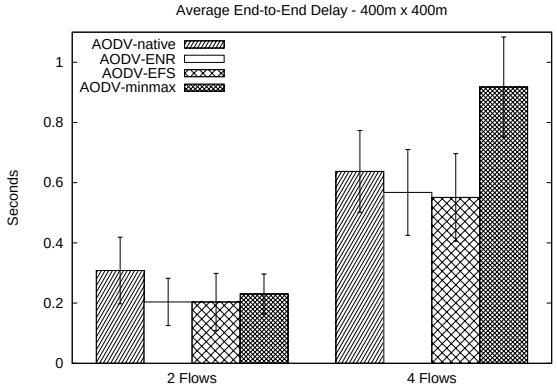


Fig. 6: Scenario-1, 400m x 400m.

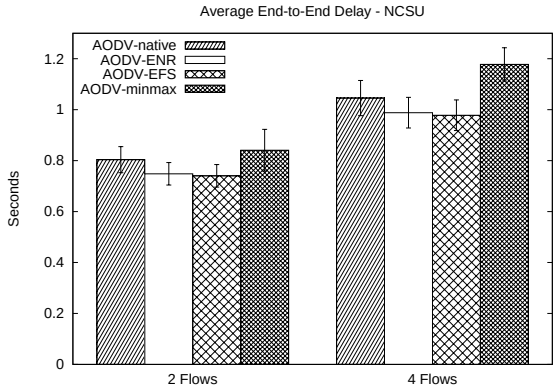


Fig. 8: Scenario-2.

Table 2: End-to-end delay improvement.

AODV	400m x 400m		800m x 800m		NCSU	
	2flows	4flows	2flows	4flows	2flows	4flows
ENR	33.9%	10.9%	2.6%	14.8%	6.9%	5.5%
EFS	34.0%	13.6%	6.4%	20.3%	7.9%	6.5%
minmax	25.2%	-44%	-3.6%	-3.3%	-4.6%	-12.6%

of the nodes of the described topologies have been set with initial energy levels picked up randomly. During the simulation we have observed that circa of 30% of nodes go down. Figures 3, 4 and 5 shows the average network lifetime for the different approaches. The X-axis represents the number of flows, while the Y-axis provides the network lifetime in seconds for Scenario-1 (cf. Figures 3 and 4) and in hours for Scenario-2 (cf. Figure 5).

Globally, as shown both metrics exhibit better results. In terms of EFS, we can see that in Scenario-1 there is no improvement when compared to ENR. We believe this occurs due to the fact that this is a very dense network and nodes travel short distances with a slow speed - the paths are quickly established and even though the nodes energy-levels are heterogeneous.

When the area changes (refer to Figure 4) then some improvement becomes visible, in particular for more congested networks. The same behavior occurs for Scenario-2, which points out that with realistic settings the improvement are similar.

While our metrics do stable behavior, the *AODV-minmax* varies according to the topology. For this concrete simulation scenarios, Table 1 provides the same results in percentage which show that the improvement provided by EFS becomes higher for multipath environments and when the network is more congested.

6.2 End-to-end Delay

As our main goal is to extend network lifetime without penalizing the end-to-end delay, throughput and packet loss,

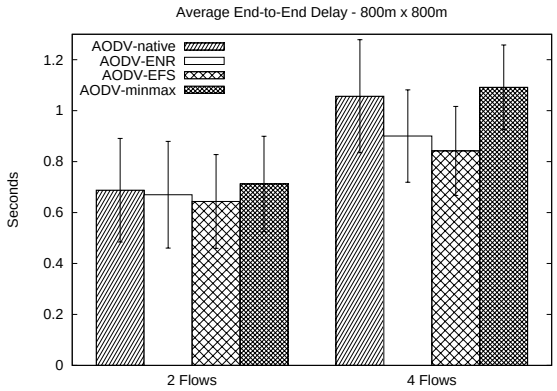


Fig. 7: Scenario-1, 800m x 800m.

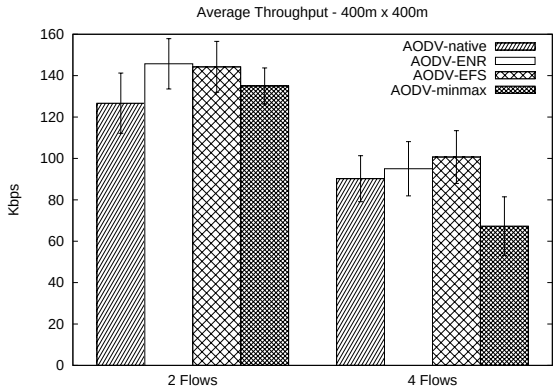


Fig. 9: Scenario-1, 400m x 400m.

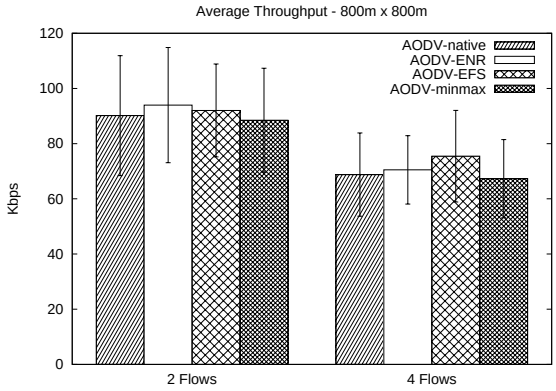


Fig. 10: Scenario-1, 800m x 800m.

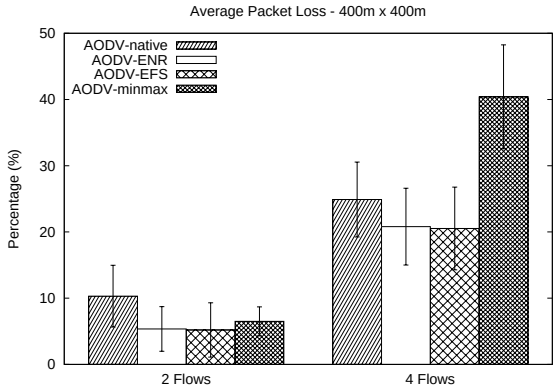


Fig. 12: Scenario-1, 400m x 400m.

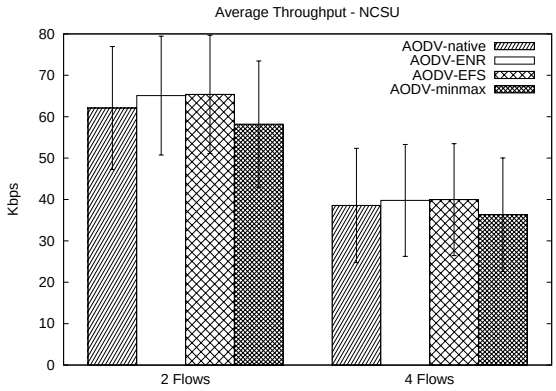


Fig. 11: Scenario-2.

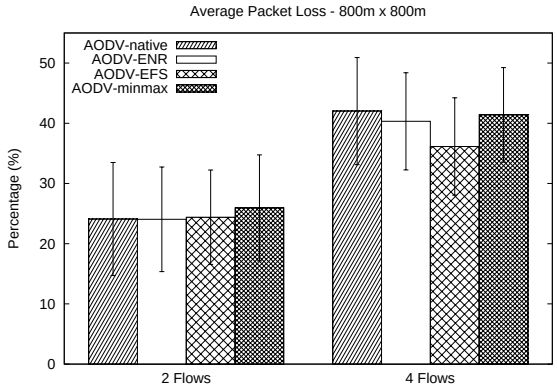


Fig. 13: Scenario-1, 800m x 800m.

Figures 6, 7 and 8 shows the average end-to-end delay achieved by applying EFS in comparison to the other approaches. We now consider only Scenario-2. The improvement margin against native AODV is provided in Table 2.

As shown both our metrics generate significantly less end-to-end delay. Across scenarios with higher load traffic, the *AODV-EFS* provides a higher gain. The reason for EFS to provide a higher gain in terms of delay reduction relates to the nodes selecting quicker a hop on the path by considering an association between two node as binding cost metric.

6.3 Throughput

We have then analyzed throughput impact and Figures 9, 10 and 11 shows the average throughput for the scenarios that have been set.

Table 3 provides the throughput gain against AODV. There is a slight gain of *AODV-ENR* and *AODV-EFS* in comparison to *AODV-native.* The gain provided by EFS is comparable to the gain provided by ENR for less congested networks. When the load on the network increases, then EFS provides slightly better values.

6.4 Packet Loss

We have analyzed the packet loss impact and Figures 12, 13 and 14 shows results obtained for all approaches, while Table 4 provides the gain derived from applying each metric, against native AODV.

For packet loss, EFS provides a higher gain across all scenarios, but becoming significantly higher for scenarios where nodes seem to be more mobile (larger distances) and the network is more congested.

Table 3: Throughput *gain.*

AODV	400m x 400m		800m x 800m		NCSU	
	2flows	4flows	2flows	4flows	2flows	4flows
ENR	15%	5.3%	4.2%	2.5%	4.8%	3.2%
EFS	13.9%	11.6%	2.1%	9.7%	5.3%	3.7%
minmax	6.5%	-25.4%	-1.9%	-2.2%	-6.3%	-5.8%

Table 4: Packet loss *gain.*

AODV	400m x 400m		800m x 800m		NCSU	
	2flows	4flows	2flows	4flows	2flows	4flows
ENR	48.0%	16.4%	0.2%	4.0%	4.1%	2.0%
EFS	49.5%	17.6%	1.3%	14.0%	5.0%	2.8%
minmax	36.8%	-62.2%	-7.7%	-1.5%	-5.2%	-4.3%

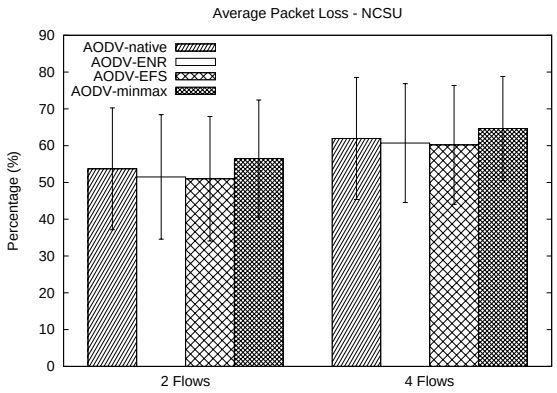


Fig. 14: Scenario-2.

7. Conclusions

In this paper we discuss and propose energy-awareness routing metrics that can provide stability in terms of network lifetime to multihop shortest-path based routing, without incurring strong penalties in terms of operational changes and maintenance. We have evaluated the our proposed metrics based on realistic settings for a specific case of on-demand routing, AODV.

We have shown that the energy-awareness metrics being proposed significantly improve AODV behavior in terms of network lifetime and also in terms of overall network operation, based upon throughput, packet loss, end-to-end delay.

In terms of the behavior of EFS vs. ENR there seems to be an improvement in particular when scenarios have larger distances, and when the network load is higher. This implies that EFS seems to provide more robustness when scenarios have more variability (e.g. more nodes moving, and several successors at disposal). In terms of network lifetime and for the scenarios evaluated, EFS results in a small improvement. We intend to explore further scenarios to analyze whether or not these conditions always hold. The end-to-end delay is also slightly improved.

The greater advantage of applying EFS instead of ENR seems to relate with an improvement in throughput and a significant improvement concerning packet loss. Our belief for this gain relates to the fact that EFS allows nodes to react quicker to energy changes on a path - resulting paths will be more robust earlier in time, assuming that nodes have a reasonable out-degree (several successors available).

We are currently carrying on this work both by fine-tuning not only scenarios but also the proposed metrics by evaluating their potential contribution for other forms of multihop routing, e.g. link state routing by *Optimized Link State Routing (OLSR)* protocol. As future work, we also intend to release the metrics explaining how they can be applied to the two main families of shortest-path routing, i.e., link state and distance-vector approaches.

Acknowledgment

This work is supported by Fundação Ciência e Tecnologia (FCT) PhD scholarship number SFRH/BD/44005/2008 and sponsored by national fundings via FCT, in the context of the UCR project PTDC/EEA-TEL/103637/2008.

References

[1] R. Sofia, P. Mendes, W. Moreira, A. Ribeiro, S. Queiroz, A. Junior, T. Jamal, N. Chama, and L. Carvalho, "UPNs: User-provided Networks, Technical Report: Living-examples, challenges, advantages, Tech. Rep. SITI-TR-11-03, March, 2011.

[2] "ULOOP: User-centric Wireless Local-Loop," EU IST FP7 Project (Grant 257418).

[3] A. Junior and R. Sofia, "Energy-efficient routing in user-centric environments," *The 2010 IEEE/ACM International Conference on Green Computing and Communications (GreenCom2010)*, Dez, 2010.

[4] A. Junior, R. Sofia, and A. Costa, "Energy-efficient routing," in *19th IEEE International Conference on Network Protocols (ICNP)*, oct. 2011, pp. 295 –297.

[5] C. E. Perkins, E. M. Belding-Royer, and S. R. Das, "Ad hoc on-demand distance vector (aodv) routing," RFC 3561, July 2003.

[6] C. Yu, B. Lee, and H. Y. Youn, "Energy efficient routing protocols for mobile ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 3, no. 8, pp. 959–973, 2003.

[7] S. Mahfoudh and P. Minet, "Survey of energy efficient strategies in wireless ad hoc and sensor networks," in *ICN 2008*, April 2008.

[8] K. Scott and N. Bambos, "Routing and channel assignment for low power transmission in pcs," in *5th IEEE International Conference on Universal Personal Communications*, vol. 2, Oct 1996, pp. 498–502.

[9] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *INFOCOM 2000*, vol. 1, 2000, pp. 22–31 vol.1.

[10] Q. Xie, C.-T. Lea, M. Golin, and R. Fleischer, "Maximum residual energy routing with reverse energy cost," in *IEEE GLOBECOM '03*, vol. 1, Dec. 2003.

[11] C.-K. Toh, "Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks," *IEEE Communications Magazine,*, vol. 39, no. 6, pp. 138–147, Jun 2001.

[12] S. Kwon and N. Shroff, "Unified energy-efficient routing for multihop wireless networks," in *INFOCOM 2008*, april 2008, pp. 430–438.

[13] T. Zhu and D. Towsley, "E2R: Energy efficient routing for multi-hop green wireless networks," in *INFOCOM (WKSHPS)*, April 2011.

[14] K. Gomez, R. Riggio, T. Rasheed, and F. Granelli, "Analysing the energy consumption behaviour of wifi networks," in *IEEE Online Conference on Green Communications (GreenCom)*, sept. 2011.

[15] A. Misra and S. Banerjee, "MRPC: maximizing network lifetime for reliable routing in wireless environments," in *IEEE Wireless Communications and Networking Conference*, Mar 2002.

[16] X. Zhang, F. Zou, E. Wang, and D. K. Sung, "Exploring the dynamic nature of mobile nodes for predicting route lifetime in mobile ad hoc networks," in *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, January 2010.

[17] J. Vasseur, M. Kim, K. Pister, N. Dejean, and D. Barthel, "Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks," RFC 6551, March 2012.

[18] A. Junior, R. Sofia, and A. Costa, "User-Centric Routing - Energy-Efficient Routing in User-Centric Environments, progress report 2010/2011," SITI/University Lusófona and University of Minho, Tech. Rep. SITI-TR-11-10, 2011.

[19] D. Kim, J. J. Garcia-Luna-Aceves, K. Obraczka, J.-C. Cano, and P. Manzoni, "Routing mechanisms for mobile ad hoc networks based on the energy drain rate," *IEEE Transactions on Mobile Computing*, vol. 2, no. 2, pp. 161–173, 2003.

[20] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "CRAWDAD data set ncsu/mobilitymodels (v. 2009-07-23)," Downloaded from http://crawdad.cs.dartmouth.edu/ncsu/mobilitymodels, July 2009.

[21] G. C. Ewing, K. Pawlikowski, and D. Mcnickle, "Akaroa2: Exploiting network computing by distributing stochastic simulation," in *International Society for Computer Simulation*, 1999, pp. 175–181.

Approximate Modeling of Wireless Channel Based on Service Process Burstiness

Shuguang Fang[1,2], **Yuning Dong**[1], **and Haixian Shi**[3]
[1] College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
[2] Electronic Engineering Department, Wuxi Institute of Commerce, Wuxi, China
[3] School of Horticulture, Nanjing Agricultural University, Nanjing, China

Abstract - *For link packet loss rate, delay and throughput etc. QoS performance analysis in wireless communication systems, queuing analysis based on measuring or estimating models for traffic flow and wireless channel is one of critical technologies. However, the complicated modern wireless communication systems and unreliable wireless channels usually cause accurate wireless channel models very complicated, which makes queuing analysis be of high computation complexity. Burstiness, induced by the effect of a fading channel process and modulation and coding technology in physical layer, is one of critical factors affecting QoS performance of wireless communication systems. To this end, this paper proposes a novel simpler modeling method, by which one can get a relatively simpler wireless channel model whose burstiness behavior is equivalent to that of the original channel model. Numerical examples are given to show the effectiveness of the proposed model.*

Keywords: Wireless Channel; QoS Performance; Burstiness; Approximate Modeling

1 Introduction

Queuing analysis technology is one of critical technologies for link packet loss rate, delay and throughput etc. QoS metrics analysis in wireless communication systems. The accuracy of link QoS metrics to a great extent depends on the accuracy of wireless channel model in queuing analysis technology, therefore it is necessary to model wireless channel accurately. However, varying wireless channel conditions due to mobility and changing environment [1] cause accurate channel modeling to be very challenging.

Wireless channel process modeling through first-order Markovian chains [2, 3, 4, 5] is simplicity and analytical tractability, but later results in [6] show that such methods can be very erroneous in several cases of interests [7]. In ref.[8] a physical-layer 2-D Markov model using both the amplitude and the rate of change of the fading envelope is presented, and based on this multidimensional physical-layer Markov model, the quality-of-service (QoS) at the data-link layer is invested relying on an analytical framework based on a discrete-time Markov chain in multi-rate adaptive modulating and coding(AMC) wireless networks, this model provides a valuable radio-link-level performance measure, but its complexity exponentially increases with the number of states of the arrival process, the number of states of the PHY Markov model, or the maximum queue length of the system. In ref. [9] a cross-layer analytical framework is presented for analyzing the QoS performance of the decode-and-forward (DF) relaying wireless networks, where the AMC is employed at the physical layer under the conditions of unsaturated traffic and finite-length queue at the data link layer. Considering the characteristic of DF relaying protocol at the physical layer, authors model a two-hop DF relaying wireless channel with AMC as an equivalent Finite State Markov Chain (FSMC) in queuing analysis, its complexity similarly increases with the number of states of the PHY Markov model, or the maximum queue length of the system.

Advanced hidden Markov models(HMM) would accurately model the wireless channel [10], note that if the HMM of channel is available, certain queuing theoretic results can be applied for more exact analysis for some specific cases[11], however this method is similarly complex in complicated modern wireless communication system.

Thus there is the need for simple modeling and approximation approaches that capture the most effect of wireless channel process characteristic on link QoS performance analysis in queuing analysis technology. The effect of wireless channel fading process and modulating and coding technology in physical-layer make wireless channel service process be burstiness, which critically affects the wireless link QoS performance [7, 12]. Based on this, we propose a novel simple approximate wireless channel modeling method based on Peakedness in discrete time[13], by which the simple approximate model Peakedness equal to the original channel model Peakedness, and verify its effective by numerical examples.

The rest of this paper organized as follows. The generalized Peakedness of process burstiness and the Peakedness in discrete time of Markov modulated batch Bernoulli process (MMBBP) are introduced in Section 2; In Section 3 we propose a novel wireless simple model approximate method based on Peakedness in discrete time,

and verify its effectiveness by numerical examples in Section 4; Finally this paper concludes in Section 5 including the main results and future works.

2 The Process Burstiness and Description

Burstiness is a term often used in traffic analysis to represent time correlation in arrival statistics [7]. System complexity of modern wireless networks causes the arrival streams at link and channel service process are highly positively correlated in time, therefore they are all bursty process, in other words, they are all burstiness, which is one of critical factors affecting link QoS performance.
The simplest burstiness measures take only the first-order properties of the process into account, in practice the peak to mean ratio and the squared coefficient of variation are the most frequently used; the burstiness measures expressing second-order properties of the process are more complex, including the autocorrelation function, the indices of dispersion and generalized peakedness [13]. Comparing with the squared coefficient of variation, the autocorrelation function and the indices of dispersion, generalized peakedness taking the first-order and second-order properties of process and the system servicing the process into account is more effective and efficient process burstiness measure.

Peakedness, defined based on the Infinite-Server Effect Principle [14], was originally developed by teletraffic engineers to characterize call arrival streams modeled as stationary point process for approximating block probability at trunk groups [15, 16], in which the service time distribution of the fictitious infinite server group usually is exponential service time distribution. In ref. [17], Eckberg defined the generalized Peakedness for any service distribution $B(t)$, and whose mean is $1/\mu$ (μ is the fictitious infinite server group service rate), then the generalized Peakedness $z\{B(t)\}$ defined as the variance-to-mean ratio of the number of busy servers in the fictitious infinite server group,

$$z_X[B] = \frac{Var[S(t)]}{E[S(t)]} \tag{1}$$

Where, $S(t)$ is the number of busy servers in the fictitious infinite server group at time t. Ref. [13] and [15] propose the peakedness in discrete time and the modified peakedness which encompasses point process and fluid models under a common framework and verify them respectively. The peakedness function in discrete time of the arrival stream with respect to geometric holding time distribution is given by [13],

$$\tilde{z}_{geo}[\mu] = 1 + \frac{K^*[1-\mu]-1}{2-\mu} \tag{2}$$

Where, $K^*[1-\mu]$ is the z-transform of the autocorrelation function of the arrival process.

Based on peakedness in discrete time, ref. [13] presents the peakedness result of Markov modulated batch Bernoulli process (MMBBP). In MMBBP, we have a discrete time Markov process as a modulating process. In each state of the modulating Markov-process, batch arrivals are generated according to a general distribution corresponding to the state. Let P and D denote the transition probability matrix and the steady-state distribution vector of the modulating Markov process respectively,

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n-1} & p_{1,n} \\ p_{2,1} & p_{2,2} & & p_{2,n-1} & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{n-1,1} & p_{n-1,2} & \cdots & p_{n-1,n-1} & p_{n-1,n} \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n-1} & p_{n,n} \end{bmatrix} \tag{3}$$

$$D = [\pi_1, \pi_2, \cdots, \pi_{n-1}, \pi_n] \tag{4}$$

Owing to the Markov property of the process, we have the following formulas,

$$\sum_{j=1}^{n} p_{i,j} = 1 \tag{5}$$

$$\sum_{j=1}^{n} \pi_j = 1 \tag{6}$$

$$DP = D \tag{7}$$

Let M_1 and M_2 be diagonal matrices corresponding to the first and second moments of the number of arrivals in the corresponding states, let I be the identity matrix and e be a vector of all ones.

$$M_1 = \begin{bmatrix} m_{1,1} & 0 & \cdots & 0 & 0 \\ 0 & m_{2,1} & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & m_{n-1,1} & 0 \\ 0 & 0 & \cdots & 0 & m_{n,1} \end{bmatrix} \tag{8}$$

$$M_2 = \begin{bmatrix} m_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & m_{2,2} & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & m_{n-1,2} & 0 \\ 0 & 0 & \cdots & 0 & m_{n,2} \end{bmatrix} \quad (9)$$

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad (10)$$

$$e = [1,1,\cdots,1,1]' \quad (11)$$

where, $m_{i,1}$ and $m_{i,2}$ are respectively the first and second moments of the number of arrivals corresponding to state i. We can express the mean number of arrivals as $m_1 = M_1 eD'$ and the second moment as $m_2 = M_2 eD'$. The auto-covariance function of the arrival process is given by $k[i] = DM_1 P^i M_1 e - m_1^2$. Then, according to (2), the peakedness function of MMBBP $\tilde{z}geo[\mu]$ with respect to geometric holding time distribution defined as (12).

$$\tilde{z}_{geo}[\mu] = 1 + \frac{1}{2-\mu}\left(\frac{2(1-\mu)DM_1 P\left(I-(1-\mu)P\right)^{-1} M_1 e + m_2}{m_1} - 1 \right) - \frac{m_1}{\mu} \quad (12)$$

3 The Wireless Channel Approximate Modeling

In this section, we propose a simple wireless channel approximate modeling method based on equivalent process burstiness, by which we can get the simple wireless channel approximate model (SWCAM), whose peakedness function value equal to that of the channel process, by measuring first-order moment m_1 and peakedness function value $\tilde{z}_{geo}[\mu]$. First, we get the first-order and second-order moment of SWCAM according to the measuring peakedness; second, we define variable Y_i whose value can be obtain by the measuring peakedness; finally, we get the SWCAM by numerical analysis. Let Em_1 and Em_2 be the first and second moment of SWCAM, EM_1 be diagonal matrices of SWCAM corresponding to the first-order moment of the number of serviced packets by channel in the corresponding states, EP and ED be the transition probability matrix and the steady-state distribution vector of the SWCAM respectively, and which similarly satisfy formula (5), (6) and (7).

3.1 The First-Order and Second-Order Moment of SWCAM

Let service rate $\mu = 1$, then formula (12) equally transformed into,

$$\tilde{z}_{geo}(1) = m_2 / m_1 - m_1 \quad (13)$$

Then, we can get the second-order moment of channel service process by measuring m_1 and $\tilde{z}_{geo}(1)$,

$$m_2 = \left(\tilde{z}geo(1) + m_1\right) m_1 \quad (14)$$

According to the principle of equivalence, the first-order and second-order moment of the SWCAM should equal to m_1 and m_2 respectively,

$$\begin{cases} Em_2 = m_2 \\ Em_1 = m_1 \end{cases} \quad (15)$$

3.2 The Variable Y_i

Let $\omega_i = 1 - \mu_i$, formula (2) may be equivalently transformed into,

$$K^*(\omega_i) = \left(\tilde{z}_{geo}(\mu_i) - 1\right)(\omega_i + 1) + 1 \quad (16)$$

Using eq. (12) and (14), we can compute,

$$K^*(\omega_i) = \frac{2\omega_i DM_1 P\left(I - \omega_i P\right) M_1 e}{m_1} + \tilde{z}_{geo}(1) - \frac{2\omega_i}{1-\omega_i} m_1 \quad (17)$$

Then, we define variable Y_i as,

$$Y_i = Y(\omega_i) = \frac{m_1}{2\omega_i}\left(K^*(\omega_i) + \frac{2\omega_i}{1-\omega_i} m_1 - \tilde{z}_{geo}(1) \right) \quad (18)$$

which can be computed by measuring m_1 and $\tilde{z}_{geo}[\mu]$ of the channel service process.

3.3 The Simple Wireless Channel Approximate Modeling

Using eq. (17), eq. (18) equivalently transformed into,

$$Y_i = Y(\omega_i) = DM_1P(I - \omega_i P)M_1 e \quad (19)$$

Defining vector M_0 as ,

$$M_0 = M_1 e = \left[m_{1,1}, m_{2,1}, \cdots, m_{n-1,1}, m_{n,1} \right]' \quad (20)$$

defining row matrix $X(\omega_i)$ as,

$$X(\omega_i) = \left[x_{i,1}, x_{i,2}, \cdots, x_{i,n-1}, x_{i,n} \right] = DM_1P(I - \omega_i P)^{-1} \quad (21)$$

Using eq. (20) and (21) , eq. (19) equally rewritten as,

$$X(\omega_i)M_0 = Y(\omega_i) \quad (22)$$

Eq. (21) can be equally transformed into ,

$$X(\omega_i)(I - \omega_i P) = DM_1P \quad (23)$$

Then, using eq. (23), we can get the system of equations (24) (given in APPENDIX). Using eq. (5) and (6), adding Left side and Right side of system of equations (24) respectively, and then we can get,

$$x_{i1} + x_{i,2} + \cdots + x_{i,n} = m_1 / \omega_i \quad (25)$$

Using system of equation (24) and eq. (25), we can get the expressions of variables $x_{i1}, x_{i,2}, \cdots, x_{i,n}$ for different ω_i , which only includes the state transition probability and constants, and using these expressions and eq. (22), we can get formula (26) (given in APPENDIX) with state transition probability for different ω_i. For the Simple Wireless Channel Approximate Modeling based on equivalent peakedness, the state transition probability and steady-state distribution of SWCAM should similarly satisfy eq. (26), i.e. satisfy eq. (27) (given in APPENDIX). And ED, $EM1$ similarly satisfy the following formula,

$$ED \times EM1 = Em1 \quad (28)$$

Then accord to the number of variables which determined by the states of SWCAM, we can have the system of equations (29) (given in APPENDIX).

When we get the SWCAM by system of equations (29), we should determine the times of $\tilde{z}_{geo}[\mu]$ measuring according to the number of states of the SWCAM. Using system of equations (29), we can get the SWCAM of original channel with equivalent peakedness by numerical method, and we will verify system of equations (29) effective by numerical examples in section 4.

4 Numerical Results

Our objective in this section is mainly to validate the simple wireless channel approximate modeling method introduced in Section 3. For giving wireless channel service process model in multi-states Markov model, we can get the correspondingly SWCAM by the methods introduced in Section 3.

Table 1 (given in APPENDIX) lists the SWCAM in simple two-state Markov model correspondingly to 10-state, 5-state and 3-state original Markov model of wireless channel model correspondingly, which shows the feasibility of the methods introduced in Section 3.

All the figures in this section compare the peakedness of multi-state original wireless channel Markov model (Multi-states OWCMM) and that of correspondingly two-state simple wireless channel approximate model (Two-state SWCAM). Figs. 1-3 reveal the peakedness equivalent of two-states SWCAM and the correspondingly wireless channel model in 10-state, 5-state and 3-state Markov models respectively listed in Table 1, and validate the simple wireless channel approximate modeling method introduced in Section Ⅲ. Comparing curves in Figures. 1-3, it can be deduced that the orders of SWCAM and original channel model more close, the peakedness of SWCAM and original channel model more consistent, and correspondingly SWCAM more accurate. In Fig.2 and Fig.3, the order difference between SWCAM and original channel model are respectively 1 and 3, which are all small, and the peakedness of SWCAM and original channel are all well consistent; but the order difference between SWCAM and original channel model in Fig.3 is 8, which are bigger than that in figures1-2, then the peakedness consistent between SWCAM and original channel is inferior to that in Figs.1-2.

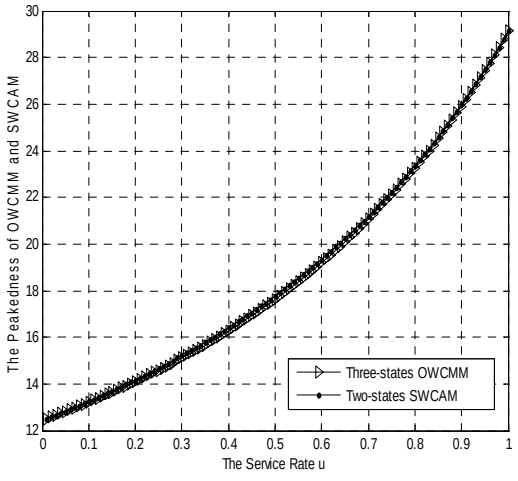


Figure 1. The peakedness of Three-states OWCMM and Two-states SWCAM

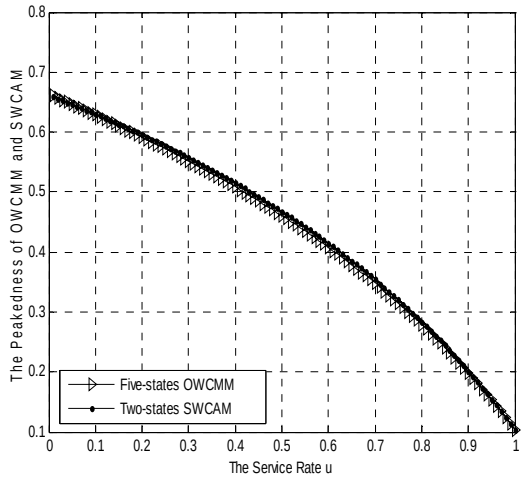


Figure 2. The peakedness of Five-states OWCMM and Two-states SWCAM

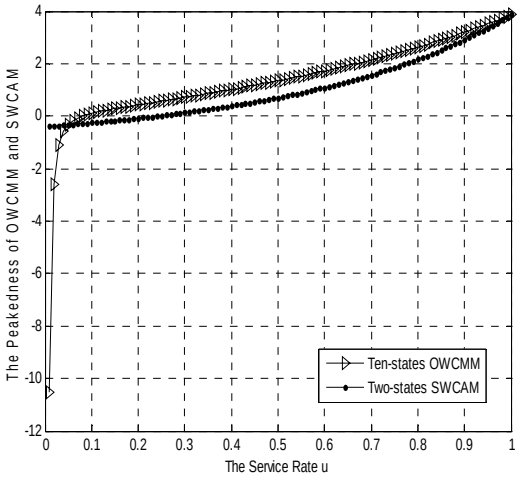


Figure 3. The peakedness of Ten-states OWCMM and Two-states SWCAM

Table 1 (given in APPENDIX) lists the SWCAM in simple two-state Markov model correspondingly to 10-state, 5-state and 3-state original Markov model of wireless channel model correspondingly, which shows the feasibility of the methods introduced in Section 3.

All the figures in this section compare the peakedness of multi-state original wireless channel Markov model (Multi-states OWCMM) and that of correspondingly two-state simple wireless channel approximate model (Two-state SWCAM). Figs. 1-3 reveal the peakedness equivalent of two-states SWCAM and the correspondingly wireless channel model in 10-state, 5-state and 3-state Markov models respectively listed in Table 1, and validate the simple wireless channel approximate modeling method introduced in Section Ⅲ. Comparing curves in Figures. 1-3, it can be deduced that the orders of SWCAM and original channel model more close, the peakedness of SWCAM and original channel model more consistent, and correspondingly SWCAM more accurate. In Fig.2 and Fig.3, the order difference between SWCAM and original channel model are respectively 1 and 3, which are all small, and the peakedness of SWCAM and original channel are all well consistent; but the order difference between SWCAM and original channel model in Fig.3 is 8, which are bigger than that in figures1-2, then the peakedness consistent between SWCAM and original channel is inferior to that in Figs.1-2.

5 Conclusions

Modern wireless communication system complexity and wireless channel variability make accurate wireless channel modeling be highly complex, which makes the computation complexity of queuing analysis in wireless channel QoS performance exponentially increase with the number of states of the arrival process, the number of states of the PHY Markov model, or the maximum queue length of the system. Meanwhile, the burstiness behavior of wireless channels service process is a critical factor which affects the wireless link QoS performance. With the proposed burstiness approximate modeling method, one can get a simpler model of the wireless channel, whose burstiness behavior is equivalent to that of the corresponding wireless channel service process. To validate this method, we analyze the peakedness equivalent of three-order, five-order and ten-order Markov wireless channel models with the corresponding two-order equivalent burstiness low complex wireless channel Markov model respectively. The application of this method to wireless channel QoS analysis will be one of our future works.

6 Acknowledgment

This work is supported in part by National Natural Science Foundation of China (No.60972038), Ministry of Education (China) Ph.D. Programs Foundation (No.20103223110001), and the graduate student scientific research innovation project of Jiangsu Province, China (No.CXZZ11_0392)

7 References

[1] Mustafa Cenk Gursoy. “On the Capacity of Training-Based Transmissions with Input Peak Power Constrain- ts”. Communications, 2008. ICC '08. IEEE International Conference on, pp. 1282-1286, 2008.

[2] H. S. Wang and N. Moayeri, “Finite-state Markov Channels - A Useful Model for Radio Communication Channels,” IEEE Trans. Veh. Technol., vol. 44, no. 1, pp. 167-171, 1995.

[3] ZHANG Q and KASSAM S.A., “Finite-state Markov model for Rayleigh fading channels,” IEEE Trans. Commun., vol. 47, no. 11, pp. 1688-1692, 1999.

[4] J. G. Kim and M. M. Krunz, “Delay analysis of selective repeat ARQ for a Markovian source over wireless channel,” *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1968-1981, Sept. 2000.

[5] L. B. Le, E. Hossain, and A. S. Alfa, “Radio link level performance evaluation in wireless networks using multi-rate transmission with ARQ-based error control,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 10, pp. 2647–2653, Oct. 2006.

[6] C. C. Tan and N. C. Beaulieu, “On First-order Markov Modeling for the Rayleigh Fading Channel,” IEEE Trans. Commun., vol. 48, no. 12, pp. 2032-2040, 2000.

[7] Amal Ekbal and John M. Cioffi, “Effect of Wireless Channel Process on Queueing Delay-Approximate Analysis Using Peakedness Function,” IEEE International Conference on Communications, pp.468 – 472, Jan. 2005.

[8] Guillem Femenias, Janume Ramis and Loren Carrasco, “Using Two-Dimensional Markov Models and the Effective-Capacity Approach for Cross-Layer Design in AMC/ARQ-Based Wireless Networks,” IEEE Trans. on Vehicular Technology, vol. 58, No. 8, Oct. 2009.

[9] K. Zheng, Y. Wang, L. Lei and W. Wang, “Cross-Layer Queuing Analysis on Multi-hop Relaying Networks with Adaptive modulation and Coding ,” IET Commun., vol. 4, no. 3, pp. 295-2, 2010.

[10] W. Turin and R. van Nobelen, “Hidden Markov Modeling of Flat Fading Channels,” IEEE J. Select. Areas Commun., vol. 16, no. 9, pp. 1809-1817, 1998.

[11] W. Turin and M. Zorzi, “Performance analysis of delay-constrained communications over slow Rayleigh fading channels,” IEEE Trans. Wireless Comm., vol. 1, no. 4, pp. 801-807, 2002.

[12] Alexei V. Nikitin, “On the Impulsive Nature of Interchannel Interference in Digital Communication Systems,” Radio and Wireless Symposium (RWS), 2011 IEEE, pp. 118-121, Jan. 2011.

[13] S. Molnar and Gy. Miklos, “Peakedness Characterization in Teletraffic,” PICS '98 Proceedings of the IFIP TC6/WG6.3 Seventh International Conference on Performance of Information and Communication Systems, 1996.

[14] David L. Jagerman and Benjamin Melanmed, “Burstiness Description of Traffic Streams: Indices of Dispersion and Peakedness,” the Conference on Information Science and System, Princeton, New Jersey, vol.1, pp.24-28, 1994.

[15] Brian I. Mark, David L. Jagerman and G. Ramamurthy, “Peakedness Measures for Traffic Characterization in High-Speed Networks,” Proceedings of the 16th IEEE Annual Conference on Computer Communications, 1997.

[16] H. Heffes and J. M. Holtzman, “Peakedness of Traffic Carried by a Finite Trunk Group With Renewal Input,” The Bell System Technical Journal, vol. 52, no. 9, pp. 1617-1642, 1973.

[17] Eckberg, A. “Approximations for Bursty (and smoothed) Arrival Delay based on Generalized Peakedness,” Proceedings of the 11-th International Teletraffic Congress, Kyoto, Japan. 1985.

8 Appendix

$$\begin{cases} x_{i1}\left(1-\omega_i p_{1,1}\right)-x_{i,2}\omega_i p_{2,1}-\cdots-x_{i,n}\omega_i p_{n,1}=\pi_1 m_{1,1}p_{1,1}+\pi_2 m_{2,1}p_{2,1}+\cdots+\pi_n m_{n,1}p_{n,1} \\ -x_{i1}\omega_i p_{1,2}+x_{i,2}\left(1-\omega_i p_{2,2}\right)-\cdots-x_{i,n}\omega_i p_{n,2}=\pi_1 m_{1,1}p_{1,2}+\pi_2 m_{2,1}p_{2,2}+\cdots+\pi_n m_{n,1}p_{n,2} \\ \ddots \\ -x_{i1}\omega_i p_{1,n}-x_{i,2}\omega_i p_{2,n}-\cdots+x_{i,n}\left(1-\omega_i p_{n,n}\right)=\pi_1 m_{1,1}p_{1,n}+\pi_2 m_{2,1}p_{2,n}+\cdots+\pi_n m_{n,1}p_{n,n} \end{cases} \tag{24}$$

$$f_{20,Y\left(\omega_i\right)}\left(p_{1,n},\cdots,p_{1,n-1};\cdots;p_{n,1},\cdots,p_{n,n-1},\pi_1,\cdots,\pi_n\right)=Y\left(\omega_i\right) \tag{26}$$

$$f_{20,Y\left(\omega_i\right)}\left(Ep_{1,n},\cdots,Ep_{1,n-1};\cdots;Ep_{n,1},\cdots,Ep_{n,n-1},E\pi_1,\cdots,E\pi_n\right)=Y\left(\omega_i\right) \tag{27}$$

$$\begin{cases} f_{20,Y(\omega_1)}\left(Ep_{1,n},\cdots,Ep_{1,n-1};\cdots;Ep_{n,1},\cdots,Ep_{n,n-1},E\pi_1,\cdots,E\pi_n\right)=Y(\omega_1) \\ \qquad\qquad \ddots \\ f_{20,Y(\omega_i)}\left(Ep_{1,n},\cdots,Ep_{1,n-1};\cdots;Ep_{n,1},\cdots,Ep_{n,n-1},E\pi_1,\cdots,E\pi_n\right)=Y(\omega_i) \\ ED\times EM1=m1 \\ \sum_{j=1}^{n}E\pi_j=1 \\ ED\times EP=ED \end{cases} \tag{29}$$

TABLE I. SERVICE PROCESS AND SWCAM OF WIRELESS CHANNEL

Wireless Channel Service Process			SWCAM	
The number of states	State Transition Matrix P	M_1	EP	EM_1
10	$\begin{bmatrix} 0.0913 & 0.0818 & 0.0838 & 0.0575 & 0.0954 & 0.0938 & 0.1534 & 0.1491 & 0.1209 & 0.0710 \\ 0.1404 & 0.0922 & 0.0667 & 0.1625 & 0.1516 & 0.0952 & 0.1077 & 0.1016 & 0.0559 & 0.0521 \\ 0.1167 & 0.0571 & 0.2092 & 0.0483 & 0.0560 & 0.0423 & 0.0564 & 0.1080 & 0.0771 & 0.2288 \\ 0.0941 & 0.0404 & 0.1979 & 0.2142 & 0.0960 & 0.0243 & 0.0564 & 0.0894 & 0.1301 & 0.0573 \\ 0.1699 & 0.2004 & 0.0625 & 0.0331 & 0.0836 & 0.0838 & 0.1195 & 0.1431 & 0.0241 & 0.0740 \\ 0.1390 & 0.0651 & 0.1662 & 0.1257 & 0.0848 & 0.1004 & 0.0442 & 0.0756 & 0.1671 & 0.0949 \\ 0.0999 & 0.0444 & 0.0937 & 0.1196 & 0.1301 & 0.0758 & 0.0704 & 0.1883 & 0.0072 & 0.1696 \\ 0.1942 & 0.1693 & 0.0210 & 0.0557 & 0.0713 & 0.1445 & 0.0290 & 0.1538 & 0.0227 & 0.1390 \\ 0.0654 & 0.1346 & 0.1225 & 0.1561 & 0.1559 & 0.0657 & 0.1207 & 0.0342 & 0.0653 & 0.1285 \\ 0.0886 & 0.0831 & 0.1566 & 0.1065 & 0.1069 & 0.1488 & 0.1394 & 0.0998 & 0.0817 & 0.0845 \end{bmatrix}$	$\begin{bmatrix} 7.6722 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.9752 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.0947 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 87.6389 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17.1912 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10.0941 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 20.9822 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 24.6944 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 85.9222 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 18.5901 \end{bmatrix}$	$\begin{bmatrix} 0.4000 & 0.6000 \\ 0.8000 & 0.2000 \end{bmatrix}$	$\begin{bmatrix} 83.6332 & 0 \\ 0 & 33.4739 \end{bmatrix}$
5	$\begin{bmatrix} 0.4000 & 0.2000 & 0.1000 & 0.2000 & 0.1000 \\ 0.1000 & 0.3000 & 0.2500 & 0.2000 & 0.1500 \\ 0.1100 & 0.2900 & 0.3000 & 0.2000 & 0.1000 \\ 0.0500 & 0.1500 & 0.1000 & 0.3000 & 0.4000 \\ 0.1000 & 0.1000 & 0.1000 & 0.3000 & 0.4000 \end{bmatrix}$	$\begin{bmatrix} 2.0769 & 0 & 0 & 0 & 0 \\ 0 & 3.1491 & 0 & 0 & 0 \\ 0 & 0 & 0.4753 & 0 & 0 \\ 0 & 0 & 0 & 0.2583 & 0 \\ 0 & 0 & 0 & 0 & 1.3966 \end{bmatrix}$	$\begin{bmatrix} 0.8000 & 0.2000 \\ 0.6000 & 0.4000 \end{bmatrix}$	$\begin{bmatrix} 1.8420 & 0 \\ 0 & 0.0323 \end{bmatrix}$
3	$\begin{bmatrix} 0.4000 & 0.2000 & 0.4000 \\ 0.2000 & 0.1000 & 0.7000 \\ 0.3000 & 0.2500 & 0.4500 \end{bmatrix}$	$\begin{bmatrix} 32.4247 & 0 & 0 \\ 0 & 32.0724 & 0 \\ 0 & 0 & 185.1371 \end{bmatrix}$	$\begin{bmatrix} 0.3000 & 0.7000 \\ 0.6000 & 0.4000 \end{bmatrix}$	$\begin{bmatrix} 144.1527 & 0 \\ 0 & 74.1985 \end{bmatrix}$

Wireless Power Transmission by Magnetic Resonance Circuits

Shahram Javadi[1], Aliasghar Mohamedi[2]
Islamic Azad University, Central Tehran Branch
[1] Assistant Professor of Electrical Engineering Dept.
[2] Electrical Engineering Dept., IAU, Central Tehran Branch

Email[1]: sh.javadi@iauctb.ac.ir
Email[2]: aa.mohamedi@gmail.com

1. Abstract

Since the advent of electricity and start using it, the transmission from the source to the consumer has always been existed. Meanwhile, wireless power transmission was an issue that was in the minds of the scientists from the beginning of generation of electricity. Until now, many efforts have been made in this regard and scientists have discovered several methods so far.

The subject of this paper is to transmit the electricity wirelessly by magnetic resonance circuits. In this paper, first of all the performance of magnetic resonance method would be described, then the relevant circuits are studied in more details, and finally the effect of each quality parameters in this method of transmission is measured by some examinations.

Keywords: wireless power transmission, magnetic resonance

2. Introduction

The initial idea of wireless power transmission by magnetic resonance method was introduced by Nikola Tesla. In 1899, Tesla achieved a great progress by designing a resonance transformer that was recognized as Tesla coil and transmitted 100 million volts of electrical energy to a distance of 26 miles and lit 200 lamps and turned on an electric motor. Unfortunately, Tesla's experiments were halted due to lack of financial resources and he failed to finish his plans.

Despite the antiquity of this method of transmission is more than a century, but the scientists didn't do so much research in this regard for a long time. One of the reasons of that break was this method was not economically commodious. Because even with today's technology the highest yield were recorded in about 40%, which was conducted by researchers at MIT University and 60 W of electrical energy transmitted to a distance of about 2 meters and a 60 watts light bulb was lit. Today with the rapid growth of science and technology, this issue has become one of the up to date issues. More articles and studies published in this issue are related to after the year 2008 and in 2009 the number of these papers is considerable. Topics such as reliability, security, being cheaper than other methods of wireless power transmission and ... Increased attention of the researchers to this method. Transferring the energy is always embedded with issues like efficiency and quality. For further investigation in this regard, in this paper the influence of parameters such as operating frequency, the diameter of the transmitter and receiver coils, the diameter of the wire of the coils, the distance between the transmitter and the receiver coils and the coils turns will practically be evaluated on the performance of designed power transmission system.

3. Theory

3.1. The concept of magnetic resonance

In physics, resonance is the tendency of the system (usually a linear system) to oscillate with maximum amplitude at certain frequencies named resonant frequency or natural frequency. If you consider an oscillating system (like a swing), we can have the maximum amplitude if we blow it with a frequency equal to its natural frequency. The main function of this method is that two separate coils with a same resonant frequency can form a resonant system and have energy exchange, while the effect of coupling between the coils with different natural frequencies would be less, so this causes the effective distance range in the case of

resonant system would increase comparing to the non-resonant mode.

The effective performance distance in this manner, depending on the factors such as operating frequency, the accuracy rate for the circuits at the desired resonant frequency and the transmitter and receiver coils size, can increase to several times of the size of the coils.

3.2. The resonance circuit

Consider a circuit comprising a capacitor and an inductor. Each of the inductor and capacitor has impedance which can be calculated as follows:

$$Z_L = j\omega L \quad (1)$$

$$Z_C = \frac{1}{jC\omega} \;\; ; \;\; \omega = 2\pi f \quad (2)$$

Now if the working frequency of the circuit is such that the absolute value of Z_L and Z_C is the same, we can say that the circuit is at resonant mode and the working frequency of the circuit in this mode is called the resonant frequency. If we equate the absolute values of Z_L and Z_C we will have:

$$Z_L = Z_C \Longrightarrow j\omega L = \frac{1}{j\omega C} \Longrightarrow f = \frac{1}{2\pi\sqrt{LC}} \quad (3)$$

The last equation is used to calculate the resonant frequency of an LC circuit. When an LC circuit is in resonant mode, the inductor and the capacitor constantly exchange energy together. This energy is stored as electric field the capacitor and magnetic field in the inductor and constantly by charging and discharging their energy exchange causing an AC current flowing in the circuit. If the circuit doesn't disturb, the energy oscillation (due to the absence of losses) will continue. If we put a resistor in the circuit above, the energy exchange in the absence of a power supply will be damped and finally stops.

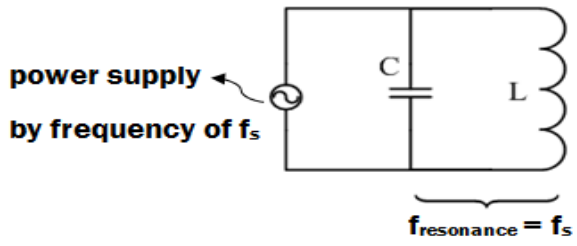


Figure (1): a resonant LC circuit

3.3. Wireless Power Transmission

In wireless power transmission by resonance method, we have two coils with a determined selfish capacity, each coupling to a capacitor with a determined capacity and meantime they have mutual inducting effects with each other. They play the role of our transmitter and receiver. One of the coils is connected to the AC power supply and the other is connected to the load. The performance of this system is that when the AC source is connected to the transmitter coil, an AC current is established. The AC current in the transmitter coil establishes an AC flux with the same frequency. The frequency of the source must be equated to resonant frequency of the transmitter and receiver circuit. Alternating flux produced by the transmitter coil, would be received by the receiver coil by mutual induction and since its frequency is equal to the resonant frequency of the receiver circuit, the maximum energy transfer amount occurs. Overview of such system is shown in figure (2).

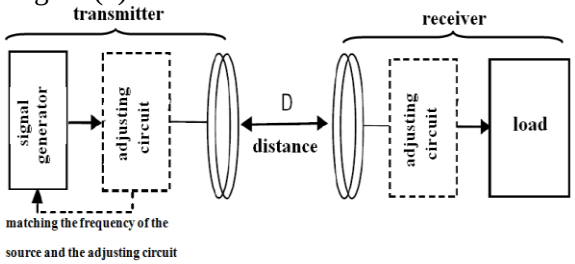


Figure (2): overview of the dismounted system

In the next section, the circuit of the system for the experiments is presented.

3.4. The overall circuit of the experiments

The overall circuit that was used for the experiments is shown in Figure (3).

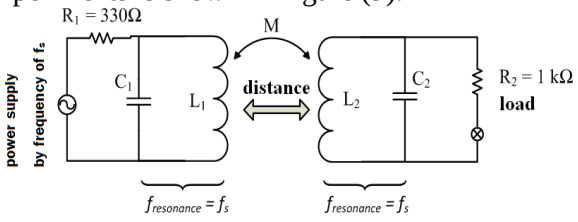


Figure (3): The overall circuit of the experiments

As shown in Figure (3), the amount of L_is and C_is (i = 1 , 2) must be configured such that the resulting resonance frequency is the frequency of the source.

4. Case Studies

In this section, the system that was closed during the tests and its theory was discussed in the previous section would be explained further.

First, a signal generator with a frequency of 1 MHZ and voltage amplitude of 23 V (peak-peak) was considered as a power supply of the circuit. The exact waveform of this signal generator is given in the test results. To dismount the circuits, a breadboard was used (one for the transmitter and one for the receiver) and any part of the circuit was mounted on the breadboard. For the transmitter and receiver, it is used of coils that were made of different materials at different dimensions which also are the selfish capacity of the circuit (Although the number of turns of coils was less but winding by a machine caused the coils to have a smooth shape and take the form of a loop). For the capacitors of the circuit, it is used of several capacitors to form a desired total capacitance. It is used of a digital oscilloscope to measure and capture the transmitting and receiving waveforms. For maintaining a constant distance between the coils and also having the coils concentric and to maintain the annular shape of the coils and have no geometrical distortion, it is used of a cardboard tube. The value of inductance, capacitance and resistance of the coils was measured using an RLC meter. Measuring the coils resistance was for calculating the power. The resistance of the coils comparing to the resistance which is placed in the circuit (either in the transmitter and the receiver) was very low, so it was ignored. (Coil resistance was about 0.1 to 0.2 ohms). In the receiver, it is used of a 1 kΩ resistance for the circuit load. It should also be added that changing the 1 kΩ resistance with the 1 MΩ, 8.2 MΩ, 16.4 MΩ and 32.8 MΩ, the amount of the received voltage in the output remained unchanged with good approximation. The amount of the capacitors and the inductors of both sides of transmitter and receiver were taken the same (I.e. in the experiments: C1 = C2 and L1 = L2).

Before presenting the results of the experiments, two points are necessary to express:

- ✓ The presented experiment is probably one of the first experiments in this regard in Iran.
- ✓ Improving the current circuit is part of future research goals.

The aim of the implemented experiments, was to measure the influence of parameters such as distance between the coils (d), the radius of the coils (r), the diameter of the wire used in the coils (D), the number of turns of the coils and power supply frequency (f) on the system performance.

Case 1:

Table (1): parameters amounts

Parameter	Amount
Frequency (f) - kHZ	1042
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	15
Distance between coils (d) – cm	50
Coil's self-inductances (L) - µH	16.1
Capacity of the capacitors (C) - nF	1.5

Figure (4) shows the overview of the dismounted system.

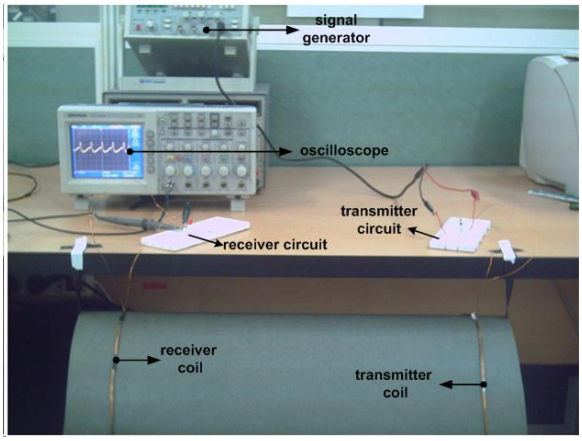


Figure (4): Overview of the dismounted system

The equation of the resonance frequency is:
(f_r is the resonance frequency of the circuit and f_s is the frequency of the power supply)

$$f_s = 1042000\ HZ \quad (4)$$

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{16.1 \times 10^{-6} \times 1.5 \times 10^{-9}}}$$

$$= 1024100\ HZ \quad (5)$$

This difference between the resonance frequency of the circuit, from the calculating methods and the one in the experiment, is for the reasons like not assuming the exact amount of equivalent impedance of the coils and the breadboard itself,

approximations like to assume the coils to be exactly the same and issues like this.

In the figures (5) to (8) the waveform of the power supply (signal generator) and the output waveform related to the terminals of the 1 kΩ resistance are shown.

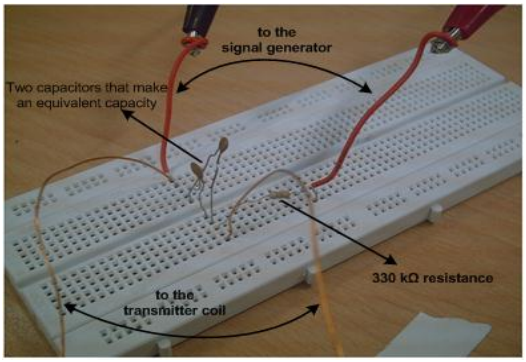


Figure (5): the transmitter circuit of case1

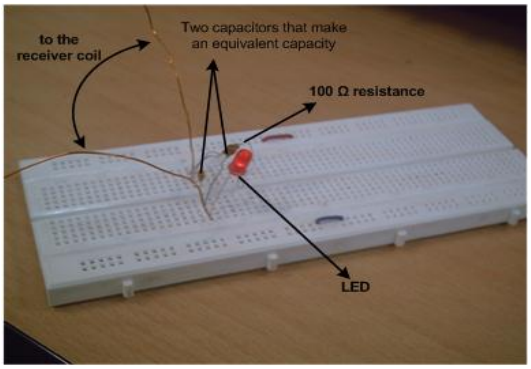


Figure (6): the receiver circuit of case1

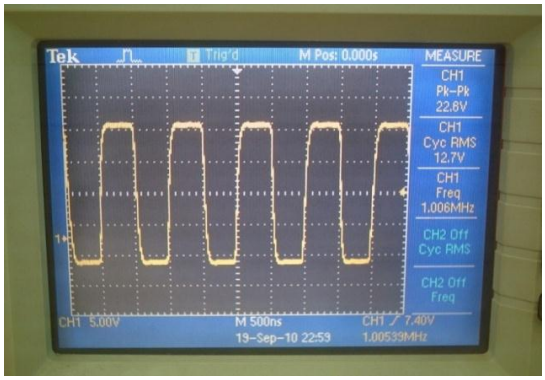


Figure (7): waveform of the signal generator

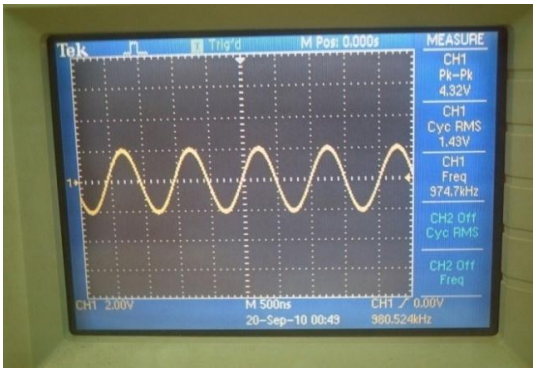


Figure (8): the output waveform

The measured quantities in this experiment are shown in the table (2):

Table (2): measured quantities of case1

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	7 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	1.43 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We have:

$$I_s = \frac{V_{R1}}{R_1} = \frac{7}{330} = 21.2\ mA \quad (6)$$

$$P_{in} = V_s.I_s.cos\varphi = 12.6 \times 21.2 \times (-1)$$

$$= 267.12\ mW \quad (7)$$

$$P_{R1} = \frac{V_{R1}^2}{R_1} = \frac{7^2}{330} = 148.5\ mW \quad (8)$$

$$P_{out} = P_{R2} = \frac{V_{R2}^2}{R_2} = \frac{1.43^2}{1\times10^3} = 2.1\ mW \quad (9)$$

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{2.1}{267.12-148.5} = 1.77\% \quad (10)$$

Case 2:

Table (3): parameters amounts of case2

Parameter	Amount
Frequency (f) - kHZ	1014
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	15
Distance between coils (d) – cm	25
Coil's self inductances (L) - µH	16.1
Capacity of the capacitors (C) - nF	1.5

The measured quantities in this experiment are shown in the table (4):

Table (4): measured quantities of case2

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	7 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	1.43 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{10.89}{236.73-116.5} = 9.06\% \quad (11)$$

Case 3:

Table (5): parameters amounts of case3

Parameter	Amount
Frequency (f) - kHZ	1050
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	10
Radius of the coils (r) – cm	15
Distance between coils (d) – cm	50
Coil's self inductances (L) - µH	93
Capacity of the capacitors (C) - nF	0.15

The measured quantities in this experiment are shown in the table (6):

Table (6): measured quantities of case3

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	7.12 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	1.71 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{2.924}{272.16-153.6} = 2.47\% \quad (12)$$

Case 4:

Table (7): parameters amounts of case4

Parameter	Amount
Frequency (f) - kHZ	1066
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	8
Distance between coils (d) – cm	25
Coil's self inductances (L) - µH	7.2
Capacity of the capacitors (C) - nF	3.1

In this experiment it is necessary to explain that with this coil with the diameter of 8 cm, when the distance between the coils was 50 cm, the amount of power loss was so that the received voltage at the receiver was very low and noisy. So, the distance was edited to 25 cm. The measured quantities in this experiment are shown in the table (8):

Table (8): measured quantities of case4

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	5.2 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	1.2 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We would have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{1.44}{199.08-81.9} = 1.23\% \quad (١٣)$$

Case 5:

Table (9): parameters amounts of case5

Parameter	Amount
Frequency (f) - kHZ	526
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	8
Distance between coils (d) – cm	5
Coil's self inductances (L) - µH	7.2
Capacity of the capacitors (C) - nF	12.4

The measured quantities in this experiment are shown in the table (10):

Table (10): measured quantities of case 5

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	6 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	4.33 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We would have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{18.75}{299.1-109.1} = 9.87\% \quad (14)$$

Case 6

Table (11): parameters amounts of case 6

Parameter	Amount
Frequency (f) - kHZ	2119
Cross section diameter of the wire (D) – mm	0.5
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	8
Distance between coils (d) – cm	5
Coil's self inductances (L) - µH	7.2
Capacity of the capacitors (C) - nF	0.75

The measured quantities in this experiment are shown in the table (12):

Table (12): measured quantities of case 6

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	6.2 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	6.81 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{46.38}{236.7-116.5} = 38.6\% \quad (15)$$

Case 7:

Table (13): parameters amounts of case7

Parameter	Amount
Frequency (f) - kHZ	1048
Cross section diameter of the wire (D) – mm	1.1
Number of turns of the coils (N)	4
Radius of the coils (r) – cm	15
Distance between coils (d) – cm	50
Coil's self inductances (L) - µH	15
Capacity of the capacitors (C) - nF	1.603

The measured quantities in this experiment are shown in the table (14):

Table (14): measured quantities of case7

Quantity	Amount
V_s (voltage of the signal generator)	12.6 V (RMS)
V_{R1} (voltage of the 330Ω resistor)	6.5 V (RMS)
V_{R2} (voltage of the 1 kΩ resistor)	1.48 V (RMS)
Phase difference of the voltage and the circuit of the source (φ)	180 degrees

We have:

$$\eta = \frac{P_{out}}{P_{in}-P_{R1}} = \frac{2.2}{248.22-128} = 1.83\% \quad (16)$$

5. Acknowledgment

The authors thank I.A.U., Central Tehran branch for all financial supports and helps to use laboratories.

6. Conclusion

As described before, the goal of performing the experiments in this paper is to evaluate the amount and the way of influence of the different parameters of the system on its performance. Table (15) shows the effect of the different parameters by comparing the experiments results.

Today, by the rapid improving of the technology, the issue of wireless power transmission, especially by the resonant circuits has become an up to date issue in the world.

In this paper, first of all, the concept of resonance had been explained, then the resonant circuits proposed, then some dismounted systems

in the laboratory were assessed and at last the effect of each of the parameters in the system has been evaluated. A brief summary of the experiments conclusion is shown in table (15).

Table (15) The effect of the different parameters by comparing the experiments results

Experiment No.	Evaluating parameter	conclusion
1 , 2	d (distance betwee the coils)	The distance between the coils in examination #1 was equal to 50 cm and in examination #2 decreased to 25 cm. we can see that the efficiency increased from 1.77% to 9.06%. This means by halving the distance, efficiency increased more than 5 times.
1 , 3	(N) number of turn of the coils	In the examination #1, the number of turns of the coils was equal to 4, but in the examination #3 it increased to 10. We can see that the efficiency increased from 1.77% to 2.47%. So, by increasing the number of turns of the coils, the efficiency increases by a normal coefficient.
1 , 4	(r) radius of the coil	The radius of the coils in the examination #1 was equal to 15 cm, but in the examination #4 was decreased to 8 cm. As we can see, it had a notably effect on the system because although the distance was decreased to 25 cm, the efficiency just had been equal to 1.23% (and even we didn't have the efficiency of examination #1)
5 , 6	(f) workin frequency	The working frequency of the examination #5 was equal to 526 kHZ that it increased to 2119 kHz in the examination #6. it can easily be seen that the working frequency has an enormous effect on the efficiency an increased it from 9.87% to 38.6% (that means more than 7 times)
1 , 7	(D) diameter of th wires	In the examination #7 the diameter of the wire was almost equal to two times of that in #1. By comparing the results we can see that increasing the wire diameter would just poorly increase the efficiency.

References

[1] R. Selvakumaran, W. Liu, B.H Soong, Luo Ming and S.Y Loon, "Design of Inductive Coil for Wireless Power Transfer", 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Singapore, July 14-17, 2009, pp. 584- 589

[2] C. Yu, R. Lu, Y. Mao, L. Ren, C. Zhu, "Research on the Model of Magnetic-Resonance Based Wireless Energy Transfer System", Vehicle Power and Propulsion Conference, 2009. VPPC '09. IEEE, 7-10 Sept. 2009, pp. 414 – 418

[3] S. Jalali Mazlouman, A. Mahanfar, B. Kaminska, "Mid-range Wireless Energy Transfer Using Inductive Resonance for

Wireless Sensors", Computer Design, 2009. ICCD 2009. IEEE International Conference, 4-7 Oct. 2009, pp. 517 – 522

[4] A. A. Trikolikar, S. L. Nalbalwar, M. P. Bhagat, "Review of Wireless Power Transmission by Using Strongly Coupled Magnetic Resonance", International Journal of Advanced Engineering & Applications, Jan. 2010, pp. 177-181

[5] C. Zhu, K. Liu, C. Yu, R. Ma, H. Cheng, "Simulation and Experimental Analysis on Wireless Energy Transfer Based on Magnetic Resonances", IEEE

Vehicle Power and Propulsion Conference (VPPC), September 3-5, 2008, Harbin, China

Deploying and configuring wireless mesh network in coexistence of highly interfering wireless LANs

Lamia Romdhani and Amr Mohamed

Computer Science and Engineering Department,
Qatar University, Doha, Qatar
Email: lamia.romdhani@qu.edu.qa and amrm@qu.edu.qa

Abstract: *In wireless mesh network, environmental conditions and location of "nodes" on the network are largely unpredictable and in a constant state of change. Heterogeneous signal propagation patterns in addition to interference from existing infrastructure pose significant challenges for a wireless mesh network design. In this paper, we explore, by experiments, link performance results produced at Qatar University (QU) wireless mesh test-bed in the coexistence of high interfering wireless transmission. We first perform signal strength measurements for path loss, shadowing, and fading at 2.4/5 GHz frequencies. Then, we investigate the effect of unique cell structure, and heavy lab machinery of the engineering building on the link performance. Furthermore, we study different wireless mesh node configurations and equipments' effect on performance and network connectivity. The causes behind link instability and performance shortcoming of the QU wireless mesh network (QUMESH) are identified. Our analysis reveals that the link performance degradation is caused by high interference with wireless LANs, and special QU College of engineering structure. Our work provides insights to network designers to realize high-performance wireless networks by considering the adequate configuration, equipments, and addressing the adaptive transmission approaches.*

Key Words**:** *Testbed, Wireless Mesh Networks, Real Experimentation, Harsh Environment, Wi-Fi Interference.*

I. INTRODUCTION

Wireless Mesh Networks (WMNs) consist of mesh routers and mesh client nodes. A mesh network testbed allows us to study how real applications perform in such networks. Besides to routing capability for gateway/repeater functions as in a conventional router, a mesh router contains additional routing functions to support mesh networking. To further improve the flexibility of mesh networking, a mesh router is usually equipped with multiple wireless interfaces built on either the same or different wireless access technologies. Compared to traditional wireless access point, a wireless mesh network can achieve the same coverage with much lower transmission power per mesh node through multihop communication. Optionally, the medium access control protocol in a mesh router is enhanced for better scalability in a multi-hop mesh environment. A mesh network testbed is critical to identifying fundamental technical problems in the mesh networking technology and validating potential solutions, as it is not possible to accurately capture the behavior of the entire network in theory or in simulators. For example, an accurate interference characterization is a prerequisite and key to designing high performance scheduling, routing, and rate control schemes. However, it is generally difficult to model wireless signal propagation and interference accurately for a given deployment of mesh network, which can potentially change constantly due to environment changes [8]. Thus, the coexistence of wireless mesh and wireless LANs poses one of the biggest design challenges.

In this paper, we provide a detailed framework of a wireless mesh network testbed deployment and configuring in coexistence of other wireless LANs on the QU campus, in the college of engineering. We present the site-specific signal strength measurement results for path loss, shadowing, and fading leveraging isolation between the WMN and existing WLAN infrastructure onsite. First, we performed a set of physical measurements to characterize the indoor channel for 802.11a/g wireless local area networks at 2.4/5 GHz frequencies. We used spectrum analyzer Rohde & Schwarz FSH8 and InSSIDer (free software for WLANs). Second, we picked up deployment locations for individual mesh routers as well as the gateway(s) to balance the connectivity and maximize the coverage and throughput. We proposed a 2-tier network architecture that optimizes the outdoor connectivity across the college cells, enhances the indoor coverage, and broadens the service provisioning for various types of end devices. We also investigated various environmental phenomena, impact, and issues on QUMESH deployment. Our studies show that, high performance wireless networking is affected by a number of internal and external variables. The interference with other Wi-Fi LANs, poses a significant challenge to the mesh network deployment, as well as node to node communication efficiency. We observed severe performance degradation especially over multihop wireless mesh transmission. The building structure,

antenna selection, and their positions also have a significant impact on the link quality performance. We study the effects of these parameters and analyze the experimental results obtained under various network scenarios with different environmental conditions. Furthermore, we developed several management and configuration tools that allow users to configure, program, troubleshoot, interact with, and receive output data from nodes in the network, filling a gap in current testbed management solutions. The aim of these studies is to provide guidelines to setup a wireless mesh testbed yielding to a significant improvement in network connectivity and overall application performance.

The remainder of this paper is organized as follows. We devote Section 2 for reviewing some related works that have been done so far to perform real testbed experimentations. In Section 3, our research work's motivations are given. The QU wireless mesh network architecture will be described in Section 4. We provide a deeper analysis of the main obtained experimental results in Section 5. Section 6 summarizes the paper and outlines the future works.

II. Related Works

In the last decade several wireless mesh network testbeds have been setup at different locations, mostly in the educational campuses. Bicket et al. [1] evaluated a 37-node 802.11b community mesh network over an area of approximately four square kilometers in Cambridge, Massachusetts. The MIT mesh network (Roofnet), adopts off-the-shelf equipment, e.g. IEEE 802.11 wireless cards and standard Omni-directional antennas. The authors evaluated multiple aspects of the architecture such as the effect of node density on connectivity and throughput as well as the characteristics of wireless links. Gambiroza et al. [12] simulated a multihop wireless backhaul network consisting of multiple Transit Access Points (TAPs), which are connected to the Internet through multiple entry points. They studied TCP/UDP fairness, while considering different parameters such as the role of the link layer protocol, antenna technology, and traffic types. Based on the findings in [12], Camp et al. [13] deployed a two-tier mesh network in Houston, Texas, that aims at providing Internet access over a wide area with minimal infrastructure. The deployed network comprises an access tier and a backhaul tier. The access tier connects mobile clients with mesh nodes, whereas the backhaul tier interconnects the mesh nodes and forwards traffic to and from the Internet. Using this network, the authors presented a measurement driven deployment strategy and a data driven model to study the impact of design and topology decisions on network-wide performance. In [15], Raniwala et al. proposed a dual-radio wireless mesh network comprising 9 PC nodes, each equipped with two IEEE 802.11a interfaces. The authors show that, by employing sophisticated channel assignment approaches, network throughput can be significantly improved. De et al. [9] proposed a mobile 12-node experimentation testbed for multihop wireless networks. Each node in the testbed comprises a wireless computing device and a mobile robot. Fixed signal attenuators are used to limit the transmission range of the mobile nodes. In [11], Eriksson et al. evaluated the feasibility of an all-wireless office mesh network consisting of 21 multi-radio mesh nodes. The authors captured user traffic on office PCs with wired Ethernet connectivity and replayed them on the mesh network. A set of parameters, such as different routing metrics and hardware settings were evaluated. Raychaudhuri et al. [16] proposed an open access research testbed called Orbit for evaluating next-generation wireless network protocols. The testbed consists of an indoor radio grid emulator for controlled experiments and outdoor field trial software for end user evaluations. Lundgren et al. reported in [14] on their experience in designing and deploying the UCSB MeshNet, a 30-node wireless mesh testbed which covers several floors inside a building. S.M. ElRakabawy et al. presented ScaleMesh [10], a 20-node scalable dual-radio wireless mesh testbed based on IEEE 802.11b/g technology. Using ScaleMesh, large-scale mesh networks can be emulated within a miniaturized experimentation area by using variable signal attenuators. Our contributions in this work include design and deployment of wireless mesh testbed under some unique characteristics: special cell structure and heavy lab machinery of the engineering building. We also, provide a detailed study and analysis of the coexistence of QUMESH and wireless LANs impact on the wireless mesh network performance. Moreover, we developed different configuration tools and scripts that enable a centralized network management.

III. Motivations

First experiments performed at QUMESH, which is mainly in the focus of research, show that links constantly disconnect and reconnect. Therefore, this issue motivates us to investigate the causes behind this instability. Our focus in particular is to provide a framework for articulating and investigating the issues and questions related to QUMESH deployment and configuring in the coexistence with high interfering wireless transmission. We aim to study the link performance under various environment conditions. This way, we expect to be able to select appropriate configuration parameters of wireless mesh nodes and address adaptive mechanisms and schemes to overcome the observed issues.

IV. QU Mesh Architecture

Each mesh node is a Dell desktop PC (specs) equipped with a Wistron Neweb CM9 Atheros 802.11a/b/g/n Dualband mPCI 5004 chipset wireless card with two external antennas (see Figure 1). The mesh nodes run Mandrake 10.2 with kernel version 2.6.11-6mdksmp and the open-source madwifi-0.9.4 driver is used to enable the wireless cards. All wireless cards operate in ad-hoc mode. All nodes support a variety of ad hoc routing protocols, such as OLSR and AODV, in addition to preparatory routing protocols [24]. Each wireless node further possesses a Gigabit Ethernet NIC, which is connected to the subnet of the QU through a Gigabit switch. This

allows for the remote management of the wireless nodes from any wired host in the subnet. Hence, wireless experiments can be managed from a remote computer and traces can be analyzed through the wired network. The IEEE 802.11b/g/n standard supports 11 different channels. According to the IEEE 802.11 specifications [17], channels 1, 6, and 11 are non-overlapping. However, in practice, non-overlapping channels strongly depend on the vendor of the corresponding network cards and may strongly vary. We have used the different channels (from 2.402 GHz to 2.483 GHz) with maximum transmit power of 19dBm, and data rate of 2Mbps.

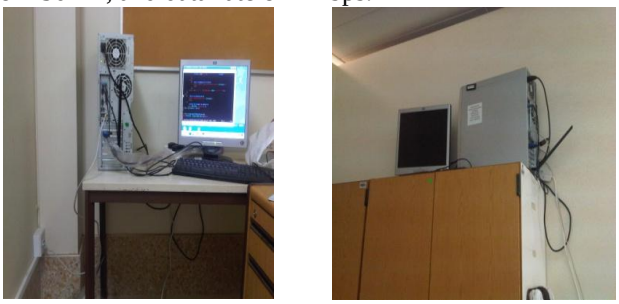


Figure 1: QUMESH nodes

A. Mesh Nodes Deployment

Environment description and challenges: The building structure of QU College of engineering is based on an irregular grid of small unconnected two-floor buildings (commonly called cell) with indoor areas separated by outdoor corridors. The detailed structure and building layout is depicted in Figure 2.

Such building architecture makes it challenging for deploying the mesh nodes because, on one hand, we cannot deploy mesh nodes in outdoor corridors for security, on the other hand, the mixed indoor/outdoor nature inherent in the building structures makes wireless channel impairments more evident due to changes in temperature between inside and outside the [*25*]. Moreover, more than 128 WLAN access points are deployed in 60 cells throughout the college, supporting 802.11g standards. The access points send frequent beacons in idle mode and have automatic channel settings, which allow APs to change channel setting dynamically. Such AP configuration makes hard to configure the mesh nodes.

Regarding the various environmental issues described above, we have conducted a set of measurements to characterize the wireless channel propagation during intra-cell and inter-cell communication [23]. These measurements helped us to identify the adequate placement of the QUMESH nodes.

Based on these observations and previous measurements, we have adopted a phased approach for deploying mesh nodes [25]. These studies helped to increase the QUMESH connectivity and connect the isolated nodes to the mesh network as shown in Figure 2.

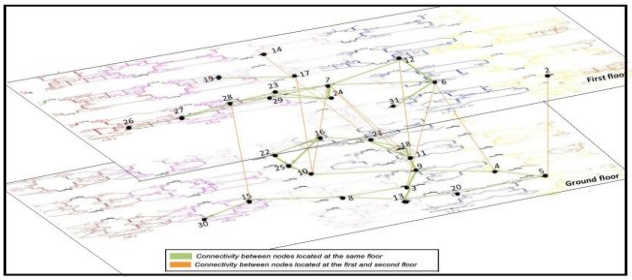


Figure 2 : Nodes connectivity in the QUMESH

B. QUMESH tools

The troubleshooting and maintenance of the testbed become a challenging job, especially, as the number of nodes increases. Thus, designing tools for automating these functionalities become evident. These tools support the following major functionalities:

QUMESH configuration tool*:*

- Automatic configuration of the wireless MadiWiFi driver on all the mesh nodes.
- Automatic configuration of the static
- Query and compare the configuration of all mesh nodes for troubleshooting.

QUMESH link state tool: This tool provides an automatic measuring of wireless link quality to get a snapshot of the testbed link state during the performance of research studies.

Detailed Analysis and Findings

In this section, we report the results of the extensive experiment sets that have been performed at QUMESH test-bed. In the first set of experiments, we study the physical measurements that have been done to characterize the indoor channel for 802.11a/g wireless local area networks at 2.4/5 GHz frequencies, regarding node location [23]. We used spectrum analyzer Rohde & Schwarz FSH8 and InSSIDer (free software for WLANs). Measurements are taken in indoor and outdoor environments at various locations on different times of the day. In the second set of experiments, we examine impacts of wireless mesh node equipment on the connectivity and packet loss ratio. Especially, we investigate the antenna selection and the their best position. In the third set of experiments, we observe the effect of WLAN interference and per channel load on the link performance. We describe results of the measured link in terms of packet loss and link latency metrics. Then, we investigate multihop wireless transmission performance at QUMESH. Finally, we provide a deeper analysis of the main test results obtained.

The experiments reported in this paper are performed with no control on the WLAN APs, which poses different levels of intereference at different parts of the Mesh network testbed. The measurements were performed using the link state tool based on the "ICMP" [22] to quantify the link quality based on physical, MAC, and network layer settings. The packet size is equal to1500 bytes. Each plotted point in the presented graphs, is the average of 30 experiment iterations. During one iteration, 100 packets are sent from source to destination mesh nodes.

C. Physical layer measurements

For configuring the mesh nodes and optimizing their location at QUMESH, we have adopted some physical layer measurement approaches [23].

2.4 GHz spectrum management: With only 3 non-overlapping channels in the 2.4 GHz band, there isn't much flexibility in AP and antenna placement. If we look at Figure 3, we can see how the 2.4 GHz spectrum is divided into three channels for 802.11b and 802.11g. However, there is a "cheat" for the 2.4 GHz spectrum in which four channels are jammed closer together to minimize interference; this method is shown in Figure 4.

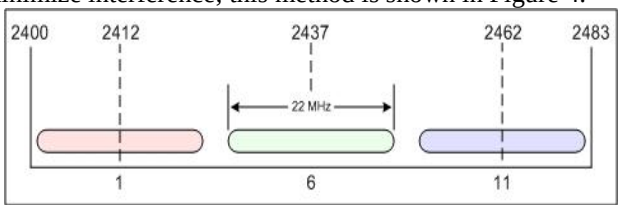


Figure 3: 3 non-overlapping channels in 2.4 GHz band

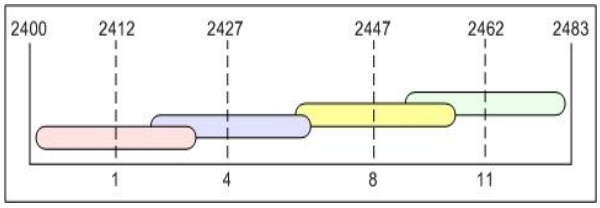


Figure 4: Four channels are jammed closer together to minimize interference

Measurement Procedure:

We aimed to perform signal strength measurements for path loss, shadowing, and fading at 2.4/5 GHz frequencies. The major problem we faced during the signal strength measurement was the interference from large number of university's access points (APs). Since the spectrum analyzer, which is a physical layer device can't differentiate signals from different access points on the basis of MAC address or AP's service set identifier (SSID), we used inSSIDer software for measurements. The purpose of this study is to characterize the indoor channel for 802.11a/g wireless local area networks based on the path loss model described in [23]. The average path loss at distance d between transmitter and receiver is given by:

$$PL(dB) = PL(d_0) + 10n \log \frac{d}{d_0}$$

Where n is the path loss exponent, and d_0 is the close-in reference distance. We observed the signal levels at different distance intervals against distance ratio $\frac{d}{d_0}$.

Knowing the channel statistical models enables designer of communication systems to devise methods to combat the fading effect. Figure 5 shows a typical fading pattern observed between two nodes placed closed to each other within the Rsearch Lab at QU.

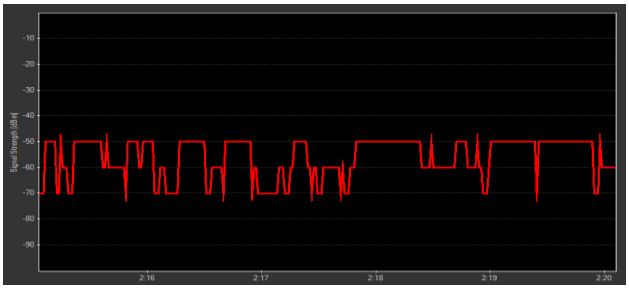


Figure 5: Signal fluctuation with time

D. Impact of node's physical equipment on link performance

Wireless network antennas typically perform best in open environments with minimum possible obstructions. The performance may be significantly affected if antennas are mounted near metal surfaces, concrete walls, or other high-density materials, which is the case of our QUMESH test-bed.

In the first phase of deployment, all wireless mesh nodes were equipped with a single 5dbi antenna *(2.4 GHz 5 dBi Rubber Duck Omni Antenna RP-TNC)*. We used the measurement tools to measure the performance of the QUMESH links during different times during the day. The results showed that some nodes are isolated and so, can't communicate with any other node. In addition, there are some links that go up and down intermittently.

To overcome this problem, we equipped nodes with 7dbi antenna *(2.4 Ghz 7dBi Rubber Duck Omni Antenna RP-TNC)* with the objective of increasing the link range and hence increase stability. However, we have found that nodes are still isolated and links were going down more frequently. We report the antenna patterns, which explain the interesting results obtained, depicted in Figures 6, 7, and 8. From these figures, we show that antenna selection is a critical step towards maintaining link stability because increasing the antenna gain may not be homogeneous across all directions, and hence may affect the link performance drastically. This is clear from Figures 6, and 7 for the 5, and 7 dbi antennas. However, the 8 dbi antenna *(8dBi 2.4GHz - ARS-N18 ALFA NETWORK)* seems to be truly omni-directional and hence, the direction does not affect the link performance as shown in Figure 7. To investigate more on this issue, we performed the following experiments.

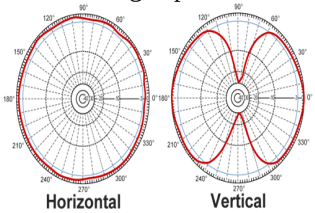


Figure 6: 5 dbi pattern

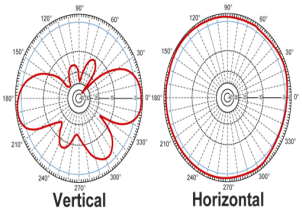


Figure 7: 7 dbi pattern

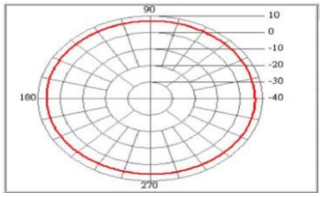


Figure 8: 8 dbi pattern

We placed two wireless mesh nodes close to each other in an indoor cell with no walls or obstructions in the range. The distance between the two nodes is 7 meters. We conducted experiments using 5 dbi,7 dbi, and 8 dbi antennas with different antenna directions.

The results seem to confirm our initial observations. Indeed, when using 7dbi antennas the direction of the antennas affects the link performance, as shown in Figure 10. The link goes down almost consistently when placed the antenna at 45° direction. However, with 5dbi antennas, the direction does not affect the link performance as shown in Figure 11. Moreover, the performance results obtained with 8 dbi antenna is similar to 5dbi as shown in Figure 9 since the two nodes are close to each other.

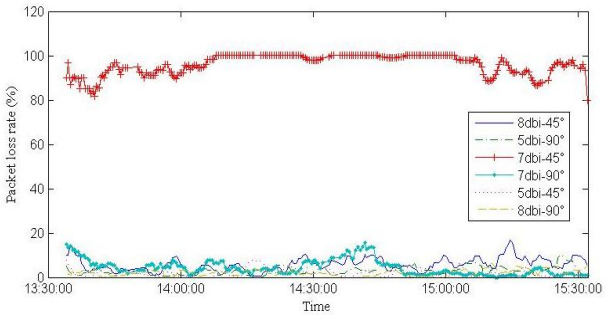
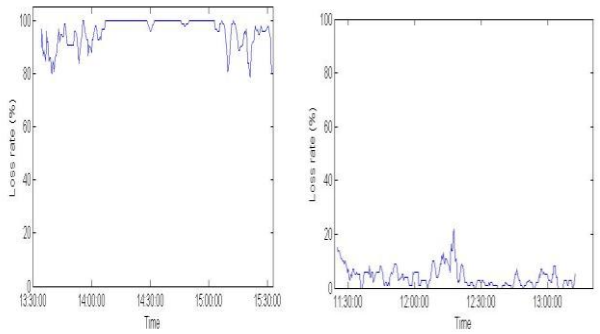
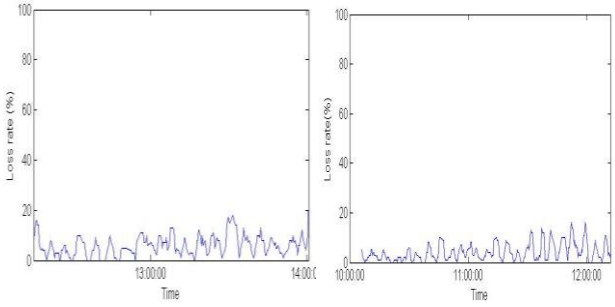


Figure 9: Loss rate results regarding different antennas positions



(a)antenna placed at 45 degree position (b)antenna placed at 90 degree position

Figure 10: Loss rate obtained with 7dbi antennas placed on different positions



(a)antenna placed at 45 degree (b)antenna placed at 90 degree

Figure 11: Loss rate obtained with 5dbi antennas placed at different positions

In this experiment, we study the link performance, using 5dbi and 8 dbi antennas, between nodes located at different floors.. We aim by this way to study the effect of this equipment on link quality in the presence of heavy building structure in the addition to WLAN interference.

Figure 12 illustrates the effect of the new used antenna on link performance, in terms of packet loss rate observed between isolated wireless mesh nodes 5 (located at ground floor) and 2(located at first floor) (when using 5dbi antennas) on channel 9. With 5dbi antennas, the packet loss rate exceeds 90 % for long time periods, during which the connectivity between the two nodes shows a significant degradation. However, using 8dbi Alpha antenna the packet loss rate remains lower than 20 % during all testing period which guarantees a good connectivity between the two nodes.

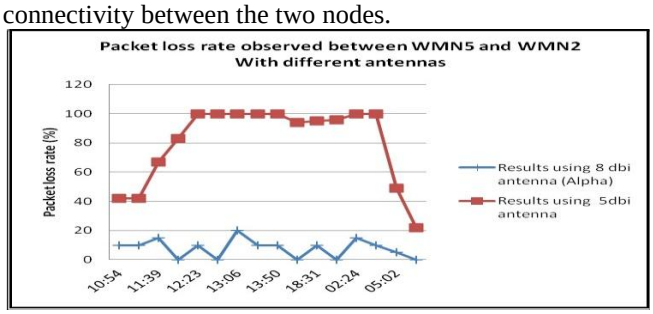


Figure 12: Packet loss rate with different antennas

The obtained experiment results show that link between wireless nodes using 5dbi antenna disconnect and connect frequently. However, with 8dbi (Alpha) antennas, we get more stable links. It also increases the reachability of all nodes to each other and so, enhance the network connectivity. Note that for the different experiment results presented in this paper, only 8dbi antennas are being used.

E. Building structure effects on link quality

Measurements carried out in various large structures, quantify both radio-signal attenuation and distortion (multipath) in the radio propagation channel, and showed that there are serious engineering college building structure issues. We devote this Subsection to disscuss these issues.

We conducted multichannel experiments considering indoor links, connecting intra-floor and inter-floor mesh nodes. We aim to highlight the impact of the different frequency channels on both scenarios.

Packet loss rate is depicted in Figure 13. The rate observed between nodes located at different floors is higher than the packet loss rate observed for nodes located at the same floor.

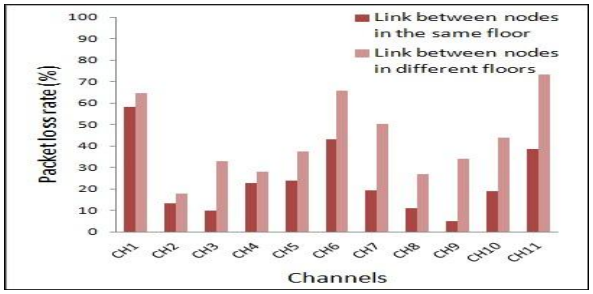


Figure 13 : Building structure effects on link quality

Although results seem to be evident, the gaps between different channel performance are large. For nodes in the same floor, the packet loss rate varies from 2% to 42 %. However, for nodes, in different floors, the loss percentage varies from 18% to 80%. Thus, one important

conclusion is that some channels are able to mitigate more on the building problem, than other channels. This behavior would be perceived as an opportunity for addressing new mechanisms to come up with the observed issues and enhance the link quality based on an inteligent channel assignement.

F. Wireless LAN traffic load effect

Due to the strong co-existence (i.e. 2.4GHz spectrum sharing) of the QUMESH testbed and the wireless LAN of QU, it has been shown; that the interference caused by the university WLAN causes the mesh network links instability. Also, it has been identified that WLAN access points are configured to jump across non-overlapping channels intermittently. Furthermore, our experiment results showed that some channels exhibit higher interference than others. This behavior creates restrictions on what channels can be used on which link.

Therefore, we use Kismet tool [21] to sniff the WLAN traffic load observed for different frequency channels (1, 9, and 11) during three days, one day for each channel. Kismet tool detects the presence of wireless networks, including those with hidden SSIDs. In parallel, we study the performance behaviour of a QUMESH link considering different channel settings. Then, we analyse the Kismet logs to identify the overhead transmitted over each channel.

To evaluate link quality, we consider an indoor mesh link connecting the wireless mesh nodes 4 and 9. We enable communication between the two nodes during 72 hours. In the first day, the nodes communicated over channel 1, the second day they used channel 9, and the third day channel 11 was assigned to these nodes.

Figure 14 shows the behaviour of the overhead that is being transmitted by WLAN traffic over the frequency channels 6, 9, and 11. Figure 15 shows the packet loss results observed for one mesh link using channels 1, 11, and 9. Figure 16 shows the obtained average delay using the three channels. From the plots, we observe that the highest overhead is transmitted over channel 1. Thus, the packet loss rate measured for this channel is higher than the packet loss rate measured for channel 9, which has the lowest overhead among the three channels, as shown in Figure 14.

Furthermore, the huge amount of WLAN traffic load can not only severely degrade the performance of a wireless mesh, but also could affect the network connectivity. For example, we observed that the packet loss rate reaches more than 90 % using channel 1, caused by high interference, which leads to a complete link failure.

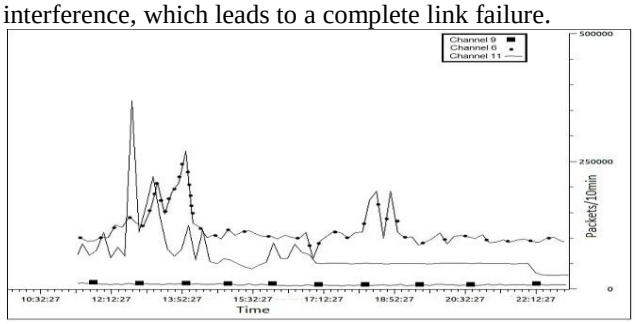


Figure 14: WLAN traffic load over different channels

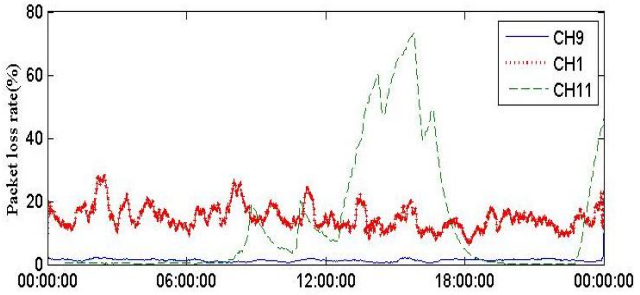


Figure 15: Packet Losses observed over different channels between nodes 6 and 7

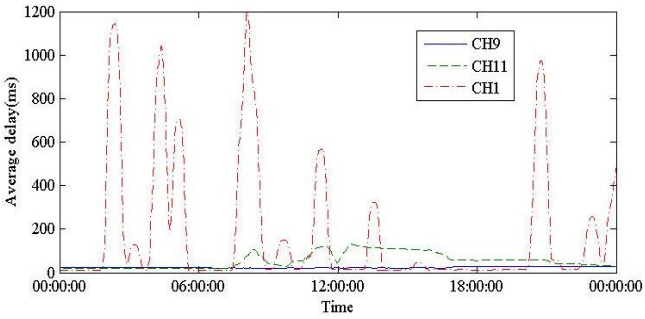


Figure 16: Average delay observed for different channels between nodes 6 and 7

G. Multi-hop wireless transmission performance at QUMESH

We performed a set of experiments to evaluate the end-to-end path performance when considering multi-hop transmission.

Figure **17** shows that, the problem becomes more complicated in multi-hop network, due partly to inter-hop interference and rate-hop count tradeoff. We observe that the packet loss rate increases, as the number of hops to travel increases. The connection between source mesh node and destination mesh node is constantly connecting and disconnecting when the number of hops is greater than 4. Moreover, increasing the number of hops increases the average end-to-end delay as shown in Figure 18. The results obtained using channel 9 outperforms the results obtained using channels 1, 11, and 3. The rapid degradation of multihop wireless transmission performance results poses a significant challenge, especially, for multimedia streaming and delay sensitive applications.

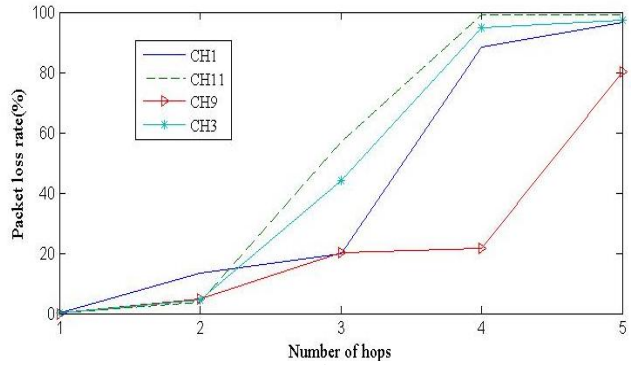


Figure 17: Packet Loss rate for muti-hop transmission

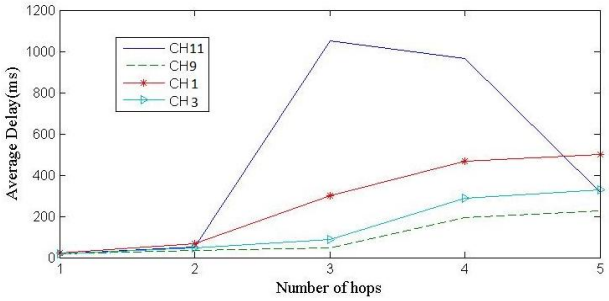


Figure 18: Average delay for multi-hop transmission

We concluded that mesh node should be configured to use the intermediate overlapping channels opportunistically. Hence, QUMESH node implementations should rely on channel switching capability of the wireless radios to ensure network connectivity and enhance real-time application performance. Such approach calls for cognitive schemes that sense and perform dynamic channel assignment to mitigate un-avoided interference. However, dynamic channel switching at mesh node level incurs overhead in terms of switching delay due to both software and hardware restrictions [18]. This can be prohibitive for many delay sensitive, real-time applications. The situation can be worse in the case of a multihop network, as every node along the traffic path may require a channel switch that adds up to the overall end-to-end delay. Therefore, it is extremely interesting to note that, from our experiment results, QUMESH outdoor/indoor links do not need to switch channel too frequently for data transmissions. For QUMESH indoor links, the channel allocation policy that should be addressed, has to consider also an efficient cross-layer routing strategies that can make use of the flexibilities of a multichannel network while favoring delay sensitive applications by routing them on low delay paths.

V. CONCLUSION

In this framework, we propose characterization studies in deployment and performance evaluation of wireless mesh network in coexistence of other wireless LANs. We described the unique aspects of our QUMESH testbed that differentiate it from other existing wireless mesh testbeds. We intensively studied the link quality performance. Specifically, we provided insight into the runtime issues from the point of view of WLAN traffic load that poses significant adverse effect on QUMESH performance and connectivity. Furthermore, we investigated the impacts of special building structure of the engineering college.

To mitigate these challenges, efficient mechanisms are needed to handle the coexistence issues. We recommend the use of intermediate channels (i.e. channels 9 and/or 4) in order to overcome the Wi-Fi interference problems. Furthermore, we should address dynamic channel assignment technique to cognitively change the channel based on sensing the channel quality. Moreover, selection and proper use of the right antenna can dramatically improve the network connectivity. More importantly, it is necessary to address more intelligent upper layer protocols to mitigate link instability and optimize performance.

In our future works, we aim to perform several evaluation protocols and address cross-layer mechanisms to enhance mesh network connectivity and performance.

ACKNOWLEDGMENT*:* This work is supported by Qatar National Research Fund (QNRF) No.08-374-2-144.

REFERENCES

[1] J. Bicket and al. Architecture and Evaluation of an Unplanned 802.11b Mesh Network. MOBICOM, 2005.
[2] Broadband and Wireless Network Lab Wireless Mesh Network.
[3] Microsoft Mesh Networks.
[4] MIT Roofnet: http://pdos.lcs.mit.edu/roofnet.
[5] Purdue University Wireless Mesh Network Testbed.
[6] TFA Rice University: http://tfa.rice.edu/.
[7] Bastian Blywis, and al. Trends, Advances, and Challenges in Testbed-based Wireless Mesh Network Research. Springer Journal of Mobile, Networks and Applications, 15(3):315-329, 2010.
[8] S. Das, and al. Understanding Wireless Routing Link Metrics Dynamics. ACM SIGCOMM/USENIX 2007.
[9] P. De, and al. MiNT-m: An Autonomous Mobile Wireless Experimentation Platform. Proc. ACM MobiSys, 2006.
[10] S. ElRakabawy and al. ScaleMesh: A Scalable Dual-Radio Wireless Mesh Testbed. SECON Workshops '08., San Francisco, CA, 2008.
[11] J. Eriksson and al. Feasibility Study of Mesh Networks for All-wireless Offices. Proc. ACM MobiSys, Sweden, 2006.
[12] V. Gambiroza, and al. End-to-End Performance and Fairness in Multihop Wireless Backhaul Networks. ACM MOBICOM, Philadelphia, PA, 2004.
[13] J. Camp, J. Robinson, C. Steger, and E. Knightly. Measurement Driven Deployment of a Two-Tier Urban Mesh Access Network. Proc. ACM MobiSys, 2006.
[14] H. Lundgren and al. Experiences from the Design, Deployment, and Usage of the UCSB MeshNet testbed. IEEE Transaction on Wireless Communications, 13(2), 2006.
[15] A. Raniwala and T. al. Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network. Proc. IEEE INFOCOM, Miami, 2005.
[16] D. Raychaudhuri, and al. Overview of the ORBIT Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols. WCNC, 2005.
[17] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification. 1999.
[18] P. Kyasanur, C. and al. Net-x: System extensions for supporting multiple channels, multiple interfaces, and other interface capabilities (2006)
[19] Gregor N. Purdy. Linux iptables Pocket Reference, O'Reilly Media: Sebastopol, CA,2004.
[20] Taleb Moazzeni, 'A Wireless Propagation Channel Model with Meteorological Quantities Using Neural Networks', IEEE CCNC 2006.
[21] http://www.kismetwireless.net/
[22] J. Postel, 'Internet Control Message Protocol' RFC 792, Sept. 1981. http://www.rfc-editor.org/rfc/rfc792.txt.
[23] I. Ahmed, and al. 'Characterization of the Indoor/outdoor Radio Propagation Channel at at 2.4 GHz'. IEEE GCC February 19-22, 2011, Dubai, UAE.
[24] D. Koutsonikolas, and al. 'Pacifier: High-Throughput, Reliable Multicast without "Crying Babies" in Wireless Mesh Networks', IEEE INFOCOM 2010.
[25] L. Romdhani, and al.' QUMESH: Wireless mesh network deployment and configuration in harsh environment', accepted in IEEE WCNC 2012 Paris, France.

SESSION

NOVEL ALGORITHMS AND PROTOCOLS + ROUTING

Chair(s)

TBA

Leveled Indoor Localization Algorithms Based on Passive RFID

Matthew Chan [1] **and Xiaowen Zhang** [1,2]
[1] Computer Science Dept., Graduate Center, CUNY
365 Fifth Ave., New York, NY 10016, U.S.A.
[2] Computer Science Dept., College of Staten Island, CUNY
2800 Victory Blvd, Staten Island, NY 10314, U.S.A.
Corresponding Email: xiaowen.zhang@csi.cuny.edu

Abstract— *We propose three indoor localization algorithms, namely the leveled nearest-neighbor, leveled multilateration and leveled Bayesian inference, to locate stationary objects (books, merchandise) by affixing passive RFID tags to them. The algorithms use a number of passive tags deployed on the known coordinates as reference points. The simulation results show that the proposed leveled detectable count RFID localization algorithms produce superb accuracy performance.*

Keywords: Localization algorithm, passive RFID, simulation.

1. Introduction

Radio Frequency Identification (RFID) has been widely utilized for the automatic identification, inventory and tracking of animals, pharmaceuticals, supply chains, merchandise, library books, and other objects. The unique identification information of any objects can be stored economically and retrieved without requiring physical contact or line-of-sight (unlike barcodes). The RFID systems generally offer coarse-grained location information about tagged objects; they can tell whether the object is present or absent in the proximity. However, more accurate location information of objects could enhance these applications, and enable a large number of other applications. For example, in a large warehouse, large library, hospital, or smart home, how to quickly and accurately locate certain objects, books, or patients, in a cost-effective manner, is highly desirable.

Locating and positioning of (moving) objects and persons are popular and important. Localizing techniques can be classified into RF-based (Radio Frequency) and non-RF-based [25]. GPS (Global Positioning System) [23] is a successful RF-based method for positioning outdoor objects. The system is based on two-dozen satellites and a few ground monitoring centers. Four or more satellites broadcast the RF signals to a GPS receiver, then it uses TOA (time-of-arrival) of signals to calculate the 3D geographic position information (latitude, longitude, and altitude) of the receiver. However, GPS receivers in general do not work in indoor settings because the satellite signals are heavily attenuated by the building walls and floors. RF-based localization approaches also include wireless local area network (WLAN), wireless sensor network (WSN), and RFID localizations. In WLAN localization, a WLAN device can be localized based on beacon signal strength information between the device and access points. Some of the major studies on WLAN localization can be found in [5], [19], [18], [24]. Generally speaking, WSN localization techniques are based on AOA (angle-of-arrival), TDOA (time-difference-of-arrival), and RSS (received signal strength). WSN localization is an important area of research and many studies have been reviewed in [2], [12], [15]. RFID localization techniques are surveyed in [25], [1], [7]. They are similar to other RF-based localizations in principle; the main advantage over WLAN and WSN approaches is the inexpensiveness and ease of deployment [25]. We will address RFID localization in detail in the subsequent sections. Non-RF-based localization techniques employ acoustic, vision, ultrasound [16], infrared [22] and laser as measuring signal. These techniques typically are built on TOA and triangulation or multilateration algorithms.

1.1 An overview of RFID technology

RFID, as an enabler of the Internet of Things (IoT), plays an important role for identifying an object and for providing coarse-grained location and other information of the object. Together with the Internet infrastructure it makes connections and communications from everything to everything else at anytime and from anywhere possible. An RFID system consists of three major components [9], [17]: the RFID tag (also known as a transponder), the RFID reader (transceiver, or interrogator), and the backend server. The tag can be passive (a small antenna connected to a microchip without integrated power source, which harvests energy emitted by the reader to operate and communicate with the reader by backscattering), active (powered by a internal battery, it transmits an RF signal in response to the reader, rather than backscattering the reader's signal), or semi-passive (using a battery to power up the microchip and also using backscatter to communicate with the reader). The reader is located between the tag and backend server. The reader sends and receives information to and from the tag and communicates with (updates) the backend server. The backend server, at the highest level of RFID infrastructure, runs applications, hosts databases, processes information

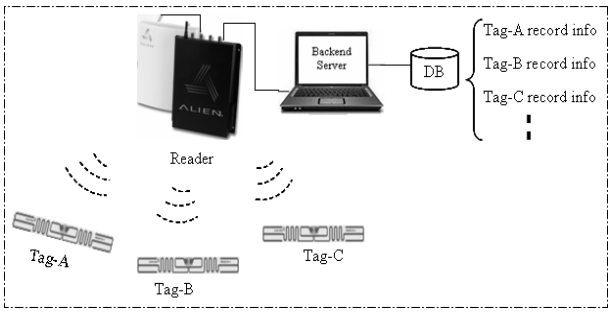


Figure 1: A typical RFID system and its components.

from the reader, and connects to the enterprise network. In a sample RFID system, shown in Fig. 1, we use ALR-8780 RFID reader/antenna and passive EPC Class 1 UHF RFID tags ALL-9338 by Alien Technology.

The rest of the paper is organized as follows. In Section 2, we introduce RFID localization systems and basic localization methods in the literature. In Section 3, we propose three leveled indoor localization algorithms based on passive RFID. In Section 4, we obtain the simulation results for the proposed algorithms. We compare and analyze the simulation results in Section 5. Finally, we conclude the paper in Section 6.

2. RFID localization systems

LANDMARC [13] is a highly cited RFID localization prototype where active RFID tags are localized based on received signal strength (RSS) measurements. A multi-level approximation of the location is achieved by dynamically changing multiple readers' transmitter powers in multiple levels and k-nearest neighbor reference tags. LANDMARC has been used to facilitate the management of hospitals in locating patients, staff, and equipment [11]. SpotON [8] is another indoor localization system based on RSS measurements to estimate inter-tag distances. RG [10], robotic guide, is a RFID-based indoor navigation robot with the help of passive RFID tags deployed in the environment. RG navigates in the building using potential fields and by finding empty spaces around itself. WLPS [20], wireless local positioning system, is intended for the design of intelligent vehicles to improve road safety by avoiding collisions. Using active RFID tags, WLPS estimates the location based on the time-of-arrival and direction-of-arrival. R-LIM [6], RFID-based library information management, is used to help find displaced library books in real-time. R-LIM is based on the aging-counter method which, in essence, is the nearest neighbor and proximity approach. RFID localization systems are generally implemented based on two classes of algorithms [1]: (Class-1) using RF propagation models to estimate the distance, e.g. RSS and TOA; (Class-2) without using RF propagation models. The Class-2 algorithms can be further classified into two sub-groups: (1) calibrate the RF distribution and then estimate the location, e.g., multilateration and Bayesian inferences; (2) direct estimate location, e.g., nearest neighbors, proximity, and kernel-based learning. In the next section we will elaborate our algorithms of choice.

3. The three RFID localization algorithms

3.1 The leveled nearest-neighbor

The proposed scheme is to combine the nearest-neighbor method [13] with multi-leveled signal strength indication to enhance accuracy. Instead of directly using RSS, which is difficult to measure by the current RFID readers, we will use detectable or non-detectable as RSSI. The nearest-neighbor means that the target tag's signal strength should be similar to that of its closest neighbors, and so is its location. In order to increase the accuracy, we expand the nearest-neighbor method into levels. The setup is as follows.

The reference passive RFID tags are deployed in a grid. The reader (mobile or quasi-mobile) has three different discrete power levels - high, medium, and low - with high-level having the longest reading range, and low-level the shortest range. The algorithm is illustrated in Fig. 2. The black dots are reference tags located on the grid with known coordinates. The solid-line square is the target. When the high-level power is used, there are four nearest neighbors: a, b, c and d, which are detectable in the large radius. The location estimate of the target is the center of $abcd$, i.e., the black dash-line square. When the power is switched to medium-level, there are two detectable neighbors, b and c, in the red dash-line radius. The new estimated location for the target is the red dash-line square. To keep the calculation simple, we just move the first estimate to the right of the half distance of the center of $abcd$ to line bc. Finally when the low-level power is used, only the single neighbor c is detected. Therefore the algorithm moves the second estimate down half distance of the center of $abcd$ to line cd to get the last estimate of the target, i.e., the black solid-line square. The estimate of the target location gradually gets more accurate.

We can also see that the leveled nearest-neighbor algorithm is indeed a weighted average approach [14], i.e., the closer the reference tag, and the more weight it will get. We can use a formula to express the estimate as

$$(x, y) \approx (\hat{x}, \hat{y}) = (\frac{\sum_{i \in N} w_i \cdot x_i}{\sum_{i \in N} w_i}, \frac{\sum_{i \in N} w_i \cdot y_i}{\sum_{i \in N} w_i}). \quad (1)$$

In Fig. 2, neighbor set $N = \{a, b, c, d\}$, (x_i, y_i) are the coordinates of the reference neighbors, where $i = a, b, c,$ or d. The weight w_i equals to the number of being detected. In the example in Fig. 2, reference tags a and d are detected once, therefore $w_a = w_d = 1$; b is detected twice, hence $w_b = 2$; and c is detected three times, thus $w_c = 3$.

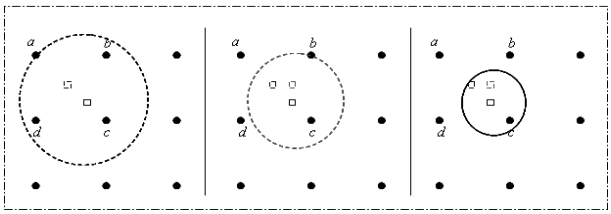


Figure 2: Leveled nearest-neighbor algorithm by three levels.

3.2 The leveled multilateration

The multilateration algorithm is a mature approach to locate objects; it requires calibrating the RF signal distribution first. In this algorithm, obtaining the weights (detectable counts as in Fig. 2 of 4.1) is the calibrating process. A larger number of detectables indicate stronger signal strength. The distances from all detected reference tags in N are replaced with the inverses of corresponding squared weights. The system of equations for the general multilateration algorithm [25] can be re-written as

$$\frac{1}{w_a^2} = (x-x_a)^2+(y-y_a)^2, \ldots, \frac{1}{w_d^2} = (x-x_d)^2+(y-y_d)^2. \quad (2)$$

The above system of quadratic equations can be reduced to linear equations by subtracting the first equation from all other $(|N|-1)$ equations [25]. Reference [4] provides the details for how to solve the system of equations by a standard least-square (LS) approach.

3.3 The leveled Bayesian inference

Bayesian inference [25], [21] is to optimally estimate the posterior probability of the hidden or unknown states (variables) in a Markov system in a recursive manner as incomplete observations become available in a Bayesian network, which is the combination of Bayes rule $(p(A|B) = p(B|A)p(A)/p(B))$ and a (directed) graph. The edges of the graph are used to represent conditional dependence and information flow. We estimate the posterior probability of the target tag t (for simplicity, with t also representing its location (x, y)) when a series of n signal strengths $s_i(i = 1, \ldots, n)$ of its reference tags $(a, b, \ldots)$ transmitted to t are available. As an example, Fig. 3 demonstrates how the posterior probability is inferred for the case of Fig. 2. The strengths of the three measures by the readable of reference tags (a, b, c, d) are $s_1 = (1, 1, 1, 1)$, $s_2 = (0, 1, 1, 0)$, and $s_3 = (0, 0, 1, 0)$, where 1 means presence and 0 means absence of the corresponding reference tag. Since the system is assumed to be a Markov process, hence given t, the probabilities of s_i are independent of each other. The probability of target tag position t, given the series of measurements of its neighbor strengths $p(t|s_1, \ldots, s_n)$, can be calculated by the following recursive equation (see [21] for details)

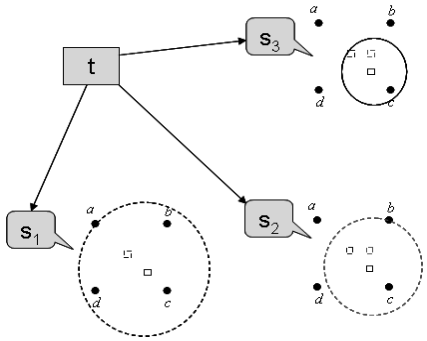


Figure 3: Leveled Bayesian inference by three levels.

$$p(t|s_1, \ldots, s_n) = \alpha p(s_n|t) \times p(t|s_1, \ldots, s_{n-1}), \quad (3)$$

where α is a normalizing factor. $p(s_i|t)$ is the probability of signal strength of measurement i given the target. Here, the signal strength will use the detectable counts as in Subsections 4.1 and 4.2. In the case of moving targets, Eq. 3 involves multi-dimensional integrals, in which Kalman filters, particle filters, and Monte Carlo methods, among other techniques, need to be utilized to approximate the solution.

For solving the recursive Eq. 3, we need to determine the *a priori* probability $p(t|s_0) \equiv p(t)$ for the target t, which represents all available information known about the target beforehand. As we only know the boundaries about the target in advance, we'll use the uniform distribution probability for $p(t|s_0)$ since it is the least biased [3]. Therefore, $p(t|s_0) = p(t) = 1/N$, where N is the total number of reference tags in the region of the target tag. In our leveled algorithm, we make three measures, i.e., $n = 3$. If we take account of the initial s_0, the Eq. 3 will become

$$p(t|s_0, s_1, \ldots, s_3) = \alpha p(s_3|t)p(s_2|t)p(s_1|t)p(t). \quad (4)$$

In the case of Fig. 3, $N = 9$ and we get the following equations for each reference tag:

$$\begin{aligned} p_a(t|s_0, s_1, s_2, s_3) &= \alpha p(s_3|t)p(s_2|t)p(s_1|t)p(t) \\ &= \frac{1}{7} \times \frac{1}{6} \times \frac{1}{4} \times \frac{\alpha}{N}, \end{aligned} \quad (5)$$

$$\begin{aligned} p_b(t|s_0, s_1, s_2, s_3) &= \alpha p(s_3|t)p(s_2|t)p(s_1|t)p(t) \\ &= \frac{2}{7} \times \frac{2}{6} \times \frac{1}{4} \times \frac{\alpha}{N}, \end{aligned} \quad (6)$$

$$\begin{aligned} p_c(t|s_0, s_1, s_2, s_3) &= \alpha p(s_3|t)p(s_2|t)p(s_1|t)p(t) \\ &= \frac{3}{7} \times \frac{2}{6} \times \frac{1}{4} \times \frac{\alpha}{N}, \end{aligned} \quad (7)$$

$$p_d(t|s_0,s_1,s_2,s_3) = \alpha p(s_3|t)p(s_2|t)p(s_1|t)p(t) = \frac{1}{7}\times\frac{1}{6}\times\frac{1}{4}\times\frac{\alpha}{N}. \quad (8)$$

Then, we obtain $\alpha = 126$ from the following normalization equation

$$p_a + p_b + p_c + p_d = \alpha(1+4+6+1)/(7\times 6\times 4\times 9) = 12\alpha/1512 = 1. \quad (9)$$

Therefore, we have $p_a = p_d = 126\times 1/1512 = 0.083$, $p_b = 126\times 4/1512 = 0.33$, $p_c = 126\times 6/1512 = 0.50$.

4. Simulation results

Uing a computer simulation program and the target location $t = (4.1, 3.6)$, we compute the location errors of the proposed multi-power level approaches with regard to different reference tag densities from 1 to 5 meters (with 1 meter step), and different radius ratios of the detection range of the high-level power which varies from 70% to 150% of the "default R (radius)."

The "default R" used is the radius of the high-level power detection range when the reference tag density is at 4 meters; i.e. $3/4$ of the length of the line $ac = 4.2426$ meters (see Figure 2). In other words, 150% of default R is $4.2426\times 1.5 = 6.3639$. However, when R is too small (70% of the default $= 4.2426\times 0.7 = 2.9698$) and the tag density is too low (5 meters), none of the reference tags gets picked up by any of the multi-level detection ranges, so no location estimate of the target tag can be ascertained.

The current ratio of the radii of the set of three detection ranges used is $4:3:2$, i.e. the small detection range is half of the large one and the medium one is the average of the large and the small. For each case when the large R changes (from 70% to 150% of the "default R"), the ratio of the three detection range radii remains unchanged at $4:3:2$.

4.1 The leveled nearest-neighbor

As demonstrated by the Figures 4 and 5, the results (localization error at: $0.0414, 0.0970$ meter) of the proposed leveled nearest-neighbor method are impressive when the reference tag density is at one meter and the large detection range is at 80% and 90% of "default R" ($3.3941, 3.8183$ meters). Furthermore, even the results at default R are quite good.

However, when the reference tag density is too low while the detection range is also too small, results can be disappointing; e.g. when the density is 5 or 4 meters and the radius is 80% of R, the errors are 1.9026 and 0.9848 meter respectively. The problem is that only the high power detection range is able to read the reference tags, but the medium and small detection ranges fall short to do so.

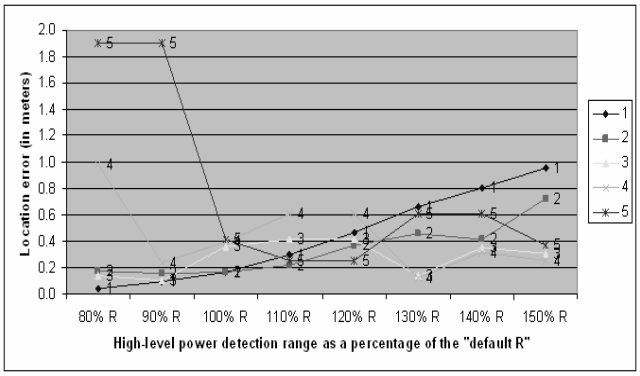


Figure 4: Leveled nearest-neighbor simulation results for different radius ratios.

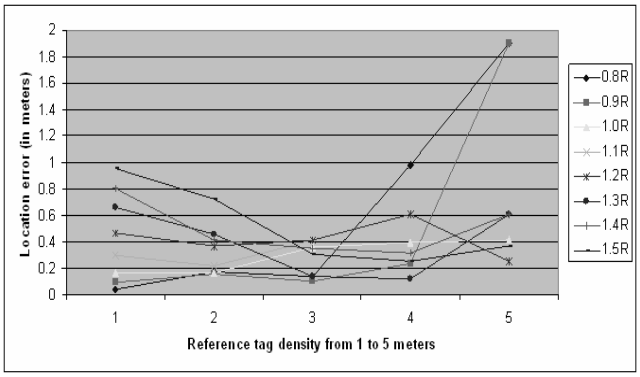


Figure 5: Leveled nearest-neighbor simulation results for different reference tag densities.

4.2 The leveled multilateration

As indicated by the graphs in Figures 6 and 7, the overall localization accuracy performance of the proposed leveled multilateration method is quite satisfactory. Furthermore, Figure 7 shows that in general the higher the reference tag density and with the greater power level of the RFID reader (which produces a larger radius of the detected zone), the better is the accuracy performance in the estimation of the location of the target entity. Nonetheless, there is always the possibility of overshooting the peak performance. When the RFID reader's power level employed is too high, as seen in Figure 6, the accuracy starts to degrade when the radius of the detection range reaches $1.5\times R$ $(= 6.3639$ meters). Lastly, Figure 8 reveals that when the parameters of the reference tag density and power level are "optimized," the results of the proposed leveled multilateration algorithm $(0.1003, 0.0408$ meter) can be exceptional (the radius at $1.40\times R$ $(= 5.9396$ meters) and the reference tag density at one or two meters).

4.3 The leveled Bayesian inference

Akin to the performance of the leveled nearest neighbor and multilateration methods, the leveled Bayesian inference

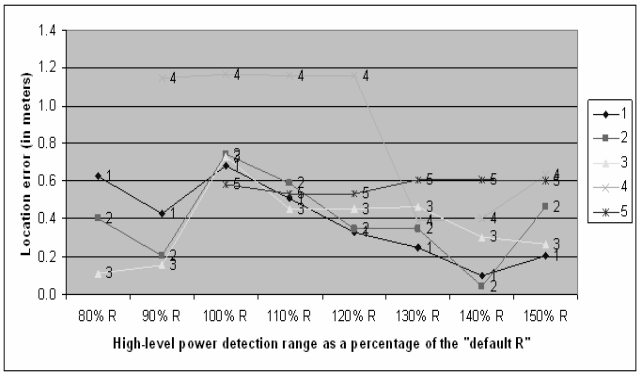


Figure 6: Leveled multilateration simulation results for different radius ratios.

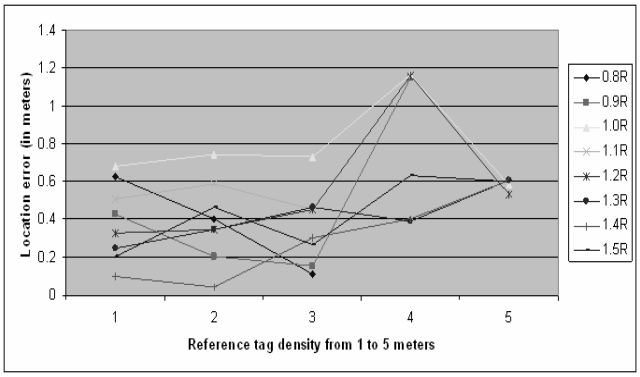


Figure 7: Leveled multilateration simulation results for different reference tag densities.

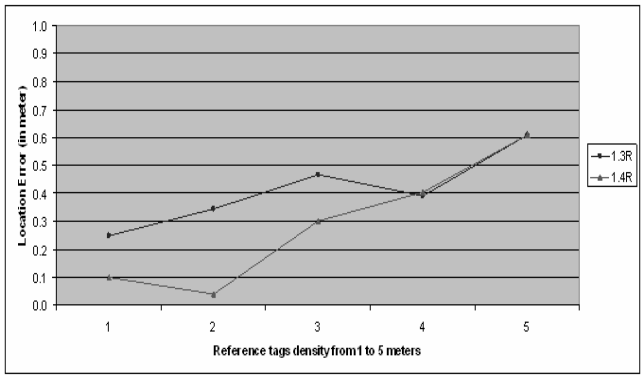


Figure 8: Leveled multilateration simulation results for different reference tag densities.

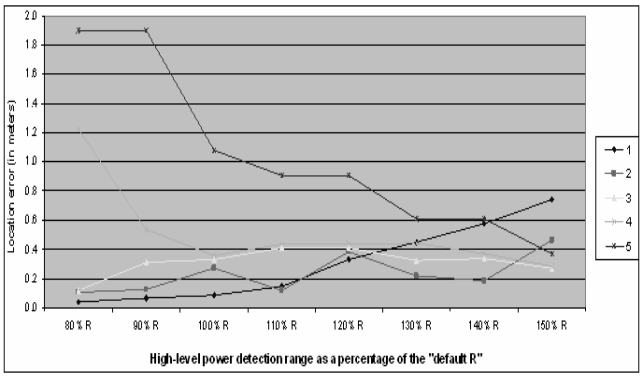


Figure 9: Leveled Bayesian inference simulation results for different radius ratios.

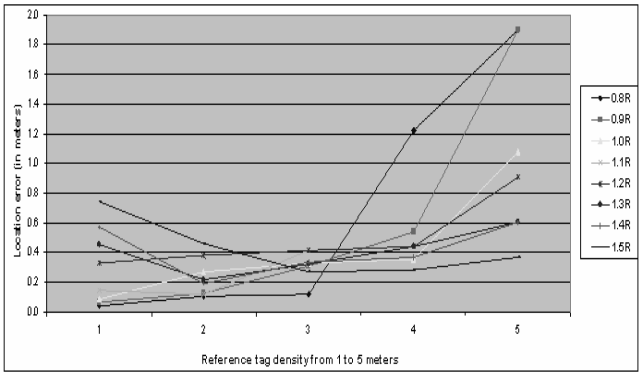


Figure 10: Leveled Bayesian inference simulation results for different reference tag densities.

algorithm performs quite well with highly satisfactory outcomes (Figures 9 and 10). More importantly, the results ($0.0389, 0.0660$, and 0.0871) are excellent when the high power detection ranges employed are relatively short (80% to 100% of the default R) and the reference tag density is high (at one meter). On the other hand, problems similar to that of the leveled nearest neighbor scheme can occur when the tag density is too high but the detection range is overly short. That causes the failure in the detection of any reference tags as they are outside the interrogation zone of the mid- and low-power levels of the RFID reader and the performance of the scheme will be significantly jeopardized.

5. Analysis

Overall as illustrated by the outcomes of the computer simulations, all three multi-level RFID localization algorithms perform quite well as long as the extreme parameter values are handled properly. Figure 11 demonstrates that the LNN (leveled nearest-neighbor) and LBI (leveled Bayesian inference) schemes obtain similar outcomes when the reference tag density used is at one meter while the LML (leveled multilateration) result has a more distinct pattern. Furthermore, the LNN and LBI methods generate much

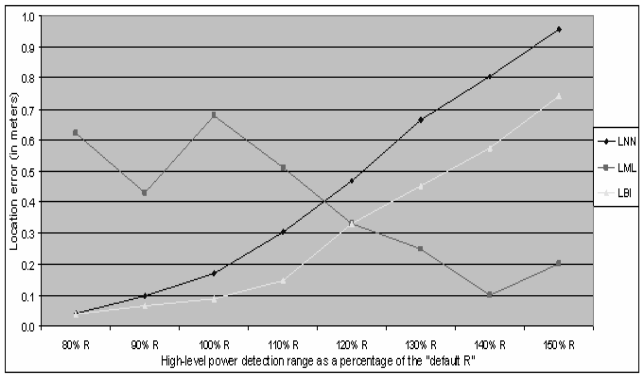


Figure 11: Proposed three algorithm simulation results at 1-meter tag density for different radius ratios.

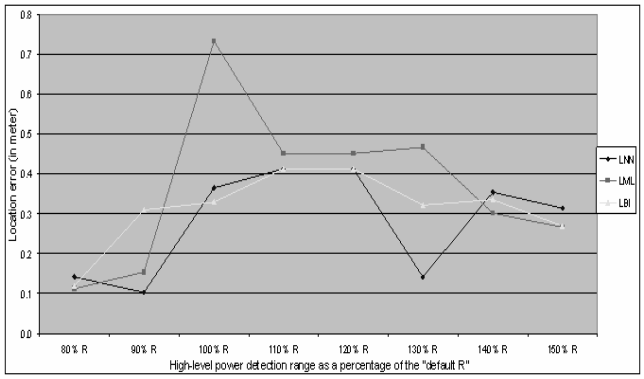


Figure 12: Proposed three algorithm simulation results at 3-meter tag density for different radius ratios.

more accurate localization results $(0.0414, 0.0389)$ than that of LML (0.6228) when the tag detection range utilized is small (80% and 90% of default R) and the reference tag density is high (at one meter). On the other hand, when the detection range is large ($150\% \times R = 6.3639$ meters), LML does significantly better than both the LNN and LML at the one meter tag density. As indicated by Figure 12, when the reference tag density employed is adjusted to three meters, all three proposed methods have comparable localization results and perform quite well even when the tag detection range used is at maximum (150% of the default R).

6. Conclusions

As clearly revealed by the simulation results, the proposed leveled detectable count RFID localization methods produce superb accuracy performance, especially when the proper values of the reference tag density and detection range are elected in conjunction with the specific multi-power level approach. In addition, our study indicates that a localization error of less than four centimeters is attainable. In contrast to other non-RFID schemes, utilizing passive RFID technology is a highly cost effective means for the acquisition of indoor positional information of both stationary and mobile entities. Furthermore, the deployment of RFID infrastructure is much less cumbersome and can be accomplished in a reasonably brief time frame.

As it is well documented in numerous RFID localization literatures that RSSI is quite susceptible to interferences and noises in the environment, our adaptation of employing detectable counts significantly increases the resistance of our algorithms to the various interference challenges imposed by the real life circumstances of RFID localization applications. To further examine the accuracy and capabilities of our approaches, we will carry out experiments and field test our methods both in a laboratory and a more realistic real-world setting. Additionally we will extend our methods of locating stationary objects to entities that travel freely in a large indoor venue.

Acknowledgments

This work was supported in part by PSC-CUNY Research Award # 65590-00 43.

References

[1] M. Bouet and A. Santos. RFID tags: positioning principles and localization techniques. In *Proc. of IFIP Wireless Days*, pages 1–5, 2008. Dubai.

[2] A. Boukerche, H. Oliveira, E. Nakamura, and A. Loureiro. Location systems for wireless sensor networks. *IEEE Wireless Communications*, 14(6):6–12, 2007.

[3] F. Bourgault, T. Furukawa, and H. Durrant-Whyte. Coordinated decentralized search for a lost target in a Bayesian world. In *Proc. of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pages 48–53, 2003. Las Vegas.

[4] J. Caffery. A new approach to the geometry of TOA location. In *Proc. of the 52 nd IEEE Vehicular Technology Conf. (IEEE VTS-Fall VTC'00)*, volume 4, pages 1943–9, 2000.

[5] P. Castro, P. Chiu, T. Kremenek, and R. Muntz. A probabilistic room location service for wireless networked environments. In *Proc. of Ubicomp'01*, pages 18–27, 2001.

[6] J. Choi, D. Oh, and I. Song. R-LIM: an affordable library search system based on RFID. In *Proc. of the 2006 Int'l Conf. on Hybrid Information Technology (ICHIT)*, pages 103–108, 2006.

[7] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34:57–66, 2001.

[8] J. Hightower, C. Vakili, G. Borriello, and R. Want. Design and calibration of the SpotON ad-hoc location sensing system. Tech. Rep., Computer Science Dept., Univ. of Washington, Aug. 2001.

[9] P. Krishna and D. Husak. RFID infrastructure. *IEEE Applications & Practice*, (9):4–10, Sept. 2007.

[10] V. Kulyukin, C. Gharpure, J. Nicholson, and S. Pavithran. RFID in robot-assist indoor navigation for the visually impaired. In *Proc. of 2004 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, pages 1979–84, 2004.

[11] C. Li, L. Liu, S. Chen, C. Wu, C. Huang, and X. Chen. Mobile healthcare services system using RFID. In *Proc. of 2004 IEEE Int'l Conf. on Networking, Sensing & Control*, volume 2, pages 1014–19, 2004.

[12] G. Mao, B. Fidan, and B. Anderson. Wireless sensor network localization techniques. *Computer Networks*, 51:2529–53, 2007.

[13] L. Ni, Y. Liu, I. Lau, and A. Patil. LANDMARC: Indoor location sensing using active RFID. *Wireless Networks*, 10(6):701–710, 2004.

[14] A. Papapostolou and H. Chaouchi. RFID-assisted indoor localization and the impact of interference on its performance. *Journal of Network and Computer Applications*, 34:902–913, 2011.

[15] N. Patwari, J. Ash, S. Kyperountas, I. Hero, A. Moses, and N. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *IEEE Signal Processing*, 22(4):54–69, 2005.

[16] N.B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *Proc. of MOBICOM 2000*, pages 32–43, 2000. Boston.

[17] S. Sarma, S. Weis, and D. Engels. RFID systems and security and privacy implications. In *Proc. of CHES 02*, volume 2523 of *LNCS*, pages 454–469, 2002.

[18] A.H. Sayed, A. Tarighat, and N. Khajehnouri. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Processing*, 22(4):24–40, 2005.

[19] A. Smaliagic and D. Kogan. Location sensing and privacy in a context-aware computing environment. *IEEE Wireless Communications*, 9(5):10–17, 2002.

[20] H. Tong and S. Zekavat. A novel wireless local positioning system via a merger of DS-CDMA and beamforming: probability-of-detection performance analysis under array perturbations. *IEEE Trans. on Vehicular Technology*, 56(3):1307–20, 2007.

[21] R. von der Merwe. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. PhD thesis, Electrical and Computer Engineering, Oregon Health & Science University, 2004.

[22] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10(1):91–102, 1992.

[23] Wikipedia. Wikipedia - the free encyclopedia. http://en.wikipedia.org.

[24] K. Yu, J. Montillet, A. Rabbachin, P. Cheong, and I. Oppermann. UWB location and tracking for wireless embedded networks. *Signal Processing*, 86:2153–71, 2006.

[25] J. Zhou and J. Shi. RFID localization algorithms and applications - a review. *Journal of Intelligent Manufacturing*, 20:695–707, 2009.

Optimization Channel Assignment Method for Maximum Throughput under Communication and Positioning Requirements

Ming Li[1], **Long Han**[1], **Weiqiang Kong**[2], **Shigeaki Tagashira**[3], **Yutaka Arakawa**[2], **and Akira Fukuda**[2]
[1]Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[2]Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[3]Faculty of Informatics, Kansai University, Osaka, Japan

Abstract— *In this paper, we first discuss a throughput estimate method based on RTS/CTS, and then propose a channel assignment method that could achieve two goals: (a) the maximum throughput of overall network can be guaranteed and (b) terminals can be located in our system. Traditionally, channel assignment is used to mitigate interference for communication. In the area of positioning, to meet the requirement of highly-accurate positioning, wireless Access Points (APs) are always assigned to a single channel. In other words, although mutual interference occurs seriously, channel assignment is not discussed in this area. To the best of our knowledge, we are the first to propose a trade-off method that gives considerations to communication and location simultaneously. To confirm its effectiveness, we evaluate our approach by simulation. The results illustrate that the throughput of our channel assignment method is higher than other methods.*

Keywords: channel assignment, communication, positioning, throughput.

1. Introduction

802.11-based wireless access is a widely used technology in public hot spots such as university campus, airports, coffee bars, and hospitals, etc. A typical development of it in recent years is Wireless LAN (WLAN) of Wi-Fi by which mobile terminals can access wireless networks smoothly through a WLAN Access Point (AP) [1]. However, with the extensive establishment of WLAN environment in public facilities, radio wave interference has become a severe problem.

On the other hand, as wireless networks rapidly gained popularity, various services that are based on context-aware of ubiquitous have been changing our daily life. As an essential part of context-aware service, positioning technologies are important for improving convenience of context-aware software. The most well-known positioning system is GPS, which is broadly used in vehicles and mobile phones etc. However, a major problem of conventional GPS is that it cannot achieve a satisfactory accuracy degree for users in indoor environment such as in buildings or underground [2]. Therefore, technologies like ultraviolet, RFID and WLAN are researched and developed for indoor positioning [2][3].

In our research, wireless APs are utilized as positioning devices since rapid development of WLAN provides a platform for positioning through APs. From our point of view, there are two primary reasons of using WLAN in positioning system.

1) WLAN is used initially for communication and Internet access. Since existing infrastructures such as APs can be used for positioning purpose in our system, we do not need to install other special devices or software.
2) WLAN can be used in indoor environment as well as outdoor environment. It would be very convenient if both indoor and outdoor LBS could be invoked by using the same mobile devices.

The foremost contribution of our research is that we are the first, to the best of our knowledge, to propose a channel assignment method that provides simultaneously wireless broadband communication service as well as precise positioning service. Based on CSMA/CA with RTS/CTS [11], we propose a throughput estimate method suitable for our research. The objective function of our channel assignment method is to achieve the maximum throughput based on the proposed estimate method. The necessity (or technical obstacles that have to be solved) of channel assignment in our research can be explained in detail as follows. To avoid radio wave interference (that impairs network throughput), it is required that an AP within the interference range of other APs is assigned with a different channel. In addition, to achieve the purpose of positioning, it is required that a mobile terminal could be able to receive radio waves from at least three APs since trilateration or multilateration location method [4] is used in our system for positioning. And thus, adjacent APs must be assigned to the same signal channel. Therefore, channel assignment is one of main work in this research.

We extend the description of our work as follows. Section 2 discusses related work. Section 3 introduces background technologies related to this paper. Section 4 defines the problem to be addressed, formalizes the problem, and proposes a channel assignment method that could achieve maximum throughput for overall network. Section 5 evaluates the proposed method and compares with other methods. Section 6 concludes the paper and mentions future work.

2. Related Works

As 802.11-based equipments have been vastly utilized, their supporting communication devices have correspondingly developed and become spatially dense. It leads to a severe mutual interference among devices which results in a poor communication throughput. Most related works thus merely target on how to realize maximum throughput for high-speed communication through optimized channel assignments. Different from those works, our research discusses channel assignments by taking both communication performance and positioning into account at the same time.

Channel assignment problem can be classified from different point of views. Based on AP management, it can be classified into *centrally managed* mode and *uncoordinated* mode. The former is mainly applied in public places such as campus or airports where all APs are managed by a central entity. The latter is often used in residential neighborhoods or private hot spots without a central manager. In both modes, an efficient mechanism for addressing interference issue to improve WLAN performance is needed.

Based on centrally managed mode, the work in [5] discusses the problems of channel assignment and AP placement simultaneously, and proposes an approximate solution with Integer Linear Programming (ILP). By adjusting the number of mobile terminals connected to APs, it evaluates network performance such as throughput. The main purpose is to find out optimized (minimized) AP locations from AP candidate points while considering a channel assignment method with a maximized throughput. However, a problem of this work is that the proposed algorithm takes an exponential computation time due to exhaustive searching. Thus, it is considered to be not suitable for large scale networks.

In [6], to investigate the fairness issue of channel assignment in resource sharing among mobile terminals, the authors proposed, based on centrally managed mode, a close-to-optimal approach called *patching algorithm*, which is better than exhausting searching w.r.t. time complexity. Furthermore, the work proposes a probability-based throughput estimation method. However, to evaluate throughput performance, the authors compare the method with other throughput evaluation standard. Due to the difference between calculation methods, it is difficult to illustrate that the method outperforms others.

The work in [7] uses centrally managed mode, in which AP placement and channel assignment problems are optimized by mathematical programing. In particular, the objective is to minimize overlapping coverage areas and to maximize network throughput. The authors propose estimated accumulation of overlapping coverage and throughput, respectively. To balance these two aspects, a trade-off parameter α was introduced into integrated functions. By tuning α, the value of multi-objective functions can be obtained.

The mode used in [8] is uncoordinated, in which a distributed, self-configure channel assignment scheme is proposed for uncoordinated WLANs. Core feature of this work is *client-assisted*, which means that APs are assisted with feedback from clients for gathering adjacent devices' traffic information. Since WLAN becomes uncoordinated and independently managed with high AP density, an automatic channel assignment method like the one in [8] becomes important. However, realization of this proposed approach is restricted by hardware development of communication devices.

All the above mentioned related works utilize *non-overlapping channels*. Another viewpoint for classifying channel assignment problem is based on *overlapping channels*. The work in [9] uses the 802.11b/g standard for channel assignment, in which Interference-factor (I-factor) is introduced to judge whether adjacent nodes are overlapping with each other. I-factor is the signal-to-noise ratio (SNR), which depends on the extent of frequency overlap between adjacent channels. Particularly, in [10], they propose an idea that partially overlapped channels are not always harmful, and they judge the interface according to threshold of SNR, not simply by the color graph theory.

3. Technology Background

WLAN can be worked in infrastructure mode or ad-hoc mode, where one of the difference between them lies on how they are connected. In infrastructure mode, APs are connected to each other with wired link to form backbone network (that can be utilized by terminals). In ad-hoc mode, APs interconnect through single/multiple hop wireless links to the backbone. And we consider the former mode in our research.

3.1 Channel Assignment and Data Rate

Based on specifications, 802.11 mainly consist of 802.11b/g and 802.11a.[1] Because of the limitation of 802.11a in outdoor environment, and thus it is not further considered in this paper. The ISM (Industrial, Scientific and Medical) band of 802.11b/g contains 11$\sim$14 channels. Inferred from Fig.1, there are only 3 non-overlapping channel in 802.11b/g (namely, channel 1, channel 6 and channel 11) [12].[2] We utilize three non-overlapping channels in this paper.

IEEE 802.11k was established and completed [13] to transfer and measure information of adjacent APs. According to 802.11k, AP channel and neighbor information can be periodically transmitted along with beacon. By this mechanism, APs can clear the channel assignment information of neighborhood. Therefore, 802.11k is useful for the development of channel assignment problem. In our

[1] Although not considered in this paper, our work is applicable to the newest specification 802.11n as well.

[2] Actually, 802.11b has one more channel – channel 14 in japan, but we do not discuss this case.

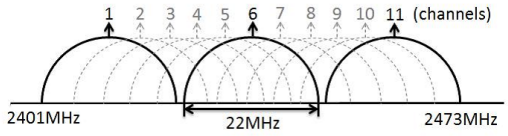


Fig. 1: 802.11b/g 2.4GHz Channels.

research, channel information can be gathered according to this protocol.

According to IEEE802.11b/g standard specification, data rate can be automatically selected by connection quality. IEEE802.11g works in 2.4MHz with a maximum raw data rate of 54Mbps. And the maximum raw date rate of IEEE802.11b is 11Mbps. The data rate is changed according to signal quality and network environment. In this paper, we assume that change of data rate is only related with distance between APs and terminals.

In general, transmission range of APs or terminals can be divided into communication range and interference range. If packets are sent/received between APs and terminals, they have to be located in a mutual communication range. Medium contentions occur if APs or terminals are in the mutual interference range. The received sensitivity thresholds of the two ranges are defined by the APs' hardware.

3.2 Medium Access Control

At present 802.11b/g is widely used in WLAN environment. According to communication protocol of Ethernet, 802.11 mechanism of MAC layer is called CSMA/CA [11]. However, it cannot avoid collision problem such as hidden terminal problem and multiply simultaneous terminal requests. Therefore, the RTS/CTS handshake-based MAC is proposed. In our research, RTS/CTS is involved when discussing Potential Restrainer. As shown in Fig.2, the two terminals (i.e., t_1 and t_2) are in the communication range of ap, and they may not know the mutual existence of each other. The general access procedure of RTS/CTS is as follows.

1) Terminal t_1 wants to send data through accessing ap by sending a message RTS (Request To Send).
2) After receiving RTS, ap sends a receipt message CTS (Clear to Send). The CTS can be detected by other communication devices (such as t_2) that are in the interference range of ap. And all devices except t_1 and ap must keep silence until the confirmation message ACK (Acknowledgment) is received by ap. Of course t_1 will wait for a period if ap is busy.
3) t_1 then obtains an authorization to send data to ap. On the contrary, t_1 can also receive data from ap.
4) When this data transmission finishes, ap sends ACK to all the other devices in its interference range, informing that they could now send RTS messages. ap becomes idle again and waits the next RTS from terminals such as t_1 or t_2.

We can summary the above access procedure as four-way handshakes of RTS-CTS-DATA-ACK. To avoid the situation that t_1 and t_2 send RTS messages simultaneously, ap waits a random time, and the terminal whose RTS is received first can send data. According to the principle, when ap sends CTS, terminals in the interference range of ap can be restrained, hence we analysis four types of Potential Restrainers in subsection 4.2 .

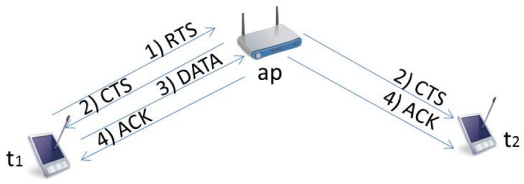


Fig. 2: CSMA/CA with RTS/CTS.

3.3 Location Method

In this subsection, we describe the mechanism of our positioning system and the multilateration positioning method [4] used in this paper.

Firstly, in our positioning system, the strength of data frame is observed and used to estimate the distance between APs and terminals. A terminal can communicate with a AP in a period of time. However, the communication (exchange data frame) strength can be observed by other APs which work in the same channel. Considering the principle of multilateration used in our system, we require a terminal must observe three APs at least for positioning.

Next, we explain the multilateration by RSSI (Received Signal Strength Indication) for positioning. RSSI can be detected by mobile devices. It is an advantage for us that we estimate accuracy by RSSI of APs, while do not depend on special hardware. It means that the positioning environment is established by existed equipment of WLAN.

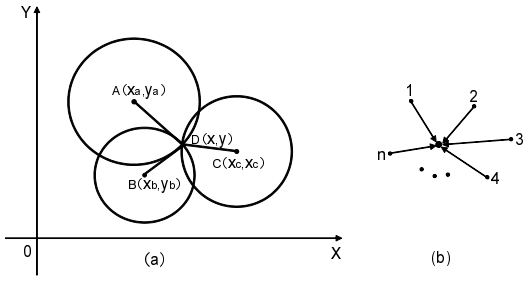


Fig. 3: Triangle and Multilateration Method for Positioning.

We can measure the distance between transmitters and receivers by RSSI and calculate coordinate by pythagorean

theorem. For example, there are three points A(x_a, y_a), B(x_b, y_b), and C(x_c, y_c) which are known, and shown as Fig.3(a). The coordinate of intersection point D(x, y) is unknown. We denote the distance form point D to A,B,C as r_a, r_b, r_c respectively. We can acquire the equations as follows and solve it.

$$\begin{cases} \sqrt{(x-x_a)^2+(y-y_a)^2} & = r_a \\ \sqrt{(x-x_b)^2+(y-y_b)^2} & = r_b \\ \sqrt{(x-x_c)^2+(y-y_c)^2} & = r_c \end{cases}$$

However, in real environment, there are not only three transmitters(APs) emitting signals. Hence we have to consider multiple known points such as n which shown as Fig.3(b), and the equations can be described as follows.

$$\begin{cases} \sqrt{(x-x_1)^2+(y-y_1)^2} & = r_1 \\ \sqrt{(x-x_2)^2+(y-y_2)^2} & = r_2 \\ \vdots & \\ \sqrt{(x-x_n)^2+(y-y_n)^2} & = r_n \end{cases}$$

Obviously, the coordinate of D is clear if the (x, y) are solved. The equations above cannot be solved directly so easily because it is a non-linear equations. However, we can solve it by Maximum-Likelihood Estimation (MLE).

4. The Proposed Method

Based on the technologies mentioned in the previous section, we first propose a throughput estimation method, and then propose an optimized channel assignment method with a maximized overall network (terminals) throughput. Meanwhile, since the best solution must satisfy the requirements of both positioning and communication, we summary some assumptions/preconditions made in this work as follows:

(a) To achieve the goal of positioning, every terminal must be able to observe in its interference range three APs of the same channel.
(b) To satisfy the requirement of communication, every terminal must be able to connect with a AP for communication that is in its communication range.
(c) The APs whose coordinates are known can be placed as a uniform deployment or not in a field beforehand.
(d) Every device (AP and terminal) has only one Network Interface Card (NIC) in this work, and thus, the channel of communication AP is the same as the channel of positioning APs.

4.1 Problem Formulation

We first define some core concepts – the sets AP and T of APs and Terminals, respectively, that are used in the work.

The set of APs: $AP = \{ap_i \mid i = 1, 2, ..., m\}$, where each ap_i is an AP.

The set of Terminals: $T = \{t_j \mid j = 1, 2, ..., n\}$, where each t_j is a terminal.

Based on the three non-overlapped channels of 802.11b/g, we consider that each AP can be assigned with a channel k ($k \in \{1, 6, 11\}$), and thus the set AP is divided into three subsets AP_k, where each $ap \in AP_k$ is assigned with channel k. For $\forall k, k'$ ($\in \{1, 6, 11\}$), $AP_k \cap AP_{k'} = \Phi$ and $\cup AP_k = AP$.

Since each AP is assumed to have one interface only and thus an AP only works with one channel. In addition, all APs are assigned with a channel k, where $k \in \{1, 6, 11\}$. The distance between an AP ap_i and a terminal t_j can be calculated by a function $D(ap_i, t_j)$.

As mentioned in subsection 3.1, every AP or terminal has an interference range and a communication range. We use R and r ($R \geq r$) to denote the corresponding radius respectively. To judge whether a terminal t_j is located in the interference range of ap_i, we define the function $AT(t_j,\ ap_i)$ as follows.

$$AT(t_j,\ ap_i) = \begin{cases} 1 & if\ \ D(ap_i,\ t_j) \leq R \\ 0 & otherwise \end{cases} \quad (1)$$

where $i \in \{1, 2, ..., m\},\ j \in \{1, 2, ..., n\}$.

To check whether an AP ap_i is assigned with a channel k, we defined the function $AC(k,\ ap_i)$ as follows:

$$AC(k,\ ap_i) = \begin{cases} 1 & if\ \ ap_i \in AP_k \\ 0 & otherwise \end{cases} \quad (2)$$

where $i \in \{1, 2, ..., m\},\ k \in \{1, 6, 11\}$. The meaning of function (2) is that if ap_i works with channel k, then $AC(k,\ ap_i) = 1$.

Next, we use S_{jk} to denote the number of APs that are assigned with channel k and are in the interference range of terminal t_j.

$$S_{jk} = \sum_{i=1}^{m} AT(t_j,\ ap_i) \cdot AC(k,\ ap_i) \quad (3)$$

where $i \in \{1, 2, ..., m\},\ j \in \{1, 2, ..., n\},\ k \in \{1, 6, 11\}$.

We formulate the two assumptions/preconditions (a) and (b) mentioned in the beginning of this section as follows. Note that we denote them as "Restriction 1" and "Restriction 2" respectively, for illustration simplicity in the algorithm in subsection 4.3.

Restriction 1 (on Positioning): The basic condition for positioning is that in the interference range of any terminal t_j, there are at least three APs working in the same channel as this terminal. So we describe the assumption as follows.

$$\forall j,\ \exists k \quad S_{jk} \geq 3. \quad (4)$$

where $j \in \{1, 2, ..., n\},\ k \in \{1, 6, 11\}$.

Restriction 2 (on Communication): Every terminal must be able to connect with a AP for communication. To satisfy

this, it is required that at least one AP is in the communication range of the terminal, which can be described as follows.

$$\forall i,\ \exists j \quad D(t_i,\ ap_j) < r \tag{5}$$

where $i \in \{1, 2, ..., m\},\ j \in \{1, 2, ..., n\}$.

In fact, there may exist multiple APs in the communication range of a terminal. So we define a set $APT_j = \{ap_i \mid D(t_j, ap_i) < r\}$. The element APs of set APT_j are those that can communicate with terminal t_j in their communication range. In addition, a set H_j, whose element APs are those that must be satisfied for both positioning and communication for terminal t_j, is defined as follows:

$$\begin{cases} \forall j \quad H_j \subseteq APT_j \ \wedge \ H_j \notin \Phi \\ \forall i \quad ap_i \in H_j \rightarrow \exists k \ AC_k(ap_i) \wedge S_{ik} \geq 3 \end{cases}$$

For providing wireless broadband, terminals are required to connect their nearest AP which is selected from the AP set that satisfies above formula. The function $O(t_i, ap_j)$ is used to define this issue.

$$O(t_j,\ ap_i) = \begin{cases} 1 & ap_i \in H_j \wedge (\exists i'\ (ap_{i'} \in H_j \wedge i' \neq i) \\ & \rightarrow D(ap_i, t_j) \leq D(ap_{i'}, t_j)) \\ 0 & otherwise \end{cases} \tag{6}$$

If ap_i is the nearest AP to t_j, the value of function $O(t_i, ap_j)$ is 1, otherwise it is 0. In brief, one terminal can only be worked in a channel, and this channel must be taken both communication and positioning into account simultaneously.

4.2 Throughput Estimation

In our research, we estimate terminal throughput according to data rate and the number of potential restrainers. We assume a WLAN environment without interference and obstacle, and that AP has no delay in switching communication from one terminal to another.

We imagine that there are one terminal and one AP merely in an ideal environment. If the terminal is in the area of 54Mbps, we consider that the throughput of this terminal is 54Mbps. On the other hand, if the number of terminals increases, the phenomenon of Medium Access Contention Constraints (MACC) [12] happens obviously. In this paper, we estimate throughput of overall network by considering accumulation of potential restrainers. Therefore, we analyze four types of potential restrainers for an arbitrary terminal t_1 that communicates with AP ap_1.

Type1: As shown in Fig.4(a), mobile terminals in the interference range of t_1 are considered as restrainers of type1. If mobile terminals near to t_1 are in its interference area, potential MACC happens directly. The number of potential type1 restrainers is denoted by RST_1, where in Fig.4(a) r_1 and r_2 are such restrainers.

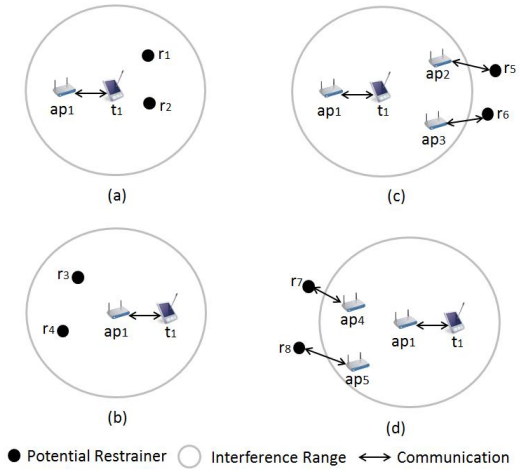


Fig. 4: Four Types of Potential Restrainers.

Type2: Different from type1, mobile terminals, which are in the interference range of ap_1 that connects with terminal t_1, are considered as restrainers of type2. potential MACC also happen in this case. The number of potential type2 restrainers is denoted by RST_2, where in Fig.4(b) r_3 and r_4 are such restrainers.

Type3: As shown in Fig.4(c), mobile terminals, which communicate with those APs that besides ap_1 but also in the interference range of t_1, are considered as restrainers of type3. According to the principle of CSMA/CA with RTS/CTS, MACC occurs between those terminals and t_1. The number of potential type3 restrainers is denoted by RST_3, where r_5 and r_6 are such restrainers.

Type4: As shown in Fig.4(d), mobile terminals, which communicate with those APs that are in the interference range of ap_1, are considered as restrainers of type4. Similar to type3, MACC also occurs between those terminals and t_1. The number of potential type4 restrainers is denoted by RST_4, where r_7 and r_8 are such restrainers.

Based on these types of potential restrainers, the function of calculating throughput for an arbitrary mobile terminal $t \in T$ can be defined as follows, where $t_j \neq t$:

$$TH(t) = \frac{DR(t)}{\sum_{j=1}^{n} RST(t, t_j)} \tag{7}$$

where

$$RST(t, t_j) = \begin{cases} 1 & if\ t_j\ is\ the\ potential\ restrainer\ of\ t \\ 0 & otherwise \end{cases} \tag{8}$$

$DR(t)$ is the data rate of t which is calculated by the RSSI (Received Signal Strength Indication) received from its connecting AP. In this research, we converse it according

to the distance between AP and terminal. For terminal t, the number of all of potential restrainers can be represented by $\sum_{j=1}^{n} RST(t, t_j)$. Function $RST(t, t_j)$ is used to check terminal whether t_j is the potential restrainer of t.

Therefore, the function of calculating throughput for overall terminals can be defined as:

$$\sum_{j=1}^{n} TH(t_j) \tag{9}$$

4.3 Channel Assignment Algorithm

The objective of the research is to find out the maximum throughput of overall terminals through channel assignment. Pseudo code of the proposed algorithm is described as follows.

Channel Assignment Algorithm

Initialization:
  All aps are assigned to channel 1;
  $MAX_{TH} = \sum_{j=1}^{n} TH(t_j)$;

Optimization:
  Exhaustive search for any possible combinations of aps' channels
  {
   If (Restriction 1 && Restriction 2)
     Calculate throughput TH of this channel combination;
     If ($TH > MAX_{TH}$)
      $MAX_{TH} = TH$;
  }

Output:
  The channel assignment when MAX_{TH};
  The value MAX_{TH}.

Initially, all aps are assigned to the same channel such as channel 1. The throughput of overall terminals is calculated according to this channel assignment. Then aps' channels are changed by exhaustive enumeration. For every possible combination of aps' channels, check whether it satisfies both Restriction 1 and Restriction 2. Finally, find out, from all satisfied combinations, the channel assignment method by which a maximum throughput is obtained.

5. Evaluation and Analysis

5.1 Simulation Parameters

We calculate the throughput of our proposed method in this section. Since we are the first to propose the channel assignment method by considering both communication and positioning and there are no similar researches, we have to compare it with single methods and random methods. By single methods we mean the methods in which all APs are assigned with the same channel. By random methods we mean the methods in which APs select three isolated channel randomly.

As shown in the table 1, we test our simulation in a 50×50 (m) area. The radiums of communication and radius of interference are set as 11m and 16m, respectively. We test every case under two specifies of 802.11b and 802.11g. And only three isolated channels can be used in our paper.

Table 1: The Simulation Parameters

Simulation Parameters	Value
Service area size (m)	50×50
Number of APs	6∼15
Number of terminals	2∼24
Channel Set	{channel 1, channel 6, channel 11}
The Employed 802.11 Specify	802.11b, 802.11g
Radius of Communication (m)	11
Radius of Interference (m)	16

5.2 Scenarios and Analysis

We consider two test scenarios for different numbers of APs and terminals, respectively. In the first scenario, we fix the number of APs as 11 such as shown in Fig.5, whereas the number of terminals is changed from 2 to 24. According to the result of simulation shown in Fig.6, status change of throughput are similar for 802.11b and 802.11g. Take the specify of 802.11g as an example, the throughput of our proposal is optimal obviously. As a quantitative analysis, the throughput of our proposed method is about three times higher than single methods and 1.5 times higher than random methods. As the the number of terminals grows, the throughput keeps a stable value in every case. However, the average throughput is calculated and shown in Fig.7. It is clear that as the number of terminals growing, the average throughput per terminal decreases for every case. Nevertheless our proposal is always better than other methods in general.

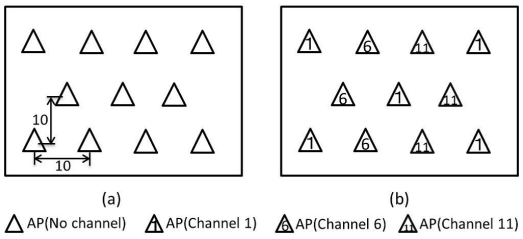


Fig. 5: An Example of AP Deployment.

On the other hand, to present an example of channel assignment, we show an assignment pattern according to our proposal when the number of APs is 11 and the number of terminals is 14 such as shown in Fig.5. Based on our assignment, the throughput of overall network is 36.67 Mbps (802.11b) and 180 Mbps (802.11g), respectively, which is better than other methods.

In the second scenario shown in Fig.8, we fixed the number of terminals, and change the number of APs from 6 to 15. As the number of APs increase, the overall throughput of proposed method is higher than others in two specifies respectively. Furthermore, we have to point out that the growth

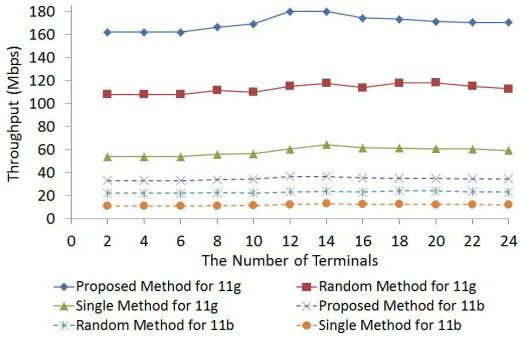


Fig. 6: Overall Throughput of Terminals for Fixed APs.

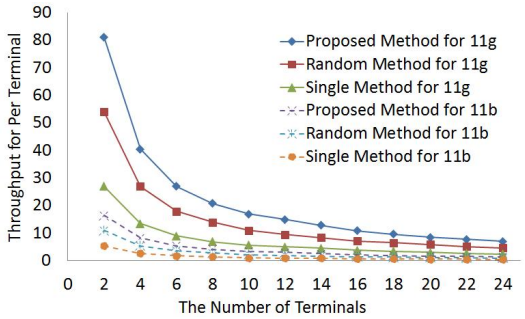


Fig. 7: Average Throughput of Terminals for Fixed APs.

rate of our proposal is optimal obviously. In this section, we evaluate our simulation by changing some parameters. The results are clear that our proposal is better than other two methods with respect to network throughput.

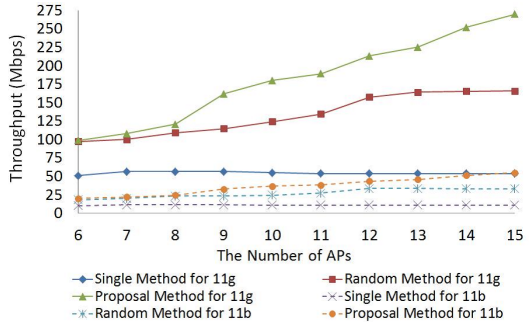


Fig. 8: Overall Throughput of Terminals for Fixed Terminals.

6. Conclusion

In this paper, a throughput estimate method according to the number of potential restrainers is proposed based on CSMA/CA with RTS/CTS. Furthermore, we propose a method for channel assignment, which, to the best of our knowledge, are the first channel assignment method that considers communication and positioning simultaneously. To verify its effectiveness, we compare the method with single methods and random methods with respect to throughput by using our proposed estimate method. The results appear that our method is better than others in every case that we tested.

In the future, we plan to evaluate the positioning aspect of our proposed method, namely whether positioning could reach the optimal objectives of communication and positioning. Moreover, we will improve the performance of channel assignment method and try to avoid extensive search of all possibilities.

References

[1] ABI Research, http://www.abiresearch.com/, 2011.

[2] P. Bahl and V.N. Padmanabhan, " RADAR: An In-Building RF-based User Location and Tracking System ", *Proc. IEEE INFOCOM 2000*, pp. 775–784, 2000.

[3] T. Kitasuka, T. Nakanishi, and A. Fukuda, " Wireless LAN based Indoor Positioning System WiPS and Its Simulation ", *Proc. 2003 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM '03)*, pp.272–275, 2003.

[4] Axel Kupper, Location-Based Services: Fundamentals and Operation, John Wiley & Sons, Ltd, 2005.

[5] Y. Lee, K. Kim, and Y. Choi, " Optimization of AP Placement and Channel assignment in Wireless LANs ", *Proc. IEEE Conf. Local Computer Networks*, pp. 831–836, 2002.

[6] X. Ling, and K. L. Yeung, " Joint Access Point Placement and Channel assignment for 802.11 Wireless LANs ", *IEEE Transactons on Wireless Communications*, Vol. 5, No. 10, pp. 2705–2711, October 2006.

[7] A. Eisenblatter, H. F. Geerdes, and I. Siomina, " Integrated Access Point Placement and Channel assignment for Wireless LANs in an Indoor Office Environment ", *Proc. 8th IEEE Intl. Symposium on a World of Wireless, Mobile and Multimedia Networks*, pp. 1–10, June 2007.

[8] Xiaonan Yue, Chi-Fai Michael Wong, and Shueng-Han Gary Chan " CACAO: Distributed Client-Assisted Channel assignment Optimization for Uncoordinated WLANs ", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 22, No. 9, pp. 1433–1440, September 2011.

[9] A. Raniwala, and T.C. Chiueh, " Architecture and Algorithms for an IEEE 802.11 Based Multi-channel Wireless Mesh Network ", *Proc. IEEE INFOCOM '05*, pp. 2223–2234, 2005.

[10] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, " Partially Overlapped Channels not Considered Harmful ", *Proc. ACM. SIGMETRICS Performance Evaluation Review*, Vol. 34, pp. 63–74, 2006.

[11] A. Colvin, " CSMA with Collision Avoidance ", *Computer Communication*, Vol.6, No.5, pp.227–235, 1983.

[12] IEEE. IEEE Std 802.11g^{TM}-2003 (amendment to IEEE std 802.11): Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, 2003.

[13] S. D. Hermann, M. Emmelmann, O. Belaifa, and A. Wolisz, " Investigation of IEEE 802.11k-based Access Point Coverage Area and Neighbor Discovery ", *Proc. 32nd IEEE Conference on Local Computer Networks*, pp. 949-954, October 2007.

Impact of TCP increase parameter in high bandwidth-delay product network with under-buffered links

Dowon Hyun, Bong Keol Shin, Tae Hyun Seong, Young Yil Kim, and Ju Wook Jang
Department of Electronic Engineering, Sogang University, Seoul, Korea

Abstract - *TCP Reno wastes thousands of RTTs ramping up to full link utilization since its congestion window size goes below the minimum congestion window size for full link utilization after packet loss (e.g., 3Dup-ACK) in high bandwidth-delay networks with under-buffered links. Many TCP variants are proposed to increase its congestion window size rapidly or maintain its congestion window size over the minimum congestion window size to achieve high link utilization in high bandwidth-delay network without considering under-buffered links. However, they have the problem of TCP fairness. In this paper, we invest the impact of the increase parameter based on our proposed TCP congestion control algorithm earlier so that we can utilize the link almost up to 100% except the first congestion avoidance cycle and have a good TCP fairness in high bandwidth-delay networks with under-buffered link. We provide ns-2 simulation results which show the effect of competing TCP Reno flow according to increase parameter.*

Keywords: Link utilization, TCP Reno, under-buffered link, congestion window, increase parameter

1 Introduction

Link utilization of TCP Reno[1] is changed according to the router buffer size. For the given link with capacity C (packet/sec) and the minimum round trip time RTT_{min} (seconds), optimal buffer size B_{opt} equals the value of C x RTT_{min} according to rule-of-thumb[2] when a drop-tail queue is used. Let B denote the buffer size of router. When $B \geq B_{opt}$, TCP Reno fully utilizes the link and it is called an over-buffered link. When $B < B_{opt}$, TCP Reno does not fully utilize the link and it called an under-buffered link. In this case, TCP Reno wastes thousands of RTTs ramping up to full link utilization since its congestion window size goes below the minimum congestion window size for full link utilization after packet loss (e.g., 3Dup-ACK) in high bandwidth-delay networks with under-buffered links.

TCP variants[3~8] are proposed to increase its congestion window size rapidly and/or maintain its congestion window size more than required minimum congestion window size to achieve high link utilization in high bandwidth-delay product network without considering under-buffered links. They have high link utilization and good intra-protocol fairness in under-buffered links. However, they have serious problem of TCP fairness when they are competing against TCP Reno flows. For example, link capacity C = 100Mbps, the minimum round trip time RTT_{min} = 80ms and the buffer size $B = B_{opt}/4$ = 250 (packets). Table. 1 shows that ns-2[9] result of throughput and fairness index[10] when TCP variant are competing against TCP Reno. TCP Reno flow starts at 0 ~ 400 sec and TCP variant flow starts at 100 ~ 500 sec.

Table. 1 Comparison of the throughput, the link utilization, and the fairness index between TCP Reno and TCP variant

	Throughput (Mbps)	Link utilization	Jain's fairness index
TCP Reno TCP Reno	48.23 37.64	0.859	0.985
TCP Reno Compound TCP[3]	21.49 72.81	0.943	0.772
TCP Reno HighSpeed TCP[4]	7.99 81.96	0.890	0.597
TCP Reno Scalable TCP[5]	3.93 84.72	0.887	0.546
TCP Reno H-TCP[6]	6.26 86.99	0.932	0.572
TCP Reno Bic TCP[7]	4.61 90.33	0.950	0.551
TCP Reno Cubic TCP[8]	11.08 82.21	0.933	0.632

As shown in Table. 1, TCP variants have high throughput and link utilization since they have higher increase parameter than TCP Reno. Relatively, competing TCP Reno has very low throughput. Thus TCP Reno and TCP variant have bad fairness index nevertheless they achieve high link utilization.

We proposed an enhanced TCP congestion control that can be fully utilized in high bandwidth-delay network with an under-buffered link in [11]. Our scheme has good TCP fairness with TCP Reno. In this paper, we analyze the impact of increase parameter for good TCP fairness using our scheme.

The rest of the paper is organized as follows. In Section 2, we give a brief overview of our proposed TCP congestion control and modify increase parameter to evaluate the impact of increase parameter. We analyze impact of increase parameter using the ns-2 in Section 3. Finally, we conclude in Section 4.

2 The enhanced TCP congestion control

We proposed the enhanced TCP congestion control algorithm in [11]. We use it to evaluate the performance of increase parameter. Our TCP scheme can be fully utilized in high bandwidth-delay network with an under-buffered link. We give a brief overview of our TCP scheme and modify increasing method as TCP Reno does.

2.1 Estimation of the congestion window size for full link utilization

RTT increases at the time when the link is full and the buffer starts to fill up. However, RTT in TCP Reno is unsuitable to detect the exact time RTT increases due to small changes and oscillation. Thus, we use a smoothed RTT using log-likelihood (LSRTT) which reflects increasing trend of RTT better than that of RTT[12]. Let $C(t)$ denote the congestion indicator as shown below:

$$C(t) = \frac{LSRTT(t)}{RTT_{min}} > \alpha \qquad (1)$$

where t is the number of transmission round and RTT_{min} is the minimum of all measured round trip times as in TCP Vegas[13]. If $C(t) \le \alpha$, we regard that there is no congestion. When $C(t)$ is larger than α where $\alpha > 1$, we deduce that congestion exists.

We denote by W_{full}^{i} the minimum congestion window size for full link utilization at the time when $C(t) > \alpha$. Superscript i is the number of the congestion avoidance cycle. W_{full}^{i} can be obtained as shown below:

$$W_{full}^{i} = cwnd_{proposed}^{i}(t^{*}) \times \beta, t^{*} = \min\{t \mid C(t) > \alpha\} \qquad (2)$$

where $cwnd_{proposed}^{i}(t)$ is the congestion window of the proposed TCP scheme. We describe $cwnd_{proposed}^{i}(t)$ in detail in next sub-section. W_{full}^{i} may be overestimated since $\alpha > 1$. The overestimated W_{full}^{i} may cause aggressive transmission against TCP Reno. Thus, we add another parameter β, which reduces overestimated W_{full}^{i} where $\beta < 1$.

2.2 Congestion window control

Figure. 1 shows the congestion window evolution schematic of the proposed TCP scheme. The shaded region in Fig. 2 is not used in TCP Reno in an under-buffered link. Therefore, our scheme sends extra packets to utilize the shaded region without disturbing TCP Reno. To do so, we add a new congestion window $extra^{i}(t)$, which transmits extra packets. The proposed TCP congestion avoidance control algorithm can be represented as shown below:

$$cwnd_{proposed}^{i}(t) = cwnd^{i}(t) + extra^{i}(t) \qquad (3)$$

where, t is the number of transmission round and i is the number of the congestion avoidance cycle. $cwnd^{i}(t)$ is the congestion window of TCP Reno. Note that we do not control $cwnd^{i}(t)$. Therefore, $cwnd^{i}(t)$ is updated in the same way as in TCP Reno.

Our TCP scheme updates $extra^{i}(t)$ is updated as follows:

- If $C(t) \le \alpha$ and $cwnd_{proposed}^{i}(t) < W_{full}^{i-1}$, $extra^{i}(t)$ increases quickly to acquire spare bandwidth (from t0 to t1 in Figure. 1).
- If $C(t) \le \alpha$, $extra^{i}(t)$ decreases linearly to maintain $cwnd_{proposed}^{i}(t) = W_{full}^{i-1}$ (from t1 to t2 in Figure. 1).
- If $C(t) > \alpha$ or $cwnd_{proposed}^{i}(t) \ge W_{full}^{i-1}$, $extra^{i}(t)$ is set to 0 (from t2 to t3 in Figure. 1).

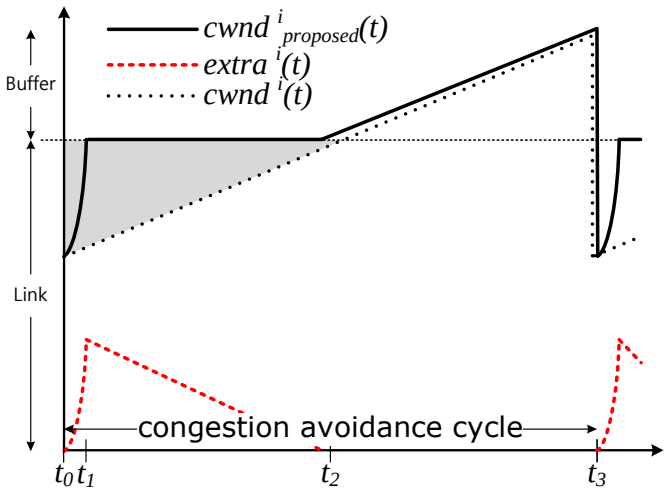


Figure. 1 The congestion window evolution and extra transmitted packets of the proposed TCP scheme during one congestion avoidance cycle.

In the first step, $extra^i(t)$ increases $cwnd^i_{proposed}(t)$ up to the value of W^{i-1}_{full} according to the increasing parameter a. It acquires available bandwidth quickly. If we set $extra^i(t)$ equal to W^{i-1}_{full} at the instant of the starting fast recovery, it can cause packet loss since the transmission speed of burst packet without ACK clocking at sender can be faster than the packet receiving speed of router. It can cause serious performance degradation. Thus, we increase $extra^i(t)$ like a slow start. In the second step, $extra^i(t)$ decreases linearly to maintain $cwnd^i_{proposed}(t)$ equal to W^{i-1}_{full} until $C(t) \le \alpha$. In this step, the proposed TCP scheme uses only an unused link capacity. Therefore, our TCP scheme neither increases RTT nor affects existing TCP Reno flows. In the third step, $extra^i(t)$ is set to 0 when $C(t) > \alpha$ or $cwnd^i_{proposed}(t) \ge W^{i-1}_{full}$. Our scheme performs additive increase by $cwnd^i(t)$ as TCP Reno does.

We summarize the proposed TCP scheme for $extra^i(t)$ as in (4).

$$extra^i(t+1) = \begin{cases} a \times cwnd^i(t), \text{ if } C(t) \le \alpha, cnwd^i_{proposed}(t) < W^{i-1}_{full} \\ \max\{0, W^{i-1}_{full} - cwnd^i(t)\}, \text{ if } C(t) \le \alpha \\ 0, \text{ if } C(t) > \alpha \text{ or } cwnd^i(t) > cnwd^i_{proposed}(t) \end{cases} \quad (4)$$

Our scheme halves its congestion window $cwnd^i_{proposed}(t)$ just as TCP Reno does when the packet loss occur.

To investigate the impact of increase parameter, we modify increasing method in (3) as TCP Reno does during $t_0 \sim t_1$ in Figure. 1. Modified $extra^i(t)$ as shown below:

$$extra^i(t+1) = \begin{cases} extra(t) + inc / cwnd^i(t), \text{ if } C(t) \le \alpha, cnwd^i_{proposed}(t) < W^{i-1}_{full} \\ \max\{0, W^{i-1}_{full} - cwnd^i(t)\}, \text{ if } C(t) \le \alpha \\ 0, \text{ if } C(t) > \alpha \text{ or } cwnd^i(t) > cnwd^i_{proposed}(t) \end{cases} \quad (5)$$

where, inc is the increase parameter.

3 Simulation result

We evaluate the performance of the TCP+ according to increase parameter using ns-2. We consider a simple topology which consists of two TCP sender, TCP receivers, and router. One of two TCP sender use TCP Reno and one of two TCP sender use TCP variant. The sender's access link is higher and faster than the receiver's bottleneck link. Bottleneck link capacity C = 100Mbps and round trip time RTT_{min} = 80ms. Router's buffer size B is 250 (packets) and the router has buffer size a single drop-tail buffer. We use maximum segment size MSS=1000Byte. In this topology, the calculated window size for full link utilization is 1000(packets). Figure. 2 shows the simulation network topology.

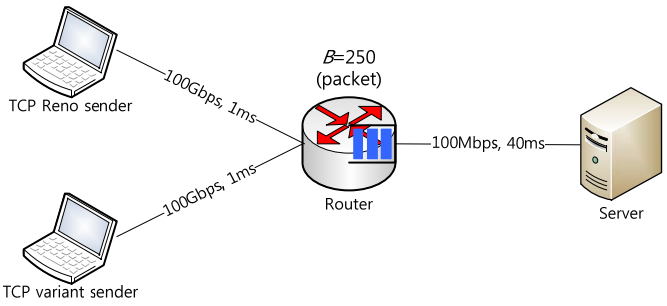


Figure. 2 Simulation network topology.

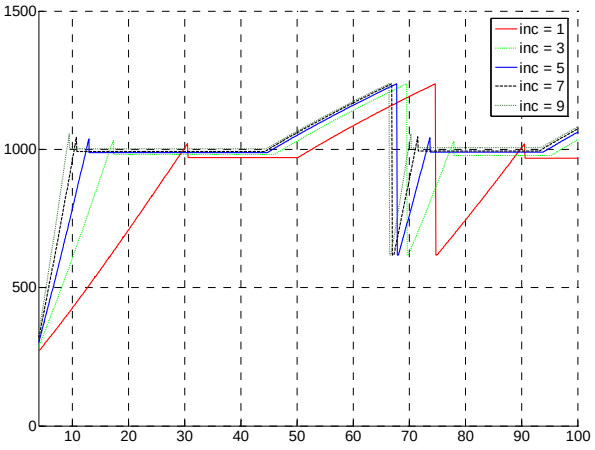


Figure. 3 Congestion window evolution with variations of inc.

Table. 2 Comparison of the link utilization with variations of *inc*.

Increase parameter	Link utilization	Link utilization ratio (%)
1	0.908	-
3	0.934	2.86
5	0.942	3.74
7	0.946	4.19
9	0.949	4.52

Figure. 3 and Table.2 show the congestion window evolution and link utilization for our TCP single flow with variations of inc. As inc increases, our TCP scheme can increase its congestion window size rapidly and it has higher

link utilization. However, high inc can cause over estimation of the congestion window size for full link utilization as shown in Figure. 3. As a result, it operates more aggressively than TCP Reno.

Table. 3 shows the throughput, the link utilization, and fairness index of our scheme and TCP Reno when two flows are competing. TCP Reno flow starts at 0~400 sec and TCP variant flow starts at 100~500 sec. The throughput of TCP Reno is decreased, the link utilization is increased, and fairness index is decreased as *inc* increases. Since our TCP scheme preoccupies more available link capacity and the buffer according as to increasing *inc*. There is trade-off between the link utilization and TCP Reno fairness.

Table. 3 Comparison of the throughput, the link utilization, and fairness index between our TCP scheme and TCP Reno.

	Throughput (Mbps)	Link utilization	Jain's Fairness index
TCP Reno TCP Reno	48.23 37.64	0.859	0.985
TCP Reno TCP+ *inc*=1	45.64 43.18	0.890	0.999
TCP Reno TCP+ *inc*=3	42.77 48.85	0.916	0.996
TCP Reno TCP+ *inc*=5	35.19 57.69	0.930	0.945
TCP Reno TCP+ *inc*=7	32.78 60.93	0.937	0.917
TCP Reno TCP+ *inc*=9	32.12 61.85	0.940	0.909

Table. 3 Comparison of the throughput, the link utilization, and fairness index between our TCP scheme and TCP Reno.

4 Conclusions

In this paper, we present the impact of increase parameter of TCP in high bandwidth-delay product network with under-buffered links. Our simulation results show that high increase parameter can cause the problem of TCP fairness. Thus, appropriate increase parameter is considered. We claim that our results are useful to design a new TCP friendless protocol in high bandwidth-delay product network with under-buffered links.

5 Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0005101).

Ju Wook Jang is corresponding author.

6 References

[1] V. Jacobson, "Congestion avoidance and control," ACM SIGCOMM, pp.314-329, 1998.

[2] C. Villamizar and C. Song, "High performance TCP in ANSNET," ACM Comput. Commun. Rev., vol.24, no.5, October 1994.

[3] K. Tan. J. Song, Q. Zhang, and M. Sridharan, "Compound TCP: A Scalable and TCP-Friendly Congestion Control for High-speed Networks", in Proc. of PFLDnet 2006, Nara, Japan, February 2006.

[4] S.Floyd, "Highspeed TCP for Large Congestion Window," RFC 3649, December 2003.

[5] Tom Kelly, "Scalable TCP: Improving Performance in High-speed Wide Area Networks." in Proc. of PFLDnet 2003, Geneva, Switzerland, February 2003.

[6] D. Leith and R. Shorten, "H-TCP: TCP for high-speed and long-distance networks," in Proc. of the PFLDNet Workshop, Argonne, IL, USA, February 2004.

[7] L. Xu, K. Harfoush, I. Rhee, "Binary increase congestion control (BIC) for fast long-distance networks," in Proc. of IEEE INFOCOM 2004, Hong Kong, China, March 2004.

[8] I. Rhee and L. Xu, "CUBIC: A new TCP-friendly high-speed TCP variant," in Proc. of the PFLDNet Workshop, Lyon, France, February 2005.

[9] The Network Simulation – NS2, http://www.isi.edu/nsnam/ns/

[10] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research, tech.rep. TR-301, Sept. 1984.

[11] Dowon Hyun and Ju Wook Jang, "Enhanced TCP Congestion Control with Higher Utilization in Under-Buffered Links," IEICE Trans. Commun., Vol.E95-B, NO.4, pp.1427-1430, April 2012.

[12] D. W. Ngwenya and G. P. Hancke, "The Effects of Using Change Detection Algorithms in Estimation of the Average RTT," in Proc. SATNAC, Western Cape, South Africa, September 2004.

L. S. Brakmo, S. W. O'Malley, and L. L. Peterson, "TCP Vegas: Net techniques for congestion detection and avoidance," in Proc. of ACM SIGCOMM, London, UK, Auust 1994.

An Experimental Study of Low Stretch Factor on Greedy Geometric routing based Algorithm

Omkar Kulkarni[1], **Huaming Zhang** [2], **and Swetha Govindaiah**[3]
[1]Computer Science Department, University of Alabama, Huntsville, Alabama, USA
[2]Computer Science Department, University of Alabama, Huntsville, Alabama, USA
[3]Computer Science Department, University of Alabama, Huntsville, Alabama, USA

Abstract— *Greedy routing has drawn a lot of attention in research community. To our best knowledge,* stretch factor *has not been studied for greedy routing algorithms. In this paper, we propose a new greedy routing algorithm and take the initiative to evaluate its stretch factor. Our algorithm generalizes the algorithm presented in [1]. It embeds an n-vertex connected graph $G(V, E)$ in a semi-metric space. With an appropriate distance notion, greedy routing works fine, just like for the case of embedding graphs in metric spaces. Theoretically, our algorithm runs in linear time. The virtual coordinates of each vertex u can be represented in $deg(u)O(logn)$ bits. The distance between any two vertices can be computed in constant time. We observe that stretch factors are always small constant numbers. We present their values and patterns in the paper. To our best knowledge, this is the first study towards a practical low-stretch-factor greedy routing algorithm.*

Keywords: Stretch factor, Greedy routing algorithm, experimental study

1. Introduction

Stretch factor is defined as the worst ratio between the length of the routing path computed by the algorithm versus the length of a shortest path between any source u and any destination w. Routing through shortest path between any two vertices is always of special interest in network routing algorithms. In this paper, the shortest path between any two vertices u and w is defined as a path connecting u and w with smallest number of edges. The number of edges in such a shortest path is denoted by $\mathcal{M}(u, v)$. Consider a routing algorithm $\mathcal{A}$. Consider any two vertices u and w, we define $\mathcal{A}(u, w)$ to be the number of edges in the routing path between u and w, as generated by the algorithm $\mathcal{A}$. We define *stretch factor* $\mathcal{S}(u, w)$ to be the ratio of $\mathcal{A}(u, w)$ over $\mathcal{M}(u, v)$. When applying an routing algorithm $\mathcal{A}$ to a network G, the stretch factor of the G is defined as the worst ratio (i.e., the maximum ratio) among all such vertex pairs. We denote it by $\mathcal{S}_A(G)$. Stretch factor is a key parameter when evaluating the efficiency of a routing algorithm. Lower stretch factors always mean lower overhead cost, better power conservation and less traffic congestion. Obviously, by applying traditional routing table, a routing algorithm can achieve the best stretch factor. Namely, the routing path computed by such a routing algorithm between any two vertices u and w is always the shortest path between u and w. Therefore, the stretch factor is always 1 and hence is optimal. It is worth noting that such an approach requires $O(nlogn)$ bits to store such a table at every vertex (where n is the number of vertices in the network). In some resources-limited networks, such as wireless sensor networks, requiring $O(nlogn)$ bits to represent a routing table at each node is simply not feasible. To combat such resource limitation, *geometric routing* using *virtual coordinates* was proposed for such networks. The hope was that, for each node u, we assign virtual coordinates to u. Then the routing path can be computed from such virtual coordinates between any pair of vertices. However, in order for geometric routing to outperform the traditional routing table approach, the virtual coordinates at each vertices should be representable in less than $O(nlogn)$ bits. Ideally, one wants the virtual coordinates to be *succinct*. Namely, the virtual coordinates can be represented in $O(logn)$ bits.

After being introduced, geometric routing has drawn a lot attention recently within the network research community [2]. *Greedy routing* is perhaps the simplest form of geometric routing. In greedy routing, a message from a source u to a destination w is simply forwarded to a neighbor v of u such that the distance from v to w is closer than the distance from u to w. The forwarding process continues until the message reaches its destination w. Obviously, greedy routing does not always succeed. A message can get stuck in a *void position* v, where all its neighbors are further away from the destination w than v. Previously, most research focused on how to assign virtual coordinates so that greedy routing always succeeds. The key was to compute virtual coordinates, with an appropriate distance notion, so that greedy routing always succeeds between any source u and any destination w. As for the general case of applying geometric routing, in order to make greedy routing algorithms practically feasible for resource limited networks, such as wireless sensor networks, one needs to find greedy routing algorithms in which the virtual coordinates assigned to each node in the network are succinctly representable in $O(logn)$ bits [3]. Since then, for greedy routing, researchers focused on designing greedy routing algorithm that always guarantee delivery. They have

tried to use different embeddings, different metric spaces and different semi-metric spaces (see [4], [5], [6], [3], [7], [8], [9], [1], [10]). However, up to now, no research has been conducted on evaluating the stretch factors, which seems odd, considering that stretch factor is such an important parameter in network routing algorithms. In this paper, we take the initiative and we do the experimental study on stretch factor of a new greedy routing algorithm.

1.1 Related work

From now on in this paper, we will interchangeably use the word network and the word graph. Rao et al.[2] proposed to use graph drawing, based on the structure of a graph G, to compute vertex coordinates in the drawing. The drawing coordinates are used as the virtual coordinates of the vertices of G. Then geometric routing algorithms rely on virtual coordinates to compute routes. *Greedy drawing* is introduced as a solution for greedy routing. Simply speaking, a greedy drawing is a drawing of the graph G into a metric space in which greedy routing works. More precisely:

Definition 1: [11] Let S be a set and $d(*,*)$ a metric function over S. Let $G = (V, E)$ be a graph.

1) A *drawing* of G in S is a mapping $\delta : V \to S$ such that $u \neq v$ implies $\delta(u) \neq \delta(v)$.
2) The drawing δ is a *greedy drawing* with respect to d if for any two vertices u, w of G ($u \neq w$), u has a neighbor v such that $d(\delta(u), \delta(w)) > d(\delta(v), \delta(w))$.

Consider a greedy drawing δ for G in a metric space S, endowed with a metric function $d(*,*)$. Let's define $d_\delta : V \to R$ as follows: $d_\delta(u, w) = d(\delta(u), \delta(w))$. Then it is easy to see that $d_\delta(*,*)$ is a metric function defined on the vertex set V of G. We will simply call this metric function the *induced metric function* for G from $d(*,*)$ and δ, and always denote it by d_δ.

Consider any two vertices $u \neq w$ of G in the greedy drawing δ. According to the definition of a greedy drawing, it is easy to see that there is a *distance decreasing path* between u and w in the drawing with respective to the metric function d_δ (see [11]). Namely, there is a path $(u = v_1, v_2, \cdots, w = v_k)$ in G such that $d_\delta(v_i, w) < d_\delta(v_{i+1}, w)$. Therefore, greedy routing simply forwards the message from u to such a neighbor v, which is closer to the destination w than u. The forwarding process continues and the distance to the destination w keeps dropping. Eventually, the distance become 0 and the message reaches the destination w. Therefore, when there is a greedy drawing of G into a metric space S, greedy routing always succeeds.

1.2 Greedy routing via embedding graphs into semi-metric spaces

Very recently, Zhang et al. [1] studied greedy routing algorithm via embedding graphs into semi-metric spaces. They proposed a new notion of greedy drawing. Instead of using greedy drawing of graphs in metric spaces, they used *greedy embedding* of graphs in *semi-metric spaces*. We have the following definition from [1]

Definition 2: Let S be a set. A semi-metric on S is a function $d : S \times S \to R$ that satisfies the following three conditions:

1) $d(x, y) \geq 0$.
2) $d(x, y) = 0$ if and only if $x = y$.
3) $d(x, y) = d(y, x)$.

Definition 3: [11] Let S be a set and $d(*,*)$ a semi-metric function over S. Let $G = (V, E)$ be a graph.

1) An *embedding* of G into S is a mapping $\delta : V \to S$ such that $u \neq v$ implies $\delta(u) \neq \delta(v)$.
2) The embedding δ is a *greedy embedding* with respect to d if for any two vertices u, w of G ($u \neq w$), u has a neighbor v such that $d(\delta(u), \delta(w)) > d(\delta(v), \delta(w))$.

Let δ be a greedy embedding of a graph $G = (V, E)$ into a semi-metric space S, whose semi-metric function is $d(*,*)$. Consider the vertex set V of G, for any two vertices u and v in V, if we define $d_\delta(u, v) = d(\delta(u), \delta(v)$, then $d_\delta(*,*)$ is a semi-metric function defined on the vertex set V of G, $d_\delta(*,*)$ is called the *induced semi-metric function* for G. Note that, given a greedy embedding into a semi-metric space, greedy routing works exactly the same as for the case of greedy drawing in metric spaces. Consider a greedy embedding δ for G into a semi-metric space S, equipped with a semi-metric function $d(*,*)$. Consider any two distinct vertices $u \neq w$ of G, there is a *distance decreasing path* between u and w in the embedding. Namely, there is a path $(u = v_1, v_2, \cdots, w = v_k)$ in G such that $d_\delta(v_i, w) < d_\delta(v_{i+1}, w)$. Therefore, greedy routing simply needs to forward the message to such a neighbor which is closer to the destination. The message keeps forwarded, the distance keep dropping until the message reaches the destination w (whose distance to w is 0).

1.3 A group of semi-metric spaces to be used

Let N be the set $\{1, 2, \cdots, n, \cdots\}$, i.e. the set of natural numbers. We use $P(N)$ to denote the power set of N. We have the following definition [1]:

Definition 4: A proper subset Π of $P(N)$ is called a *perfect sub-power set of* N if Π satisfies all the following conditions:

1) $\Pi \neq \emptyset$ and Π is a finite set.
2) $\emptyset \notin \Pi$.
3) For any two distinct elements x and $y \in \Pi$, $x \cap y = \emptyset$.
4) Every element $x \in \Pi$ is a finite set of natural numbers. We define the *thickness* of Π to be the maximum cardinality of all its elements, which will be denoted by $\mathcal{T}_\Pi$.

We have the following technical lemma from [1]:

Lemma 1: Let Π be a perfect sub-power set of N. x and y be two elements in Π. Let $d(x, y) = min\{|x_i - y_j|: x_i \in x$ and $y_j \in y\}$. Then $d(*,*)$ is a semi-metric function defined on Π.

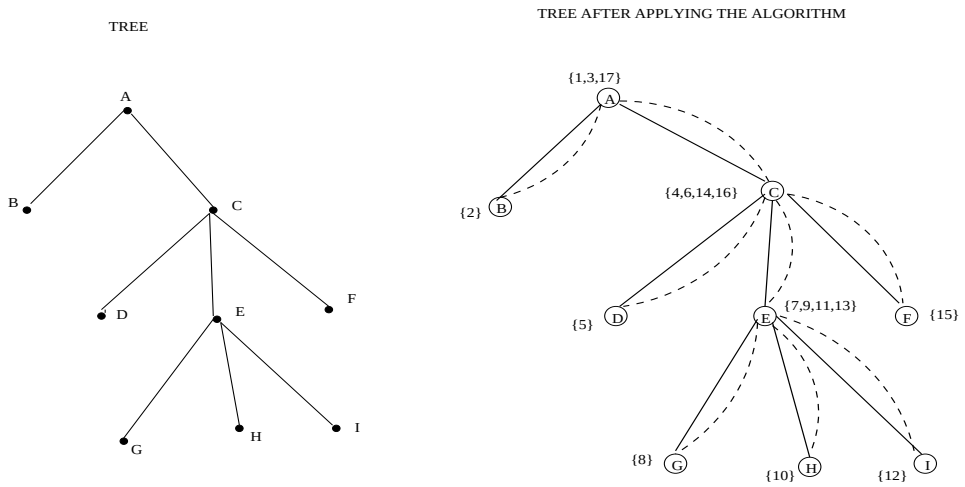


Fig. 1: (1) A tree T. (2) Virtual coordinates for each vertex in T.

Zhang et al. [1] studied greedy embedding of any connected tree onto such a perfect sub-power set of N. Let $T = (V, E)$ be a tree with n vertices. r be the root of T. The greedy embedding $\delta : V \to \delta(V) = \Pi$ simply assigns numbers $\{1, \cdots, 2n-1\}$ to the vertices of the tree while walking along the facial cycle of the tree in counterclockwise direction. The walk starts from the root r at the time 1 and comes back to r, whenever a vertex v is visited, the number is inserted into $\delta(v)$. See Figure 1[1] for an illustration. In such a greedy embedding, each vertex of the tree T is mapped to a finite set of natural numbers from 1 to $2n-1$. The distance between the vertices can be computed using Definition 4. Zhang et al. proved that, using this greedy embedding and this notion of distance, for any two vertices u and v in the tree T, the unique path between u and v is a distance decreasing path between u and v. The following lemma summarizes the results from [1].

Next, we have the following lemma [1]:

Lemma 2: 1) Let $T = (V, E)$ be a tree. The embedding $\delta : V \to \Pi = \{\delta(v) : v \in V\}$ is a greedy embedding from T onto $\Pi = \{\delta(v) : v \in V\}$ with respect to its semi-metric function d. The embedding can be obtained in linear time. The thickness $\mathcal{T}_\Pi$ of Π, which is defined as the maximum cardinality of $\delta(v), v \in V$, is at most $\leq \Delta(T)+1$. When $|V| = n > 2$, the thickness $\mathcal{T}_\Pi$ of Π could be reduced to $\Delta(T)$ if one chooses a degree-1 vertex r as its root.

2) Let $G = (V, E)$ be a connected graph with more than 2 vertices. Then G admits a greedy embedding into a perfect sub-power set with thickness bounded by $\Delta(G)$. This greedy embedding is constructible in linear time.

3) For any vertex u, the time of u to find a neighbor to forward a message is bounded by $\Delta(G)^2$.

[1]See [1]

2. Greedy Embedding by using one spanning tree and multiple other trees

In this section we introduce a simple greedy routing algorithm by using one spanning tree and multiple other trees to compute virtual coordinates for the vertices. We will still use greedy embedding into perfect sub-power set. Our algorithm extends the greedy routing algorithm presented in [1]. Let $G = (V, E)$ be the input graph with n vertices. Our algorithm works as follows: First, we compute a spanning tree T_1 of G. Then we use T_1 to assign virtual coordinates to the vertices. Then consider $G_1 = G - T_1$ (i.e., G_1 is the graph after deleting all the edges in T_1). Then we compute a tree from G_1, which we call T_2. Note that, T_2 is not necessarily a spanning tree of G. It is not always a spanning tree of G_1 as well since G_1 might not be connected. From T_2,we will add new virtual coordinates to the vertices of G, by starting from $4n$ (the reason will be clear later). Then we can define $G_2 = G - T_1 - T_2$. We repeat this process until all edges are consumed.

In [1], the authors proved that the virtual coordinates computed from T_1 is a greedy embedding. In the following, we will prove that the virtual coordinates computed from T_1 and T_2 is also a greedy embedding.

T_1 defines greedy embedding $\delta_1 : V \to \Pi_1$ as in subsection 1.3. Note that, all the numbers in $\{1, 2, \cdots, 2n-1\}$ have been used as virtual coordinates in the embedding. We will use d_1 to denote the semi-metric function on Π_1.

Consider T_2. It defines an embedding of $\delta_2 : V \to \Pi_2$, in which the virtual coordinates are from $\{(4n-1)+1, (4n-1)+2, \cdots, (4n-1)+(2k-1)\}$ (Namely, the virtual coordinates start from $4n$). Here note that k is the number of vertices in tree T_2. $k \leq n$, meaning tree T_2 may or may not be a spanning tree of graph G. We shift the virtual coordinates by $4n-1$ in order to avoid interaction between virtual coordinates of two different iterations.

Note that, for any $u, w \in V$, $\delta_1(u) \cap \delta_2(w) = \emptyset$. In addition, $|u_i - w_j| \geq 2n$ for any $u_i \in \delta_1(u)$ and $w_j \in \delta_2(w)$.

Let $\delta : V \to \Pi = \{\delta_1(v) \cup \delta_2(v) : v \in V\}$ be defined as follows: for each $v \in V$, $\delta(v)$ is defined as $\delta_1(v) \cup \delta_2(v)$. Obviously Π is a perfect sub-power set. We can define the same semi-metric function $d(*,*)$ as in subsection 1.2 for elements in Π. Consider $u \neq w$ be two vertices in G. We will use d_2 to denote the semi-metric function defined on Π_2. Here, $d_2(\delta_2(u), \delta_2(w))\}$ is defined as ∞ if either $\delta_2(u)$ or $\delta_2(w)$ is not defined.

We have the following claim:

Claim: $d(\delta(u), \delta(w)) = min\{d_1(\delta_1(u), \delta_1(w)), d_2(\delta_2(u), \delta_2(w))\}$.

Proof of the Claim:

First, we note that $d(\delta(u), \delta(w)) \leq min\{d_1(\delta_1(u), \delta_1(w)), d_2(\delta_2(u), \delta_2(w))\} < (2n-1)$.

Let $\delta(u) = \{u_1, \cdots, u_{k_u}\}$ and $\delta(w) = \{w_1, \cdots, w_{k_w}\}$. Suppose $d(\delta(u), \delta(w)) = |u_a - w_b|$ for $1 \leq a \leq k_u$ and $1 \leq b \leq k_w$.

We have the following several cases.

Case 1: $u_a \in \delta_1(u)$ and $w_b \in \delta_2(w)$, then $|u_a - w_b| \geq 2n$, which is not possible, since $d(\delta(u), \delta(w))$ is always $< 2n - 1$.

Case 2: $u_a \in \delta_2(u)$ and $w_b \in \delta_1(w)$, then $|u_a - w_b| \geq 2n$, which is not possible either. Therefore, either $u_a \in \delta_1(u)$ and $w_b \in \delta_1(w)$, or $u_a \in \delta_2(u)$ and $w_b \in \delta_2(w)$. Therefore, $d(\delta(u), \delta(w)) \geq min\{d_1(\delta_1(u), \delta_1(w)), d_2(\delta_2(u), \delta_2(w))\}$. Hence, $d(\delta(u), \delta(w)) = min\{d_1(\delta_1(u), \delta_1(w)), d_2(\delta_2(u), \delta_2(w))\}$.

End of the proof of the claim. What the claim really means is that the distance between any two vertices $u \neq w$ is either their distance in the greedy embedding from T_1, or their distance in the greedy embedding from T_2. Suppose that $d(\delta(u), \delta(w)) = d_i(\delta_i(u), \delta_i(w)), i \in \{1, 2\}$. Since both δ_1 and δ_2 are greedy embeddings, it is easy to see that there is a neighbor v of u such that $d(\delta(v), \delta(w)) < d(\delta(u), \delta(w))$. Therefore, there is a distance decreasing path between u and w in G. Therefore, $\delta : V \to \Pi$ is a greedy embedding of G onto the perfect sub-power set Π.

Apparently, by mathematical induction the above process can be easily generalized to use k additional trees that can be extracted from graph G. Tree will be extracted and its edges will be deleted from graph G. Recursively, the process will end up in empty graph with all vertices n possessing a set of virtual coordinates. Note that, this greedy routing algorithm takes advantage of all the trees. In some sense, the message is forwarded to a neighbor which is the nearest with respect to all the greedy embeddings obtained from all the trees. In other words, trees can be *merged* to make a new greedy embedding. This is a feature that none of the other greedy routing algorithms has. It is conceivable that, if we select trees on the basis of highest virtual coordinate distance $d(\delta(u), \delta(w))$, the forwarding process will speed up and hence the stretch factor will be lowered.

The algorithms to allocate virtual coordinates to tree runs in linear time. The virtual coordinate for each vertex v is a finite set consisting of up to $2deg(v)$ natural numbers (the root of each true has one more virtual coordinate than its degree in the tree). Each tree inserts at most $2n - 1$ virtual numbers. In the recursion of adding additional trees, each tree consumes at least one edge. Therefore, totally, we have at most quadratic number in terms of n of additional trees. Therefore, every virtual coordinate used is bounded by a polynomial of n, which can be represented in $O(logn)$ bits. Note that, for two vertices u and w, they have at most $2deg(u)$ and $2deg(w)$ virtual coordinates respectively. Note that, the virtual coordinates are in inserted in sorted order. By modifying merge sub procedure in the Merge Sort, it is easy to see that $\delta(u, w)$ can be computed in $max(deg(u), deg(w)$ time, it is bounded by $\Delta(G)$. Therefore, for one node to find a neighbor to forward a message, the time is bounded by $\Delta(G)^2$. Note that, if the graph G is a bounded degree graph, all the virtual coordinates for a vertex can be represented succinctly, and the time to find a neighbor to forward a message is in constant time.

3. Experimental Study

In this paper, we conduct an experiment to test how $\mathcal{S}(u, w)$ behaves with the proposed algorithm. For the purpose of this experimental study on connected n-node graph, we coded program in java. Program is mainly divided in two phases.

3.1 Random Graph Generation

In first phase, random graph is generated within 3 steps. In **Step 1**, we create n node connected graph by using function *random graph generator*. The function accepts two inputs as nodes n and edges e to populate two dimensional adjacency matrix. The size of the matrix is constrained by the number of nodes passed into the function. Two random numbers n_1 and n_2 representing nodes are then generated within the range of 1 to $n - 1$. These nodes are chosen in such a way that $n_1 \neq n_2$ and there exist no edge between n_1 and n_2. If both conditions are true then an edge is established between them. This edge is represented in matrix by inserting 1 at n_1th column and n_2th row as well as n_2th column and n_1th row. After executing this function, another function *isConnected()* is invoked to asserts if all the nodes of graph G are connected. If the function returns true, it means all nodes are now connected. Then the control is navigated away from the loop. If the function returns false then the process is repeated until all nodes gets connected. Within *isConnected()* function, the breadth first tree traversal algorithm is executed to ensure the graph is connected. Every time node 1 is selected as root node to initiate traversal. Once every node is connected we exit the while loop. **Step 2** of the random graph generation makes sure that the connected graph has specified number of edge

count *e*. Function *addEdge()* and *removeEdge()* handles the responsibility to produce exact edge count *e* in n node connected graph *G*. The graph generated by step 1 contains e' edges, where $e' < e$ or $e' > e$. In this case, one of the two methods is executed to increase or decrease the edge count in the graph. The *removeEdge()* method calls *isConnected()* to ensure the graph is still connected after the removal of the specified edges. In case the graph is not connected, this results in the production of a forest. To remedy this setback, the *isGraphConnected()* function replaces the removed edge with an alike edge that keeps the graph united. **Step 3** of the random graph generation is the last step of the phase I. In this step, a graph in the form of adjacency matrix is written to a file in binary format. This file is then stored on a hard disk for forthcoming analysis in phase II.

3.2 Execution of greedy routing algorithm

In the second phase, we use the existing file created in phase I. The binary file is read by the program to create and populate the adjacency matrix. This adjacency matrix represents a graph. For calculating the stretch factor of a graph, we follow these five steps.

3.2.1 Extract tree from a graph

In this step, a tree (T) is extracted from a graph (G) by using BFS traversal. Virtual coordinates are then allocated to tree T. This tree is stored in adjacency matrix. The root node for BFS traversal is decided randomly. At the same time, we create a co-tree matrix (T') to denote all the edges of graph G which are not in the tree, but exist in the graph.

3.2.2 Execute Floyd-Warshell's algorithm

In this step, we execute Floyd's algorithm to calculate distance between tree nodes. An adjacency matrix is populated with all the distances. This algorithm runs in n cubic time. This adjacency matrix contains all pair shortest path.

3.2.3 Allocate virtual coordinates to tree

From this step onwards, we use our proposed algorithm to calculate the distance between two nodes. In this step, first we allocate virtual coordinates to all nodes of tree (T). Each node n in a tree has a list of virtual coordinates. The virtual distance $\delta(u, v)$ between two connected adjacent nodes u, v will always be 1. The maximum cardinality $\mathcal{T}_\Pi$ of any node v in a tree cannot be more than $deg(v)$.

3.2.4 Calculate stretch factor

In this step, the *stretch factor* $\mathcal{S}(u, w)$ of graph G is calculated by considering maximum ratio of $\delta(u, v)$ and $E(u, v)$. At this stage, graph contains many edges E' which have been explored by Floyd's algorithm but not by our algorithm.

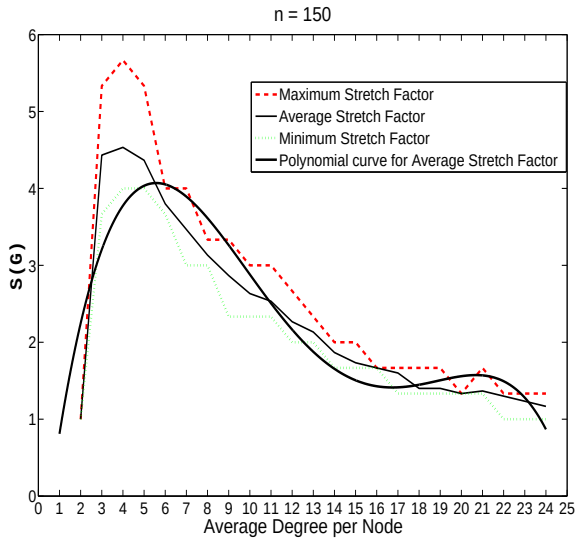


Fig. 2: Behavior for $node = 150$ at different average degrees per node.

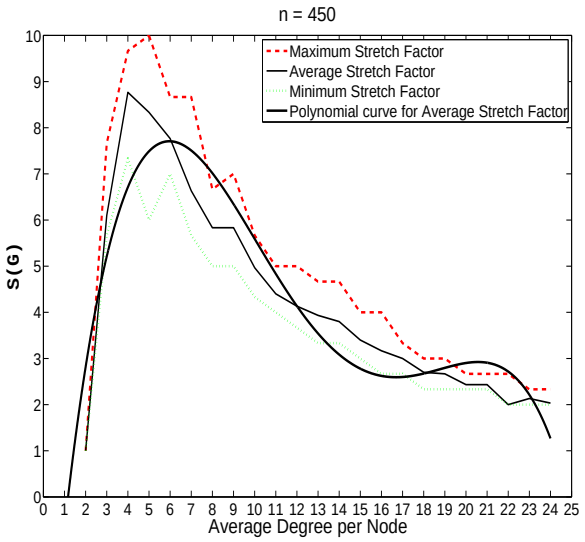


Fig. 3: Behavior for $node = 450$ at different average degrees per node.

3.2.5 Allocate virtual coordinates to co-tree

In this step, the algorithm is executed on all remaining edges of co-tree matrix T'. This co-tree matrix can be a tree, a forest or a graph. In order to explore all edges of the co-tree, we recursively execute following steps; **1.** Extract a tree such that this tree contains an edge with *maximum virtual distance* $\delta(u, v)$ from E'. **2.** Consider either of the nodes of maximum virtual distance edge as root node while extracting a tree. **3.** Execute the algorithm to allocate virtual coordinates to tree nodes. This process may add new virtual

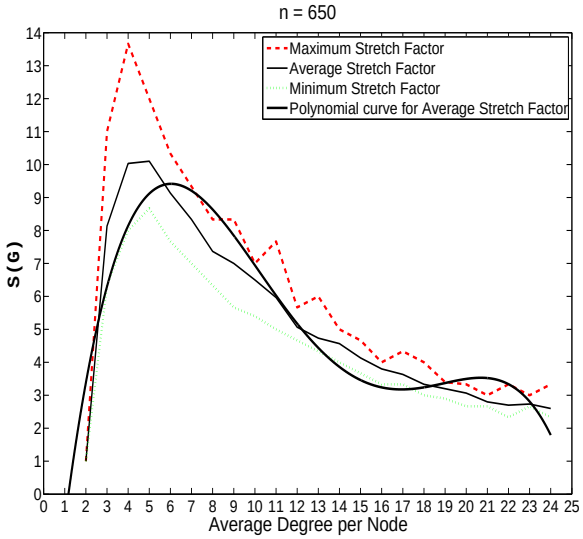


Fig. 4: Behavior for $node = 650$ at different average degrees per node.

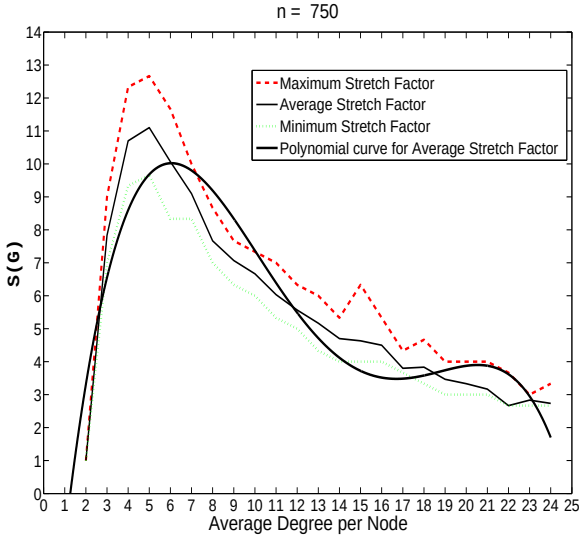


Fig. 5: Behavior for $node = 750$ at different average degrees per node.

coordinates to nodes of graph. We remove all the tree edges from matrix. **4.** Calculate the stretch factor of a graph. The recursive call explores all edges of graph G. The program is halted if all edges gets explored or the *stretch factor* $\mathcal{S}(u, w)$ of a graph reaches to 1. In later case there is no use of executing algorithm repeatedly because stretch factor cannot be improved further. We keep a track of number of iterations required to reach stretch factor up to 1. Basic idea while allocating virtual coordinates is to bring farther nodes closer which have an existing edge between them. It reduces the routing distance between such adjacent nodes to 1. We also

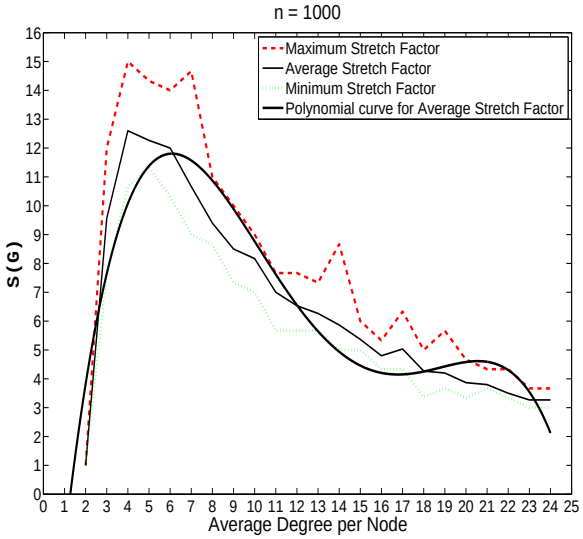


Fig. 6: Behavior for $node = 1000$ at different average degrees per node.

calculate the *cardinality* of a graph. The degree of a node n possessing maximum virtual coordinates is also measured.

4. Experimental results

For our experimental study, we used a laboratory controlled environment consisting of 15 computers with $i7$ processors on a Windows 7 operating system. Each program was executed in the Eclipse IDE. It took roughly 1 hour for a graph of size 1000 nodes and 10000 edges to reach stretch factor $\mathcal{S}(u, w) = 1$. Based on the experimental data produced by each program we plotted the following graphs.
a) Average degree per node Vs stretch factor.
b) Degree fix : Number of nodes per graph Vs stretch factor.
c) Iterations Vs stretch factor.
d) Fix ratio of node and degree Vs stretch factor.

4.1 Fig. 2 - Fig. 6

Fig. 2 is plotted with average degree per node in graph on x-axis and stretch factor $\mathcal{S}(u, w)$ on the y-axis. It represents an average degree varying from 1 to 24 for 150 nodes. It also has 3 graphs representing maximum; which is worst case stretch factor, the minimum; which is best case stretch factor and an average case stretch factor behavior. The single point in a graph is calculated by taking the average of 10 graphs of the same degree. The maximum point for a degree specifies the maximum stretch factor received out of 10, whereas the minimum point for a degree specifies the lowest stretch factor degree received for the same degree. The average for a degree is determined by taking the average of 10 stretch factors.

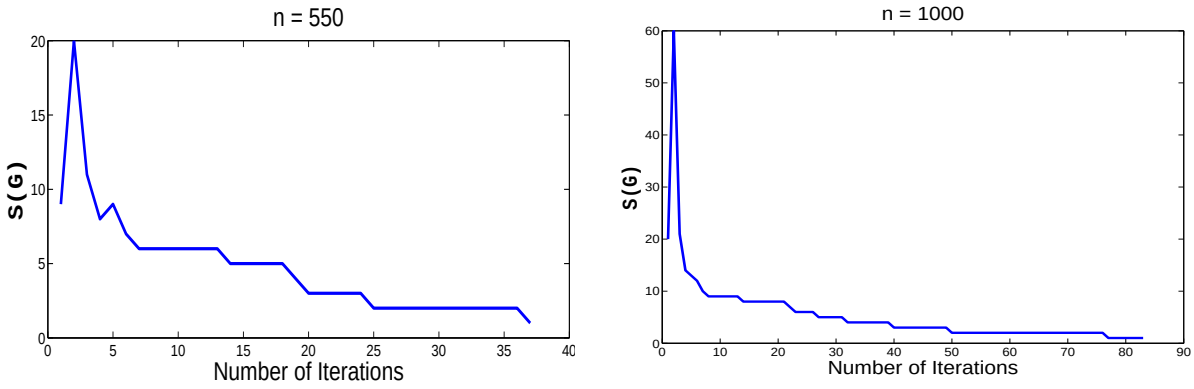


Fig. 7: Iterations for nodes 550 and 1000

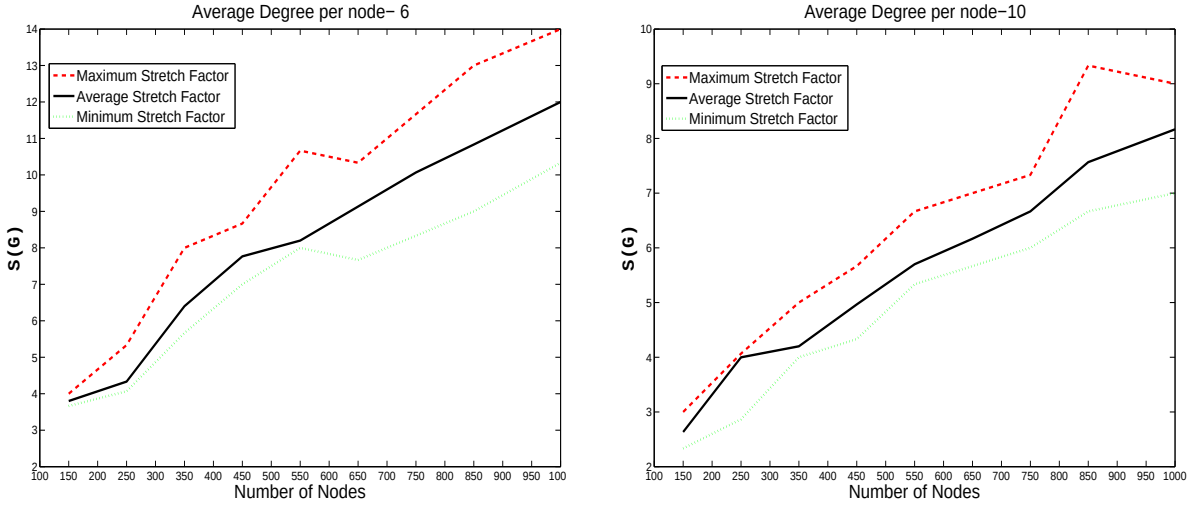


Fig. 8: Behavior for $degree = 6$ and $degree = 10$

In addition, a polynomial trend line with degree 4 is drawn for average case stretch factor $\mathcal{S}(u, w)$ graph. It can be observed from Fig. 2 that the trend line follows a certain pattern in which it spikes up and slowly declines until reaching a stabilized state of stretch factor 1. Moreover, an additional observation was identified pertaining to the stretch factor. In which, if the node count is lower then the stretch factor is also lower during the initial spike. It is observed that, as number of nodes in a graph increases, the maximum stretch factor also increases accordingly. High pick in stretch factor is observed when the graph is less dense. As the number of edges in graph increases, then stretch factor decreases accordingly. The remaining Fig. 3 to Fig. 6 reflects the same profile and behavior as described above.

4.2 Fig. 7

Fig. 7 represents the number of iterations required until stretch factor reaches to 1 for node 550 and 1000. X-axis is plotted with number of iterations and the stretch factor is on the y-axis with a constant node count. We have observed that, after the first iteration, the stretch factor reaches it's maximum which indicates a high spike in the graph. However, every iteration thereafter leads to a decline in the stretch factor which eventually leads to a stretch factor of 1, assuming the graph is dense.

4.3 Fig. 8

Fig. 8 is plotted with fix average degree 6 and degree 10 per node in the graph. The x-axis represents an edge to node ratio whereas the y-axis represents the stretch factor. It's observed that, for a fixed average degree per node, the stretch factor increases with constant rate. After certain number of iterations, stretch factor improves to 1 in case of dense graphs. The average degree of a node in graph is directly proportional to the cardinality of a node.

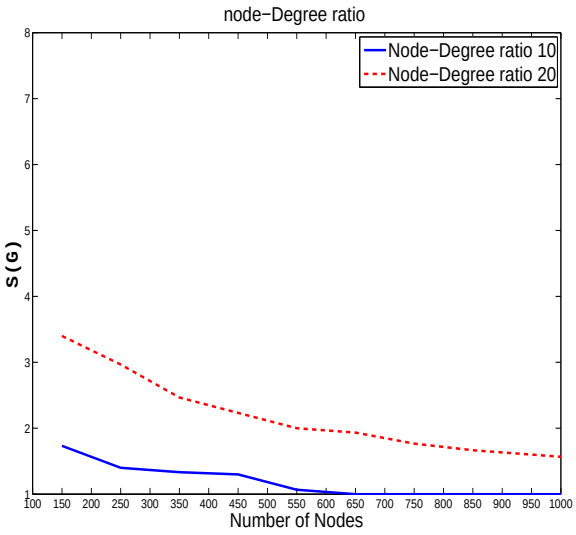


Fig. 9: Constant node/degree ratio.

4.4 Fig. 9

Fig. 9 is plotted with fix ratio of nodes and average degree against stretch factor. For Fig. 9, the ratio is considered as 10 for lower continuous (blue) line and 20 for higher dotted (red) line. We have observed that, if the number of nodes are increased in proportion with degree, stretch factor stabilizes and tends to be almost constant.

5. Conclusions

In this paper, we conducted an experimental study on the greedy routing algorithm by embedding a graph G in semi-metric spaces. We iteratively insert virtual coordinates to nodes by exploring sub-trees of G. The virtual coordinates of each node u can be represented in $deg(u)O(logn)$ bits. The distance between any two nodes u and w can be computed in $max(deg(u), deg(w))$ time. Therefore, for a bounded-degree graph (for example, in general, UDGs are bounded-degree graphs), the virtual coordinates can be represented succinctly and the distance between any two nodes can be computed in constant time. Empirically, we extensively evaluated stretch factors of our greedy routing algorithm. We observed that stretch factors are always small constant numbers and they show certain patterns. To our best knowledge, this is the first study towards a practical low-stretch-factor greedy routing algorithm.

References

[1] H. Zhang and S. Govindaiah, "Greedy routing via embedding graphs onto semi-metric spaces," in *Frontiers in Algorithmics and Algorithmic Aspects in Information and Management*, ser. Lecture Notes in Computer Science, M. Atallah, X.-Y. Li, and B. Zhu, Eds. Springer Berlin / Heidelberg, 2011, vol. 6681, pp. 58–69.

[2] A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica, "Geographic routing without location information," in *Proceedings of the 9th annual international conference on Mobile computing and networking*, ser. MobiCom '03. New York, NY, USA: ACM, 2003, pp. 96–108.

[3] D. Eppstein and M. T. Goodrich, "Graph drawing," in *In Proc. 16th Int. Symp. Graph Drawing*, I. G. Tollis and M. Patrignani, Eds. Berlin, Heidelberg: Springer-Verlag, 2009, ch. Succinct Greedy Graph Drawing in the Hyperbolic Plane, pp. 14–25.

[4] P. Angelini, G. Di Battista, and F. Frati, "Succinct greedy drawings do not always exist," in *Graph Drawing*, ser. Lecture Notes in Computer Science, D. Eppstein and E. Gansner, Eds. Springer Berlin / Heidelberg, 2010, vol. 5849, pp. 171–182.

[5] P. Angelini, F. Frati, and L. Grilli, "An algorithm to construct greedy drawings of triangulations," in *Graph Drawing*, ser. Lecture Notes in Computer Science, I. Tollis and M. Patrignani, Eds. Springer Berlin / Heidelberg, 2009, vol. 5417, pp. 26–37.

[6] R. Dhandapani, "Greedy drawings of triangulations," in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, ser. SODA '08. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2008, pp. 102–111.

[7] M. Goodrich and D. Strash, "Succinct greedy geometric routing in the euclidean plane," in *Algorithms and Computation*, ser. Lecture Notes in Computer Science, Y. Dong, D.-Z. Du, and O. Ibarra, Eds. Springer Berlin / Heidelberg, 2009, vol. 5878, pp. 781–791.

[8] A. Moitra and T. Leighton, "Some results on greedy embeddings in metric spaces," in *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 337–346.

[9] X. He and H. Zhang, "Schnyder greedy routing algorithm," in *Theory and Applications of Models of Computation*, ser. Lecture Notes in Computer Science, J. Kratochvíl, A. Li, J. Fiala, and P. Kolman, Eds. Springer Berlin / Heidelberg, 2010, vol. 6108, pp. 271–283.

[10] X.He and H. Zhang, "On succinct convex greedy drawing of 3-connected plane graphs." in *SODA*, D. Randall, Ed. SIAM, 2011, pp. 1477–1486.

[11] C. H. Papadimitriou and D. Ratajczak, "On a conjecture related to geometric routing," *Theor. Comput. Sci.*, vol. 344, no. 1, pp. 3–14, Nov. 2005.

[12] R. B. Muhammad, "A distributed geometric routing algorithm for ad hocwireless networks," in *Proceedings of the International Conference on Information Technology*, ser. ITNG '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 961–963.

[13] R. Kleinberg, "Geographic routing using hyperbolic space," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, may 2007, pp. 1902 –1909.

[14] M. B. Chen, C. Gotsman, and C. Wormser, "Distributed computation of virtual coordinates," in *Proceedings of the twenty-third annual symposium on Computational geometry*, ser. SCG '07. New York, NY, USA: ACM, 2007, pp. 210–219.

Enhanced Multipath Routing And Proactive Route Maintenance For Congestion Minimization And Transient Link Failures

Selvamani K[1], Kanimozhi S[2], Elakkiya R[3] and Kannan A[4]
[1,3]Department of Computer Science and Engineering, Anna University, Chennai, Tamilnadu, India
[2,4]Department of Information Science and Technology, Anna University, Chennai, Tamilnadu, India

Abstract - *The usage of internet is highly increasing for mission-critical applications and hence internet is always expected to be available for the users. But unfortunately service disruptions happened even in well designed and managed networks due to link failures and most of the failures are transient. Existing works on multipath routing scheme has focused only on the heuristic methods. But in this proposed work, we focus on optimal congestion reduction schemes to minimize the congestion in a network. This deployed link state routing protocols react to link failures through the global link state and the routing table re-computation by causing significant discontinuity in forwarding after the failure. This paper aims in providing fast global re-routing called Failure Insensitive Routing (FIR) for handling these transient link failures and efficient multipath routing for congestion minimization. This proposed approach detects the link failures and forwards the packets using interface-specific forwarding technique and suppresses the link state advertisement and triggers the local re-routing using backward table. This approach also aims to formulate multipath routing to restrict the quality (length) of the selected paths and the number of routing paths per destination.*

Keywords: Fast Rerouting; Transient Failures; Multipath Routing; Congestion Avoidance;

1 Introduction

Currently deployed protocols called Link-state protocols perform global routing table to route around the failures which usually takes few seconds. It also raises the congesting situation due to network link failures. Whenever a link fails in a network, the traffic is rerouted around the failure leading to congestion on a link, which will increase the traffic than ever before. In such cases, a mechanism called congestion control is used to maximize the throughput and to minimize the impact of the packet loss. In real-time applications, such as VoIP which has emerged in recent years requires a fast rerouting mechanism before all routers on the network update their routing tables. In addition to this fast rerouting is more appropriate than global routing the link when failures are transient. Moreover, to support time-sensitive applications, the network needs to survive failures with minimal disruption to the service. Hence it is necessary to develop a fast local routing protocol for link state protocols to handle transient link failures and efficient multipath routing algorithm for congestion minimization.

2 Related Works

Multi-Protocol Label Switching (MPLS) based approaches to failure recovery [5] leverage explicit routing for fast rerouting. An explicitly routed protection LSP (Label Switched Path) is set up to provide a backup path for each vulnerable physical link and acts as a parallel virtual link. When the physical link fails, the upstream node switches traffic from the physical link to the virtual link. The label stacking capability of MPLS is used to re-route all the LSPs that used to go over the failed link by nesting them into the protection LSP. Since rerouting is done locally at the point of failure without the need to perform any signaling at the time of failure, MPLS can handle transient failures effectively with minimal disruption to forwarding of data. However, deployment of MPLS necessitates changes in the forwarding plane of traditional routers. Our objective is to provide fast local rerouting to deal with transient link failures with minimal changes to the current networking infrastructure.

Previous studies and proposals on Multi-Path routing in the previous context have focused on heuristic methods. In [2], a Multi-Path routing scheme, termed Equal Cost Multi-Path (ECMP), has been proposed for balancing the load along multiple shortest paths using a simple round-robin distribution. By limiting itself to shortest paths, ECMP considerably reduces the load balancing capabilities of multipath routing; moreover, the equal partition of flows along the (shortest) paths (resulting from the round robin distribution) further limits the ability to decrease congestion through load balancing. OSPF-OMP [2] allows splitting traffic among paths unevenly; however, that often results in an inefficient flow distribution. Both [1] and [2] considered multipath routing as an optimization problem with an objective function that minimizes the congestion of the most utilized link in the network. In this paper, we focus on

multipath routing algorithms that both select the routing paths and split traffic among them.

The pathology of link overload [9] in the network is studied and has proposed a deflection routing algorithm to alleviate overload by exploiting the highly meshed nature of the backbone and a judicious use of link weights. This proposal is based on the Sprint network which is built using a pure IP philosophy, though it can be applied to other networks, say those enabled with MPLS. In blacklist-based interface-specific [7] forwarding (BISF) is proposed that infers a blacklist, a list of links that might have failed, based on a packet's incoming interface and its destination, and determines the next-hop by excluding the blacklisted links. To enhance failure resiliency without jeopardizing routing stability, a local rerouting based approach [8] called failure insensitive routing is demonstrated. In Tabu-search heuristic [10] is proposed for choosing link weights which allow a network to function almost optimally during short link failures. The heuristic takes into account possible link failure scenarios when choosing weights, thereby mitigating the effect of such failures.

3 Proposed Methodology

This paper aims in combining the enhanced FIR protocol for handling link failures and multipath algorithm for congestion minimization which provides better quality of source over link failures and network congestion. The proposed techniques for handling link failures and congestion minimization are explained as follows.

3.1 Handling Transient Link Failures

In this proposed system, local rerouting called Enhanced Failure Insensitive Routing (EFIR) for handling the link failures and enhanced Multipath algorithms for congestion minimization are considered together to obtain better quality of service in case of network failure. Moreover, the idea is to calculate backup paths in advance before the network link fails. Whenever a failure is detected, the interrupted packets are immediately forwarded through the identified backup paths in order to minimize the service disruption. The two main key ideas that underpin this Proactive enhanced FIR approach based on local rerouting are interface-specific forwarding and Failure Inferencing. Under FIR, routers infer link failures based on packets and pre-compute interface-specific forwarding tables (backup paths) in a distributed manner and trigger local rerouting without relying on network-wide link-state advertisements. Hence, this proposed EFIR enhances failure resiliency and routing stability by suppressing the advertisement of transient failures and locally rerouting packets along loop-free paths during the suppression period.

3.2 Congestion Minimization

This second technique that performs enhanced Multipath routing called Restricted K-Path Routing (RKPR) for congestion minimization which combines the two algorithms called Restricted Multipath (RMP) and K-Path Routing (KPR). The main idea is to select paths based on quality of service and number of paths detected in the network. To minimize the route congestion and to minimize the service disruption, the packets are forwarded through multiple paths. In this work, two fundamental constraints are considered. First, each path that has been selected for routing should have adequate quality for sending the packets. Hence it is necessary to prohibit the substantially inferior paths which are considered as longer path. Second, practical restriction is on the number of routing paths per destination. Therefore, in practice, it is desirable to use few paths as possible to minimize the network congestion. Through comprehensive simulations, this system proves that Multipath solutions obtained by optimal congestion reduction schemes are fundamentally more efficient than solutions obtained by heuristics. This approach reduces the overheads and congestion resulted in earlier works.

4 Concept of Enhanced FIR

In EFIR technique, when a link fails, the adjacent node to it locally reroute the packets to the affected destinations and all other nodes simply forward packets according to their pre-computed interface-specific forwarding tables without being explicitly aware of the failure. Once the failed link comes up again, forwarding resumes over the recovered link. Figure 1.1 shows the topology used for simulation of our technique EFIR which is considered in [7].

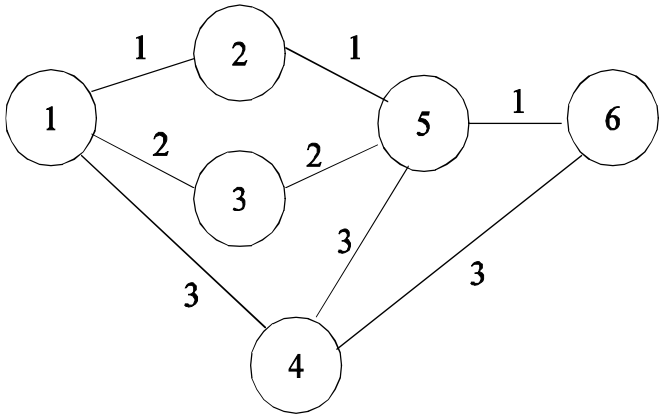


Figure 1.1 Topology used for simulation of EFIR

The algorithm for the proposed EFIR technique is documented in figure 1.2.

EFIR Implementation

Input : Transient Link Failure.

Output: Local rerouting of packets during failure.

Algorithm:

i. Start

ii. Identify the key links (set of links whose individual or combined failure causes the node to reroute the packet to its parent).

iii. Compute the shortest path by eliminating the key links with the help of fuzzy rules.

iv. Perform routing through the path found in step-iii.

v. Repeat the step-ii, iii and iv during each transient failure.

vi. Stop

Figure 1.2 Algorithm for EFIR

Using the conventional forwarding tables, local rerouting is not viable as it causes forwarding loops. Under enhanced technique of EFIR, forwarding loops are avoided by inferring a link's failure from packet's incoming interface. When the path 2–5 is down in the figure 1.1, node 2 locally reroutes packets from nodes 1 to 6 back to node 1 instead of dropping them. When a packet destined to 6 arrives at node 1 from node 2, then node 1 can infer that some link along its shortest path to 6 must have failed, as otherwise, 2 should never forward packets destined to 6 to node 1. The key links associated with the interface and destination 6 is 2–5, 5–6. It should be noted that these inferences about potential link failures are made not on the fly but in advance.

5 Multipath Algorithms

The Paths for rerouting the packets from source to destination are selected based on these two main algorithms namely Restricted Multi-Path algorithm (RMP) and K-Path Routing algorithm. These two algorithms are illustrated as follows.

5.1 RMP (Restricted MultiPath)

Consider a network G(V,E), two nodes s,t $\in$ V , a length le>0, and a capacity for each link ce>0, a demand , and a length restriction L for each routing path. Find a feasible path flow that minimizes the network congestion factor such that, if P$\subseteq$ P(s, t) is the set of paths in P(s, t) that are assigned a positive flow, then, for each p$\in$ P, it holds that L(p)$\leq$ L.

Remark 1: For convenience, and without loss of generality, we assume that the length le of each link e$\in$ E is not larger than the length restriction L. Clearly, links that are longer than L are erased.

5.2 KPR (K-Path Routing)

In a given network G(V,E), the two nodes s, t $\in$ V, a capacity ce>0 for each link e$\in$ E , a demand λ > 0, and a restriction on the number of routing paths K. Find a feasible path flow that minimizes the network congestion factor, such that, if P$\subseteq$ P(s, t) is the set of paths in P(s, t) that are assigned a positive flow, then $|P| \leq K$

Input : Network traffic from network layer, length, flow rate.

Output: Multipath routing with minimum number of path and reduced flow rate.

Algorithm of the RKPR:

i. Start

ii. Create a traffic type of RMP and KPR.

iii. Create the integrated traffic type that selects maximum of three paths to any destination in the network using c++.

iv. Also, reduce the flow rate using the agent created using c++ for the path selected in step-iii.

v. Multicast the packets using the traffic agent created to the destination.

vi. Stop

Figure 1.3 Algorithm for RKPR

Remark 2: In both problems, the source to destination pair (s, t) is assumed to be connected, i.e., $|P(s, t)| \geq 1$.

Remark 3: In both problem formulations, it is possible to limit the link congestion factor fe/ce of each e$\in$ E to any desired congestion level αe by replacing the given capacity value ce with a new capacity value αe•ce. Clearly, the capacity constraint fe $\leq$ αe•ce (that both problems must satisfy) assures that the link congestion factor would be at most αe.

6 Results and Performance Evaluation

Our performance evaluations are based on the simulations of seven nodes as shown in figure 1.4 for handling link failure and 25 nodes for congestion minimization scenario that form a network with symmetric links over a rectangular (1500m * 500m) flat space in 100 sec of simulation time. The minimum speed for each node as 2ms except for the static case, and vary the maximum speed to 15ms to evaluate the impact of Transient Failure on our routing performance. Figure 1.5 shows the routing of enhanced FIR during failure of link 1-4. Figure 1.6 shows the ratio between the network congestion produced by an optimal Multipath routing assignment and the network congestion produced by ECMP using wax man topology.

Table 1.1 shows the implementation analysis of FIR that compares the throughput of OSPF with and without FIR.

TABLE 1.1 ANALYSIS OF LINK FAILURE WITH AND WITHOUT FIR

Time in sec	No. of packets received in bytes		
	OSPF without link failure	During link link failure without FIR	During link failure with FIR
3	645518	530020	566923
12	1154200	764520	1033032
27	1154230	1153620	1153930
36	1153684	1134142	1153600
69	91	91	91

The following graphs shows the simulated results obtained from the implementation using NS-2 for the proposed method. This hybrid approach proves in considerable amount of performance over the existing individual methods.

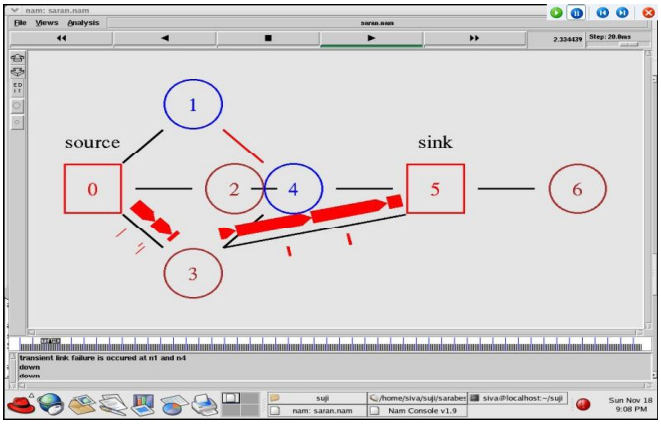


Figure 1.4 Rerouting by EFIR in NS-2

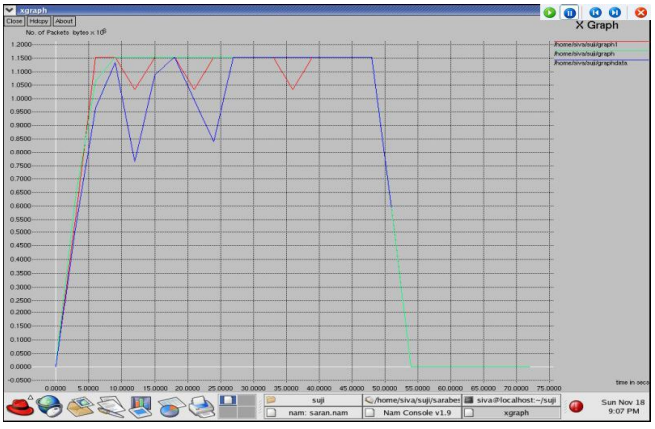


Figure 1.5 Xgraph with and without EFIR

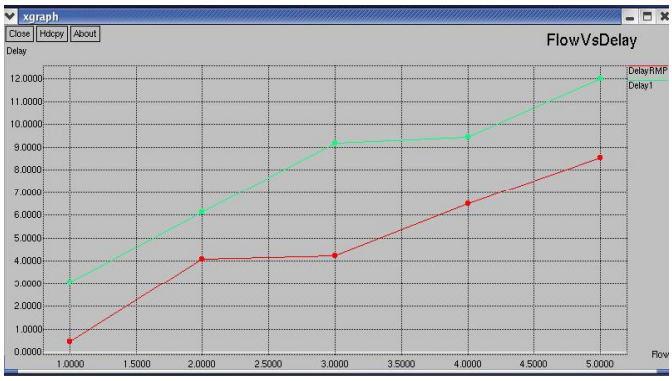


Figure 1.6 Congestion Rate for OMP and RKPR

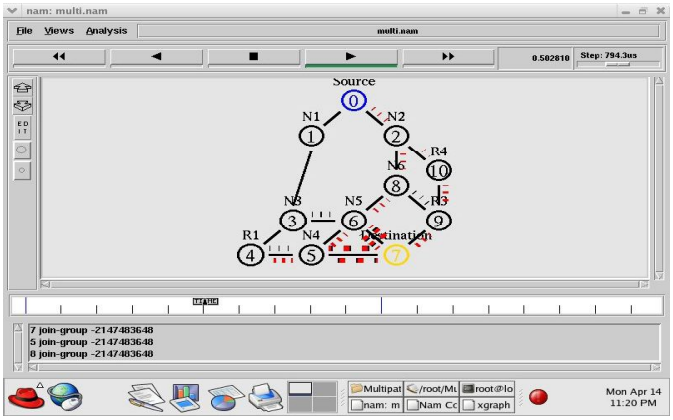


Fig 1.7 Simulation of RKPR in NS-2

7 Conclusion and Future Work

One fundamental challenge for IP Backbone network is to provide uninterrupted service availability during transient link failures and congestion. This proposed work explores a novel Pro-active approach for handling such transient link failures and an optimal congestion reduction scheme for congestion minimization. This algorithm can drastically reduce the network congestion at the price of routing along minimum number of paths that are slightly longer than the shortest path.

In this work, we have presented a network layer routing solution that performs local rerouting and a Multipath routing in a unified framework. RMP-KPR algorithm restricts the number of path and flow rate to reduce the congestion along the path. We have also developed a formal model to analyze the routing stability and network availability under both Multipath and RMP-KPR approaches through simulations and graph comparison. Our results indicate that the improvement due to hybrid RMP-KPR algorithm is markedly better when the path and flow rate are restricted.

Our routing mechanism assumes a forwarding table per each interface and is applicable to links with symmetric weights. One possible extension is to deploy FIR in networks with asymmetric weights, broadcast LANs, and multiple areas. The distributed implementation of Algorithm RMP remains an open issue for future investigation. Also there is a possibility of providing security and energy efficient to make ease of the multipath routing that can minimize the network target.

8 References

[1] Ron Banner, Ariel Orda, "Multipath Routing Algorithms for Congestion Minimization", IEEE/ACM Transaction on Networking, Vol. 15, Issue 2, pp. 413-424, Apr 2007.

[2] H.Han, S.Shakkottai, C.V.Hollot, R.Srikanth, and D.Towsley, "Multipath TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet", IEEE/ACM Transaction on Networking, Vol. 14, Issue 6, pp.1260-1271, Dec 2006.

[3] Merindol Pascal, Pansiot Jean-Jacques and Cateloin Stephane, "Path Computation for Incoming Interface Multipath Routing" in ECUMN'07: Proceedings of the Fourth European Conference on Universal MultiService Networks, 2007.

[4] Kyeongia Lee, Armand Toguveni and Ahmed Rahmani, "Hybrid Multipath Routing Algorithms for Load Balancing in MPLS Based IP Networks" in AINA'06: Proceedings of the 20th International Conference on Advanced Information Networking and Application, 2006.

[5] Wei Lin, Bin Liu and Yi Tang, "Traffic Distribution over Equal-Costs-Multi- Paths using LRU-based Caching with Counting Scheme", in AINA'06: Proceedings of the 20th International Conference on Advanced Information Networking and Application, 2006

[6] Larry L.Peterson & Bruce S.Davie, "Computer Networks – A systems Approach", Morgan Kaufmann publishers, second edition, 2000.

[7] S.Nelakuditi, Sanghwan Lee, Yinzhe Yu, Zhi-Li Zhang and Chen-Nee Chuah, "Fast Local Rerouting for Handling Transient Link Failures", IEEE/ACM Transactions on Networking, Vol. 15, Issue 2, pp: 359-372, April 2007.

[8] http://ftp.cse.sc.edu/reports/drafts/2007-001-bisf.pdf, "Handling Failures in IP Networks through Interface Specific forwarding"

[9] A. Nucci, B. Schroeder, S.Bhattacharyya, N. Taft, and C. Diot, "IGP Link Weight Assignment for Transient Link Failures", Presented at the ITC18, Berlin, Germany, 2003.

[10] S. Iyer, S. Bhattacharyya, N. Taft, and C. Diot, "An Approach to Alleviate Link Overload as Observed on an IP Backbone," in Proceedings of IEEE INFOCOM, pp. 406–416, 2003.

CFO Estimation Schemes Using the Cyclic Prefix for OFDM Systems in Non-Gaussian Noise Environments

Changha Yu, Jong In Park, and Seokho Yoon†
College of Information and Communication Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do, Korea
†Corresponding author

Abstract— *In this paper, carrier frequency offset (CFO) estimation schemes robust to the non-Gaussian noise for orthogonal frequency division multiplexing (OFDM) systems are proposed. Applying the probability density function of the cyclic prefix of OFDM symbols to the maximum-likelihood (ML) criterion, we propose the ML and low-complexity ML estimation schemes. Simulation results show that the proposed schemes offer a robustness and a substantial performance improvement over the conventional estimation scheme using cyclic prefix in non-gaussian noise environments.*

Keywords: carrier frequency offset; cyclic prefix; maximum-likelihood; non-Gaussian noise; OFDM

1. Introduction

Due to its immunity to multipath fading and high spectral efficiency, orthogonal frequency division multiplexing (OFDM) has been adopted as a modulation format in a wide variety of wireless systems such as digital video broadcasting-terrestrial (DVB-T), wireless local area network (WLAN), and worldwide interoperability for microwave access (WiMAX) [1]-[3]. However, OFDM is very sensitive to the carrier frequency offset (CFO) caused by Doppler shift or oscillator instabilities, and thus, the frequency offset estimation is one of the most important technical issues in OFDM systems [1], [4]. Specifically, we are concerned about the FO estimation based on the blind approach, which uses the cyclic prefix (CP) of OFDM symbols [4].

Conventionally, the CFO estimation schemes have been proposed under the assumption that the ambient noise is a Gaussian process [5], which is generally justified with the central limit theorem. However, it has been observed that the ambient noise often exhibits non-Gaussian nature in wireless channels, mostly due to the impulsive nature originated from various sources such as car ignitions, moving obstacles, lightning in the atmosphere, and reflections from sea waves [6], [7]. The conventional estimation schemes developed under the Gaussian assumption on the ambient noise could suffer from severe performance degradation under such non-Gaussian noise environments.

In this paper, we propose robust blind CFO estimation schemes in non-Gaussian noise environments. Based on the CP structure of OFDM, we first derive a maximum-likelihood (ML) CFO estimation scheme in non-gaussian noise modeled as a complex isotropic Cauchy noise, and then, derive a simpler blind estimation scheme with a lower complexity. From simulation results, the proposed schemes are confirmed to offer a substantial performance improvement over conventional blind estimation scheme in non-Gaussian noise environments.

2. Signal Model

The kth received OFDM sample $r(k)$ can be expressed as

$$r(k) = x(k)e^{j2\pi k\varepsilon/N} + n(k) \tag{1}$$

for $k = -G, \cdots, -1, 0, 1, \cdots, N-1$, where $x(k)$ is the kth sample of the transmitted OFDM symbol generated by the inverse fast Fourier transform (IFFT) of size N, G is the size of the CP, ε is the CFO normalized to the subcarrier spacing $1/N$, and $n(k)$ is the kth sample of additive noise.

In this paper, we adopt the complex isotropic symmetric α stable (CISαS) model for the independent and identically distributed noise samples $\{n(k)\}_{k=0}^{N-1}$; this model has been widely employed due to its strong agreement with experimental data [8], [9]. The probability density function (pdf) of $n(k)$ is then given by [8]

$$f_n(\rho) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\gamma\left(u^2+v^2\right)^{\frac{\alpha}{2}} - j\Re\{\rho(u-jv)\}}dudv, \tag{2}$$

where $\Re\{\cdot\}$ denotes the real part, the dispersion $\gamma > 0$ is related to the spread of the pdf, and the characteristic exponent $\alpha \in (0, 2]$ is related to the heaviness of the tails of the pdf: A smaller value of α indicates a higher degree of impulsiveness, whereas a value closer to 2 indicates a more Gaussian behavior.

A closed-form expression of (2) is not known to exist except for the special cases of $\alpha = 1$ (complex isotropic Cauchy) and $\alpha = 2$ (complex isotropic Gaussian). In particular, we have

$$f_n(\rho) = \begin{cases} \frac{\gamma}{2\pi}\left(|\rho|^2+\gamma^2\right)^{-\frac{3}{2}}, & \text{when } \alpha = 1 \\ \frac{1}{4\pi\gamma}\exp\left(-\frac{|\rho|^2}{4\gamma}\right), & \text{when } \alpha = 2. \end{cases} \tag{3}$$

Due to such a lack of closed-form expressions, we concentrate on the case of $\alpha = 1$: We shall see in Section 4 that the

estimation schemes obtained for $\alpha = 1$ are not only more robust to the variation of α, but they also provide a better performance for most values of α, than the conventional estimation scheme.

3. Proposed Schemes

3.1 Maximum-likelihood CFO Estimation Scheme

In estimating the CFO, we consider a property of the CP structure of OFDM, i.e., $x(k) = x(k+N)$ for $k = -G, -G+1 \cdots, -1$ as in [5]. Then, from (1), we have

$$r(k+N) - r(k)e^{j2\pi\varepsilon} = n(k+N) - n(k)e^{j2\pi\varepsilon} \quad (4)$$

for $k = -G, -G+1 \cdots, -1$. Observing that $n(k+N) - n(k)e^{j2\pi\varepsilon}$ obeys the complex isotropic Cauchy distribution with dispersion 2γ (since the distribution of $-n(k)e^{j2\pi\varepsilon}$ is the same as that of $n(k)$), we obtain the pdf

$$f_{\boldsymbol{r}}(\boldsymbol{r}|\varepsilon) = \prod_{k=-G}^{-1} \frac{\gamma}{\pi\left(\left|r(k+N) - r(k)e^{j2\pi\varepsilon}\right|^2 + 4\gamma^2\right)^{\frac{3}{2}}} \quad (5)$$

of $\boldsymbol{r} = \{r(k+N) - r(k)e^{j2\pi\varepsilon}\}_{k=-G}^{-1}$ conditioned on ε. The ML estimation is then to choose $\hat{\varepsilon}$ such that

$$\begin{aligned}\hat{\varepsilon} &= \arg\max_{\tilde{\varepsilon}}[\log f_{\boldsymbol{r}}(\boldsymbol{r}|\tilde{\varepsilon})] \\ &= \arg\min_{\tilde{\varepsilon}} \Lambda(\tilde{\varepsilon}), \end{aligned} \quad (6)$$

where $\tilde{\varepsilon}$ denotes the candidate value of ε and the log-likelihood function $\Lambda(\tilde{\varepsilon}) = \sum_{k=-G}^{-1} \log\left\{\left|r(k+N) - r(k)e^{j2\pi\tilde{\varepsilon}}\right|^2 + 4\gamma^2\right\}$ is a periodic function of $\tilde{\varepsilon}$ with period 1: The minima of $\Lambda(\tilde{\varepsilon})$ occur at a distance of 1 from each other, causing an ambiguity in estimation. Assuming that ε is distributed equally over positive and negative sides around zero, the valid estimation range of the ML estimation scheme can be set to $-0.5 < \varepsilon \leq 0.5$, as in [5]. The estimation scheme (6) will be called the Cauchy ML blind estimation (CMBE) scheme.

3.2 Low-complexity CFO Estimation Scheme

The CMBE scheme is based on the exhaustive search over the whole estimation range ($|\varepsilon| < 0.5$), which requires high computational complexity. Thus, we propose a low-complexity CFO estimation scheme with the reduced set of the candidate values.

In order to obtain the reduced set of the candidate values, we exploit the fact that $\varepsilon = \frac{1}{2\pi}\angle\{x^*(k)x(k+N)\} = \frac{1}{2\pi}\angle\{r^*(k)r(k+N)\}$ for $k = -G, -G+1 \cdots, -1$ in the absence of noise. Based on this property, we obtain the set of the candidate values

$$\bar{\varepsilon}(k) = \frac{1}{2\pi}\angle\{r^*(k)r(k+N)\}, \text{ for } k = -G, -G+1 \cdots, -1. \quad (7)$$

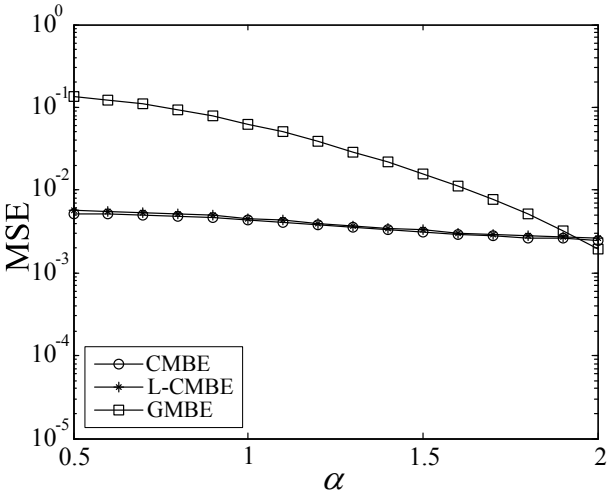


Fig. 1: The MSE performances of the CMBE, L-CMBE, and GMBE schemes as a function of α when the GSNR is 5 dB.

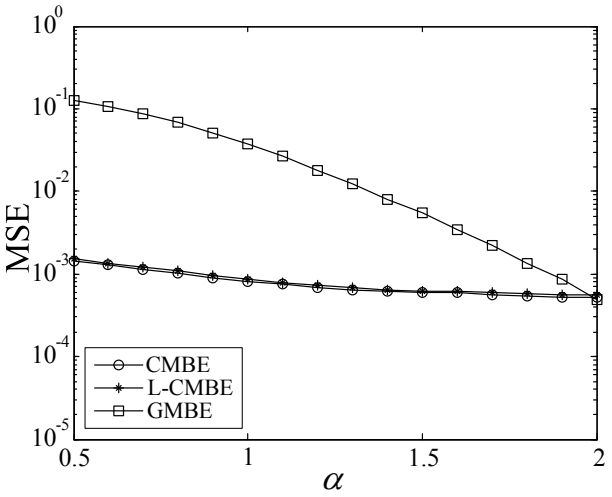


Fig. 2: The MSE performances of the CMBE, L-CMBE, and GMBE schemes as a function of α when the GSNR is 10 dB.

Exploiting the set of the candidate values in (7), the CFO estimate $\hat{\varepsilon}_L$ can be obtained as follows

$$\hat{\varepsilon}_L = \arg\min_{\bar{\varepsilon}(k)} \Lambda\big(\bar{\varepsilon}(k)\big), \text{ for } k = -G, -G+1 \cdots, -1. \quad (8)$$

In the following, (8) is denoted as the low-complexity CMBE (L-CMBE) scheme. Using only $N/2$ candidate values, the L-CMBE scheme can offer an almost same performance as the CMBE scheme with the exhaustive search, which is verified by simulation results in Section 4.

4. Simulation Results

In this section, the proposed CMBE and L-CMBE schemes are compared with the Gaussian ML blind estima-

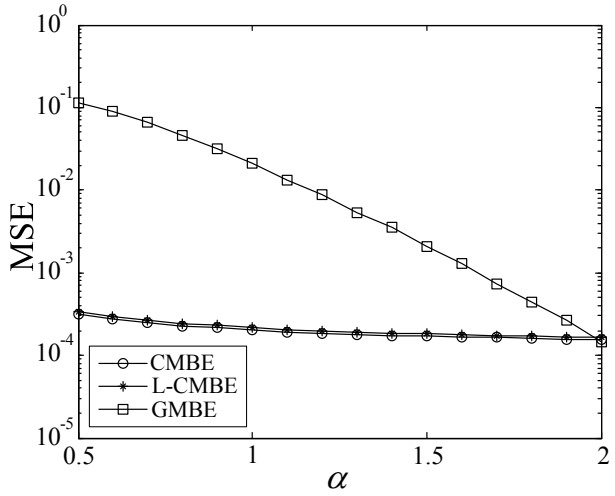


Fig. 3: The MSE performances of the CMBE, L-CMBE, and GMBE schemes as a function of α when the GSNR is 15 dB.

tion (GMBE) scheme in [5] in terms of the mean squared error (MSE). We assume the following parameters: The IFFT size $N = 64$, CFO $\varepsilon = 0.25$, the search spacing of 0.001 for the CMBE scheme, and a multipath Rayleigh fading channel with length $L = 8$ and an exponential power delay profile of $\mathbf{E}[|h(l)|^2] = \exp(-l/L)/\{\sum_{l=0}^{L-1} \exp(-l/L)\}$ for $l = 0, 1, \cdots, 7$, where $h(l)$ is the lth channel coefficient of a multipath channel and $\mathbf{E}[\cdot]$ denotes the statistical expectation. Since CISαS noise with $\alpha < 2$ has an infinite variance, the standard signal-to-noise ratio (SNR) becomes meaningless for such a noise. Thus, we employ the geometric SNR (GSNR) defined as $\mathbf{E}[|x(k)|^2]/(4C^{-1+2/\alpha}\gamma^{2/\alpha})$, where $C = \exp\{\lim_{m\to\infty}(\sum_{i=1}^{m} \frac{1}{i} - \ln m)\} \simeq 1.78$ is the exponential of the Euler constant [10]. The GSNR indicates the relative strength between the information-bearing signal and the CISαS noise with $\alpha < 2$. Clearly, the GSNR becomes the standard SNR when $\alpha = 2$. Since γ can be easily and exactly estimated using only the sample mean and variance of the received samples [11], it may be regarded as a known value: Thus, γ is set to 1 without loss of generality.

Figs. 1-3 show the MSE performances of the CMBE, L-CMBE, and GMBE schemes as a function of α when the GSNR is 5, 10, and 15 dB, respectively. From the figures, we can clearly observe that the proposed schemes not only outperform the conventional scheme for most values of α, except for those close to 2, but also provide a robustness to the variation of the value of α. Another important observation is that the estimation performance of the L-CMBE scheme is almost same as that of the CMBE scheme. From this observation, it is confirmed that the candidate values for the L-CMBE scheme is reasonable.

5. Conclusion

In this paper, we have proposed CFO estimation schemes using CP in non-Gaussian noise environments. We have first obtained the pdf of the CP of OFDM symbols, and subsequently, applyed the pdf to the ML criterion to derive the ML CFO estimation scheme in non-gaussian noise modeled as a complex isotropic Cauchy noise. Then, we also have derived a simpler CFO estimation scheme with a lower complexity. From simulation results, it has been confirmed that the proposed schemes offer a robustness and a substantial performance improvement over the conventional estimation scheme in non-gaussian noise environments.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012-0005066 with funding from the Ministry of Education, Science and Technology (MEST), Korea, by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency under Grant NIPA-2012-H0301-12-1005 with funding from the Ministry of Knowledge Economy (MKE), Korea, and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Boston, MA: Artech House, 2000.

[2] Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification: spectrum and transmit power management extensions in the 5GHz band in Europe, IEEE, 802.11h, 2003.

[3] M. Morelli, C.-C. J. Kuo, and M.-O. Pun, "Synchronization techniques for orthogonal frequency division multiple access (OFDMA): a tutorial review," *Proc. IEEE*, vol. 95, pp. 1394–1427, July 2007.

[4] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: a survey," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 1673-1694, May 2009.

[5] J.-J. Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Sig. Process.,* vol. 45, pp. 1800-1805, July 1997.

[6] T. K. Blankenship and T. S. Rappaport, "Characteristics of impulsive noise in the 450-MHz band in hospitals and clinics," *IEEE Trans. Antennas, Propagat.*, vol. 46, pp. 194-203, Feb. 1998.

[7] P. Torío and M. G. Sánchez, "A study of the correlation between horizontal and vertical polarizations of impulsive noise in UHF," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 2844-2849, Sep. 2007.

[8] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*, New York, NY: Wiley, 1995.

[9] H. G. Kang, I. Song, S. Yoon, and Y. H. Kim, "A class of spectrum-sensing schemes for cognitive radio under impulsive noise circumstances: structure and performance in nonfading and fading environments," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 4322-4339, Nov. 2010.

[10] T. C. Chuah, B. S. Sharif, and O. R. Hinton, "Nonlinear decorrelator for multiuser detection in non-Gaussian impulsive environments," *Electron. Lett.*, vol. 36, pp. 920-922, May 2000.

[11] X. Ma and C. L. Nikias, "Parameter estimation and blind channel identification in impulsive signal environments," *IEEE Trans. Sig. Process.*, vol. 43, pp. 2884-2897, Dec. 1995.

IEC 61850 GOOSE Message over WAN

Carlos Henrique R. de Oliveira, Andre Pereira Bowen
Telecommunications Research and Development Center (CPqD)[1], Companhia Hidro Elétrica do São Francisco (CHESF)
carloshe@cpqd.com.br, abowen@chesf.gov.br

***Abstract*—The publication of IEC 61850-90-1 "Use of IEC 61850 for the communication between substations" and the draft of IEC 61850-90-5 "Use of IEC 61850 to transmit synchrophasor information" opened the possibility to study IEC 61850 GOOSE Message over WAN not only in the layer 2 (link layer) but also in the layer 3 (network layer) in the OSI model. In this paper we examine different possibilities to make feasible teleprotection in the network layer over WAN sharing the communication channel with automation, management and maintenance convergence services among electrical energy substations.**

***Index Terms*— IEC 61850, GOOSE, WAN.**

I. Introduction

Convergence of services like teleprotection (IEC 61850-90-1), network management (IEC 61850-90-2), condition monitoring (IEC 61850-90-3), synchrophasor information (IEC 61850-90-5) and transmission line maintenance support (Quadruple-play Services) among electrical energy substations sharing the same WAN (Wide Area Network) is probably a new tendency and a possibility that needs a deeper investigation. Nowadays there are some references of communication among substations for example NASPInet demonstration of IP Multicast routing of phasor data across the WAN [1] and some available products like Cisco 2520 Connected Grid Switch [2], Korenix JetNet 5828G IEC61850 Layer 3 Managed Ethernet Switch [3] and RuggedCom Network Router Designed for Harsh Environments [4].

In this paper we examine different possibilities to make feasible teleprotection in the layer 3 (network layer) over WAN sharing the communication channel with different electrical energy substation automation services prioritizing the mission-critical GOOSE (Generic Object Oriented Substation Events) message considering the QoS (Quality of Service) specially latency (transfer time) requirement and the convergence of services.

The next Sections are organized as follows. In Section II it is presented a work that was useful to support the proposal of this paper. Section III contains information about advanced multicast routing protocol. In Section IV it is presented the performance requirement to GOOSE message communication. In Section V it is presented the network topology. Finally Section VI presents the conclusions.

[1] This work was supported by CHESF (Companhia Hidro Elétrica do São Francisco – Brazil). At the time of writing this paper, CHESF and CPqD's project of R&D (Research and Development) has been started to explore the issues presented in this paper.

II. Reference paper

The paper "IP and Ethernet Communication Technologies and Topologies for IED networks" [5] was an important research source to support this work. Some parts are present as follows:

The authors of this paper gathered the reason for not sending GOOSE messages over TCP/IP talking to various people in the industry who are also part of the 61850 (TC 57 working group) as follows:

a) No need for IP as there was the feeling that the 61850 GOOSE/SV traffic would be contained inside a substation. According to the authors [5], this is no longer true, for example 61850-90-1, inter substation Teleprotection, PMU traffic streaming, etc are examples of traffic not contained inside a substation;

b) There is the impression that adding IP headers increases the latency of messaging in the network. According to the authors [5]:

- This point too is no longer true with the commercially available off the shelf Ethernet technology. Since the late 1990s Cisco Switches have been forwarding Ethernet and Ethernet+IPv4/IPv6 packets at the same wire speed forwarding rates and with the same switching latencies;
- The Cisco Connected Grid Switch has forwarding latencies between 8 micro seconds to 25 micro seconds irrespective of whether the packets have an Ethernet or IPv4/v6 or TCP or UDP headers on it. The latency range comes from the size of the Ethernet Packet which can range from a frame size of 64 bytes to 1518 bytes.

A. IEC 61850 Traffic outside the Substation [5]

According to the authors currently use cases like distance protection, teleprotection and phasor measure measurement units send IEC 61850 based GOOSE and Sampled Value messages outside the substation. The message exchanges may be between multiple peer substations or between the substation and the control center. The big challenge in either of these scenarios is backhauling Ethernet traffic across the WAN network for the following reasons:

- Ethernet is not and was not built for Wide area communication;
- Ethernet is not a routable protocol;
- If one examines the current Ethernet and IP based networks, the reason these networks scale to the size of the internet and even larger is because of the key attributes of containment and hierarchy;
- In a layer 2 domain (Ethernet bridge domain), traffic reaches the destination by getting forwarded through bridge tables (which are populated based on mac address learning) or flooding the traffic when the destination is unknown. Large bridge domains can potentially de-stabilize the network and recovery times are very difficult to determine.

Some of the techniques used to connect the different Substations networks over a WAN are:

a) Tunneling
- GRE Tunnel. Tunneling the Ethernet frames on top of an IP GRE tunnel;
- Layer 2 Tunneling Protocol.

b) Encapsulations
- Multi Protocol Label Switching (MPLS);
- Pseudo Wires;
- Virtual Private LAN Service (VPLS).

And then, they conclude saying: Hence we need profiles for IEC 61850 GOOSE and Sampled Value messages on top of IPv4 and IPv6.

III. Advanced Multicast Routing Protocol

A. *GOOSE and Routers [6]*

The role of a router in an Ethernet network is, as the name implies, to examine incoming messages and send them to the intended address. Core routers perform transport while edge routers also serve to isolate LAN-only traffic from the WAN, such as broadcast and multicast messages.

GOOSE is a multicast message on layer 2 (data layer) in the OSI model. The router is designed to prevent broadcast and multicast packets from leaving the LAN (in order not to swamp the WAN) and only passes IP packets on layer 3 (network layer).

While this is not a problem within a substation LAN, any external GOOSE traffic needs to be forwarded by a router. The most logical way of doing this is to use VLAN technique. The GOOSE message to be sent to the outer world from the teleprotection device adds a VLAN tag that is recognized by the router. The router converts the GOOSE into a routable IP packet and transports it to its destination address. Care needs to be taken in the network configuration so that the shortest route is used for time critical messages such as teleprotection signaling. The network configuration also needs to provide sufficient bandwidth of the VLAN to ensure short latency.

B. *Advanced Multicast Routing Protocol*

Both [2] and [3] give support to Multicast Routing Protocol via PIM (Protocol Independent Multicast). There are four variants of PIM: PIM Sparse Mode (PIM-SM), PIM Dense Mode (PIM-DM), Bidirectional PIM (Bidir-PIM) and PIM source-specific multicast (PIM-SSM). More details can be found in [7].

Multicast Routing Protocol represents the third possibility (added to tunneling and encapsulation schemes presented in II A) to be investigate as an alternative to GOOSE communication in the layer 3 (network layer) over WAN.

IV. Communication Performance

According to the IEC 61850-90-1, the requirements for the transfer time i.e. the communication performance are basically the same in one bay, between bays and also between substations. Therefore, the same classification scheme shall be used for all links compliant with IEC 61850. For digital communication beyond the substation, transfer times ≤ 10 ms may be accepted according to the message performance class TR2.

Based on this of transfer time limit, these three schemes (tunneling, encapsulation and multicast routing) need to be investigated but not only these but also if there will be significant delay increase by using multicast filtering like IGMP snooping and GMRP (GARP Multicast Registration Protocol) at layer 2 (link layer). Generic Attribute Registration Protocol (GARP) was replaced by Multiple Registration Protocol (MRP), which is a generic registration framework defined by the IEEE 802.1ak amendment to the IEEE 802.1Q standard.

V. Network Topology

The network topology to implement a testbed of the schemes mentioned in IV is presented in Figure 1.

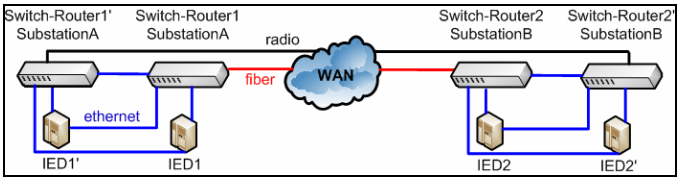


Figure 1. Testbed network topology

Switch-router1, switch-router2, IED1, IED2 and fiber are active primary network. Switch-router1', switch-router2', IED1', IED2' and radio are active redundant network.

The WAN physical link proposed in this work is optical fiber cable and radio to allow QoS performance analysis in both cases extending such analysis to the WAN redundancy case.

Considering typical fiber delay of $5\,\mu s/km$ [8], if a connection to a substation is 100 km away, the optical channel time delay is 0.5 ms that is not a deep impact in the

transfer time.

Nowadays there is available Metro Ethernet IP radio with ultra low latency (< 0.15 ms), full capacity over all 6-38 GHz frequency bands, optical GbE or electrical 10/100/1000BaseT and QoS that enables smart multi-layer packet queuing and prioritization allowing QoS performance analysis to mission-critical GOOSE message in radio channel.

It is worth noticed that redundancy mechanism to high availability automation networks (IEC 62439) as PRP (Parallel Redundancy Protocol) and HSR (High-availability Seamless Ring) needs to be available in the IEDs (Intelligent Electronic Devices) with two ethernet ports. Alternatively, redundancy mechanism in the testbed network topology can be analyzed using switch-routers that support eRSTP™ (Enhanced Rapid Spanning Tree Protocol) [9].

VI. Conclusions

Even if the testbed results of the schemes mentioned in IV show the requirement of transfer time limit was exceed, these tests will show the alternatives that need to be improved and will confirm the necessity of profiles for IEC 61850 GOOSE messages over IPv4 and IPv6 with appropriated protocol in the layer 4 (transport layer) allowing teleprotection, automation, management and maintenance convergence services among electrical energy substations.

References

[1] http://intelligrid.epri.com/Smart_Grid_Information_Sharing_Calls/2011/110510/Smarter%20Tx-Webcast-Rev-1.pdf
[2] http://www.cisco.com/en/US/prod/collateral/switches/ps10968/ps10978/data_sheet_c78_593672.html
[3] http://www.korenix.com/jetnet-ethernet-switch-5828G-overview.htm
[4] http://www.ruggedcom.com/pdfs/news/upgrade_of_ruggedrouter_network_20071126.pdf
[5] http://www.gridwiseac.org/pdfs/forum_papers10/kapadia_gi10.pdf
[6] http://www.rflelect.com/pdf_files/Inside%20the%20Cloud%20-%20Network%20Communications%20Basics%20for%20the%20Relay%20Engineer.pdf
[7] http://www.cisco.com/en/US/docs/ios/solutions_docs/ip_multicast/White_papers/mcst_ovr.pdf
[8] http://www.pes-psrc.org/Reports/H6Paper-App%20Consider%20of%20IEC61850&UCA_072205_083105.pdf
[9] http://www.ruggedcom.com/products/eRSTP/

SESSION

SYSTEM ANALYSIS + QUALITY OF SERVICE + OPTIMIZATION METHODS

Chair(s)

TBA

A Connection Admission Control Algorithm for IEEE 802.16e Networks Based on Bandwidth Reservation and Dynamic Thresholds Adjustment

Sílvio Martins Reis
Faculty of Electrical Engineering
Federal University of Uberlândia (UFU)
Uberlândia – MG - Brazil
smreis@hotmail.com

Paulo Roberto Guardieiro
Faculty of Electrical Engineering
Federal University of Uberlândia (UFU)
Uberlândia – MG - Brazil
prguardieiro@ufu.br

Abstrac – **In this paper, we present a Connection Admission Control (CAC) algorithm for the IEEE 802.16e standard, based on dynamic bandwidth reservation. These reserves are obtained by segmenting the amount of the channel's bandwidth by thresholds, which are dynamically adjusted according to the admissions of handoff and new connections. Simulations were performed using the NS-2, demonstrating that the proposed algorithm can avoid the waste of network resources, increase its efficiency and provide QoS, in terms of bandwidth, for applications.**

Index Terms: **CAC, threshold, dynamic adjustment, Quality of Service, QoS, IEEE 802.16e, WiMAX**

I. INTRODUCTION

The IEEE 802.16e standard [1] for wireless broadband, also known as mobile WiMAX (Worldwide Interoperability for Microwave Access) is an important solution to provide multimedia wireless mobility, including Quality of Services (QoS) warranties for real-time applications and for those that require high transmission rates. In IEEE 802.16e networks, the system resources are controlled by a Base Station (BS), which can effectively guarantee QoS for the Mobile Stations (MS) applying a Connection Admission Control (CAC) scheme, which determines whether or not a connection can be established according to the network resources availability [2], a traffic policing scheme and also a packet scheduling scheme that defines which service class should be served with the higher priority, based on predetermined criteria. The 802.16e standard does not specify how to implement such mechanisms, letting it open for each equipments manufacturer to create its own scheme and thereby encourage the competitiveness of each product.

In 802.16e networks, when a MS moves away from the original serving BS, the quality of the communication degrades, what makes the MS transfer the connection to a neighboring BS with a better signal quality. This process is called handoff. Generally, CAC schemes will prioritize an existing user in handoff process over a new user, in order to provide a better QoE (Quality of Experience) perceived by the user that is already connected. The design of a handoff mechanism should take into account the need of available bandwidth to meet the minimum QoS requirements of handoff connections. As a result, the BS must reserve a certain bandwidth amount exclusively for handoff connections and dedicate the remainder for new connections. However, if a fixed bandwidth amount is reserved for handoff users and this can never be used by new connections, there will probably be a waste of network resources. Therefore, a CAC scheme for 802.16e networks should take into account the need of bandwidth for handoff connections with the challenge of not wasting network resources [3] and accept the maximum possible number of new connections.

In this paper, we propose a CAC algorithm that performs a dynamic bandwidth reservation for the connections, based on the different service classes and taking into account the handoff connections. These reserves are dynamically adjusted in order to minimize the waste of network resources, improve its efficiency, provide justice in the admission of connections and ensure QoS in terms of bandwidth, for applications. To demonstrate these features of the proposed algorithm, simulations were conducted, returning satisfactory results in all evaluated parameters.

The remainder of this paper is organized as follows: Section II presents the IEEE 802.16e Service Classes. Section III identifies the related works, followed by Section IV that describes the proposed mechanism. Section V defines the network scenario and the simulation parameters, and Section VI contains the results analysis. Finally, Section VII presents the conclusions of this work.

II. IEEE 802.16E SERVICE CLASSES

The MAC layer specified in IEEE 802.16e standard provides different levels of QoS for each service flow, which is an unidirectional sequence of packets that is associated to a specific level of QoS, according to the service class that is was assigned. The standard specifies five types of service classes, which are described as follows:

(1) Unsolicited Grant Service (UGS): UGS is designed to support real-time service flows that generate fixed-size data packets on periodic basis, such as VoIP (Voice-over-Internet-Protocol) applications without silence suppression. This service allocates grants with fixed amounts of bandwidth for CBR (Constant Bit Rate) applications, without any requests. (2) Real-time Polling Service (rtPS): rtPS is designed to support real-time service flows with variable packet size, generated at periodic intervals (i.e. VBR - Variable Bit Rate), such as MPEG (Motion Pictures Experts Group) videos. The MSs request bandwidth periodically through a mechanism known as unicast polling.

(3) Extended Real-time Polling Service (ertPS): This service uses a grant mechanism similar to that used by UGS. However, the allocated grants can be used to periodically send

requests to inform the BS about the need for a new grant size. The BS does not change the grant size until it receives from the MS a bandwidth request [4]. ertPS is designed to support real-time service flows with variable data rate, such as VoIP with silence suppression.

(4) Non-real-time Polling Service (nrtPS): This service is designed for non-real-time applications, which require regular grants of variable length, such as FTP application. The service offers unicast polling, but less frequently than rtPS service.

(5) Best Effort (BE): BE is designed for applications which have no QoS requirements. The MS can use the unicast slots or contention slots to request bandwidth.

The QoS parameters and the application types supported by each service class are described in Table I.

TABLE I. IEEE 802.16E CLASSES, APPLICATIONS AND QOS PARAMETERS

Classes	Applications	QoS parameters
UGS	VoIP without silence suppression.	Max. Sustained Traffic Rate; Max. Latency; Jitter;
rtPS	Video Streaming.	Max. Sustained Traffic Rate; Min. Reserved Traffic Rate; Max. Latency;
ertPS	VoIP with silence suppression.	Max. Sustained Traffic Rate; Min. Reserved Traffic Rate; Max. Latency; Jitter;
nrtPS	FTP.	Max. Sustained Traffic Rate; Min. Reserved Traffic Rate;
BE	Web browsing, e-mail.	Max. Sustained Traffic Rate;

III. RELATED WORKS

In [5] the authors propose a CAC scheme that differentiates the connections by its service class. It has 03 modules: traffic classifier, dispatcher and the CAC Decision Maker. The Decision Maker module is based on the maximum rate for UGS and ertPS connections, the average rate for rtPS, the minimum rate for nrtPS and the average rate divided by 2 for BE connections. A connection will be accepted if: *[(Total bandwidth) – (bandwidth allocated to the current connections) – (requested bandwidth)] > 0.* The proposed algorithm also differentiates the connections by its service classes and uses bandwidth reserves for attending the connections according to service class to which they belong.

The CAC scheme proposed in [6] is based on bandwidth reserves with fixed thresholds. These thresholds split the bandwidth into reserves for the connections belonging to the different service classes. In our proposal, the algorithm also uses bandwidth reserves with thresholds, but their values are not fixed, these are dynamically adjusted based on admissions of the connections.

It is proposed in [7] a CAC scheme with a dynamic bandwidth reserve for handoff connections. This reserve varies according to the admission of new handoff connections and the ending of handoffs already admitted. In our work, the variations of the thresholds are caused by admissions of handoffs as well as new connections.

In [8] the authors propose an adaptive CAC algorithm named AACA - Adaptive Admission Control Algorithm, which dynamically determines the reserved bandwidth for handoff connections according to the arrival distributions of both handoff and new connections. When a handoff connection is accepted, the reserve is extended and when a new connection is admitted, the reserve is reduced. In our proposal, besides the reservation for handoff connections, there are also reserves for new connections of real-time, non-real-time and BE traffic. These reserves allow the differentiation in the treatment of each traffic type, in terms of the amount of bandwidth that is destined for each one. Furthermore, these reserves will change only if their occupation reaches a predetermined threshold value. As a result, more connections can be admitted in the network.

IV. PROPOSED CONNECTION ADMISSION CONTROL (CAC) ALGORITHM

In this paper, we propose an algorithm for connection admission control (CAC) based on bandwidth reservations, with thresholds that are dynamically adjusted in order to avoid the waste of reserved bandwidth and maintain the QoS of already admitted connections. The bandwidth reserves are destinated for handoff connections and the new connections of real-time, non-real-time and Best Effort traffic. Fig. 1 depicts the proposed bandwidth reservation scheme.

Denote the total bandwidth that the BS can allocate to the connections as B, the boundary between the reserves of handoff and real-time connections (UGS and rtPS) as "*thHandoff*" (handoff threshold), the boundary between the reserves of the real-time and non-real-time connections (nrtPS) as "*th*" (threshold) and the boundary between the reserves of non-real-time and Best Effort (BE) connections as "*thBE*". Denote as b_{ho}, b_{ugs}, b_{rtps} and b_{nrtps} the portion of the bandwidth B already allocated to the existing handoff, UGS, rtPS and nrtPS connections, respectively and b_{req} as the amount of bandwidth that a new connection requires before being admitted.

The threshold "*thHandoff*" varies within the range [*thMax, thHandoffMax*] and its initial value is [(*thHandoffMax* - *thMax*) * 0.8]. The threshold "*th*" varies within the range [*thMin, thMax*] and its initial value is [(*thMax* – *thMin*)/2].

B is splited into segments to provide bandwidth reserves to the different types of traffic. The bandwidth reserved for handoff connections corresponds to (*B* - *thHandoff*), the reserve for the real-time connections is (*thHandoff* - *th*), for the non-real-time is (*th* - *thBE*) and finally for the BE connections is "*thBE*". The admission of a connection by the BS obeys the following priority order: *handoff connection > UGS connection > rtPS connection > nrtPS connection*. All BE connections are allowed and a little portion of the bandwidth (*thBE*) is reserved for them, in order to avoid the "bandwidth starvation" of the BE traffic in the scheduling process.

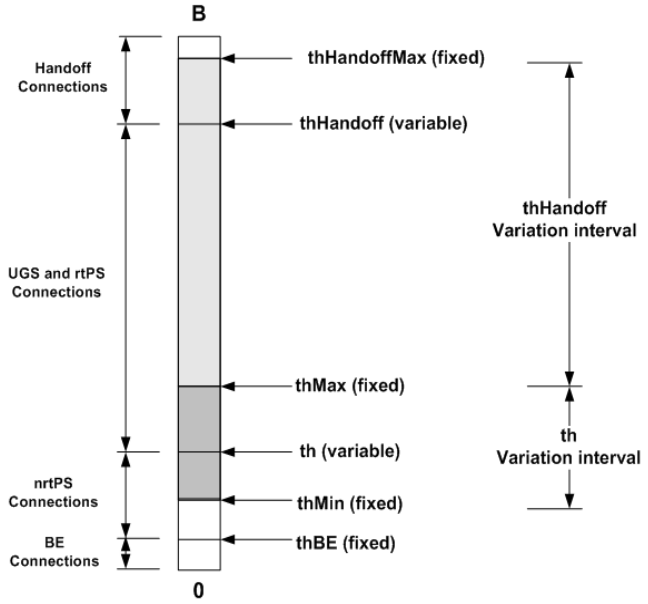


Figure 1. Proposed bandwidth reservation scheme.

A handoff connection will be admitted if:

$$(b_{req} + b_{ho}) \leq (B - thHandoff) \quad (1)$$

If the connection is admitted, b_{req} will be added to b_{ho}, that is:

$$b_{ho} = b_{ho} + b_{req}; \quad (2)$$

After the update of b_{ho}, if the following condition is satisfied:

$$((b_{req} + b_{ho}) \geq (B\text{-}thHandoff)/2 \text{ and } (b_{ugs} + b_{rtps}) < (thHandoff - th - b_{req})) \quad (3)$$

thHandoff will be reduced by b_{req} (limited to the value of thMax), that is:

$$thHandoff = thHandoff - b_{req}; \quad (4)$$

The purpose of this condition is to increase the size of the reserve for the handoff connections, if the bandwidth dedicated to the already admitted handoff connections approaches the defined limit for the change (first term in condition (3)), respecting the reserves for the UGS and rtPS connections (the second term of condition (3)). Unlike the exposed in [8], in which the threshold changes every time a connection is accepted, in this proposal the threshold *thHandoff* only changes if the occupation of the reserve for handoff connections reaches the value [(*B-thHandoff) / 2*]. This avoids the decrease of the reserve for real-time connections before the occupation of half of the reserve for handoff connections, giving more admission opportunities to real-time connections.

An UGS or rtPS connection will be admitted if:

$$(b_{req} + b_{ugs} + b_{rtps}) \leq (thHandoff - th)) \quad (5)$$

If the connection is admitted, b_{req} will be added to b_{ugs} or b_{rtps}, that is:

$$b_{ugs} = b_{ugs} + b_{req}; \quad (If\ UGS) \quad (6)$$

$$b_{rtps} = b_{rtps} + b_{req}; \quad (If\ rtPS)$$

After the update of b_{ugs} or b_{rtps}, if the following condition is satisfied:

$$((b_{req} + b_{ugs} + b_{rtps}) \geq (thHandoff\text{-}th) \text{ and } (b_{ho} \leq B - thHandoff - b_{req})) \quad (7)$$

thHandoff will be increased by b_{req} (limited to the value of *thHandoffMax),* that is:

$$thHandoff = thHandoff + b_{req}; \quad (8)$$

and, if the following condition is satisfied:

$$((b_{req} + b_{ugs} + b_{rtps}) \geq (thHandoff\text{-}th) \text{ and } (b_{nrtps} < (th - thBE - b_{req}))) \quad (9)$$

th will be decreased by b_{req}, (limited to the value of *thMin)* that is:

$$th = th - b_{req}; \quad (10)$$

The purpose of these conditions is to increase the size of the reserve for real-time connections.

A nrtPS connection will be admitted if:

$$((b_{req} + b_{nrtps}) \leq (th\text{-}thBE)) \quad (11)$$

If the connection is admitted, b_{req} will be added to b_{nrtps}, that is:

$$b_{nrtps} = b_{nrtps} + b_{req}; \quad (12)$$

After the update of b_{nrtps}, if the following condition is satisfied:

$$((b_{req} + b_{nrtps}) \geq (th - thBE) \text{ and } (b_{ugs} + b_{rtps}) < (thHandoff - (th + 4 * b_{req}))) \quad (13)$$

th will be increased by b_{req} (limited to the value of *thMax*), that is:

$$th = th + b_{req}; \quad (14)$$

The purpose of this condition is to increase the size of the reserve for nrtPS connections. Finally, all BE connections will be admitted and a little portion of the bandwidth (*thBE*) will be reserved for them, in order to avoid the "bandwidth starvation" of the BE traffic in the scheduling process. The values of the fixed thresholds (*thHandoffMax, thMax, thMin, thBE*) may be assigned by the network administrator in accordance with the traffic profile of the users.

Fig. 2 depicts the proposed CAC algorithm.

```
1: begin
2: B ← total bandwidth that the BS can allocate to the connections;
3: thHandoffMax ← 0.9*B;  //Defined by the network administrator;
4: thHandoff ← (thHandoffMax  - thMax) * 0.8
5: thMax ← 0.4*B; // Defined by the network administrator;
6: thMin ← 0.1*B; // Defined by the network administrator;
7: th ← (thMax - thMin)/2;
8: thBE ← 0.02*B; // Defined by the network administrator;
9: for all the pending connections do:
10: if (connection= handoff) then
11:    if ((b_req + b_ho) ≤ (B - thHandoff)) then
12:       b_ho ← b_ho + b_req;
13:       accept Handoff;
14:    if ((b_req + b_ho) ≥ (B-thHandoff)/2 and (b_ugs + b_rtps) < (thHandoff -th -
          b_req)) then
15:       thHandoff ← getmax(thMax, (thHandoff - b_req))
16:    endif
17:   else
18:    reject Handoff;
19:   endif
20: endif
21: if (connection= UGS)  or  (connection= rtPS) then
22:   if ((b_req + b_ugs + b_rtps) ≤ (thHandoff - th)) then
23:      if (connection= UGS)  then
24:         b_ugs ← b_ugs + b_req;
25:         accept UGS;
26:      endif
27:      if (connection= rtPS)  then
28:         b_rtps ← b_rtps + b_req;
29:         accept rtPS;
30:      endif
31:      if ((b_req+b_ugs+ b_rtps) ≥ (thHandoff - th) and (b_ho≤ B - thHandoff -
          b_req)) then
32:         thHandoff ←  getmin(thHandoffMax, (thHandoff + b_req));
33:      endif
34:      if ((b_req+b_ugs+b_rtps) ≥(thHandoff-th) and (b_nrtps < (th - thBE - breq)))
         then
35:         th ← getmax(thMin, (th - b_req));
36:      endif
37:   else
38:    reject UGS or rtPS;
39:   endif
40: endif
41: if (connection= nrtPS) then
42:    if ((b_req + b_nrtps) ≤ (th-thBE)) then
43:       b_nrtps ← b_nrtps + b_req;
44:       accept nrtPS;
45:    if ((b_req + b_nrtps) ≥ (th - thBE) and (b_ugs + b_rtps) < (thHandoff - (th
          +4*b_req))) then
46:       th ← getmin(thMax, (th + b_req));
47:       endif
48:    else
49:    reject nrtPS;
50:    endif
51: endif
52: if (connection= BE) then
53:   accept BE;
54:  endif
55:  end for
56:  end
```

Figure 2. Proposed CAC Algorithm.

V. MODELING AND SIMULATION

To evaluate the proposed CAC algorithm, we used the simulation tool NS-2 [9] with the WIMAX module developed by NIST [10]. It was necessary to extend this module to include the proposed CAC model and the AACA described in [8]. The considered scenarios involve multiple MSs entering in the network at regular intervals and random positions. Each MS was assigned to a type of traffic and the handoff connections were also considered. The main parameters of the simulation are shown in Table II:

TABLE II. MAIN SIMULATION PARAMETERS

Parameter	Value
Uplink Transmission Rate	10 Mbps;
UGS traffic	CBR Traffic Rate = 96 Kbps;
rtPS traffic	Video Streaming MPEG; Average rate = 480 Kbps;
nrtPS traffic	FTP (Min. rate = 160 Kbps; Max. rate = 800 Kbps);
BE traffic	HTTP traffic (Average rate = 64 Kbps);
Handoff connections	CBR Traffic Rate = 96 Kbps;

VI. RESULTS ANALYSIS

The results presented in this section refer to simulations performed with the use of the AACA algorithm described in [8] and the proposed algorithm, in order to compare the performance of each one. At the experiments in which the AACA algorithm was used, the considered thresholds were: *thMax = 0.9 * B; thMin = 0.1 * B; thad = [(thMin + thMax)/2]*. When the proposed algorithm was used, the considered thresholds were: *thHandoffMax = 0.9*B; thHandoff = [(thHandoffMax- thMax)*0.8]; thMax = 0.4*B; thMin = 0.1*B; th = [(thMax - thMin)/2]; thBE = 0.02*B.*

We considered the service classes UGS, rtPS, nrtPS and BE. For the handoff connections, the CBR traffic and UGS service class was considered. The minimum rate was adopted as the connection admission criteria. We adopted the ratio of one handoff connection attempt to eight new connections attempts. All were uniformly distributed over the simulation time. The performance of each CAC algorithm will be compared in terms of the number of admitted connections for each service class and the connections blocking rates.

The total simulation time was 50 seconds and the presented results are the average outcome of 10 simulations runs. The connection attempts started after 15 seconds of simulation. The results indicated that on average, the AACA algorithm admitted the total of 53 connections with 10 handoffs, 12 UGS, 9 rtPS, 12 nrtPS and 10 BE, whereas the proposed algorithm admitted the total of 68 connections with 10 handoffs, 14 UGS, 11 rtPS, 13 nrtPS and 20 BE, resulting in the admission of 15 more connections than the AACA algorithm, that is, an increase of network efficiency by 16.7%.

Fig. 3 depicts the blocking rates over time for handoff and the real-time connections. The blocking rate of non-real-time and BE connections are illustrated in Fig. 4. It is shown in these figures that the blocking rates remain equal to zero when the network load is low (the simulation time is less than 30 seconds), for both used algorithms. As the connections are

admitted, the network load increases and the blocking rates rise progressively due to the lack of available bandwidth to be reserved.

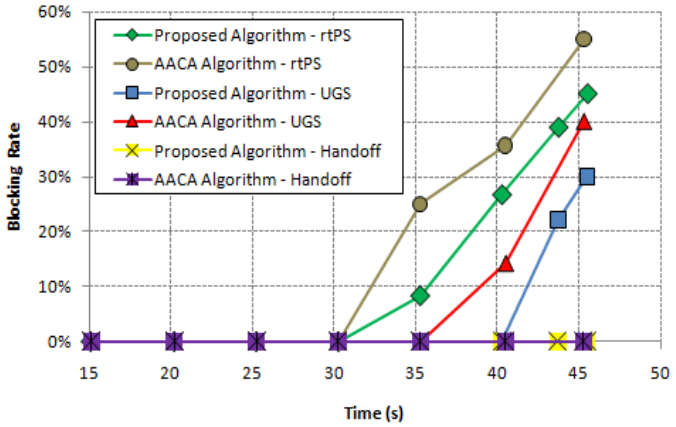


Figure 3. Handoff and real-time connections blocking rates.

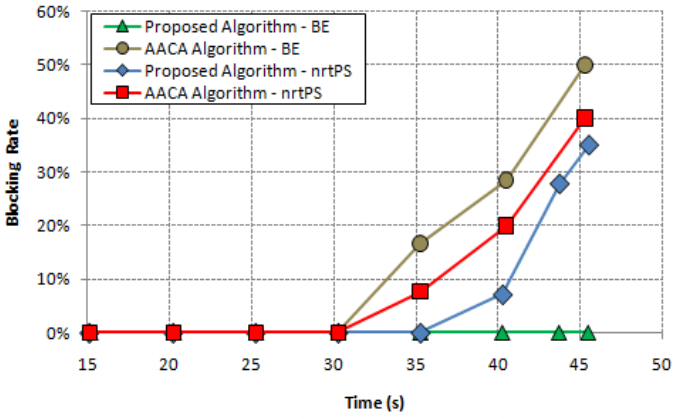


Figure 4. Non-real-time and best effort connections blocking rates.

Fig. 3 shows that the blocking rate of handoff connections remained equal to zero through all the simulation time for both used algorithms, demonstrating that both were effective in admission of handoff connections. It also shows that when the AACA algorithm is used, UGS connections started being blocked after about 35 seconds of simulation and the blocking rate increases, reaching a total of 40% when the simulation time is around 45 seconds. When the proposed algorithm is used, UGS connections started being blocked after about 40 seconds of simulation, reaching a maximum blocking rate of 30% when the simulation time is around 45 seconds. A similar behavior is observed for the rtPS connections in Fig. 3 and nrtPS connections in Fig. 4, where the final blocking rates are lower when the proposed algorithm is used. Fig. 4 shows that when the AACA algorithm is used, the blocking rate of BE connections reaches 50% when the simulation time is approximately 45 seconds. On the other hand, when the proposed algorithm is used, the BE blocking rate remains equal to zero throughout all the simulation, since this algorithm admits all the BE connections even if all the bandwidth has been already reserved for the other connections.

Table III shows the blocking rates at the end of the simulation. The new connections blocking rates obtained by the proposed algorithm were lower than those obtained by the AACA algorithm, which results in a smaller number of blocked connections and greater network efficiency.

TABLE III. FINAL BLOCKING RATES

	Blocking Rate - AACA	Blocking Rate - Proposed
Handoff	0%	0%
UGS	40%	30%
rtPS	55%	45%
nrtPS	40%	35%
BE	50%	0%

VII. CONCLUSIONS

In this paper, we presented a Connection Admission Control (CAC) algorithm for the IEEE 802.16e standard that performs bandwidth reservation for the connections, taking into account the different service classes and handoff connections. These reserves are dynamically adjusted in order to minimize the waste of network resources, improve its efficiency, provide justice in the admission of connections and ensure QoS in terms of bandwidth, for applications. The performance of this algorithm was evaluated in terms of the number of admitted connections and the blocking rates of new and handoff connections. The simulation results showed that the proposed CAC algorithm outperforms the CAC algorithm based on adaptive bandwidth reservation, with the admission of a larger number of new connections to the network and a decrease of the new and handoff connections blocking rates.

REFERENCES

[1] IEEE 802.16e 2005, "IEEE Standard for local and metropolitan area networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems", Fev, 2006.

[2] K. Wongthavarawat, e A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems", International Journal of Communication Systems, pp. 81- 96, Fev. 2003.

[3] R. Laishram, I. S. Misra, "A bandwidth efficient adaptive call admission control scheme for QoS provisioning in IEEE 802.16e mobile networks", Int J Wireless Inf Networks, 18:108-116, 2011.

[4] J. Freitag, "Mechanisms for quality of service provision in IEEE 802.16 networks", B644m Campinas, [S.P. : s.n.], Institute of Computing - UNICAMP, 2010.

[5] S. Yang, C. Cheng e R. Wu, "Enhanced CAC with QoS scheme for improving the efficiency of resource allocation on the IEEE 802.16 network", Workshops of International Conference on Advanced Information Networking and Applications, 2011

[6] C. L. Soares, P. R. Guardieiro, "Threshold-Based connection admission control for IEEE 802.16 standard", International Workshop on Wireless Multimedia Networking and Applications (WMNA'09), Wrexham, United Kingdom, Sep. 2009, in press.

[7] S. B. Chaudhry and R. K. Guha, "Adaptive connection admission control and packet scheduling for QoS provisioning in mobile WiMAX", Proc. of IEEE International Conference on Signal Processing and Communication (ICSPC), pp. 1355 - 1358, Nov. 2007.

[8] C. Wang, H. Lin e H. Lo, "Adaptive admission control algorithm in IEEE 802.16e broadband wireless access networks", IEEE 6th World Congress on Services, 2010.

[9] Network Simulator 2, http://www.isi.edu/nsnam/ns, 2010.

[10] IEEE 802.16 module for NS-2 - National Institute of Standards and Technology, http://www.antd.nist.gov/seamlessandsecure.shtml, 2011.

A Code Acquisition Scheme Based on Multiple Threshold for Optical CDMA Systems

Changha Yu, Youngpo Lee, and Seokho Yoon†
College of Information and Communication Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do, Korea
†Corresponding author

Abstract— *In this paper, we propose an enhanced code acquisition scheme for optical code division multiple access (CDMA) systems. By using multiple thresholds, the proposed scheme provides a shorter mean acquisition time (MAT) than that of the conventional multiple-shift (MS) scheme. The simulation results demonstrate that the MAT of the proposed scheme is shorter than that of the conventional MS scheme in both single-user and multi-user environments.*

Keywords: Optical CDMA; acquisition; MAT; multiple-shift

1. Introduction

In code division multiple access (CDMA) based systems, the data demodulation is possible only after a code synchronization is completed. Therefore, the code synchronization is one of the most important tasks in CDMA based systems [1]. Generally, the code synchronization consists of two stages: code acquisition and tracking. Achieving the code synchronization is called code acquisition and maintaining the code synchronization is called tracking [2], of which the former is dealt with in this paper. In code acquisition process, the most significant performance measure is mean acquisition time (MAT), which is a mean time that elapses prior to acquisition.

An optical CDMA system uses a spreading code called optical orthogonal code (OOC) proposed by Salehi [3]. Due to its good auto-correlation and cross-correlation properties, the OOC has been widely used for various CDMA based systems including optical CDMA systems [4], [5]. Keshavarzian and Salehi introduced the serial-search (SS) [4] scheme using the OOC, which is simple; however, its MAT increases as the code length becomes longer. Thus, the SS scheme is not suitable for rapid acquisition of a long code that is essential for multi-user environments. In order to overcome this drawback, in [5], the same authors proposed the multiple-shift (MS) scheme using the OOC, which consists of two stages and offers a shorter MAT compared with that of the SS scheme.

In this paper, we propose a novel code acquisition scheme referred to as the enhanced multiple-shift (EMS). The EMS scheme also consists of two stages like the MS scheme, however, by using multiple thresholds and modified local code, the EMS scheme provides a shorter MAT compared with that of the MS scheme.

The remainder of this paper is organized as follows. Section 2 describes the system model. In Section 3, we present the conventional MS and proposed EMS schemes. Section 4 analyzes the MAT performance of the EMS scheme. In Section 5, the simulation results show the MATs of MS and EMS schemes in single-user and multi-user environments. Section 6 concludes this paper.

2. System Model

In an optical CDMA channel, there exist various kinds of impairments such as noise, multipath signals, and multiple access interference (MAI). The influences of noise and multipath signals can be almost completely mitigated by using fiber-optic medium; however, that of MAI should be alleviated in the receiver [6], [7]. In this paper, thus, we consider a multi-user environment without the influences of noise and multipath signals. Then, the received signal $r(t)$ can be written as

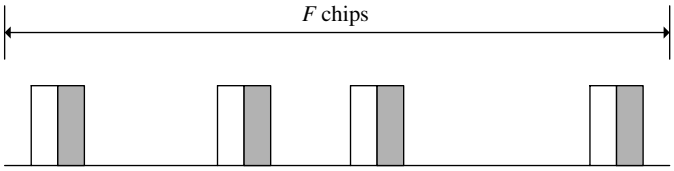
$$r(t) = \sum_{n=1}^{N} s^{(n)}(t - \tau^{(n)}), \tag{1}$$

where $s^{(n)}(t)$ is the transmitted signal of the nth user; $\tau^{(n)} \in [0, T_b)$ denotes the time delay of the nth user with bit duration T_b; and N is the number of total users. We consider the on-off-keying (OOK) modulation and assume that the bit rate is the same for all users. Thus, transmitted signal $s^{(n)}(t)$ can be expressed as

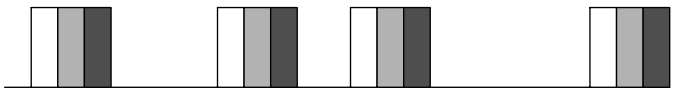
$$s^{(n)}(t) = \sum_{i=-\infty}^{\infty} b_i^{(n)} c^{(n)}(t - iT_b), \tag{2}$$

where $b_i^{(n)}$ is the ith binary data bit of the nth user and $c^{(n)}(t) = \sum_{j=0}^{F-1} a_j^{(n)} p(t - jT_c)$ is the OOC of the nth user with chip duration T_c and sequence $a_j^{(n)} \in \{0, 1\}$ of length F and weight K (the total number of '1's in $a_j^{(n)}$) with the rectangular pulse $p(t)$ of length T_c defined as

$$p(t) = \begin{cases} 1, & 0 \le t < T_c, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$



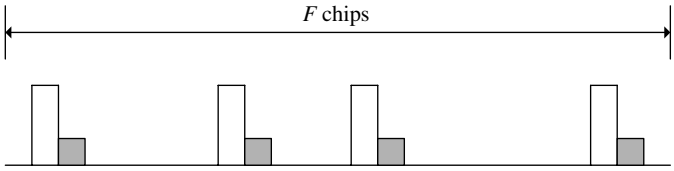
(a) The first stage local code of the MS scheme (M = 2)



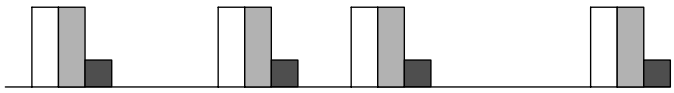
(b) The first stage local code of the MS scheme (M = 3)

(c) The second stage local code of the MS scheme (original OOC)

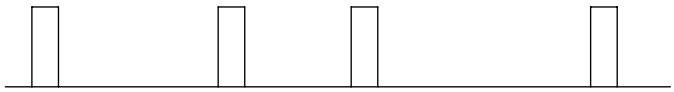
Fig. 1: The local codes of the MS scheme when OOC of (32,4,1,1) is used.



(a) The first stage local code of the EMS scheme (M = 2)



(b) The first stage local code of the EMS scheme (M = 3)



(c) The second stage local code of the EMS scheme (original OOC)

Fig. 2: The local codes of the EMS scheme when OOC of (32,4,1,1) is used.

Generally, the OOC can be denoted by its parameters, i.e., $(F, K, \lambda_a, \lambda_c)$, where λ_a and λ_c are auto-correlation and cross-correlation constraints, respectively [3]. For ideal strict orthogonality of OOC, both λ_a and λ_c have to be zero; however, since an OOC consists of 0 and 1, the ideal strict orthogonality cannot be satisfied. Thus, in this paper, both λ_a and λ_c are set to 1.

3. Code Acquisition Schemes

3.1 Multiple-Shift (MS) Scheme

In the MS scheme, total F cells in the search space are divided into Q groups, each of which contains M cells. The relation of Q and M is given by

$$Q = \left\lceil \frac{F}{M} \right\rceil, \tag{4}$$

where $\lceil \cdot \rceil$ denotes the ceiling operation. Note that the upper closest integer to the ratio of F/M is chosen as the value of Q when M is not a divisor of F.

The MS scheme consists of two stages. In the first stage, the received signal $r(t)$ is correlated with the first stage local code shown in Fig. 1. The correlation is repeated on a group-by-group basis. If the correlation value corresponding to a certain group exceeds a given threshold $TH_{MS,first}$, the group is declared to be the correct group having the time delay $\tau^{(n)}$ and the process is transferred to the second stage. In the second stage, the correlation-based search is performed again with the second stage local code (original OOC) on a cell-by-cell basis over M cells in the correct group. As in the first stage, when the correlation value corresponding to a certain cell exceeds a given threshold $TH_{MS,second}$, which is different from the value in the first stage, the cell is declared to be an estimate of the time delay $\tau^{(n)}$.

Using the definition in [5], the MAT of MS scheme T_{MS} is given by

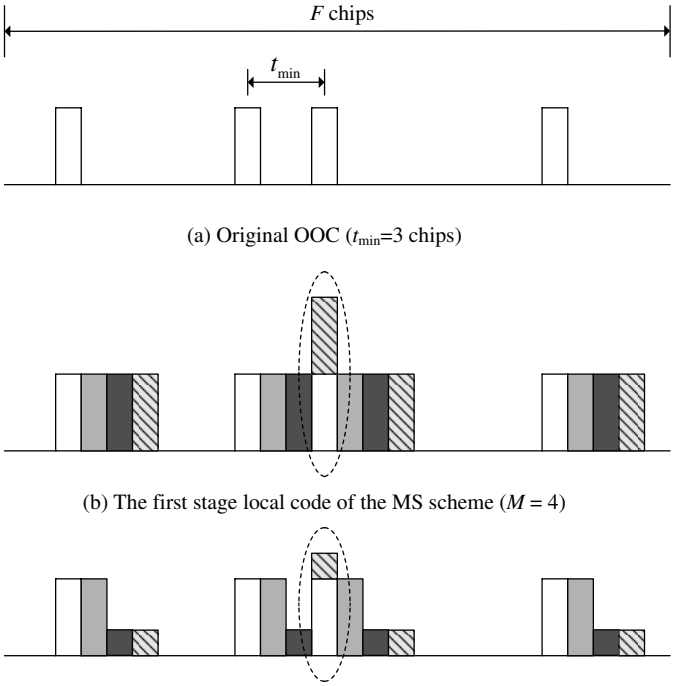
$$T_{MS} = \frac{Q+1}{2} + \frac{M+1}{2}. \tag{5}$$

From (4) and (5), we can find that the minimum value of T_{MS} equals to $\sqrt{F}$ when $M = \sqrt{F}$.

3.2 Enhanced Multiple-Shift (EMS) Scheme

The EMS scheme proposed in this paper also consists of two stages as the MS scheme. However, using multiple thresholds and modified local code, the EMS scheme has a shorter MAT than that of the MS scheme.

In the first stage, the first stage local code shown in Fig. 2 is used instead of the conventional local code for EMS scheme. The first stage local code consists of large and small chips. When M is an even number, the number of large chips is equal to the number of small chips; otherwise, the number of large chips is larger by 1 than the number of small chips. The power of the small chips is determined with the



(a) Original OOC (t_{min}=3 chips)

(b) The first stage local code of the MS scheme (M = 4)

(c) The first stage local code of the EMS scheme (M = 4)

Fig. 3: The local codes of the MS and EMS scheme when $M > t_{min}$.

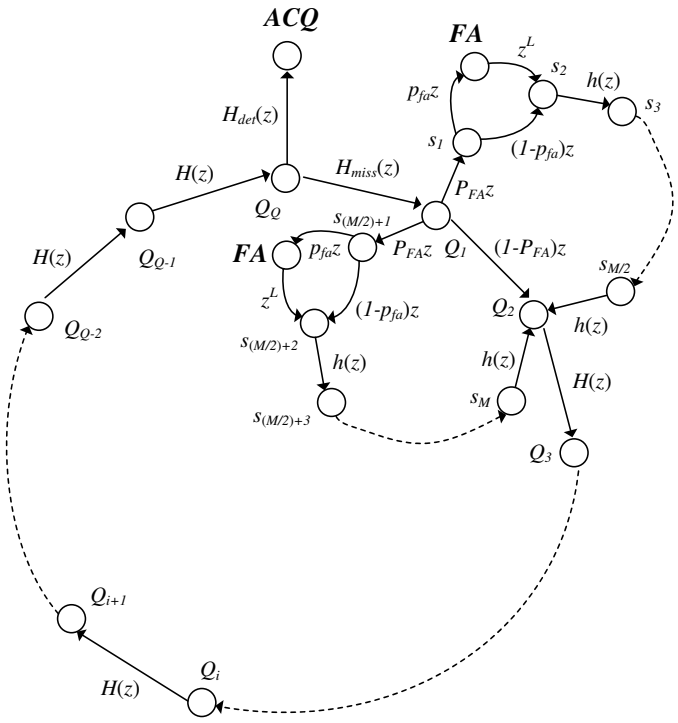


Fig. 4: Markov state model for EMS scheme.

following conditions

$$\text{Conditions of } \alpha = \begin{cases} \alpha < 1, \\ \alpha > \frac{\lambda_a}{K}, \\ \alpha > \frac{\lambda_c}{K}, \end{cases} \tag{6}$$

where α represents the power of small chips. On the other hand, the power of large chips is always set to 1.

The correlation is repeated on a group-by-group basis as in the first stage of the MS scheme. If the correlation value corresponding to a certain group exceeds a given thresholds $TH_{EMS,first}$ or $th_{EMS,first}$, the group is declared to be the correct group having the time delay $\tau^{(n)}$ and the process is transferred to the second stage. In the second stage, the correlation-based search is performed again with the second stage local code (original OOC) on a cell-by-cell basis over M cells in the correct group. As in the first stage, when the correlation value corresponding to a certain cell exceeds a given threshold $TH_{EMS,second}$, the cell is declared to be an estimate of the time delay $\tau^{(n)}$.

In the first stage, if the correlation value of the correct group is equal to or larger than $TH_{EMS,first}$, we only search the first half of M cells of the correct group in the second stage; otherwise and if the correlation value of the correct group is equal to or larger than $th_{EMS,first}$, the search is performed over the second half of M cells of the correct group in the second stage. In the EMS scheme, two thresholds, $TH_{EMS,first}$ and $th_{EMS,first}$ can be determined based on the power of large and small chips in the first stage local code.

Using the definition in [5], the MAT of EMS scheme T_{EMS} is given by

$$T_{EMS} = \frac{Q+1}{2} + \frac{M+2}{4}. \tag{7}$$

As we can see in (7), the MAT in the second stage is reduced by half and the minimum value of T_{EMS} equals to $\sqrt{F/2}+1$ when $M = \sqrt{2F}$.

For a correct operation of the proposed EMS and conventional MS schemes, M has to be equal to or smaller than t_{min}, where t_{min} is the minimum chip interval of the OOC. Otherwise, chips of local code are overlapped as shown in Fig. 3, where $t_{min} = 3$ and $M = 4$. In this case, the EMS scheme cannot guarantee a correct operation and good performance.

4. Performance Analysis

In this section, we derive the MAT expression of the EMS scheme by modeling the acquisition process as a discrete-time Markov process [1] with the circular flow graph diagram shown in Figs. 4 and 5. In Fig. 4, 'ACQ' and 'FA' represent the acquisition and false alarm states, respectively, and $P_D(p_d)$ and $P_{FA}(p_{fa})$ denote the detection and false alarm probabilities in the first and second stages,

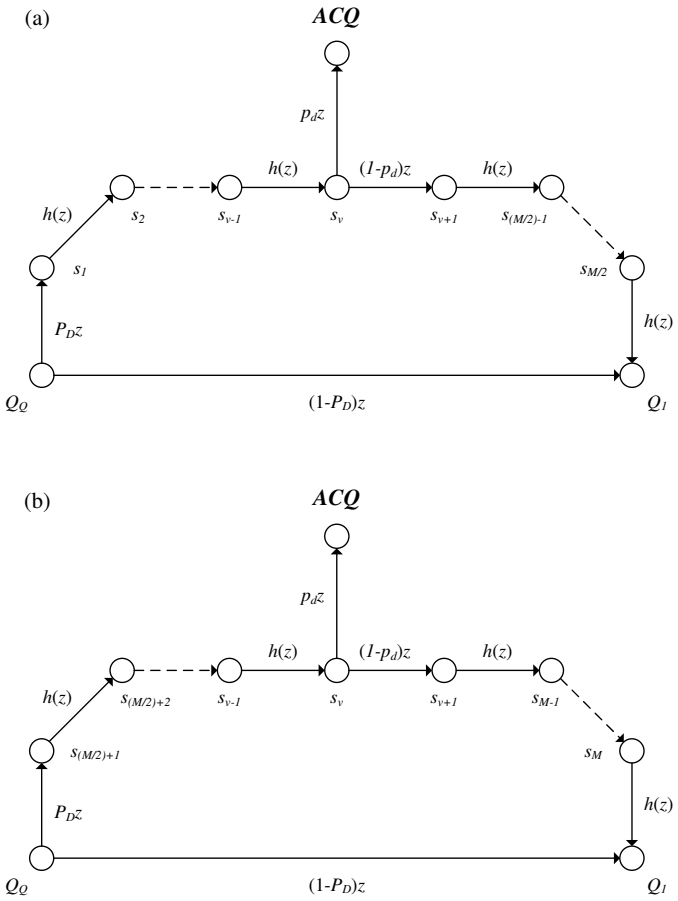


Fig. 5: (a) Large group from Q_Q to Q_1 (b) Small group from Q_Q to Q_1.

respectively. Fig. 5 represents the transition process diagrams of $H_{det}(z)$ in Fig. 4.

In Fig. 4, states $Q_1, Q_2, \cdots, Q_{Q-1}$ correspond to the incorrect groups and a state Q_Q corresponds to the correct group, where the acquisition can be achieved. States $s_1, s_2, \cdots, s_{M/2}$ and states $s_{(M/2)+1}, s_{(M/2)+2}, \cdots, s_M$ correspond to cells in the second stage, when the correlation value of the correct group is large (when the correlation value $\geq TH_{EMS,first}$) and small (when $th_{EMS,first} \leq$ the correlation value $< TH_{EMS,first}$), respectively. The gains $H_{det}(z)$ and $H_{miss}(z)$ represent the transition gains from Q_Q to ACQ (detection) and from Q_Q to Q_1 (miss), respectively. The gains $H(z)$ and $h(z)$ represent the transition gains between the successful incorrect groups and between the successful incorrect cells, respectively. L is the penalty time factor due to the false alarm. The gains $h(z), H(z), H_{det}(z)$, and $H_{miss}(z)$ can be expressed as follows:

$$h(z) = (1 - p_{fa})z + p_{fa}z^{L+1}, \quad (8)$$

$$H(z) = (1 - P_{FA})z + P_{FA}zh^{M/2}(z), \quad (9)$$

$$H_{det}^{(v)}(z) = p_d P_D z^2 h^{v-1}(z), \quad (10)$$

and

$$H_{miss}(z) = (1 - P_D)z + P_D(1 - p_d)z^2 h^{(M/2)-1}(z), \quad (11)$$

where $v \in \{1, 2, \cdots, M/2\}$ is distributed uniformly over $[1, M/2]$ and represents the correct state.

Let us assume that the search is stated at Q_i, then the transfer function between Q_i and ACQ nodes can be written as

$$U_i^{(v)}(z) = \frac{H^{Q-i}(z)H_{det}^{(v)}(z)}{1 - H_{miss}(z)H^{Q-1}(z)}. \quad (12)$$

In (12), $i \in \{1, 2, \cdots, Q\}$ is assumed to be distributed uniformly over $[1, Q]$. Thus, after averaging $U_i^{(v)}(z)$ over the probability density function (pdf) of i and v, we can re-write (12) as

$$U(z) = \frac{H_{det}(z)}{1 - H_{miss}(z)H^{Q-1}(z)} \frac{1}{Q} \sum_{i=1}^{Q} H^{Q-i}(z), \quad (13)$$

where $H_{det}(z) = \frac{2}{M} \sum_{v=1}^{M/2} H_{det}^{(v)}(z)$. Using the moment generating function, we can obtain the following relationship between $U(Z)$ and T_{EMS}.

$$U(z) = E(z^{T_{EMS}}), \quad (14)$$

where $E(\cdot)$ denotes the statistical expectation operation. In (14), $E(T_{EMS})$ can be computed by its moment generating function as

$$E(T_{EMS}) = \frac{dU(z)}{dz}\Big|_{z=1} = U'(1), \quad (15)$$

and thus, can be obtained as

$$E(T_{EMS}) = \frac{H'_{det}(1) + H'_{miss}(1)}{p_d P_D} + (Q-1)H'(1)\frac{2 - p_d P_D}{2 p_d P_D}, \quad (16)$$

where $H'(1)$, $H'_{det}(1)$, and $H'_{miss}(1)$ can be expressed as

$$H'(1) = 1 + \frac{M}{2} P_{FA}(1 + L p_{fa}), \quad (17)$$

$$H'_{det}(1) = 2p_d P_D + \frac{(M/2) - 1}{2} p_d P_D (1 + L p_{fa}), \quad (18)$$

and

$$H'_{miss}(1) = (1 - P_D) + P_D(1 - p_d)[(M/2) + 1 + \{(M/2) - 1\} L p_{fa}], \quad (19)$$

respectively.

When $p_d = P_D = 1$ and $p_{fa} = P_{FA} = 0$ (i.e., single-user case), (16) is re-written as

$$E(T_{EMS}) = \frac{Q + 1}{2} + \frac{M + 2}{4}. \quad (20)$$

After differentiating (20), we finally obtain the optimum M and minimum T_{EMS} as $\sqrt{2F}$ and $\sqrt{\frac{F}{2}} + 1$, respectively, when $F \gg 1$.

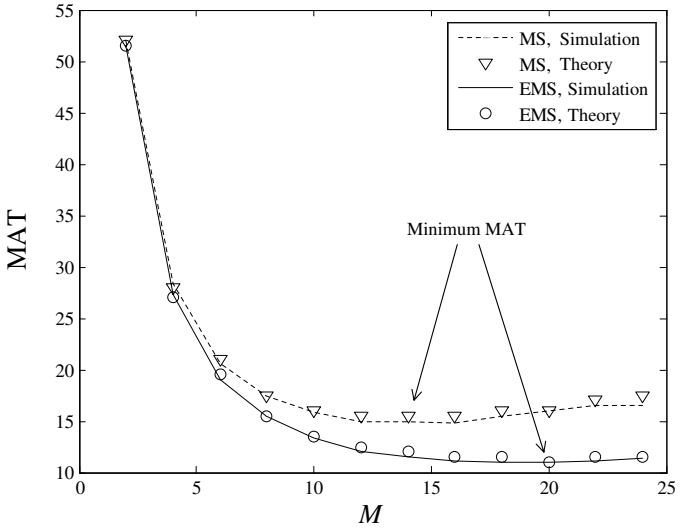


Fig. 6: The MATs of the MS and EMS schemes in a single-user environment.

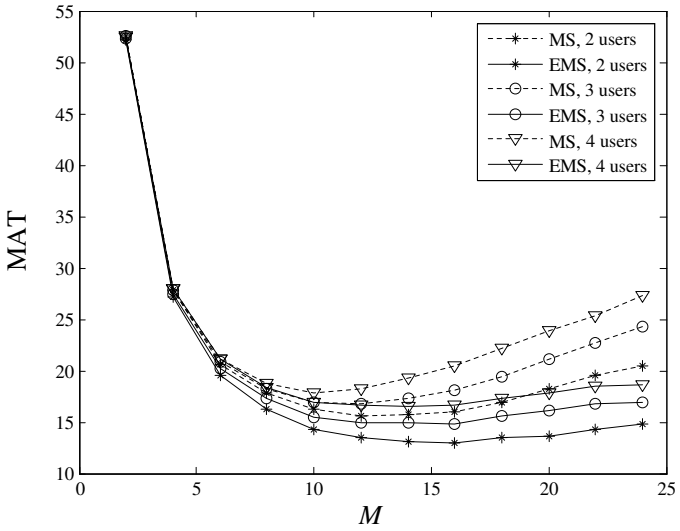


Fig. 7: The MATs of the MS and EMS schemes in a multi-user environment.

5. Simulation Results

In this section, we compare the MAT of the conventional MS scheme with that of the proposed EMS scheme in single-user and multi-user environments. Simulation parameters are as follows: $F = 200$; $K = 5$; $\lambda_a = \lambda_c = 1$; $N = 1 \sim 4$; $\alpha = 0.75$; $TH_{MS,first} = TH_{MS,second} = TH_{EMS,first} = TH_{EMS,second} = K$; $th_{EMS,first} = \alpha K$; and $L = 10$. We assume that each user transmits the data 1 or 0 with equal probability and chip is synchronized perfectly.

Fig. 6 shows the MAT performance of the MS and EMS schemes as a function of M in a single-user environment. In the figure, the dotted and solid lines represent the simulation results of the MS and EMS schemes, respectively, and, the markers ∇ and O represent the theoretical results of the MS and EMS schemes, respectively. From Fig. 6, we can observe that the MAT of the EMS scheme is shorter than that of the MS scheme, and confirm that the EMS and MS schemes provide the minimum MAT when $M = 14$ ($\sqrt{F} \simeq 14$) and $M = 20$ ($\sqrt{2F} = 20$), respectively. The difference of MAT between the MS and EMS schemes increases as M increases.

Fig. 7 shows the MAT performance of the MS and EMS schemes as a function of M in a multi-user environment. From Fig. 7, we can observe that the MAT of the EMS scheme is shorter than that of the MS scheme, and the difference of MAI between the MS and EMS schemes increases as M increases as in the single-user environment.

6. Conclusion

In this paper, we have proposed a novel code acquisition scheme called EMS for optical CDMA systems. Exploiting the multiple thresholds and modified local code, the proposed EMS scheme can provide a shorter MAT compared with that of the MS scheme. The performance of the EMS scheme has been analyzed using the circular flow graph diagram. The simulation results have confirmed that the EMS scheme offers a shorter MAT compared with that of the MS scheme in both single-user and multi-user environments.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012-0005066 with funding from the Ministry of Education, Science and Technology (MEST), Korea, by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency under Grant NIPA-2012-H0301-12-1005 with funding from the Ministry of Knowledge Economy (MKE), Korea, and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] A. Polydoros and C. L. Weber, "A unified approach to serial-search spread spectrum code acquisition-Part I: General theory," *IEEE Trans. Commun.*, vol. 32, pp. 542-549, May 1984.

[2] D. Chong, B. Lee, S. Kim, Y. -B. Joung, I. Song, and S. Yoon, "Phase-shift-network-based differential sequential estimation for code acquisition in CDMA systems," *Journal of Korea Inform. and Commun. Society*, vol. 32, pp. 281-289, Mar. 2007.

[3] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks - Part I: Fundamental principles," *IEEE Trans. Commun.*, vol. 37, pp. 824-833, Aug. 1989.

[4] A. Keshavarzian and J. A. Salehi, "Optical orthogonal code acquisition in fiber-optic CDMA systems via the simple serial-search method," *IEEE Trans. Commun.*, vol. 50, pp. 473-483, Mar. 2002.

[5] A. Keshavarzian and J. A. Salehi, "Multiple-shift code acquisition of optical orthogonal codes in optical CDMA systems," *IEEE Trans. Commun.*, vol. 53, pp. 687-697, Apr. 2005.

[6] J. A. Salehi and C. A. Brackett, “Code division multiple-access techniques in optical fiber networks - Part II: Systems performance analysis," *IEEE Trans. Commun*., vol. 37, pp. 834-842, Aug. 1989.
[7] A. Stok and E. H. Sargent, “Lighting the local area: Optical code-division multiple access and quality of service provisioning," *IEEE Networking*, vol. 14, pp. 42-46, Dec. 2000.

Joint-Design of PHY/MAC Layers for Throughput Optimization of Opportunistic Relay Networks

E. Adebola, O. Olabiyi and A. Annamalai
Center of Excellence for Communication Systems Technology Research (CECSTR)
Department of Electrical and Computer Engineering, Prairie View A&M University, TX 77446

Abstract - *This article develops a unified analytical framework for maximizing the user throughput of an opportunistic multi-relay network over generalized wireless channels using the symbol rate, packet length and constellation size of M-ary PSK/QAM digital modulation schemes as optimization variables. Optimization equations for each of the above degrees of freedom are derived in closed-form. A low-complexity discrete optimization algorithm for finding the "optimal" parameter-triplet is developed to solve the resulting non-linear joint-optimization problem. Numerical results reveal that the opportunistic routing protocols outperform the traditional cooperative system. Results also showed that increasing the number of opportunistic relays increases the throughput contrary to the case for the 'traditional' cooperative relays. Our results also reveal that it is sufficient to adapt the symbol rate and number of opportunistic relays in the low SNR regime while a joint-optimization of the packet length, constellation size and the number of opportunistic relays is desirable in the high SNR regime for maximum throughput.*

Keywords: *cross-layer optimization, adaptive PHY/MAC, generalized fading channels*

1 Introduction

It is well-known that adaptive transmission techniques can dramatically improve the spectrum utilization efficiency over time-varying channels [1]. The underlying premise is to "match" the transmission design parameters (e.g. signaling rate, power, coding rate, constellation size, packet length, etc.) to the prevailing channel conditions (i.e., based on the feedback of channel side information from the receiver to the sender). While the literature on adaptive link layer techniques are quite extensive that span over four decades, the art of adaptive link layer has received a renewed research interest in recent years, but in a cross-layer design framework, since the publications of [2]-[3]. For instance, [4]-[5] examined the throughput maximization problem via cross-layer adaptive designs following the approach in [2]-[3] for OFDM and MIMO systems, respectively. In [6]-[7], the efficacy of adaptive asymmetric (multi-resolution) modulation is investigated for multimedia data transmission. More recently, [8] has generalized the analytical framework for single user throughput optimization at the MAC layer in [2] (restricted to M-QAM and AWGN/Rayleigh environment) to different fading environments and modulation schemes.

Yet all of the above articles are restricted to point-to-point non-cooperative wireless network. Since cooperative diversity strategies can overcome the practical implementation issue of packing multiple antenna elements on a small-sized wireless device through cooperation among spatially distributed nodes ("virtual MIMO"), there has been some recent work on combining cooperative diversity protocols with adaptive PHY/MAC designs. Some of the notable contributions include [9]-[10]. However, their analyses are limited to a Rayleigh fading environment, and more critically the source rate adaptation was not considered. However, as pointed out in [2], [3] and [8], this parameter adaptation is extremely critical in the low SNR regime (e.g., tactical-edge or cell boundaries). Moreover, cooperative diversity is also most attractive to enhance the throughput in the low SNR regime. Reference [15] compares different cooperative diversity and opportunistic routing protocols. In this article we also extend the framework in [15] (restricted to Nakagami independent and identically distributed (i.i.d) environment) to different fading i.e. Rayleigh and Rice environments with independent and non-identically distributed (i.n.d) fading statistics.

The key objectives of this paper is to achieve the following: (i) develop a unified framework for throughput maximization of relay networks using opportunistic routing protocols with cross-layer adaptive designs at the PHY/MAC layers; and (ii) derive optimization equations in closed-form for optimal joint-adaptation of design parameters over generalized fading channels. To the best of our knowledge, these are still open research problems. To get the maximum throughput, we optimize the symbol rate R_s and packet length L (which are MAC layer parameters) and the constellation size $M = 2b$ (which is a PHY layer parameter) over a myriad of fading environments (Rice, Rayleigh, etc.) in a unified manner. Hence, this problem bears the characteristics of a cross-layer optimization problem across PHY and MAC layers. Our new framework allows us to gain insights on the impact of fade distribution on the optimal choice of design parameters and the optimized throughput performance for cooperative relay networks as well as comparison of different routing protocols

This work is supported in part by funding from the US Army Research Office (W911NF-04-2-0054), Air Force Research Laboratory/Clarkson Aerospace, and the National Science Foundation (0931679).

The remainder of the article is organized as follows. In Section II, we discuss the system model and the optimization framework while Section III derives the optimization equations in closed-form for a myriad of fading channel models and M-PSK modulation scheme. Selected computational results and concluding remarks are provided in Section IV and V respectively.

2 System Model & Optimization Framework

Consider a cross-layer design with the different protocol stack being considered in the proposed cross-layer design and optimization. The application layer generates the information at the rate R_s required by the data link layer. The network layer handles the packet routing and utilizes the preferred routing protocol for the application. Here, we assume the opportunistic routing is done at the network layer. Whichever case, the transmission route is either the combination of the entire *N* relays (maximum ratio combining) or selection of the relay in the best route. This relay selection can take place at the source in the route discovery process or can be distributed among the relays (through channel state dependent back off time). Also, the relay selection can take place at the destination and in this case it resembles traditional selection diversity combining. Each of these network layer protocols has different effect on the overall achievable spectral efficiency of the system and different channel side information requirement. The medium access control (MAC) and data link layers handle the packetization and determine the payload (i.e. packet length *L*) of the physical system. Lastly, the physical layer maps the bits in the packet to symbols based on the chosen modulation scheme. The number of bits per symbols b, determines the efficiency of the physical layer. In short, the adaptation of these four layers enables spectrally efficient and robust data transmission over time-varying channels i.e. 'match' transmission design parameters to the prevailing channel or network conditions. Therefore, efficient resource utilization strategies for handling multimedia data over wireless networks can benefit substantially from this cross-layer paradigm.

Consider a source node *S* with application/signaling rate R_s, trying to deliver a packet of length *L* using modulation rate *b*, over a cooperative multi-relay network with *N* relays to the destination node *D*. If we assume that the packet length *L* bits, includes *C* CRC bits (to ensure bits received in error are detected at the receiver) and $K = L - C$ information bits (payload). We further assume that the CRC decoder only exist at the destination node and is able to detect all errors in the received packet (since the probability of undetected errors is negligible for reasonable values of *C*). Hence, we assume here that the relays in this system are passive and do not decode the information but only amplify the received signal and forward it towards the destination. The additional assumptions made here is that the transmission parameters such as the signaling rate, modulation scheme, constellation size, packet/payload length are identical at both the source and the relays, irrespective of their wireless link conditions to the destination node, to reduce the implementation complexity. Similar to the assumptions in [3] and [8] and for ease of analysis, we also assume that both positive acknowledgements (ACK) and negative acknowledgments (NACK) from the destination node *D* are sent through a separate control channel and arrives at the transmitter S instantaneously and without error.

If the mean channel statistics are assumed to be the same during first transmission and retransmissions of a packet, then the throughput of a selective repeat automatic repeat request (SR-ARQ) protocol has been shown in [8] to be independent of the number of retransmissions per packet. Thus, the normalized throughput (average spectral efficiency) equation for cooperative diversity/opportunistic routing protocols with orthogonal transmissions is given by

$$\frac{\eta}{W} = \frac{L-C}{Lf(N)} bR_s(1-P_B) = \frac{L-C}{Lf(N)} b\frac{R_s}{W}(1-\overline{P}_s(b,\Omega_s))^{L/b} \quad (1)$$

where $\overline{P}_s(b,\Omega_s)$ denotes the average symbol error rate (ASER) of a specified *M*-ary modulation over fading channels, Ω_s is the mean received end-to-end SNR, *W* is the system bandwidth, $f(N)$ is the number of time slot or channels usage per each source transmission. This function is dependent on the choice of opportunistic routing protocol as will be shown later. Eq. (1) allows us to investigate the throughput optimization problem for cooperative relay networks without delving into the underlying protocol implementation as long as the end-to-end symbol error rate can be evaluated or known. The additional assumptions made here is that the transmission parameters such as the source rate, modulation scheme, constellation size, packet/payload length are identical at both the source and the relays, irrespective of their wireless link conditions to the destination node, to reduce the implementation complexity

In order to obtain the optimum throughput performance, the three design parameters (*b*, R_s, *L*) will be jointly-adapted. Before exploring the various combinations of design parameter optimization, we will first derive the moment generating function (MGF) of the relayed path that will later be utilized in evaluating the end-to-end average symbol error rate (ASER) of opportunistic relay networks over generalized fading channels and with any digital modulation scheme.

2.1 MGF of Opportunistic Routing Networks

2.1.1 AF Relaying with Opportunistic Route Selection and SDC (ORS-SDC) at the Source

In this protocol implementation, the best route is selected at the source based on the end-to-end relay SNR. The selection process can be done during the route discovery or in a distributed fashion similar to the proposition in [16]. Therefore the statistics of the best route selection is the same as the selection diversity combining at the destination as the best of *N*+1 links is being selected. However, the channel usage per source transmission in the case is $f(N) = 2$. Therefore, the spectral efficiency here does not reduce with

increasing number of relay as in the case of traditional cooperative diversity. Also the amount of channel side information and implementation complexity is highly reduced. The effective SNR is given by

$$\begin{aligned}\gamma_T^{ORS-SDC} &= \max(\gamma_{s,d}, \gamma_1, ...\gamma_N) \\ &\approx \max\left(\gamma_{s,d}, \min(\gamma_{s,1}\gamma_{1,d}), ...\min(\gamma_{s,N}\gamma_{N,d})\right)\end{aligned} \quad (2)$$

The cumulative distribution function (CDF) of the end-to-end SNR is then given by

$$\begin{aligned}F_{\gamma_T^{SDC}}(\gamma) &= F_{\gamma_{s,d}}(\gamma)\prod_{r=1}^{N} F_{\gamma_r}(\gamma) \\ &\approx F_{\gamma_{s,d}}(\gamma)\prod_{r=1}^{N}\left(1-[1-F_{\gamma_{s,r}}(\gamma)][1-F_{\gamma_{r,d}}(\gamma)]\right)\end{aligned} \quad (3)$$

The effective MGF can then be evaluated using the differentiation property of the Laplace transform via a single integral expression

$$\begin{aligned}\phi_{\gamma_T^{SDC}}(s) &= s\int_0^{\infty} e^{-s\gamma} F_{\gamma_{s,d}}(\gamma)\prod_{r=1}^{N} F_{\gamma_r}(\gamma)d\gamma \\ &\approx s\int_0^{\infty} e^{-s\gamma} F_{\gamma_{s,d}}(\gamma)\prod_{r=1}^{N}\left(1-[1-F_{\gamma_{s,r}}(\gamma)][1-F_{\gamma_{r,d}}(\gamma)]\right)d\gamma\end{aligned} \quad (4)$$

For special case of independent and identically distributed (i.i.d.) Rayleigh channel, the MGF can be reduced to

$$\phi_{\gamma_T^{SDC}}(s) \approx s\int_0^{\infty} e^{-s\gamma}\left(1-e^{-\gamma/\Omega_{s,d}}\right)\prod_{r=1}^{N}\left(1-e^{-\gamma\frac{\Omega_{s,r}+\Omega_{r,d}}{\Omega_{s,r}\Omega_{r,d}}}\right)d\gamma \quad (5)$$

2.1.2 AF Relaying with Opportunistic Relay Selection and MRC (ORS-MRC) at Destination

This protocol implementation takes advantage of the half duplex nature of relay transmission to achieve better performance than the ORS-SDC protocol. Here, since the source transmits in the first transmission phase, and due to the broadcast nature of wireless channel, the destination can be close enough to receive this signal before receiving the signal from the relay. This is particularly true in the distributed ORS protocol implementation proposed in [16]. Therefore, if the channels side information of both links is available, the received signal can be combined with maximum ratio combining scheme at the destination. Note that the transmission channel usage $f(N) = 2$ per each source transmission but the statistics is slightly different from the ORS-SDC protocol. The effective end-to-end SNR of ORS-MRC protocol can be expressed as

$$\begin{aligned}\gamma_T^{ORS-MRC} &= \gamma_{s,d} + \max(\gamma_1, \gamma_2, ...\gamma_N) \\ &\approx \gamma_{s,d} + \max\left(\min(\gamma_{s,1}\gamma_{1,d}), ...\min(\gamma_{s,N}\gamma_{N,d})\right)\end{aligned} \quad (6)$$

The effective MGF can then be evaluated using the addition and differentiation property of the Laplace transform via a single integral expression

$$\begin{aligned}\phi_{\gamma_T^{ORS-MRC}}(s) &= s\phi_{\gamma_{s,d}}(s)\int_0^{\infty} e^{-s\gamma}\prod_{r=1}^{N} F_{\gamma_r}(\gamma)d\gamma \\ &\approx s\phi_{\gamma_{s,d}}(s)\int_0^{\infty} e^{-s\gamma}\prod_{r=1}^{N}\left(1-[1-F_{\gamma_{s,r}}(\gamma)][1-F_{\gamma_{r,d}}(\gamma)]\right)d\gamma\end{aligned} \quad (7)$$

For special case of independent and identically distributed (i.i.d.) Rayleigh channel, the MGF can be reduced to

$$\phi_{\gamma_T^{ORS-MRC}}(s) \approx \frac{s}{1+s\Omega_{s,d}}\int_0^{\infty} e^{-s\gamma}\prod_{r=1}^{N}\left(1-e^{-\gamma\frac{\Omega_{s,r}+\Omega_{r,d}}{\Omega_{s,r}\Omega_{r,d}}}\right)d\gamma \quad (8)$$

The respective CDF and MGF expressions for different fade distributions are given in Table 2.

2.2 Symbol Rate Optimization

In order to determine the optimum value for symbol rate R_s (that maximizes the throughput) in a generalized fading channel, we need to differentiate (1) with respect to Ω_s, set the resulting expression to zero, and then solve it to obtain the optimum SNR/symbol Ω_s^*, viz.,

$$\Omega_s^* \left.\frac{d\overline{P}_s(b,\Omega_s)}{d\Omega_s}\right|_{\Omega_s=\Omega_s^*} = -\frac{1-\overline{P}_s(b,\Omega_s^*)}{L/b} \quad (9)$$

Once Ω_s^* is determined (from (9)), we can then find R_s using relation in (10), viz,

$$R_s^* = \frac{P_r}{\Omega_s^* N_0} = \Omega_r \frac{W}{\Omega_s^*} \quad (10)$$

It is important to note that in the low SNR regime (i.e., when the received SNR/symbol Ω_r is lower than the "optimum" SNR/symbol Ω_s^*), we should select R_s^* that satisfies (10) to maximize the throughput. However, in the high SNR regime (i.e., when $\Omega_r > \Omega_s^*$), the bandwidth constraint prevents us from increasing the symbol rate beyond a certain limit ($R_s \le W$). In this case, we set $R_s = W$ and pack more bits per digital modulation symbol (i.e., to maximize the spectral efficiency at the prevailing channel condition).

2.3 Packet Length Optimization

In order to find an analytic solution for the optimal packet length L^*, we need to assume that L takes continuous values. Differentiating (1) with respect to L and solving the resulting expression leads to a closed-form formula for L^*, viz., [2]

$$L^* = \frac{C}{2} + \frac{1}{2}\sqrt{C^2 - \frac{4bC}{\ln(1-\overline{P}_s(b,\Omega_s))}} \quad (11)$$

From (11), it can be seen that the optimal packet length L^* depends on the constellation size and ASER (which in turn depends on the SNR per symbol Ω_s). In practice, the packet lengths are generally in the form of integer multiples of the number of bits per symbol (i.e., $L = b, 2b, 3b, \ldots$, and so on).

2.4 Constellation Size Optimization

Optimum value of b for which the throughput is maximized leads to the optimum constellation size. In order to derive an expression for finding the optimum b, we assumed that b is a continuous variable. Hence differentiating (1) with respect to b and solving the resulting to find its saddle-point, we obtain

$$\left.\frac{d\overline{P}_s(b,\Omega_s)}{db}\right|_{b=b^*} = -\frac{1-\overline{P}_s(b^*,\Omega_s)}{L} \tag{12}$$

In a practical discrete-rate adaptive modulation, b should assume a positive integer value. Thus the value b obtained from solving the transcendental equation (12) should be rounded to the nearest positive integer value.

2.5 Joint Parameter-Triplet Optimization

Joint optimization can be done by jointly varying the three optimization parameters (b, Ω_s , L). Alternatively, we can find an optimal triplet set (for every channel realization) that simultaneously satisfies equations (9), (11) and (12). From our discussions in the preceding subsections, it can shown that the optimal throughput is given by

$$\eta^* = \begin{cases} \frac{L^*-C}{L^*} b^* R_s^* (1-\overline{P}_s(b^*,\Omega_s^*))^{L^*/b^*} & \Omega_r \le \Omega_s^* \\ \frac{L^*-C}{L^*} b^* W (1-\overline{P}_s(b^*,\Omega_r))^{L^*/b^*} & \Omega_r > \Omega_s^* \end{cases} \tag{13}$$

From (9) and (12), it can be observed that (1) is strictly concave for all range of Ω_s and b and therefore there is singular maximum point for both Ω_s^* and b*. However, since (9), (11) and (12) are non-linear equations but b and L are integers for practical purposes, it might be difficult to solve the combinatorial programming problem. Here, we propose a very simple routine that can yield highly accurate optimal points. The proposed method circumvents the need to solve the non-linear equation in the solution proposed in [8]. The routine steps are given as follows.

A Simple Algorithm for Joint Optimization of Parameter Triplet (b, Ω_s, L)

Step 1:
Specify a range for b. The choice of b should be constrained to integer values for practical discrete modulation schemes. In this work, we assume b = [1, 2, 3, ..., 5].

Step 2:
Substitute (10) (i.e., $R_s = (\xi P_r/N_o)/\Omega)$) and (11) into (1) where Ω denotes the mean received SNR for the *S-D* link while ξ is a positive constant that relates Ω_s to Ω. For any fixed P_r/N_o, the resulting auxiliary expression is a concave function with respect to Ω. For each b specified in Step 1, sweep and find Ω^* that maximizes this auxiliary function.

Perform the Steps 3-5 for each mean received SNR Ω.

Step 3:
Obtain R_s^* from $R_s = W\Omega/\Omega^*$ using Ω^* from Step 2.

Step 4:
Substitute R_s^* from Step 3 and (11) into (13) with the condition that $L^* = L_{\max}$ if $L^* \ge L_{\max}$ to obtain b^* (that maximizes the throughput).

Step 5:
Compute L^* using (13) (with the condition that $L^* = L_{\max}$ if $L^* \ge L_{\max}$) with b^* from Step 4.

The above algorithm greatly simplifies the combinatorial programming problem for the parameter-triplet optimization and can be readily applied to a number of other cases of interest including the shadowed fading environments.

3 Optimization Equations in Closed-Form

In order to derive a closed-form expression for end-to-end ASER of opportunistic routing networks, we exploit a desirable exponential form for the conditional error probability (CEP) for various digital modulation schemes. For instance, the CEP for the MPSK in AWGN is given by [13, Table II]

$$P_s = ae^{-b_0\gamma_s \sin^2(\pi/2^b)} + ce^{-2b_0\gamma_s \sin^2(\pi/2^b)}, \quad b \ge 2 \tag{14}$$

where coefficients a, b_0 and c are tabulated in [13, Table I]. This form is very attractive since the averaging of the CEP shown in (14) over the probability density function (PDF) of the fading SNR can be simplified dramatically (i.e., the ASER is simply a function of the Laplace transform of the PDF). Therefore, the end-to-end ASER can be expressed as

$$\overline{P_s} = a\phi(s,\Omega) + c\phi(2s,\Omega) \tag{15}$$

where $s = b_0 \sin^2(\pi/M) = b_0 \sin^2(\pi 2^{-b})$ and $\phi(.,.)$ is the MGF of end-to-end SNR. Substituting (4) and (7) into (15), we obtain a unified expression for the ASER in a myriad of fading environments (including mixed-fading).

4 Numerical Results

For the purpose of illustration, we have considered a triple opportunistic relay system in our numerical analysis although the developed framework is applicable to any arbitrary number of relays. Unless otherwise stated, the parameters in Table 1 will be used to generate the plots in this section.

TABLE 1. SIMULATION PARAMETERS

Description	Value
CRC, C	16 bits
Packet Length $L_{\max}$	512 bits
Bandwidth W	1024 kHz
Constellation size $b_{\max}$	5 bits
Power factor	$\delta_{s,r} = \delta_{r,d} = 2$

Fig. 1 shows the plot of L* as a function of average SNR for different values of b. It can be observed that the optimum L* increases as the average SNR increases. However, in practical implementation, the value L* cannot be more than $L_{\max}$ as this can lead to more delay and larger memory requirement. Also, we see L* is maximum when b = 1 although the throughput is not maximum, as will be shown in Fig. 2.

Fig. 2 shows the throughput vs. average SNR using the L^* values obtained from (11). It can be observed that lower b values give the best performance at low SNR and a higher b gives better performance as the SNR increases. This is an example of two parameter optimization at fixed R_s.

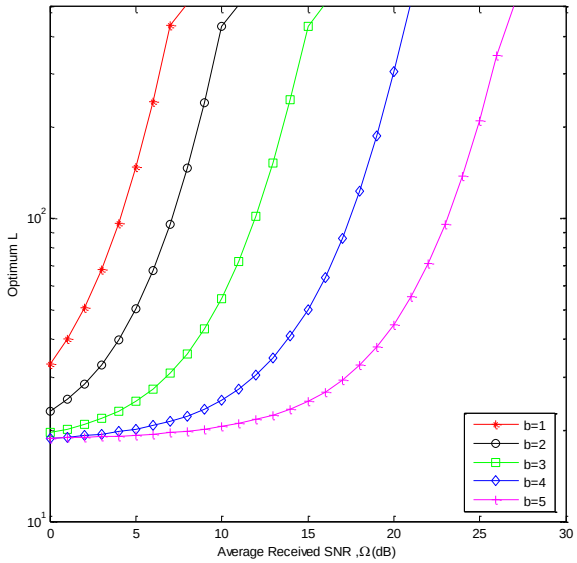


Fig. 1 Optimum L^* versus average SNR with different b values over Rayleigh fading for a triple opportunistic relay system

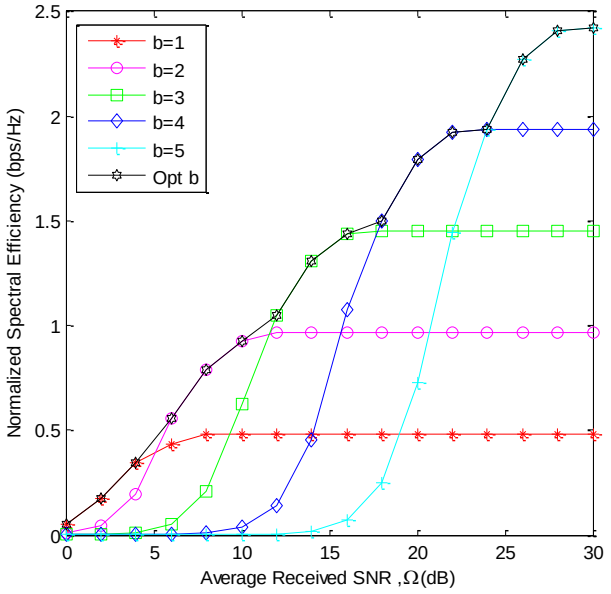


Fig. 2 Throughput vs. average SNR at L^* with R_s = 1024 kbps and different b values over Rice (K=4) fading channel for a triple opportunistic relay system.

In Fig. 3 we illustrate the auxiliary function (in step 2 of the algorithm for joint triplet parameter optimization) vs. average SNR at L^* and variable b. This plot shows the existence of a preferred mean SNR Ω_s^* for each b and can be used to determine the optimum R_s^*. Note that Ω_s^* for each of the constellation size is obtained from the peak throughput. For instance, $\Omega_s^* = 6$ dB for b = 2. Therefore, when $\Omega_r \geq \Omega_s^*$, $R_s^* = W = 1024kbps$, and when $\Omega_r < \Omega_s^*$, we obtain R_s^* using (10). In this example, $R_s^* = 646kbps$ at 4 dB.

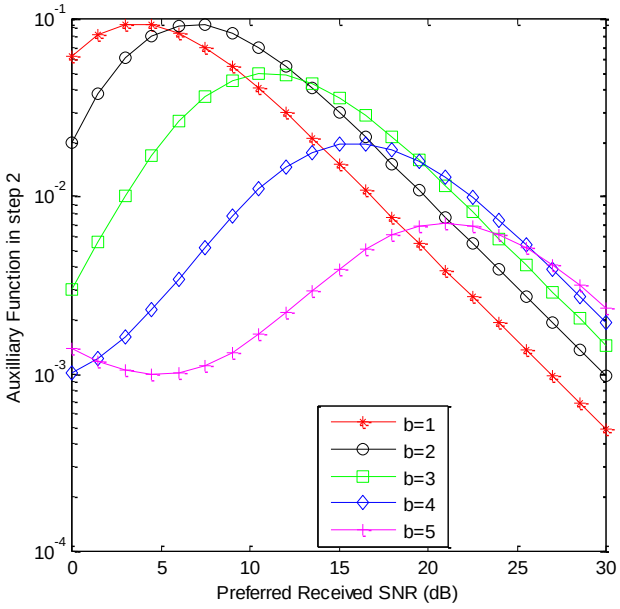


Fig. 3 Auxiliary Function vs. preferred average SNR at L^* for different b over Rayleigh channel for a triple opportunistic relay system

Fig. 4 shows the throughput against the average SNR with R_s^* and L^* values for different values of b. The optimum curve shows the optimum values of b across the range of average SNR and therefore shows the maximum achievable throughput when all the triplet parameters have been optimized.

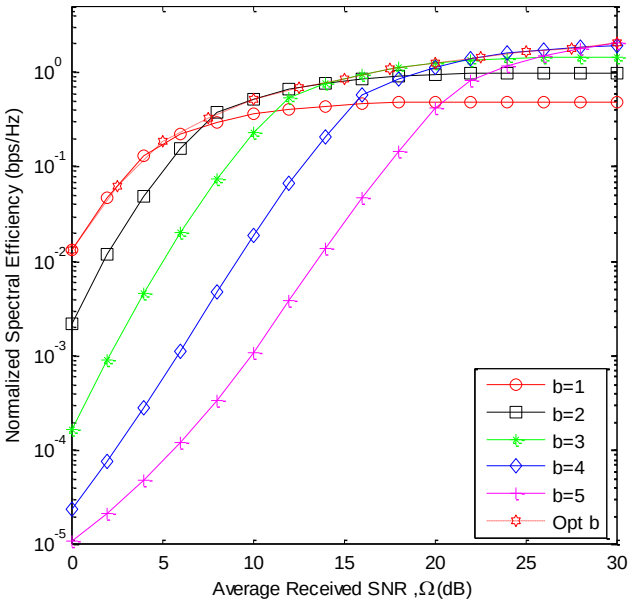


Fig. 4 Throughput vs. average SNR with R_s*, L^* for different b over Rice (K= 4) fading channel for a triple cooperative relay.

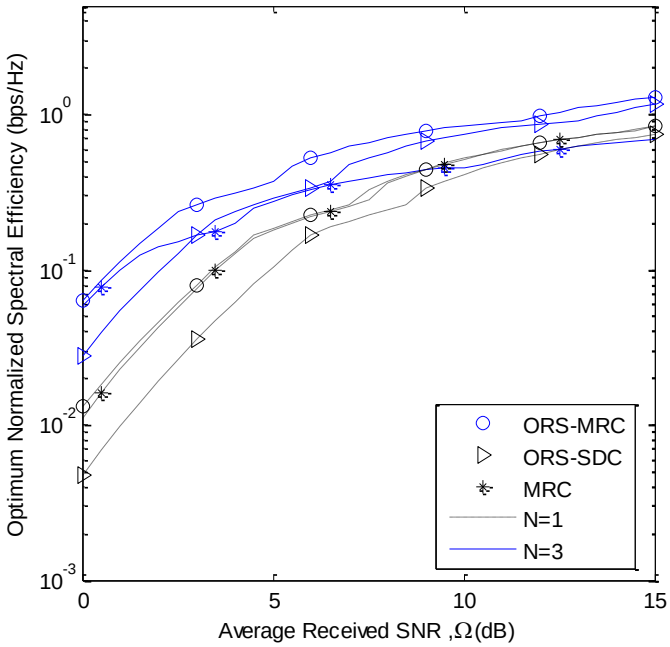


Fig. 5 Comparison of the efficacy of different cooperative diversity and opportunistic routing protocols on the cross-layer design optimization with Rayleigh channel, and $N=\{1, 3\}$.

Fig.5 compares the efficiency of different network layer protocols on the cross layer-design. Several interesting observations can be drawn from this figure. First, the figure shows that, for single relay case, the ORS-MRC gives the same performance as MRC scheme. However, as the number of relay increases, the ORS-MRC protocol performs better than MRC protocol at all range of received SNR. This is because irrespective of the number of relay in participation, the total channel usage for ORS-MRC is kept constant at two timeslots per source transmission whereas, the channel usage for the 'traditional' schemes (i.e. MRC and SDC scheme) increases with increasing number of relays. In all, the ORS-MRC gives the best overall performance. Therefore, it could be concluded that, the ORS-MRC and MRC protocols are recommended for nodes at the tactical edge or at the cell boundary where the received signal strength is very low. However, the ORS-MRC and ORS-SDC are of choice at higher received SNR and less severe fading environment.

5 Conclusion

In this article, we have examined the impact of opportunistic route selection and the joint-optimization of symbol rate, packet length and constellation size on the user throughput of cooperative relay networks. We concluded that the opportunistic routing is the more advantageous than the cooperative diversity at high SNR regime and better channel condition, and with large number of participating relays. Thus the number of cooperating relays and the source rate can be jointly optimized for maximizing the throughput at the tactical-edge or cell boundaries, while packet length, constellation size and number of cooperating relays should be jointly optimized to maximize the throughput in better channel conditions.

6 References

[1] A. Goldsmith and S. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels";*IEEE Trans. Communications*, Vol. 45, pp. 1218-1230, Oct. 1997.

[2] T. Yoo, R. Lavery, A. Goldsmith and D. Goodman, "Throughput Optimization using Adaptive Techniques"; Internet draft, May 2005.

[3] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-Layer Design of Ad-Hoc Networks for Real-Time Video Streaming"; *IEEE Wireless Communications*, Vol. 12, No.4, pp. 59- 65, Aug. 2005.

[4] Y. Fakhri, B. Nsiri, D. Aboutajdine and J. Vidal, "Throughput Optimization via the Packet Length and Transmission Rate for Wireless OFDM System in Downlink Transmission"; *International Journal of Computer Science and Network Security*, Vol. 6, No. 3B, March 2006.

[5] Y. Cho, M. Kim and F. Tobagi, "Adaptive Throughput Optimization for MIMO Systems in Rayleigh Fading Channels"; *Proc. IEEE ICC'07*, pp. 2574-2579, June 2007.

[6] M. Pursley and J. Shea, "Adaptive Nonuniform Phase-Shift-Key Modulation for Multimedia Traffic in Wireless Networks"; *IEEE Journal Select. Areas Communications*, pp. 1394-1407, Aug. 2000.

[7] J. James, O. Odejide and A. Annamalai, "Adaptive Multiresolution Modulation for Multimedia Traffic"; *Proc. IEEE CCNC'11,* Las Vegas, Jan. 2011.

[8] O. Odejide and A. Annamalai, "Further Results on Throughput Optimization using Adaptive PHY/MAC/APP Layer Techniques"; *Proc. IEEE MILCOM'11*, Baltimore, Nov. 2011.

[9] L. Dai and K. B. Letaief, "Throughput Maximization of Ad-Hoc Wireless Networks using Adaptive Cooperative Diversity and Truncated ARQ"; *IEEE Trans. Communications*, Vol. 56, No. 11, pp. 1907-1918, Nov. 2008.

[10] A. Vosoughi and Y. Jia "Joint Adaptive Modulation-Coding and Cooperative ARQ for Wireless Relay Networks"; *Proc. 33rd IEEE Sarnoff'10*, Princeton, April 2010.

[11] A. Annamalai, B. Modi, R. Palat and J. Matyjas "Tight Bounds on the Ergodic Capacity of Cooperative Analog Relaying with Adaptive Source Transmission Techniques"; *Proc. IEEE PIMRC'10*, pp. 18-23, Istanbul, Sept. 2010.

[12] B. C. Modi, O. Olabiyi and A. Annamalai "On Ergodic Capacity of Cooperative Amplify-and-Forward Relay Networks in Rice Fading Environments"; *Proc. IEEE GLOBECOM'11*, Houston, Dec. 2011.

[13] O. Olabiyi, and A. Annamalai, "ASER Analysis of Cooperative Non-Regenerative Relay Systems over Generalized Fading Channels"; *Proc. IEEE ICCCN'11,* Maui, Aug. 2011.

[14] T. Wang, A. Cano and G. Giannakis, "High-Performance Cooperative Demodulation with Decode-and-Forward Relays"; *IEEE Trans. Communications*, Vol. 55, No. 7, pp. 1427-1437, July 2007.

[15] O. Olabiyi, A. Annamalai, O. Odejide and E. Adebola "Integrated Design of APP/NET/PHY/MAC Layers for cooperative Relay Networks"

[16] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman "A simple cooperative diversity method based on network path selection," *IEEE Journal Select. Areas Communications*, Vol. 24, No. 3, pp. 659-672, Mar. 2006.

[17] M. K. Simon and M-S Alouini "*Digital Communication over Fading Channels"*. New York: Wiley, 2nd edition, 2005.

TABLE 2

MGF OF SNR FOR SEVERAL COMMON STOCHASTIC CHANNEL MODELS

Channel Model	**MGF of SNR** $\phi_\gamma(s)=\int_0^\infty f(\gamma)e^{-s\gamma}d\gamma$	**CDF of SNR** $P_\gamma(\gamma)$
Rayleigh	$(1+s\Omega_s)^{-1}$	$1-e^{-\gamma/\Omega_s}$
Nakagami–n (Rice: K=n^2)	$\frac{1+K}{1+K+s\Omega_s}\exp\left(\frac{-Ks\Omega_s}{1+K+s\Omega_s}\right)$	$1-Q_1\left(\sqrt{2K},\sqrt{2(1+K)\gamma/\Omega_s}\right)$
Nakagami-m	$\left(1+\frac{s\Omega_s}{m}\right)^{-m}$	$1-\frac{\Gamma\left(m,m\gamma/\Omega_s\right)}{\Gamma(m)}$
G-distribution	$m\sum_{r=0}^{m+k-1}\binom{m+k-1}{r}\frac{(-1)^{r+1}s^{-k}}{(r+1-k)!}\sum_{p=0}^{r}\binom{r}{p}\left(2\sqrt{\frac{\alpha s}{\beta}}\right)^{r-p+1}$ $\Gamma(r+p+1)H_{-(r+p+1)}\left(\frac{b}{2}\sqrt{\frac{b}{s}}+\sqrt{\frac{\alpha s}{\beta}}\right)$ $\eta=\frac{\exp(\mu)}{2\sinh\left(\frac{\sigma^2}{2}\right)},\ \theta=\exp\left(\mu+\frac{\sigma^2}{2}\right)$ $\beta=2m\theta,\ \alpha=\eta\Omega_s,\ b=\frac{1}{\theta}\sqrt{\frac{\eta}{\Omega_s}}$ $H_\nu(x)$ = *Hermite function*	$-A\Gamma(m)\sum_{k=1}^{m}\frac{2^k\gamma^{m-k}}{(\beta b)^k(m-k)!}\frac{K_{m-k+\frac{1}{2}}\left(b\sqrt{\alpha+\beta\gamma}\right)}{\left(\sqrt{\alpha+\beta\gamma}\right)^{m-k+\frac{1}{2}}}$ $A=\frac{(\eta\Omega_s)^{\frac{1+2m}{4}}}{\Gamma(m)}\sqrt{\frac{2\eta}{\pi\theta}}\exp\left(\frac{\eta}{\theta}\right)\left(m/\Omega_s\right)^m$ $K_\nu(x)$ = Modified Bessel function of the second order v

where K and m are the fading indices for rice and nakagami channels. $\mu(\text{dB})$ and $\sigma(\text{dB})$ are the mean and the standard deviation of $10\log_{10}\Omega_s$ respectively.

An Efficient Software Update Method for WSNs

Hyeyeong Jeong[1], Yeungmoon Kwon[1], Byoungchul Ahn[1] and Bruce Kim[2]
[1]Dept. of Computer Engineering, Yeungnam University, Gyungsan, Korea
[2]Dept. of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL, USA

Abstract - *Wireless Sensor Networks(WSNs) are applied to many monitoring areas and are deployed for long periods of time using batteries. Present sensor nodes can perform many functions at the same time and contain complex software. During the lifetime of sensor nodes, they are required to reprogram their software because of their new functions, software, software bug fixes. The nodes are inaccessible physically or it is very difficult to upgrade their software by one by one. To upgrade the software of sensor nodes in WSNs remotely, this paper presents an energy efficient method by selecting an optimal relay node. The proposed method is compared with the flooding method and the SPIN method. Three methods are simulated in NS-2 with the same environmental parameters. Their performances are measured by upgrade times, the number of relay nodes, energy consumptions and error rates according to several packet sizes. Simulation results show that the energy consumption of the proposed method is less than 50% that those of other method. The update time of the proposed method is 33% faster than others*.

Keywords: *Software upgrade, Remote upgrade, flooding, gossiping, WSN software*

1. Introduction

Since Wireless Sensor Networks(WSNs) are applied to many different domains such as forest fire monitoring, agriculture monitoring and control, environment monitoring and so on, their node software is required to be updated by adding functions, removing bugs and reconfiguring their functions. A few years ago, their performance and functions were very limited and were discarded without reusing or reprogramming. But sensor nodes for WSNs have been developed to reduce power consumption and can be reused without discarding them by the recent technical advances

Therefore, the deployed nodes need to be updated software to fix software bugs, or to reprogram new functions without discarding or removing them. But it is very difficult to collect them, reprogram them and deploy them back in real world. Also a lot of budgets and efforts are wasted if they are recollected and deployed them again. Therefore, all nodes need to be upgraded their software remotely.

This paper presents an energy efficient software update method by reducing the number of relay nodes from the base node to all nodes of WSNs.

2. Related Work

Stephen *et al*. suggest a model for upgrading software in WSNs[8][9]. This model presents the theoretical approaches to upgrade software. They have not presented any simulation results or experiments to verify their model. They validate their model against three different systems, representing three classes of software update: static/monolithic updating(MOAP), dynamic/mobile agent-based updating(Mate) and dynamic/component-based updating(Impala)

Since control software contains execution code for a processor of sensor nodes, it is very important to maintain reliable data transfer. A method for reliable data transfer in WSNs is developed for 1:1 communication such as S-TCP[3] and RMTS[4]. But 1:1 communication methods are inefficient to upgrade many nodes of WSNs. If these methods are used for re-programing sensor nodes of WSNs, each node must be updated first and retransmit the software upgrade data to another node one by one. Therefore it is necessary to develop an efficient upgrade method for sensor nodes with fast upgrade time and small data retransmission.

The direction to upgrade control software is the opposite direction of normal data transfer[5][6]. It is necessary to study for large data transfer from one node to many nodes efficiently. There are some researches about upgrades for sensor nodes. But they are focused on system management, not an upgrade itself [7]. In this paper, an energy efficient software upgrade method is described by comparing the other methods.

3. Proposed Upgrade Model

All sensor nodes of WSNs are assumed to be the same model with the memory size and the same processor. It means that all sensor nodes use the same software version. And a distance between two nodes is the same and the location of nodes is fixed.

3.1 Data Relay

Generally, the normal mode of a node is operating in sensing data mode with low power consumption. Each node transmits its sensed data to the sink node. When a sink node starts to transfer its software update data to others, all nodes stop their sensing operation and switch their mode to the software upgrade mode. In case of software data transmission, data size is very large and must be transmitted very fast and continuously. If a node detects the software upgrade protocol, it should switch the normal sensing mode to the software upgrade mode and prepare for the software upgrade. When the node finishes receiving all software upgrade data, it requests lost packets to its source node. After lost data are received again, the node reprograms its own flash memory and restarts its operation again. After upgrading its software, the node must relay the software upgrade data to other nodes. But it is not necessary for all nodes to participate in relaying the software upgrade process to another node in WSNs. Only a few nodes need to relay the software upgrade data to other nodes. Therefore, it is very important to choose relay nodes because of overall performance.

The total number, *N*, of relay nodes is calculated by *l* / *r*, where *l* is the maximum distance of a sensor network and *r* is radio radius of a node. All relay nodes are placed in multiples of radio radius of the location. The other nodes only receive the software upgrade data and reprogram themselves. When software is updated by *N*, it is the optimal topology. But in real WSNs, each node is located on 2-dimensional plain. For 2-dimensional plain at each step, the total number of relay nodes, *N*, is calculated by (1).

$$N = \begin{cases} 1, & (l <= r) \\ 1 + \sum_{n=1}^{n=\lfloor d \rfloor} \left\lceil \frac{2\pi \cdot n \cdot r}{r} \cdot \frac{1}{4} \right\rceil = 1 + \sum_{n=1}^{n=\lfloor d \rfloor} \left\lceil \frac{\pi \cdot n}{2} \right\rceil & (l > r) \end{cases} \quad (1)$$

r is a radio radius of a node
l is distance between start node and last node
d = l/r

To upgrade all nodes in the field at least *N* nodes should participate in relay operation. When the start node is located in the corner of WSNs, Equation (1) is acceptable. If the start node is located in the center of WSNs, the number of relay nodes, *N*, is increased up to four times

3.2 Power Consumption and Upgrade times

The power consumption of relay nodes is calculated by the sum of receiving data and transmitting data, and is calculated as Equation(2). At Equation (2), J_{nr} and J_{ns} are the power consumption of receiving data and transmitting data to other nodes. Some nodes located in duplicated radio area are received a few multiple times of data size of the software upgrade data. Therefore, the total energy consumption of all nodes is *J* in (3).

$$\begin{aligned} J_{nr} &= J_r \cdot file\_size\,(1+e) \\ J_{ns} &= J_s \cdot file\_size\,(1+e) + J_{nr} \end{aligned} \quad (2)$$

J_r *is the energe for datas end*
J_s *is the energe for datas send*
J_{nr} *is the energy consumped by receiving node*
J_{ns} *is the energy consumped by relaying node*
e is trasnmissi on error rate
file_size is the size of firmware data

$$J = (J_s + J_r \frac{N_t}{Area\_of\_field} \cdot \pi \cdot r^2 + e \cdot J_s) file\_size \cdot N \quad (3)$$

N_t *is total number of nodes in a filed*
N is a number of relaying nodes

Equation (3) represents that all parameters are fixed except *e* and *N*. It is very important to reduce transmission error and the number of relaying nodes. The time to upgrade all nodes of WSNs depends upon the step count of relay. The time is calculated by Equation (4).

$$T = \lceil d \rceil \cdot (t_s + e \cdot t_s + t_u) \quad (4)$$

d is relay step count (l/r)
t_S *is firmware file sned time (file_size / bandwidth)*
t_u *is update time*
e is transmission error rate

From Equation (4), the total software update time of WSNs depends upon relay step count, *d*, and transmission error rate, *e*. If all relay nodes send software upgrade data to other nodes at the same time, they make a lot of collision or interference by radio signals and the transmission error rate is increased. It is necessary to prevent all nodes from relaying data at the same time in each step. To reduce energy consumption and upgrade time, it is very important to select a relay node at each step.

3.3 Proposed Method

It is very important to reduce the number of relay nodes and prevent neighbor nodes from retransmitting data at the same time. It saves energy consumptions as well as data error rates. The proposed method is to select a node which have many neighbor nodes.

As soon as a relay node transmits the software upgrade data, it should select the next relay node to propagate the software update operation. If the next relay node is located on the boundary line of radio radius of the present relay node, it is the most effective relay node. But in real WSNs, there are a few nodes on the boundary line of radio radius. It is hard to recognize if the nodes are located on boundary line of radio radius. Selection methods of relay nodes are carefully considered.

The software upgrade protocol has two steps. The first step is pre-transmission step. A relay node should know the

status of all neighbor nodes before it starts sending the software upgrade data. All neighbor nodes broadcast their status to the relay node periodically in the first step. The status data includes the upgrade version information, the number of neighbor nodes, energy and the node status.

The next step is to receive the software upgrade data and update it. Before the relay node starts to send the software upgrade data, neighbor nodes respond by sending a "*receive-start*" message. And neighbor nodes start to receive the software upgrade data and store it to their memory. After the upgrade process, neighbor nodes must request the relay node to retransmit lost packets. After they finish receiving the lost packets, they reprogram by themselves and send back "*reprogram-done*" message to the relay node. If a time-out occurs or it receives messages from all participated neighbor nodes, the relay node should select the next relay node from neighbor nodes and sends "*relay-start*" message.

```
Data send node
Broadcast status information periodically
If (receive "relay-start" && exist proper neighbor node) {
  Send "data-start" to neighbor nodes.
  Send data
}
While(No. data receive nodes > 0) {
  If(receive "reprogram-done") {
    Select one node.
    Send "relay-start" to selected node.
    Decrease No. of data receive node.
  }
}
Go to 1.1
```

```
Data receive node
Broadcast status information periodically
If (receive "data-start") {
  If (receive firmware ver. > stored firmware ver.) {
    Receive data and store it to memory
    Request missing or lost packet
    Reprogramming it-self.
    Send "reprogram-don" to data send node.
  }
}
If( receive "relay-start")
  Go to 1.1
Else
  Go to 2.1
```

Figure 1. Proposed method algorithm

3.4 Comparison Methods

In order to evaluate the proposed methods, two methods are used. One is flooding and gossiping method by Brown [8][9] and the other is SPIN by Kulik. Flooding and gossiping are two classical mechanisms to relay data in sensor networks without the need for any routing algorithms and topology maintenance. SPIN efficiently disseminates information among sensors in an energy-constrained wireless sensor network. SPIN nodes can base their communication decisions both upon application-specific knowledge of the data and upon knowledge of the resources that are available to them.

3.5 Performance Metrics

In order to evaluate the proposed methods, following metrics are used.

- The number of relay nodes - Depending on the selection method of relay nodes, the number of relay nodes might be varied. It is the most important factor.
- Energy consumption - Since sensor nodes in WSN are operated by battery power, the energy consumption is important factor with total upgrade time.
- Total upgrade time to all nodes - This factor shows the effectiveness of the software upgrade in the WSNs. Depending upon relaying nodes, the total upgrade time is much different.
- Data loss rate – This factor shows the performance of each selection method. Typically, collision and interference leads to data packet loss in WSNs.

4. Simulation Results

4.1 The Number of Relay Nodes

The number of relay nodes to upgrade software is calculated by (2). Sensor nodes in WSNs are deployed in rectangular. It is very difficult to calculate exact relay nodes for the Maximum-Distance method. By Equation (2), the number of relay nodes is about 37. Figure 2 shows the number of relay nodes. In all case, the number of relay nodes is between 33 ~ 49 nodes. When the packet size is 256bytes, the total number of relay nodes is 33 nodes. When the packet size are 128bytes and 192bytes, the total number of relay nodes are 37 nodes and 36nodes respectively. The difference of relay node is six and it does not impact the number of relay nodes

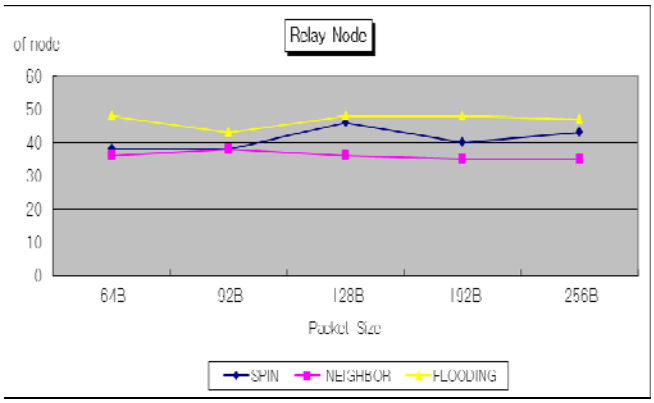


Figure 2. Number of relay nodes

4.2 Data Loss

Figure 3 shows the size of lost data. When the packet sizes are 64bytes or 92 bytes, nodes transmit many data packets and it generates so many collisions. And the collision induces energy consumption and total upgrade time. When packet sizes are between 192bytes and 224 bytes, data loss is less than 150KB.

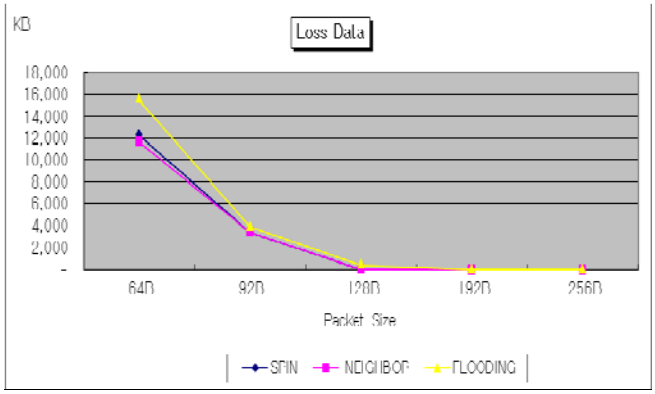


Figure 3. Data loss rate

4.3 Energy Consumption

The energy consumed by each node is calculated with (3). Figure 4 shows total energy consumption to upgrade all nodes. When the packet size is 256 bytes, the energy consumption is the lowest. When the packet sizes are between 64bytes and 92 bytes, the energy consumptions are high when they are compared to 192 bytes or longer packet sizes. The energy consumption is related to the data loss rate..

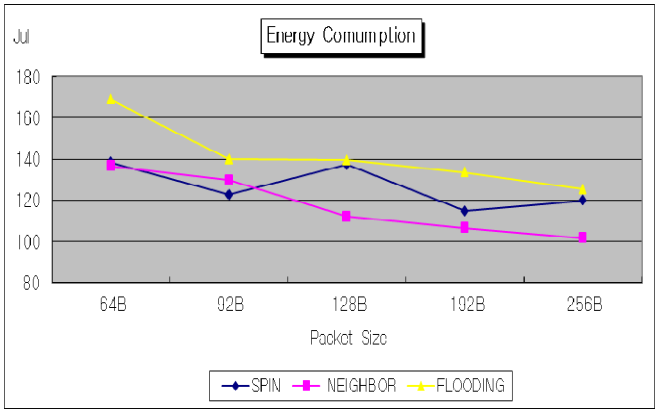


Figure 4. Energy Consumption

4.4 Upgrade Time

Total time to upgrade software is estimated as 51.2 sec by (5). This value is the minimum value. But in the simulations, some additional time are considered such as error recovery time, message wait time and so on. Figure 5 shows the time to upgrade software to all nodes. When the packet size is 256 bytes, the upgrade time is 102sec. When the packet sizes are between 128 bytes and 256 bytes, the upgrade times are between 135sec and 105 sec.

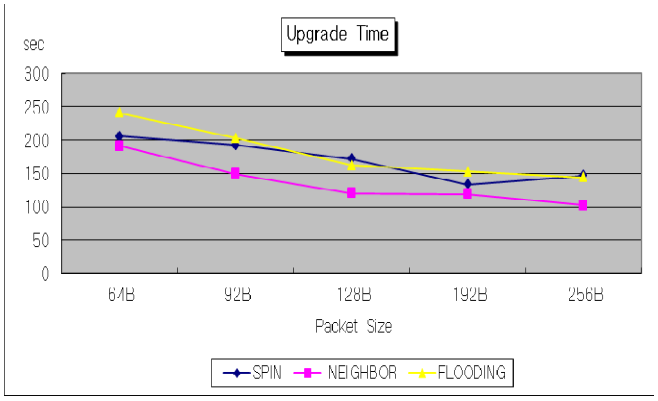


Figure 5. Upgrade Time

5. Conclusions

In this paper, three software update methods are simulated using NS-2 and their results are compared. Their performances are measured by the number of relay nodes, data loss, energy consumption and upgrade time and are analyzed by the packet sizes. Simulation results show that the energy consumption of the proposed method is less than 50% that those of other methods. The update time of the proposed method is 33% faster than others. 256 bytes of the packet size shows best performance in upgrade time, energy consumption and data loss rate.

Please address any questions related to this paper to Prof. Ahn by Email (b.ahn@yu.ac.kr).

6. References

[1] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," Proc. of the 6th annual international conference on Mobile computing and networking(Mobicom '00), pp.56-67, 2000.

[2] Wei Ye, J. Heidemann, and D. Estrin, "Sensor-MAC (S-MAC): Medium Access Control for Wireless Sensor Networks," Proc. of the 21st International Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), vol.3, pp.1567-1576, 2002.

[3] Y. G. Iyer, S. Gandham, and S. Venkatesan, “STCP: A Generic Transport Layer Protocol for Wireless Sensor Networks,” Proc. of 14th International Conference on Computer Communications and Networks, pp.449-454, 2005.

[4] F. Stann, and J. Heidemann, "RMST: Reliable data transport in sensor networks," Proc. of the First IEEE. 2003 IEEE International Workshop on Sensor Network Protocols and Applications, pp. 102 -112, 2003.

[5] W. Chen, P. Chen, W. Lee, and C. Huang, "Design and Implementation of a Real Time Video Surveillance System with Wireless Sensor Networks," Proc. of Vehicular Technology Conference, 2008, pp.218-222. 2008.

[5] Honggang Wang , Dongming Peng , Wei Wang , Hamid Sharif , Hsiao-Hwa Chen , "Image transmissions with security enhancement based on region and path diversity in wireless sensor networks", IEEE Transactions on Wireless Communications, vol. 8, no. 2, pp.757-765, 2009.

[6] C-C. Han, R. Kumar, R. Shea, M. Srivastavam, "Sensor Network Software Update Management: a Survey," Intl. Journal of Network Management, no. 15, No. 4, John Wiley & Sons, pp. 283-294, 2005.

[7] S. Brown and C. Sreenan, "A New Nodel for Updating Softeare in Wireless Sensor Networks," IEEE Network Nov/Dec. pp.42-47, 2006.

[8] S. Brown and C. Screenan, "Software Update Recovery for Wireless Sensor Networks," Proc. of International Conference on Sensor Networks Applications, Experimentation and Logistics (SENSAPPEAL) ICST 2009

[9] Y. Kwon, Ph.D. Dissertation, Yeungnam University, 2011

[10] S. Hedetniemi, A. Liestman, A survey of gossiping and broadcasting in communication networks, Networks 18 (4) (1988) 319–49.

[11] W. Heinzelman, J. Kulik, H. Balakrishnan, Adaptive protocols for information dissemination in wireless sensor networks, in: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom99), Seattle, WA, August 1999.

[12] J. Kulik, W.R. Heinzelman, H. Balakrishnan: Negotiation–Based Protocols for Disseminating Information in Wireless Sensor Networks. In: Wireless Networks, Vol 8, pp. 169-185, 2002.

[13] J.N. Al-Karaki, A.E. Kamal: Routing Techniques in Wireless Sensor Networks: A Survey. In: IEEE Wireless Communications, Vol. 11, no. 6, pp. 6-28, 2004

[14] Zeenat Rehena1, Krishanu Kumar2, Sarbani Roy2, Nandini Mukherjee2 : SPIN Implementation in TinyOS Environment using nesC. In : IEEE, 2010

[15] Zeenat Rehena, Sarbani Roy, Nandini Mukherjee : A Modified SPIN for Wireless Sensor Networks. In: IEEE, 2011

SESSION

NOVEL APPLICATIONS AND RELATED ISSUES

Chair(s)

TBA

Utilizing Automated Robots to Recalibrate WiFi Fingerprint Maps for Indoor Location Estimation

Reza Farivar†, Victor Wu‡, Ellick Chan*, Roy H. Campbell†
†University of Illinois at Urbana-Champaign
‡BMO Harris Bank
*Stanford University
{farivar2,rhc}@illinois.edu, wu.victor@gmail.com, emchan@stanford.edu

***Abstract*—WiFi fingerprinting is a localization technology that leverages existing wireless infrastructures in indoor settings. Although fingerprints can be calibrated to a high degree of accuracy, the precision of such maps changes due to environment changes. We propose a system that automatically re-calibrates the WiFi fingerprint maps using a mobile robot. As the robot traverses the building, it encounters RFID anchor tags embedded in the floor that helps the robot to determine its exact location. At these locations, the robot samples a WiFi fingerprint and associates it with the RFID tag location in a location database. Using this technique, we are able to overcome difficulties in fingerprinting calibration including human error and lack of samples. With this system we are able to attain twice the accuracy of state of the art systems: from 10 feet of accuracy to 3.5 feet.**

I. Introduction

Advances in mobile technologies have resulted in a proliferation of location-based services. These services often require a location determination mechanism to provide accurate real-time estimates of users' locations. While GPS (global positioning system) has emerged as the dominant technology for outdoor environments, multiple solutions exist (without a dominant winner) for indoor settings. 802.11 WiFi location estimation systems (in particular WiFi fingerprinting[1]) leverage existing WiFi infrastructures in a building, providing cost-effective systems that require minimal additional hardware.

In most practical WiFi fingerprinting systems such as [1], [2], [3], [4], [5], [9], a WiFi fingerprint map is created by measuring the received signal strengths of all WiFi access points in a building during an off-line training phase. A fingerprint consists of a location coordinate and a vector of observed signal strengths. To estimate the unknown location of a user during the on-line phase, the user's WiFi enabled mobile device scans the WiFi radio channels at the unknown location and creates a vector of radio signal strengths. The device then estimates its location by finding the closest vector in the fingerprint map to this vector (e.g. by using Euclidean distance). The location associated with the closest vector is the location estimate.

A major practical hurdle in WiFi fingerprinting systems that is often overlooked in practice and previous literature is the pre-deployment calibration of the system, which is typically a labor intensive manual process. Usually a human operator has to cover the building step by step, and it each location enter their location and record a WiFi fingerprint reading. The calibration of such systems, being such a tedious manual task, is usually justified as a one-time process. However, since wireless channel propagation characteristics are continually changing due to physical changes in the environment (even minor changes such as moving the furniture around), and deployment of new access points or removal of old ones, WiFi fingerprint maps are quickly rendered stale. Previous methods to address this problem include creating multiple maps [2] and introducing additional transceivers in the physical space to refine the map [5].

In this paper, we propose a more direct approach to solving this problem. We propose eliminating the need for a human operator to record the WiFi fingerprints, and instead use autonomous robots to periodically update the WiFi fingerprint maps. The robots can move throughout the building and collect received signal strength vectors throughout the building. To determine its own location (remember that each fingerprint vector requires a location associated to it), a robot scans passive RFID tags embedded in the flooring or underneath the carpets at strategic locations. The tags provide absolute location values as "anchor points". The robot then uses dead reckoning navigation methods between the tags to approximate its location from the closest scanned RFID tag.

Being passive devices, the lifetime of RFID tags are at least 20 years [7], making them suitable to embed in a building at desired locations, as a true one-time phase.

In this project, we have used an "iRobot Create" robot and "Touchatag RFID" kit to implement our system on top of the existing WiFi infrastructure in the Computer science building of our University. Our results show an average accuracy of 3.5 feet in our 802.11 localization system with an RFID tag deployment grid of 7 feet, and with a simple dead reckoning algorithm in our robot.

The rest of the paper is organized as follows. We first provide literature related to WiFi fingerprinting, RFID localization technologies, and robot area coverage (specifically for iRobot robots). Next, we describe our system architecture. We then evaluate our system through experiments and discuss the results, and finally we conclude and provide future work.

II. Related Work and Background

The system reported in this paper is based on a few pre-existing technologies and ideas. This section intends to give a brief summary on each, while presenting the published related work in the literature.

A. WiFi Fingerprinting

WiFi fingerprinting is well-researched, and various techniques have been employed to increase the location estimation

[1]In this paper, we focus on WiFi fingerprinting. Triangulation [3] is another well-studied WiFi location estimation technique. However, due to strong multi-path effects in indoor settings, it often results in poor location estimates.

accuracy, [21], [22]. Our WiFi location estimation system described in the next section provides an overview of such a system. The majority of WiFi fingerprinting literature focus on the on-line phase of estimating a user's location using a previously created fingerprint map. From our own experience in deploying a wireless localization system in several buildings in a 6 month span [19], [20], generating and updating the wireless map is a non-trivial problem, and should not be overlooked. WiFi signal strength in an office environment change over time because new access points are introduced in or removed from the environment and furniture gets moved. In many buildings the wireless channel propagation characteristics are continually changing (due to continually changing environments and infrastructure), WiFi fingerprint maps are quickly rendered stale, leading to inaccurate location estimates mere weeks after the map creation. Creating the wireless maps is a tedious manual task; requiring humans to constantly re-calibrate the map becomes a tedious solution that ultimately is not feasible. This becomes a larger problem, especially if there are a large number of predetermined locations. In [2], the authors propose a solution of creating multiple maps, reflecting different possible channel conditions. Access points compare the received signal strengths of transmissions from each other to choose the best map to use in the on-line phase. This is updated as channel conditions vary. However, a chosen map may not accurately approximate the channel conditions at a given time, since the environments change drastically and in an unpredictable manner. In [5], the authors introduce additional transmitters and receivers in the physical space at well-known fixed locations. The transmitters periodically broadcast packets. The receivers listen for these packets as well as mobile users' packets to refine the fingerprint map when estimating the location of users. Like [5], our solution also introduces additional hardware. However, we provide a more direct approach by frequently updating the fingerprint map.

B. RFID Technologies

RFID technology is used in many localization technologies. For example, the authors in [23] use a mobile robot to determine the locations of fixed tags in space. This creates an RFID map that the robot uses to localize itself, as well as to track the movements of other mobile objects. Similarly, in [10], the authors embed tags in the floor, similar to our set up. A small mobile vehicle equipped with an RFID reader moves through the field of tags randomly. As it scans the tags, it creates a map of the field. Other mobile vehicles can use this map to navigate through the tag field. In [24], the authors store the location information in tags. They also store meta-data such as the building and neighborhood. This information is shown in a display upon tag interrogation on a hand-held device.

C. iRobot Roomba and Create

Mobile robots have been subject of intense interest in the research community for quite some time. As a mobile robot platform, we selected the iRobot Create [12], a research friendly version of the iRobot Roomba [14] product. By default, these robots take a purely heuristic approach to coverage planning. Motivated by cost-constraints and simplicity in design, these robots are able to use a basic set of behaviors, such as spiraling and wall following, to cover the required area, without the need for complex sensors or other large computational resources [17]. [15] and [16] introduce the "backtracking spiral algorithm" (BSA) for robot coverage, and a hybrid version of the it is used in iRobot Roombas [18]. Spirals are a typical method of coverage since a robot does not need to have a lot of information about its current orientation to move.

The downside of the default algorithm is the time taken to cover a given area. In section III-D we provide an analysis showing how spiral trajectories can be used to trade-off coverage time with coverage area. Moreover, these algorithms are designed to ensure the robot covers a given area, but they do not provide information about the location of the robot within that area. In our system, we need to know the location of the robot at each moment on top of covering the area, therefore we use a dead-reckoning algorithm, described in section III-D instead off relying on the default coverage algorithm of iRobot, with experimental results shown in section IV-A.

III. System Architecture Description

Our system provides an inexpensive way to automatically maintain up to date the fundamental information required by many WiFi in-door localization systems: the map of the signal strength in an area. It integrates RFID tags and WiFi fingerprinting in an autonomous mobile system that collects WiFi signal strength surveys.

The system is composed of readily available commercial components: an iRobot Create mobile robot unit, a Touchatag RFID reader with the respective RFID tags and an iPod Touch. The mobile robotic platform is equipped with an RFID reader in charge of reading a set of RFID tags embedded in known fixed positions of a building. The robot moves in the area where it is supposed to survey. Every time the reader detects an RFID tag, the system uses the iPod WiFi radio to record a set of samples of the WiFi signal strength at that known position. This information is integrated with previous information to maintain the WiFi signal strength map up-to-date after the environment changes, with movement of furniture or relocation of access points.

The next sections describe more in detail each component of our system: the WiFi location estimation mechanism, the RFID system and the iRobot Create robotic platform.

A. WiFi Location Estimation

The system acts as a support for an indoor localization system that provides location-aware services to users in a building. The WiFi location estimation mechanism employed by our system is composed in two phases. The first phase is an off-line calibration; during this period of time the map of the WiFi fingerprinting of the building is created or updated. This phase is performed autonomously by the mobile robot. The second phase is an on-line phase, representing the normal usage of the system: users can use the system to locate themselves within the environment and to obtain access to location-based services through their 802.11 enabled mobile handsets, such as Apple iPhone, Blackberry or Google Android devices. In our implementation, we have used an Apple iPod touch (3rd generation).

During the off-line calibration phase we create multiple WiFi fingerprint maps of the building. Initially, passive RFID tags are distributed throughout the survey area. They can be embedded beneath carpet or placed on top of the tiles. The location of each tag is stored in the server as a true one-time human-involved

calibration phase. Thereafter, during each calibration phase, a robot equipped with RFID readers and a network connected mobile device traverses through the building. At each RFID tag and in between them, the robot records a WiFi fingerprint for the mobile device mounted on the robot. A WiFi fingerprint is a record containing: the physical location of the robot in the building (retrieved using the RFID tag) when the fingerprint is taken; the mobile device identifier, and a set of vectors of Received Signal Strength Indication (RSSI) of all the access points in the building visible in that point. If the robot is not within range of an access point, the associated signal strength is assigned a value of zero. To compensate for transient wireless noises and the mobile device reader noise, a set of signal strength vectors (8 consecutive samples in our implementation) are collected by the robot at each point and are filtered using a Kalman filter [21]. The results are finally submitted to a central fingerprint storage server. The central server stores all the samples and makes them available to other users.

In the on-line phase, the location of a mobile user is estimated using the WiFi fingerprint map. The mechanism is implemented at the client side. The mobile device scans the access points and forms a vector of received signal strengths, using a value of zero for access points beyond its communication range. This vector is then correlated with the fingerprint map associated with the device, and the wireless signal space distance to each pre-stored fingerprint is calculated. Next, the 3 nearest neighboring fingerprints in the signal space are selected, and their physical location is averaged to estimate the user's location. Again, we use Kalman filtering in the on-line phase to reduce the noise effects. Due to the interactivity requirements of the on-line system, we use fewer samples to perform the Kalman filtering. In our current implementation, we chose to perform the Kalman filter process with 3 samples.

The research community may question why mobile users should not interrogate tags directly to determine location information. We note that this is not a viable solution due to hardware constraints. Dedicated hand-held RFID readers are expensive and not ready for consumer consumption, while many smart-phones are already equipped with 802.11 WiFi capabilities, and are seamlessly deployed worldwide. Including an RFID reader in existing mobile devices is not cost effective either. Specifically, the mobile industry is already finding it difficult to include increasingly more radios (cellular, GPS, WiFi, Bluetooth, FM) in devices, without raising costs and significantly changing the form factors of the devices themselves. Similarly UHF (ultra high frequency) RFID readers are expensive and not easily miniaturizable to include them as part of conventional mobile devices. And even though near field communications (NFC) RFID readers are beginning to appear in smart-phones manufactured by Samsung, Nokia and Motorola [11], NFC read range is at most on the order of 10s of centimeters, out of the range of regular human users standing or walking.

B. Touchatag RFID System

In our implementation, we use the Touchatag [8] passive RFID system to re-calibrate the WiFi fingerprint map, a popular consumer-oriented RFID solution. The RFID system operates at 13.56 MHz (high frequency), and are physically small. A reader measures 9.7 cm x 6.4 cm and a tag is 3.1 cm in diameter, small enough to be embedded on the bottom of a robot. The read range is a few centimeters, according to our experiments. These characteristics make Touchatag the ideal choice for our implementation. We removed the casing from the three readers and mounted the circuit boards directly on the underside of the robot. This provides sufficient range for the readers to read tags placed on the floor. While too small of a range would render our system non-functional; too large of a range would cause interference problems as the Create would scan tags in its immediate vicinity, thus confusing the robot, especially around a high spatial tag density.

Initially the readers were unable to scan tags quickly enough, since in our experiments the robot passes over the tags at approximately 300 millimeters per second [13]. We solved this problem by decreasing the polling period of readers in the Touchatag software from 500 milliseconds to 10 milliseconds.

C. iRobot Create System

In our implementation, we use the iRobot Create [12] as our mobile robot platform to re-calibrate the WiFi fingerprint map. The Create is similar to the iRobot Roomba [14]. It is manufactured specifically for the research community, enabling them to develop customizable programs for it.

We affixed a plexiglass two-level platform on top the of the chassis. On the bottom level, we place a netbook computer (Everex Cloudbook, Via C7-M processor). On the top level, we place an Apple iPod touch. Three Touchatag readers (stripped of their casings) are mounted on the underside of the chassis. We use three readers to increase the probability of a successful tag read as the robot passes over a tag.

D. Robot Coverage Algorithm

An important problem to consider when designing the system is the density of RFID tags that need to be deployed in an environment like a building. Even though RFID tags are inexpensive, the manual deployment and calibration of the RFID tags is still a labor intensive task, even though it needs to be performed only once. Therefore, the question of how and in which pattern the RFID tags are deployed becomes an important topic to consider. Furthermore, the robot's movement pattern and algorithm should be designed in conjunction with the tag deployment pattern to optimize both the robot trajectory time as well as the coverage of the building. Through careful deployment of a few RFID tags relative to the target area, and dead reckoning navigating in between tags, we were able to create a virtual grid of finer granularity than just using the RFID tags deployment grid, as we will describe below. Increasing the granularity decreases the error in location estimates.

As shown in Figure 1, the RFID tags are deployed in two interlinked grids. The tags in each grid are spaced 84 inches (7 feet) apart in each direction. This density of tags is enough to cover an area of 25 feet by 50 feet (1250 square feet) with only 28 tags. One naive option for the robot trajectory is random motion. However, [18]states that random movement takes 5 times as long as a deterministic movement pattern to cover 98% of a typical room without obstacles. We choose to program the robot to move in a hybrid zigzag and spiral pattern. Figure 1 shows a detailed view of the pattern our robot uses to cover the survey area. The robot uses a simple motif to cover the area. Figure 2 shows one motif in detail. The robot moves on a snake like path in between two RFID tags, using dead reckoning to locate itself. During the motif pattern, it makes

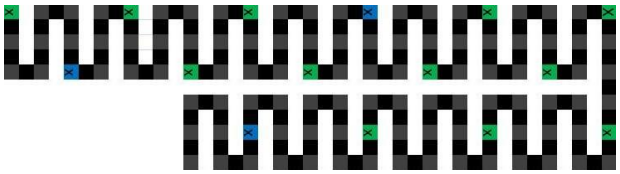


Fig. 1. Tags are distributed in two interlinked grids. The robot follows the zig-zag path.

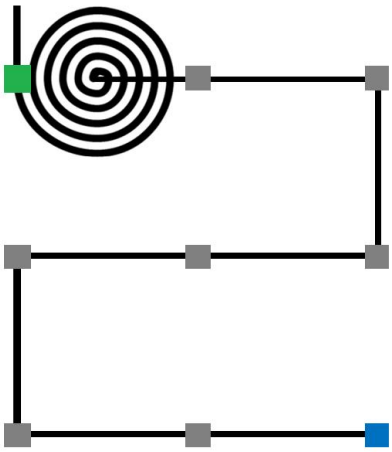


Fig. 2. A single motif in the robot path (starting from bottom right point). After finishing the zig-zag part of the motif, the robot goes into spiral mode to find the next tag.

frequent stops to take WiFi samples and associate them with its location from dead reckoning. As is common among mobile robots, dead reckoning is not an accurate location estimation measure, especially after extended periods of relying on it. This is due to the accumulated error. To compensate for this, as the robot nears the end of each motif, it changes its operating mode to a spiral mode and searches for the next RFID tag, which should be close to where it almost finished the motif. By using a "tighter" spiral, the robot is able to find the next tag with higher probability, since area coverage is increased. However, a tighter spiral increases the time taken by the robot. The following simple analysis quantifies this trade-off.

One practical implication of our hybrid navigation algorithm, consisting of a snake like motion and then a spiral to find the next RFID tag, is that when a tag is found, the robot could be facing any arbitrary direction. To continue to the next motif, the robot needs to re-align itself towards magnetic North after finding a RFID tag in a spiral. In our design, we used a Google Android G1 phone as a networked magnetometer, since the Apple iPod touch device doesn't employ a magnetometer. It should be noted that magnetometers can be found in many available smart-phones, including recent Apple iPhones.

When moving forward, the robot has an effective lateral length of c, where c is the interrogation footprint of the "Touchatag" readers attached on the bottom of the robot. In other words, if the robot moves a distance of l, it covers an area of lc. Figure 3 shows the dotted path of the Archimedean spiral (a spiral with constant separation distance between successive turns) with trajectory of the the robot (specifically, the center

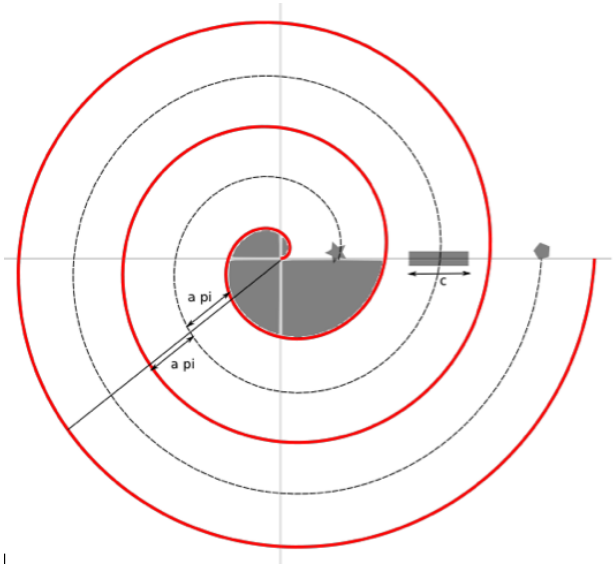


Fig. 3. Archimedean spiral.

of the robot). This is represented by the parametric formula

$$r = a\theta + a\pi, \tag{1}$$

where $\theta \in [0, \hat{\theta}]$ represents the evolution of the trajectory, $2a\pi$ is the distance between successive turns, and r is the radial distance from the coordinate system origin. In the figure, the star is the beginning of the trajectory and the pentagon is the end. (In this specific case, $\hat{\theta} = 4\pi$.) To avoid the robot covering area more than once, we require $2a\pi \geq c$, or

$$a \geq \frac{c}{2\pi}. \tag{2}$$

For ease of exposition, we ignore the darkened area at the center. (It can be shown that this area is $\frac{4}{3}a^2\pi^2$, which can easily be incorporated in the following analysis, if it represents a significant fraction of the coverage area.) If $a > \frac{c}{2\pi}$, some area is left uncovered by the robot. If $a = \frac{c}{2\pi}$, all the area is covered by the robot. Specifically, suppose A is the area of the physical space the robot wishes to cover with its spiral. Then A is $2a\pi$ multiplied by the length of the dotted spiral. That is,

$$A = 2a\pi \int_0^{\hat{\theta}} r d\theta = 2a\pi \left(\frac{1}{2} a\hat{\theta}^2 + a\pi\hat{\theta} \right). \tag{3}$$

The fraction of the area A that the robot actually covers is

$$f = \frac{c\left(\frac{1}{2}a\hat{\theta}^2 + a\pi\hat{\theta}\right)}{2a\pi\left(\frac{1}{2}a\hat{\theta}^2 + a\pi\hat{\theta}\right)} = \frac{c}{2a\pi}. \tag{4}$$

Assuming the robot moves with speed s, the trajectory time, t, is the length of the dotted spiral divided by s. Using (3), we have

$$t = \frac{A}{s2a\pi}. \tag{5}$$

Therefore, we have a tradeoff. If we decrease a (i.e. tighten the spiral, down to at most $a = \frac{c}{2\pi}$), we increase the fraction

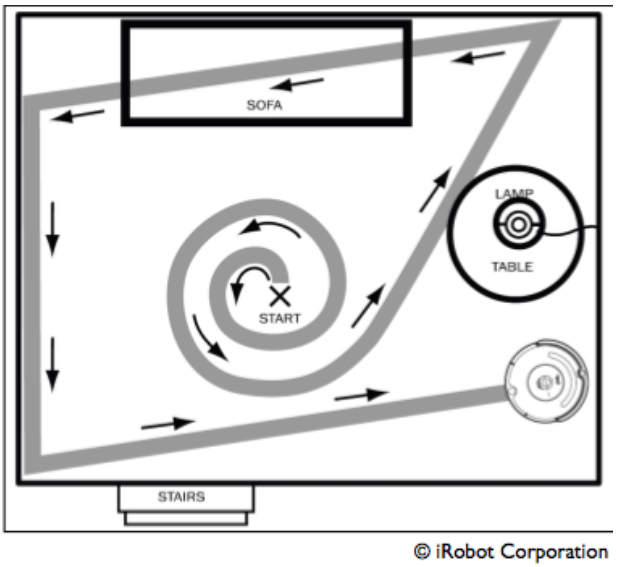


Fig. 4. The built-in navigation of the iRobot Roomba/Create.

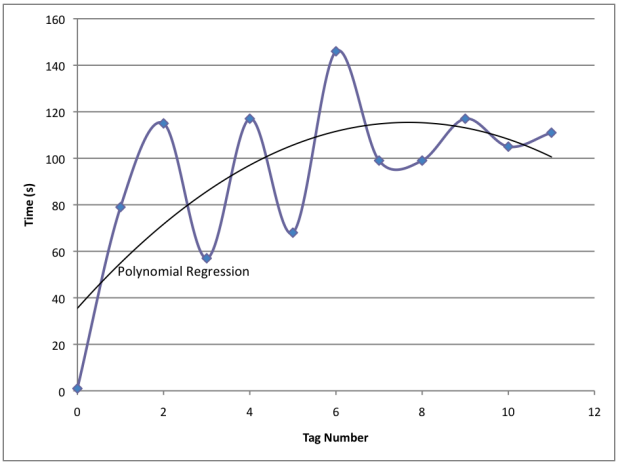


Fig. 5. Differential Acquisition Time: Run 1.

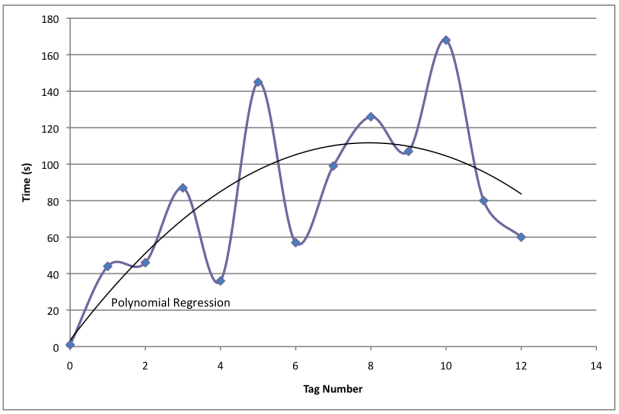


Fig. 6. Differential Acquisition Time: Run 2.

of area covered, but increase the trajectory time. Specifically, if we require the entire area to be covered, the trajectory time would be $t_{entire} = \frac{A}{sc}$.

IV. Experimental Evaluation

Two experiments were performed to evaluate the system. The first experiment was designed to discover the quality and usefulness of the "iRobot Create's" in-built navigation algorithm.The second experiment uses a custom navigation algorithm, designed as a combination of a simple movement pattern and a spiral movement for dead reckoning navigation in between multiple RFID tags. This experiment is meant to show and assess the feasibility of using a robot to take WiFi fingerprint samples.

A. Time to Locate a Tag Using the Built-in Navigation Algorithm vs our hybrid scheme

The aim of this first experiment was to determine how much time it would take for the built-in navigation algorithm of the "iRobot Create" to find all the RFID tags embedded in a test area. We set up a small test field of 10 feet by 8 feet. The robot was set to find 20 tags arranged in a grid of 2 feet granularity. The built-in navigation algorithm of the iRobot Roomba family (with the "Create" being a member) has two phases, as depicted in Figure 4. In the first phase, the robot starts from its initial position in an outward counter-clockwise spiral pattern for about 5 turns, and then enters the second phase. In the second phase, the robot moves in straight trajectories, and "bounces" off walls when it hits one.

The built-in algorithm was modified to stop the robot when it detects a tag (tags are continuously scanned), allowing ample time for the iPod Touch to take 8 WiFi signal strength samples, calculate a Kalman filtered vector and send them to a server. Afterward, the robot goes into spiral mode and then straight trajectory mode (as above) until it detects the next tag. (Note that the robot may detect the next tag even before the spiral mode ends.) To prevent multiple readings of a tag, the robot remembers what tags it has read so far. If this measure is not implemented, the same tag is often read multiple times during spiral mode. Figures 5 and 6 chart the differential acquisition time required for each tag to be found in two subsequent runs of the experiment. Even though the underlying data in both graphs have large variance, the polynomial regressions of both of the charts are quite similar. Looking at these trends, it is apparent that the time to find the next tag increases as the number of remaining tags in the area decreases. This trend continues up to when 8 tags have been read. From there, the trend reverses in both experiments. This is because with fewer tags remaining (smaller spatial tag density), the built-in algorithm of the robot gets a chance to move into its second phase, where the robot moves much faster than the first phase.

Table IV-A shows the amount of time it takes in average to find tags in our hybrid scheme, introduced in section III-D. The first row shows the amount of time in average that the robot takes to find the next tag starting on top of another tag. To find this number, we ran the experiment 10 times and computed the average. The second row shows the time it took to find 10 tags using this algorithm. It should be noted that during this experiment, the robot was instructed to only move according to the coverage pattern of section III-D, and it would not stop

TABLE I. Acquisition Time for our hybrid pattern

Average time to find the next RFID tag	8.4s
Time to find 10 tags	91s

to gather WiFi fingerprints.

Looking at the second row of table III-D, we note that the time taken to find 10 tags is slightly smaller than the time it takes the default algorithm. It should be noted that in our scheme the acquisition time is linear based on the number of tags to find, while the it is harder to predict how much time the default algorithm would take. But it should be noted that the main reason behind our custom coverage pattern is not faster coverage time, but the ability to estimate the robot's location using dead reckoning, and that the semi-random nature of the default algorithm would make this task impossible.

B. Fingerprinting Using the Mobile Robot

Recall from the design section that our robot is programmed to cover an area with RFID tags spread in a double interlinked grid pattern, in a snake-like motion motif, with a spiral movement at the end of the motif. During each motif, the robot makes frequent stops, as shown in Figure 2, to take additional WiFi fingerprints. The robot uses dead reckoning to navigate in between RFID tags, which serve as anchor points in the movement.

In this experiment, we evaluate the accuracy of the WiFi location estimation system, using the training fingerprints gathered by the robot. The robot measures the WiFi fingerprint at each tag and associates it with the location of the tag. It also uses dead reckoning to locate itself between tags, and therefore, measure additional WiFi fingerprints at these non-tagged locations, in an attempt to increase the accuracy of our system. It is important to point out that the dead reckoning navigation has some errors in determining the location the robot, especially in a simple method of only relying on sending movement commands to the robot. Furthermore, this experiment was performed in an area covered with slate blocks, forcing the robot to periodically slip on the slate surfaces. However, we observed that the robot would typically end up less than 1 foot of where it was programmed to finish its zigzag motion, and the largest drift observed was less than 2 feet. The second phase of the robot's algorithm, where it spirals to find the target RFID tag, would consistently bring it to the correct starting point.

One implication of the dead reckoning error is that the WiFi fingerprints obtained at the dead reckoning locations (the non-tagged locations) will have location error, typically less than one foot. This is acceptable, since the location estimation error of our WiFi fingerprinting system is about 3 feet on average. As such, the dead reckoning location error is within about 30% of the WiFi estimation error, and therefore within practical engineering tolerances. Thus, the effect of the dead reckoning error on the error of the WiFi location estimation is safely covered. This is a practical compromise to achieve a finer granularity grid of fingerprint samples than what we could achieve if we only wanted to rely on the RFID tags locations for fingerprints.

We validate our system using cross evaluation of the fingerprint data sets. To evaluate the accuracy of the system, one sample is randomly taken out from the set of fingerprints of each run of the experiment, and is compared to other samples

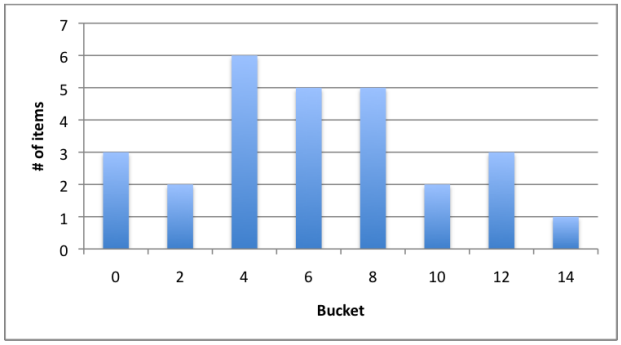


Fig. 7. Histogram of location errors without dead reckoning. The x-axis is the error in feet. The y-axis is the frequency.

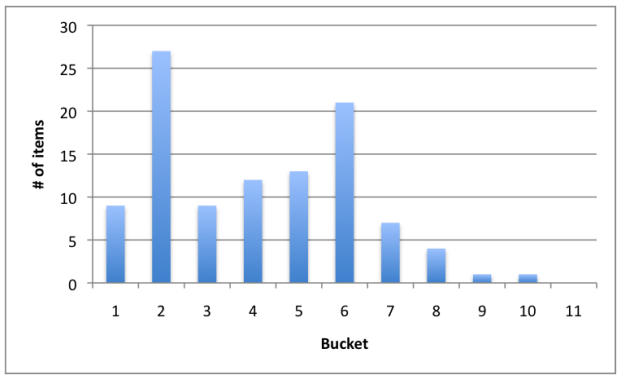


Fig. 8. Histogram of location errors with additional samples from dead reckoning. The x-axis is the error in feet. The y-axis is the frequency.

to estimate a location for it. Then, this estimated location is compared with the real location of the sample, and the distance is reported as error. This procedure is randomly repeated for all of the samples, and the results are used to calculate the average error, minimum and maximum errors.

Two runs of the experiment were performed in this stage. In the first run, fingerprint samples were taken only on the RFID tag locations, which, as mentioned earlier are spread on two interlinked grids where each cell of the grid is 84 inches (7 feet) in each direction. These points are shown on figure 1 with an "X". There were a total of 28 samples taken in this run. Results of this run of the experiment are shown in a histogram of location errors, as depicted in Figure 7. The mean of error distance in this case was 7.17 feet, with the maximum observed error being about 12 feet. It is notable that from the results, the average error is roughly proportional to the distance between the sampling points on the grid, which is 7 feet. The standard deviation of the distance errors in the first run was equal to 3.79 feet.

In the second run of this experiment, the robot was programmed to stop at points in between the RFID tags and take WiFi fingerprint samples. Results of this run of the experiment are shown in a histogram of location errors, as depicted in Figure 8. The distance error is now much smaller, with a mean of 3.52 feet and a standard deviation of 2.27 feet. This is an interesting result, since the size of the finer grid created by dead reckoning in this experiment run is 3.5 feet apart, showing again that the average error in the accuracy of the system is

proportional to the granularity of the sampling grid. This fact alone shows the necessity of having a finer sampling grid, which we have achieved by navigating the robot using dead reckoning.

The graph also shows that almost 34% of the trials in the cross evaluation have an error of less than 2 feet, and 95% of the trials have an error less than 7 feet. As mentioned earlier, the average error in the estimated location is 3.5 feet.

V. Conclusions, Discussion and Future Works

Although WiFi fingerprinting is a powerful and versatile technique for mapping large swathes of physical spaces, the process requires significant effort on the part of an engineering team to calibrate points on the map to wireless fingerprints. In this paper, we have presented a technique that automates this process to a great extent. By using RFID tags to anchor WiFi calibration points, an autonomous robot can cover vast expanses of space effectively.

Our current research is primarily concerned with the operation of a single robot in a closed space. Future considerations include the use of multiple robots to cover the area of a building in a shorter time duration and the use of robot memory where the previous calibration results and geometry of the room are considered. In this case, the robot would leverage existing calibration information to test hypotheses regarding its location. If the hypotheses are accurate, the robot makes fewer measurements and assumes that the underlying WiFi fingerprint remains largely unchanged.

Finally, by integrating more ambient sensors, such as wind flow, temperature and sound, our robot can be used to detect various conditions in different environments for scientific experiments and location-based services.

References

[1] P. Bahl and V. N. Padmanabhan, "RADAR: an In-building RF-based User Location and Tracking System," in *Proc. 2000 IEEE International Conference on Computer Communication*, Tel Aviv, Israel, Mar. 2000, vol. 2, pp. 775-784.

[2] P. Bahl, A. Balachandran, and V. N. Padmanabhan, "Enhancements to the RADAR User Location and Tracking system," *Technical Report*, Feb. 2000.

[3] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, "On Indoor Position Location with Wireless LANs," in *Proc. 2002 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Pavilhão Atlântico, Portugal, Sep. 2002, vol. 2, pp. 720-724.

[4] K. Kaemarungsi and P. Krishnamurthy, "Modeling of Indoor Positioning Systems Based on Location Fingerprinting," in *Proc. 2004 IEEE International Conference on Computer Communication*, Hong Kong, China, Mar. 2004, vol. 2, pp. 1012-1022.

[5] P. Krishnan, A. S. Krishnakumar, W.-H. Ju, C. Mallows, and S.N. Gamt, "A System for LEASE: Location Estimation Assisted by Stationary Emitters for Indoor RF Wireless Networks," in *Proc. 2004 IEEE International Conference on Computer Communication*, Hong Kong, China, Mar. 2004, vol. 2, pp. 1001-1011.

[6] B. Li, A. G. Dempster, C. Rizos, and J. Barnes, "Hybrid Method for Localization using WiFi," in *Proc. 2005 Spatial Sciences Institute Biennial Conference*, Melbourne, Australia, Sep. 2005, pp. 341-350.

[7] "Passive RFID Tag (or Passive Tag)," *Technovelgy*. Available:http://www.technovelgy.com/ct/Technology-Article.asp?ArtNum=47

[8] *Touchatag*. Available: http://www.touchatag.com

[9] B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor Positioning Techniques Based on Wireless LAN," in *Proc. 2006 IEEE Australian Conference on Wireless Broadband and Ultra Wideband Communications*, Sydney, Australia, Mar. 2006.

[10] J. Bohn and F. Mattern,"Super-distributed RFID Tag Infrastructures," *Lecture Notes in Computer Science*, vol. 3295, pp. 1-12, 2004.

[11] "NFC Research: Devices," *Near Field Communication Research Lab*. Available: http://www.nfc-research.at/index.php?id=45

[12] "iRobot Create," *iRobot*. Available: http://store.irobot.com/product/index.jsp?productId=2586252&cp=2591511&

[13] "iRobot Create Owner's Guide," *iRobot*. Available: http://www.irobot.com/filelibrary/create/Create Manual_Final.pdf

[14] "iRobot Roomba," *iRobot*. Available: http://store.irobot.com/shop/index.jsp?categoryId=2804605

[15] E. Gonzalez, M. Alarcon, P. Aristizabal, and C. Parra, "BSA: A Coverage Algorithm," *Proc. 2003 IEEE International Conference on Intelligent Robots and Systems*, Las Vegas, NV, Oct. 2003, vol. 2, pp. 1679-1684.

[16] E. Gonzalez, O. Alvarez, Y. Diaz, Carlos Parra, and C. Bustacara, "BSA: A Complete Coverage Algorithm," *Proc. 2005 IEEE International Conference on Robotics and Automation*, Barcelona, Spain, Apr. 2005, pp. 2040-2044.

[17] J. L. Jones, "Robots at the Tipping point: the Road to iRobot Roomba," *IEEE Robotics and Automation Magazine*, vol. 13, issue 1, pp. 76-68, Mar. 2006.

[18] J. L. Jones and P. R. Mass, "Method and System for Multi-mode Coverage for an Autonomous Robot," US Patent 6809490, to iRobot Corp., Oct. 26, 2004.

[19] R. Farivar, D. Wiczer, A. Gutierrez, and R. H. Campbell, "A Statistical Study on the Impact of Wireless Signals' Behavior on Location Estimation Accuracy in 802.11 Fingerprinting Systems," in *Proc. PMEO-UCNS'2009 Workshop*, in conjunction with the *23rd International Parallel and Distributed Processing Symposium (IPDPS 09)*

[20] R. Farivar, M. Montanari, E. Chan, and R. H. Campbell, "An Automatic User Study Demo in Indoor Environments and its Privacy Implications," in *Proc. 2009 IEEE International Conference on Pervasive Computing and Communications*, Galveston, TX, Mar. 2009.

[21] J. Yim, C. Park, J. Joo, and S. Jeong, "Extended Kalman Filter for Wireless LAN Based Indoor Positioning," *Decision Support Systems*, vol. 45, issue 4, pp. 960-971, Nov. 2008.

[22] F. Linaker and M. Ishikawa, "Real-time Appearance-based Monte Carlo Localization," *Robotics and Autonomous Systems*, vol. 54, issue 3, pp. 205-220, Mar. 2006.

[23] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, M. Philipose, "Mapping and Localization with RFID Technology," in *2004 IEEE International Conference on Robotics and Automation*, New Orleans, LA, May-Apr. 2004, vol. 1, pp. 1015-1020.

[24] A. Loffer, U. Wissendheit, H. Gerhauser, and D. Kuznetsova, "GIDS - A System for Combining RFID-based Site Information and Web-based Data for Virtually Displaying the Location on Handheld Devices ," in *2008 IEEE International Conference on RFID*, Las Vegas, NV, Apr. 2008.

hymnMark: Towards Efficient Digital Watermarking on Android Smartphones

Nai Miao*, Yutao He**, Jane Dong*

*Department of Electrical and Computer Engineering
California State University, Los Angeles
5151 State University Drive, Los Angeles, CA 90032 USA
**Jet Propulsion Laboratory/California Institute of Technology
4800 Oak Grove Drive, Pasadena, California 91109
miaonai1229@gmail.com,Yutao.He@jpl.nasa.gov,jdong2@calstatela.edu

ABSTRACT

Fast growth in ubiquitous use of digital-camera-equipped smartphones in our daily life has generated large amount of multimedia data such as images, audio, and video clips that need to be processed, stored, and transmitted on battery-powered mobile devices. Yet little research has been done to protect those multimedia data on smartphone platforms. This paper presents design and implementation of an efficient digital watermarking application, called hymnMark, to perform watermark embedding and detection for digital images on the Android platform. Preliminary evaluation shows that hymnMark can successfully embed in color images different types of watermarks with good resistance to noise as well as a number of digital signal processing attacks, in the meantime entail low power consumption.

Keywords: Digital Watermarking, Power efficiency, Smartphone, Android, Discrete Cosine Transform.

1. INTRODUCTION

With the widespread of mobile networks, smartphone applications become more and more popular in recent years. The high mobility of the smartphones makes them ideal end platforms for multimedia applications such as web video, image browsing, photo sharing, etc. Since these digital media are highly subject to attacks including content modification, it is critical to better protect data integrity.

Digital watermarking is an effective technology to achieve authentication, copyright protection, and integrity of multimedia data and has been extensively studied in the past decades [1]. However, digital watermarking algorithms are computation-intensive and power hungry. How to develop an efficient digital watermarking application on resource-constrained mobile devices like smartphones requires a decent balance between performance and power consumption yet it has received little research attention. The research described in this paper aims to tackle the problem by developing a robust and power-efficient digital watermarking application targeted to smartphone platforms.

In our research, first, characteristics of wide range of digital watermarking algorithms have been investigated in the context of the energy-constrained Android smartphone platform. As a result of the study this set has been down-selected to one algorithm with optimal power efficiency for smartphone platforms. Second, the selected algorithm has been further optimized to reduce the computation cost with respect to Android computing environment. As a proof-of-concept, it is then implemented in an Android app called *hymnMark*. It features a user-friendly GUI to allow easy watermark generation, embedding and detection in Android-powered smartphones. Third, comprehensive empirical evaluation has been conducted to measure the algorithm's performance (i.e., robustness against various attacks) and power consumption.

In summary, our research makes the following main contributions:

1. We have developed the first Android app that performs digital watermarking completely on the smartphone. Our research has shown that effectiveness of a digital watermark algorithm for smartphones not only depends on its performance but also its power efficiency, since it will limit its sustained performance to protecting multimedia data.
2. We have identified a set of practical optimization techniques that proves to be effective in the smartphone environment. We believe that they can be applied to other digital watermarking applications on smartphones.
3. We have developed a micro-level instrumentation methodology that allows measurement of power consumptions inside one application. It enables fine-grained power profiling which in turns helps pinpoint the *hotspot* of one application for further optimization.

The rest of the paper is organized as below. Section 2 describes the related work. Section 3 provides an overview of digital watermarking technologies and the hardware configuration of Android environment. Section 4 presents in details the design of the hymnMark system. Complete

evaluation and results are given in Section 5, followed by conclusions and future in Section 6.

2. RELATED WORK

The computation complexity of watermarking algorithms varies significantly with different embedding approaches. However, to achieve good resistance to noise, compression and other signal processing attacks, a common practice is to embed the watermark in the transform domain [1]. Some robust watermarking approaches also require spread spectrum analysis. Hence, the computational cost can be fairly high, which will also lead to high power consumption. Among the existing efforts of developing good watermarking system on low power devices, Arun Kejariwal at el made valuable contributions via evaluating a number of existing watermarking approaches in embedded environment. Their research provided a good perspective of the power consumption of various watermarking algorithms [2]. According to their experimental results, the Koch and Bruyn approach has the lowest power consumption and shortest execution time, especially for host images with high resolutions.

Takao Nakamura [3] described a fast and robust watermarking detection scheme on cellular phones. However, it only worked with 16-bit watermark and images with resolution 288*352. In 2011, J. Jeedella and H. Al-Ahmad [4] at Khalifa University of Science, Technology & Research proposed an algorithm for watermarking mobile phone color images using BCH code. This algorithm demonstrated good robust level through benchmark tests for attacks and the watermarked image had high PSNR. However, this method required the watermark to be in the format of numbers. Particularly, the implemented algorithm utilized cell-phone numbers as the watermark.

It is worthwhile to mention two available applications on Android platform for watermark detection, namely Digital Space [5] and Digimarc Discover [6]. These two applications are very similar and allow the user to hold the camera mounted on the smart-phone about 5-7" away from the image until cell-phone "bee" to finish the detection. After "Bee", the application will tell the user whether there is a watermark in the image. Users of these applications need to register online, embed watermarks into images, and save them in their accounts. Only watermark detection on the registered images is performed on a smart-phone.

3. BACKGROUND INFORMATION

3.1 Overview of Digital Watermarking

Digital watermarking is the process of embedding information into a digital signal, like audio, image, and video, which can be detected for authentication and identification. The embedded watermark can be number, characters, image, or any other identification information [1].

Digital watermarking systems can be categorized into different types. In terms of perceptibility, there are visible and imperceptible digital watermark. Since invisible watermarks are typically used for authentication and data integration, we only consider this type of watermarks in our research. In terms of robustness, there are three types of digital watermark, namely robust, semi-fragile, and fragile watermark. Robust watermark are widely used for copyright protection, while the other two are used for data integrity and authentication. Specifically, due to its ability to detect attacks as well as its good resistance to channel noise and compression, semi-fragile watermark has become a desirable approach for authentication.

The embedding process of digital watermark also varies a lot. In general, the watermark can be embedded in spatial domain, transform domain, or both. Embedding approaches involving transform domain analysis usually provide better resistance against compression. DCT (discrete cosine transformation) DWT (discrete wavelet transformation) are two widely used transformations in watermark embedding. Both have their own advantages and disadvantages. Since DCT is used in compression standards such as JPEG and MPEG, DCT domain embedding offers significant convenience for JPEG images, while the multi-resolution nature of DWT offers good means for spread-spectrum analysis and thus enhance the robustness of the embedded watermark, or provides support to localize the regions being attacked.

3.2 Smart-phone configuration requirements and constraints

To design a digital watermarking application on an Android smart-phone platform, it is important to take into accounts its hardware constraints. Table 1 lists features of three different Android smart-phones. Specifically, the screen resolution, the processor power, the memory size and the power efficiency are critical in a design. Ideally, a developed watermarking system on an Android platform should be fast and responsive, power efficient, uses less memory space, and provides seamless user experience to achieve the target security functions. To meet the above design goal, an appropriate watermarking algorithm needs to be selected that for implementation under hardware constraints listed in Table 1 [7].

4. DESIGN OF HYMNMARK

4.1 Watermarking algorithm selection

The first step of our research is to study existing watermarking algorithms and identify suitable algorithms with good performance-computation balance that can be implemented in low power devices. In comparison with characteristics of a number of algorithms, Koch's algorithm [8] has been selected as the baseline algorithms, due to its

lowest power consumption, shortest execution time, and greater robustness against common attacks.

The embedding process of Koch's algorithm can be described briefly as follows: First, DCT transformation is applied to the entire image. The next step is to generate position sequence that maps the watermark bit to the pixel locations. Next, Randomly Sequenced Pulse Position Modulated Code (RSPPMC) [8] will be embedded into the locations in image blocks that are selected in the first two steps. Lastly, inverse DCT and de-quantization is applied to the embedded blocks. Among all the steps, the largest computation power is spent in the RSPPMC embedding part.

There are some limitations of Koch's algorithm. First, it is non-blind watermark algorithm, that is, watermark cannot be detected without side information. In particular, detection process of Koch's algorithm needs a key file to indicate the location sequence and watermark length in order to detect watermark. Secondly, Koch's algorithm does not support multi-resolution images because of nature of the DCT transformation. Thirdly, the watermark length is also limited. Since only one bit is embedded into one 8*8 block of an image, the watermark length is bounded by the number of 8x8 blocks in a host image. As a result, modifications are required to further improve its performance. To the best of our knowledge, our work is the first effort to implement and test Koch's algorithm in Android platform. Therefore, design and implementation of the proposed watermarking application together with the evaluation results will provide useful insights and guidelines for future research in the area.

	HTC Sensation 4G	Samsung Galaxy S II	HTC EVO 3D
Android version	Gingerbread (2.3)	Gingerbread (2.3)	Gingerbread (2.3)
Skin	Sense 3.0	TouchWiz 4.0	Sense 3.0
US carrier	T-Mobile	TBD	Sprint
Display	4.3-inch Super LCD	4.3-inch Super AMOLED Plus	4.3-inch Super LCD with glasses-free 3D
Resolution	540 x 960	480 x 800	540 x 960
Dual-core processor	1.2GHz Qualcomm MSM8260	1.2GHz Samsung Exynos 4210*	1.2GHz Qualcomm MSM8660
RAM	768MB	1GB	1GB
Storage	1GB internal with 8GB MicroSD card	16GB or 32GB internal with MicroSD slot	4GB internal with MicroSD slot**
Front camera	VGA	2 megapixel	1.3 megapixel
Rear camera	8 megapixel,dual LED flash	8 megapixel,LED flash	2x 5 megapixel, 3D images and video
Video	1080p at 30fps	1080p at 30fps	1080p at 24fps (2D), 720p at 30fps (3D)
4G internet	14.4Mbps HSPA+	21.1Mbps HSPA+	WiMAX
Accelerometer	Yes	Yes	Yes
Gyroscope	Yes	Yes	Yes
Battery	1520mAh	1650mAh	1730mAh
Thickness	11.3mm	8.49mm	12.1mm
Weight	148g / 5.22 ounces	116g / 4.09 ounces	170g / 6 ounces

Table.1 Three Typical Smart-phone Hardware Configurations

4.2 hymnMark Architecture

The hymnMark conceptual flow-chart is shown in Figure.2. In the top level hymnMark includes two processes: watermark embedding and detection. The embedding process includes four major functions: 1) import image; 2) select watermark; 3) RSPPMC embedding; and 4) save/export watermarked image. In our current implementation, host images can only be imported from local memory. A watermarked image can be saved to a file on the SDCard of a smart-phone. The detection process includes four essential functions: 1) import image; 2) import a key file; 3) watermark detection; 4) display the retrieved watermark.

4.3 Implementation

4.3.1 Overview

hymnMark is implemented as an app on Android operating system using JAVA with Android Plug-In. It has been tested on a real Android Phone NEXUS One with Android system 2.3.3. Its detailed configuration is: Android SDK 2.2, Eclipse jdk-6u24-linux-x64, eclipse-jee-helios-SR2-linux-gtk-x86_64, and Ubuntu 10.10 Linux.

4.3.2 GUI Front-End

hymnMark features a GUI-based front-end. It uses the popular MVC (Model-View-Control) framework for GUI applications. In this framework, user interface and the models, which are usually called "classes" in Java, are separately designed to cooperate with each other through controllers. Generally, an "XML" layout is a view interacting with users; a "Java" class is a model that will be called by a controller when needed; an "activity" is a "View and Controller", so as a "View", it corresponds with an XML-layout as a "controller".

4.3.3 Digital Watermark Kernel

Details of the watermark embedding and detection processes of Koch's algorithm are described as follows.

4.3.3.1 Embedding Process:

The embedding process takes three inputs, namely *Host image*, *selected Watermark*, and *JPEG Quality factor*, and produced the watermarked image. After the host image is loaded, color space transformation will be applied such that the image will be represented in YUV color space instead of RGB. In hymnMark, the digital watermark is only embedded in Y component for a true color host image.

After color space transform, block-based DCT is performed followed by JPEG-alike quantization to each 8x8 block of DCT coefficients. To embed RSPPMC, two DCT coefficients will be randomly selected in the low-middle frequency range per block while the MSE of the original block and embedded block meets the minimum requirement, which is $1e^{-3}$.

$$MSE = \frac{1}{mn}\sum_{y=1}^{m}\sum_{x=1}^{n}[I(x,y) - I'(x,y)]^2$$

where *I(x,y)* and *I'(x,y)* are the Y component values at the index of (x,y) of the original and the embedded images, respectively.

After the embedding process, de-quantization and inverse DCT are conducted to each block. Then, the embedded blocks are multiplexed to create the Y layer. YUV to RGB color space conversion will be conducted to get the watermarked image.

4.3.3.2 Detection Process:

The detection process takes two inputs, *host image* and *key*. Like the embedding process, the detection process also requires color space transformation, DCT transformation, and JPEG-like quantization prior to actual detection.. Based on the key stored in a file, essential information for watermark detection, such as location sequence, watermark length, watermark type, and quality factor, is obtained. The selected DC coefficients blocks and the coefficients in each block can be recovered in order of the "seed" from the key file. Then, inverse RSPPMC is conducted to each pair of coefficients in each block; the process is repeated until the full length of watermark is recovered. The result of inverse RSPPMC is a bitstream that is further converted to a text or an image with respect to the watermark type. The detected watermark can also be saved in a new file on the SD Card of a smart-phone.

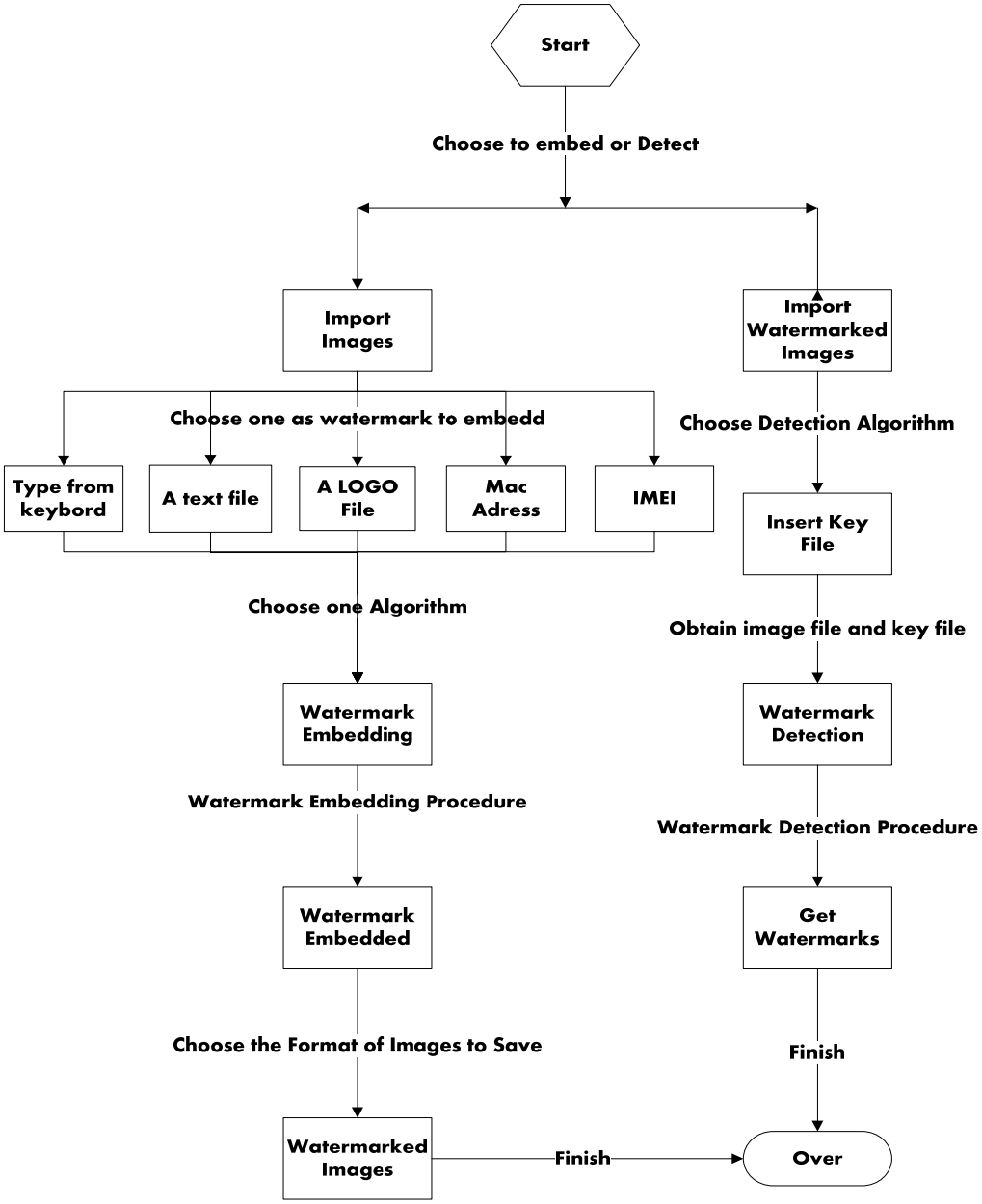
4.4 Optimizations to the Koch Algorithm

In order to reduce the computation cost, save power consumption, and to accommodate color images, the following modifications have been made in the implementation.

First, the order of DCT transformation and the selection of image blocks has been swapped. In the original Koch's algorithm, DCT transformation is applied to the entire picture, which is unnecessary since we only embed the watermark bits in a subset of DCT blocks. A disadvantage of the original Koch's algorithm is that if the image is big, the DCT transformation will cause large amount of computation. By switching the order, we only need to apply the DCT transformation to the selected 8*8 blocks. As a result, computation cost is reduced, and so does the power consumption.

Secondly, our implementation narrows down the range of pixel selection for watermark embedding within each block. Watermark bits are supposed to be embedded into the low-middle frequency of DC coefficients. In an 8*8 block, the random selection range could be shrank to {3,4,5,10,11,12,14,17,18,19,20,24,25,26,27,28,32,33,34,35, 40,41} instead of {0-63}. In the optimized implementation, 34.38% of computational power for DC coefficients selection is saved compared to the original Koch's Algorithm.

Thirdly, we add an image size adjustment approach after a host image is read from SD Card to avoid the out-of-memory (OOM) problem that is common for Android applications. To work with the memory constraints for an Android application (16MB), hymnMark automatically reduces the image size based on the device screen. For example, Android Nexus One's screen size is 600*480; then a host image of 2096*2096 will be resized to 480*480. In this way, we are able to control the memory usage of the application while handling multiple images



Figue.2 hymnMark Application User Flow-Chart

In addition to reducing computation, we have modified the algorithm to support color images. In particular, to achieve good perceptual performance, color space conversion is performed first and the watermark is embedded in Y components of a color image.

5 RESULTS AND ANALYSIS

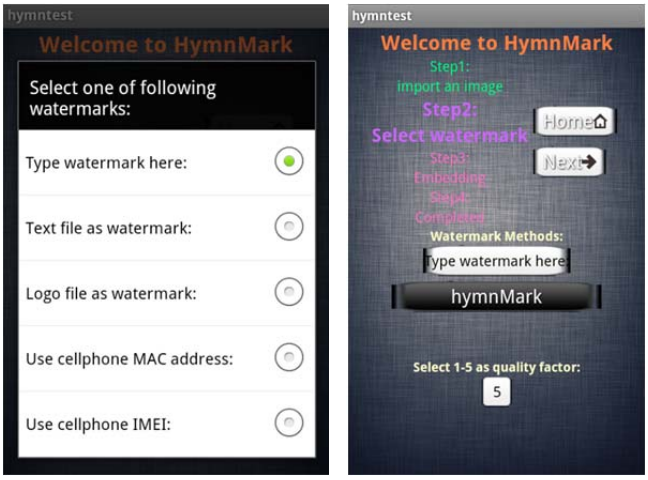
5. 1 User Interface

hymnMark has a user-friendly GUI. Figures 3 to 6 illustrate the key steps of using hymnMark to embed and detect watermark on a smart-phone.

As shown in Figure 3a, a user can click ImageView to import a host image. Five types of watermark are supported in hymnMark: plain text, a text file, a logo image file, the MAC address of the smart-phone, and the IMEI number of the smart-phone. A user can select the preferred watermark types through the GUI. In addition, a user can input quality factor (in range of 1 to 5) to indicate the watermarking robustness. Figure 3b shows that a user types text message *hymnMar*k as the watermark and selects 5 as the quality factor (the most robust). Figure.4a shows the embedding progress and the watermarked image after the embedding process is completed. If a user is satisfied with the

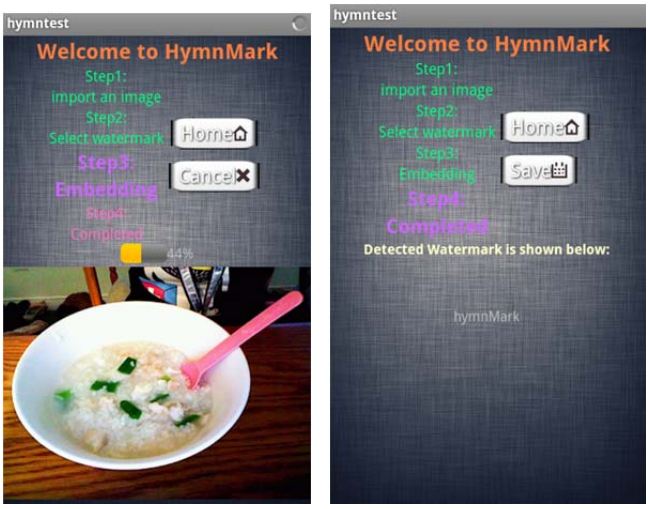
watermarked image , the user can click on "save" to save it onto the SD Card.

From the main page, if a user chooses "detect", a detection page will show up. The embedded image will be displayed in the "ImageView" and the application will ask for a key file in order to detect the watermark in the image. After the key file is loaded, the detection process will start. Figure 4b displays the detected watermark. If the watermarked image has been severely damaged, the detection may not be successful, and the detected watermark may contain errors or may be illegible.



(a) (b)

Figure.3 hymnMark Interface illustration: a) Five types of watermark message Selection; b) Typing Message as Watermark



(a) (b)

Figure.4 hymnMark Interface illustration :a) Embedding Page; b) Detected watermark shown page

5.2 Performance Analysis

To evaluate the performance of the hymnMark app with respect to quality of watermarked images, a number of tests have been conducted. Table 2 lists the quality analysis results for host images of different sizes. It is obvious that if the watermark is short (1 bit), the impact is small and the resulted watermarked image has higher quality. Quality factor indicates the robustness of the watermark. To make the watermark more robust, it needs to be embedded into the DCT coefficients of lower frequency, which will consequently have more impact on the image quality. Our test results show that even for quality factor of 5, the watermarked image still has excellent quality (PSNR > 35 dB).

Host Image Size	Watermark Length	Quality Factor	PSNR
32kb	1b	1	50.55
32kb	1b	2	50.42
32kb	1b	3	50.27
32kb	1b	4	50.01
32kb	1b	5	49.97
200kb	3.1kb	1	37.47
200kb	3.1kb	2	37.33
200kb	3.1kb	3	37.11
200kb	3.1kb	4	37.03
200kb	3.1kb	5	36.87

Table.2 PSNR for Various Quality Factors

5.3 Robustness Analysis through Multiple Attacks

In [9], Johnson C. Lee analyzed the attacks on common watermark techniques. Following his analysis, we have evaluated the performance of hymnMark system under some common attacks including rotation, cropping, scaling, mosaic, Gaussian, contrast, chrominance, luminance and compression. All the attacks are executed through Adobe Photoshop CS4 version 11.0.

Table 3 summarizes the results of the robustness tests. In the table, *robust range* means that among the specified parameter settings, the watermark can be detected correctly. Take compression as an example, the max setting in Photoshop for users to modify is 0-12, which 0 stands for the worst quality, while 12 means the best quality. The watermarked images are tested in the Photoshop and be detected for the watermark message. When the compression range is in 4 to 12, the watermark can be detected completely. Therefore, its correspondent robust range is 4 to 12. The testing results show that hymnMark has no resistance towards attacks such as rotation, cropping and scaling, but has fairly good resistance against contrast change and compression.

Attacks:	Robust Range	Max setting in PS
Rotation	None	0 to 360
Cropping	None	Any
Scaling	None	Any
Mosaic	None	2 to 200
Gaussian	0.0 to 5	0.0 to 250
Luminance	-25 to 25	-50 to 150
Chrominance	-10 to 10	-180 to 180
Contrast	-50 to 50	-50 to 100
Compression	4 to 12	0 to 12

Table.3 Robustness Analysis

5.4 Power Consumption Analysis

The power consumption of hymnMark can be measured at two different levels, *macro-level* and *micro-level*. Macro-level measurement shows total power consumption of hymnMark compared to other applications on the smart-phone. Micro-level measurement, on the other hand, provides a close-up view of power profiling of hymnMark, which is capable of showing the hotspot of hymnMark, that is, where it consumes most power.

5.4.1 Smart-phone power model:

Efficient energy management requires good understanding of where and how power is consumed, including how much the whole system uses and how much each component uses. In [12], Aaron Carroll and Gernot Heiser tested energy consumption of CPU/RAM, screen display, GSM (Global System for Mobile Communication, originally from Group Special Mobile), flash storage, network and GPS through different applications. The results in Table.4 show that the majority power consumption is used in GSM module and screen display. In these experiments, it is easy to see that brightness of display is the most significant factor that affects the power consumption of a mobile device. , followed by the CPU power consumption.

On the other hand, smart-phones are considered as personal portable computers nowadays and their users expect fast-responsive time of apps. For example, people who get lost want to find a right direction fast when they ask for help from a map application on smart-phone.

	Power (% of total)							Battery life
Workload	GSM	CPU	RAM	Graphics	LCD	Backlight	Rest	[hours]
Suspend	45	19	4	13	1	0	19	49
Casual	47	16	4	12	2	3	16	40
Regular	44	14	4	14	4	7	13	27
Business	51	11	3	11	4	11	10	21
PMD	31	19	5	17	6	6	14	29

Table.4 Daily energy usage and battery life under a number of usage patterns [12]

5.4.2 Power Measurement Methodology

The most popular power consumption measurement tools are PowerTutor and PowerProfile. PowerTutor is an Android app working on Google phones that calculates the power consumption of CPU, display, Wi-Fi, and user applications running on the platform. To access the power consumption measurement, it uses Android inner resources such as:

- *android.content.Context;*
- *com.android.internal.util.XmlUtils;*
- *org.xmlpull.v1.XmlPullParser.*

However, PowerTutor and PowerProfile only calculate the power usage based on components not within each application. As a result, they fail to provide fine-grained measurement and insights on power consumptions of functions such as *read-image*, *read-watermark*, color space conversion, block selection, *DCT transformation*, *quantization*, and *embedding/detection*, *de-quantization*, *iDCT transformation*, *inverse color space conversion*, *store images/watermark.*.

Research described in [13] has developed a fine-grained power measurement tool called *Eprof* but it is not available to the community. As a result, we have developed the micro-level power consumption analysis method based on Android EXTRA_LEVEL and EXTRA_SCALE APIs EXTRA_LEVEL measures the current power level, and EXTRA_SCALE measure the maximum level of the smart-phone. The methodology details are described as follows:

1. Set up a test project to evaluate "EXRTA_LEVEL" and "EXTRA_SCALE" variables, and verify their functions;.
2. Once the verification is passed, the micro-level power consumption analysis should be conducted following the steps illustrated in Figure 9.

The current power level of the smart-phone is measured before and after the execution of each code block. This allows us to measure the power consumption of each part of the application; and we can also study the impact of application parameters (such as the quality factor) on the power usage. Therefore, the micro-level power consumption analysis is very useful to optimize the implementation of each part of the application under power constraint.

5.4.3 Power consumption of hymnMark

To analyze the power consumption of hymnMark in the macro-level, "PowerTutor"[10] is utilized. During our tests, we have found that the total power of the fully charged cell-phone, a Nexus one, was 1144mAh (4.12 Volt). The watermark embedding and detection process have been measured through continuous execution. On average, the embedding process consumes about 66.7mW power, and the detection process consumes 38.4mW. Hence, the average of the entire hymnMark is (66.7+38.4)/2=52.55mW.

Using the same experiment setting, we have also measured an Android default browser's power consumption, which is around 282mW. In comparison, our developed hymnMark system consumed less power than a regular web browsing application.

Currently we are in the process of tuning the micro-level instrumentation, and we hope to report the preliminary results of the micro-level power analysis for hymnMark during the conference presentation.

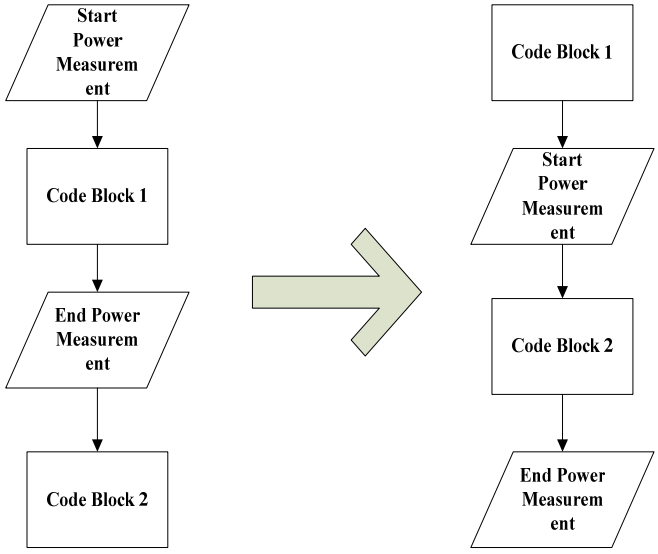


Figure.9 Procedure of Micro-level power consumption analysis

6. CONCLUSION AND FUTURE WORK

This paper presents the first Androi SmartPhone watermark app. The core of the system is the watermark embedding and detection processes based on Koch's algorithm. Optimizations have been made to reduce the computation cost and power consumption on the Android SmartPhone platform. Comprehensive testing has been conducted to evaluate quality, robustness, and power consumption of the implementation. Experimental results demonstrate that the watermarked images have excellent visual quality; and the power consumption is lower than a web browser app on a smart-phone platform. In the future, we will further optimize the algorithm to reduce the power consumption and execution time. More power consumption analysis will be conducted to gauge the energy efficiency of each internal function block.

References

[1]. Watermarking digital image and video data, A-State-of-the-Art Overview, Gerhard C. Langelaar, Iwan Setyawan, and Reginald L. Lagendijk, IEEE image processing magazine, 2000

[2]. Energy Analysis of Multimedia Watermarking on Mobile Handheld Devices, Arun Kejariwal, Sumit Gupta, Alexandru Nicolau, Nikil Dutt, Rajesh Gupta, School of Information and Computer Science Tensilica Inc. Dept. of Computer Science and Engineering University of California at Irvine, IEEE Conference Publications,2005

[3]. A Fast and Robust Digital Watermark Detection Scheme for Cellular Phones, Takao Nakamura, Atsushi Katayama, Ryo Kitahara, and Kenji Nakazawa, NTT Cyber Space Laboratories Yokosuka-shi, 239-0847 Japan, 2006

[4]. An Algorithm For Watermarking Mobile Phone Color Images Using BCH Code, J. Jeedella and H. Al-Ahmad, Khalifa University of Science, Technology & Research, IEEE Conference Publications, 2011

[5]. Digital Space, https://play.google.com/store/apps/developer?id=Digital+Space

[6]. DigiMarc Discover, https://play.google.com/store/search?q=digimarc+discover&c=apps

[7]. Android System Smart-Phone Hardware Configuration. http://www.engadget.com/2011/04/15/htc-sensation-versus-the-rest-of-the-dual-core-world-smartphone/

[8]. Towards Robust and Hidden Image Copyright Labeling, E. Koch & J. Zhao, Fraunhofer Institute for Computer Graphics Wilhelminenstr. 7, 64283 Darmstadt, Germany, Proc. of 1995 IEEE Workshop on Nonlinear Signal and Image Processing (Neos Marmaras, Greece, June 20-22, 1995)

[9]. Analysis of Attacks on Common Watermarking Techniques, Johnson C. Lee, Student Member, IEEE Electrical and Computer Engineering Department University of British Columbia 2356 Main Mall, Vancouver, BC Canada V6T 1Z4, IEEE Conference Publications, 2011

[10]. Power Tutor, http://ziyang.eecs.umich.edu/projects/powertutor/

[11]. How Fast Is Your Mobile App? Gomez Knows, Charles Babcock, InformationWeek November 04, 2011.

[12]. An Analysis of Power Consumption in a Smartphone, Aaron Carroll, NICTA and University of New South Wales, Gernot Heiser NICTA, 2010 USENIX Annual Technical Conference, 2010.

[13] Abhinav Pathak, Y. Charlie Hu, and Ming Zhang, *Where is the energy spent inside my app? Fine Grained Energy Accounting on Smartphones with Eprof* , Eurosys 2012.

Distributed Collaboration for Effective Cognitive Radio Networks Implementation

Obeten O. Ekabua
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(obeten.ekabua@nwu.ac.za)

Nnenna C. Eric-Nwonye
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(23989696@nwu.ac.za)

Abstract: *Cognitive radios are radios that easily adapt to their operating environment making it possible for unlicensed users to temporarily access, or make use of a licensed frequency spectrum when the licensed user is not using it. This therefore calls for methods that will ensure that the unlicensed user's activities or transmissions, does not interrupt or degrade the licensed user's activities. The ability to reliably identify idle frequency bands is a challenge to individual radios due to the random nature of mobile networks. Radios therefore need to collaborate for performance improvement. Reported in this paper is the development of algorithms for collaborative spectrum sensing, relaying and neighbor selection amongst cognitive radio networks. Included also, is the design of a distributed collaboration strategy for spectrum sensing, through the formation of cognitive coalitions for users to autonomously collaborate and self-organize into independent coalitions, in order to maximize their detection capability. Additionally, we introduce proactive networking protocols such as Destination-Sequenced Distance Vector (DSDV) and Optimized Link State Routing (OLSR), for the radios' communication, relaying and neighbor selection. Finally, for the proposed collaboration strategies to be effective, protocols for auto configuration, spectrum coordination and management were introduced, so that the collaborating radios can conduct their activities in a conflict- free, fair and organized manner.*

1. Introduction

Cognitive radios (CR) are radios or wireless devices capable of sensing and reacting to its operating environment by dynamically adapting itself for good application and network performance. CR improves spectrum efficiency, spectrum utilization, and make spectrum available to new technologies by operating on unused spectrum channels in their local neighborhoods without disrupting the operations of licensed users [1]. The rapid growth and explosion of wireless communications has made the problem of spectrum utilization more critical, creating an ever increasing demand for more radio spectrum [1, 11]. The increasing diversity of communication applications have resulted in overcrowding of the allocated spectrum bands and scarcity of spectrums, resulting to poor service delivery and network problems.

Spectrum is a finite resource which is carefully managed nationally and internationally. This is done by statically allocating spectrum bands for certain users and for delegated use. The downside of static allocation of frequency bands is that it leads to very inefficient use of the spectrum, because many allocated frequency bands are significantly underutilized. This allocation of spectrum bands has created a spectrum shortage that hinders the growth of new wireless applications.

Cognitive radios are the ideal solution to the spectrum scarcity problem because they are radios with the ability to reliably and autonomously identify unused frequency bands. Cognitive radios allow for usage of idle licensed frequency bands by unlicensed (cognitive) users [4]. In order to allow for maximization of spectrum utilization in cognitive radios, it is necessary not to allow unlicensed users to cause interruption or degradation of service to the original license holder. The unlicensed (secondary) users need to monitor the spectrum activities continuously to find a suitable spectrum band for possible utilization and to avoid possible interference to the licensed (primary) users [5, 11].

This spectrum monitoring is mostly done through spectrum sensing, which must be performed before the cognitive users use the licensed spectrum. Spectrum sensing have been identified as a key enabling functionality to ensure that cognitive radios would not interfere with the primary user, by reliably detecting primary user signals. It relies on the secondary systems

to identify free spectrum, through direct sensing of the licensed bands [1, 7].

However, the primary signal may be weak or faded, resulting in degraded signals which will be difficult for individual cognitive radio to detect. The two major sources of degraded signals are multipath and shadowing. For example, a cognitive radio can assume the absence of a primary user if it does not see energy in a particular band, which might be a mistake if the cognitive radio suffers severe shadowing with respect to the primary transmitter. Also, in fading channels, single radio sensing requirements are set by the worst case channel conditions introduced by multipath, shadowing and local interference. These conditions could easily lead to a situation where detection of the primary signal will not be possible [4, 12].

These conditions can be avoided if multiple radios share their individual sensing measurements. Each cognitive radio node has only limited local observation to the whole spectrum due to various constraints, therefore, collaboration among cognitive radio nodes are important for acquiring the complete spectrum information. Cooperative spectrum sensing is conducted among cognitive users so as to detect the primary user accurately.

The presence of multiple radios help to reduce the effects of severe multipath at a single radio since each radio will give an independent result of the same situation. With multiple sensing, the probability that all cognitive radios will see deep fades is extremely low. Cooperation allows radios to achieve spectrum sensing robustness to fading environments without drastic requirements on individual radios.

Cooperative communication techniques with cognitive radios hold the promise of promoting efficient spectrum usage and sharing. Also, collaboration among cognitive radios solves the problems of the main challenges of primary user activity detection, which is an important issue in cognitive radio network implementation [4, 10, 12].

2. Dynamic Configuration for Collaboration

Collaboration among cognitive radio networks is increasingly regarded as a key technology for tackling the challenges of a practical implementation of cognitive radio and also for significant performance improvement. Cognitive radios can collaborate through exchange of information, performing tasks cooperatively, negotiating with peers and using peer information to determine their operating settings.

Presented in this paper, are distributed collaboration approaches and algorithms for collaborative spectrum sensing and signal processing, relaying, and service discovery, using various networking protocols. For spectrum sensing, the radios are organized into groups or coalitions with one of the radios as the cognitive head (CH) of each coalition. The radios are working in a distributed fashion, therefore there is no base station, and instead another cognitive radio will act as the coordination channel for information exchange and organization amongst the different coalitions. The radios determine the occupancy state of the spectrum by comparing it to a set threshold.

This paper also provides a method for information exchange whereby, the cognitive nodes use proactive protocols such as Destination-Sequenced Distance Vector (DSDV) and Optimized Link State Routing (OLSR), for relaying, communication and neighbor selection. The radios communicate with each other, update their status and obtain updated information about other radios, using DSDV which is a table-driven protocol. OLSR protocol is used by the radios for neighbor selection and relaying information.

3. Collaborative Spectrum Sensing and Signal Processing

During spectrum sensing, the cognitive radios form different coalition groups with one of the radios as the coalition head of each coalition. Another cognitive radio also serves as a coordination channel or collection point for the whole coalition. This coordination channel is needed to enable the exchange of information among the coalitions. Different users sense the spectrum and share their sensing results or measurements locally and will cooperatively decide on the occupancy of the spectrum. Each radio's signal and local measurement is collected and sent to the coalition head where it is processed into a decision regarding the occupancy state of the primary band. The local decision is made by comparing the results with a prefixed threshold Y. The decision in turn will be broadcast to all cognitive users in the coalition.

To buttress this point, let us represent spectrum sensing and signal processing as follows:

let;

Y = a set threshold

S = result obtained from sensing the spectrum

if S > Y then Primary User is Present

else

if S < Y then Primary User is Absent

Each radio sends its local decision to the coalition head of their group, the result obtained is Boolean in nature (ie 0 signal is absent, and 1 signal is present), based on the predetermined set threshold Y. If n number of radios combines independent local measurements, then probability of correctly detecting the status of the system increases. The set threshold Y will be used to determine and measure the reliability of the collected results. When the collected signal S*i* exceeds the threshold Y, decision 1 will be made which assumes that the primary user is present; otherwise, decision 0 will be made.

Every cognitive user *i*, for *i* = 1, . . ., N, conducts spectrum sensing individually and collects the energy S*i*. It then reports to the coalition head, a decision D*i*, which is given by

D*i* = 0; where 0 < S*i* < Y

and

D*i* = 1; where S*i* > Y

The coalition head will then make a decision based on the measurements received from each radio in the coalition and send to the coordination channel. Let us assume that the coordination channel receives *k* out of *N* local decisions reported from the coalition heads, it will then make a final decision. The spectrum is assumed to be available only when all the *k* reporting decisions are 0.

Below is an algorithm for distributed collaborative spectrum sensing. This algorithm is used to construct a framework for spectrum sensing.

3.1 Distributed Collaborative Sensing Algorithm

For the proposed distributed collaborative sensing method, let us establish an algorithm as follows:

Initial State: CR = { CR1,. . ., CRN}

CR1 = {CR11,. . ., CR1N}

.

.

.

CRn = {CRn1, . . . ,CRnN}

Where CR1 to CRn are each a coalition made up of n number of cognitive radios. The algorithm is grouped into four stages and different activities takes place during each stage.

Stage 1: Individual Sensing

During individual sensing stage, each individual cognitive radio senses the spectrum and computes its local spectrum sensing observation or result for reporting to the coalition head.

Stage 2: Sensing Result Processing by the Coalition Head

In the second stage, each cognitive radio will report its sensing result to the coalition head, who compiles and processes the signals received from each cognitive radio in the coalition.

The coalition head then makes a tentative Boolean decision of 0 or 1 based on the signals. The signal bit and decision are then sent to the coordination channel

Stage 3: Coordination Channel Decision

During the third stage, the coordination channel receives signal bits and results from each coalition head, compares the results received with the prefixed threshold Y and then make a final Boolean decision of 0 or 1 on the occupancy state of the spectrum

Stage 4: Coordination Channel Broadcast

In stage four, the coordination channel broadcasts the decision to the coalition heads who in turn broadcasts the message to each radio in their coalition. Finally, it updates the spectrum database with the new result.

4. Channel Monitoring

During channel monitoring, each cognitive radio node is equipped with transceivers, which are used for spectrum monitoring and for distributing sensing results over to the coalition head (Fig 1). The spectrum monitor will keep detecting any active signals in the specified frequency range and will also update a signal database continuously. Once the primary user appears, the secondary receiver will pick the best available channel from the spectrum database to continue previous communication.

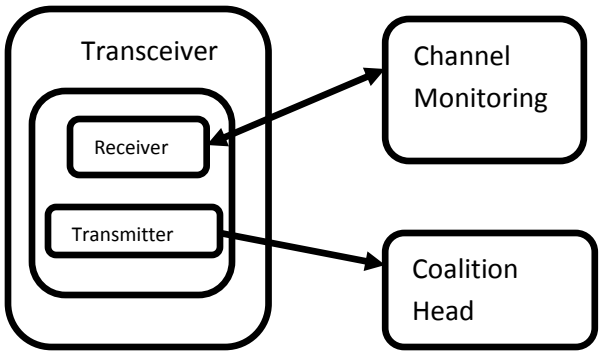


Fig 1: Channel Monitoring

Channel monitoring is done using Energy detection and Cyclostationary detection methods. The radios use an energy detector to measure the energy received on a primary band and use cyclostationary detector to differentiate primary signal from noise and interference. The cognitive radio nodes access the licensed bands in two phases; the sensing phase and the transmission phase. In the sensing phase, a cognitive radio node senses the channel using its receiver and then exchanges sensed channel information with other cognitive radio nodes and the coalition head using the transmitter.

5. Neighbor Detection and Multipoint Relay (MPR) Selection

For neighbor detection and multipoint relay selection, the radios use proactive protocols such as Destination-Sequenced Distance Vector (DSDV) and Optimized Link State Routing (OLSR) for regular exchange of topology information with other nodes in the group. DSDV is a table driven protocol therefore, each radio node advertises to its neighboring nodes in the coalition group its own routing table and information. This is done periodically and each new update must have a sequence number (time stamp) so that receiving nodes can distinguish old information from current information.

OLSR enables it to work in a completely distributed manner without depending on any central entity and makes it possible for nodes to broadcast messages. The nodes would broadcast 'Hello' messages which are used for neighbor detection. They detect their neighbors through link sensing which is accomplished through periodic emission of 'Hello' messages over the interfaces. Individual nodes can use this topology information to compute next hop destinations for all nodes in the group using shortest hop forwarding path. It then selects its multipoint relays (MPR) based on the one hop node that offers the best routes to the two hop nodes.

The idea of multipoint relays (MPR) is to minimize the overhead of flooding messages in the network by reducing redundant retransmissions in the same region. Each node in the network would select a set of nodes in its symmetric one-hop neighborhood, which may retransmit its messages. This set of selected neighbor nodes is called the MPR set of that node. The neighbors of node A which are not in its MPR set would be able to receive and broadcast messages but would not retransmit broadcast messages received from node A.

Each node selects its MPR set from among its one-hop symmetric neighbors. This set is selected such that it covers all symmetric two-hop nodes in terms of radio range. Each node maintains information about the set of neighbors that have selected it as an MPR. A node will obtain this information from periodic 'Hello' messages received from the neighbors. Upon receipt of this MPR selector information, each node will calculate and update its route to each destination. The objective of MPR selection is for a node to select a subset of its neighbors such that a broadcast message, retransmitted by these selected neighbors, will be received by all nodes two hops away from it (fig 2).

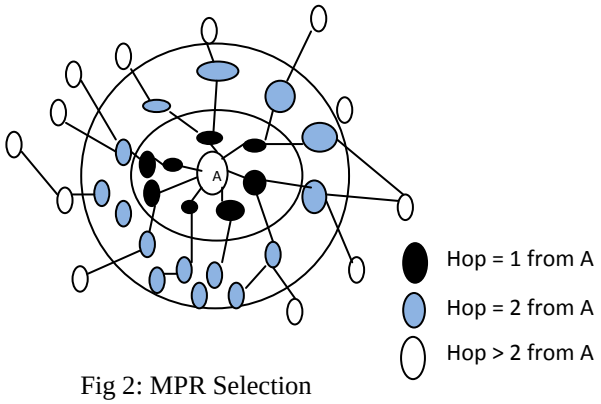


Fig 2: MPR Selection

The 'Hello' messages will include the sender's identifiers and the identifiers of its k-hop neighbors. With this information, a cognitive radio device knows the nodes with which it can directly communicate with over the channel or that it can reach in up to n hops. For the proposed MPR selection, an algorithm is derived below.

5.1 Neighbor Detection and MPR selection Algorithm

For nodes to identify their neighbors and select nodes to relay their messages, we propose an algorithm as follows:

Stage 1: Node Advertisement

Coalition 1 = {CR11, . . . , CR1n}

Coalition 2 = {CR21, . . . , CR2n}

.
.
.

Coalition N = {CRN1, . . .,CRNn}

The activities involved in neighbor detection are also grouped in stages beginning with nodes advertising their routing tables.
In the first stage, a node, 'A' sends information and advertises its routing table with a sequence number added to each advert for other nodes

Stage 2: Receiving Nodes

During the second stage, each neighboring node receives the broadcast and then transmits its response back to Node A

Stage 3: Link Sensing and Neighbor Discovery

In the final stage, Node A processes the responses and identifies its one hop neighbors. Through its one hop neighbors, it identifies its two hop neighbors and finally selects its Multipoint Relay (MPR) node.

6. Cognitive Relay

For relaying signals and messages amongst the radios, each coalition group is created with agreement to forward each other's packets. One node (the relay) forwards the transmission received from another node (the source) towards its destination. Some packets from the primary user will also be delivered by the secondary nodes, which aim at enhancing secondary throughput through the increase of transmission opportunities for the secondary.

Each cognitive radio node is also utilized as a relay node to convey the signal transmitted from the primary user PU to the coalition head CH (fig 3). The coalition head then relay the signals received to the coordination channel. The idea is to utilize relay nodes to convey the signal transmitted from the primary user to the coordination channel, which will make estimation of the presence or absence of primary activities. Given n cognitive relays between the primary user and the coordination channel; the cognitive relays simultaneously receive primary user's signal through independent fading channels and each cognitive relay then amplifies the received primary signal and forward to the cognitive head. The relay node uses one of its transceivers to receive data and uses the other transceiver to forward data to its destination (fig 1).

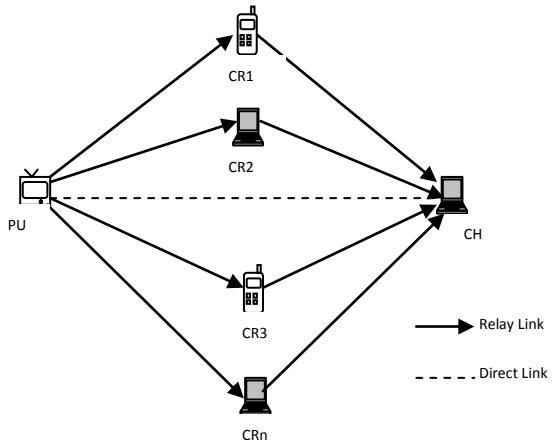


Fig 3: Cooperative Network with Cognitive Relays

7. Service Discovery

During service discovery, the coordination channel broadcasts information about current network status to the cognitive radios. This information is broadcast through table updates and beacons. It also gives information about new radios joining a coalition. The coordination channel also serves as a router between the coalitions and also as router between multiple network layers within each coalition.

A new node joining a coalition listens for beacons and updates on the control channel, to obtain service discovery and connectivity information to initialize its radio parameters. Current nodes will provide performance information in the beacons because as a radio joins a cognitive coalition, connectivity will also be established with the neighboring radios or coalition and a message sent to the neighbors for a service discovery overlay. This allows the new radio to discover how to access all of the various services. The control head then exposes the new radio's identifier to the coalitions through the appropriate name service. When a node receives a new 'join' request or data, it stores the source ID and the sequence number of the packet in its table.

For the purpose of service discovery, the algorithm below is derived.

7.1 Naming and Service Discovery Algorithm

For a node to discover services from the other nodes, we propose an algorithm in two stages as follows:

Let CRnew be a new cognitive node

Stage 1:

During the first stage, CRnew sends a 'join' message and joins a cognitive radio coalition

The coalition head sends a message to the neighboring radios and the coalition heads of the neighboring radios sends the service discovery settings to CRnew through beacons

Stage 2:

On receiving the settings, CRnew starts up

It listens for beacons on the control channel and obtains connectivity information from the beacons and tables

And finally, initializes its radio parameters.

8. Local Coordination and Bargaining

For local coordination and bargaining, each coalition of cognitive radios is also a bargaining group and each group modifies spectrum assignment within the group to improve system utility which is measured in terms of proportional fairness among the cognitive radios. Each coalition also performs local coordination to modify their spectrum usage in order to achieve a conflict free spectrum sharing. The coalitions negotiate spectrum usage through message exchanges and the radios exchange information among themselves locally about how to bargain with other coalitions for mutual gain. The radios or coalitions also need to observe the behaviors of neighboring nodes and adjust spectrum usage according to predefined rules. With these, fair solutions for dynamic spectrum access can then be achieved in a self-organized way.

The protocols that will be necessary to make these proposed collaboration strategies effective are:

9. Auto Configuration Protocols

Auto configuration requires a radio node to be aware of itself, the surrounding nodes and current network status when it starts up. In this collaboration approach, It does this by obtaining reachability and performance information by listening for beacons and table update. The new node also negotiates with existing coalitions for name and service discovery. Dynamic channel assignment method is used for communication among the cognitive radio nodes because; the set of available channels could change over time. The distributed MAC-layer configuration will enable nodes to dynamically discover the network topology and physical location of each node in the network. The nodes would invoke the MAC-layer configuration operation periodically to maintain accuracy despite changes in network topology, changes in channel availability set maintained by individual nodes, and node movements. When a radio is turned on, it will remain silent until the first execution of the MAC-layer

configuration protocol. During the MAC-layer configuration, time is split into intervals and each interval is further divided into time slots of equal length. A node is allowed to transmit during its allocated timeslot in each interval and all other nodes are in receive mode during that time. This ensures that every node in the network gets a chance to transmit without collisions during each interval.

10. Spectrum Coordination and Management Protocols

Spectrum etiquette and coordination policies are implemented and enforced using the coordination channel. Each radio sends information about spectrum usage through beacons and table update so that neighboring cognitive radios or nodes can avoid using the same frequency. Cognitive radio devices will use the coordination channel routing tables to set up links with other nodes. A message cache is also maintained by each node to detect duplicate messages. All radios will use this policy to communicate with the coordination channel. The coordination channel can obtain information about the environment through measurements and information sent or obtained by different radio terminals. It then proffers suggestion or make decisions for efficient coordination.

Control information exchange among users is important. Channel selection and negotiation for data communication will be done by control information exchanging among users. All users use the coordination channel for information and control packet exchanges. Whenever a user wants to initiate communication with any other user or wants to send data, it will negotiate with the intended receiver by exchanging necessary control information on the coordination channels. Using this method, the radios' activities can be conducted in an orderly and conflict-free way.

11. Conclusion

Cognitive radios are fully programmable wireless devices that can sense and dynamically adapt to their operating environment to enhance system performance. Instead of using statically assigned spectrum, cognitive radios can operate on unused spectrum channels in their local neighborhoods without disrupting the operations of existing spectrum owners (primary users). Frequency bands are statically allocated to certain users for specific use, but most of the allocated frequencies are grossly underutilized while at the same time the emergence of new technologies and applications have placed more demand on the frequency spectrum, causing scarcity of frequency. Cognitive radios attempt to alleviate the problem of inefficient utilization of the frequency spectrum by opportunistically using the licensed bands when the licensed user is not using it. Therefore, reported in this paper is the development of distributed collaboration approaches for cognitive radio networks using updating and table driven networking protocols- Destination-Sequenced Distance Vector (DSDV) and Optimized Link State Routing (OLSR).

For these collaboration strategies to be effective, auto configuration, spectrum coordination and management protocols were introduced. Auto configuration protocol makes a radio node to be aware of itself, the surrounding nodes and network status when it starts up. This is achieved by obtaining reachability and performance information from other radio nodes through beacons and table updates. Spectrum coordination and management is implemented using the coordination channel, which proffers suggestions and makes decisions for efficient coordination.

References

[1]. NSF Workshop Report. Future Directions in Cognitive Radio Network (CRN) Research. March 2009

[2]. Federal Communication Commission (FCC). Spectrum Policy Task Force Report. Report ET Docket No 02 – 135, November 2002. www.fcc.gov/oet/info/database/spectrum/

[3]. Federal Communication Commission (FCC). Notice of Proposed Rule Making and Order. ET Docket No 03 – 322, December 2003. www.fcc.gov

[4]. Amir Ghasemi, Elvino S. Sousa. Spectrum Sensing in Cognitive Radio Networks: Requirements, Challenges and Design Trade-offs. IEEE Communications Magazine, April 2008.

[5]. Jun Ma, Goeffrey Ye Li. Signal Processing in Cognitive Radio. Proceedings of the IEEE, Volume 97, No 5, May 2009.

[6]. Joseph Mitola III. Cognitive Radio: An Integrated Agent Architecture for Software Defined Radios (SDR). PHD

Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[7]. Danijela Cabric, Artem Tkachenko, Robert W Brodersen. Experimental Study of Spectrum Sensing Based on Energy Detection and Network Cooperation.

[8]. Cabric D., S.M Mishara, R.W Brodersen. Implementation Issues in Spectrum Sensing. In Proc, 38th Annual Asilomar Conference on Signal, Systems and Computers. Pacific Grove, CA, USA, November 2004. PP. 772 – 776.

[9]. Cabric D., et al. A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum. In Proc. Of 14th IST Mobile Wireless Communications Summit. Dresden, Germany, June 2005.

[10]. Oscar Filio-Rodriguez, et al. Collaborative Spectrum Sensing for Cognitive Radio: Diversity Combining Approach. Wireless Sensor Network, 3, 24 – 37. www.SCiRP.org/journal/wsn. 2011.

[11]. Danijela Cabric, Artem Tkachenko, Robert W Brodersen. Spectrum Sensing Measurements of Pilot, Energy and Collaborative Detection. Berkeley Wireless Research Center, University of California, Berkeley. www.eecs.berkeley.edu. 2003.

[12]. Shridhar Mubaraq Mishra, Anant Sahai, Robert W Brodersen. Cooperative Sensing Among Cognitive Radios. University of California, Berkeley. www.eecs.berkeley.edu.

[13]. Simone O., Gambini J., Bar-ness Y. Cooperation and Cognitive Radios. CWCSPR, NJIT, University Heights, NJ 07102 USA.

[14]. Wireless Working Group. Technical Document on Cognitive Radio Networks. Draft, Version 1.0. September 15, 2006.

[15]. Vibhav Krashan Chaurasiya. International Journal of Enterprise Computing and Business Systems. Volume 1, Issue 2. www.ijecbs.com. July 2011.

[16]. Ying-Chang Lian, et al. Advanced Signal Processing for Cognitive Radio Networks.EURASIP journal on Advances in Signal Processing. Volume 2010, Article ID 715987. January 2010.

[17]. Ghasemi A. and Sousa E. S. Spectrum Sensing in Cognitive Radio Networks: The Cooperation-Processing Trade-Off. Wiley Wireless Communication and Mobile Comp. Special Issue on Cognitive Radio, Software-Defined Radio, and Adaptive Wireless Systems, vol. 7, no. 9, November 2007.

[18]. Zhi Quan, Shuguang Cui, Ali H. Sayed. Optimal Linear Cooperation for Spectrum Sensing in Cognitive Radio Networks. IEEE Journal of Selected Topics in Signal processing. Volume 2, No 1. February 2008.

[19]. Siva Ram Murthy C. and B. S. Manoj. Ad Hoc Wireless Networks, Architecture and Protocols. Prentice Hall. 2008.

[20]. Diomidis S. Michalopoulos and and George K. Karagiannidis. PHY-Layer Fairness in Amplify and Forward Cooperative Diversity Systems. IEEE Transactions on Wireless Communications. Volume 7, No. 3. March 2008.

Cell Broadcast Simulation in Android

Uzzal Das and Kay Zemoudeh
School of Computer Science and Engineering, California State University, San Bernardino, CA, USA

Abstract - *In text-based cell broadcast [5] a message is sent to a large number of recipients. The Cell Broadcast Simulation App on Android platform [1] is an attempt to make cell broadcast possible. The app simulates important concepts of broadcasting using text messages or Short Message Service [8]. The user of the app is not required to maintain or select contacts. The app maintains and updates the contacts automatically. The user just installs the app and sends a message and whoever has already installed the app will receive the message. There is a radius feature. This feature gives complete control over the broadcast range. The app has a message filtering and saving system. A user can set his/her keywords to filter the incoming messages. Incoming messages with a filter keyword will be saved under that keyword and the user can access those messages at a later time.*

Keywords: Cell Broadcast, One-to-many, SMS, Android, Java

1 Introduction

Cell phones use radio frequencies to communicate. Two different radio frequencies create a channel to make communication possible between cell phones. In general, this type of communication is a point-to-point connection and each connection uses a single channel to make a call or send a text message. Point-to-point has many limitations. Only two users can use it to communicate. So, if many users want to communicate with each other through text messages at the same time, then point-to-point communication is not very efficient. For a large group, many point-to-point connections have to be established for proper communication. There is a more efficient solution to this problem: Broadcasting. Only a single channel is required to broadcast data or messages among the cell phones (Mobile Phones). In broadcasting, many cell phones can listen to a particular channel provided by the Service Provider. Broadcasting is a simple way to reach many users (one-to-many) by using only a single channel with less traffic and overhead. Since cell phones have become a common utility and most people are using smart phone nowadays, it is easier to reach an individual using cell phone. In the time of emergency, when it is required to reach a large number of people, cell broadcasting is the most effective way. Social communication at the local level with less traffic is another important application of cell broadcasting and there are many other uses.

Since no promising application of Broadcasting exists, service providers do not provide broadcast channels in the US [5]. A prototype can be very helpful to investigate the applications of broadcasting, and Cell Broadcast Simulation app will serve this purpose. It simulates all the capabilities of SMS broadcasting. Using this app a large number of users can be easily reached. GPS feature and message filtering gives an indication of the many features that can be integrated with this promising smart phone application. Cell Broadcast Simulation has these basic features of the broadcasting along with other features. This app will act as a proof of concept that there are lots of practical uses of cell broadcast and upon availability of the broadcast channels this app will become the standard with which one can perform many real life activities all with little overload or traffic. The app is simple yet has many integrated features. The main focus of this app is to make cell broadcasting possible by giving the user a positive experience and expose them to its many potential uses.

For example, one can use this app to solicit the advice of his/her neighbors on a good local handyman. She will simply set the radius to one mile and sends a broadcast message such as "does anyone know a good handyman?" Those who have already installed the app will receive the message and if they are inclined they will respond with the phone number of a handyman that they had a good experience with. The sender of the message could wait for the responses by monitoring his/her phone, or set a new filter "handyman" to save away any message containing the word handyman and view them at a later time. As another example, one could solicit the traffic status from other drivers to find out the source of the traffic jam up the road and approximately how long it would take for the traffic to open up. All this happens anonymously without the parties involved knowing each other.

Using regular point-to-point channel to reach many people via text messages involves heavy traffic. But using a broadcast channel, it can be very easily done. All cell phones will listen to a particular radio frequency or broadcast channel. When a text message is sent on this channel, all those listening will receive the message at the cost of sending one message. Different level of broadcasting is possible. It can be limited to a particular radius. It can be limited to a cell area. It can be limited to a particular purpose. For example, it is very easy to discard unrelated messages, so that unauthorized users will not be able to see a message. In that way, the broadcasting can be used in a secure way. Social communication is another important application of text message broadcasting. Emergency broadcast is another important application.

Cell Broadcast Simulation app is implemented in a way so that the user can experience the cell broadcasting. When a

broadcasting channel is available, only the channel has to be integrated and nothing else needs to be changed in the app. Then the simulation app will be a fully functional broadcast app. Until then, this simulation app performs its functionality through the SMS system. Therefore, currently the broadcast is simulated using send and receive through SMS. It is important to note that users do not know how the broadcast is performed. The app will maintain the contact, so that only by installing the app, users can communicate with each other. The app has a radius feature which is the broadcasting area. With filtering and message-saving feature, users have total control over how they want to see their message. This app exposes the users to the functionality of the cell broadcasting and illustrates how promising the application would be. This app will be distributed to different service providers so that they can see the real application of the cell broadcasting and may agree to use their broadcast channel so that cell broadcasting can be implemented.

2 System Environment

Using this app on an Android cell phone, a user will be able to send an SMS to all the users who have also installed the app. She can also select a radius. When she sends an SMS, her current GPS position and her selected radius are added to the message. The receiver app will receive the message and calculates the distance with the sender's GPS location and receiver's GPS location. If the distance falls within both the sender's and the receiver's radius, then the message will be displayed on the receiver app. As shown in Figure 1, only user 4 receives the message because both users cover each other's radius. If the message has any word that matches the receiver's filter keyword, the message will be also saved and the receiver will be able to recall and see it later.

The process of adding a new user is shown in Figure 2.There will be a root user for this app. Root's mobile phone number will be used to introduce a new user to others. When a new user installs this app, the new user's phone will automatically send a message to the root to add her in the contact list. After receiving the add request message from the new user, the root updates its contact with the new user's mobile number and sends an add request to all other existing users. After receiving the add request from the root, existing user apps will automatically include the new user in their contact and send an add request with their own mobile number to the new user. The new user will receive those add requests and adds all of them to its contact list. All these processes happen in the background automatically without user involvement.

The root user can delete a user from the contact. The deletion processes is shown in Figure 3. To delete a user, the root's app will notify all other users about the deletion. Upon receiving the delete request from the root, all the other users will delete that particular user from their contact. The deletion process is also automatic without user involvement.

3 Implementation

The app is implemented on Android platform using Java, XML, SQLite, Eclipse IDE with Android Emulator [2, 3, 4, 6]. Any smart phone with Android 2.1 software platform or above can install the app. There should be GPS and SMS system available for the proper functioning of the app. The following four tables are used in the Database:

Contact:

Id	Contact

Message:

Id	Message	Phone	Time	Filter_Status

Distance	Latitude	Longitude

Filter:

Id	Filter

Filtered_Messages:

Id	Message	Phone	Filter_Key

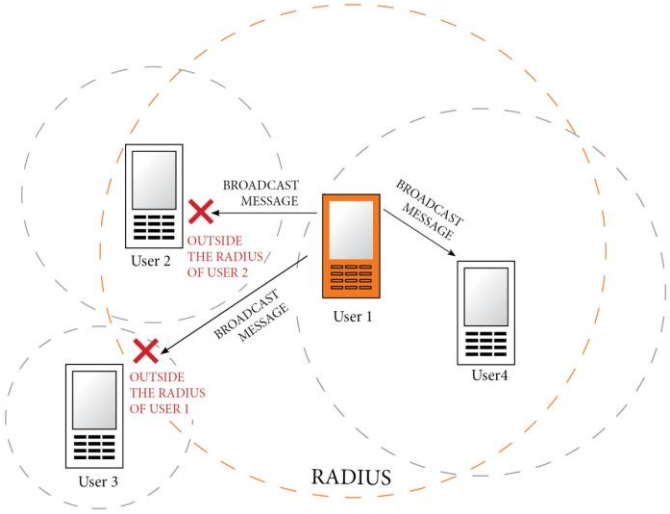


Figure 1. System Environment

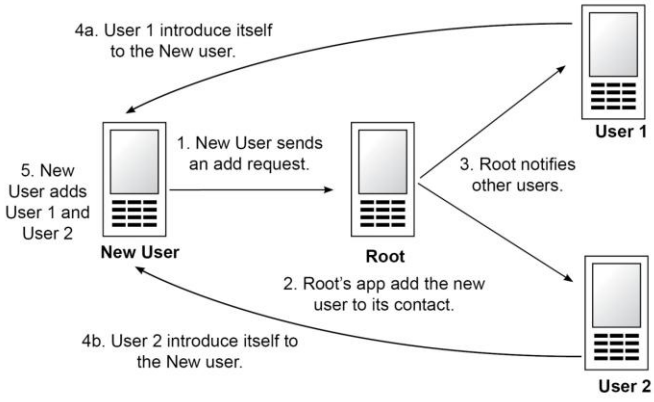


Figure 2. Adding a New User

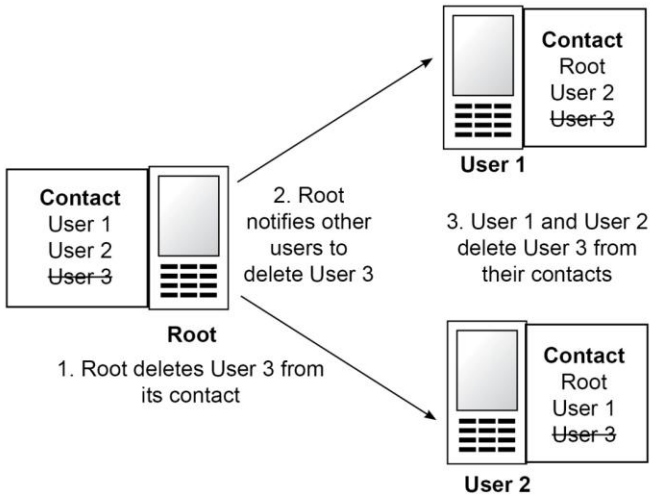


Figure 3. Deleting a User

There are several different screens with option menu and dialogs created with XML files. Icons are custom made. Navigation between different screens is very intuitive, and users can easily learn how to use the app.

There are many screens. The main focus is to keep the operation simple and similar to give the user a seamless experience. We will discuss four screens here. As soon as the app opens, the main screen appears, Figure 4. This screen is designed to display all the necessary information that is required to operate the app properly. There is a place to input the message and a button to send the message. The main screen also displays radius in terms of mile and the current location in terms of latitude and longitude [7]. It will also display a few filtered keywords that the user has already entered.

After installing the app, the user has to set the radius from the dialog shown in Figure 5. By clicking the menu button of the Android device, the user can get the option menu shown in Figure 6. The menu has six options to select. The "GPS Radius" option is for setting the GPS radius. The "Other Users" option opens up a new screen and displays the other users list. "Filter" option is for setting filter keyword and also for accessing saved filtered messages. "Delete All" option deletes all the messages in the main screen. "Exit" option shutdowns the app and the app will no longer use any resource of the device such as GPS service. When a message arrives, the app will notify the user through the notification service. It will make a sound and display the message in the notification bar. This feature can be turned off through the Notification ON/OFF option. If the notification service of the app is on, this option turns off the notification; otherwise, notification is on.

Figure 7 displays the "Filter" screen. When the user presses the "Filter" option from the option menu, this screen is displayed. One can put the filter keyword in the input field and press the add button to add the filter keyword. Messages with specified keywords will be saved under that keyword. One can see those messages by clicking on the displayed

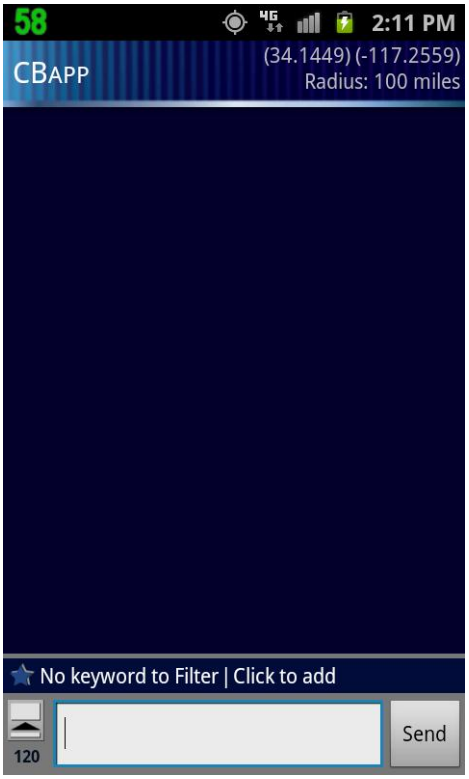


Figure 4. Main Screen

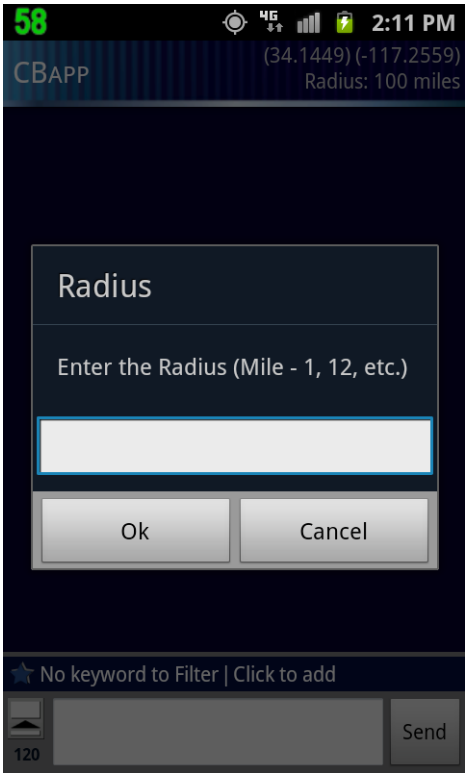


Figure 5. Radius Dialog

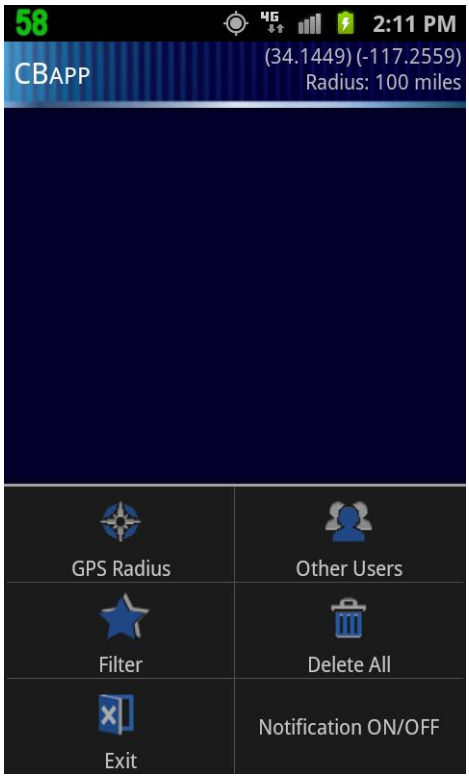


Figure 6. Option Menu

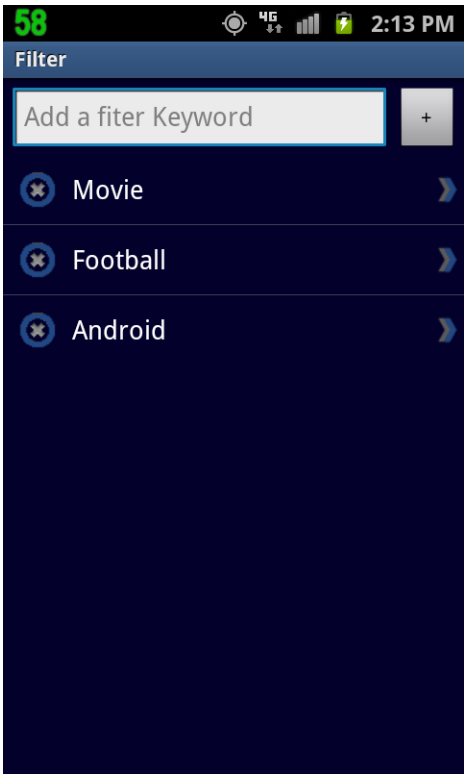


Figure 7. Filter Screen

keywords. One can also delete any keyword from this screen by pressing the delete button at the left of the keyword. All the messages saved under the deleted keyword will be also deleted.

The app has a root user, who receives the add and delete requests and notifies other users to add or delete users. The root has the authority to delete any user. Without root the app will not operate. Root should not be disconnected, and must be up at all times.

4 Conclusions

The app is developed to give the user the best experience to understand the usefulness of the text broadcast. Although a broadcast channel doesn't exist, the user would not be able to notice a difference. This app serves as a prototype of a real broadcast app.

Future work includes integrating a broadcast channel. The app is designed to easily replace the use of the current SMS system with the broadcast channel. This will reduce traffic for the carrier and cost for the user. It would also be beneficial to develop the same app for iPhone. This will give the app a wider audience and exposure.

5 References

[1] Open Handset Alliance, “Android”, openhandsetalliance.com, [Online]. Available: http://www.openhandsetalliance.com/android_overview.html [Accessed: 14 Nov, 2011].

[2] QuinStreet Inc, "Android SDK", webopedia.com, [Online]. Available: http://www.webopedia.com/TERM/A/Android_SDK.html [Accessed: 17 Dec. 2011].

[3] The Eclipse Foundation, "About the Eclipse Foundation", eclipse.org, [Online]. Available: http://www.eclipse.org/org/ [Accessed: 17 Dec. 2011].

[4] QuinStreet Inc, "Integrated Development Environment" webopedia.com, [Online]. Available: http://www.webopedia.com/TERM/I/integrated_development_environment.html [Accessed: 23 Dec. 2011].

[5] Cell Broadcast Forum, "What is Cell Broadcast", cellbroadcastforum.org, [Online]. Available: http://www.cellbroadcastforum.org/whatisCB/index.html [Accessed: 2 Jan. 2012].

[6] Google, "ADT Plugin for Eclipse", android.com, [Online]. Available: http://developer.android.com/sdk/eclipse-adt.html [Accessed: 2 Jan. 2012].

[7] NASA, "What is the definition of latitude and longitude?", nasa.gov, [Online]. Available: http://jwocky.gsfc.nasa.gov/teacher/latlonarchive.html" [Accessed: 2 Jan. 2012].

[8] Hillebrand. F et al., "SHORT MESSAGE SERVICE (SMS): THE CREATION OF PERSONAL GLOBAL TEXT MESSAGING", Chichester, West Sussex: Wiley, 2010.

Experimenting a Bluetooth-Based Positioning and Tracking Service of a Football Player

E. Basar, M. Salamah, and M. Kizildag,
{erden.basar; muhammed.salamah; mehmet.kizildag @emu.edu.tr; }

Eastern Mediterranean University, Computer Engineering Department
Gazimagusa, K.K.T.C. Mersin 10 – Turkey

Abstract

This paper presents an application for positioning and tracking of a mobile node in a Bluetooth-Based wireless networks. The addressed experimental real time application is tracking a football player in a match. The places of a player are calculated using triangulation positioning techniques via a small Body-Tag attached to the player. Furthermore, using this new methodology of positioning, the overall performances of a player (like, speed, fatigue, etc) can be calculated. The given methodology can also be used for other application areas as well.

1. Introduction

In wireless networks, mobility of a node has a great importance for many intuitive, commercial and technical applications. Computing of a mobile node's (MN) position is an important issue in telecommunication and became one of its standards. Positioning can be achieved via different technologies. Bluetooth positioning [1], Global Positioning System (GPS) [2], Ultra Wide Band (UWB), Radio Frequency Identification (RFID) [3] and sensor networks [4] are some examples. Many methods for locating MNs have been proposed and discussed in the literature. These include Angle of Arrival (AOA), Time of Arrival (TOA), Time Difference of Arrival (TDOA), and Received Signal Strength (RSS) [5].

Generally, location determination schemes can be divided into two major categories, namely, network-based positioning systems and handset-based positioning systems. In last decade, a variety of positioning technologies in both categories have been proposed and developed. Handset-based positioning methods use a modified handset to calculate its own position. These modified handsets use a fully or partially-equipped global positioning system receiver to position the object. Although network based methods for mobile positioning in wireless communication systems have their advantage, nevertheless, network based methods, when compared with handset based methods, are less complex to modify and generally less accurate [6]. Hence, improvement of accuracy, when using network based methods for mobile positioning, becomes necessary. Increasing accuracy expands the scope of applications where such a system can be used. Performance analysis of a player in football area can be a good example of such an application. A player's positions can be calculated in a time difference and statistical information can be calculated using positioning in wireless networking systems.

This research work focuses on the performance analysis and tracking of mobile nodes in location sensing wireless networks [7]. The wireless communication is provided using a Bluetooth-based short range transmission oriented network system. The given system can be used to develop applications in a variety of areas, from sensing a doctor's location in an emergency room of a hospital, to understanding and evaluating a player's performance in a football ground. In this study, a player is considered as a mobile node carrying specially designed tag capable of communicating with receivers that monitor the player.

The rest of the paper is organized as follows; section 2 discusses the architecture of the system. Section 3 details the performed

experiments. Section 4 highlights the discussion on the obtained results and finally we present conclusion in section 5.

2. System Architecture

The main purpose of this system is to obtain X and Y coordinates (physical position) of a mobile object. Bluetooth is used to calculate the position of the mobile node.
Bluetooth has two required Host Controller Interface (HCI) commands that can retrieve information from Bluetooth radio about the received signal. The HCI Read RSSI command and returns, for a specified Asynchouronous Connectionless (ACL) connection, a signed 8-bit integer giving values between -128 and +127 [8]. The HCI Read Get Link Quality command returns, for a specified ACL connection, an 8-bit unsigned integer giving values between 0 and 255 [8]. The HCI_"Get_Link_Quality" command returns a number that is directly equivalent to Bit Error Rate (BER) on the following scale [8]:

- Between 255 and 215 each bit difference represents 0.0025 % BER (255 means 0.0000 % BER, 254 means 0.0025 % BER, 253 means 0.0050 % BER etc.)
- Between 215 and 90 each bit difference represents 0.08 % BER (215 means 0.1000 % BER, 214 means 0.18 % BER, 213 means 0.26 % BER etc.)
- Between 90 and 0 each bit distance represents 0.64% BER.

The Body Tag is an intelligent Bluetooth Tag which provides an ideal solution for tracking and or positioning of a mobile object. It is made of shock-resistant plastic as seen in Fig.1, and can be attached to an individual's clothing or worn around the neck using a strap [9]. The tag acts as the located object. An individual wearing or carrying the Body Tag will thus be registered by access points covering the area in question.
Body Tag can be used, for example for [9]:
• Tracking: The Body Tag functioning as a means to track an individual or object and to keep a constant overview of a group of individuals or objects within a confined area.
• Positioning: The Body Tag functioning as a means to locate an individual or object and determine the objects position accurately
• Notification: The end-user can receive notifications or alerts of different kinds, like via alerts directly from the Body Tag (LED's and buzzer)
• Identification: The Body Tag functioning as a means to authenticate the identity of individuals or objects.
• Security: The Body Tag functioning as access control, only allowing individuals with proper credentials to access restricted areas
• Data storage: Relevant data can be stored directly in the Body Tag and can be easily transferred between the Tag and other securely paired Bluetooth devices.

Figure 1. Body Tag [9]

The Body Tag uses a protocol that interfaces to a command interpreter embedded in the firmware. The protocol is not part of the Bluetooth standard. The command interpreter protocol, BTCMD, allows another Bluetooth device to control and monitor the Body Tag. The protocol is currently applicable when using a connection between the Body Tag and another Bluetooth device using the Serial Port Profile (SPP). BTCMD mechanism is shown in the Figure 2.

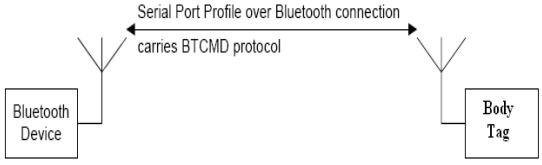


Figure 2. BTCMD Protocol [9]

The Protocol is modeled on a conventional client/server, RPC (Remote Procedure Call)-like structure. The client (a Bluetooth device) sends requests to the server (the Body Tag), the Body Tag attempts to service the requests and returns

responses to the client. Only one request can be in train at a time. The protocol has two message types:
Client Request: Client → Server Client requests to server.
Server Response: Server → Response from server to client.
Both message types uses same format which is shown in the Figure 3.

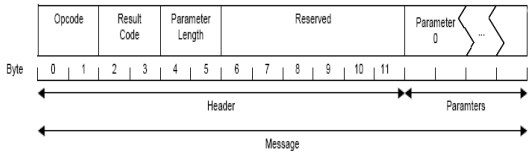


Figure 3. BTCMD Message Format [9]

Opcode field designates what command is requested.
The result code field is used in responses (Server Response) to tell the result of a request. It is not used for (Client Request) and here the field must be set to zero.
The parameter length field designates the length of all parameter measured in bytes. If there are no parameters in a request (Client Request) the field must be set to zero.
The parameters contain any values that must be supplied to a request or it is returned from a response.
There are many important commands embedded on the Body Tag. The most important command is GetRSSIValue which indicates Received Signal Strength Indicator (RSSI) but this command is not implemented yet. So, in the experiments RFSignalLinkQuality command is used to made a look up table for positioning. The command indicates the Radio Frequency Signal Link Quality between Bluetooth devices.
Bluetooth oriented architecture of the system, used in our experiments, is illustrated in Figure 4. There are three laptop computers which are connected via a LAN, equipped with USB Bluetooth dongles and one of them (Base station #1) has a database to record positions of the player. The player wears Body Tag and the aim is to calculate its position (X,Y). The computers are in known places depicted by (X1, Y1), (X2, Y2) and (X3, Y3) respectively, where each coordinate gives the location of a Base Station.

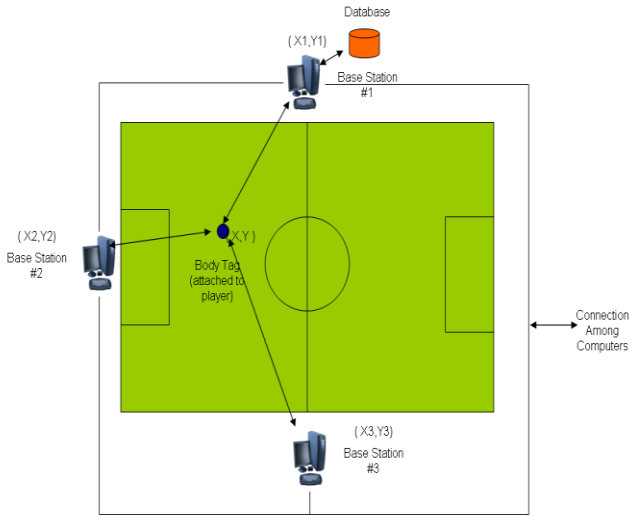


Figure 4. System Architecture.

3. Experiments

The current position of the player (wearing the Tag) is calculated using the triangulation method as shown in Figure 5.

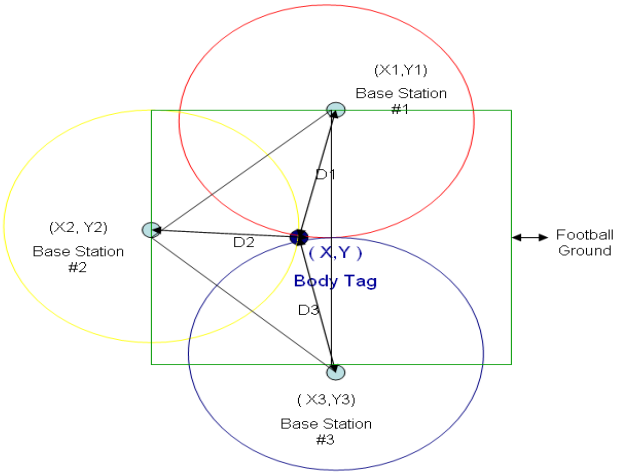


Figure 5. Triangulation Method.

Body Tag's BTCMD interpreter is used to find distances D1, D2 and D3 which are between Base Station # 1, Base Station # 2, Base Station # 3, and Body Tag respectively. The position of the player (X,Y) can be calculated as follows for every time difference.

$$A:(X-X1)^2+(Y-Y1)^2=D1^2 \quad (1)$$

$$B:(X-X2)^2+(Y-Y2)^2=D2^2 \quad (2)$$

$$C:(X-X3)^2+(Y-Y3)^2=D3^2 \quad (3)$$

Equations (1), (2), and (3) can be reduced to:

$$A - B : aX + bY = e \quad (4)$$

$$B - C : cX + dY = f \quad (5)$$

Where a, b, c, d, e, f are some constants.

$$e = aX + bY \quad (6)$$

$$f = cX + dY \quad (7)$$

Solving linear equations by Cramer's Rule;

$$X = \frac{Det\begin{vmatrix} e & b \\ f & d \end{vmatrix}}{Det\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \quad (8)$$

$$Y = \frac{Det\begin{vmatrix} a & e \\ c & f \end{vmatrix}}{Det\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \quad (9)$$

To find the distances D1, D2 and D3 we used a look up table that contains the location of the base stations. The coordinates of the player are recorded to the database in each time difference and performance analysis can be easily calculated accordingly.

Microsoft Visual Studio and C# programming language are used to provide the graphical user interface for the database and serial port communication. Furthermore IVT BLUESOLEIL software is used for discovering the Body Tag devices and the services on the Body Tags.

All electromagnetic radiation from radio waves to x-rays (RF Signals) travel at the speed of light (3×10^8 m/s). Hence, the signal travels 1 meter in approximately 3.3 nanoseconds. Nanoseconds measurement is solved by using Kernel functions of Windows in C#. The following experiments are carried out:

Experiment 1: A Dummy Data Send and Receive

In this experiment, the computer sends 12 packets as a request to the Body Tag, and the Body Tag deals with the request and sends the response to the computer that writes the response through the serial port. The problem with this experiment is that the total time for the event is approximately 80 milliseconds. Furthermore, the time for traveling of an RF signal with 12 packets, the time for writing through a serial port, the time needed by the Body Tag to work with RF signal and MS C# execution time, all affect the overall execution time. Hence it is not possible to use this method for nanoseconds measurements.

Experiment 2: Only Serial Port Writing Time

In this experiment only the time required for writing on a serial port is calculated for a request. Unfortunately, the writing on the serial port time is approximately 1.4 milliseconds for each request. Also in this experiment average writing time on serial port is taken into account but results are same and hence this method can not be used also.

Experiment 3: Try RSS Indication

As RSS indication is not implemented on the Body Tag hence getting results through RSS is not possible as well.

Experiment 4: RF Signal Link Quality Indication

In this experiment tag is replaced at different positions starting from 0 to 90 meters and a look up table for each meter is prepared. A number between 0 and 255 is given as a response to the Body Tag for RF Link Quality request. The RF Link Quality response value decreases when Body Tag goes further. The scheme results for 1, 10 and 25 iterations are shown in Figures 6, 7, and 8 respectively. It should be noted that there are 30 iterations in each meter.

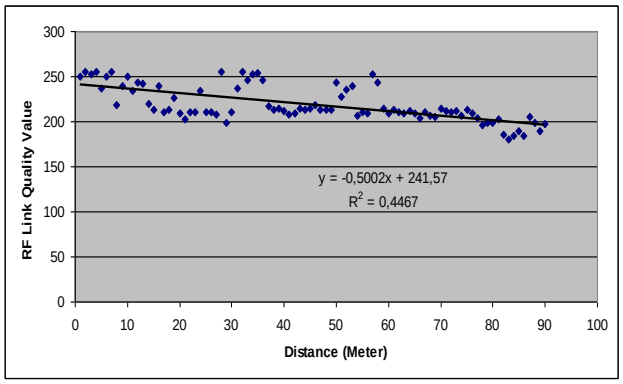


Figure 6. First 1 Iteration Outcome of Experiment 4.

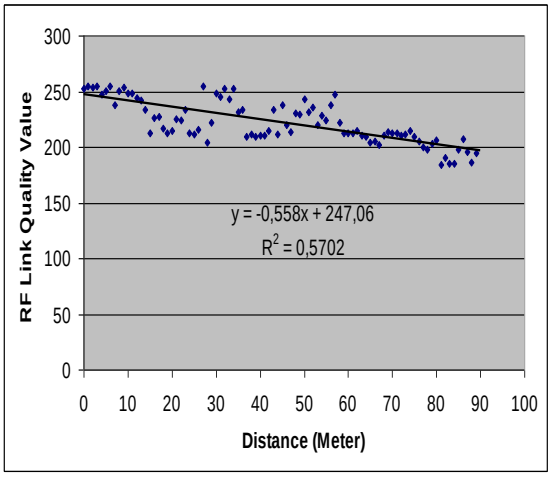


Figure 7. First 10 Iterations Outcome of Experiment 4.

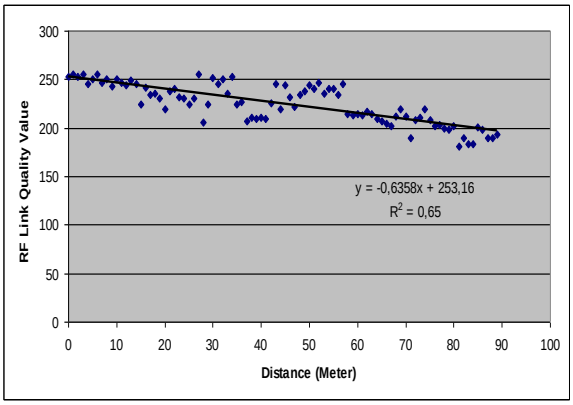


Figure 8. First 25 Iterations Outcome of Experiment 4.

The system reaches better results when the iteration number is increased. The best result in the look up table is first 25 iterations of the outcome of experiment. The slope is 0.6358 and in the look up table y = -0.6358x + 253.16 and R^2 = 0.65 value which is defined as the ratio of the sum of squares explained by a regression model and the total sum of squares around the mean [19].

Experiment 5: RF Signal Link Quality Indication in the Line of Sight

In this experiment tag is replaced between meters starting from 0 to 99 meters and again a look up table for each meter is prepared. A number in the range of 0 to 255 is given as a response to the request for RF Link Quality request. The RF Link Quality response value decreases as the Body Tag distance increases. The scheme results for 1, 10 and 25 iterations are shown in Figures 9, 10, and 11 respectively. In the experiment line of sight property is taken into account, and there are 50 iterations in each meter.

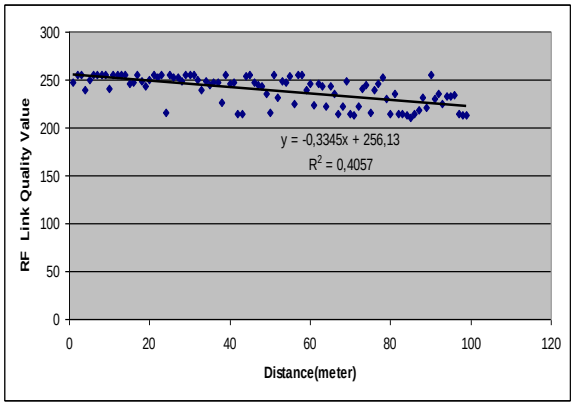


Figure 9. First 1 Iterations Outcome of Experiment 5.

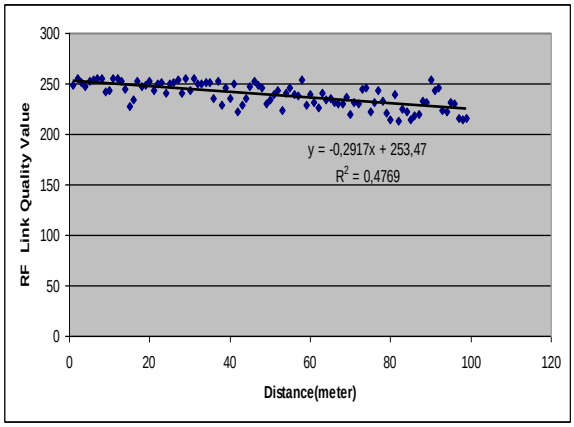


Figure 10. First 10 Iterations Outcome of Experiment 5.

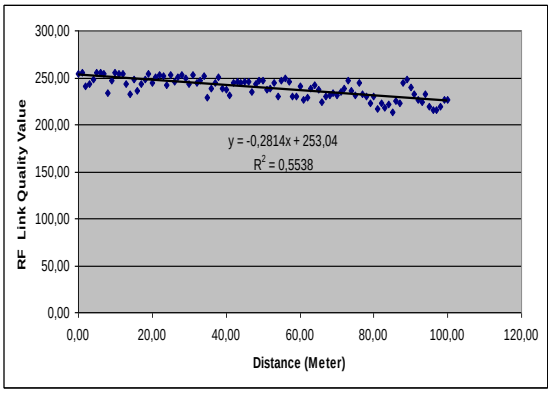


Figure 11. First 25 Iterations Outcome of Experiment 5.

Here the slope is less than the other experiments because the line of sight is an important factor in achieving better results. In the last experiment the look up table is changed to y = -0,2814x + 253,04 and R^2 = 0,5538 has approximate values.

4. Results Accuracy

In the experiments, triangulation method is used to calculate the position of the tag. The tag's position is calculated using 50 iterations for each place of the tag by all base stations. A typical trajectory of the tag is shown in Figure 12.

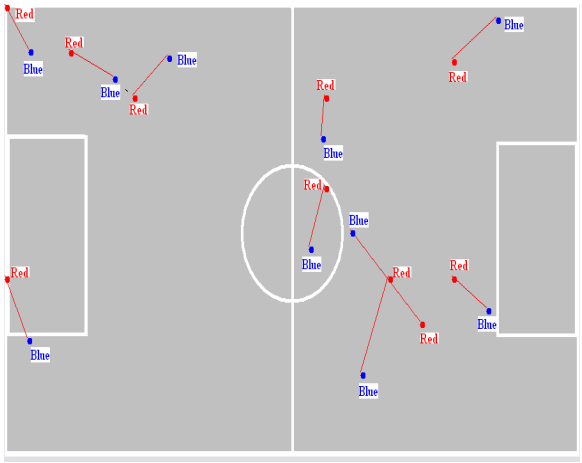


Figure 12. Calculated Positions of the Tag

The red points are the real position of the tag and blue points are the calculated position of the tag. The differences between red and blue points show the attained accuracy (Figure 13) for 10 different trials.

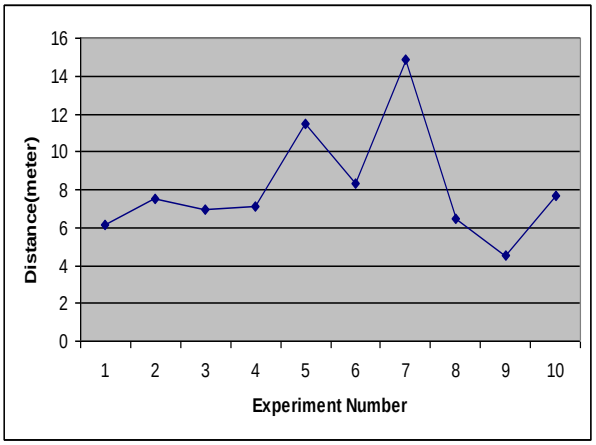


Figure 13. Distances between Red and Blue Points

The average accuracy is found as 6.89 meters. The experimented trajectory for a player is shown in the Figure 14 which is similar trajectory as in the simulation.

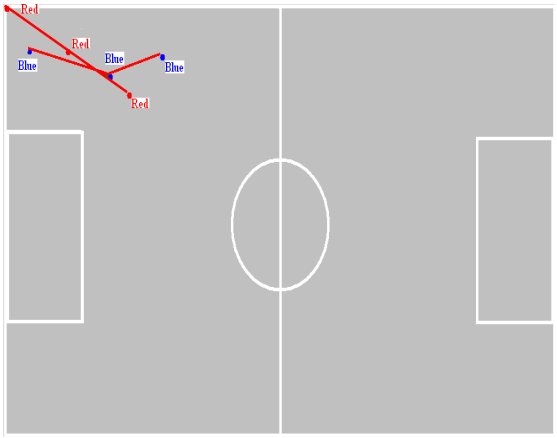


Figure 14. Experimented Trajectory

5. Conclusion

In this research work, a new application for the Sapphire Dart system model for football players is simulated. The implemented system architecture and experiments and their results are discussed. The simulation and the experiment results of the implemented system can be helpful for the football teams to check their players' performances in a digital environment.

The average speeds, and total displacements of players are very important for a coach to compare their performances. This record can help the coach to substitute a player during a match. Furthermore, this can guide a coach in determining which offence player is suitable for helping the defense of his team, when needed and which defense player is suitable for helping in offense of his team, when needed.

The simulation program finds the player's speed and displacement; it may be desirable to keep more statistics such as the total calorie dissipated by the player. These calculations can be future work of this study.

Moreover, all assumptions in the simulation model are based on 2D motion. The study is being extended to contain 3D motion. This will require more multifaceted calculations, for example calculation of the calorie dissipation

will become more complex as the jumping motions of a player should be considered.
Finally, it is worth to mention that the position location system can also be applied to the ball in order to track its motion. This can help in refereeing process of the game.

References

[1] K.Hallberg, J.Nilsson, M.Synnes, "Positioning with Bluetooth**,"** Proc. of the 10th International Conference on Telecommunications (ICT 2003), Vol 2, pp.:954 - 958, Feb. 2003.

[2] Jyh-Ching Juang, "On GPS positioning and integrity monitoring," IEEE Transactions on Aerospace and Electronic Systems, Vol. 36, Issue 1, pp.:327 – 336, Jan. 2000.

[3] Fontana, R.J., Gunderson, S.J., "Ultra-wideband precision asset location system", Proc. of IEEE Conference on Ultra Wideband Systems and Technologies, Vol 7, pp.147-150, May 2002.

[4] Xiang Ji, Hongyuan Zha, "Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling," 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004) Vol. 4, pp. :2652 - 2661, March 2004.

[5] Abhishek Pramod Patil, "Performance Analysis of Bluetooth Technologies and Their Applications to Location Sensing," Master Thesis, Michigan State University, 2002.

[6] Vossiek, M. et. al. "Wireless local positioning - concepts, solutions, and applications**,"** Proc. of the Radio and Wireless Conference (RAWCON'03), pp.:219 – 224, Aug. 2003

[7] Mehmet Kizildag, Performance Analysis and Tracking of Mobile Nodes in Location Sensing Wireless Networks, Master Thesis, Eastern Mediterranean University, 2007.

[8] http://www.bluetooth.com/, "The official Bluetooth Wireless Info. site".

[9] http://www.bluelon.com/, "Bluelon Inc. site".

Introduce Mural Artistry into Cartoon Material Creation

Peng Ge-gang, LI Xin- yu, SONG Ying, XIANG Li- sheng, SHEN Qing, Li Ren-fa
Hunan University and Talkweb Information System CO. LTD, China，410205

Abstract: Mural is one of the earliest art in human history, which directly painted on the wall recorded people's superb skill and creating artistry. Our ancestors drew up a variety of graphics with charred branches (charcoal) or yellow soil on the wall to record their important events, such as hunting, dance, ritual and war, and form the ancient murals. Subsequently, the murals in Dunhuang of China and pyramids of Egypt developed superbly. They surround some special themes, showed superb artistry and formed the world-renowned historical and cultural heritages. We aim this project to such a long history, high performance capability art form. By self-designed image processing algorithms, the image merge and color-transform algorithm, we create a newly pseudo-archaic mural recreating technology. Based on such a technology, a traditional digital image may be re-created to have a special view similar to an antique mural, so any user may re-create his-own cartoon character or background by himself with some common digital material.

Key words: Mural; Digital Cartoon；Digital Image Processing

1. Introduction

The animation product has a dual character of cultural product and information product. As a cultural product, it is asked to be as funny as possible to make the most users like to see and hear. As an information product, it is asked to be as cheap as possible to get the most benefit. On the other hand, digital media departments have accumulated a huge of different kind materials. How to reuse them as the software reuse does becomes one of the goals in animation industry. Can these digital media reused as an animation material[1]? We are deal with it in this paper.

Mural is one of the earliest art in human history, which directly painted on the wall recorded people's superb skill and creating artistry. Long ago, our ancestors drew up a variety of graphics with charred branches (charcoal) or yellow soil on the wall to record their important events, such as hunting, dance, ritual and war, and form the ancient murals. Subsequently, the murals in Dunhuang of China and pyramids of Egypt developed superbly. They surround some special themes, showing superb artistry and formed the world-renowned historical and cultural heritages. But after the vicissitudes, natural calamities and man-made misfortunes, these historical and cultural heritages are being damaged inevitably. The use of modern technology to in-paint these ancient murals become computer workers problem, and have achieved gratifying results[2, 3].

We aim this project to such a long history, high performance capability art form in another way that is reusing those digital pictures to re-create cartoon material. Based on self-designed image processing algorithms, the image merge and color-transform algorithm, we create a newly pseudo-archaic mural recreating technology. Based on such a technology, a traditional digital image may be re-created to have a special view similar to an antique mural that is a pseudo-archaic mural. So any user may re-create his-own cartoon character or background himself with some common digital material.

2. Technical Procedure

The project's core algorithm includes image merge and color-transform.

The former algorithm realizes the local modified image merges with the global image seamlessly. With a user determined parameter "variation threshold", this algorithm determines the scale for color merge. In other word, the larger the threshold is, the larger the area is modified according the creating desire. Several kind of re-creating may be done after your operation. For example, if you are desire to show a local damage on the mural, you may modify the original color of these pixels (in a digital image) to earth yellow or gray white, as to simulate the natural color of the ancient mural after a long time natural calamities.

The latter algorithm may get a visual simulation of a mural under a long time vicissitudes, from the original bright colorful to dark yellow fade. We are going to describe the algorithm in detail as definition 1 later (see Figure 1 and figure 2 also).

1(a) true mural. 1(b) re-creating mural.

Fig 1 Visual compare of colorful murals

2(a) true charcoal mural. 2(b) re-creating mural.

Fig 2 Visual compare of charcoal drawing murals

2.1 Definition

Definition 1. When modify the color of a pixel from the original value to a product with a coefficient, said it implements a pixel color-transform operation. The coefficient, as a multiplier here, may be a negative one or positive one. In particular, when a non-negative coefficient is used, it named as a darken color-transform operation; otherwise, named as a whiten color-transform operation (see Figure 3, figure 4). Detail definition may give as below:

A. The original brightness value Y of the pixel is calculated by:

$$Y = 0.257 * r_{in} + 0.504 * g_{in} + 0.98 * b_{in} \quad (1)$$

Here r_{in}, g_{in}, b_{in} is the RGB value of this pixel.

B. Color-transforming coefficient M set is an effective coefficient and correlated with the brightness of Y.

C. The output of this pixel after a color-transform operation is given as:

$$r_{out} = abs(r_{in} - (M_1 * r_{in})) \quad (2)$$

$$g_{out} = abs(g_{in} - (M_2 * g_{in})) \quad (3)$$

$$b_{out} = abs(b_{in} - (M_3 * b_{in})) \quad (4)$$

3(a) original one. 3(b) after a transform operation

Fig 3 A darken color-transform operation

4(a) original one. 4(b) after a transform operation

Fig 4 A whiten color-transform operation

Definition 2. It is said an image merge operation may be done successfully if the two conditions below are satisfied. First, if the difference between a sample pixel and some pre-selected pixels is less than threshold T. Second, if there is no visual confusion after these (selected) pixels accomplished a color-transform operation (as defined in definition 1). Detail describe may give as below:

A. The visual confusion can be measurement with two indexes below:

(1) Color-transform range: It is measured by the rate of transform pixels among the total pixels. It correlates with the threshold T. That means, larger the threshold T and larger the range after a color-transform.

(2) Color variation: It correlates with the coefficient M (in definition 1). That means, larger the coefficient M (average) and larger color variation.

B. By select a suitable T and M, a new image, may be created. Such an image has an allowed color variation but has a different visual scene and in a recognizable manner. (See fig. 3 and fig. 4)

2.2 Set parameters

In order to meet user DIY needs, many parameters, including iteration times, variation threshold, are reserved.

1. Iteration times (iterativeTimes): Because of long time exposed in a natural environment, mural color becomes darken. According the selected iteration times, the program iterates several times to simulate the darker and darker procedure.

2. Color sample pixels (colorSamples): User may choice two pixels as color samples. If the difference between the samples and some pre-selected pixels is less than threshold T, the program executes image merge operation (according to definition 2); otherwise, executes image color-transform operation (according to definition 1).

3. A variation threshold (variationThreshold): This parameter defines a threshold between the samples and the pre-selected pixels (both in RGB component). It turns, larger the threshold and larger color extending region.

2.3 recreating a pseudo-archaic mural

Algorithm 1. The purpose of this algorithm is to create a pseudo-archaic mural. Correlating parameters are iteration times, colorSamples and variation threshold. First, calculate the RGB value of colorSamples and named as color1 and color2 (r_1, g_1, b_1 and r_2, g_2, b_2). And then, if the difference (in RGB value) between samples and those being tested pixels is less than threshold T, the program retains the original color; otherwise do darken color-transform.

Step (1) calculates the RGB value of colorSamples and named as color1 and color2;

Step (2) by completing steps (3) and decrease iterativeTimes until iterativeTimes equal 0;

Step (3) by completing step (4) ~ (6) column by column and line by line to get the output of every pixel;

Step (4) if the difference between the RGB value of current pixel (r_{in}, g_{in}, b_{in}) and the samples (r_1, g_1, b_1 or r_2, g_2, b_2) is less than threshold T, the output of current pixel ($r_{out}, g_{out}, b_{out}$) keeps the original value (r_{in}, g_{in}, b_{in}). Otherwise turn to step (5);

Step (5) gets the original brightness value Y from r_{in}, g_{in}, b_{in}, according formula (1). Then get M from Y.

Step (6) call a subroutine to do a darken color-transform operation. This subroutine is based on formula (2)~ (4).

Algorithm 2. The purpose of this algorithm is to create a charcoal stick mural (as fig 2(b)).

Correlating parameters are colorSamples and variation threshold. First, calculate the RGB value of colorSamples and named as color1 and color2 (r_1, g_1, b_1 and r_2, g_2, b_2). And then, if the difference (in RGB value) between samples and those being tested pixels is less than threshold T, the program do darken color-transform, otherwise do darken color-transform.

Step (1) takes a grayish operation to the whole picture.

Step (2) set the whole reserved output image data buffer to white.

Step (3) calculates the RGB value of colorSamples and named as color1 and color2;

Step (4) by completing step (4) ~ (7) column by column and line by line to get the output of every pixel;

Step (5) if the difference between the RGB value of current pixel (r_{in}, g_{in}, b_{in}) and the samples (r_1, g_1, b_1 or r_2, g_2, b_2) is less than threshold T, do nothing. Because the whole reserved output image data buffer is set to white, so do nothing here equal to set the output of current pixel to white (same as $r_{out} = 255, g_{out} = 255, b_{out} = 255$). Otherwise turn to step (6).

Step (6) according formula (1) gets the original brightness value Y from r_{in}, g_{in}, b_{in} . Then get M from Y.

Step (7) call a subroutine to do a darken color-transform operation.

5(a) the result of Photoshop 5(b) this project.

Fig5 compare the result with Photoshop, adobe

3. Conclusion

Through this research we found: after an image merge and color-transform operation, a similar to an ancient mural visual effect may get.

Different color-transform coefficient M may affect the effect of color-transform. In contrast to figure 1(b) and figure 3(b) may find a different visual effect from the same source image, due to the different darken color-transform.

We are glad to say that pictures processed by our independently designed technology and algorithm have the similar effect of Photoshop, a well-known image processing software of Adobe company. Picture 5(a) and 5(b) respectively represent the two technical results for comparison.

Further comparison and identification are welcome, as well as inquiries and acquisitions of source code from the company the

authors' work for.

We hope to get criticisms and corrections on deficiencies in this article.

Thank the national science and technology support plan and the State Ministry of culture for supporting this project.

References

1. Li Xinyu, Song Ying, Xiang Lisheng, Shen Qing, A Mobile Cartoon Creating Scheme Based on the Materials Reuse, ICWN'09.
2. Huang Wei, Wang Shuwen, Yang Xiaoping, Jia Jianfang, Dunhuang murals in-painting based on image decomposition, Journal of Shandong university (engineering science), 2010.4
3. Yang Xiaoping, Wang Shuwen, Dunhuang mural in-painting in intricate disrepaired design based on improvement of priority algorithm, Journal of computer-aided design and computer graphics, 2011.

A Novel Correlation Function for CBOC Signal Synchronization

Changha Yu, Youngpo Lee, and Seokho Yoon†
College of Information and Communication Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do, Korea
†Corresponding author

Abstract— *Binary offset carrier (BOC) signal synchronization is based on the correlation between the received and locally generated BOC signals. Thus, the multiple side-peaks in BOC autocorrelation are one of the main error sources in synchronizing BOC signals. Recently, new correlation functions with no side-peak were proposed for sine and cosine phased BOC signal synchronization, respectively, by the authors [3]. In this paper, we propose new correlation functions with no side-peak for composite BOC (CBOC) signals by using the similar approach to the previous work.*

Keywords: CBOC; side-peak; ambiguity problem; synchronization

1. Introduction

In binary offset carrier (BOC) modulation, the BOC signal is created by the product of the data signal, a pseudo random noise (PRN) code, and a sub-carrier. Due to its capability to resist multipath and its separated spectrum from that of the global positioning system (GPS) signal [1], the BOC has been adopted as the modulation method for the global navigation satellite systems (GNSSs) including the European Galileo and the GPS III systems [2].

In GNSSs, a timing error from the synchronization process can result in a critical positioning error. Thus, timing synchronization is crucial for reliable GNSS-based communications. The BOC signal synchronization is generally carried out in two stages: acquisition and tracking. In acquisition stage, first, time phase of the locally generated BOC signal is aligned with that of the received signal within the allowable tracking range, and then, the fine adjustment is performed to achieve synchronization in tracking stage. Fig. 1 shows the autocorrelation and early-late discriminator output for BOC signal in the acquisition and tracking stages, respectively, where T_b is the PRN code chip duration, d denotes an early-late spacing, and false alarm is the event that an autocorrelation value outside the allowable tracking range exceeds a specified threshold. From the figure, we can see that the autocorrelation has multiple side-peaks, which would increase the false alarm probability, and consequently, might cause the synchronization process converge to a false lock point. This is called the *ambiguity problem*.

In order to solve the ambiguity problem, new correlation functions with no side-peak were proposed for original BOC signals including sine phased BOC (SinBOC) and cosine phased BOC (CosBOC) [3], whose sub-carriers are the sine and cosine phased square waves having each binary ± 1 values, respectively. However, the correlation function with no side-peak proposed for original BOC signals cannot be employed for composite BOC (CBOC) signals specialized for use on Galileo open service [2]. In this paper, thus, we propose new correlation function for signal synchronization of CBOC signal with a similar approach as in [3].

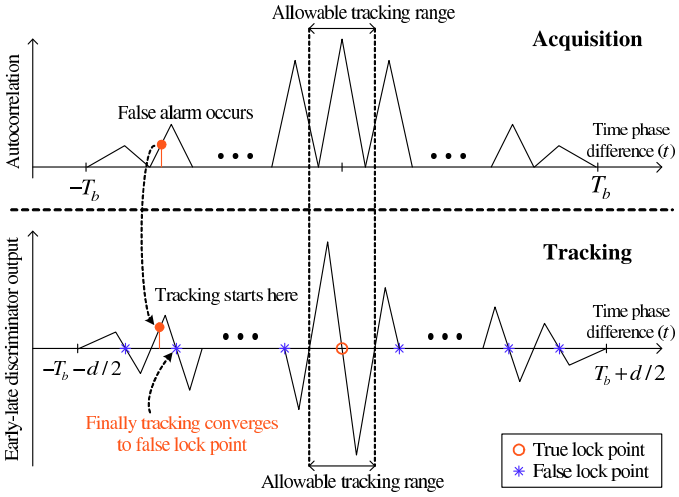


Fig. 1: Ambiguity problem in BOC signal synchronization.

The remainder of this paper is organized as follows. Section 2 describes the CBOC signal model and proposes a new correlation function for CBOC signals Section 3 presents simulation results, and finally, Section 4 concludes this paper.

2. Proposed Correlation Function

The CBOC signal CBOC(u, v, γ) is obtained from a weighted sum of two BOC signals SinBOC$(u, 1)$ and SinBOC$(v, 1)$ with the power split ratio γ, where SinBOC(kn, n) is defined as the product of the navigation data, a PRN code, and a square wave sub-carrier $\mathrm{sgn}\{\sin(2\pi knt \times 1.023 \times 10^6)\}$. For example, CBOC$(6, 1, 1/11)$ is generated through the combination of SinBOC$(6, 1)$ and SinBOC$(1, 1)$ with spectrum components

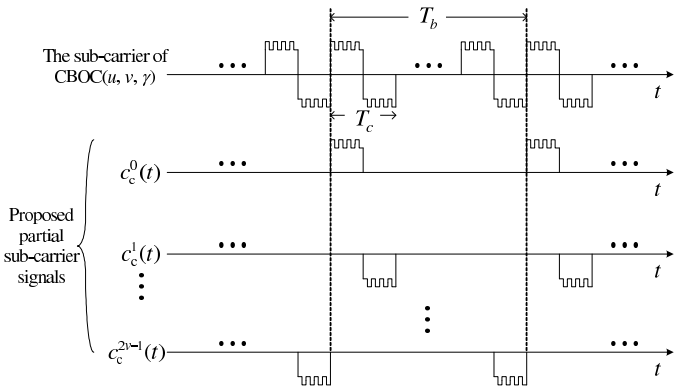


Fig. 2: Waveforms of the conventional and proposed sub-carrier signals for CBOC(u, v, γ).

given by

$$G_{\text{CBOC}}(f) = \frac{1}{11}G_{\text{SinBOC(6,1)}}(f) + \frac{10}{11}G_{\text{SinBOC(1,1)}}(f), \quad (1)$$

where $G_{\text{SinBOC}(\cdot,\cdot)}(f)$ is the unit power spectrum density of a SinBOC defined in [4]. Thus, it is natural to consider u and v having u/v $(u > v)$ even number to guarantee orthogonality between two SinBOC (fundamentally, sub-carrier) signals, then, the CBOC(u, v, γ) signal can be expressed as

$$\begin{aligned} S_{\text{c}}(t) &= c(t)d(t)c_{\text{c}}(t) \\ &= c(t)d(t)\{\sqrt{\gamma}c_{\text{sin}(u,1)}(t) + \sqrt{1-\gamma}c_{\text{sin}(v,1)}(t)\}, \end{aligned} \quad (2)$$

where the CBOC sub-carrier $c_{\text{c}}(t)$ is the weighted sum of two square wave sub-carrier $c_{\text{sin}(u,1)}(t)$ $(= \text{sgn}\{\sin(2\pi ut \times 1.023 \times 10^6)\})$ and $c_{\text{sin}(v,1)}(t)$ $(= \text{sgn}\{\sin(2\pi vt \times 1.023 \times 10^6)\})$. Fig. 2 shows the waveform of the conventional and proposed partial sub-carriers for CBOC(u, v, γ), where T_c denotes the period of the sub-carrier $c_{\text{sin}(v,1)}(t)$. In this paper, we assume that $d(t) = 1$ as in a pilot channel. A GNSS often includes a pilot channel to achieve rapid synchronization in the absence of data modulation on the transmitted signal [5].

Fig. 3 shows the processes of generating the proposed correlation function for CBOC(u, v, γ), where $R_{\text{c}}^0, R_{\text{c}}^1, \cdots, R_{\text{c}}^{2v-1}$ denote correlations between the received CBOC(u, v, γ) signal and partial sub-carrier signals $c_{\text{c}}^0(t)$, $c_{\text{c}}^1(t)$, $\cdots$, $c_{\text{c}}^{2v-1}(t)$ over T_b, respectively. From the figure, we can see that the operation $|R_{\text{c}}^q| + |R_{\text{c}}^{2v-q-1}|$ for $q = 0, 1, \cdots, v-1$ create the correlation function with $4v - 2q - 1$ peaks including a main-peak and $2q$ side-peaks in the same shape as the main-peak. On the other hand, the operation $|R_{\text{c}}^q - R_{\text{c}}^{2v-q-1}|$ creates the correlation function with $4v - 4q - 2$ side-peaks only where the main-peak and the $2q$ side-peaks are removed. Thus, with the operation $|R_{\text{c}}^q| + |R_{\text{c}}^{2v-q-1}| - |R_{\text{c}}^q - R_{\text{c}}^{2v-q-1}|$, we can obtain a correlation function with a main-peak and $2q$ side-peaks in the same shape as the main-peak. Finally, removing $2q$

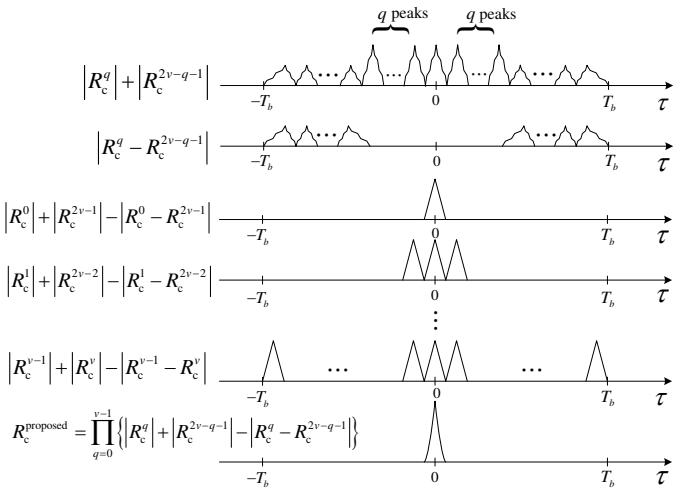


Fig. 3: The process of the proposed correlation function generation.

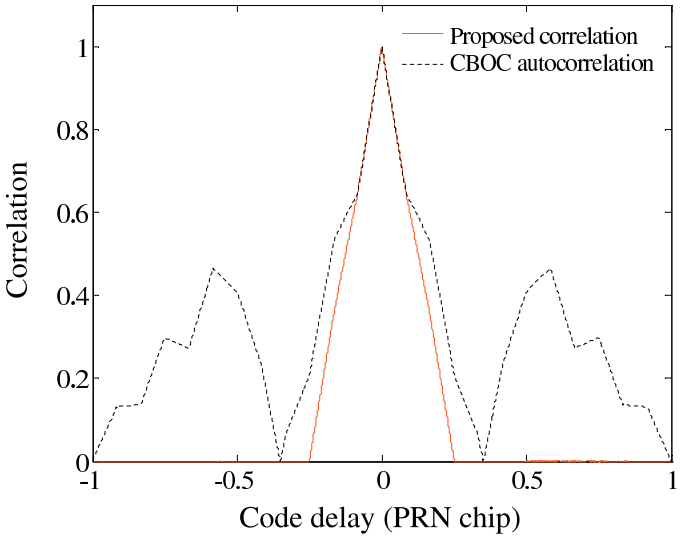


Fig. 4: The proposed correlation function and autocorrelation function of CBOC$(6, 1, 1/11)$.

side-peaks and increasing the main-peak magnitude, we can obtain a correlation function with no side-peak as as

$$R_{\text{c}}^{\text{proposed}} = \prod_{q=0}^{v-1}\{|R_{\text{c}}^q| + |R_{\text{c}}^{2v-q-1}| - |R_{\text{c}}^q - R_{\text{c}}^{2v-q-1}|\}. \quad (3)$$

Fig. 4 shows the proposed correlation function and autocorrelation function of CBOC$(6, 1, 1/11)$, and we can observe that side-peaks can be completely removed with the proposed correlation function.

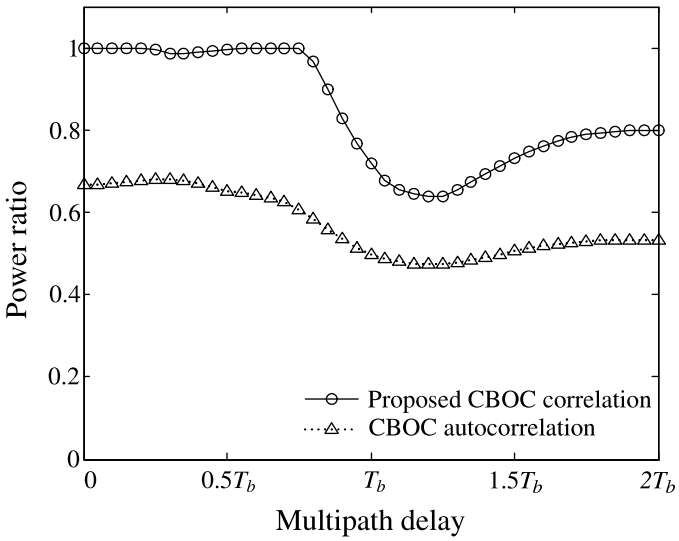


Fig. 5: Power ratio of the proposed correlation and autocorrelation functions for CBOC$(6, 1, 1/11)$ in a multipath channel.

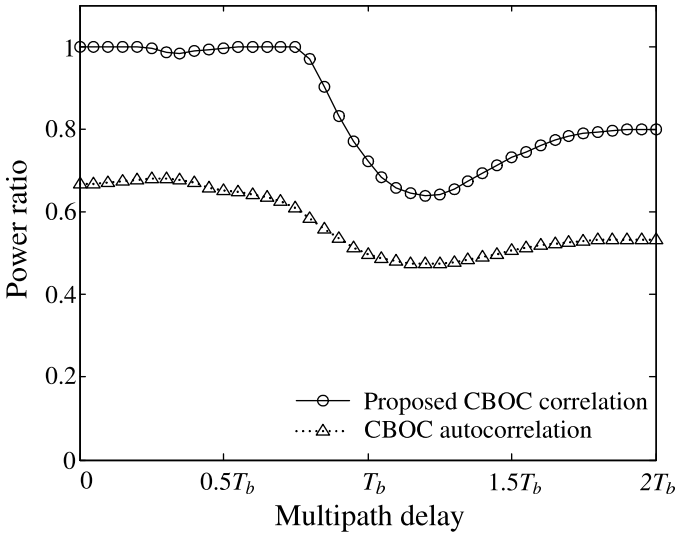


Fig. 6: Power ratio of the proposed correlation and autocorrelation functions for CBOC$(6, 1, 4/33)$ in a multipath channel.

3. Simulation Results

In this section, proposed correlation functions are compared with autocorrelation functions in terms of the power ratio between a main-peak and all peaks including the main-peak defined as

$$\text{power ratio} = \frac{\text{power in a main-peak}}{\text{power in all peaks}} \tag{4}$$

It should be noted that the power ratio of 1 means that correlation functions do not have any side-peak. For simulation, we assume the two-path channel model [6], whose impulse response is defined as

$$h(t) = \delta(t) + \alpha\delta(t - \beta), \tag{5}$$

where α $(\alpha < 1)$ denotes the attenuation factor of second path which is set to 0.5 in this paper, β $(\beta > 0)$ is the time difference between the first and second path, and $\delta(t)$ represents the Dirac-delta function.

Figs. 5 and 6 show the power ratio of the proposed correlation and autocorrelation function for CBOC$(6, 1, 1/11)$ and CBOC$(6, 1, 4/33)$, respectively, in a two-path channel. From the figures, we can see that the proposed CBOC correlation function has a better power ratio than that of the CBOC autocorrelation function in the two-path channel.

4. Conclusion

In this paper, new correlation functions with no side-peak have been proposed for CBOC signal synchronization. We have first created new sub-carrier signals by dividing the conventional sub-carrier signals of CBOC signal. Then, new correlation functions with no side-peak have been obtained by combining the correlation between received CBOC signal and divided sub-carrier signals. In a multipath channel, the proposed correlation functions have the power ratio less than 1; however, they have much better power ratios than those of the autocorrelation functions.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012-0005066 with funding from the Ministry of Education, Science and Technology (MEST), Korea, by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency under Grant NIPA-2012-H0301-12-1005 with funding from the Ministry of Knowledge Economy (MKE), Korea, and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] O. Julien, C. Macabiau, M. E. Cannon, and G. Lachapelle, "ASPeCT: Unambiguous sine-BOC(n,n) acquisition/tracking technique for navigation applications," *IEEE Trans. Aerospace and Electronic Systems*, vol. 43, pp. 150-162, Jan. 2007.

[2] M. Fantino, G. Marucco, P. Mulassano, and M. Pini, "Performance analysis of MBOC, AltBOC and BOC modulations in terms of multipath effects on the carrier tracking loop within GNSS receivers," in *Proc IEEE/ION Position Location and Navigation Symposium (PLANS)'08*, 2008, pp. 369-376.

[3] S. Kim, D. Chong, Y.-B. Joung, S. Ahn, Y. Lee, S. Y. Kim, and S. Yoon, "A novel unambiguous BOC signal synchronization scheme for global navigation satellite systems," in *Proc. IASTED Commun. Internet, Inform. Technol (CIIT)'07*, 2007, pp. 43-48.

[4] J. W. Betz, "Binary offset carrier modulations for radionavigations," *Journal of The Institute of Navigation*, vol. 48, pp. 227-246, Winter 2001-2002.

[5] E. S. Lohan, A. Lakhzouri, and M. Renfors, “Complex double-binary-offset-carrier modulation for a unitary characterisation of Galileo and GPS signals," *IEE Proc.-Radar Sonar Navig.*, vol. 153, pp. 403-408, Oct. 2006.

[6] J. Soubielle, I. Fijalkow, P. Duvaut, and A. Bibaut, “GPS positioning in a multipath environment,” *IEEE Trans. Signal Process.*, vol. 50, pp. 141-150, Jan. 2002.

Towards Enhancing Situational Awareness in Pervasive Systems using Multimodal Sensation

Mohamed A. Abd el Samie, Sherif G. Aly
Department of Computer Science and Engineering,
The American University in Cairo

Abstract

This paper aims at introducing a situation aware model that utilizes multimodal sensor data provided through various sensation capabilities available on current handheld smart phones. The model makes use of ten different virtual and physical sensors available on mobile phones, to develop a wider selection of possible parameters that identify a situation for one of five predefined context scenarios namely: In meeting, Driving, In Party, In Theatre and Sleeping. In order to understand how the various contextual parameters will contribute to the identification of the aforementioned situations, a detailed survey was conducted as part of the human perspective context acquisition to conclude the users' preferences in the case of each context situation. The survey was divided into 5 phases where participants identified the most relevant sensor attribute that would best describe each context situation. Details and results of the survey are mentioned.

Keywords

Situational awareness, Pervasive computing, Context awareness, Mobile Computing

1. Introduction

As the massive shift from dedicated to mobile computing continues to rapidly spread, the computational power and performance of handheld devices have also exponentially increased. This recent evolution in mobile computing created a motivating platform for situation aware applications, due to the high set of features and sensation abilities currently available on most handheld devices.

Developing situation aware applications is challenging. Collecting, representing, and handling context is crucial to the success of such kinds of applications. The limitation of context data acquisition in situational awareness may negatively affect the accuracy of inferred data, especially because of incomplete and incorrect context information, which in turn leads to misperception and incorrect application adaptation behavior.

As a result, we propose a situational awareness model capable of identifying a predetermined set of situations with higher accuracy compared to existing approaches. More relevant sensation available on current handheld devices will be used, along with a more concrete understanding on how such sensation can be used to identify situations. A thorough survey is conducted that will allow the researchers to better utilize the added sensation to enhance the accuracy of situation identification. The situation awareness model will establish integration between situation templates [8] as means of representation and the evolve-pool-collaborate model [4] as means of reasoning.

Employing situational templates [8] as an approach to represent various context scenarios will result in a machine readable description of a specific situation that can be used across multiple applications. In addition, utilizing the evolve-pool-collaborate model [4] will result in a reduction of cost of reasoning through the reuse of situation templates and sharing of context information across multiple handheld devices.

Finally, we propose imposing a habitat sensitive implementation to our model and as a result a comprehensive survey was conducted to capture the preferences of mobile device users in a defined habitat/geographical region, results of the survey will be shared in a further section.

This paper is structured as follows, section 2 reviews related work. In section 3 we detail the proposed model, and in section 4 we illustrate the survey methodology. Section 5 showcases the results of the survey, while section 6 concludes the paper.

2. Related work

In this section, we introduce the notion of context awareness, as well as ways by which context is sensed modeled, and architected. We also introduce the idea of situation templates that we will use in our research, as well as Darwin Phones that we will use for evaluation.

2.1 Context awareness

A system can be defined as context aware if it has the ability to extract, interpret and use context data to adapt its functionality to the current context of use [20]. In addition, any context aware application may support three general kinds of features: 1) Presentation of information and services to a user, 2) Automatic execution of a service and 3) Tagging context information for later referencing [1].

Context aware applications normally use various sensors to infer the user's activity. One of the main obstacles facing applications of the sort is the methodology of detecting the appropriate context form noisy and ambiguous sensor data. Any form of adaptation as a result of context awareness must be relevant to the user's expectation, or otherwise will be considered as intrusive behavior. As a result, sensor data which acts as the seed for any context aware application must be accurate to avoid any misinterpretation [2].

2.2 Sensing context

Sensors provide means to collect data or information about the physical world. This knowledge provides computer systems with a mechanism to infer actions most suitable for the physical situation at hand. Figure (1) showcases various sensor types and their possible attributes. A combination of multiple sensors may reveal more information for the computer system to reason with, constructing a more comprehensive and accurate image of the physical world [7][10].

Gathering sensor data can be done through a number of ways illustrated in [9][11] as:

- Device-databases (e.g. calendars, to-do-lists, address books and profile information).
- Direct input to the application running (notepad and taking notes).
- Active environments (active badges, IR-networks and cameras).
- Sensing context using sensors (Sensor Badges, GPS, cameras, microphones, etc.)

Sensor Type	Possible Attributes
Light Sensors	Ambient light, indoors brightness, outdoors, …
Accelerometers	Motion, vibration, physical state, …
Video Cameras	Facial expressions, gaze, behavior, presence, …
Microphones	Speech, noise, music, decibel levels, …
Motion Detectors	Presence, single or multiple users, …
Pressure Sensors	Pressed, occupied, hand gestures, …
RFID	Location, activity, situation, ...
GPS	Location, …
Environmental Sensors (temp, humidity, etc.)	Weather and other environmental conditions
Event Monitors	GUI events, schedules, notifications, errors, updates, ...
Action Listeners	
Data Loggers	
Agents, Processes	

Figure (1) List of possible sensors [6]

2.3 Modeling context

This refers to the process where raw sensor data is modeled to reflect physical entities in an environment which could be manipulated. This modeling intends to build knowledge concepts from the information provided by the sensors in an environment. Significant information should be interpreted from raw data attained by the sensors.

According to [5] low level context data provided by sensors may be modeled to high level context information by a number of methods:

- *One-to-One:* when one low level context value matches one high level context dimension. The sensed aspects of a context are combined and compared with a model to provide a value.
- *Context Fusion:* When several low level context values match one high level context dimension.
- *Context Fission:* When one low level context value matches several high level dimensions

2.4 Context Aware Architecture

Context aware systems should be composed of a number of elements handling the following issues: Context Discovery, Context Management, Context Representation and Adaptation.

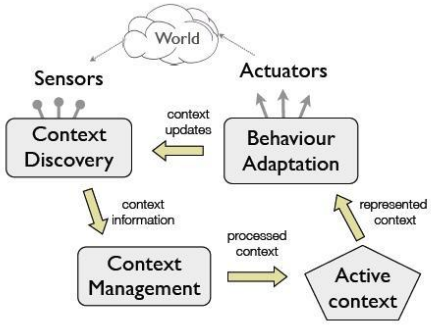


Figure (2) Context aware system architecture [3]

Cádiz et al. [3] proposed a context aware architecture which addresses the elements above, creating a foundation for the design context aware systems.

- *Context Discovery:* Responsible for gathering raw context data from different sensors and translating this data into information of useful nature.
- *Context Management:* This component is responsible for conflict resolution and processing raw data into consistent meaningful information.
- *Context Representation:* This component focuses on grouping context information of similar nature together.
- *Behavior Adaptation:* In charge of adapting the application's behavior to meet the users' needs in a constantly changing environment.

2.5 Situation Templates

"A situation template is an abstract, machine readable description of a certain situation type, which could be used by different applications to evaluate their situation" [8]. A situation template is composed of an accurate description of context information considered relevant to a given situation type and means to infer the existence of a solid situation from given data or values. Figure (3) shows a clear representation of the situation template model.

Each situation template is part of a larger template library which could be referenced by various applications to evaluate their current situations.

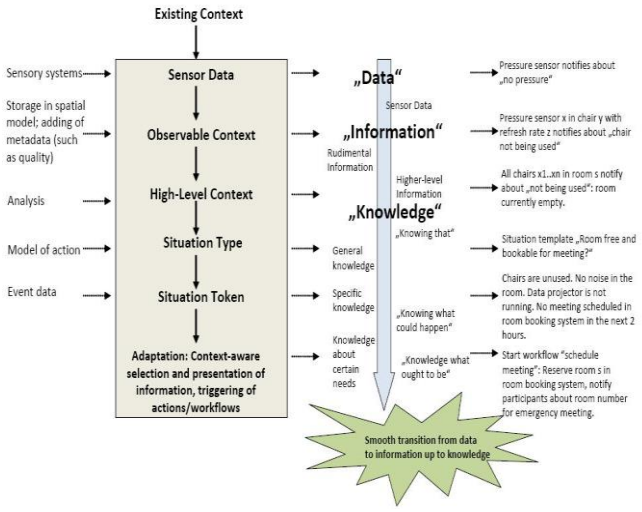


Figure (3) Situation Templates [8]

The work in [8] describes a *situation type* as the characterization of an explicit, recurring circumstance or event in the real world that could be expressed in an idealized way and that acts as an evaluation method for the adaption process of context-aware applications. The idealized description of those certain conditions including different parameters considered as relevant, their thresholds, and instructions of how to infer the existence of a situation from these parameters is referred to as *Situation Template***.**

Situation templates separate between two uncertainty metrics probability metric and confidence metric to address the issue of uncertainty of context ranging from sensor values, inferred context and situation recognition process.

- *Probability Metric:* a value, which represents the probability of the occurrence of the situation from the recognition-process view. A Higher value indicates a higher assumption of the occurrence of a certain situation [8].
- *Confidence Metric:* A normalized value between zero and one, which reflects the quality or the correctness of the used situation-template. The higher value indicates that the template will detect the situation with high levels of correctness [8].

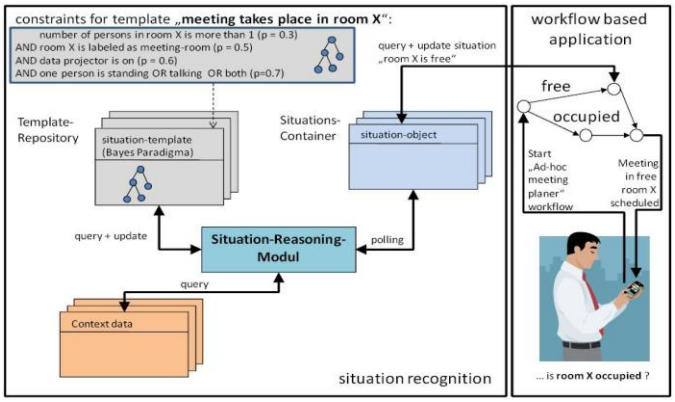


Figure (4) Situation Recognition Component [8]

Figure (4) illustrates how the situation recognition component is integrated into the existing model's infrastructure. An example of a predefined situation template is shown to infer a situation "whether a meeting takes place in the room X" or not. All predefined templates are stored in a Template-Repository. Each situation maps to a specific situation object that is attached to a situation template.

The user is using a simple meeting planer application that initiates a situation aware workflow. After the workflow is initiated, it looks up the situation of all available meeting rooms and reserves the first free room. The user is able to input feedback regarding the correctness and accuracy of the situation recognition component [8].

2.6 Darwin Phones

2.6.1 The Model

The Darwin Phone presents a collaborative reasoning system that utilizes sensors available on a mobile device to automatically infer various aspects of a person's life while achieving better accuracy and scalability, at lower cost to the user [4].

Darwin combines three different computational steps to achieve its goal as seen in figure (5): 1- Evolution, 2- Pooling and 3- Collaborative Inference.

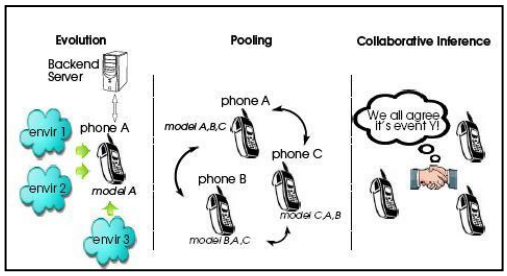


Figure (5) Darwin Model [4]

- *Classifier Evolution:* Is an automated approach using self-evolving classification models over time such that the classifiers are robust to the constantly changing sensing conditions common to mobile phones.

- *Model Pooling:* A novel approach that allows mobile phones to exchange classification models if the model is obtainable from a different phone, thus, allowing mobile phones to promptly increase their classification capabilities; that is, if a given mobile phone does not have a model for a certain situation there will be no need to recreate it, it can merely reuse another phone's models.

- *Collaborative Inference:* combines the classification results from multiple phones to reach a better, more accurate inference with high confidence in the sensing result.

2.6.2 Darwin operations

Step 1: Every Mobile device creates a number of models for the events to be sensed in the seeding phase. As time goes by the seed models are used to gather new data and evolution takes place. The reason for this step is that, by increasingly gathering new samples, the models will recruit data in different environments and hence, become more robust to environmental variations [4].

Step 2: When a number of mobile devices are in close proximity they exchange their models so that each device will contain its own original set of models in addition to the co-located device's model. Different devices may share their knowledge base enabling them to undergo larger classification tasks. For example: in the case of speaker recognition, enhancing performance from recognizing only the owner of the phone to recognizing all the people around in conversation with the same application [4].

Step 3: Collaborative inference exploits this diversity of different phone sensing context viewpoints to increase reasoning capabilities by utilizing all the multiple sensation of the different mobile devices [4].

3. The Proposed Model

3.1 Background

A common off the shelf smart phone contains a multitude of physical and virtual sensors. As previously identified, most current situation aware applications fail to utilize much of the available sensing capabilities on a given mobile devices. The under utilization of sensor data can produce incomplete and incorrect context information which may lead to misperception and incorrect adapting behavior. Additionally, situation aware applications may not consider that preferences of users may differ from one geographical location to another.

As a result, this paper proposes constructing a situation aware model that utilizes a large number of available sensors on an off the shelf smart phone to identify the occurrence of predefined context scenarios. The model will enhance situational awareness with habitat sensitive information. In the sense that context perception will consider variations in context meaning throughout different geographical regions and context reasoning will be influenced based on the preferences of users in a specific geographical area.

3.2 Method of Implementation

As a first phase five context scenarios will be defined: Meeting, Driving, Party, Sleeping and Movie Theatre. Each context scenario will be inferred by gathering a set of parameters that identify a situation. The parameters of each situation are viewed as the set of multimodal sensor data that is obtained from various sensors found on an off the shelf smart phone. The parameter set will be determined by a detailed survey conducted to gather the wisdom of the crowd leading to the identification of such situations.

In order to understand the wisdom of the crowd a localized survey was conducted with 155 subjects as part of the human perspective context acquisition to determine the users' preference and experience in the case of each context situation. Each participant identified the sensor attributes with the highest priority in relevance to each one of the predefined context scenarios. The context model was complemented with this localized survey to reach a habitat sensitive approach, in the sense that the context reasoning will be based on the preferences of users located in a specific geographic region.

The proposed model will implement the concept of situation templates described in [8] as the means for representing a specific event. Where, a situation template is a certain situation type holding different thresholds and their coherences for automatically spotting the existence of a particular situation. Different situation types will be available for each of the 5 predefined scenarios.

This work also intends to utilize the evolve-pool-collaborative inference model previously mentioned with some variations [4]. We intend to make use of both the classifier evolution and the pooling modules to complement our research. "The classifier evolution is an automated approach to up-dating models over time such that the classifiers are robust to the variability in sensing conditions common to mobile phones" [4]. As for the pooling module it is implemented to reuse context models already defined on other phones. "With pooling, mobile phones exchange context models whenever a model is available from another phone, thus, allowing mobile phones to quickly expand their classification capabilities" [4].

A confidence metric will be constructed where confidence metric is described as a normalized value between zero and one, which reflects the quality or the correctness of the used situation-template.

4. Survey design and implementation

This survey was conducted to gather the user's experience and preferences in each of the predefined context scenario. The survey was divided into 5 phases where participants identified the most relevant sensor attribute that would best describe a given context situation.

4.1 Survey objective

The primary research objective is to answer the major research questions and validate or negate the alleged hypothesis. The research was in the form of a survey/questionnaire as a research instrument. Primary targets for data collection were fellow researchers, random mobile users, experts in the field of mobile computing, telecom equipment suppliers and mobile operators.

4.2 Survey major research questions

1. What are the factors that indicate the occurrence of each one of the following context situations:
    a. Meeting
    b. Party
    c. In Car
    d. Movie Theatre
    e. Sleeping
2. What is the level of importance of each factor?
3. What are the parameters of each factor that would specify the previously mentioned activities?

4.3 Survey research method

This type of research is simply an analytical predictive research. The research involves collecting and analyzing data, applying statistical tests, survey/questionnaire was used to collect numerical data and hypothesis testing is required, therefore, the used process for this research is quantitative.

4.4 Survey sampling method

The survey conducted implemented two different methods of sampling to collect data from the proposed target segment:

- *Snowball Sampling:* Uses recommendations to find people with the specific range of skills that have been determined as being useful. In this method study subjects recruit future subjects from within their own contacts.

- *Stratified Sampling:* The population is divided into homogeneous subgroups before sampling. Each subgroup from the population will produce representatives.

4.5 Initial Hypotheses

Figure (6) shows the initial set of hypotheses used in this research:

Activity	Context Parameters	
Party	• **Sound Level:** High • **Sound Type:** Noise/Music • **Motion:** High • **Light Intensity:** Low • **Day:** Weekend	• **Time:** Night • **People in Proximity:** High • **Absolute Position:** Regularly Changing • **Phone Ringer:** High • **Location:** Irrelevant
Meeting	• **Sound Level:** Medium • **Sound Type:** Speech/Conversation • **Motion:** Low • **Light Intensity:** Low • **Day:** Weekday	• **Time:** Morning/Afternoon • **People in Proximity:** Low • **Absolute Position:** Stable • **Phone Ringer:** Vibrate/Silent • **Location:** Work Place
In Car	• **Sound Level:** Medium • **Sound Type:** Engine • **Motion:** Medium • **Light Intensity:** Irrelevant • **Day:** Irrelevant	• **Time:** Irrelevant • **People in Proximity:** Low • **Absolute Position:** Rapidly Changing • **Phone Ringer:** Irrelevant • **Location:** Irrelevant
Theatre	• **Sound Level:** High • **Sound Type:** Speech/Noise/Music • **Motion:** Low • **Light Intensity:** Low • **Day:** Irrelevant	• **Time:** Theatre display schedule • **People in Proximity:** High • **Absolute Position:** Stationary • **Phone Ringer:** Irrelevant • **Location:** Irrelevant
Sleeping	• **Sound Level:** Low • **Sound Type:** Irrelevant • **Motion:** Low • **Light Intensity:** Low • **Day:** Irrelevant	• **Time:** Night • **People in Proximity:** Low • **Absolute Position:** Stationary • **Phone Ringer:** Silent/Vibrate • **Location:** Home

Figure (6) Initial Hypothesis

5. Results

This section illustrates the knowledge of the crowd obtained through the comprehensive survey conducted among 155 subjects from different backgrounds and age groups. The aim of the survey was to gather the users' preferences and experience in dealing with each of the predefined context scenarios, where each subject answered the major research questions by indicating which of the identified sensor attributes has the highest priority in relevance to each one of the predefined context scenarios. These results will be integrated to the inference module of our proposed model forming a habitat sensitive situation aware model.

Each of the following figures shows a graphical representation of the users' preferences to each context scenario and its defined sensor attributes. The figures from (7-11) identify the top two values in terms of user preference for each sensor attribute in each of the predefined context situations.

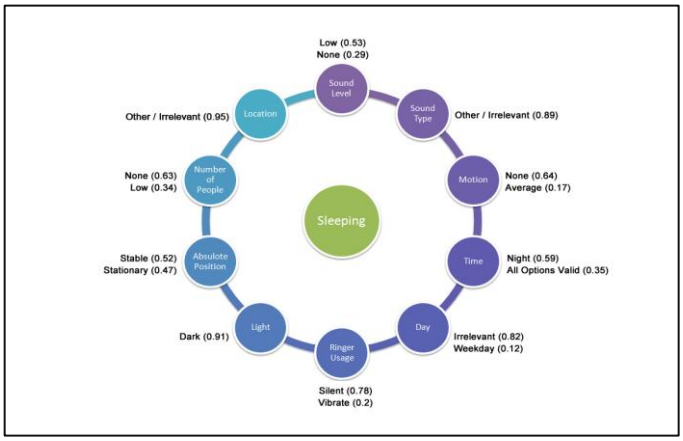


Figure (9) Sleeping

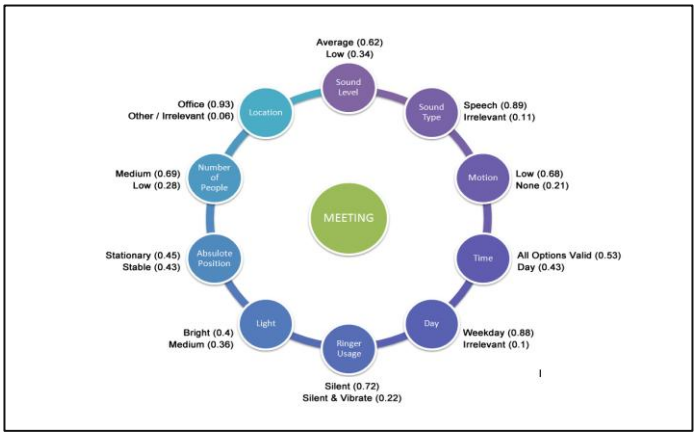


Figure (7) Meeting

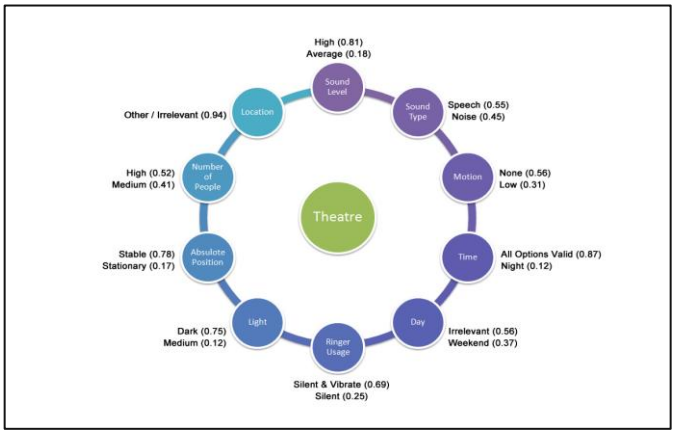


Figure (10) Movie Theatre

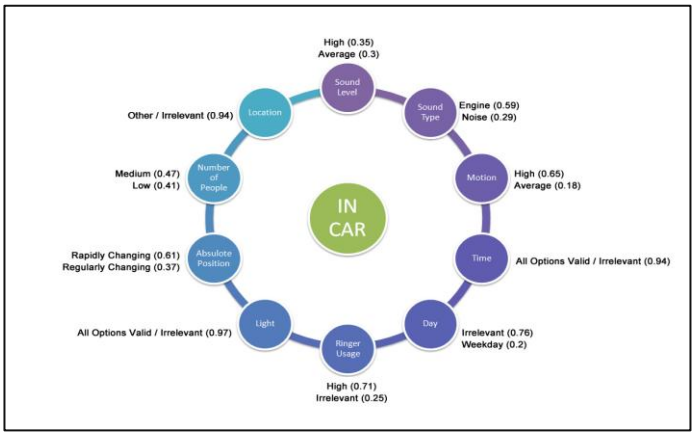


Figure (8) Driving

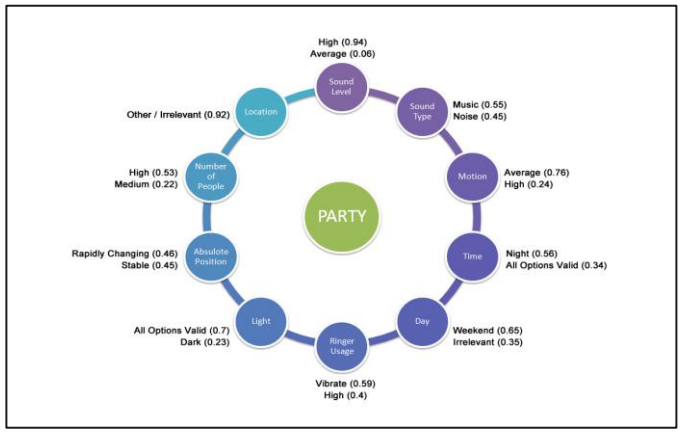


Figure (11) Party

6. Conclusion

This research aspired to construct a situational awareness model that makes use of a fairly higher amount of sensation which can be found on off the shelf smart phones. The proposed model aims at using ten various virtual and physical sensors to accurately represent the occurrence of a specific predefined context situation. A quantitative research approach was employed using surveys as a tool for data gathering to understand the wisdom of the crowd. The conducted survey expressed the user's experience regarding the context parameters that identify a certain situation. The survey was applied to 155 subjects generating statistical data that varied in some cases from the initial activity hypothesis; the results of the survey will be used to complement the researcher's initial assumptions to reach a minimal collection of context parameters. The research also proposed an approach that compliments situational awareness with habitat sensitive context reasoning. In the sense, that context inference will consider variations in the meaning of context in different geographical regions.

7. References

[1] A. K. Dey. "Understanding and using context". Personal and Ubiquitous Computing, Special issue on Situated Interaction and Ubiquitous Computing 5, 1, 2001.

[2] A. K. Dey and G. D. Abowd. "Towards a Better Understanding of Context and Context Awareness," In proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness, ACM Press, 2001.

[3] Alfredo Cádiz, Sebastián González, and Kim Mens. "Orchestrating Context-Aware Systems: A Design Perspective". In Proceedings of the first international workshop on Context-aware software technology and applications (CASTA '09), USA, 2009.

[4] E. Miluzzo, et al. "Darwin phones: the evolution of sensing and inference on mobile phones", In proceedings of the 8th international conference on Mobile systems, applications, and services, June, USA, 2010.

[5] Soylu, A., De Causmaecker, P., Desmet, P. "Context and Adaptivity in Context-Aware Pervasive Computing Environments," Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, 2009.

[6] Godbole, A.; Smari, W.W. "Human Perspective Based Context Acquisition, Learning and Awareness in the Design of Context Aware Systems," Military Communications Conference, 2006.

[7] H. Gellersen et al. "Multi-Sensor Context-Awareness in Mobile Devices and Smart Artifacts," Mobile Networks and Applications (Monet), vol. 7, no. 5, pp. 341–351, 2002.

[8] Häussermann, et al. "Understanding and designing situation-aware mobile and ubiquitous computing systems - an interdisciplinary analysis on the recognition of situation with uncertain data using situation templates", In Proceedings of the International Conference on Mobile Ubiquitous and Pervasive Computing, pp. 329-339, 2010.

[9] Khedo, K.K. "Context-Aware Systems for Mobile and Ubiquitous Networks", International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006.

[10] Loke, S. *Context-aware pervasive systems: architectures for a new breed of applications.* Auerbach Publications. 2006.

[11] Schmidt, A. "Implicit Human Computer Interaction through Context", Personal Technologies Volume 4, pp. 191-199, June 2000.

Based on Chromatic Relief Technology to DIY a Wonderful Cartoon Material

PENG Ge-gang, LI Xin-yu, SONG Ying, XIANG Li-sheng, SHEN Qing
(Talkweb Information System Co.Ltd., Changsha, Hunan 410205, China)

Abstract: Relief is a combinative product of sculpture and painting. Artists can carve out works with 3D perception in a plane, and gorgeously affix them on the surface of the wall or a container. So it is widely used in architectural design or appliance binding. We aim our project to such a long history and strong artistic tools to re-create cartoon materials. Based on a self-designed adaptive chromatic fusion and image project algorithm, the software transforms a conventional digital image into a pseudo 3D colorful relief (much like a sculpture), with visual sense of hierarchy. Users can DIY a wonderful cartoon material him-self. It appears a unique visual effect and endless possibilities, that is "a picture is worth a thousand words".

Key words: Relief, Digital Cartoon, Image Project, Digital Image Processing

1. INTRODUCTION

Recently, the "Plan of Promoting Culture Industry" issued by the State Council is the first special plan in the culture industry of China. It marks the culture industry has become a national strategic industry. The "Plan" lists the cartoon industry as one of the culture industry categories which have been greatly developed by the government years before. It provides a rare opportunity for the development of cartoon industry.

At the same time, the Ministry of Culture has established a national award named as "cartoon award". Which enjoy an equal status as the famous Wenhua Award of China. This award covers caricature, animation, new media and many other categories of arts.

These two pieces information above excited the cartoon field, which described a clear signal that the state is encouraging and paying a great attention to the cartoon industry. We can reasonably speculate that cartoon industry in China might enter a period of rapid development. We explore our research creatively and hold our effect firmly is a good choice indeed [1,2].

Relief is an ancient art which sculptures some fluctuant images on a plane. It combines and inherits the creation techniques of carving and painting, and it combines a 2D virtual painting and a 3D round-sculpture on a small and compressed space. Consequently, it is a high difficulty creation [3].

2、TECHNICAL COURSE

The traditional relief is carved out of materials such as marble, granite, boxwood, etc. so it usually presents a single color tone. On the other hand, colorful photos present brilliant colors. If we re-create a pseudo 3D relief from colorful digital photos, but abandon those beautiful colors and only leave a Monochrome sculpture, it may become a pale and dim art. So we decide to set a software transforms those conventional digital image into a pseudo 3D colorful relief (much like a sculpture), with visual sense of hierarchy.

We formed a technical way as below: design a self-designed adaptive chromatic fusion algorithm, and an image project algorithm to transform a conventional digital image into a pseudo 3D colorful relief and with a hierarchy visual sense. By such a software, a mobile user may DIY a wonderful cartoon material (character or background) them-self with a result of unique visual effect.

2.1 Parameters Setting

To meet the different needs of DIY users, we set a number of unique parameters aside the program. They including: keeping colors, projection directions, carving depth, (two) color sample points (for the integration operation later), color integration interval, color differential degree and so on.

1. Keeping colors (keepingColors). The goal of these parameters is cover a thin layer of pseudo-color onto the relief surface. There are four selections: None,

Red, Green and Blue. The meaning respectively: without Keeping color and cover pseudo-color with red, green or blue.

2. Projection direction (projectDirection). Choose a different incident light direction can result a different hierarchy projection effect. There are 8 selections: RightTop, Right, RightBottom, Bottom, LeftBottom, Left, LeftTop and Top. For example, when projectDirection = RightTop means the incident light is setting on the right-top corner and results a shadow on at left-bottom corner.

3. Carving depth (carvingDepth). The greater the carvingDepth is, the higher 3D visual perception.

4. Colors integration sample points (colorIntegPoint). Color fusion (or integration) is the key point to obtain chromatic relief effect on this project. WYSIWYG operation makes user can choose two sample points for the color integration later. Such integration points used for the program to detect and get relative RGB values, and applied them on the full-image. In this way, our colorful relief algorithm forms some color fusion zones to achieve seamless fusion between the samples' pixels and the other pixels. As a result, the colorful relief achieves a spectacular visual effect when the traditional relief only has gray color! Appended figures later show different sceneries and portraits as a chromatic relief's work. All the relief blends well with the original background.

5. Color integration interval (colorMixArea). The larger this value, the larger the color blending area spread by the color sample.

6. Color differential degree (colorDiff). This value decide the differential degree between the source point and the current processing points when color integrating. Based on the color differential between the "source point" and "processing points", we can project pseudo 3D hierarchy representations of the "source point" onto "processing points" to get a result of hierarchy visual sense by applied formula (1), (2) and (3) below.

7. An alternative option of gray relief or chromatic relief. If a "Monochrome relief" (gray relief) is chosen, only 256-level grayscale used to show a traditional relief. If "chromatic relief" is chosen, the picture may show you a wonderful scene that is a complex not only an overlapping carves but also of all hues landscape. Farther, a pseudo-color image can formed when the reservedColor parameter choose Red, Green or Blue.

2.2 Operations to form a relief

According a self-designed adaptive chromatic fusion and image project algorithm, we can transform a conventional digital image into a pseudo 3D colorful relief, with hierarchy visual sense.

We refer the current being processed point in the image as a "processing point", and its coordinates named as $x0$, $y0$. Contrast, an image point is on the incident light direction and produce a project to the "processing points" is referred as a "source point", and its coordinates named as $x0 + xShift$, $y0 + yShift$. Here, the value of xShift and yShift is proportional to the value of carvingDepth and the symbols of xShift and yShift is related with the parameter projectDirection. For example, when set the projectDirection to LeftTop, that results in xShift=carvingDepth and yShift =－carvingDepth。

Preparation: As said above, according to the optional value of carvingDepth and projectDirection, we can get the value of xShift and yShift; Based on the position of 2 sample pixels (colorIntegPoint), we can get blue, green and red values of them, they are denoted as b1, g1, r1, and b2, g2, r2 respectively, and they are used in step 3 and step4.

Processing:

Step 1. Set the processing point coordinates at the initial upper left corner of the image and repeat the following steps 2 to Step 8, row by row and line by line, to push forward the procedures.

Step 2. Get the blue, green and red values of the "processing point", and denoted as b, g, r;

Step 3. If abs(b-b1), abs(g-g1) and abs(r-r1) are all less then colorMixArea, then take b1, g1, r1 as color output of the "processing point", go to step (7). Here, abs is an absolute value function.

Step 4. If abs(b-b2), abs(g-g2) and abs (r-r2) are all less then colorMixArea, take b2, g2, r2 as the color output of the "processing point", go to step (7). Here, abs is an absolute value function too.

Step 5. According the coordinate of "source point" $(x0 + xShift,\ y0 + yShift)$ get the pixel data (blue, green, red values), and denoted as b3, g3, r3. Get the pixel data of "processing point", b, g, r, and do differential operation as formula (1),(2) and (3) to get b4, g4 and r4.

b4= abs(b+colorDiff-b3) (1)

g4 = abs(g + colorDiff-g3) (2)

r4 = abs(r +colorDiff-r3) (3)

Step 6. If the option of keepingColors 为 None, then take b4, g4, r4 as the output of the "processing point". If the option of keepingColors 为 Blue, then take b, g4, r4 as the output of the "processing point". If the option of keepingColors 为 Green, then take b4, g, r4 as the output of the "processing point". If the option of keepingColors 为 Red, then take b4, g4, r as the output of the "processing point". Then go to step 7.

Step 7. If the X coordinate plus 1 is less than the width of image still, then turn to step 2. Otherwise, set the X coordinate to zero, and Y coordinate plus 1, turn to step 8.

Step 8. If the Y coordinate plus 1 is less than the height of image still, then turn to step 2. Otherwise, this program is stop.

The practical effect of this project is shown in figures below:

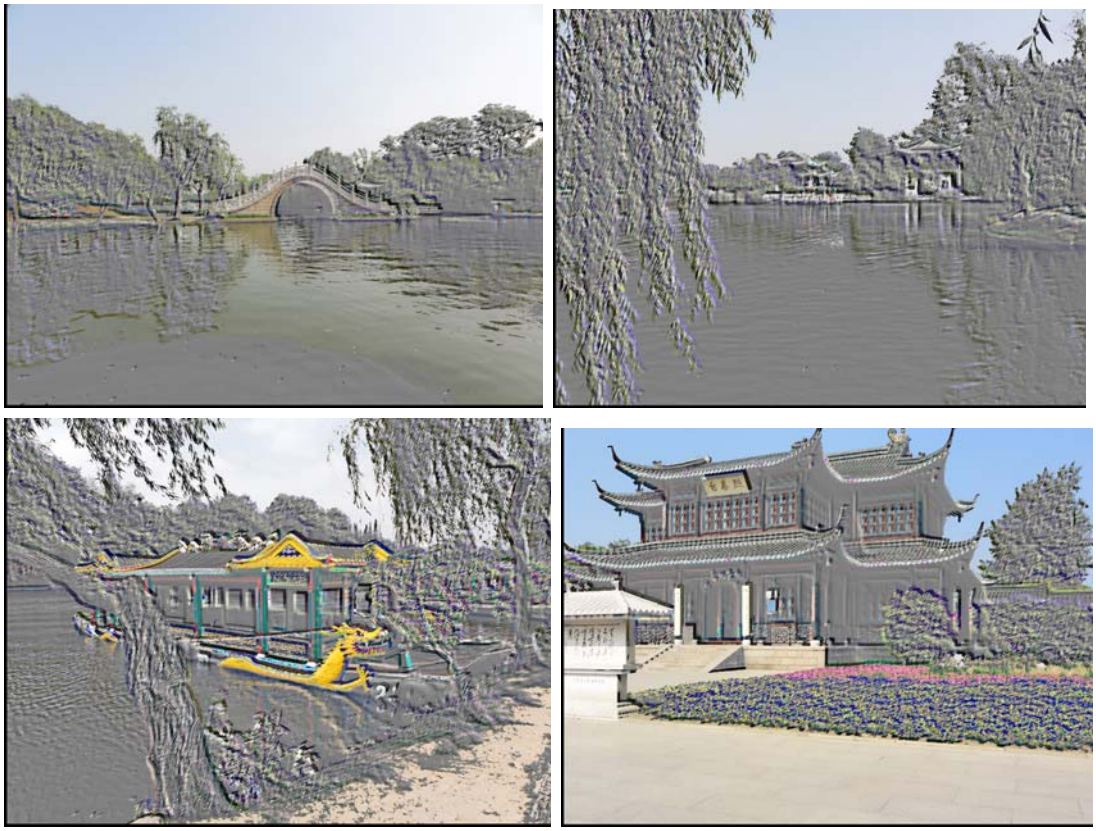


Fig 1 Pseudo 3D colorful relief comes from digital images

(a, b, c) our results

(d, e) the source image

Fig 2 Pseudo 3D colorful relieves

(a) Result of PhotoShop (b, c) our results (d) the source image

Fig 3 Hierarchy monochrome relieves

3. CONCLUSION

Through the research of this subject, we can find that unexpected effect can be received through selecting and retaining the regional color in original image. All results in Fig 1 are all processed from digital photos and they all have dual appreciation effect of relief and scenery postcard; 2(d) and 2(e) are the original image of 2(a,b) and 2(c). Compared with 2(c) and 2(d), 2(a,b) and 2(c) has a different feeling indeed. Without this article, I am afraid it is hard to imagine how picture 2(a,b) was formed. Please compare 2(a,b) with 2(d) and pay attention to the 3D shadow of the arms in 2(a,b). There is no shadow in the original picture.

We are glad to say that pictures processed by our independently designed technology and algorithm have the similar effect of Photoshop, a well-known image processing software of Adobe company. Picture 3(a) and 3(b) respectively represent the two technologies for comparison. Further comparison and identification are welcome, as well as inquiries and acquisitions of source code from the company the authors' work for.

We hope to get criticisms and corrections on deficiencies in this article. Thank the national science and technology support plan and the State Ministry of

culture for supporting this project.

References

1. Li Xinyu, Song Ying, Xiang Lisheng, Shen Qing, Based on Boundary Location Technology to Separate an Image Semi-automatically, ICWN'11.

2. PENG Ge-gang, Li Xinyu, Song Ying, Xiang Lisheng, Shen Qing, Based on Image Twist Technology to Create a Funny Mobile Cartoon Character，ICWN'11.

3. Lu Ying，The charm of relief，Family of the Drama，2010.5

Authors

Peng Gegang: Peng is a post-doctor in computer applictation technology.He is a senior architectural designer and interested in systemarchitecture, mobile communication, etc. Email: penggegang@hotmail.com

Li Xinyu： Li is the general manager of Talkweb Information System Corp. He is a senior architectural designer and received a BS degree in Computer and Information Engineering from HoHai University, Nanjing in 1988. He is interested in system architecture, mobile communication, etc.

Song Ying: Song is the vice general manager of Talkweb Information System Corp. He received a BS degree in Chemistry from China University of Geosciences, Wuhan in 1989. He is interested in mobile communication, multimedia, etc.

Xiang Lisheng: Xiang is a vice technical supervisor in Talkweb Information System corp. He received a BS degree in computer science from Hunan University in 1996. He is interested in MM, animation, etc. E-Mail: 13974838381@hnmcc.com

Shen Qing: Shen received a BS degree in electronic instrument from Changchun Geologic Institute. He was a professor in the Institute of Computer Science at National University of Defense Technology, China. Now he is the first chief technical consultant of Hunan Talkweb Information System corp. He is interested in pattern recognition, AI, MM, animation, etc. E-Mail:sq1950224@yahoo.com.cn

SESSION

COMMUNICATION SYSTEMS: NOVEL APPLICATIONS, ALGORITHMS, AND SYSTEMS

Chair(s)

Prof. Hamid R. Arabnia

Proactive Channel Allocation for Ad-Hoc Networks

Yosi Ben-Asher[1] **and Yehuda Ezra**[1]
[1]CS Department, University of Haifa, Haifa, Israel

Abstract— *In this work we consider the problem of creating multi-edge channels in ad-hoc networks wherein each node can use multiple frequencies for sending and receiving packets by filling the slots of an OFDMA matrix. We consider the problem of how to fill-up the slots of a $frequency \times time$ matrix (FTM) at each node such that maximal collision free $k > 1$ path cover (CFkPC) of the communication graph is obtained. This problem is a variant of edge and node disjoint path cover in graphs extended to include collisions caused by the hidden terminal problem where two nodes A and B which are not in transmission range collide by transmitting at the same frequency and time to an intermediate node C. The proposed solution for filling the FTM is based on using maximum independent set in a suitable conflict graph. Our simulations compare between the proposed CFkPC approach and a version of the DCA protocol [12] that was extended to use multiple send/receive slots. The results show significant advantage of using CFkPC versus the adaptive approach of the extended DCA. We assume that GPS coordinate and GPS time synchronization is available.*

Keywords: Wireless, Ad hoc, multi-channel,routing

1. Introduction

We consider the possibility of sending and receiving packets in multi-frequency channels by the mobile nodes of Ad-Hoc networks. In Ad-Hoc networks, due to movements of users, the connections between nodes frequently change creating different topologies of the communication graph between the nodes of the network. This creates a challenging situation for suitable protocol that can decide, for a given node v, in which frequencies it should transmit and at what time slot such that minimal latency and maximal throughput of packets is obtained. We assume that all nodes use a GPS for location and perfect time synchronization. We model the use of multi-frequencies channels in ad-hoc networks by assuming that each node v needs to fill up slots in a $frequency \times time$ matrix M (called FTM) indicating whether it will transmit $M^v_{f,t} = send$ or receive $M^v_{f,t} = receive$ a packet at this $frequency \times time$ slot. FTM is repeatedly used by each node for sending and receiving packets and thus it is an abstraction of a strategy for using multi frequency channels. Depending on how the FTM at each node are filled, communication channels between neighboring nodes are generated. Thus if $u \longrightarrow v$ are two nodes in communication range and $M^u_{f,t} = send, M^v_{f,t} = receive$, then possibly u can transmit packets to v at this $frequency \times time$ slot. The channel created between u and v by setting $M^u_{f,t} = send, M^v_{f,t} = receive$ is subject to two problems:

- Frequent movements of u and v may move/remove u and v in/out-of communication range.
- Transmissions of other neighboring nodes at the same $frequency \times time$ slot can collide with packets sent by u (known as the hidden terminal problem).

The goal in this work is to find a strategy for filling $M^v_{f,t}$ such that we maximize the number of channels of length $k > 1$ while minimizing packet-lost caused by collisions (due to the hidden terminal problem [2]).

For the purpose of this paper a channel is an agreement between two nodes u and v to use a certain frequency to transmit packets from u to v in a given frequency for a certain period of time or another terminating condition. As such it is possible to generalize the channel from a single hop to a multi-hop channel where u sends packets to v through a sequence of nodes which are all part of the channel agreement. Consider for example a tree like communication graph that has seven nodes given in figure 1. Each node x, u, w, r, z, v, y has the ability to use two frequencies and two time slots (depicted by the 2×2 FTMs in figure 1). Following the previous discussion we would like to configure the slots (transmit/receive) such that multi-hop communication paths will be generated. These communication paths will allow us to pipeline packets of data streams to longer distances faster than what could be obtained had we used single edge channels. Practically, pipelining or streaming of messages implies that packets are transmitted along such a path without acknowledgment from the receiver. The communication paths that have been selected to cover the communication graph of figure 1 contain two paths $p0, p1$ each containing four edges.

We compare two strategies for channel allocation in ad hoc networks:
Extended Dynamic Channel Allocation (EDCA) is a strategy where by using a control channel with request-to-send (RTS) and clear-to-send (CTS) messages, a node u can negotiate with a neighboring node v for a slot $M^{u/v}_{f,t}$ that is currently free. By snooping on other requests on the control channel, both u, v can verify that $M^{u/v}_{f,t}$ is not used by any neighboring node of u or v (possibly preventing collision with other nodes). EDCA is an extension of a well known protocol DCA that have been extensively used in other works. Another extension of DCA is the multi channel

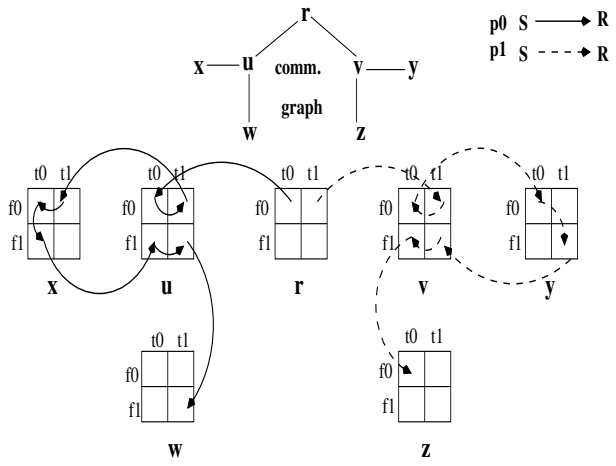


Fig. 1: Configuring $frequency \times time$ slots to create multi-channel paths for streaming of packets

DCA [12] which allowed a node dynamically jumps from one frequency to another according to the traffic of the network, but still do not allow transmitting or receiving on multi channels simultaneously. Thus, in this work we extend the original DCA algorithm [15] to handle receiving and transmitting packets through multi set of channels at same time slot.

Proactive Channel Allocation (PCA) is what we propose here, a strategy based on finding a path-cover by a maximal number of paths of length k such that each path passes through $2k$ transmit/receive slots of the FTMs available at each node. For example the path-cover depicted in figure 1 is a solution to the coverage problem for $k = 4$ as follows:

- There is no legal cover with more than two paths of length $k = 4$ (must be verified by checking all possible cases). This is also a maximal cover since no path, even of length one, can be added to it.
- Each path forms a legal scheduling for pipelining a stream of packets traversing along this path edges (as can be seen there are edges that seems to go back in time but this will be explained later on).
- There are no interferences between the paths, i.e., no node can receive more than one packet at any f_i, t_j slot. For example, node v is in communication range with nodes r, y, z and is scheduled to receive a packet at $f0, t1$ and at $f1, t1$. Indeed only node r broadcasts at $f0, t1$ and only node y broadcasts at $f1, t1$.

As opposed to the EDCA approach the channels that are created by these paths contain more than one edge allowing forwarding of packets to longer distances without dynamically checking which neighbor is "free". In addition unlike the EDCA, for PCA, the FTMs are filled once for a specific graph topology regardless of the communication sessions that are currently going on.

Intuitively the PCA method proposed here significantly differs from current techniques to perform routing and channel allocation in ad hoc networks as follows:

1) It creates multi-channels containing more than one edge, actually transforming the communication graph to a hyper-graphs which is a graph whose edges connects more than two nodes. This stand in contrast to current techniques that create channels only between neighboring nodes. By the term channel we prefer to a group of nodes that agrees on a set of frequency/time slots through which data-packets can be broadcasted.
2) We focus on streams rather than on individual packets, we thus assume that a channel, once created, will be used to transfer a stream of packets. Moreover, we assume that pipelining a stream of n packets through a channel of length k in a pipeline mode is faster than $n \cdot k$ namely sending each packet through k distinct channels. Thus, by creating channels of length $k > 1$ we support pipelining of a stream of packets where packets are being sent along the edges of the path without acknowledgement.
3) Further, all current methods for routing in ad hoc networks use adaptive and dynamic creation of channels tracking the frequent changes in the topology of the communication graph. In contrast, PCA samples the communication graph to create channels and attempts to use them for a relatively long time regardless of the frequent changes in the topology. There are two reasons as to why a sample "road-map" will be useful for a relatively large time before an update of it is needed:
   - There are many changes caused by frequent movement that cancel one another, e.g., a node moves out of a channel but another may take its place listening and transmitting at the right frequencies and time slots.
   - The sampled "road-map" may be wrong about some of its roads (channels) but by attempting to use it a packet will travel at the direction it needs to go. Thus the sampled map is a good approximation of the true road-map that currently exits.

   Thus PCA creates an approximated "road map" which is likely to be useful in most cases even when the "roads" are frequently changing.

Some of the related works are as follows. Dual Busy Tone Multiple Access [7] is a method that divides a common channel into two sub-channels. Wu et al [15], propose MAC protocol called Dynamic The main drawback of this protocol concealed in the fact that RTS, CTS and ACK packets exchange are necessary for every pair nodes on every communication path. Jain et al [8], propose a protocol that achieves throughput improvements by intelligently selecting

the data channel, but also required RTS and CTS and ACK packets exchanged for every pair nodes as described in DCA. Finaly, [12] propose a protocol that uses all channels as data channels where nodes negotiate channels with their destination nodes during a time-window. In this window, every node must listen to the default channel. [14] proposed a Usage-Prediction Based Channel Allocation scheme for GSM networks.

2. The Problem of Finding a Collision-Free Maximal Path Coverage

The problem of finding a maximal FTM scheduling for a given communication graph with communication paths of length k is equivalent to finding maximal coverage of a graph by disjoint paths of length k.

Definition 2.1: Let FTG be an undirected graph resulting from expanding a communication graph G such that each node has been replaced by a clique of internal edges representing the FTM slots as depicted in figure 2. Each communication edge of G have been expanded to a set of external-edges connecting the suitable f,t nodes of two neighboring cliques. By definition FTGs are not general graphs and in particular each path in an FTG consists of alternating external and internal edges (see the FTG graph of figure 2 for an illustration). Let a $collision\ free\ k\ path\ cover$ (CFkPC) of an FTG be any cover of it's edges by a maximal number of directed paths $p_1, p_2, \ldots, p_n$ of length k ($k \geq 1$) such that:

- Each path p_i corresponds to some path of FTG.
- No node/edge is shared by two paths (node/edge disjoint paths).
- All the nodes of a path are distinct (i.e., paths cannot visit a node more than once).
- Each path should start with an external-edge and end with an external edge.
- Paths should not be connected by an external edge that connects two of their external edges. Thus if u, v are two external edges in two different selected paths then there can be no external edge in G that connects them (we later show that this is sufficient to prevent hidden terminal problems).

We can sharpen this definition to include paths of length smaller than k by grading a path-cover of G by a vector with k coordinates $grade = < x_k, \ldots, x_1 >$ where x_i counts the number of paths of length i in the cover and two grades are compared lexicographically. A maximum CFkPC of an FTG is a CFkPC with the maximal number of paths of length k or the one that achieves the maximal grade.

The CFkPC of graphs is similar to the multi-dimensional matching [1] and to $cover - by - disjoint - paths$ in graphs which are all NP-complete problems. Thus, CFkPC is likely to be NP-complete. Figure 2 illustrates how the communication graph with 3×2 FTMs at each node is extend to include the $M_{f,t}$ slots as nodes forming an undirected graph called FTG. The FTG is formed by:

1) For each node v, every slot in its FTM $M^v_{f,t}$ is made a node of FTG.
2) For each edge (v,u) and f,t slot add all edges $\left(M^v_{f,t}, M^u_{f,t}\right)$ (indicating a possible send-receive operation between v and u using this f,t slot).
3) Add an internal edge between every two slots of the same node $M^u_{f,t}$ and $M^u_{f',t'}$ where either $f \neq f'$ or $t \neq t'$ (indicating that a message received at $M^u_{f,t}$ will be sent at $M^u_{f',t'}$ or vice-versa).

A possible cover of the resulting FTG by paths of up to $k = 5$ edges is given at the upper part of figure 2. This cover, does not satisfy the hidden terminal requirement and several paths can collide, e.g., the receive in $M^u_{f0,t1}$ collide with the Send of both $M^w_{f0,t1}$ and $M^z_{f0,t1}$.

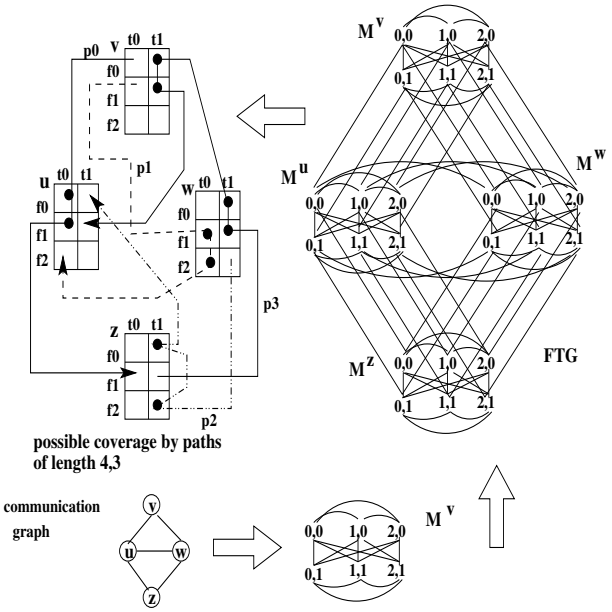


Fig. 2: Building the FTG of a communication graph and a possible $k - path - cover$ for it with 3×2 FTM

3. The algorithm for computing maximal CFkPC

We turn now to the algorithm for computing maximum CFkPC of an FTG. First we show that finding a CFkPC allowing collision free broadcasts:

Lemma 3.1: The conditions of definition 2.1 grantee that if we broadcast packets along the paths of a CFkPC then no two packets will collide.

Proof omitted due to space limitations.

Next consider a possible lower bound on the number of paths in the CFkPC of a given FTG:

Lemma 3.2: Let a conflict graph CG of a given FTG be a graph whose nodes correspond to the edges of the FTG and an edge is inserted between two nodes u, v if the corresponding edges $e_u, e_v \in FTG$ collide under the conditions of definition 2.1. It follows that the number of paths of length k in a maximum CFkPC of the given FTG is smaller than $\frac{|MIS(CG)|}{k}$ where $MIS(CG)$ is the maximum independent set of CG.
Proof omitted due to space limitations.

Finding an MIS of a graph [13] is well known and if we settle for maximal independent set [11] we can use fast distributed algorithms suitable for ad-hoc networks. Let G be a given graph representing a communication graph where each node uses an $F \times T$ FTM. The algorithm for maximizing use of multi-channels of length k uses the following steps:

1) Let each node internally maintains its part of the FTG by representing the FTM as a clique as described in figure 2. This requires that each node will identify its neighbors.
2) Compute conflict graph CG using the two rules of definition 2.1: 1) two connected edges of the same type and 2) three connected external edges. Note that this can be done distributively letting each node acquire information on conflicting edges from its neighbors.
3) Find maximum independent set (MIS) in the resulting CG. Statically this can be computed by an exact exponential algorithm [9], [13] or an approximated by a polynomial approximation algorithm for finding the biggest colored class as the MIS [10]. For a distributed algorithm that can be used in ad-hoc networks it is reasonable to use an algorithm for finding a maximal independent set (an IS that cannot be increased by adding any other node to it) to approximate the MIS. Distributed maximal IS protocols have been used in ad hoc networks extensively to find a connected dominate set in the communication graph [3], [4]. We select the algorithm of [11] that completes in $\log^*$ steps on unit-disc graphs (using GPS addresses as unique node ids). The size of the MIS is also computed and broadcast to every node.
4) Finally multi-channels can be computed using a simple protocol that grows paths randomly:
   a) Each node learns which edges of the FTG and the MIS resides in its neighbors.
   b) External edges of the MIS are selected at each node to be the head of a path. This is done with probability $\frac{k}{2|MIS|}$ so that most edges of the MIS are not selected and are marked as free-edges.
   c) For each path L whose head is currently at node v we randomly select an internal edge from the MIS that will continue it. Next we select the external edge from the MIS that will continue it to the next node u and send a continuing request for L to u.
   d) When a node u receives a continuing request for path L from v it checks to see if $< u, v >$ is a free external edge which is not used for any other path. If so u returns an acknowledgment to v and marks $< u, v >$ as used. Otherwise, u returns a fail message to v which ends the path L.
   e) This is continued for each active path until this path reaches a length of k or reaches a node for which there is no free edge in the MIS that can be added to it.
   f) A backward chain of messages from the last active end-point of each path L (either reached length k or terminated) is sent back through the active end-point of a path/channel to configure the FTMs at each node.
   g) There is a fixed time budget (order of k steps) after which each node assumes that this phase is over and packet routing is performed.

Figure 3 illustrates how the algorithm works. The input FTG is a grid of alternating levels of external and internal edges. The edges are marked with numbers $0-11$ and also by their type $'ex/in'$ for external/internal. Figure 3 right side contains the resulting conflict graph (CG) with nodes corresponding to the FTG edges $0-11$ and edges between any two conflicting edges. The MIS for the CG of figure 3 is marked by circled nodes. In order to create the paths (marked by dashed arrows) for this MIS, we start with edge-0 which can be continued only to edge-6 and finally end this path with edge-10. The second path starts with edge-3 and can be continued with two internal edges (edge-7) and (edge-08) out of which we cannot continue with edge-7 since edge-10 is not free so we attempt to continue it through edge-8. However edge-8 is an internal edge that cannot be continued by an external edge in the MIS so the resulting second path contains only edge-3.

For a given ad hoc network with n nodes, transmission radius r, F frequency channels, T time slots and field-size $L \times L$, we can use a maximal CFkPC of the communication graph as the base for an algorithm to send streams of packet over ad hoc networks. The algorithm is called PCA works as follows:

1) We globally set the value of k to be $2 \cdot \frac{field\_length}{transmission\_range}$ which is the maximal Manhattan distance between any two nodes assuming uniform distribution of the nodes in the given field (also assuming that there are enough nodes to form this density).
2) CFkPC is performed globally and the resulting paths are held in suitable routing table in every node. Thus every node has knowledge of the CFkPC paths that pass through it (including their length, start/end nodes and FTM slots that are used to realize them). Note that the CFkPC may include paths of length 1.

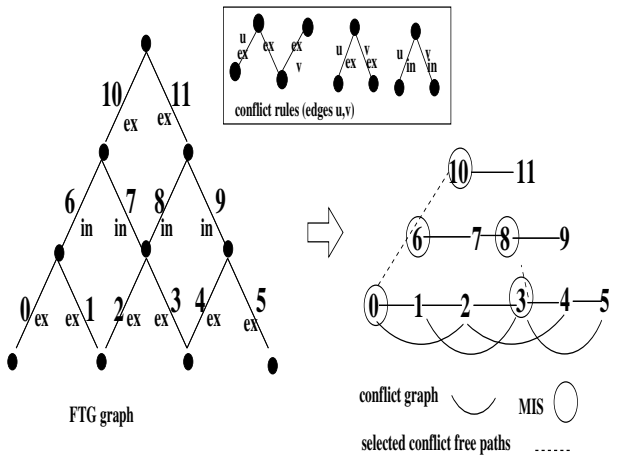


Fig. 3: Conflict graph and the rules to create it for a grid like FTG

3) Streams of packets are generated arbitrary in source nodes where each stream contains a relatively large number of packets which should be sent to the same destination.
4) At each cycle all nodes receive packets through their FTM slots and send them further as follows. The packet is sent to the next node in its current path if that node gets it closer to its destination. A packet will switch to another path if the next hop in its current path does not get it closer but there is another path in the routing table that will and that path is not used. Otherwise the packet proceeds in its current path if there is a next hop for that path. In case that the packet reach to the end of its path and there is no unused path through which it can continue it is moved to a waiting queue to wait until there is an unused path through which it can continue.
5) One slot is reserved for beacon messages where each node transmits its ID, its location, and a list of un-used F-T slots through which packets can be received "outside" the scope of the CFkPC paths. Each node constantly listen and collect beacon messages from its current set of neighbors and collects their un-used slots.
6) When a node u does not detect beacon message from a neighbor v for a certain period of time it assumes that v moved out of transmission range and the set of paths that passed through v are now "broken". Packets that needs to continue through a broken path are sent through an un-used slot to a suitable neighbor. If no un-used slot is available and no un-used path can be used packets from broken paths are added to the waiting queue.
7) A new neighbor v may move to the vicinity of a node u and is thus detected by u through its beacon message. In that case u and v will attempt to join their broken paths by connecting them together.
8) A packet may be dropped if its TTL expires, receiving queue is full, or if it collides at some node with another packet that have been transmitted at the same FT-slot.
9) The CFkPC is recalculated periodically every fix period of time called the re-calculation time.

4. Experimental Results

In here we discuss the experimental evaluation of the proposed PCA algorithm in comparison to the EDCA technique. We first summarize our assumptions: Time is divided to T frames such that for every time frame, F frequencies are available for use and none of the frequencies overlap. transmitted on different frequencies do not interfere with each other. Each host can listen or transmit on more than one channel at a time. Fixed-channel-bandwidth model is used. Each host is equipped with a single full-duplex transceiver [5]. Nodes are synchronized using GPS [6]. Simulations are performed in multi-hop networks scenario of varied nodes that are randomly placed in a 500m $\times$ 500m area. Source and destination nodes for streams are randomly chosen with probability 0.5. A node may be the source for multiple destinations and a node may be the destination for multiple sources. Each simulation was performed for duration of one minute. Packets have been partitioned to streams of $50 \ldots 250$ packets. Each data point in the result graphs is an average of 15 runs. Packet size is 512 bytes. Packet TTL was 10 implying that packets are dropped after 10 hops. Queues for receiving packets were limited to 50 packets while queues for transmission have unlimited capacity. We used geographic routing algorithm where nodes know their geographical coordinates and also their one hop neighbors coordinates. The parameters we vary are: speed of nodes, number of frequency slots F, number of time slots T, size of flows (in packets), transmission range and number of nodes in the network. We use the throughput performance metrics in our simulation:

$$Throughput = \frac{Packet\_Length * Num\_Of\_Received\_Packets}{Total\_Time}$$

where the throughput measures the rate in which packets have been received at their final destination.

It immediately follows that there are too many parameter values that should be considered, since each measurement of the PCA requires setting a value to: n number of nodes, r the transmission range, $L \times L$ size of the field, F number of available frequency slots, and T number of time slots. However, assuming uniform distribution of the nodes at any given time in the field, T can be approximated by $\frac{T \cdot F}{2} = \frac{\beta \cdot \alpha}{4 \cdot n}$ where $\alpha = \frac{\pi \cdot r^2}{L^2} \cdot n$ and $\beta = \frac{\gamma \cdot n}{2} \cdot \delta$. This formula is based on the following assumptions:

- α is the average number of nodes that are in communication range of any node (v).

- β is the approximate number of transmissions in the field such that:
  - γ is the probability that a node will start a stream of packets (0.5 in our experiments).
  - $\frac{n}{2}$ is due to the fact that each stream has two end points.
  - δ is the average number of hops a stream will follow and is approximately equal to $L/r - 1$ assuming $n > (L/r)^2$ and that streams travels in the shortest Manhattan distance between their end-points that have been selected at random.
- The term $\frac{\beta \cdot \alpha}{n}$ is the expected number of transmissions from the nodes that are in the communication range of v. Out of this number we assume that transmissions are sent equally to all four directions out of which only $1/4$ will be sent to v.
- Thus at maximum throughput each node v should have enough $F \cdot T$ slots to pass streams from/to all the nodes that reside in its transmission range yielding eq. 4.

The first experiment we consider is designed to verify the formula of eq. 4. We run the PCA algorithm with $F = 4$, $r = 100m$, $L = 500m$, $n = 60, 90, 120, 150$ and measured the throughput of PCA for $T = 2, 3, 4, 5, 6$ (with random way-point and 50 packets per stream). For this setting of eq. 4 we get that $\beta = \frac{0.5 \cdot n}{2} \cdot 8 = 2 \cdot n$, and $\alpha = 0.12 \cdot n$ yielding that $T \cdot F = 0.12 \cdot n$ and $T \cdot = 0.03 \cdot n$ The results in figure 4 show that for $n = 150$ the throughput ceased to improve for $T > 5$ which is indeed what eq. 4 determines ($0.0314 \cdot 150 = 4.71$). For $n = 120$ throughput in figure 4 ceased to improve for $T > 4$ which agrees with $0.0314 \cdot 120 = 3.76$. Similarly for $n = 90$ the experimental result is $T > 3$ which agrees with $0.0314 \cdot 90 = 2.82$ of eq. 4, and so is the result for $n = 60$ we have $T > 2$ which agrees with $0.0314 \cdot 60 = 1.88$.

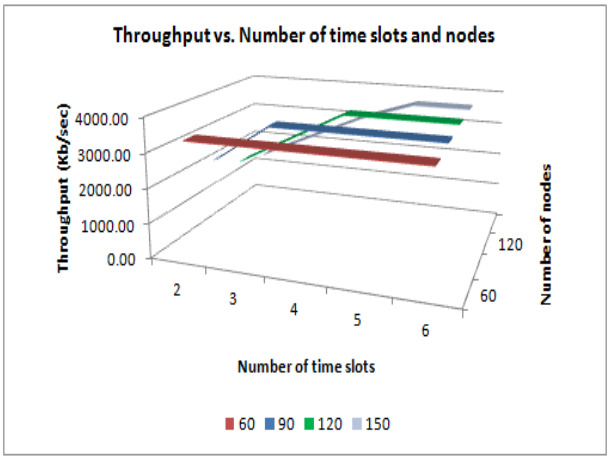


Fig. 4: PCA Throughput vs. Number of time slots T and number of nodes n.

Next we compare between PCA and EDCA selecting F, T values which are bellow the saturation point of eq. 4 $\frac{T \cdot F}{2} \leq \frac{\beta \cdot \alpha}{4 \cdot n}$. In this set of experiments we compared the throughput of PCA and EDCA for $n = 120$, $r = 250m$, $L = 500m, \#stream = 50packets$ and tested for several $< F, T >$ values. Figure 5 presents the results for $F = 4, 8, 12, 16$ and $T = 24, 12, 12, 6$. The results show that when at $F = 12$ EDCA cease to improve while the PCA continues to improve the throughput (about 40%) and maintain the channel utilization. This was also obtained for $\#stream = 250packets$ (Omitted due to space limitations).

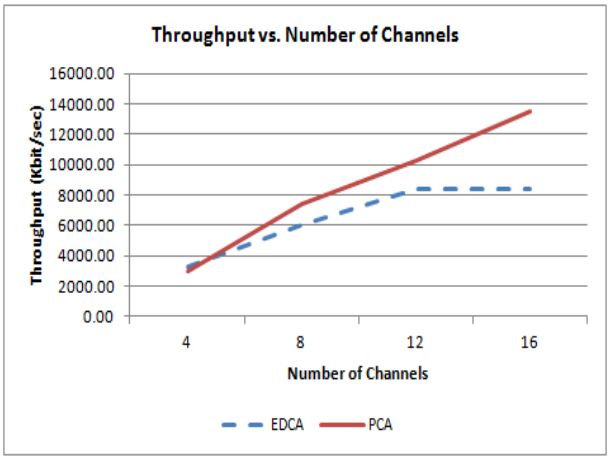


Fig. 5: Throughput vs. increased F, T values of EDCA and PCA with 50 packets per stream

The explanation for this behavior is that when the network load become very high, the control channel becomes a bottleneck for the EDCA but not for PCA. The throughput ratio between EDCA and PCA is significant. When using only 4 channels, the ratio is about 8% for EDCA and increased to 37% in favor of the PCA when using 20 channels. In a different experiment (omitted due to space limtations) we set the number of packets of each flow to 200 and found even more throughput improvement using PCA algorithm. Results of this test show that when using 4 channels, the ratio is about 18% and increased to 40% when using 20 channels. The difference between two tests is because in PCA, long communication paths are already exist and ready for streams transferring without any overhead of RTS/CTS protocols. Clearly without increasing the amount of streams β even PCA will cease to improve (as follows from eq. 4).

We have also tested the effect of the mobility rate on network throughput using the Random Waypoint mobility model. The results show that the more the speed of nodes increases the more the network throughput decreases and that there is a constant gap of 17% more throughput of PCA versus EDCA. Another experiment we done is to compare packets latency between the two algorithms. The latency was

computed for 100 packets that were selected randomly. We varied the number of generated streams in the network from 50 streams up to 500. The results show that for PCA the average latency of a packet is 6 simulation cycles compare to 17 for the EDCA. The last experiment measures the effect of the CFkPC recalculation on the network throughput. A degradation in the throughput is expected when the recalculation time increases however this experiment shows that for certain range of values this degradation does not harm the usefulness of the PCA. Due to space limitations the results of these three sets of experiments can not be included.

References

[1] *3-dimensional matching*. http://en.wikipedia.org/wiki/3-dimensional_matching, 2011.

[2] *Hidden node problem*. http://en.wikipedia.org/wiki/Hidden_node_problem, 2011.

[3] A. Agrawal, P. Klein, and R. Ravi. When trees collide: An approximation algorithm for the generalized steiner problem on networks. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 134–144. ACM, 1991.

[4] K.M. Alzoubi, PJ Wan, and O. Frieder. Maximal independent set, weakly-connected dominating set, and induced spanners in wireless ad hoc networks. *International Journal of Foundations of Computer Science*, 14(2):287–303, 2003.

[5] D. W. Bliss, P. A. Parker, and A. R. Margetts. Simultaneous transmission and reception for improvedwireless network performance. In *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, pages 478 –482, 2007.

[6] I.A. Getting. The global positioning system, 1993.

[7] Zygmunt J. Haas, Senior Member, Jing Deng, and Student Member. Dual busy tone multiple access (dbtma) - a multiple access control scheme for ad hoc networks. In *IEEE Transactions on Communications*, pages 975–985, 2002.

[8] Nitin Jain, Samir R. Das, and Asis Nasipuri. A multichannel csma mac protocol with receiver-based channel selection for multihop wireless networks. In *In IEEE IC3N*, pages 432–439, 2001.

[9] T. Jian. An $o(2^{0.304 \cdot n})$ algorithm for solving maximum independent set problem. *Computers, IEEE Transactions on*, 100(9):847–851, 1986.

[10] T. F. More. *Estimation of sparse hessian matrices and graph coloring problems*. Springer, 1982.

[11] J. Schneider and R. Wattenhofer. A log-star distributed maximal independent set algorithm for growth-bounded graphs. In *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing*, pages 35–44. ACM, 2008.

[12] Jungmin So and Nitin H. Vaidya. Multi-channel mac for ad hoc networks handling multi-channel hidden terminals using a single transceiver. *ACM MobiHoc*, 2004.

[13] R.E. Tarjan and A.E. Trojanowski. Finding a maximum independent set. *SIAM J. Comput.*, 6(3):537–546, 1977.

[14] Jinsu Wang, Sharad Mehrotra, and Nalini Venkatasubramanian. Pbca - prediction based channel allocation. In *GLOBECOM*, pages 4801–4806. IEEE, 2007.

[15] Shih-Lin Wu, Chih-Yu Lin, Yu-Chee Tseng, and Jang-Ping Sheu. A new multi-channel mac protocol with on-demand channel assignment for multi-hop mobile ad hoc networks, 2000.

Minimum Latency Gossiping in Wireless Sensor Networks

Min Kyung An, Nhat X. Lam, Dung T. Huynh and Trac N. Nguyen
Department of Computer Science
University of Texas at Dallas
Richardson, Texas 75080
Email: {mka081000, lxnhat, huynh, nguyentn}@utdallas.edu

***Abstract*—Gossiping is one of the most crucial applications in Wireless Sensor Networks (WSNs) which has been the focus of many researchers. A main issue of gossiping is how to assign timeslots to nodes for interference-free data transmission. There are three models concerning gossiping in WSNs: unit-sized, bounded-sized, or unbounded-sized messages. For these models, the problem of constructing minimum latency gossiping schedules has been widely studied in the literature although most of the existing studies are based on the graph model.**

In this paper, we study the *Minimum Latency Gossiping (MLG)* problem with unbounded-sized messages in the graph model as well as the more realistic physical interference model known as Signal-to-Interference-Noise-Ratio (SINR) where there exist relatively few works [1] on the three gossiping models. In the SINR model, we prove the NP-hardness of the MLG problem with unbounded-sized messages. In both the graph model and SINR model, we propose a constant factor approximation algorithm that yields schedules whose latency is bounded by $O(\Delta + R)$, where Δ is the maximum node degree of a network and R is its radius. We also study the performance of the algorithm through simulation.

***Index Terms*—Gossiping, All-to-All Broadcast, NP-hardness, Geometric Graph, Signal-to-Interference-Noise-Ratio, Approximation Algorithm**

I. Introduction

The *gossiping* problem has been the focus of many researchers as it is one of the crucial applications in Wireless Sensor Networks (WSNs) along with the *broadcast* problem. In a WSN that consists of a number of sensor nodes, the gossiping problem, which is also known as *all-to-all broadcast*, is to distribute the message of each node to all the other nodes in the network, whereas the broadcast problem is to distribute a unique message from a source node to all the other nodes. As the small-sized sensors have limited energy resources, researchers have focused on reducing energy consumption while distributing data in a network so that the network life time is extended. An interesting approach is to assign *timeslots* to nodes to obtain a good (short) *schedule* thereby avoiding unnecessary transmissions. As data distribution may occur periodically, reducing the *latency* of the schedule, that is, constructing schedules with a minimum number of timeslots, has been a fundamental issue in such applications.

In the literature on the gossiping problem, there are three models: *unit-sized*, *bounded-sized*, or *unbounded-sized* messages. In the unit-sized-message model, a node can send a single unit-sized message, and therefore combining messages is not allowed. In the bounded-sized-message model, a node can combine messages that have been received so far (up to some limit), whereas in the unbounded-sized-message model, there is no limit on the length of the combined message.

The gossiping problem in the *graph model* has been investigated by many researchers over the last several years. In the *collision-free graph model*, for the unit-sized-message model, [2] introduced a 1974-approximation algorithm, and [3] proposed an optimal randomized schedule with a latency bounded by $O(n \log n)$, where n is the number of sensor nodes in the network. Later, [4] introduced a 27-approximation algorithm which produces gossiping schedules with a latency bounded by $27(n + R - 1)$, where R is the network radius. In [5], two approximation algorithms with constant factors of 20 and 34 have been studied. For the bounded-sized-message model, [6] studied the gossiping problem where messages can be combined into a single message whose size is bounded by $\log n$, and [7] gave an improvement over [6]. For the unbounded-sized-message model, [8] showed that their algorithm produces gossiping schedules with $O(g + \frac{\nu \log n}{\log \nu - \log \log n})$ timeslots, where $\nu = \Omega(\log n)$ and g is the network diameter, and [9] introduced a constant factor approximation algorithm whose latency is bounded by $7\Delta + 258R$. Although there have been many studies in the collision-free graph model for the gossiping problem, surprisingly there exists no study in the *collision-interference-free graph model*, to the best of our knowledge.

The graph model which has been used in many studies, however, is not an adequate model since *cumulative interference* caused by all the other concurrently transmitting nodes is ignored. Thus, researchers have started investigating problems in WSNs in the more realistic *physical interference model* which is known as the Signal-to-Interference-Noise-Ratio (SINR) model since its introduction by Gupta et al. in [10]. For the SINR model, [1] introduced a constant factor approximation algorithm for the gossiping problem in the unit-sized-message model, and [11] proposed a $O(\log n)$-approximation algorithm for the unbounded-sized-message model. However, [11] considered only the concurrently sending nodes within some predefined interference area from a receiver, and therefore, some interference caused by senders located far away is ignored.

While these studies have been concerned with gossiping,

some other researchers have focused on related applications such as *data aggregation* and *broadcast*. Table I shows a summary of some related works. Although there have been several approximation algorithms for these related problems, there are surprisingly few studies regarding the complexity of the problems. For the problem of data aggregation, NP-hardness was proved for the collision-free model by [12], and [13] and [14] showed not only an $\Omega(\log n)$ approximation lower bound for the problem, but also the NP-hardness in the collision-interference-free graph model and the geometric SINR model, respectively. For the broadcast problem, [2] proved its NP-hardness in the collision-free model. However, to the best of our knowledge, the NP-hardness of the broadcast problem in both the collision-interference-free and SINR models as well as that of the gossiping problem in all network models remain open. (Note that the existing NP-hardness of the broadcast problem in [2] holds for the collision-free graph model, but not for the other models.)

Models	Data Aggregation	Broadcast
C-Free	[12], [15], [16]	[2], [17], [18], [19]
C-I-Free	[13]	[20], [21], [22]
SINR	[23], [24], [14], [25], [26]	[21]

TABLE I: Summary of Works on Problems Related to Gossiping

In this paper, we continue the study of the gossiping problem with *unbounded-sized messages* in the graph model and the geometric SINR model. Extending the proof in [14], we show that this problem is NP-hard in the geometric SINR model. Additionally, for the uniform power model, we introduce a constant factor approximation algorithm yielding schedules whose latency is bounded by $O(\Delta + R)$ in both the graph model and the geometric SINR model. Our algorithm gives an improved approximation ratio of 44 over the existing ratio of 285 given by [9] in the collision-free graph model. In the collision-interference-free graph model and the SINR model, our approximation algorithm is the first one to the best of our knowledge. Moreover, regarding broadcast, the broadcast subroutine of our gossiping algorithm provides broadcast schedules with a better approximation ratio than existing ones given by [20], [21] for the collision-interference-free model.

This paper is organized as follows. Section II describes our network models and defines the Minimum Latency Gossiping (MLG) problem with unbounded-sized messages. In Section III, we show the NP-hardness of MLG in the geometric SINR model. Section IV introduces a constant factor approximation algorithm for the MLG problem and shows some simulation results. Section V contains some concluding remarks.

II. Preliminaries

A. Network Models

In our paper, we model a wireless sensor network as (V, D, p), where V represents a set of n nodes, and $D : V \times V \rightarrow R^+$ represents the distance function between nodes. Letting $p_{max} : V \rightarrow R^+$ be the maximum power level function, we define a power assignment function as $p : V \rightarrow R^+$, $p(u) \leq p_{max}(u)$, $u \in V$.

1) Graph Model: In the graph model, given a transmission power level $p(u)$ for each node u, let $R^u_{p(u)} = \{v | v \in V, D(u,v) \leq p(u)\}$ denote the set of all nodes that can be reached by u with the power level $p(u)$. Two nodes u and v can communicate only if they are in the coverage area of each other, i.e., $u \in R^v_{p(v)}$ and $v \in R^u_{p(u)}$. However, we also need to consider the collision or interference that interferes with the communication. Given a power level $p(u)$ of u, the interference range of u is defined as $\rho \cdot p(u)$, where $\rho \geq 1$ is the interference factor. Given $\rho \geq 1$, let $F^u_{p(u)} = \{v | v \in V, D(u,v) \leq \rho \cdot p(u)\}$ denote the set of all nodes in the interference range of u. Then, *collision* (or *conflict*) is said to occur at a receiver node w if there exist other concurrently sending nodes u and v such that $w \in R^u_{p(u)} \cap F^v_{p(v)}$, where $\rho = 1$. On the other hand, *interference* is said to occur at w if there exist other concurrently sending nodes u and v such that $w \in R^u_{p(u)} \cap F^v_{p(v)}$, where $\rho > 1$.

In the literature, the graph model concerning only collision (i.e., when $\rho = 1$) is called the *collision-free model*, whereas the graph model concerning both collision and interference (i.e., when $\rho \geq 1$) is called the *collision-interference-free model*. In the graph model, the communication graph can be modeled as a bidirectional graph $G(V, E)$, where $E = \{(u,v) | u, v \in V, D(u,v) \leq p(u)$ and $D(v,u) \leq p(v)\}$.

2) SINR Model: In the physical interference model (SINR) [10], if a node u transmits with its power level $p(u)$, then the power received at another node v is $p(u) \cdot D(u,v)^{-\alpha}$, where α, the *path loss exponent*, is commonly assumed to be in the interval $[2, 6]$. In order that node v can receive and decode data sent by u, the ratio of the received power at v to the interference caused by all the other concurrently transmitting nodes and background noise must be beyond an SINR threshold $\beta \geq 1$. Formally, node v can receive data successfully via the link (u, v) only if

$$SINR(u,v) = \frac{p(u) \cdot D(u,v)^{-\alpha}}{N + I_v} \geq \beta$$

where $N > 0$ is the background noise, and

$$I_v = \sum_{w \notin \{u,v\}, w \in X} p(w) \cdot D(w,v)^{-\alpha}$$

is the cumulative interference at v caused by nodes in X that is the set of other concurrently transmitting nodes. Observing that u can send its data to the nodes within the distance $(\frac{p(u)}{NB})^{\frac{1}{\alpha}}$, the communication graph can be modeled as a directed graph $G(V, E)$, where $E = \{(u \rightarrow v) | u, v \in V, D(u,v) \leq (\frac{p(u)}{NB})^{\frac{1}{\alpha}}\}$.

B. Problem Definition

In this paper, we are concerned with gossiping in the *unbounded-sized-message* model, i.e., we assume that multiple messages can be combined as a single message, and there is no limit on the length of a message that one node can transmit.

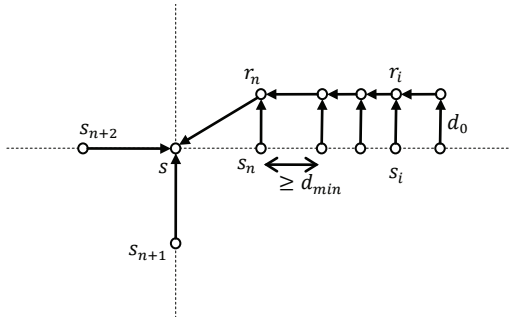


Fig. 1: The corresponding geometric MLG instance

The Minimum Latency Gossiping (MLG) problem is defined as follows. A schedule is defined as a sequence of *timeslots*, at each of which, several nodes are scheduled to send its data to its receivers. Formally, at each timeslot t, we have an *assignment vector* $\pi_t = \langle (s_{t_1}, p(s_{t_1})), \ldots, (s_{t_k}, p(s_{t_k})) \rangle$ in which s_{t_i} is assigned to send data with its power level $p(s_{t_i})$, $1 \leq i \leq k$, and

1) (Graph Model) neither collision nor interference occurs at any receiver r, or
2) (SINR Model) the SINR threshold inequality is satisfied for all receivers r,

where (s_{t_i}, r) is an edge in the communication graph $G(V, E)$, i.e., all the senders s_{t_i} can transmit concurrently.

A *schedule* is a sequence of assignment vectors $\Pi = (\pi_1, \pi_2, \ldots, \pi_M)$, where M is the length of the schedule which is also known as its *latency*. A schedule Π is *successful* if the message $m(v)$ of each node $v \in V$ is received by all the other nodes in the network. In a schedule, a node may be scheduled at several timeslots with different power levels. The MLG problem is defined as follows:

Input. A set of nodes V, a distance function D which is defined as the Euclidean distance between nodes, a maximum power level function p.

Output. A successful schedule of minimum length.

III. NP-Hardness

In this section, we prove the NP-hardness for the minimum latency gossiping (MLG) problem. The structure of this proof is similar to the one in the proof of the NP-hardness of the minimum latency aggregation problem in [14], which, as the proof of the NP-hardness of scheduling with power control in geometric SINR [27], is based on a construction in [28]. We restate the construction in [14] for reader's convenience and omit the details.

In order to prove MLG's NP-hardness, we construct a polynomial time reduction from the *Partition* problem which was proven NP-complete in [29]. This decision problem is defined as follows. Given a finite set of distinct and positive integers, the objective is to determine if it is possible to divide this set into two subsets such that the sums of all integers in each subset are equal.

Let I_P be an instance of Partition which consists of a set S of n distinct and positive integers $a_1, a_2, ...a_n$. Without loss of generality, assume that $a_1 < a_2 < ... < a_n$. We construct in polynomial time an instance I_M of the MLG problem as follows.

In the instance I_M, we have $2n + 3$ nodes including $2n$ nodes s_i and r_i, $1 \leq i \leq n$, 2 nodes s_{n+1} and s_{n+2} and a center node s. These nodes are deployed on the plane at the following positions.

$$\begin{aligned} pos(s) &= (0, 0) \\ pos(s_{n+1}) &= (0, -(\tfrac{24P}{N\beta(A^\alpha\beta+24)})^{\frac{1}{\alpha}}) \\ pos(s_{n+2}) &= (-(\tfrac{24P}{N\beta(A^\alpha\beta+24)})^{\frac{1}{\alpha}}, 0) \end{aligned}$$

and, for all $1 \leq i \leq n$,

$$\begin{aligned} pos(s_i) &= ((\tfrac{P}{b_i N\beta})^{\frac{1}{\alpha}}, 0) \\ pos(r_i) &= ((\tfrac{P}{b_i N\beta})^{\frac{1}{\alpha}}, d_0) \end{aligned}$$

where P is the maximum power value to be defined below. Let $\sigma = \sum_{i=1}^{n} a_i$, $A = ((\frac{1}{a_n - 1})^{\frac{1}{\alpha}} - (\frac{1}{a_n})^{\frac{1}{\alpha}})$, $b_i = \frac{a_i A^\alpha}{12\sigma}$ and $d_0 = (\frac{12P\sigma}{12\sigma N\beta + nN\beta^2})^{\frac{1}{\alpha}}$.

With $d(u, v)$ denoting the Euclidean distance between u and v, we define the maximum power level for the nodes as follow:

$$\begin{aligned} &p_{max}(s_i) = p_{max}(s_{n+1}) = p_{max}(s_{n+2}) = P \\ &p_{max}(r_i) = N\beta d(r_i, r_{i+1})^\alpha, 1 \leq i \leq n, r_{n+1} \equiv s \\ &p_{max}(s) = N\beta d(s, s_1)^\alpha = \tfrac{P}{b_1} = \tfrac{12P\sigma}{a_1 A^\alpha} > P \end{aligned}$$

Fact 1. *Let $T_i = \{s_j | 1 \leq j \leq n + 1 \wedge i \neq j\}$. It holds for all $1 \leq i \leq n$ that $SINR(s_i, r_i)$ exceeds β when node s_i is assigned to send data to r_i at the same timeslot as the nodes in T_i.*

Fact 2. *For all $1 \leq i \leq n$, s_i can send data only to r_i.*

Fact 3. *s_{n+1} and s_{n+2} can send data only to s.*

Fact 4. *r_{i+1} can receive data from r_i (where $r_{n+1} \equiv s$) if and only if there is no other nodes sending in r_i's timeslot.*

Fact 5. *It holds for all $1 \leq i < n$ that r_i can send data to s through r_{i+1} only.*

Fact 6. *s_1 can receive data from s if and only if there is no other node sending in that timeslot.*

Lemma 1. *I_P has a solution if and only if I_M has an optimal gossiping schedule of length $n + 3$.*

Proof: Omitted. ■

From Lemma 1 we obtain

Theorem 2. *The MLG problem is NP-hard.*

IV. Constant Factor Approximation Algorithm

In this section, we introduce a constant factor approximation algorithm for the MLG problem in the graph model and the physical interference (SINR) model, assuming a uniform power level P, i.e., for all $v \in V$, $p(v) = P$. In those models, we make the following assumptions:

For the *graph model*, we set the maximum link length $r =$

P, and assume that the undirected graph $G = (V, E)$, where $E = \{(u, v)|d(u, v) \leq r\}$, is connected and $\rho \geq 1$.

For the *SINR model*, define $r_{max} = (\frac{P}{N\beta})^{\frac{1}{\alpha}}$ and notice that if node u on link (u, v) of length r_{max} is transmitting, then none of the remaining nodes can transmit concurrently. Thus we are interested in links (u, v), where $d(u, v) \leq \delta(\frac{P}{N\beta})^{\frac{1}{\alpha}}$ for some constant $\delta \in (0, 1)$ as considered in [23]. Thus, in the SINR model, we let r be $\delta(\frac{P}{N\beta})^{\frac{1}{\alpha}}$, and assume that the undirected graph $G = (V, E)$, where $E = \{(u, v)|d(u, v) \leq r\}$, is connected and $\alpha > 2$ [10].

A. Algorithm

We first introduce some definitions and notations that are used subsequently:

- *Graph Center*: Given a communication graph $G = (V, E)$, we call node c a *center* node if the distance from c to the farthest node from c is minimum.
- *Maximal Independent Set* (MIS): A subset $V' \subseteq V$ of the graph G is said to be *independent* if for any vertices $u, v \in V', (u, v) \notin E$. An independent set is said to be *maximal* if it is not a proper subset of another independent set.
- *Connected Dominating Set* (CDS): A *dominating set* (DS) is a subset $V' \subseteq V$ such that every vertex v is either in V' or adjacent to a vertex in V'. A DS is said to be *connected* if it induces a connected subgraph.

1) Gossip Tree Construction: Our algorithm assigns timeslots to nodes based on a gossip tree whose construction is based on that of a data aggregation tree in [23]. After a center node c is chosen, we construct a breadth-first-search (BFS) tree (cf. [30]) on G rooted at node c so that the latency can be bounded in terms of the network radius R rather than its diameter. [23] finds an MIS on G using an algorithm in [31] based on the BFS tree. We call nodes in the MIS *dominators* and the others *dominatees*. The constructed MIS satisfies the property that the distance between any pair of its complementary subset is *exactly two hops* [31]. Next, we connect the dominators using some *connectors* that were originally dominatees to obtain a CDS of G. If there exist some remaining dominatees that are not connected to the CDS, then each of such dominatees is connected to its neighboring dominator that has the smallest hop distance to c in the BFS tree. We denote the newly formed tree by T, and use it as the gossip tree in our algorithm.

2) Gossip Scheduling Algorithm: Our algorithm (Algorithm 1) starts by partitioning the network into square cells each of which has diagonal length r, and labels each cell with the label $CL(i, j)$ if its upper-left corner has coordinate (i, j). Once the gossip tree T is obtained, we gather data from each node to the center node c (Steps 4 – 5 in Algorithm 1). Node c then broadcasts the collected data combined with its own based on T (Steps 6 – 14 in Algorithm 1). Assigning timeslots for data gathering and broadcasting is based on a constant C which guarantees that two senders (receivers) can transmit (receive) data successfully if they are at least C cells apart from each other. The constant C is set as follows in the two different models:

- Graph model: $C = \lceil \rho \cdot \sqrt{2} + 2 \rceil$
- SINR model:

$$C = \lceil (\frac{P \cdot 2\pi}{N(\delta^{-\alpha} - 1)(\alpha - 2)})^{\frac{1}{\alpha - 2}} \cdot \delta^{-1} (\frac{N\beta}{P})^{\frac{1}{\alpha}} + 1 \rceil$$

Algorithm 1 Gossiping

Input: A set V of nodes with a uniform power level P
Output: Length of schedule

1: Partition the network into square cells each of which has diagonal length r.
2: Construct a gossip tree T using the algorithm in [23].
3: Set the first timeslot $t \leftarrow 1$.
4: // Data Gathering starts.
5: $t \leftarrow$ **CCA**(V, T, t)
6: // Broadcasting start.
7: $TS(c) \leftarrow TS(c) \cup \{t\}$
8: $t \leftarrow t + 1$
9: **for** $i = 1$ to $R - 1$ **do**
10: Let $S_i \subseteq V$ be the set of connectors at level i in T.
11: **if** $S_i \neq \emptyset$ **then** $t \leftarrow$ **RBS**(S_i, t) **end if**
12: Let $S_{i+1} \subseteq V$ be the set of dominators at level $i + 1$ in T.
13: **if** $S_{i+1} \neq \emptyset$ **then** $t \leftarrow$ **SBS**(S_{i+1}, t) **end if**
14: **end for**
15: **return** $t - 1$

Data Gathering. In order to gather data to c, we use the data aggregation algorithm called *Cell Coloring* algorithm in [13], which is included in Algorithm 2. The *Cell Coloring* algorithm, which is based on an algorithm in [23], is originally built for the graph model assuming $\rho \geq 1$. We use the algorithm not only for the graph model, but also for SINR model where the constant C is defined accordingly.

Broadcast. Once all the data is gathered to c, c broadcasts the collected data combined with its own to the whole network. We introduce a new broadcast scheduling algorithm which is also based on the data aggregation algorithm of [23]. The details of our broadcast algorithm are contained in Algorithm 1 (Steps 6 – 14). In Step 7, c (which is also a dominator) broadcasts the data to its neighbors (lower level connectors). In Steps 10 – 11, the connectors which have just received data from the upper level dominators relay the data to their lower level dominators. These steps are based on the receivers' locations (Algorithm 4), i.e., the connectors whose receivers (lower level dominators) are C cells apart from each other are assigned to the same timeslot. Then, in Steps 12 – 13, the dominators relay the data to its dominatees and lower level connectors. These dominators are scheduled based on their (senders') locations (Algorithm 3), i.e., the dominators which are C cells apart from each other are assigned to the same timeslot. These Steps 10 – 13 are repeated until the data is

disseminated to the whole network.

In Algorithms 1, 3 and 4, $TS(v)$ denotes the set of timeslots at which node v is activated to send data.

Algorithm 2 Cell Coloring Algorithm (**CCA**) [13]

Input: A set V of nodes, a tree T and a starting timeslot t
Output: Timeslot t

1: Let $S \subseteq V$ be the set of dominatees in T.
2: **if** $S \neq \emptyset$ **then** $t \leftarrow$ **SBS**(S, t) **end if**
3: **for** $i = R$ to 2 **do**
4: Let $S_i \subseteq V$ be the set of dominators at level i in T.
5: **if** $S_i \neq \emptyset$ **then** $t \leftarrow$ **SBS**(S_i, t) **end if**
6: Let $S_{i-1} \subseteq V$ be the set of connectors at level $i-1$ in T.
7: **if** $S_{i-1} \neq \emptyset$ **then** $t \leftarrow$ **SBS**(S_{i-1}, t) **end if**
8: **end for**
9: **return** t

Algorithm 3 AssignTimeSlot (**SBS**) [13]

Input: A set S of nodes and a starting timeslot t
Output: Timeslot t

1: **while** $S \neq \emptyset$ **do**
2: Pick one node $v_s \in S$ in each cell. Let $S' \subseteq S$ be the set of such nodes.
3: **for** $c_1 = 0, ..., C-1$ and $c_2 = 0, ..., C-1$ **do**
4: $X \leftarrow \emptyset$, $X \leftarrow \{v_s | v_s \in S'$ with $CL(x, y)$ such that $c_1 = x \mod C$ and $c_2 = y \mod C\}$
5: **if** $X \neq \emptyset$ **then**
6: **for** each $v_s \in X$ **do**
7: $TS(v_s) \leftarrow TS(v_s) \cup \{t\}$
8: **end for**
9: $t \leftarrow t + 1$, $S \leftarrow S \setminus X$
10: **end if**
11: **end for**
12: **end while**
13: **return** t

B. Analysis

In this section, we analyze Algorithm 1 and bound the latency of the gossip schedules produced by it. First, we set the constant value C for the graph model and the SINR model based on [13], [32].

Lemma 3 (Graph Model)**.** *[13] Let $C = \lceil \rho \cdot \sqrt{2} + 2 \rceil$, where $\rho \geq 1$ is the interference factor. Then any two sender (receiver) nodes that are at least C cells apart from each other can concurrently send (receive) data without any collision and interference.*

Lemma 4 (SINR Model)**.** *[32] For SINR threshold $\beta \geq 1$, path loss exponent $\alpha > 2$, background noise $N > 0$, and some constant $\delta \in (0, 1)$, let*

$$C = \lceil (\tfrac{P \cdot 2\pi}{N(\delta^{-\alpha} - 1)(\alpha - 2)})^{\frac{1}{\alpha - 2}} \cdot \delta^{-1} (\tfrac{N\beta}{P})^{\frac{1}{\alpha}} + 1 \rceil$$

Algorithm 4 Receiver-based Scheduling (**RBS**)

Input: A set S of nodes and a starting timeslot t
Output: Timeslot t

1: **for** each $v_s \in S$ **do**
2: $Z \leftarrow Z \cup$ v_s's lower level dominators in T
3: **end for**
4: **for** $c_1 = 0, ..., C-1$ and $c_2 = 0, ..., C-1$ **do**
5: $X \leftarrow \emptyset$, $X \leftarrow \{v_r | v_r \in Z$ with $CL(x, y)$ such that $c_1 = x \mod C$ and $c_2 = y \mod C\}$
6: **if** $X \neq \emptyset$ **then**
7: **for** each $v_r \in X$ **do**
8: $v_s \leftarrow v_r$'s upper level connector in T
9: $TS(v_s) \leftarrow TS(v_s) \cup \{t\}$
10: **end for**
11: $t \leftarrow t + 1$
12: **end if**
13: **end for**
14: **return** t

Then any two sender nodes that are at least C cells apart from each other can concurrently send (receive) data without interference.

Next, we prove that the latency of a gossip schedule found by Algorithm 1 is bounded by $O(\Delta + R)$. We need the following lemmas.

Lemma 5. *[23] The number of connectors in a cell is at most 12.*

Lemma 6. *In Algorithm 1, gathering data from all the other nodes to center node c takes at most $\Delta \cdot C^2 + 6 \cdot R$ timeslots.*

Proof: First consider the dominatees in each cell and their dominator v (Steps 1–2 in Algorithm 2). Obviously, there are at most Δ dominatees in each cell, and one of those Δ dominatees must be a connector to connect the dominator v to another dominator. Therefore, the number of dominatees is bounded by $\Delta - 1$, and gathering data from all the dominatees to the corresponding dominators takes at most $(\Delta - 1)C^2$ timeslots.

Next, consider the dominators at level i (Steps 4–5 in Algorithm 2). Since there is at most one dominator in each cell, gathering data from all the dominators at level i to the connectors at level $i-1$ takes C^2 timeslots. As this process is repeated at most $\frac{R}{2}$ times, gathering data from all dominators to upper level connectors takes at most $\frac{R}{2} \cdot C^2$ timeslots.

Now, let us consider only the connectors at level j, where $1 < j < R$ (Steps 6–7 in Algorithm 2). In one cell, at most 11 of those connectors at level j have the role of sending the collected data to their dominators at level $j-1$; one remaining connector in the cell must relay data from the dominator at level $j-1$ to another dominator at level $j-3$. Therefore, gathering data from the connectors at level j to the dominators at level $j-1$ takes at most $11C^2$ timeslots. As this process is repeated $\frac{R-2}{2}$ times, it takes at most $\frac{R-2}{2} \cdot 11C^2$ timeslots. On the other hand, all the connectors at level 1 send data to

a sink, and this requires at most $12C^2$ timeslots.

Thus, the latency of data gathering is at most $(\Delta-1)C^2 + \frac{R}{2}\cdot C^2 + \frac{R-2}{2}\cdot 11\cdot C^2 + 12C^2 = \Delta\cdot C^2 + 6R$. ■

Lemma 7. *In Algorithm 1, broadcasting data takes at most* $1 + C^2\cdot(R-1)$ *timeslots.*

Proof: First, note that in Steps 7–8 in Algorithm 1, the center node sends data to its neighbors in one timeslot. Next, consider the connectors at level i (Steps 10–11). Any two of those connectors cannot share one dominator at level $i+1$ on the gossip tree T; otherwise a cycle would result. This means that those connectors can be scheduled based on the receivers' locations, i.e., the locations of the corresponding dominators at level $i+1$. As there is at most 1 dominator in each cell, relaying data from the connectors at level i to the dominators at level $i+1$ takes at most C^2 timeslots. Since Steps 10–11 are repeated $\frac{R-1}{2}$ times, at most $\frac{R-1}{2}\cdot C^2$ timeslots are needed.

Now consider the dominators at level j (Steps 12–13). Since at most one dominator can be located in a cell, sending data from the dominators to the connectors at level $j+1$ takes at most C^2 timeslots. As the Steps 12–13 are repeated $\frac{R-1}{2}$ times, it takes at most $\frac{R-1}{2}\cdot C^2$ timeslots.

Thus, the latency of broadcasting is at most $1 + \frac{R-1}{2}\cdot C^2 + \frac{R-1}{2}\cdot C^2 = 1 + C^2\cdot(R-1)$. ■

Lemma 8 (Lower bound)**.** *[9] If* Δ *is the maximum node degree in a network, then every gossip schedule with unbounded-size messages has at least* $\Delta + R - 1$ *timeslots.*

Theorem 9. *Algorithm 1 produces gossip schedules whose latency is bounded by* $C^2\cdot\Delta + (C^2+6)R + (1-C^2) = O(\Delta+R)$*, and it is a constant-factor approximation with the factor of* $2(C^2+6)$.

Proof: By Lemma 6 and Lemma 7, the latency, denoted by SOL, of schedules produced by Algorithm 1 is bounded by $C^2\cdot\Delta+(C^2+6)R+(1-C^2)$. Next, without loss of generality, assume that $n \geq 2$, and therefore $\Delta \geq 1$ and $R \geq 1$. Then, denoting the lower bound by OPT, the approximation ratio is

$$\begin{aligned}\frac{SOL}{OPT} &\leq \frac{C^2\Delta+(C^2+6)R+(1-C^2)}{\Delta+R-1} \leq \frac{(C^2+6)\Delta+(C^2+6)R}{\Delta+R-1}\\ &\leq \frac{(C^2+6)(\Delta+R-1)+(C^2+6)(R+\Delta-1)}{\Delta+R-1}\\ &= 2(C^2+6)\end{aligned}$$

■

Note that this approximation ratio is 44 in the collision-free graph model, an improvement on the approximation ratio of 258 given by [9]. To the best of our knowledge, in the collision-interference-free graph model and the SINR model, our results are the first constant-factor approximation algorithms.

Also note that our broadcast scheduling algorithm yields an approximation ratio of C^2 by Lemma 7. In the collision-interference-free graph model, it is $C^2 = \lceil\rho\sqrt{2}+2\rceil^2$ which improves on the approximation ratio of $6\lceil\frac{2}{3}(\rho+2)\rceil^2$ given by [21]. In addition, [20] studies the broadcast algorithm with $\rho = 2$ giving an approximation ratio of 26, whereas our approximation ratio is 25.

C. Simulation

In our simulation, networks are generated randomly in the Euclidean plane where the number of nodes is 500. The nodes are randomly deployed on an area of size $m \times m$, where $m = 300, 400$ and 500. For each m, we generate 100 different networks, and average the latencies produced by the algorithm over the networks. For the simulation, we set the various parameters as follows:

1) Graph model:
   - Choice of ρ: We use $\rho = \{1, 2, \cdots, 10\}$
   - Initial power assignment: We first use Kruskal's algorithm [30] to find the minimum spanning tree T_{MST} using edge weights defined as the distance between any two nodes. Then, we set the initial power $P = r$, where r is the length of the longest edge in T_{MST}. Given the initial power assignment, we obtain the initial graph $G = (V, E)$, where $E = \{(u,v)|d(u,v) \leq r\}$.
2) SINR model:
   - Choices of SINR parameters: We use $\alpha = 5$, $N = 1$ and $\beta = 1$.
   - Choice of δ: We use $\delta = \{0.1, 0.2, \cdots, 0.9\}$.
   - Initial power assignment: Given a uniform power assignment, if δ is too small, the graph may not be connected. In order to make the initial graph connected even with the smallest $\delta = 0.1$ in our simulation, we set the initial power $P = \beta N(\frac{r}{0.1})^{\frac{1}{\alpha}}$, and obtain the initial graph $G = (V, E)$, where $E = \{(u \to v)|d(u,v) \leq \delta(\frac{P}{N\beta})^{\frac{1}{\alpha}}\}$.

Table II shows the performance of Algorithm 1 in the graph model. For fixed ρ, as the network becomes denser (i.e., the network node degree becomes larger and the network radius smaller) the latency decreases, whereas as the network becomes sparser (i.e., the network node degree becomes smaller and the network radius larger) the latency increases. For fixed network density, as ρ becomes larger (i.e., the interference range becomes wider), the latency increases.

ρ	300×300	400×400	500×500
1	313.13	321.52	321.59
2	357.85	368.71	369.98
3	430.32	440.18	440.23
4	461.79	469.39	469.41
5	513.79	522.64	523.09
6	536.89	542.38	544.81
7	557.04	562.19	564.57
8	596.11	599.17	601.54
9	613.90	617.61	619.30
10	642.07	645.58	647.34

TABLE II: Latencies of Algorithm 1 in Graph Model

Table III shows the performance of Algorithm 1 in the SINR model. For fixed δ, as the network becomes denser (i.e., the network node degree becomes larger, but the network radius becomes smaller) the latency decreases, whereas as

the network becomes sparser (i.e., the network node degree becomes smaller, but the network radius becomes larger) the latency increases. For fixed density, as δ becomes smaller (i.e., the network node degree becomes smaller, but the network radius becomes larger), the latency increases, whereas as δ becomes larger (i.e., the network node degree becomes larger, but the network radius becomes smaller), the latency decreases.

δ	300×300	400×400	500×500
0.1	654.91	662.98	663.09
0.2	651.21	660.08	662.20
0.3	646.19	648.97	651.39
0.4	605.93	615.83	615.86
0.5	512.63	523.91	524.17
0.6	401.23	414.80	414.91
0.7	341.18	354.35	355.08
0.8	295.06	305.52	307.22
0.9	264.31	274.54	275.66

TABLE III: Latencies of Algorithm 1 in SINR Model

From these tables, we can observe that having smaller network radius rather than having smaller network node degree gives better results. This is because the variation of network node degree affects only the latencies of data aggregation schedules, whereas the variation of the network radius affects the latencies of the schedules for data aggregation as well as broadcast.

V. Conclusion

In this paper, we have studied the Minimum Latency Gossiping (MLG) problem with unbounded-sized messages in the graph model and SINR model. We have proved the NP-hardness of the problem in the SINR model, and proposed a constant factor approximation algorithm whose latency is bounded by $O(\Delta + R)$ in both graph and SINR models assuming a uniform power level. We have also studied the performance of the algorithm through simulation. As to future work, we plan to study the gossiping problem with unit-sized and bounded-sized messages.

References

[1] P.-J. Wan, L. Wang, and O. Frieder, "Fast Group Communications in Multihop Wireless Networks Subject to Physical Interference," in *IEEE MASS*, 2009, pp. 526 –533.

[2] R. Gandhi, S. Parthasarathy, and A. Mishra, "Minimizing Broadcast Latency and Redundancy in Ad Hoc Networks," in *ACM MOBIHOC*, 2003, pp. 222–232.

[3] F. Manne and Q. Xin, "Optimal Gossiping with Unit Size Messages in Known Topology Radio Networks." in *CAAN*, 2006, pp. 125–134.

[4] S. C.-H. Huang, H. Du, and E.-K. Park, "Minimum-Latency Gossiping in Multi-hop Wireless Networks," in *MOBIHOC*, 2008, pp. 323–330.

[5] R. Gandhi, Y.-A. Kim, S. Lee, J. Ryu, and P.-J. Wan, "Approximation Algorithms for Data Broadcast in Wireless Networks," in *IEEE INFOCOM*, 2009, pp. 2681–2685.

[6] R. Bar-yehuda, A. Israeli, and A. Itai, "Multiple Communication in Multi-Hop Radio Networks," *SIAM Journal on Computing*, vol. 22, pp. 875–887, 1993.

[7] M. Christersson, L. Gasieniec, and A. Lingas, "Gossiping with Bounded Size Messages in Ad Hoc Radio Networks," in *Proc. of the 29th Intern. Colloquium on Automata, Languages and Programming (ICALP)*, 2002, pp. 377–389.

[8] F. Cicalese, F. Manne, and Q. Xin, "Faster Centralized Communication in Radio Networks." in *ISAAC'06*, 2006, pp. 339–348.

[9] K. Krzywdzinski, "Fast Construction of Broadcast Scheduling and Gossiping in Dynamic Ad Hoc Networks," in *IMCSIT*, 2010, pp. 879–884.

[10] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. on Information Theory*, vol. 46, pp. 388–404, 2000.

[11] Q. Xin, "Minimum-Latency Communication in Wireless Mesh Networks under Physical Interference Model," in *ICC*, 2010, pp. 1–6.

[12] X. Chen, X. Hu, and J. Zhu, "Minimum Data Aggregation Time Problem in Wireless Sensor Networks," in *MSN*, 2005, pp. 133–142.

[13] M. K. An, N. X. Lam, D. T. Huynh, and T. N. Nguyen, "Minimum Data Aggregation Schedule in Wireless Sensor Networks," *International Journal of Computers and Their Applications (IJCA)*, vol. 18, pp. 254–262, Dec. 2011.

[14] N. X. Lam, M. K. An, D. T. Huynh, and T. N. Nguyen, "Minimum Latency Data Aggregation in the Physical Interference Model," in *ACM MSWiM*, 2011, pp. 93–102.

[15] S. C. H. Huang, P.-J. Wan, C. T. Vu, Y. Li, and F. Yao, "Nearly Constant Approximation for Data Aggregation Scheduling," in *IEEE INFOCOM 2007*, 2007, pp. 6–12.

[16] X. Xu, X. Y. Li, X. Mao, S. Tang, and S. Wang, "A Delay-Efficient Algorithm for Data Aggregation in Multihop Wireless Sensor Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, pp. 163–175, 2011.

[17] S. C.-H. Huang, P.-J. Wan, X. Jia, and H. Du, "Low-Latency Broadcast Scheduling in Ad Hoc Networks," in *WASA*, 2006, pp. 527–538.

[18] S. C.-H. Huang, P.-J. Wan, X. Jia, H. Du, and W. Shang, "Minimum-Latency Broadcast Scheduling in Wireless Ad Hoc Networks," in *INFOCOM*, 2007, pp. 733–739.

[19] R. Gandhi, A. Mishra, and S. Parthasarathy, "Minimizing Broadcast Latency and Redundancy in Ad Hoc Networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 840–851, 2008.

[20] Z. Chen, C. Qiao, J. Xu, and T. Taekkyeun Lee, "A Constant Approximation Algorithm for Interference Aware Broadcast in Wireless Networks," in *IEEE INFOCOM*, 2007, pp. 740 –748.

[21] S. C. H. Huang, P.-J. Wan, J. Deng, and Y. S. Han, "Broadcast Scheduling in Interference Environment," *IEEE Transactions on Mobile Computing*, vol. 7, pp. 1338–1348, November 2008.

[22] G. Calinescu and S. Tongngam, "Interference-aware Broadcast Scheduling in Wireless Networks," *Ad Hoc Netw.*, vol. 9, no. 7, pp. 1069–1082, Sept. 2011.

[23] X.-Y. Li, X. Xu, S. Wang, S. Tang, G. Dai, J. Zhao, and Y. Qi, "Efficient Data Aggregation in Multi-hop Wireless Sensor Networks under Physical Interference Model," in *MASS 2009*, pp. 353-362.

[24] H. Li, Q. S. Hua, C. Wu, and F. C. M. Lau, "Minimum-Latency Aggregation Scheduling in Wireless Sensor Networks under Physical Interference Model," in *Proc. of 13th ACM Intern. conf. on Modeling, analysis, and simulation of wireless and mobile systems*, 2010, pp. 360–367.

[25] M. M. Halldórsson and P. Mitra, "Wireless Connectivity and Capacity," in *SODA*, 2012, pp. 516–526.

[26] N. Hobbs, Y. Wang, Q.-S. Hua, D. Yu, and F. C. M. Lau, "Deterministic Distributed Data Aggregation under the SINR Model," in *TAMC*, 2012, pp. 385–399.

[27] M. Vlker, B. Katz, and D. Wagner, "On the Complexity of Scheduling with Power Control in Geometric SINR," Tech. Rep., 2009.

[28] O. Goussevskaia, Y. A. Oswald, and R. Wattenhofer, "Complexity in Geometric SINR," in *MobiHoc*, 2007, pp. 100–109.

[29] R. Karp, "Reducibility Among Combinatorial Problems," in *Complexity of Computer Computations*, R. Miller and J. Thatcher, Eds. Plenum Press, 1972, pp. 85–103.

[30] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. MIT Press and McGraw-Hill, 2009.

[31] P.-J. Wan, K. Alzoubi, and O. Frieder, "Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks," in *INFOCOM*, 2002, pp. 1597–1604.

[32] M. K. An, N. X. Lam, D. T. Huynh, and T. N. Nguyen, "Connectivity in Wireless Sensor Networks in the SINR Model," in *Proc. of 20th Annual IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2012, to appear.

Simple Closed-Form Approximations for the ASER of Digital Modulations over Fading Channels

A. Annamalai, E. Adebola and O. Olabiyi

Center of Excellence for Communication Systems Technology Research
Dept. of Electrical & Computer Engineering, Prairie View A&M University, Texas 77446

Abstract – *In this article we apply two distinct methods to obtain simple closed-form approximations for the average symbol error rate (ASER) performance metric of a broad class of coherent digital modulations in a myriad of fading environments, which are known to be analytically involved as they require evaluation of the expectation of the Gaussian Q-function and/or its integer powers. In the first approach, we exploit the shifting property of Dirac delta approximations of the Q-function to circumvent the need for integration. In the second approach, we introduce tight exponential-type approximations for the Q-function that directly lead to the development of closed-form expressions for the ASER in terms of only the moment generating function (MGF) of the received signal-to-noise ratio (SNR) random variable. Numerical results reveal that our proposed solutions based on the MGF method are much more versatile and can yield better accuracy compared to our approximations derived via the Dirac delta approximation technique.*

Keywords: *unified analysis of digital communications over fading channels, moment generating function method, Dirac delta approximation, Gaussian quadratures*

1 Introduction

The Gaussian Q-function is defined as

$$Q(x) = \frac{1}{2}\text{erfc}(\frac{x}{\sqrt{2}}) = \int_x^\infty \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)dy,\ x \geq 0 \qquad (1)$$

which corresponds to the complement of the cumulative distribution function (CDF) of a normalized (zero-mean, unit variance) Gaussian random variable. This mathematical function plays a vital role in the analysis and design of digital communications since the conditional error probability (CEP) of a broad class of coherent modulation schemes can be expressed either in terms of $Q(x)$ alone or as a weighted sum of its integer powers (e.g., see Table 1, [1, Eqs. (8.36)-(8.39)], [2, Chapter 4]). In addition, system performance measures such as the average symbol, bit or block error probabilities in fading channels typically involve taking the statistical expectation of $Q(x)$ and its integer powers with respect to the random variable that characterizes the fading channel (i.e., $E[Q^p(x)]$). However, analytical difficulties associated with evaluating $E[Q^p(x)]$ from its canonical integral representation of (1) (owing to the presence of the argument of the function in the lower limit of the integral) have led to the development of alternative exponential-type integral representations for the Q-function and its integer powers [1, Eqs. (4.2), (4.9), (4.31) and (4.32)], analytically simple and tight closed-form bounds and approximations for $Q(x)$ [3]-[16], characteristic function method [21] and asymptotic analysis [17]-[19].

Table 1. CEP of several coherent digital modulation schemes.

Modulation	Conditional Symbol Error Probability $P_s(\gamma)$
BPSK	$Q(\sqrt{2\Omega\gamma})$
M-PSK	$2Q(\sqrt{2\Omega\gamma}\sin(\pi/M)),\ M \geq 4$ (approx.)
M-QAM	$4\lambda Q\left(\sqrt{3\Omega\gamma/(M-1)}\right) - 4\lambda^2 Q^2\left(\sqrt{3\Omega\gamma/(M-1)}\right)$, where $\lambda = (\sqrt{M}-1)/M$
Differentially Encoded BPSK (DE-BPSK)	$2Q\left(\sqrt{2\Omega\gamma}\right) - 2Q^2\left(\sqrt{2\Omega\gamma}\right)$
DE-QPSK	$4Q\left(\sqrt{\Omega\gamma}\right) - 8Q^2\left(\sqrt{\Omega\gamma}\right) + 8Q^3\left(\sqrt{\Omega\gamma}\right) - 4Q^4\left(\sqrt{\Omega\gamma}\right)$

Note: $\Omega = E_s/N_o$ denotes the received signal-to-noise ratio (SNR).

But the appeal of the single exponential-type integral with finite integration limits for Q-function is restricted to the first four powers of $Q(x)$. Also, majority of existing bounds and approximations (with the exception of [4], [8], [12]-[14]) are not in the "desirable exponential form" for facilitating the task of statistical averaging of the CEP in generalized multipath and multichannel fading environments. Moreover, only [5], [6], [9], [11] and [15] have considered the problem of finding $E[Q^p(x)]$ in closed-form, and subsequently applied their results for ASER analysis of differentially-encoded BPSK and QPSK digital modulation schemes in Nakagami-m fading. Although [6] may be extended to other fading environments, their final expression will involve the computation of higher-order derivatives of the MGF of SNR. It was also pointed out in [11] that the accuracy of [6, Eqs. (10)-(11)] deteriorates considerably with the increasing value of Nakagami-m fading severity index. While [11, Eq. (10)] (which utilizes a semi-infinite Gauss-Hermite quadrature approximation for $Q(x)$) can achieve better accuracy compared to [5], [6] and [9], its solution is limited to only Nakagami-m fading. In [16], Jang suggested using an asymptotic Dirac delta approximation $Q(\sqrt{x}) \doteq 0.5\delta(x-2)$ [16, Eq. (63)] instead to eliminate the need for integration involving coherent BPSK modulation. Whereas in [15], the same author suggested decomposing the integrand of $E[Q^p(x)]$ into a product of a generalized function $g(x)$ (i.e., "nascent" delta function) and an auxiliary function

This work is supported in part by funding from the US Army Research Office (W911NF-10-1-0087), Air Force Research Laboratory/Clarkson Aerospace, and the National Science Foundation (0931679 & 1040207).

after replacing the Q-function with its approximation [15, Eq. (4)], and then simplifying the resulting integral by invoking an asymptotic Dirac delta approximation for $g(x)$ as

$$g(x) = x^{c-1}\exp(-ax) \doteq \frac{\Gamma(c)}{a^c}\delta(x - c/a) \tag{2}$$

where notation "$\doteq$" denotes the Dirac delta approximation. However, the ASER analysis of digital modulations via Dirac delta approximation technique has thus far been restricted to coherent BPSK and differentially-encoded coherent BPSK and QPSK modulation schemes. Furthermore, recognizing that the shifting point can directly impact the overall accuracy of the resulting ASER approximation, we anticipate that the use of a tighter series approximation for $Q(x)$ (in lieu of [15, Eq. (4)]) may lead to an improved ASER approximation. Hence, one of the objectives of this paper is to extend [15] and [16] by deriving simple closed-form expressions for the ASER of a wide range of digital modulation schemes in conjunction with improved Q-function approximations given in [9, Eq. (25)] and [14]. In particular, our results in Section 2 generalize [15]-[16] to higher order signal constellations (i.e., may be utilized to predict the ASER of M-QAM, M-PSK as well as differentially encoded M-PSK) in addition to yielding slightly better approximations even for the specific cases considered in [15]. In Section 3, we also derive analytically simple and tight closed-form ASER approximations using the MGF method. While exponential-type approximations for $Q(x)$ presented in [4], [8] and [12]-[14] are already in a desirable exponential form, to the best of our knowledge, their utility in ASER analysis of DE-BPSK and DE-QPSK over generalized fading channels have not been reported previously. In the Appendix, we develop a rapidly converging series expression for a generic integral via Gauss-Chebyshev quadrature (GCQ) numerical integration technique and subsequently highlight some of its application including the development of an efficient and asymptotically exact ASER formula for differentially-encoded M-PSK over fading channels via the MGF method. Selected computational results and comparisons between various ASER approximations for different M-ary modulation schemes and fading environments are provided in Section 4.

2 Dirac Delta Approximation

Similar to [15], we consider a normalized probability density function (PDF) of the fading channel SNR in the form

$$p_\gamma(\gamma) = K\exp(-h\gamma)\gamma^{c-1}f(\gamma),\ \ \gamma \geq 0 \tag{3}$$

where γ denotes the squared magnitude of the channel fading amplitude, K is a constant, and $f(\gamma)$ is an auxiliary function that depends on fading characteristics. The coefficients K, h, c and $f(\gamma)$ for several different wireless channel models are also summarized in Table 2. From Table 1, we can also write down a generic expression for the symbol error probability of a wide range of digital modulation schemes in AWGN as

$$P_s(\gamma) = \sum_{z=1}^{Z}\alpha_z Q^{p_z}(\sqrt{\beta_z\Omega\gamma}) \tag{4}$$

where α_z and β_z are constants that depend on a specified digital modulation, and $\Omega = E_s/N_0$ corresponds to the received symbol SNR. To compute the ASER, we need to find the statistical expectation of (4) with respect to the fading random variable γ, viz.,

$$\begin{aligned}\overline{P_s} &= \sum_{z=1}^{Z}\alpha_z K\int_0^\infty Q^{p_z}(\sqrt{\beta_z\Omega\gamma})\exp(-h\gamma)\ \ \gamma^{c-1}f(\gamma)d\gamma \\ &= \sum_{z=1}^{Z}\frac{\alpha_z K}{(\beta_z\Omega)^c}\int_0^\infty Q^{p_z}(\sqrt{x})\exp(\frac{-hx}{\beta_z\Omega})\ \ x^{c-1}f(\frac{x}{\beta_z\Omega})dx\end{aligned} \tag{5}$$

In the remaining part of this section, we will consider various approximations for $Q(x)$ in (5), and simplify the integration task by exploiting the asymptotic Dirac delta approximation (2) with the shifting property of $\delta(.)$.

2.1 Jang's Q-function Approximation

Expanding the integer powers of [15, Eq. (4)] using the binomial theorem, we obtain

$$\begin{aligned}Q^{p_z}(x) &\cong \frac{\exp(-x^2\ p_z/2)}{(2\pi)^{p_z/2}}\sum_{k=0}^{p_z}\binom{p_z}{k}\left(-\frac{\exp(-x-1/2)}{x+1}\right)^k\left(\frac{1}{x}\right)^{p_z-k} \\ &= \frac{\exp(-x^2\ p_z/2)}{(2\pi)^{p_z/2}}\left[\frac{1}{x} - \frac{\exp(-x-0.5)}{x+1}\right]^{p_z}\end{aligned} \tag{6}$$

Substituting (6) into (5), and then simplifying the resulting expression using (2), we immediately arrive at

$$\begin{aligned}\overline{P_s} \doteq \sum_{z=1}^{Z}&\frac{K\alpha_z\Gamma(c)}{(2\pi)^{p_z/2}(a_z\beta_z\Omega)^c}f(\frac{c}{a_z\beta_z\Omega}) \\ &\times\left[\sqrt{\frac{a_z}{c}} - \left(\sqrt{\frac{c}{a_z}}+1\right)^{-1}\exp\left(-\sqrt{\frac{c}{a_z}}-\frac{1}{2}\right)\right]^{p_z}\end{aligned} \tag{7}$$

where $a_z = p_z/2 + h/(\beta_z\Omega)$ and $\Gamma(c) = \int_0^\infty x^{c-1}e^{-x}dx$ denotes the Gamma function. For the readers convenience, we have also summarized the fading/modulation parameter selections in Table 2 (i.e., obtained by setting b = 1). It is important to highlight that (7) generalizes [15, eq. (11)] to a broader class of digital modulation schemes (see Table 1).

2.2 Boyd's Q-function Approximation

In this subsection, we will develop two new Dirac delta approximations for (5) based on Boyd's upper and lower bounds for $Q(x)$ [9, Eq. (25)]. Our work is motivated by the fact that these bounds are much tighter than [15, Eq. (4)] (especially near zero), and also due to their simple form that leads to upper and lower bounds for the integer powers of $Q(x)$ in an identical form, viz.,

$$F^p(x,\pi-1) \leq Q^p(x) \leq F^p(x,2/(\pi-2)) \tag{8}$$

where the auxiliary function $F^p(x,\psi)$ is defined as

$$F^p(x,\psi) = \left(\frac{(\psi+1)/\sqrt{2\pi}}{\psi x + \sqrt{x^2+2(\psi+1)^2/\pi}}\right)^p\exp(-px^2/2) \tag{9}$$

The lower bound is slightly tighter than the upper bound when the argument $x > 1.23$ for $p = 1$. Nevertheless, Dirac delta approximations for both the lower and the upper bounds can be obtained by an appropriate substitution for the coefficient ψ in (9). Substituting (9) into (5), and then invoking the Dirac delta approximation (2), we obtain

Table 2. Coefficients for various stochastic channel models.

Channel Model	K	c	h	a	$f(c/(a\beta\Omega))$
Nakagami-m	$m^m/\Gamma(m)$	m	m	$\frac{bp}{2}+\frac{m}{\beta\Omega}$	1
Nakagami-n	$(1+n^2)\exp(-n^2)$	1	1	$\frac{bp}{2}+\frac{1}{\beta\Omega}$	$\exp\left(-\frac{n^2}{a\beta\Omega}\right) I_0\left(2n\sqrt{\frac{1+n^2}{a\beta\Omega}}\right)$
Nakagami-q	$(1+q^2)/(2q)$	1	$\frac{(1+q^2)^2}{4q^2}$	$\frac{bp}{2}+\frac{(1+q^2)^2}{4q^2\beta\Omega}$	$I_0\left(\frac{1-q^4}{4q^2 a\beta\Omega}\right)$

$$\overline{P_s} \doteq \sum_{z=1}^{Z} \frac{K\alpha_z \Gamma(c)}{(a_z\beta_z\Omega)^c} f(\frac{c}{a_z\beta_z\Omega}) \left(\frac{(\psi+1)/\sqrt{2\pi}}{\psi\sqrt{\frac{c}{a_z}}+\sqrt{\frac{c}{a_z}+\frac{2}{\pi}(\psi+1)^2}} \right)^{p_z} \quad (10)$$

The coefficients required for evaluating (10) are provided in Table 2 (by setting $b = 1$).

2.3 Olabiyi's *Q*-function Approximation

More recently, [13]-[14] have developed accurate and invertible exponential-type approximations (up to the third order) for $Q(x)$ by approximating the erfc(.) function as a weighted sum of powers of an exponential function. This form is particularly suitable for finding the statistical expectation of the CEP (including integer powers of $Q(.)$) over the PDF of fading SNR. However, in this subsection we will investigate the efficacy of this $Q(x)$ approximation for deriving simple and accurate closed-form ASER formulas for DE-BPSK and DE-QPSK via Dirac delta approximation approach. Since the $Q(x)$ approximation in [15, Eq. (4)] contains two exponential terms, we will only consider the second order exponential-type approximation for $Q(x)$ in our comparisons, viz., [14, Eq. (2) and Table 2]

$$Q\left(\sqrt{x}\right) \cong \frac{w_1}{2}\exp(\frac{-bx}{2}) + \frac{w_2}{2}\exp(-bx) \quad (11)$$

where $w_1 = 0.3017$, $w_2 = 0.4389$, and $b = 1.0510$. It is also worth mentioning that the invertible property of (11) is not very critical in our current application, and thus other exponential-type approximations such as [8, Eq. (13c)] can be used, if desired. Using the binomial theorem expansion, it is quite straight-forward to show that

$$Q^{p_z}\left(\sqrt{x}\right) \cong \left(\frac{w_1}{2}\right)^{p_z} \sum_{k=0}^{p_z} \binom{p_z}{k} \left(\frac{w_2}{w_1}\right)^k \exp\left(-\frac{xb(p_z+k)}{2}\right) \quad (12)$$

Next substituting (12) into (5), and then simplifying the resulting expression using (2) and the shifting property of Dirac delta function, we arrive at

$$\overline{P_s} \doteq \sum_{z=1}^{Z} K\alpha_z \left(\frac{w_1}{2}\right)^{p_z} \sum_{k=0}^{p_z} \binom{p_z}{k} \left(\frac{w_2}{w_1}\right)^k \frac{\Gamma(c)}{(\widetilde{a_z}\beta_z\Omega)^c} f(\frac{c}{\widetilde{a_z}\beta_z\Omega}) \quad (13)$$

where $\widetilde{a_z} = b(p_z+k)/2 + h/(\beta_z\Omega) = a_z + bk/2$. But a slightly more compact ASER formula can be derived if we first decompose the exponential term in (12) into a product of two exponential terms (i.e., $\exp(-xbp_z/2)\exp(-xbk/2)$ that allows one to simplify the binomial sum) before invoking the Dirac delta approximation (2), viz.,

$$\overline{P_s} \doteq \sum_{z=1}^{Z} \frac{K\alpha_z \Gamma(c)}{(a_z\beta_z\Omega)^c 2^{p_z}} f(\frac{c}{a_z\beta_z\Omega}) \left[w_1 + w_2 \exp\left(-\frac{bc}{2a_z}\right) \right]^{p_z} \quad (14)$$

where the coefficients K, a_z and c for different fading environments are summarized in Table 2 (with $b = 1.0510$).

3 Moment Generating Function Method

In this section, we present yet another method for deriving simple and tight closed-form approximations for the ASER of DE-BPSK and DE-QPSK over fading channels. Specifically, we take advantage of a tight exponential-type approximation for $Q^p(x)$ (see Eq. (12)) in (5) to express the ASER as a weighted sum of the MGF of SNR, viz.,

$$\overline{P_s} \cong \sum_{z=1}^{Z} \alpha_z K \left(\frac{w_1}{2}\right)^{p_z} \sum_{k=0}^{p_z} \binom{p_z}{k} \left(\frac{w_2}{w_1}\right)^k \phi_\gamma(\frac{\Omega\beta_z b(p_z+k)}{2}) \quad (15)$$

by recognizing that the resulting integral,

$$\int_0^\infty e^{-s\gamma} p_\gamma(\gamma)\, d\gamma = \phi_\gamma(s) \quad (16)$$

is simply the Laplace transform of the PDF of SNR. This result (15) is rather interesting especially considering that most prior work on ASER analysis of DE-QPSK/DE-BPSK with/without maximal-ratio diversity receiver (e.g., [5], [6] and [11]) are quite restrictive and limited to the Nakagami-m channel with independent and identically distributed (i.i.d) fading statistics. In contrast, (15) can be readily applied to characterize the ASER of DE-QPSK/DE-BPSK over a myriad of fading environments (including non-identically distributed fading link statistics) and diversity transceivers. Furthermore, in our case the ASER expressions are mostly expressed in terms of elementary functions. For instance, the MGF of SNR in a Nakagami-m channel is given by $\phi_\gamma(s) = m^m/(m+s)^m$.

Although we recognize that the accuracy of a series approximation for $Q(x)$ can be improved by considering more number of terms in that series (e.g., [8, Eq. (13d)]), but its efficiency will be determined by the tightness combined with the number of terms in that series. Hence, we have also developed highly accurate and computationally efficient series approximations for integer powers of the *Q*-function and the CEP of differentially-encoded *M*-PSK in the Appendix, with the aid of GCQ approximation and multinomial theorem. These asymptotically exact closed-form approximations are also in a desirable exponential form and thus will facilitate ASER analysis over generalized fading channels via the MGF method.

4 Numerical Results

In this section, selected numerical results are provided to investigate the efficacies of our new approximations (7), (10), (13), (14) and (15) for ASER analyses of several coherent modulations in a myriad of fading environments.

Fig. 1 shows a comparison of various ASER approximations for QPSK in different Nakagami-m fading channels. The exact performance curve is generated using [1, Eq. (5.78)]. It is apparent that (10), (13) and (14) performs considerably better than (7) when the channel experience more severe fading (i.e., smaller fading severity index m) especially at lower values of the mean channel SNR Ω. The choice of ψ in (8) (that corresponds to the upper and lower bounds) appears to cause only a negligible effect on the overall tightness of the final ASER approximations using (10). Furthermore, the curve corresponding to (15) (MGF method) is also quite close the exact ASER curve for a wide range of m and Ω.

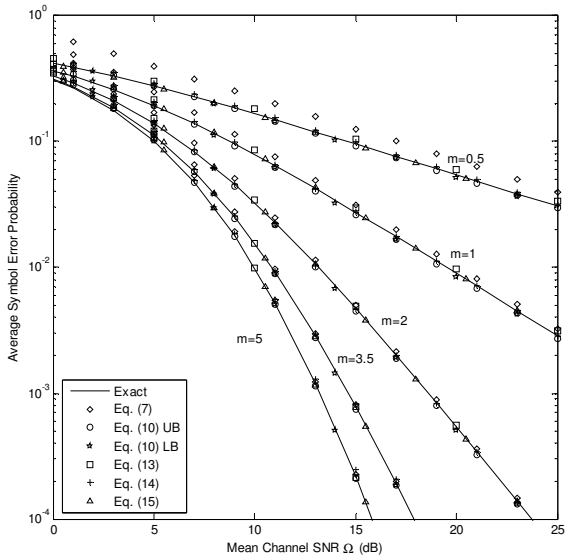


Fig. 1 ASER of QPSK in Nakagami-m fading (m = 0.5, 1, 2, 3.5, 5).

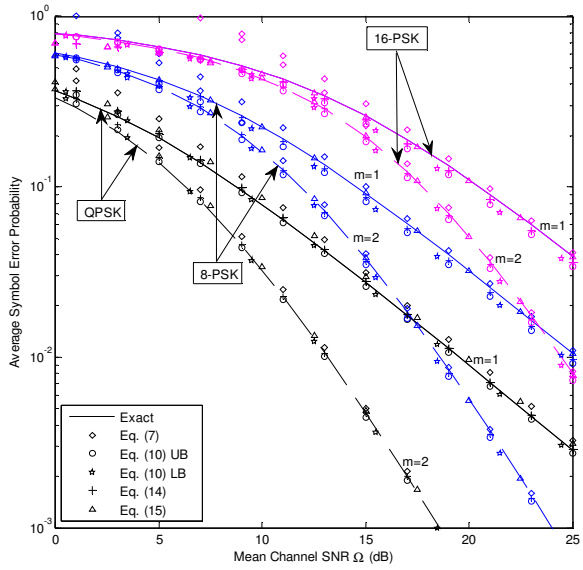


Fig. 2 ASER of M-PSK (M = 4, 8, 16) in Nakagami-m fading (m = 1, 2).

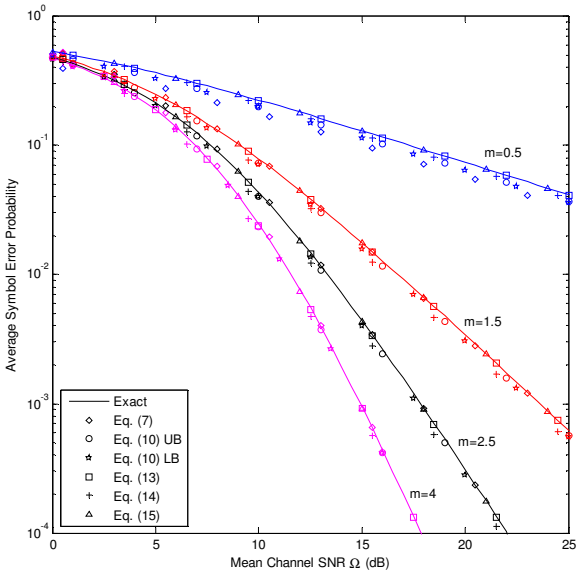


Fig. 3 ASER of DE-QPSK in Nakagami-m fading (m = 0.5, 1.5, 2.5, 4).

Table 3. ASER of DE-BPSK in Nakagami-m fading (m = 0.5, 3).

Ω	0 dB	5 dB	10 dB	20 dB	25 dB
Fading severity index m = 0.5					
Exact	2.677e-1	1.683e-1	9.863e-2	3.176e-2	1.789e-2
Eq. (7)	2.236e-1	1.381e-1	8.008e-2	2.565e-2	1.444e-2
Eq. (10) UB	2.183e-1	1.425e-1	8.500e-2	2.761e-2	1.556e-2
Eq. (10) LB	2.159e-1	1.408e-1	8.398e-2	2.728e-2	1.537e-2
Eq. (13)	2.634e-1	1.659e-1	9.727e-2	3.133e-2	1.765e-2
Eq. (14)	2.286e-1	1.497e-1	8.909e-2	2.888e-2	1.627e-2
Eq. (15)	2.634e-1	1.659e-1	9.727e-2	3.133e-2	1.765e-2
Fading severity index m = 3					
Exact	1.770e-1	4.211e-2	4.027e-3	7.537e-6	2.515e-7
Eq. (7)	2.021e-1	4.460e-2	4.112e-3	7.553e-6	2.517e-7
Eq. (10) UB	3.470e-2	1.340e-2	2.160e-3	6.498e-6	2.281e-7
Eq. (10) LB	3.423e-2	1.321e-2	2.133e-3	6.428e-6	2.257e-7
Eq. (13)	1.802e-1	4.181e-2	3.955e-3	7.399e-6	2.470e-7
Eq. (14)	3.495e-2	1.279e-2	1.981e-3	6.029e-6	2.124e-7
Eq. (15)	1.802e-1	4.181e-2	3.955e-3	7.399e-6	2.470e-7

In Fig. 2, we investigate the accuracies of various Dirac delta approximations when applied to higher order constellations, since the prior work on M-PSK is restricted to only M = 2. It is evident that the curves generated using (7) (i.e., direct generalization of [15]) virtually breaks-down at small values of Ω as constellation size M increases. This can be attributed to the increasing relative error of [5, Eq. (4)] with decreasing value of its argument. As anticipated, (10) yields better approximation than (7) in this case. The results in this Fig. 2 are also interesting because we have now demonstrated that it is possible to derive relatively simple and reasonably accurate closed-form approximations for M-ary modulations via the Dirac delta approximation technique.

Fig. 3 depicts the ASER performance for DE-QPSK in different Nakagami-m channels. It is interesting note that both (13) and (15) tend to yield very good ASER approximations over a wide range of Ω and m values, compared to all other approximations. It is also evident that (7) becomes very accurate as m increases. To investigate this trend further, in Table 1 we summarize the ASER values for various ASER

approximations for DE-BPSK in Nakagami-m fading. The trends observed from Fig. 3 are also apparent from Table 3.

To apply Dirac delta approximation technique to Nakagami-n and Nakagami-q channels, it has been suggested in [15] to choose an auxiliary function as flat as possible so that the final ASER approximation will be robust to the sampling point error. The corresponding auxiliary functions are summarized in Table 2. At this point, we would like to emphasize that similar "ad-hoc" manipulations are not required for the MGF method discussed in Section 3. Furthermore, (15) can be applied directly (i.e., without any further manipulations) to study the multichannel reception case (e.g., maximal-ratio diversity, etc.). The same cannot be said for the Dirac delta approximation technique.

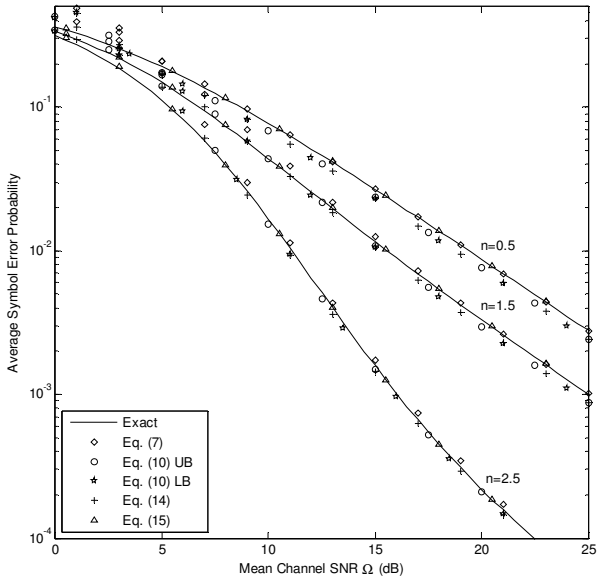


Fig. 4 ASER of 4-QAM in Nakagami-*n* fading (n = 0.5, 1.5, 2.5).

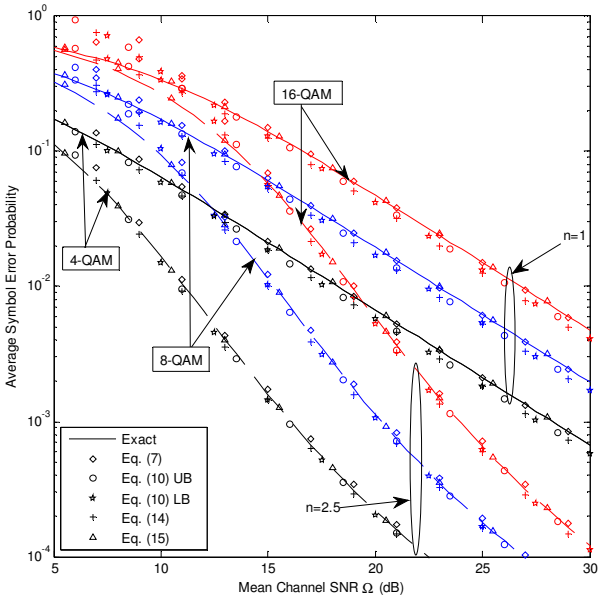


Fig. 5 ASER of *M*-QAM in Nakagami-*n* fading (n = 1, 2.5).

The impact of fade distribution (i.e., fading parameter n in the Nakagami-n channel) and the constellation size of *M*-QAM on various ASER approximations developed in this paper are illustrated in Fig. 4 and Fig. 5, respectively. It is apparent that the accuracies of (7), (10) and (14) deteriorate with decreasing values of n and Ω. However, ASER approximations (10) and (14) are slightly better than (7). These trends are somewhat consistent with our observations for *M*-PSK in Nakagami-m fading (see Fig. 1 and Fig. 2). Interestingly, the curves generated via closed-form approximation (15) are virtually indistinguishable compared to their exact performance curves obtained using an integral formula in [1].

5 Concluding Remarks

In this article, we have studied two novel method for efficient evaluation of ASER for a broad class of coherently detected digital modulation schemes in Nakagami-m, Nakagami-n and Nakagami-q channels. Our new closed-form Dirac delta approximations achieves better accuracy than those reported in [15] and [16], besides generalizing their results to higher order constellations. In addition, we have highlighted the advantages and limitations of Dirac delta approximation method for ASER analysis. We have also demonstrated that our closed-form ASER approximation based on the MGF method (15) is highly accurate and is more versatile than all other asymptotic Dirac delta approximations. In the Appendix, we have shown that asymptotically exact ASER and/or average block error rate in fading channels can be computed efficiently by exploiting a Gaussian quadrature method in conjunction with multinomial theorem, albeit at additional computational cost.

6 Appendix

In this appendix, we will consider evaluating a generic integral of the form depicted in (A.1) via Gauss-Chebyshev quadrature (GCQ) numerical integration technique. This approach leads to a rapidly converging series approximation for the integral besides circumventing the need for a look-up table to store the weights and abscissas associated with other variants of Gaussian quadrature methods. Although a small number of integration points are sufficient to yield a good accuracy, one may readily increase the order N to satisfy a prescribed relative error without any difficulty since the weights and abscissas for the GCQ approximation are in closed-form. Interestingly, this result can be also utilized to derive exponential-type series approximations for higher integer powers of the Gaussian probability integral $Q(x)$ and other related functions.

Let us consider a generic integral of the form

$$g[\delta,\alpha,\beta]=\int_0^{\alpha\pi}\delta f(\beta,\theta)d\theta \qquad \text{(A.1)}$$

where the integrand is a function of the integration variable θ and constants δ and β. Applying the variable substitution $t=\cos(\theta/\alpha)$ (i.e., $\theta=\alpha\cos^{-1}(t)$ and $d\theta=-dt\ \alpha/\sqrt{1-t^2}$) in (A.1), we obtain

$$g[\delta,\alpha,\beta]=\alpha\delta\int_{-1}^{1}\frac{f(\beta,\alpha\cos^{-1}(t))}{\sqrt{1-t^2}}dt \qquad \text{(A.2)}$$

It is now straight-forward to evaluate (A.2) using the GCQ method with abscissas $\cos[(2k-1)\pi/(2N)]$ (which are zeros

of the N^{th} degree Chebyshev polynomial of the first kind) and weights π/N, viz.,

$$g[\delta,\alpha,\beta] \approx \frac{\alpha\delta\pi}{N}\sum_{k=1}^{N} f(\beta, \frac{\alpha(2k-1)\pi}{2N}) \quad \text{(A.3)}$$

Next, we will present several applications of (A.3) to simplify the evaluation of the integral (A.1).

Example 1:

In the first example, we will highlight the utility of (A.3) for deriving simple closed-from ASER approximations for M-ary phase shift keying (*M*-PSK) and M-ary differential phase shift keying (*M*-DPSK) modulation schemes over generalized stochastic fading environments. Using the MGF approach for performance analysis of *M*-PSK with diversity receivers, we have [1, Eq. (9.15)]

$$\overline{P}_s = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \phi_\gamma(\frac{\Omega\sin^2(\pi/M)}{\sin^2(\theta)})d\theta \quad \text{(A.4)}$$

where $\phi_\gamma(s)$ denotes the MGF of SNR in a specified fading environment. Substituting $\alpha=(M-1)/M$, $\beta=\Omega\sin^2(\pi/M)$, $\delta=1/\pi$ and $f(\beta,\theta)=\phi_\gamma(\beta\csc^2(\theta))$ in (A.1), we get

$$\overline{P}_s \approx \frac{M-1}{MN}\sum_{k=1}^{N}\phi_\gamma(\frac{\Omega\sin^2(\pi/M)}{\sin^2((M-1)(2k-1)\pi/(2NM))}) \quad \text{(A.5)}$$

Mimicking the above steps, we can also derive a closed-form approximation for the ASER of *M*-DPSK [1, Eq. (8.200)] as

$$\overline{P}_s \approx \frac{M-1}{MN}\sum_{k=1}^{N}\phi_\gamma(\frac{\Omega\sin^2(\pi/M)}{1+\cos(\pi/M)\cos((M-1)(2k-1)\pi/(2NM))}) \quad \text{(A.6)}$$

The above technique may be readily used to simplify the performance evaluation of any arbitrary two-dimensional signal constellations with/without diversity receivers, but the details are omitted here for brevity.

Example 2:

In our second example, we will employ (A.3) to derive simple exponential-type approximations for $Q(x)$ and its integer powers. One of the benefits of applying GCQ approximation to the conditional error probability (prior to taking statistical expectations over the fading density functions) when dealing with the powers of $g[\delta, \alpha, \beta]$ is that multinomial theorem can be exploited to minimize the number of resulting summation terms compared to simplifying the resulting multi-fold integral after performing the statistical averaging over the fading SNR. This result is of interest in the ASER analysis of differentially-encoded *M*-PSK schemes over fading channels. For instance, the CEP of coherently detected DE-BPSK and DE-QPSK are depicted in (A.7) and (A.8) respectively:

$$P_s = 2Q(\sqrt{2\Omega\gamma}) - 2Q^2(\sqrt{2\Omega\gamma}) \quad \text{(A.7)}$$

$$P_s = 4Q(\sqrt{\Omega\gamma}) - 8Q^2(\sqrt{\Omega\gamma}) + 8Q^3(\sqrt{\Omega\gamma}) - 4Q^4(\sqrt{\Omega\gamma}) \quad \text{(A.8)}$$

Substituting $\alpha=1/2$, $\beta=x^2/2$, $f(\beta,\theta)=\exp(-\beta\csc^2(\theta))$ and $\delta=1/\pi$ [2, Eq. (4.2)] in (A.3), we get a rapidly converging exponential-type series approximation for $Q(x)$, viz.,

$$Q(x) \approx \frac{1}{2N}\sum_{k=1}^{N}\exp\left[\frac{-x^2}{2\sin^2\left(\frac{\pi}{4N}(2k-1)\right)}\right] \quad \text{(A.9)}$$

It is important to highlight that the above series converges to its exact value considerably faster than the corresponding Reinmann integration sum presented in [4, Eq. (8)]. Besides, it completely eliminates the need for finding the optimized coefficients in the Prony's approximation and their associated computational difficulties (e.g., selection of the range as well as the data points for "curve-fitting") especially for higher order approximations (thereby, providing a greater flexibility for attaining a prescribed accuracy).

Exponential-type approximations for higher integer powers ($p \geq 1$) of Gaussian probability integral can be attained via multinomial expansion of (A.9), i.e.,

$$Q^p(x) \approx \frac{1}{(2N)^p}\sum_{k_1+\ldots+k_N=p}\binom{p}{k_1,\ldots,k_N} \times \exp\left[-x^2\sum_{t=1}^{N}\frac{k_t}{2\sin^2\left(\frac{\pi}{4N}(2t-1)\right)}\right] \quad \text{(A.10)}$$

where $\binom{p}{k_1,k_2,\ldots,k_N}=\frac{p!}{k_1!k_2!\ldots k_N!}$ and $k_i \in \{0,1,\ldots,p\}$. The number of terms in the multinomial sum is given by $\frac{(p+N-1)!}{p!(N-1)!}$. Hence it is apparent that (A.10) requires significantly fewer number summation terms compared to repeatedly multiplying (A.9) to achieve the same level of accuracy. For instance, when $N = 5$ and $p = 3$, the number of summation terms in (A.10) and that of repeatedly multiplying (A.9) are given by 7!/(3!4!) = 35 and 5^3 = 125, respectively. Nevertheless, a much simpler closed-form GCQ approximation for $Q^2(x)$, $Q^3(x)$ and $Q^4(x)$ can be derived from their respective single exponential-type integral representation depicted in [1, Eq. (4.9)], [1, Eq. (4.31)] and [1, Eq. (4.32)]. The results for $Q^2(x)$ and $Q^4(x)$ are summarized below as illustrative examples:

$$Q^2(x) \approx \frac{1}{4N}\sum_{k=1}^{N}\exp\left[\frac{-x^2}{2\sin^2\left(\frac{\pi}{8N}(2k-1)\right)}\right] \quad \text{(A.11)}$$

$$Q^4(x) \approx \frac{1}{6\pi N}\sum_{k=1}^{N}\cos^{-1}\left[\frac{3\cos(\frac{\pi}{6N}(2k-1))-1}{2\cos^3(\frac{\pi}{6N}(2k-1))}-1\right] \times \exp\left[\frac{-x^2}{2\sin^2\left(\frac{\pi}{12N}(2k-1)\right)}\right] \quad \text{(A.12)}$$

Hence a simple exponential-type approximation for $Q^5(x)$ can be obtained by multiplying the series approximations (A.12) and (A.9). It is evident that the resulting series approximation is simpler than (A.10). It is also important to recognize that a rapidly converging exponential-type series approximation for (A.8) can be obtained with the aid of [1, Eqs. (4.2), (4.9), (4.31)-(4.32)] and (A.3), viz.,

$$P_s \approx \frac{2}{N}\sum_{k=1}^{N}\exp\left[\frac{-\Omega\gamma}{2\sin^2\left(\frac{\pi}{4N}(2k-1)\right)}\right] - \frac{2}{N}\sum_{k=1}^{N}\exp\left[\frac{-\Omega\gamma}{2\sin^2\left(\frac{\pi}{8N}(2k-1)\right)}\right] + \frac{2}{3N\pi}\sum_{k=1}^{N}\cos^{-1}\left[\frac{3\cos(\frac{\pi}{6N}(2k-1))-1}{2\cos^3(\frac{\pi}{6N}(2k-1))}-1\right]\exp\left[\frac{-\Omega\gamma}{2\sin^2\left(\frac{\pi}{12N}(2k-1)\right)}\right] + \frac{4\sin^{-1}(1/\sqrt{3})}{N\pi^2}\sum_{k=1}^{N}\exp\left[\frac{-\Omega\gamma}{2\sin^2\left(\sin^{-1}(1/\sqrt{3})(2k-1)/(2N)\right)}\right] \times \left\{\pi - \cos^{-1}\left[\frac{3\cos(\sin^{-1}(1/\sqrt{3})(2k-1)/N)-1}{2\cos^3(\sin^{-1}(1/\sqrt{3})(2k-1)/N)}-1\right]\right\} \quad \text{(A.13)}$$

The above result is quite interesting in that several researchers had in the last five years developed various non-exponential type approximations for $Q(x)$ and subsequently applied their approximations to derive closed-form approximations for the ASER of DE-QPSK modulation with maximal-ratio diversity receiver (e.g., [5], [6] and [11]). However, their results were restricted to Nakagami-m channels with independent and identically distributed (i.i.d) fading statistics. In contrast, our closed-form approximation (A.13) can be readily applied to characterize the ASER of DE-QPSK modulation over a myriad of wireless fading channels (including non-identically distributed fading link statistics).

Example 3:

In our third example, we will demonstrate the efficacy of (A.3) and its multinomial expansion for facilitating the ASER analysis of coherently detected differentially-encoded M-PSK particularly when the constellation size M is greater than 4. In this case, simplifications of the conditional error probability similar to (A.7), (A.8) or (A.13) do not seem feasible. If we define $f(\beta,\theta)=\exp(-\beta/\sin^2(\theta))$ in (A.1), then the desired symbol error probability in an AWGN channel [1, Eq. (8.36)] can be expressed as

$$\begin{aligned} P_s = {} & 2g[\frac{1}{\pi},\frac{M-1}{M},\Omega\gamma\sin^2(\frac{\pi}{M})]-g^2[\frac{1}{\pi},\frac{M-1}{M},\Omega\gamma\sin^2(\frac{\pi}{M})] \\ & -\sum_{j=1}^{M-1}\Bigg\{\frac{1}{4}g^2[\frac{1}{\pi},1-\frac{2j-1}{M},\Omega\gamma\sin^2((2j-1)\pi/M)] \\ & +\frac{1}{4}g^2[\frac{1}{\pi},1-\frac{2j+1}{M},\Omega\gamma\sin^2((2j+1)\pi/M)] \\ & -\Bigg(g[\frac{1}{\pi},1-\frac{2j-1}{M},\Omega\gamma\sin^2((2j-1)\pi/M)] \\ & \times\frac{1}{2}g[\frac{1}{\pi},1-\frac{2j+1}{M},\Omega\gamma\sin^2((2j+1)\pi/M)]\Bigg)\Bigg\} \end{aligned} \tag{A.14}$$

As highlighted in [1, pp. 235] the fact that the second and third terms of (A.14) involve squaring of integrals still poses some difficulties in terms of their extension to the fading channel. Hence the application of (A.3) in conjunction with the multinomial theorem [20] is particularly attractive in this case since it can overcome the above-mentioned issue. In fact, it is rather straight-forward to obtain GCQ approximations for $g[1/\pi,\alpha,\beta]$ and $g^p[1/\pi,\alpha,\beta]$, viz.,

$$g[\frac{1}{\pi},\alpha,\beta]\approx\frac{\alpha}{N}\sum_{k=1}^{N}\exp\left[\frac{-\beta}{\sin^2\left(\frac{\pi}{2N}(2k-1)\alpha\right)}\right] \tag{A.15}$$

$$\begin{aligned} g^p[\frac{1}{\pi},\alpha,\beta]\approx{} & \left(\frac{\alpha}{N}\right)^p\sum_{k_1+\ldots+k_N=p}\binom{p}{k_1,\ldots,k_N} \\ & \times\exp\left[-\beta\sum_{t=1}^{N}\frac{k_t}{\sin^2\left(\frac{\pi}{2N}(2t-1)\alpha\right)}\right] \end{aligned} \tag{A.16}$$

When p = 2, there will be a total of $N(N+1)/2$ terms in the above multinomial sum. It is also apparent that (A.15) and (A.16) are in a desirable exponential form and therefore, will facilitate statistical averaging over the density function of the fading SNR in a myriad of fading environments.

7 References

[1] M. K. Simon and M.S. Alouini, *Digital Communication over Fading Channels*, New York: Wiley, 2nd Edition, 2005.

[2] J. G. Proakis and M. Salehi, *Digital Communications,* McGraw Hill, Fifth Edition, 2007.

[3] P. O. Borjesson and C. E. Sundberg, "Simple Approximations of the Error Function Q(x) for Communications Applications," *IEEE Trans. Communications.,* vol. COM-27, no. 3, pp. 639-643, Mar. 1979.

[4] M. Chiani, D. Dardari, and M. K. Simon, "New Exponential Bounds and Approximations for the Computation of Error Probability in Fading Channels," *IEEE Trans. Wireless Communications,* vol. 2, no. 4, pp. 840-845, July 2003.

[5] G. K. Karagiannidis and A. S. Lioumpas, "An Improved Approximation for the Gaussian Q-Function," *IEEE Commun. Letters.*, vol. 11, no. 8, pp. 644-646, Aug. 2007.

[6] Y. Isukapalli and B. D. Rao, "An Analytically Tractable Approximation for Gaussian Q-Function," *IEEE Commun. Letters.*, vol. 12, no. 9, pp. 669-671, Sep. 2008.

[7] Y. Chen and N. C. Beaulieu, "A Simple Polynomial Approximation to the Gaussian Q-function and Its Application," *IEEE Commun.Letters,* vol. 13, pp. 124-126, Feb. 2009.

[8] P. Loskot and N. C. Beaulieu, "Prony and Polynomial Approximations for Evaluation of the Average Probability of Error over Slow-Fading Channels," *IEEE Trans. Vehic. Tech.*, pp. 1269-1280, Mar. 2009.

[9] W. M. Jang, "A Simple Upper Bound of the Gaussian Q-Function with Closed-Form Error Bound ," *IEEE Commun. Letters.,*vol. 15, no. 2, pp. 157-159, Feb. 2011.

[10] M. Lopez-Benitez and F. Casadevall, "Versatile, Accurate, and Analytically Tractable Approximation for the Gaussian Q-Function," *IEEE Trans. Communications*, vol. 59, April 2011, pp. 917-922.

[11] Q. Shi and Y. Karasawa, "An Accurate and Efficient Approximation to the Gaussian Q-Function and its Applications in Performance Analysis in Nakagami-m Fading," *IEEE Commun. Letters,* vol. 15, pp. 479-481, May 2011.

[12] S. Chang, P. Cosman and L. Milstein, "Chernoff-Type Bounds for the Gaussian Error Function," *IEEE Trans. Commununications,.* Vol. 59, pp. 2939-2944, Nov. 2011.

[13] O. Olabiyi, and A. Annamalai, "ASER Analysis of Cooperative Non-Regenerative Relay Systems over Generalized Fading Channels," *Proc. IEEE ICCCN'11,* Maui, Aug. 2011.

[14] Olabiyi and Annamalai, "Invertible Exponential-Type Approximations for the Gaussian Probability Intergral Q(x) with Applications," accepted for publication in the *IEEE Wireless Comm. Letters*, 2012.

[15] W. M. Jang, "Quantifying Performance in Fading Channels Using the Sampling Property of a Delta Function," *IEEE Commun. Letters.,*vol. 15, no. 3, pp. 266-268, March. 2011.

[16] W. M. Jang, "Quantifying Performance of Cooperative Diversity using the Sampling Property of a Delta Function," *IEEE Trans. Wireless Commun.,*vol. 10, no. 7, pp. 2034-2039, July. 2011.

[17] H. S. Abdel-Ghaffar and S. Pasupathy, "Asymptotical Performance of M-ary and Binary Signals over Multipath/Multichannel Rayleigh and Rician Fading," *IEEE Trans. Commun.*, vol. 43, 1995, pp. 2721-2731.

[18] Z. Wang and G. B. Giannakis, "A Simple and General Parameterization Quantifying Performance in Fading Channels,"*IEEE Trans. Communications,* vol. 51, no. 8, pp. 1389-1398, Aug. 2003.

[19] A. Nasri, R. Schober and Y. Ma, "Unified Asymptotic Analysis of Linearly Modulated Signals in Fading, Non-Gaussian Noise, and Interference," *IEEE Trans. Commun.*, vol. 56, June 2008, pp. 980-990.

[20] http://en.wikipedia.org/wiki/Multinomial theorem.

[21] A. Annamalai, C. Tellambura and V. K. Bhargava, "A General Method for Calculating Error Probabilities over Fading Channels," *IEEE Trans. Communications*, vol. 53, no. 5, May 2005, pp. 841-852.

Study of MAC Protocols for a Real Underwater Sensor Network Application

S. Climent, A. Sanchez, J.V. Capella, J.J. Serrano
ITACA, Universitat Politècnica de València
46022 València, Spain
+34963877000 Ext. 88251

Abstract— *Simulations have proven to be useful in aiding the research in wireless and underwater sensor networks. Simulations can be very convenient also when planning the development and deployment of an underwater sensor network for a real application. However, in order to achieve trustworthy results and to be able to extract correct conclusions, accurate models and real parameters should be used.*

This paper studies the behaviour of different medium access techniques when they are applied to a real monitoring application. To this end, different MACs where implemented and tested by means of simulation, using accurate models fed with real data.

Keywords: Simulation, Protocol analysis, Low-power design, MAC protocols, Sensor networks.

1. Introduction

Underwater sensor networks have many interesting and challenging applications like submarine surveillance or environmental monitoring. These applications, and specially the long-term ones, have to be carefully planed before the actual deployment of sensor nodes.

Carefully planning the necessary hardware and algorithms that are going to be executed can avoid node or even network failures once the application is already deployed. This is very important, since retrieving the already deployed nodes can be very expensive and sometimes even impossible.

Simulations are a very convenient way of testing algorithms and protocols before their actual deployment. However, one has to be certain that the simulation results are as much accurate as possible. The use of, for example, a too much simple physical model or supposing synchronization among all nodes, can alter the behaviour of the protocol and sometimes lead to wrong or incomplete conclusions [1].

In this paper we aim to give some insights on which medium access protocol might be more appropriate to apply on a given application. This application is enclosed under a Spanish research project of an unattended monitoring installation in an offshore fish farming facility.

The hardware that is going to be employed in this application is a low-cost, low-power modem with wake-up capabilities and limited computational resources [2]. This forbids us from using complex medium access protocols which would need higher computational power.

In order to carry out this study, a model of the actual hardware modem for the ns-3 simulator is employed [3]. Moreover, since the simulations accuracy is a major requirement, they are performed using the existing Bellhop propagation model for the ns-3 simulator. In addition, Sound Speed Profile (SSP) and bathymetry data where extracted from the location where the application is going to be deployed by using the WOSS API [4].

The rest of the paper is organized as follows: Section 2 introduces the target application of this study. Section 3 highlights some of the more interesting medium access protocols for our application. Section 4 discuses the selected protocols for this study and Section 5 shows their implementation. Finally, Section 6 analyzes the simulation results and in Section 7 conclusions and future work are drawn.

2. Target Application

The target application of this study is enclosed under a Spanish research project of an unattended monitoring installation in an offshore fish farming facility. The general architecture is depicted in Figure 1, where different sensor nodes are placed at fish nets and to the sea bottom also. These nodes are capable of measuring different environmental variables on demand and send these data to the sink.

Sink nodes (there might be more than one sink node depending on the installation requirements) are placed at buoys at the sea surface and equipped with solar energy-harvesting capabilities. They also include some radio modem in order to communicate with an onshore installation.

A sink might require a group of nodes to send some environmental variables periodically during a certain amount of time. In order to keep the architecture as flexible as possible, there is no need for these nodes to be known a priori. For example, the sink can send a message asking for certain information and the desired sample

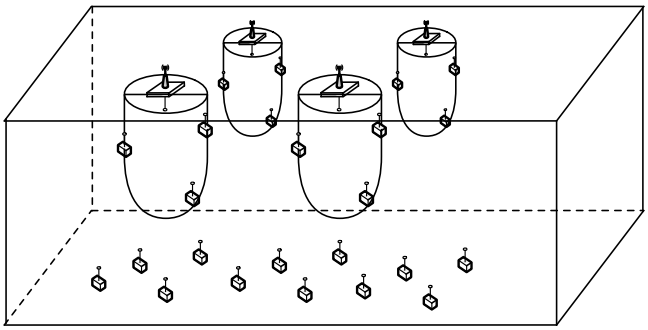


Fig. 1: Target Application Diagram

frequency. After that, the nodes capable of providing this information have to compete to acquire the channel and send the information.

Finally, regarding the underwater communication capabilities, all nodes are equipped with a low-power, low-cost underwater acoustic modem with integrated wake-up capabilities presented by Sanchez et al. in [2].

The aim of this paper is to give insights on which medium access protocol might be more adequate to implement for this application. While bearing in mind the application constrains, the wake-up capabilities of the underwater modem and the fact that it has to be implemented in a low-cost micro-controller with limited computational performance. To this end, in the following section several well known medium access control protocols are going to be introduced.

3. Related work on MAC protocols

Medium Access Control (MAC) protocols are mainly in charge of avoiding packet collisions but, generally speaking, they have to deal also with other factors like, energy efficiency, scalability or latency. These protocols can be divided into two main groups, contention-free and contention-based.

Among the contention-free MAC protocols, time division, frequency division and code division are the three basic types.

In Time Division Multiple Access (TDMA), each node is assigned a fixed slot for it to transmit hence, all nodes must be synchronized in time. Moreover, time guards must be included in order to take into account the propagation delay and the synchronization accuracy [5].

Frequency Division Multiple Access (FDMA) divides the available bandwidth into different frequency bands, allowing different nodes to transmit and receive at the same time while avoiding collisions. In underwater acoustic networks, this scheme is considered unsuitable due to the narrow bandwidth available for communications [6].

Finally, the last basic type of contention-free protocols is the Code Division Multiple Access (CDMA). These type of protocols use binary codes to modulate the signal using a spread-spectrum technique, allowing different nodes to communicate at the same time using different codes. The drawback of using this technique comes with the reduced data rates due to the generally long codes needed to achieve low cross-correlation [7].

On the other hand, contention-based MAC protocols avoid the pre-allocation of resources by allowing the nodes to compete with each other and obtain the medium access on demand. This group of protocols, usually relay on a random access to distribute transmissions and normally they also include some recovery mechanism in case a collision occurs. Two well known and basic protocols in this group are ALOHA and CSMA.

In order to improve efficiency over these simple protocols, different authors have proposed numerous ones which, although maintaining the random access mechanism, perform a channel reservation before sending the actual data packet.

Multiple Access Collision Avoidance (MACA) [8] and its adaptation to underwater acoustic networks MACA-U [9] are well-known protocols that can be classified in this group. Nodes using these protocols perform a handshake exchanging RTS and CTS packets prior to the actual data transmission. Although they can avoid some collisions, they are not data collision-free protocols and some data packets might be lost.

A different proposal is the Propagation-delay-tolerant Collision Avoidance Protocol (PCAP) [10], which also performs a RTS/CTS handshake, but deferring the CTS packet transmission so that it reaches the data transmitter after twice the maximum propagation delay. The major disadvantage is that the protocol needs all nodes to be synchronized.

Another random access with reservation protocol is the Distance-aware Collision Avoidance Protocol (DACAP), proposed in [11]. This protocol tries to avoid Data-RTS collisions by deferring the data transmission for t seconds after sending the RTS. It also introduces a short warning packet sent by the receiver if it overhears an RTS after sending a CTS. The main drawback of this protocol is that the time t has to be set up in advanced and cannot be adapted to network changes.

One more approach is the one given by the original FAMA (Floor Acquisition Multiple Access) protocol [12], which prevents packet collisions provided that the RTS and CTS frames are long enough. Given the long propagation delay of the underwater acoustic medium, these packet lengths might be very long, hence Molins et. al. propose in [13] the Slotted FAMA MAC protocol. This protocol provides some energy savings since nodes do not have to transmit long RTS/CTS frames. However in this case, the slot length needs to be equal to the maximum propagation delay plus the transmission time of a CTS

packet, which can lead to a low channel utilization and also requires synchronization among all nodes.

Finally, T-Lohi (Tone-Lohi), a hybrid between RTS/CTS and CSMA, is proposed in [14]. This protocol adapts automatically the contention time to the number of contending nodes. The nodes send a short packet called tone prior to the actual data packet to count the number of terminals contending for the channel. If a node does not receive any other tones, it starts the transmission. However, if it receives more tones, it adapts its backoff time depending on the number of tones received.

4. Chosen Protocols

Given the application constrains where the underwater nodes have to remain operative and unattended during large periods of time and the network load is not periodically and uniformly distributed, TDMA was discarded from this study. Moreover, the use of any sort of time slots for transmitting data would imply time synchronization, which would increment the energy communication costs and increase its complexity.

The FDMA alternative was also abandoned due to the reduced bandwidth of the hardware modem (1 KHz) and the fact that previous empirical experiments discarded it too [15].

CDMA, although promising, requires more computational power than the one that the implemented node can provide. Moreover, the employed modem can achieve a transmission speed of 1 Kbps and implementing CDMA would decrease this speed considerably [7].

From the other studied protocols in the previous section, PCAP also requires synchronization among all nodes. DACAP, although interesting, needs a fixed t parameter to be set up before deployment and given the changing topology of the network (the number of contending nodes might vary depending on what parameters is asking the sink for), it seemed unsuitable for this application.

ALOHA and MACA-like protocols are the ones that seemed more appropriate since they do not need time synchronization and, using the backoff mechanism, can adapt to the number of contending nodes. Although very interesting, the implementation and study of T-Lohi is left for future work.

For these reasons, ALOHA-CS, MACA and FAMA are the protocols implemented and studied in this work. These protocols have been adapted, when possible, to take advantage of the modem wake-up capabilities.

5. Protocol implementation

In order to evaluate these protocols, they have been implemented under the ns-3 simulator [16], using the model of the underwater modem with wake-up capabilities, previously presented in [3].

The wake-up system has two operational modes, a tone mode and a pattern mode. When configured to use a tone mode, the wake-up subsystem wakes-up the main receiving circuitry whenever a tone is present in the channel. Hence, when using this mode, a node sends a wake-up tone prior to the actual packet and all nodes within the receiving distance would wake-up and receive the packet.

When using the pattern mode, a node only wakes-up when it receives a predefined pattern (which can be the node address). Following the same procedure, the transmitter sends a wake-up pattern prior to the actual packet but only the intended receiver will receive this packet.

This wake-up subsystem also allows to assess the channel state continuously without waking up the main radio circuitry.

In what follows, we introduce how the different protocols have been adapted to use the underwater wake-up modem model and implemented under the ns-3 simulator. Mainly, each time the MAC layer needs to send a packet, a wake-up signal is sent first, so the receiving nodes can wake-up their main radios and receive the packet.

5.1 ALOHA-CS

This protocol is the most simple one of the implemented protocols in this work. Each time there is a packet to be sent, the protocol asserts the channel and, if it is free, it begins the transmission. If the channel is busy, it waits and backs-off for some predefined amount of time when it starts the procedure again. The packets are kept in a FIFO queue until they are sent.

The same protocol was implemented also but including acknowledgements. In this case, the packets are kept in the FIFO queue until the acknowledgement is correctly received.

5.2 MACA-like protocol

Before sending the actual data packet, a RTS/CTS exchange is done between sender and receiver. These packets include the estimated amount of time that will take the complete communication. This time is estimated by knowing the maximum propagation time and the length of the data packets being sent.

Like the previous one, packets are kept in a FIFO queue and a second version of the protocol has been implemented including acknowledgements.

5.3 FAMA-like protocol

Like the previous one, before sending the data packet, a RTS/CTS exchange is done between sender and receiver. As specified by the FAMA protocol, these packets

Table 1: Underwater modem energy consumption

MODE	Wake-up	Modem
TX mode	120 mW	120 mW
RX mode	8.1μW	24 mW
IDLE mode	8.1μW	24 mW
SLEEP mode	-	3μW

have to be of an specific length in order to avoid data packets colliding with other packets.

The FAMA protocol is designed in a way that all nodes must remain awake listening the channel and updating their timers.

In order to take full advantage of the wake-up capabilities, the node will remain in a low-power state until a packet is directly sent to it. Instead of listening the channel continuously through the main radio, it takes advantage of the carrier sensing capabilities of the wake-up modem, updating the timers accordingly to the channel state.

This protocol implementation also includes a FIFO queue and a second version with acknowledgements. Moreover, following the original specifications, a burst mode is supported where, once the node has acquired the channel it can send more than one data packet.

6. Simulation Results

In order to conduct the experiments, we implemented the previous protocols using the ns-3 simulator and adapting them, when possible, to make use of the underwater modem wake-up capabilities.

The Bellhop model was fed with the SSP and bathymetry data, obtained using the WOSS API [4], from the coast of Burriana (Spain) where the application is going to be deployed.

The modem consumption parameters were the ones shown in Table 1 and the transmission speed was set to 1000 bps [2].

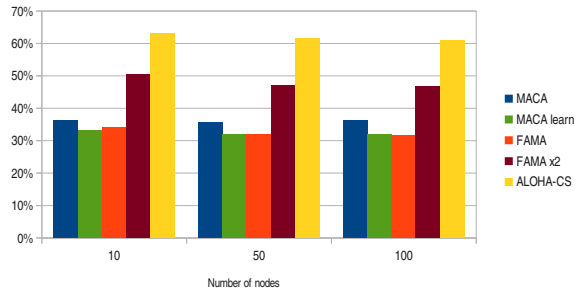
An scenario of 100x100 meters was tested with 10, 50 and 100 nodes randomly deployed. Only one sink was placed at the center of the scenario. Each scenario was simulated several times in order to achieve a confidence interval of ±1 with a confidence level of 95%.

All simulations where seeded using the number 1330703057 and each repetition was done advancing the run number [16].

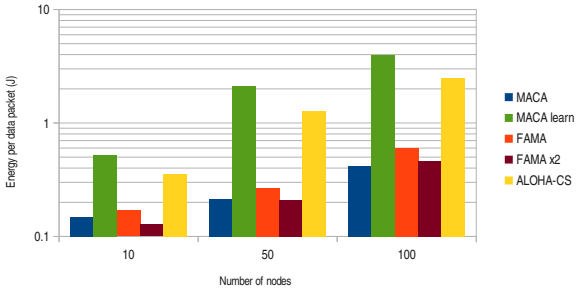
The simulation stop time was set to 30 minutes.

All traffic was directed to the sink node and was generated by a Poisson distribution with an average packet inter-arrival time of 0.6 seconds, which assured that the network was working under saturation. Each packet was generated by this function and assigned to a source node using a random uniform distribution.

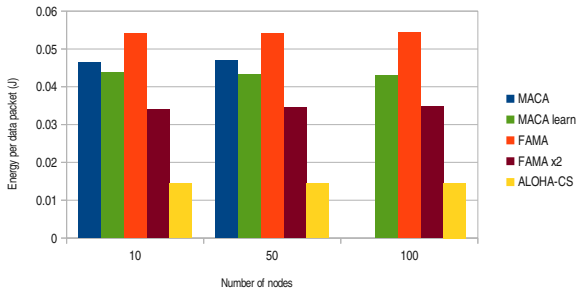
In order to obtain the best possible results for each protocol regardless of the backoff algorithm used, a set of simulations was performs for each protocol varying the backoff time from 0.5 seconds to 20 seconds and the best results in terms of packet delivery ratio are the ones shown in the following figures.



(a) Normalized Received Packets



(b) Energy per data packet. Node



(c) Energy per data packet. Sink

Fig. 2: Results without acknowledgement

In what follows the results of the simulations are going to be introduced for each one of the implemented protocols using and not using acknowledgements.

6.1 Without acknowledgement

Figure 2(a) depicts the number of correctly received packets by the sink normalized to the theoretical maximum. As can be seen, ALOHA-CS is the one that achieves the maximum throughput followed by FAMA x2, which is FAMA but sending two data packets instead of one.

The MACA learn protocol is exactly the same as the MACA-like implemented protocol but using the wake-up tone instead of the pattern mode. This way, all nodes can learn about the on going transmissions and try to avoid collisions. The MACA-like protocol achieves slightly higher packet delivery ratios than the remaining two.

When taking a look to the power consumption in Figure 2(b), one can see how the wake-up tone use of the MACA learn protocol highly increases the energy consumption of the nodes since they wake-up for almost every transmission done.

It can be seen also how ALOHA-CS has a huge energy consumption per each correctly received packet. On the other hand, the remaining three protocols have a similar energy consumption, being FAMA x2 the lowest one.

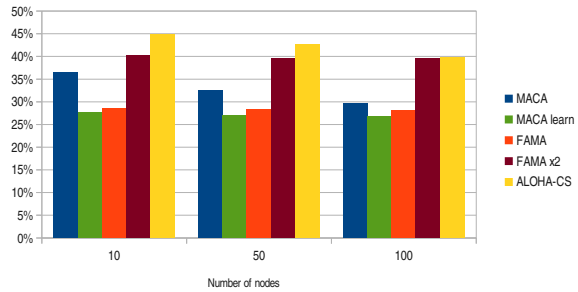
Focusing on the sink energy consumption as depicted in Figure 2(c), ALOHA-CS is the protocol that achieves the lowest one because the sink does not have the CTS sending overload.

6.2 With acknowledgement

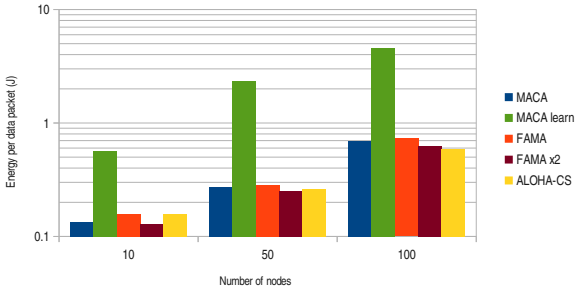
When acknowledgements come into play, the difference in the number of correctly received packets is greatly reduced between ALOHA-CS and FAMA x2, as shown in Figure 3(a).

Figure 3(b) shows how the lowest energy consumption is achieved by FAMA x2, except for the 100 nodes scenario where the ALOHA-CS consumption is a bit lower. Regarding the other protocols MACA learn has the highest energy consumption since it uses the wake-up tone mode.

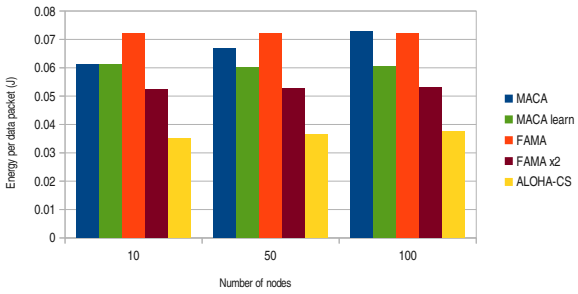
Finally, sink energy consumption per data packet is depicted in Figure 3(c) and as expected, since ALOHA-CS does not have the CTS overhead it has the lowest energy consumption. FAMA and FAMA x2 have a stable energy consumption, independently of the number of nodes, which might be useful when trying to implement energy-neutral operations at the sink node.



(a) Normalized Received Packets



(b) Energy per data packet. Node



(c) Energy per data packet. Sink

Fig. 3: Results using acknowledgement

7. Conclusions

This paper studies the suitability of different MAC protocols to a real underwater sensor network application. To this end, we study different MAC proposals and, bearing in mind the hardware characteristics and restrictions, we have implemented and tested them under the ns-3 simulator.

In order to achieve as accurate simulations as possible, we have used real SSP and bathymetry data from the location where the application is going to be deployed, as well as, real energy consumptions measured from the actual hardware. Moreover, accurate models like the Bellhop propagation model and the modem model where used.

Future work will include the study of more MAC protocols like T-Lohi and real tests at the actual offshore installation using real hardware nodes.

Acknowledgements

The authors gratefully acknowledge financial support from the CICYT (research projects CTM2011-29691-C02-01, TIN2011-28435-C03-01).

References

[1] P. Casari, F. E. Lapiccirella, and M. Zorzi, "A Detailed Simulation Study of the UWAN-MAC Protocol for Underwater Acoustic Networks," in *Oceans 2007*. Vancouver, BC: IEEE, 2007, pp. 1–6.

[2] A. Sanchez, S. Blanc, P. Yuste, and J. J. Serrano, "An Ultra-Low Power and Flexible Physical Modem Design to Develop Energy-Efficient Underwater Sensor Networks," *Sensors*, no. In Press, 2012.

[3] S. Climent, A. Sanchez, J. V. Capella, S. Blanc, and J. J. Serrano, "Modelling and Simulation of Underwater Low-Power Wake-Up Systems," in *S-CUBE 2012*, 2012.

[4] F. Guerra, "World Ocean Simulation System (WOSS): a simulation tool for underwater networks with realistic propagation modeling," *Proceedings of the Fourth ACM International Workshop on UnderWater Networks*, 2009.

[5] J. Proakis, E. Sozer, J. Rice, and M. Stojanovic, "Shallow water acoustic networks," *IEEE Communications Magazine*, vol. 39, no. 11, pp. 114–119, 2001.

[6] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, May 2005.

[7] M. Stojanovic and L. Freitag, "Multichannel Detection for Wideband Underwater Acoustic CDMA Communications," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 3, pp. 685–695, Jul. 2006.

[8] P. Karn, "MACA - a new channel access method for packet radio," in *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, 1990, pp. 134–40.

[9] H.-H. Ng, W.-S. Soh, and M. Motani, "MACA-U: A Media Access Protocol for Underwater Acoustic Networks," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*. IEEE, 2008, pp. 1–5.

[10] X. Guo, "A propagation-delay-tolerant collision avoidance protocol for underwater acoustic sensor networks," *OCEANS 2006-Asia Pacific*, 2007.

[11] B. Peleato and M. Stojanovic, "Distance aware collision avoidance protocol for ad-hoc underwater acoustic sensor networks," *IEEE Communications Letters*, vol. 11, no. 12, pp. 1025–1027, Dec. 2007.

[12] C. L. Fullmer and J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) for packet-radio networks," *ACM SIGCOMM Computer Communication Review*, 1995.

[13] M. Molins and M. Stojanovic, "Slotted FAMA: a MAC protocol for underwater acoustic networks," *OCEANS 2006-Asia Pacific*, 2007.

[14] A. Syed, W. Ye, and J. Heidemann, "T-Lohi: A new class of MAC protocols for underwater acoustic sensor networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, pp. 231–235.

[15] J. Rice, B. Creber, C. Fletcher, P. Baxley, K. Rogers, K. McDonald, D. Rees, M. Wolf, S. Merriam, R. Mehio, J. Proakis, K. Scussel, D. Porta, J. Baker, J. Hardiman, and D. Green, "Evolution of Seaweb underwater acoustic networking," in *OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No.00CH37158)*, vol. 3. IEEE, 2000, pp. 2007–2017.

[16] Ns-3, "ns-3," 2012. [Online]. Available: http://www.nsnam.org

A Novel Wireless Channel Allocation Strategy Along the Railway

Shiya Wen, Hao Wu
State Key Laboratory of Rail Traffic Control and Safety
Beijing Jiaotong University
Beijing, P. R. China 100044
Email: 10120162,hwu@bjtu.edu.cn

***Abstract*—Future mobile communications systems demand for higher data rates and better service quality compared to state-of-the-art systems. One way to achieve this ambitious goal is to use relaying which improves spectrum efficiency by adding one or more intermediate nodes support communication pairs. They can be flexibly deployed than additional base stations while the infrastructure costs is being kept lower. Besides, the concurrent usage of different strategies is likely to be realized in the next generation mobile communications systems, such that according on the actual channel conditions the most beneficial strategy can be chosen. In this paper, a novel strategy of wireless channel allocation is proposed. The target system is the celluar mobile communication system along the railway with relay technology. The numerical results show that: it has good performance in the low complexity ,such as high stability, high capacity and channel utilization rate.**

***Index Terms*—relay; channel allocation; utilization rate; capacity.**

I. Introduction

The relay system, with its many merits, is one of the most promising techniques for the future of mobile communications systems [1] and [2]. The relay-based cellular system largely consists of three elements: a base station (BS), a relay station (RS), and a mobile station (MS). MS can connect to either BS or RS, according to the location of MS. Communication with the MS connected to RS must be performed through the RS. Use of relay in cellular systems were studied in [3] and [4] and [5] and [6]. Most of these works considered solely data traffic. The aim in [3] and [4] was improving the aggregate throughput or spectral efficiency by using relays in broadband systems, while in [5] fair sub-channel allocation algorithms were proposed and in [6] a heuristic was proposed to jointly improve outage probability and throughput. Unlike all these works we distinguish different traffic types and propose a frame-by-frame scheduler, where in each time slot, the time slots, subcarriers and power in a frame are jointly allocated to each transmission. Our algorithm provides proportional fairness for data users, while satisfying delay requirements for real time sessions. Proportional fairness is a good balance between throughput and fairness for data users. We also convert delay requirements of real time users into rate requirements and treat them as constraints in the optimization problem as in our previously proposed resource allocation scheme for a system without RS in [7]. Wireless resource allocation is an important technology in link layer. It plays an indispensable role to increase capacity of wireless networks. Wireless resource allocation put performance of a single user or all users within the district as a criterion in wireless system. It allocated wireless resource to each user. In 4G system, channel allocation and packet scheduling is combination, a consideration of channels quality and business QoS (Quality of Service) requirements. Many theoretical studies have attempted to give the best compromise between frequency utilization efficiency and business performance [8].

The existing common algorithm about wireless channel allocation should balance the throughput and fairness. Literature [9] introduces CoMP-MU-MIMO (Coordinative Multiple Point-Multi User-Multiple Input Multiple Output) schemes and the performance evaluation corresponding uplink system level. Literature [10] introduces a downlink implementation plan of CoMP (Coordinative Multiple Point) systems. With the new wireless business continuing to appear, the QoS also highlights the characteristics of diversity. The LTE (Long Term Evolution) system is required to support different requirements of business QoS. We should consider the throughput and fairness in the design of algorithm of channel allocation. QoS requirements also should be put into consideration in the design for LTE-A(Long Term Evolution-Advanced)system allocation algorithm. The collaborative scheduling algorithm is not perfect, cooperative scheduling does not have an algorithm own advantage and use resource reasonable after relay joining the queue algorithm yet. There still exist a lot of problems to be solved in base station and layer three relay cooperative scheduling, such as interaction of the information between base stations, and how to balance the performance, complexity, amount of feedback, backward compatibility and other aspects. So the research of packet scheduling and channel allocation is a key in the implementation of the LTE-A .

The cells along the railway are set in line. The information between mobile station on the train and the network side only is referred to the cell forward and the one backward. But in urban area, we will meet the wireless resource allocation problem when the train is driving in the cell along the railway, and this problem basically involves the current cell and the next cell where the train will arrive. During the whole traveling

process the communication link between the train and ground must be guaranteed. The cells along the railway will provide services for users on the train by using the resources, but it will form negative effect to the users in the cell along the railway, such as communication interruption, failure to access and lack of wireless channels.

Therefore, a novel strategy of wireless channel allocation is proposed in this paper based on the study of the existing channel allocation scheme, which is a wireless channel allocation scheme based on the cells along the railway, It aims to solve the issue of shortage of resources in the cells along the railway. The numerical results indicate that, the novel strategy could reduce the block probability of voice and improve the channel utilization rate better.

II. System Structure

A. The existing mechanism of channel allocation

(1) FCAS (Fixed Channel Allocation Scheme): The number of current remaining channel is represented as N, and data packet to be transmitted needs M channels. If $N > M$ system will allocate M channels for packets. Otherwise, transmission request of the GPRS packet will be denied. And voice calls will be blocked only in the absence of channel [11].

(2) DCAS (Dynamical Channel Allocation Scheme): We are assuming that establishing a call for data that transmission with a maximum of M channel, according to the current available channel resources. DRA (Dynamical Resource Allocation) will allocate M channels for this call. If the remaining channel number $N \geq M$, system will allocate M channels; if $0 \leq N < M$, system will assign N channels for it; if N=0, this call will be blocked. DRA strategy actually set the minimum number of using channel 1 or 2 as default. Voice calls will be blocked only in the absence of channel.

(3) HCAS (Hybrid Channel Allocation Scheme). This strategy makes a combination of the fixed channel assignment and dynamic channel allocation strategy. Each system assigns a fixed set of channels and reserves a group of flexible dynamic channels. Schedule allocating method is mainly based on the situation of known business. These flexible channels will be allocated to the predetermined system to solve the predictable changes of known business. The use of this allocation system, each volume of business is detected on continuous or periodic. Then different business will be allocated to these channels according to these measured values. Therefore, HCAS is the product which combines FCAS and DCAS [12]. Because of the shortage of wireless resource, we can choose the strategy among the following: using FCAS with fixed GPRS package, using the least number of channels that the GPRS package default (DCAS), or existing frequency interference in the actual (HCAS). All these will reduce the efficiency of wireless resource and the utilization ratio of the channel.

B. the novel allocation strategy of radio resource

In LTE system, sub-channels can be allocated to different users to exploit the channel condition, and hence maximize the achievable data rate. At the meantime, relay aided system also attracts attention in recent years. Relays are often considered as a means to improve the performance of the infrastructure based network by increasing their coverage area and exploiting the spatial diversity. Due to the two hops in the relay aided system, the resource allocation has become an essential component. Therefore, for a relay system along the railway, in order to achieve better performance, the resource allocation is of great interest.

We put forward a novel wireless channel allocation scheme based on the mixed cells along the railway. System model is shown as Fig 1, the model includes the cells along the railway, base station in the center of cell and the relay stations at the cell edge. Based on the linear area, the fixed driving direction, and a combination of train index, we propose a novel predictive wireless channel allocation strategy. Specifically, route of driving is fixed, so according to the train index, in a network database we can store a cell list along the railway which is under the control of network side. When a train is ready to start, it needs to be registered in the network, network side can learn the information of the cells where the train will pass according to the train index [13].

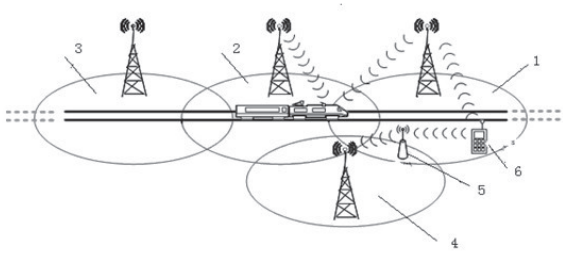


Fig. 1: The diagram of cellular model along the railway

Wireless channel allocation strategy based on the cell along the railway is shown in Fig 2, there is mobile management equipment on the train. Function of mobile management equipment includes GPS, transceiver and information collecting. Train mobile management equipment sends request to the network side to register train index, the network side can obtain the train index; according to the train index and the data along the cell previously stored, the network side makes cell list along the railway and coding. The mobile management equipment on the train sends the related information of the train to network side periodic, including the speed of train, train's location, distance to next cell, information of train stops as well as wireless resources amount in the prior two cells along the railway. After the data processing, network can get the time T1 when reaching the next area. We preset time threshold value is T2. If $T1 < T2$, network side sends a message to base station which in the next cell, informing it reserved channels resource for the upcoming train passengers. At the same time through the acquired information, we could calculate statistics reserved resource quantity of prior two cells along the railway and send the result back to the next cell, notify the cell to reserve the channel resource reasonably. The BS of the next cell recycles the wireless resource and reserves

them for the passengers according to the periodic updating information and the storage data along the railway. In this case, the quality of reserved wireless resources can be confirmed through reservation information feedback of prior two cells along the railway.

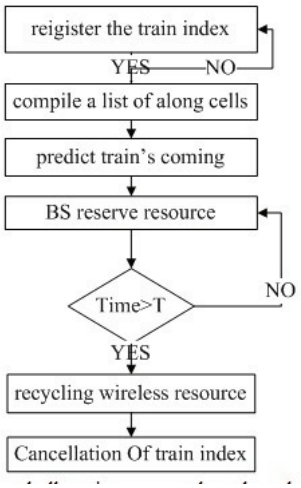


Fig. 2: Wireless channel allocation process based on the cell along the railway

As shown in Fig 1.There are communication needs in the cells where the train is pass. An access request S1 will be sent to the BS1; since the channels of recycle serves passengers on the train, BS1 feedback information indicates that channel resource is full to users, so the users in the cell tries to request the RS5 at the cell edge to access and transmits the feedback information of BS1 to the RS at the same time. RS4 receives access request and feedback information of BS1 and applies for wireless resource in short time, BS4 in the adjacent cell allocating wireless channel resource in time to serve the users in the cell along the railway through the RS4. When train left the cell at some time, BS1 recycles wireless resources for providing normal communication service to the users in the cell along the railway. At the same time, BS4 in adjacent cell recycles resource S4 through RS4. When the train arrives, network side logs off the train index. The interactive process is shown in Fig 3

III. Numerical analysis and simulation

A. Numerical analysis

In interference-limited environment, the SIR (Signal to Interference Ratio) which MS at different location suffers is a very important problem. In this part, we make analysis to the SIR of resource allocation in this system model [14]. Due to borrow channel resource from adjacent cell, interference along the railway is the largest, a link along the railway, we assume interference source number is L. d and γ_d are represented as distance of link and coefficient of path loss. P_D is the transmitting power on the link by transmitter. $d_{I,1}$ is the distance that transmitter should consider in the first link from L interference sources ,$\gamma_{I,l}$ is coefficient of path loss that

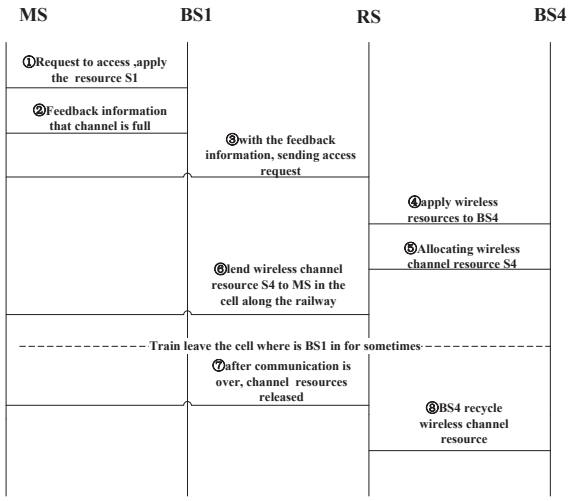


Fig. 3: Diagram of the channel allocation

transmitter should consider in the first link from L interference sources, $P_{I,L}$ is transmit power of the first interference source. SIR is given by:

$$\Gamma(d,\gamma_D,d_{I,l},\gamma_{I,l}) = \frac{P_D/PL(d,\gamma_D)}{\sum_{l=1}^{L}\left[P_{I,l}/PL(d_{I,l},\gamma_{I,l})\right]} \quad (1)$$

By our analysis, in order to obtain the maximum interference, we assume to produce maximum interference by L interference source at different location. The main interference sources that relay suffered are from adjacent cells, seeing the traditional cellular network structure, the distance from each interference source to the relay is $3R-R_0$. (R_0 is the distance from relay to mobile node, that is $R_0=\frac{2}{3}R$). When the point to point communication is complete by a two hop link, we consider SIR depends on the smaller one of two link, therefore, SIR can be expressed as the following form:

$$\Gamma_2(d_F) = \min(\Gamma_{FM}(d_F),\Gamma_{BF}) \quad (2)$$

In the formula above,$\Gamma_{FM}(d_F)$ is SIR from MS to RS when the distance from MS to RS is d_F.Γ_{BF} is SIR from BS to RS. When we consider the SIR on the link between the relay and the MS, the interference is sent from the position where the distance from RS is $3R-R_0$, therefore, we can get the SIR function on the relay's link:

$$\Gamma_{FM} = \frac{{d_F}^{-3.5}}{6\times(3R-R_0)^{-3.5}} \quad (3)$$

We focus the SIR on the link from BS to relay now, this SIR is a constant:

$$\Gamma_{BF} = \frac{(\frac{2}{3}R)^{-2}}{6\times(3R-2R_0)^{-3.5}} \quad (4)$$

Where there is a direct link between the MS and BS, we set the distance from MS to BS is a constant d_{MB}, the SIR in this link is given:

$$\Gamma_{BM} = \frac{{d_{MB}}^{-3.5}}{6 \times (3R - R_0)^{-3.5}} \quad (5)$$

System resources includes time slot, frequency and code, the distribution of Erlang B is a kind of typical business model. We can analysis the block rate through Erlang B model. If the resources of system available, we can allocate resource to the new call, but the resource is occupation, the next call will be blocked. There is no queue cache. call reach obey Poisson distribution, business time obey negative exponential distribution. In theory system is in balance state. The block rate of Erlang B is given by:

$$p(x) = \frac{A_x/X!}{\sum_{i=0}^{N} \frac{A^i}{i!}} \quad (6)$$

Where, A is the number of calls in average time, N is the number of system total source.

Here we could study the spectrum utilization rate. The spectrum utilization rate is defined by the average number of bits per second per hertz in unit area [15].

When we make a calculation of spectrum utilization rate, we use Monte Carlo simulation algorithm, at each simulation, when there is a direct link between MS and BS, we obtained the capacity of the channel:

$$C_{1,E} = B_{CH,E} \log_2^{(1+\Gamma_{MB}(d_{BM}))} \quad (7)$$

$\Gamma_{MB}(d_{BM})$ is the SIR on the link between BS and MS when the distance from BS to MS is d_{BM} . $B_{CH,E}$ is the channel bandwidth when using the distribution program by relay. Similarly, when MS establishes a two-hop link, the channel capacity can be expressed as:

$$C_{2,E}(d_F) = B_{CH,E} \log_2^{(1+\min(\Gamma_{FM}(d_F),\Gamma_{BF}))} \quad (8)$$

Where d_F is the distance from MS to RS. $\Gamma_{FM}(d_F)$is the SIR on the link between MS and RS, Γ_{BF} is the SIR on the link between BS and RS.

We assume that there are N1 direct link between MS and BS in the cells along de railway, and there are N2 two-hop links between MS and RS in the adjacent cells. Therefore, the capacity of cell is:

$$C_{CELL} = \sum_{M=1}^{N_1} C_1(d_B) + \sum_{N=1}^{N_2} C_2(d_F) \quad (9)$$

In this system model, we borrow the channel resource from adjacent cells, so there are three cells in each group, so the spectrum utilization rate of unit area is:

$$\eta_E = \frac{3 \cdot C_{CELL}}{A \cdot B_{TOT}} \quad (10)$$

Where A is total area each group, B_{TOT} is total bandwidth each group.

B. System simulation and analysis

At this stage, we will simulate for evaluation the effect of the scheme we proposed. We propose the following parameters: $B_{TOT} = 108MHZ$, R=500m, bandwidth of each carrier channel is equal, as 200 kHz, the transmitting power of the mobile stations is equal to the transmission power of relay. When calculating the spectrum utilization rate we assume that the cell is in full load.

1) simulation of blocking probability: We assume the total bandwidth in each group is 108MHz, so number M of the carrier channel in each group is equal to 540. We assume that there are 27 group channels, so there are twenty sub-channels in each group

We make experiment exceeding one thousand times by using Monte Carlo model. In every simulation experiment, users are randomly distributed in the cell. Through computer simulation, we can get the result as is shown in Fig 5.

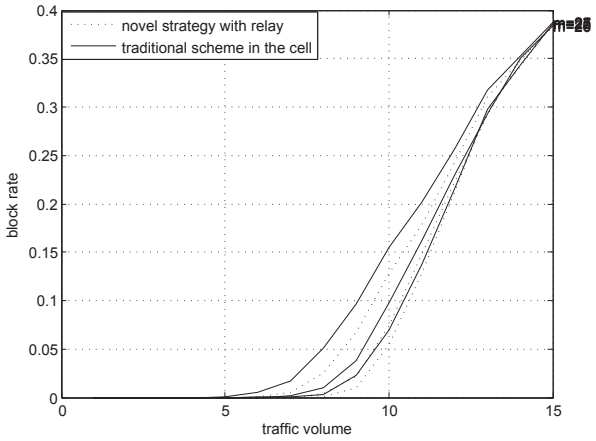


Fig. 4: The simulation of blocking probability

The horizontal coordinate is the traffic volume and vertical coordinate is blocking rate. We can learn from the Fig 6, when the traffic is small relatively, such as from zero to ten, that is to say the flow density A is between zero and one. The blocking rate of novel allocation strategy is lower than that of traditional channel allocation strategy. In the case of low traffic flow, the novel channel allocation strategy is better than the traditional. But with the increasing of A, such as from one to one point five. The performance of novel allocation strategy is close to the traditional scheme. Even later blocking rate is higher than the traditional mode. Because without channel can be borrowed.

On the other hand, there are M channels in each cell. As we can see from the fig the more m is large, the smaller is blocking rate. This is consistent with the general imagination, the more channel numbers, the larger the call numbers of accommodate. But with the increasing of A, such as from one to one point five, the density of call reach is greater than the density of call release. Due to the observation for a long time, even if the m point increases a little, It's no use to such a long time (it will be done soon ,then continue to block). But in the case

of lower density, it is the most effective way to reduce the blocking rate.

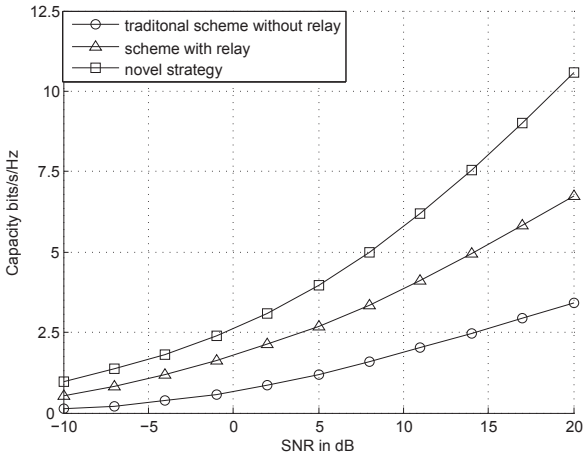


Fig. 5: Three distribution draw of spectrum utilization

2) Simulation of the spectrum utilization rate per unit area: The horizontal coordinate is the value of spectrum utilization rate. And vertical coordinate is SNR (signal to noise ratio). We calculate the average spectrum utilization rate. Results show that: average spectrum utilization rate of novel scheme based on the cell along the railway is 5.30 bits/Sec/Hz/km^2; average spectrum utilization rate of the scheme in the cell based on relay is 4.02 bits/Sec/Hz/km^2 as well as in the cell along the railway. Average spectrum utilization rate of traditional scheme is 2.38 bits/Sec/Hz/km^2. In all, the performance of novel strategy is optimal. Spectrum utilization rate improves 31 percent than the scheme with relay.

IV. Conclusion

In a view of performance, Relay channels improve the performance through spatial diversity by using additional paths between source and destination [16]. the proposed novel wireless channel allocation strategy according to the train index and mobile management equipment on the train to allocate the wireless channel predicatively. The numerical results show that: in the low complexity, it has good performance that was proven by computer simulations. The novel scheme has a high stability as well as a high capacity and channel utilization rate. It also improves the performance of SIR in cells.

Acknowledgment

This paper is supported by the State Key Laboratory of Rail Traffic Control and Safety(Contract No. RCS2010ZZ004), Beijing Jiaotong University. This paper are also supported by Program for Changjiang Scholars and Innovative Research Team in University under Grant No. IRT0949 and the Joint State Key Program of the National Natural Science Foundation of China and the National Railway Ministry of China (Grant No. 60830001). The authors would like to thank the anonymous reviewers for their helpful comments that improved this paper.

References

[1] N. S. Nourizadeh. H and T. R. Daly, "Performance evaluation of cellular networks with mobile and fixed relay station." *IEEE Vehicular Technology Conference.64th.*, 2006.9.

[2] C. H.-F. M. Mehul., "The capacity of several new classes of semi-deterministic relay channels[j]," *IEEE TRANSACTIONS ON INFORMATION THEORY.*, pp. 6397–6404, 2011,10.

[3] R. P. T. Imic, D. C. Schultz and P. Wienert., "Capacity of a relaying infrastructure for broadband radio coverage of urban areas." *IEEE 58th IEEE Vehicular Technology Conference.*, 2003-Fall.

[4] D. C. S. S. J. A. K. Park, C. G. Kang and J. Ihm., "Relay-enhanced cellular performance of ofdmatdd system for mobile wireless broadband services." *ICCCN 2007, In Proc. of 16th*, pp. 13–16, 2007,8.

[5] O. Oyman., "Ofdm2a: A centralized resource allocation policy for cellular multi-hop networks." *40th Asilomar Conference on Signals, Systems and Computers,*, pp. 656–660.

[6] M. Kaneko and P. Popovski., "Adaptive resource allocation in cellular ofdma system with multiple relay stations." *Proc. of IEEE 65th VTC*, pp. 3026–3030, 2007-Spring.

[7] J. A. T. Girici, C. Zhu and A. Ephremides., "Proportional fair scheduling algorithm in ofdma-based wireless systems with qos constraints." *12th International OFDM Workshop (Inowo 07),Hamburg, German*, 2007.

[8] S. A. D. R. G. M., "Wavelet-based downlink scheduling and resource allocation for long-term evolution cellular systems." *IET COMMUNICATIONS.*, pp. 2091–2095, 2011.9.23.

[9] J. L. G. L. C. C. Dajie. Jiang, Qixing. Wang, "Uplink coordinated multi-point reception for lte-advanced systems[j]," *Wireless Communications,*, 2009.

[10] X. Y. L. C. Jianchi. Zhu, Xiaoming. She, "A practical design of downlink coordinated multi-point transmission for lte-advanced[j]," *Vehicular Technology Conference*, 2010 ,71.

[11] B. G. Logrippo, "Understanding gprsthe gsm packet radio service." *Computer Networks*, vol. 34, pp. 763–779.

[12] M. S. B. M. ErmelT. MtillerJ. Schiller, "Performance of gsm networks with general packet radio services," *Performance Evaluation*, vol. 48, pp. 285–310, May 2002.

[13] M. K. Kim and H. S. Lee., "Radio resource management for a two-hop ofdma relay system in downlink." *Computers and Communications,ISCC 2007.*, pp. 25–31, 2007.

[14] Z. A. H. C. L. Jiang, "On the achievable diversity multiplexing tradeoff for the optimal time allocation in the two-way channel," *IEICE TRANSACTIONS ON COMMUNICATIONS.*, pp. 2624–2628., 2011.9.

[15] H. Hu and H. Yanikomeroglu, "Range extension without capacity penalty in cellular networks with digital fixed relays," *IEEE globecom 2004*, pp. 3053–3057.

[16] V. V. Y. Liang and H. Poor, "Resource allocaiton for wireless fading relay channels: Max-min solution," *IEEE Trans. Inf. Theory,*, vol. 53, pp. 3432–3453.

Enhancement of Security in Cognitive Radio Network

Nouman Maqbool Rao[1], Rao Zeeshan Maqbool[2] and Rao Imran Maqbool[3]

[1]Higher Education Commission Islamabad Pakistan
[2]Oman Telecom, Muscat, Oman
[3]Dascon Engineering Lahore, Pakistan

Abstract – *Reference to the latest developments the spectrum shortage problems are because of the wireless communication. For the future networks the most practical, scientific and systematic challenges is the use of licensed or unlicensed wireless networks with opportunistic use of the spectrum with, without and limited rules. Since different wireless networks are using different frequency bands. So there is a need to use lessening bands when there is no activity on them. Cognitive radio is a new technology which leads to solve these problems through dynamically utilization of rules and spectrum. Several spectrum sharing schemes have been proposed. Now a day's security in cognitive radio network becomes a major and challenging issue, and chances are prearranged to the attackers in cognitive radio technology as compared to the wireless networks in a general form. In cognitive radio, mobile station equipment may switch to any available frequency band, as it makes a list of available free channel and make handoff decision accordingly. So whenever handoff is made whether soft or hard there will be a chance that malicious attacker may hack ongoing traffic or he may even interrupt established traffic by imitating any kind of passive or active attack like interception, spoofing denial of service etc. This paper explore the key challenges to provide security in cognitive radio networks, and discusses the current security carriage of emerging IEEE 802.22 cognitive radio typical and recognizes security threats and vulnerabilities along with the countermeasures and solutions.*

Key Words— *Cognitive Radio, IEEE 802.22, Security Threats, Sensing*

1. Introduction

Communication is growing changing and increasing subscriber base. The occurrence of high data throughput application continue to increase the fast growing request for broadband wireless service these have led to the expansion of several wireless technologies which continuously grow with ever-increasing competencies. Under licensed frequency band IEEE has offered multiple standards. While 802.16a/d/e and 802.20 have focused on providing the necessary infrastructure to create wireless metropolitan area networks (MAN) which has the radius of approximately 1km to 5 kms, 802.22 is pursuing to define a standard Capable of serving vast regions up to 100km in size [14]. Frequency band of unlicensed frequency is being utilized by using IEEE 802.22. A typical working group (WG) is working to finalize the standard of 802.22 after the Federal Communication Commission (FCC), which passed the resolution. 802.22 is also called wireless radio area network (WRAN) or cognitive radio network (CRN) by IEEE [13].

Most of the radio function as a software base that run on microprocessor and programmable electronic devices. These technology is referred to software define radio (SDR). Cognitive Radio (CR) is further enhance as compared to SDR by employing software for measurement of the vacant portion of the wireless spectrum which is already there and operate that spectrum in a way that bound the interfering with other devices [11]. In dynamic spectrum Access (DSA) licensed; user is stated to be as a primary / key user or occupants. The user who didn't have any licensed that got the permission for spectrum opportunistically are referred to as secondary user [4].

If we compared with the typical radio networks, Cognitive Radio (CR) is more flexible and exposed to wireless network. Therefore more threats and vulnerabilities found then traditional radio environment. For example CR first senses the spectrum which is scanning a certain range of the spectrum to identify unoccupied range / spectrum. During this methodology the secondary user can determine that which spectrum can be used either radio or not. When the result of spectrum sensing in altered maliciously network activities which are normal will be disabled, even whole traffic may be broken down. [3] [7] CR is the main technique which realizes DSA policy

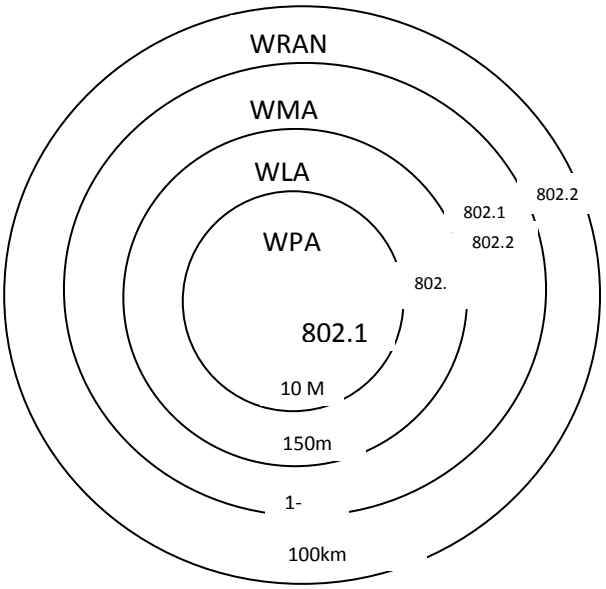


Fig. 1: Categorization of the various

The composition of this research paper is as follows. Under section II Cognitive Radio Architecture is reviewed. In III Security overview is presented. In section IV security threats and vulnerabilities are focused. In section V literature review will be discuss. In section VI countermeasures and solutions are describe. In section VII conclusion and future work is given.

2. Cognitive Radio Architectures

Basic component of the cognitive radio is shown in figure II. The operating system (OS) characterizes the higher-layer communication; Operating system can generate and receives the traffic information. The sensing component measures the parameter of the radio atmosphere and transforms the parameters to the cognitive engine. The cognitive engine than combine the information received from sensor and with policy information to make an appropriate decision about how and when it will transfer / communicate by using the radio transmitter and receiver. Some Cognitive Radios (CRs) also depend on information of transmitter location which is provided by a geo locater.

Cognitive Radios could be broadly categorized into one of three network architectures which is as shown in Fig. III. They could range with reference to architectures which include all the six components in one single non-cooperating device to networked architectures where none of the CR components may be co-located with each other. This architecture includes multiple examples of each module. Further to this there are many distributed CRs which may select to share the information such as measurements, location, or policy in order to make more knowledgeable and synchronized communication decisions. In the cognitive engine, the other CRs are effectively sensing, geo-location, or communication extensions. [3] [11].

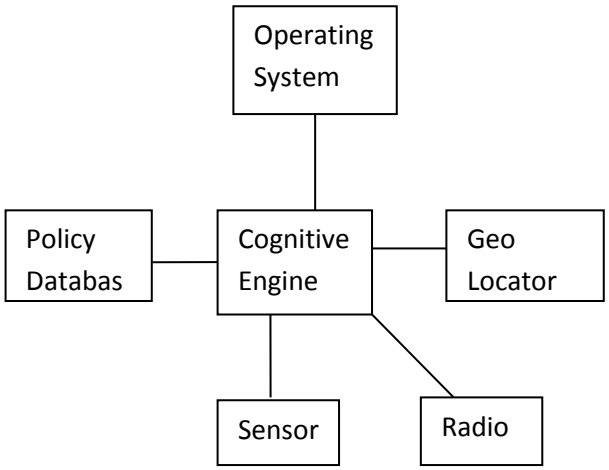


Fig. II: Cognitive Radio Component [11]

The security vulnerabilities will be occurred when any cognitive radio components are not synchronized with each other.

3. Security Overview

Cognitive Radio itself is a broader term with many potential senses. Whereas the security requirements may be different with application environment; usually there are some common requirements providing basic safety controls like cognitive radio networks. For which the security requirements are the same as in a general wireless networks because of the nature of operating on wireless media. These security requirements are outlined below. [9]

3.1. **Access Control:** The transmission of information from an object to a subject is called access. Control of access to a resource is one of the major objectives of security. Access control addresses more than just controlling which users can access which files or services. The relationships between subjects and objects are generally covered under the term of Access Control.

3.2. **Confidentiality:** Confidentiality is defined by the International Organization of Standardization (OSI). Confidentiality involves by make it sure that each part of a system is appropriately secured and Accessible only by subjects who need it.

3.3. **Authentication:** It is a process of verifying or testing that the demanded identity is valid. Authentication requires that the subject provide additional information that must exactly correspond to the identity mentioned. In this regard Password is the most common form of authentication.

3.4. **Integrity:** Integrity it offers a high level of assurance that the data, objects, and resources are unaltered from their original protected state. This includes alterations occurring while the object is in storage, in transit, or in process.

There is multiple consideration factors when the security is implemented in cognitive radio network due to the nature of CR communication, such as the flexible utilization of frequency range / spectrum and the appearances of different licensed user is unscheduled. So the additional special security issues need to be considered especially. For example it will be more difficult to authenticate the identity of the licensed user at present there are still not completed and the final solution to solve the security problems bought by CRN [3].

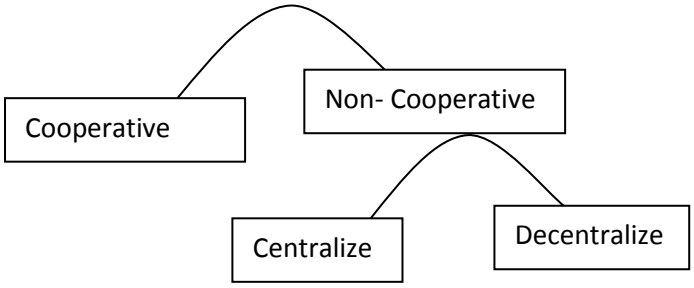


Fig. III: Cognitive Radio Network

4. Security Threats & Vulnerabilities

4.1 Rogue Base Station Attack

The attackers lie in the center of the Base Station Offer (BSO) and Base Station Renter (BSR). The attacker could take off himself as an occupant and can send back a resource return message to the offer. In the same way the attacker could pose as offer and request for resource allocation. Attacker can generate the BS-IDs and can forge the network. Attacker can get BS-IDs and negotiated channels between the offer and renter at the time of resource sharing. The attacker abuses this information and sends bogus messages [5]

4.2 Replay Attack

The attacker can capture the packet and resend these packets maliciously after passing some delay for mismanagement of the network characteristic & resources. Attacker can retransmit the packet after certain time to cause DoS [5] [8].

4.3 The Motivation of Attacks

The motivation of attack in cognitive radio network are classifies into four types [6] [9].

Selfish Attack: The attacker can access the spectrum with high priority. He occupies the spectrum resource as he wants.

Malicious Attack: The opponent slows down the unlicensed user from spectrum usage and caused denial of services (DoS).

Cheat Attack: The attacker increase his effectiveness function as well as at the same time decrease contestant's profit.

Misbehaving: CR didn't follow any general rule for sensing and managing of range of the spectrum.

4.4 Denial of Service (DoS)

During the period of sensing any attacker / hacker can flood the spectrum with inconsistent data to show off that the range is unavailable for the same because of this the sensing during is the most vulnerable to DoS. The major objective of DoS attack is to put the burden on the resources and to stop the utilization of the resources for the nodes which are also present in the network [7] [8].

4.5 Incumbent Emulation (IE) Attacks

Under the Incumbent Emulation the user tried to get the priority on the other secondaries by sending signals which could help to emulate the features of an incumbent. The impact of Incumbent Emulation attack is the genuine secondary's abilities for differentiate the signals of attacker with the actual incumbent signals during the sensing period [4] [8].

4.6 Spectrum Sensing Data Falsification (SSDF) Attacks

The hacker can spoof or mask primary user and transmit incorrect spectrum deduct sensing outcomes to a data collector, which is a major source of incorrect spectrum detection decision taken by the data collector [3] [4].

4.7 Policy Radio Threats

This is artificial intelligence (AI) threat comes by two aspect absence of any policy as well as the failure while a policy is being used. The attacker can block the policy or jam the radio which announces the policies or Attacker can modify policy or implement false policies which are in his favor [3] [10].

4.8 Learning Radio Threats

In this AI threat CR can be mature from the past experience as well as with the current situation to predict and assume the future environment and to identify the optimum processes. Attacker can change and alter the past statistics of a network for impact of the CR prediction efficiency. [3] [10].

4.9 Parameters Threats

By using different parameters Cognitive Radio has to control the operation and assess its performance. There is mixture of characteristics for these types of functions. Some parameters are used to estimate the performance of CR as well as the weigh. Attacker can alter these

characteristics which can be a reason for sub-optimal for the CR along with wrong operation [3][10].

4.10 Spectrum Management Threats

The function of spectrum organization is categorized as analysis and decision. Threat here comes from the possibility of incorrect parameters which affects the result for spectrum decision and analysis. CR may select wrong spectrum as a result the communication performance within a network may be damage [3] [10].

4.11 Common Control Channel (CCC) Attacks

The successful jamming may stop or delay communication across a large frequency range in this regard the DoS attacks are the target for the Common Control Channels. [11].

4.12 Transmitter / Receiver Failures

The attacker could get the control on the transmitter and can restrict to practice of an attack that is possible with any radio network to send as well as restrict with the licensed users of the Radio Network. This is also open the probability for the Sybil attack whenever it transmits the radio by using multiple characteristics, out of those some behaves while some others are misbehave [11].

4.13 Spurious Transmissions in QPs

The attacker can post any threat by using 802.22 as a fake transmission which ultimately results as a congestion in quiet periods (QPs). The opposite party can be interfere by multiple co-occurrence approved during the Quiet Periods which relates to the controlled procedures and he cause hardware or software defects [8].

5. Literature Review

[1] Discuss the idea of cognitive radio network by J. Mitola from software define radio (SDR) which is originally considered to improve the spectrum utilization. CR is an intelligent communication system which is aware of the environment. CR completes two major objectives which are extremely trustworthy communication when and wherever needed and effective operation of the radio spectrum. Paper discuss major three type of network architecture in CR. 1-Infrastructure, 2- Add-hoc and 3- Mesh architectures.

The author use pictorial diagram which helps the reader to easily understand the topics the paper describes. Paper used vast referencing in his paper. This paper cannot complete the whole architecture of cognitive radio network. More work related to CRN is pending.

In [2] author of the paper said that network layer can use to integrate MAC's and PHY layer for better service. And it discusses the mathematical framework for routing trust in CRN. Network layer structure can provide better service in MAC and PHYs and it integrate transport and application layer. Paper discusses the location management handoff management, security attacks and misbehaviour and security services. Papers discuss the security service which is provided in CRN. For all those services of security deliver the security enable environment as compare to malicious threats and trouble sameness. In this paper author used best and updated referencing. Paper used mathematical preposition in trust CRN. Author cannot simulate to his mathematical propositions. And cannot propose any countermeasure against malicious attack and misbehaviours.

The security threats in CR/CRN as discussed in [3]. The attacks are artificial intelligence behavior threats and dynamic spectrum access threats. Author Use the best referencing in his short paper. Paper is categorized and arrange in the appropriate manner. Author of this paper use mathematical notation. Author cannot propose any model to prove his point of view. The countermeasure of the attacks cannot discus.

The author of the paper [4] defends Incumbent emulation attack and SSDF attack. Two technique are used one is DRT which is distance Ratio Test while the second is DDT which is Distance Different Test to eliminate the IE attacks. Author use two levels in the first all native spectrums deducting result should be validated by the data receiver. Whereas second layer of protection is placement of data synthesis schemes which are strongly against attack of SSDF. The author use simulation result to prove his result and to prove his simulation the author use diagrams. No mathematical formula is used to prove his simulation.

Paper [5] discuss the potential attacks these attack are rouge base station attack and replay attacks to eliminate these attacks author proposed three solutions . These solutions are Time Stamp, Nonce, and Digital Signature. Most of the issues are discussed in detail along with appropriate proposed solution. The major weakness which was figured out is its limited vision in context of defining solution analysing criteria.

Paper [6] author describe the access point(AP) attack and misbehaviour and to encounter these attack and misbehaviour he proposed Locdef technique verifies that given signal is incumbent transmitter which estimates its location and signal individuality. So trust relationship is proposed to avoid unauthorized nodes attack to CR. Paper is categorized in good referencing. Paper discusses misbehaviour and attacks in detail tabular format which is easy to understand. Paper cannot fulfill the countermeasure of the Access Point misbehaving and Access Point attacks.

To eliminate the DOS attack flow control can be introduced [7] at MAC layer to validate the genuine nodes of the network during the channel compromise phase. In this phase no major authority or reliable outsider party is involved. The paper proposed a result to identify malicious node. Updated referencing is used in this paper. No Verification with its examination is discussed in this paper. No simulation work is done .The author point of view is very rare.

Paper [8] describe the security threat and it present two solutions which is Key management Infrastructure and Distributed key Management. Graphical Simulation and pictorial diagram used. Author cannot complete the arguments of Possible DoS Threats in Cognitive Radio Network and their Countermeasure

Author of the paper [9] discuss two approaches which is protection based layer considering different protocol layers and 2nd is detection based layer considering different protocol layers. Paper discusses protection and detection on different layer on more detail. Pictorial diagram can be used. No main security threats is discuss in this paper Countermeasures are not full filling the whole security parameters.

In [10] author describe the threats to CR and dynamic spectrum access threats and proposed Three Mitigation technique robust Sensory, Mitigation in Individual Radios and Mitigation in Network. The paper provides well background knowledge about the CR architecture and pictorial representation to clarify the security issues. Security threats and solutions are well defined, respectively. We did not found any significant weakness in the paper except the proven theory may be more validated by simulation in NS-2 or MatLab.

Paper [11] paper discuss 7 remedies against the failures they are negotiated supportive CR, CCC attacks, occupancy failures by spectrum, policy, location, sensor and by transmitter as well as receiver. Paper discuss the main security threats and as well their countermeasure. Different security can be followed against the attack. Author cannot prove his countermeasure with any model through simulation. No pictorial diagram is used in this paper.

6. Countermeasures & solutions

The countermeasure of Rogue Base Station and the return attack. For succeed on those declared attacks, our research paper [5] proposed three constraints or strategies to secure the sharing of network; which are Digital Signature, Nonce and Timestamp.

Time stamp help to prevent the replay attack. The mixture of data along with the time of dispatcher is the Time Stamp if these packets are newly generated then it is receive otherwise it is discarded.

In nonce repeat packet is discarded so DoS and replay attacks can be eliminated. Digital Signatures are used to validate the dispatcher and for recognize the alternation of received packet. Applying DS built verification of the dispatcher is actual to escape the above mention attacks.

In [6] author has proposed that LocDef arrangement authenticates either a given signal is that of an incumbent transmitter with the approximating of its position as well as detecting the signal features. LocDef could be helpful to remove or moderate some of the above-mentioned drawback. This scheme can eliminate the motivation of attacks.

Malicious nodes could be thrown the undesirable packets on the channels to halt this undesirable packets [7] discuss the impression of flow control which could be initiated at MAC level with the inclusion of time limitation. Receiver describes the monitoring of Time Interval (TI) that is why the sender is unable to transmit the data regularly. If sender spreads the data on the high rate and receiver is receiving packet regularly its means that the mentioned TI and the receiver identify the misconduct by one point / node which spread the information about malicious node.

The key point is to protecting beside IE threats is to develop any new technique which could able to handle these situations and for validating the genuineness of the incumbent signal. Paper [4] discusses the solution of IE attacks. One approach is a signature which is embedding in the incumbent signal. One more process is to work and verification procedure with incumbent transmitter and an authenticator. Two techniques are being used the first is DRT which is Distance Radio Test this use RSS which is received signal strength quantities gained from the location verifier (LV). Other technique is known as DDT which is distance difference test. This procedure is being used whenever the signals are being transmitted by a signal point to LVs, the virtual phase variance could be identified whenever the signal influences the two LVs because of opposing locations from the sender.

Two prevent SSDF attack paper [4] proposed two level of defense. The first phase of all native spectrum deducting result must be validated from data collector. The main objective is to avoid the return attacks of untruthful data inoculation by the objects outside the networks. Second phase of protection is placement of data synthesis arrangement that is forceful with compare to attacks of SSDF.

In case of policy attack paper [11] suggests that in cooperative policy can be freely exchanged and in non-cooperative nodes policy updates and renewals can require infrequent. Effective rules could be replaced freely and with self-assurance and kept for long time. It is difficult that attacker stops a CR even presence of some rules and regulations. Paper [10] elaborates that without the knowledge of policy attacker can use different funny and obvious techniques to suppose about policy. This comes into picture that the radio rule and regulations should be carefully check and validated to defend against the threats.

To improvement against learning, parameters and spectrum management threats paper [10] present a solution robust sensory input and mitigation in Network. In vigorous sensory the data entry educating sensor, input can be considerably in helping in reduction of the acceptance of CR. In scattered situation the network of CR can fuses sensor data to increase throughput. All sensor contribution would consider noisy with or without the occurrence of attackers, statistic can sometimes incorrect.

Author of the paper [11] defends against the common control channel use a robust coding of different spread spectrum. The schemes of the media access would be vigorous which could provide the fair access of data on the network. This fairness had to be brought around by the multiple layers and the simple access arrangements which should have focus on the control channels for which the need is preferable.

In Key management infrastructure the security sub layer didn't resolve the issues of inter-cell key management. Whereas the sub layer comprises with PKM protocol, this protocol can handle the intra-cell solutions which didn't allow the rations to care the administration of inter-cell solutions. The workable method for planning of inter-cell solution is to operate the backhaul substructure which would be able to interconnect many WRANs. These are linked for ACR via backhaul stations. If a common backhaul infrastructure amongst cells is not available the scattered significant organization system is required over there. For these types of arrangements 802.22 BSs helpfully operate a disseminated algorithm for inter-cell managements. Key management scheme is in two types; one is contributory while the other is distributive. Contributory group is well-defined for incorporate scheme as an output of joint struggle of different points / nodes. The distributive group contains arrangements there all important invents by one nodes [8].

Paper [11] discusses related arrangements, not reliable third party who is accountable for the generation and circulation of the cryptographic keys. And in the scattered arrangements all the nodes produces a unique key and issues it to the others nodes of the network.

In [9] author has proposed that Physical layer attacks. Jamming of signals can be prevented by spread spectrum scheme. Scrambling is another Physical layer attack that can be a countermeasure for monitoring and deducting of system anomalies. Second and secure is MAC layer that uses X.509 certificate for authentication mechanism between Mobile Station (MS) and BS. X.509 certificate holds the Public Key (PK) of MS.

Another attack which has been identified is rogue BS. Its countermeasure is done by joint verification at user-network level. Mutual Authentication can be performed after scanning, achievement of channel explanation as well as reaching and competence concession which are built on EAP with specific authentication method as EAP-Transport Layer Security [9].

7. Conclusion and future work

Cognitive radio introduces a new level of sophistication to wireless communication technology. Still CR procedures and methods are at the initial ages. Security is an important part in CR. Under this research paper we discussed in detailed a comprehensive variety of security attacks for IEEE 802.22. These threats and attacks on IEEE 802.22 can be potentially carriage a danger to the performance and sustainability of CR network. Literature review based on references material proposes CR facing bunch of threats. The security in spectrum sensing threats may be proposed in future work.

8. References

[1] K.-C Chen, Y.-J Peng, N. Parasad, Y.-C Liang, S. Sun "Cognitive Radio Network Architecture: Part I – General Structure" ACM 2008

[2] K.-C Chen, Y.-J Peng, N. Parasad, Y.-C Liang, S. Sun "Cognitive Radio Network Architecture: Part I I – Trusted Network Layer Structure" ACM 2008

[3] Yuan Zhang Gaochao Xu Xiaozhong Geng " Security Threats in Cognitive Radio Networks" High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference.

[4] Ruiliang Chen Jung-Min Park Hou, Y.T. Reed, "Toward secure distributed spectrum sensing in cognitive radio networks" Communications Magazine, IEEE Publication.

[5] Shaukat, R. Khan, S.A. Ahmed, A. "Threats Identification and their Solution in Inter-Basestation Dynamic Resource Sharing IEE-802.22" IEEE International Conference on Convergence and Hybrid Information Technology 2008.

[6] Arkoulis, S. Kazatzopoulos, L. Delakouridis, C. Marias "Cognitive Spectrum and Its Security Issues" This paper appears in: Next Generation Mobile Applications, Services and

Technologies, 2008. NGMAST '08. The Second IEEE International Conference.

[7] Shaukat, R. Khan, S.A. Ahmed, A. "Augmented Security in IEEE 802.22 MAC layer Protocol" Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th IEEE International Conference.

[8] Kaigui Bian and Jung-Min "Security Vulnerabilities In IEEE 802.22" ACM 4th Annual International Conference on Wireless Internet

[9] Xueying Zhang, Cheng Li "The security in cognitive radio networks: a survey" ACM 2009 International Conference on Communications and Mobile Computing.

[10] Clancy, T.C. Goergen, N. "Security in Cognitive Radio Networks: Threats and Mitigation" Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd IEEE International Conference.

[11] Timothy X Brown, Amita Sethi "Potential Cognitive Radio Denial of Service Attack and Remedies"

[12] Cordeiro, C.; Challapali, K.; Birru, D.; Sai Shankar, N. "IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios "New Frontiers in Dynamic Spectrum Access Networks, 2005 First IEEE International Symposium

[13] Krenik, W. Batra, A. "Cognitive radio techniques for wide area networks "Design Automation IEEE Conference, 2005.

[14] Justin Thiel "Metropolitan and Regional Wireless Networking: 802.16, 802.20 and 802.22"

Physical Layer based LTE and WiMax signal Auto-Detection using Correlation based Parameter Estimation

Muhammad Salman Khan, Sana Siddiqui

Department of Electronic Engineering, NED University of Engineering & Technology, Karachi Pakistan

Abstract *– In this paper, we propose a baseband, correlation based blind estimation method for the identification of LTE and WiMAX signals in the uplink communication scheme. LTE and WiMax standards are centerd around orthogonality using the FFT and IFFT techniques. The time parameter estimation based algorithms sequentially estimate the symbol duration and the cyclic prefix whereby useful timeslots can be calculated through these estimates. The proposed technique works over LTE and WiMax standard with a limited sample of data without any prior information like SNR level and symbol rate. The Matlab simulations display fast and accurate performance for both AWGN channel and Rayleigh fading channel. The proposed algorithm can also be used for identifying various radio access systems and hence a single terminal can be used for the reception of multiple air interfaces.*

Keywords: LTE, WiMax, Blind Estimation, ODFMA, SCFDMA, Parameter Estimation

1 Introduction

Cognitive Radios or Software Defined Radios are the most viable solutions to handle spectrum scarcity problem which arises due to the ever increasing augmentation in number of users and applications that require very high data rates. These software defined reconfigurable radio terminals can recognize different wireless networks and are capable of estimating the basic parameters of various air interfaces.

Orthogonal Frequency Division Multiple Access (OFDMA) is one of such promising techniques that enables high spectrum efficiency as well as robustness and is very effective for RF dispersive environments. OFDM is already employed as the modulation scheme for wireless communication systems like Wireless Local Area Network (WLAN) IEEE 802.11 and Wireless Metropolitan Area Network (WMAN) based on IEEE 802.16. Single Carrier - Frequency Division Multiple Access (SC-FDMA) is another derivative of OFDMA that utilizes single carrier modulation. It has an additional benefit of low peak-to-average power ratio (PAPR) compared to OFDMA. SC-FDMA is currently adopted as the uplink multiple access scheme for 3GPP LTE.

Some of the most promising spectrum sensing algorithms exploits the cyclostationarity property of communication signals. Cyclostationarity based spectrum sensing algorithms have been proposed in various papers. These algorithms do not require any explicit assumptions on the data or noise distributions. [10]

2 Background

The paper [1] exploits the cyclostationarity of OFDM signals and estimates the symbol duration by maximizing a sum of modulus squares of cyclic correlation estimates in the cyclic domain. In paper [2], under the assumption of low signal-to-noise ratio, the joint maximum-likelihood (ML) phase offset and Symbol Timing estimator for additive white Gaussian noise (AWGN) channel is proposed. The estimate depends on both the non-conjugate and the conjugate correlation function of the transmitted OFDM/OQAM signal and exploits the cyclostationarity of the OFDM/OQAM signal. Paper [3] discusses the blind time parameter estimation of OFDM signals in multipath fading channels based on correlation algorithm. Correlations of the trial source with alterable correlation length are calculated and a peak is acquired when the correlation length equals useful symbol duration. The site of correlation peak equals the length of useful symbol. Symbol duration is estimated by finding the distance between the peaks at the correlation and the guard time can be worked out. Paper [4]-[5] is also based on the blind time parameter estimation for OFDM signals using correlation and exploits the cyclostationarity property of these signals. Most systems are based on OFDM modulations but differ from their inter carrier spacing used in OFDM modulation**.** In Paper [6]-[8], efficient algorithms

based on second order statistics, matched filter and maximum likelihood estimation has been proposed respectively, for the estimation of inter carrier spacing of OFDM systems. Paper [9] discusses the estimation of symbol duration and cyclic prefix length using maximum likelihood estimator.

In the presented research, the proposed blind estimation method is divided into three steps: First of all, symbol duration is estimated by finding the distance between the peaks of the correlation result then, CP duration is estimated by performing a correlation test and finally the useful time can be worked out from the two results.

The paper is organized as follows. Section 3 describes the system model. The cyclostationarity of the OFDM and SC-FDMA signal is discussed in section 4 with the theory and mathematics for autocorrelation. Proposed single and multi carrier signal identification and parameter extraction algorithms are given in Section 5, followed by simulation results in Section 6. Finally, the paper is concluded in Section 7.

3 System Model

OFDM converts serial data stream into parallel blocks of size *NFFT* and modulates these blocks using inverse discrete Fourier transform (IDFT). The equivalent complex representation of OFDM signal is

$$y(t) = \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \frac{a_{n,k}}{\sqrt{NK}} e^{j2\pi \frac{n(t-T_s)}{NT_s}} g(t-kT_s) \quad (1)$$

where N is the number of carriers, K is the number of transmit signals, $a_{n,k}$ is the symbol sequence, T_s is the symbol duration of an OFDM symbol and $T_s=T_u+T_g$, T_u and T_g are the "useful time" and the "guard time", respectively, as shown in fig (1). The shaping filter g(t) is assumed to be equal to 1 for $0 \leq t < Ts$. The received signal is as follows:

$$r(t)=y(t)+w(t) \quad (2)$$

$$r(t)= \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \frac{a_{n,k}}{\sqrt{NK}} e^{j2\pi \frac{n(t-T_s)}{NT_s}} g(t-kT_s) + w(t) \quad (3)$$

w(t) is the Additive White Gaussian Noise with mean zero and any variance σ^2

In the wireless multipath fading channel, the received signal is given by the equation

$$r(t)=\sum_{l=0}^{L-1} h_l(t) y(t-\tau_l) + w(t) \quad (4)$$

where $h_l(t)$ is the path gain of the multipath "l" and it is assumed that individual path gains are uncorrelated, τ_l is the path delays, L the total number of paths and w(t) the additive white Gaussian noise.

For OFDM based signal this equation is represented as follows:

$$r(t)= \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \frac{a_{n,k}}{\sqrt{NK}} e^{j2\pi \frac{n(t-T_s-\tau_l)}{NT_s}} h_l g(t-kT_s-\tau_l) + w(t) \quad (5)$$

Similarly, SC-FDMA converts serial data into parallel blocks and convert it into frequency domain using discrete fourier transform (DFT). These samples after subcarrier mapping are converted to time domain symbols using IDFT. The SC-FDMA received signal is given by:

$$y(t)= e^{j w_c t} \sum_{m=o}^{M-1} \frac{x_m}{\sqrt{M}} g(t-kT_s) \quad (6)$$

$$r(t)= e^{j w_c t} \sum_{m=o}^{M-1} \frac{x_m}{\sqrt{M}} g(t-kT_s)+w(t) \quad (7)$$

w_c is the carrier frequency, x_m is the transmitted symbol, T_s is the symbol time. g(t) is same as described above.

In a multipath fading environment (without Doppler effect), this equation is given by

$$r(t)= \sum_{l=0}^{L-1} \sum_{m=o}^{M-1} \frac{x_m}{\sqrt{M}} e^{j w_c (t-\tau_l)} h_l(t) g(t-kT_s-\tau_l)+w(t) \quad (8)$$

4 Cyclostationarity and Correlation

Let us consider the formation of OFDMA and SC-FDMA signals and the results of their correlation, conceptually. Suppose, user 1 says to user 2 "How Are You Feeling". In real time system, it will be converted from analog to digital and then the process of OFDMA or SC-FDMA symbol creation will be followed. But for the sake of simplicity, let us consider each letter of the statement as a single bit being input to the system. The phrase will be converted from serial to parallel, the output will be mapped to subcarriers and Inverse Fourier Transform will be taken. Now let us emphasize on the addition of cyclic prefix to this symbol. Assuming four data subcarriers (neglecting pilot and null carriers as an assumption), the first four alphabets "howa" will form a symbol and the last letter will be copied to the front, making CP (cyclic prefix) to useful symbol (UT) ratio ¼. The sequence of symbols shown in fig. (1) will be obtained making the signal second order cyclostationary for which the condition is:

$$R_x(t+T, \lambda) = R_x(t, \lambda) \quad (9)$$

where,

$$R_x(t,\lambda)=E[r(t).r^*(t+\lambda)] \quad (10)$$

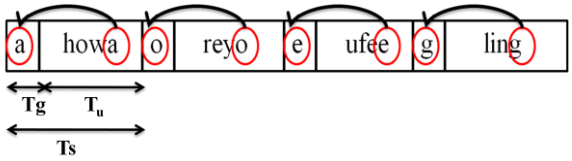


Fig (1): Addition of CP to make the signal cyclostationary

It will then pass through the channel, and will have fading effects and the white Gaussian noise will be added as well. The received signal will be autocorrelated with different lags and we will get a peak each time cyclic prefix matches the last part of the symbol as shown in figure (2). The difference between these peaks is equal to the symbol length or symbol duration in terms of samples.

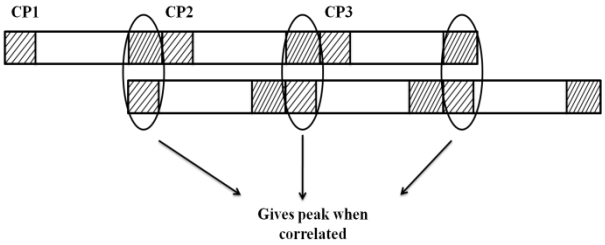


Fig (2): Autocorrelation of the received signal

This paper proposes a novel dynamic method to estimate the cyclic prefix using cross correlation and hypothesis testing. The dynamic algorithm takes one value from the set of predefined values for cyclic prefix according to the standards of WiMAX and LTE and performs cross correlation as shown in the figure (3). If the result is greater than the threshold that depends on SNR and the correlation Length, the value being tested is chosen as the Cyclic Prefix.

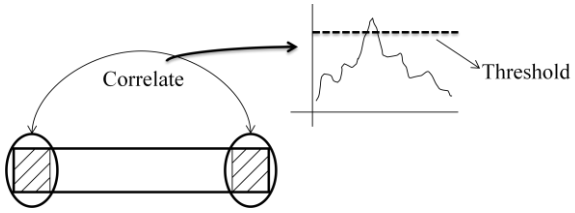


Fig (3): Cross correlation of the cyclic prefix and the last part of symbol

The autocorrelation of the received OFDMA signal in AWGN channel is given by:

$$R_r(m,\lambda)=E[r(m).r^*(m+\lambda)] \quad (11)$$

$$E[r(m).r^*(m+\lambda)]=\begin{cases}\sigma_a^2+\sigma_n^2 & \lambda=0\\ \sigma_a^2 e^{-j2\pi n N_u/NT_s} & \lambda=N_u\\ 0 & otherwise\end{cases} \quad (12)$$

Where σ_a^2 =E[r^2(m)] and σ_n^2 =E[w^2(m)] i.e. variances of signal and AWGN respectively. N_u is the length of useful symbol and r(m) is the m^{th} data.

Similarly, for SC-FDMA the result for equation (4) is :

$$E[r(m).r^*(m+\lambda)]=\begin{cases}\sigma_m^2+\sigma_n^2 & \lambda=0\\ \sigma_m^2 e^{-j2\pi n N_u} & \lambda=N_u\\ 0 & otherwise\end{cases} \quad (13)$$

In the radio propagation channel, the presence of reflectors and scattering features in the channel creates a constantly varying environment that dissipates the signal energy in amplitude, phase and time. These effects result in multiple paths of the transmitted signals that arrive at the receiver. [11] Rayleigh fading process is characterized by the Gaussian WSS uncorrelated scattering fading model [12], where the fading process is modeled as a complex Gaussian process. By above assumption the autocorrelation of the channel impulse response can be expressed as

$$R(\lambda) = \sigma^2 J_0(2\pi f_d [\lambda]) \quad (14)$$

Where, σ^2 is the rms value of the envelope of the waveform and J_0 is the zeroeth-order Bessel function of the first kind.

The autocorrelation function of the received OFDMA signal via fading channel is:

$$E[r(m).r^*(m+\lambda)]=\begin{cases}\sigma_a^2+\sigma_n^2 & \lambda=0\\ \sigma_a^2 e^{-\frac{j2\pi n N_u}{NT_s}} J_o(2\pi f_d N_u) & \lambda=N_u\\ 0 & otherwise\end{cases} \quad (15)$$

Similarly, the SCFDMA signal is:

$$E[r(m).r^*(m+\lambda)]=\begin{cases}\sigma_m^2+\sigma_n^2 & \lambda=0\\ \sigma_m^2 e^{-j2\pi n N_u} J_o(2\pi f_d N_u) & \lambda=N_u\\ 0 & otherwise\end{cases} \quad (16)$$

5 Parameter Estimation

5.1 Estimation of Symbol Duration

Cyclic prefix, when correlated with the data source, will result in varying peaks corresponding to the degree of match between CP and the data source. The distance between consecutive peaks equals the symbol duration N_s.

$$R_s = E[r(m).r^*(m+j)] \quad j=1,2,3,4, \ldots, L \qquad (17)$$

where L is the correlation length.

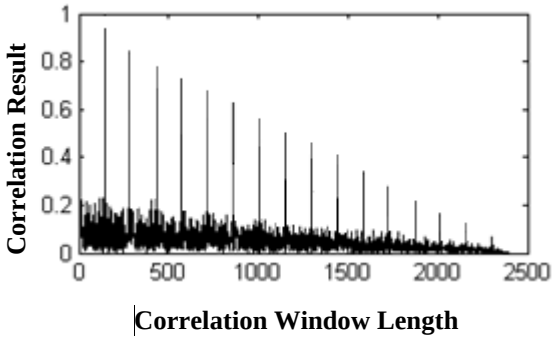


Fig (4a): Correlation results for received OFDMA signal (FFT size=128) for the estimation of its symbol duration when it is perturbed by AWGN and Rayleigh fading

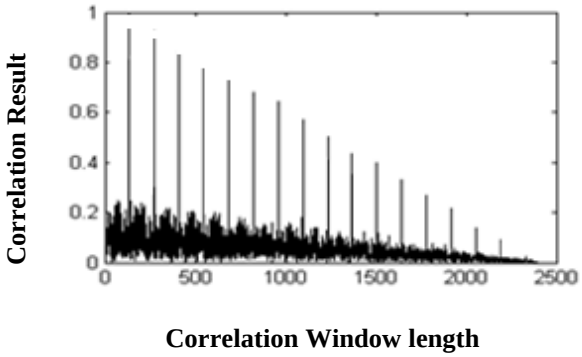


Fig (4b): Correlation results for received SC-FDMA signal (FFT size=128) for the estimation of its "Symbol Duration" when it is perturbed by AWGN and Rayleigh fading

5.2 Estimation of Cyclic Prefix

Once the symbol duration is estimated, we perform correlation test to find out the cyclic prefix. We have a predefined set of values for cyclic prefix according to the standards for WiMAX (.16e) and LTE. We check each of the values of CP by correlating the CP with the last part of the symbol. When the correlation test for these values exceeds the given threshold T_h, the value being tested is selected as the estimated CP for the given OFDMA or SC-FDMA signal.

A value of cyclic prefix is assumed with the help of estimated value of symbol duration, a correlation test is performed between the cyclic prefix and the last part of the symbol. The maximum value of the correlation result is compared with a given threshold, if it is greater than that threshold; the value being tested is the estimated cyclic prefix value. The threshold is crossed if and only if the guard time is similar to the last part of the symbol otherwise the value remains below the threshold. The correlation results for the cyclic prefix are shown in the fig (5a) and (5b).

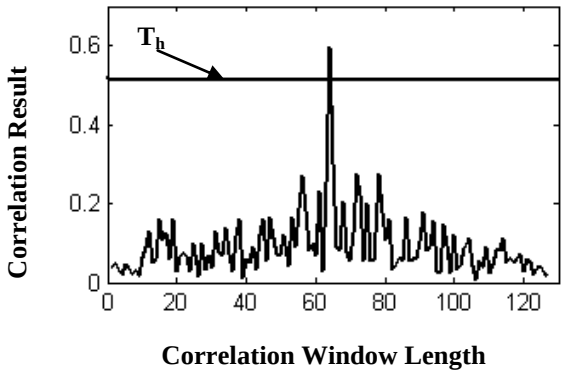


Fig (5a): The cyclic prefix correlation results for OFDMA based signals

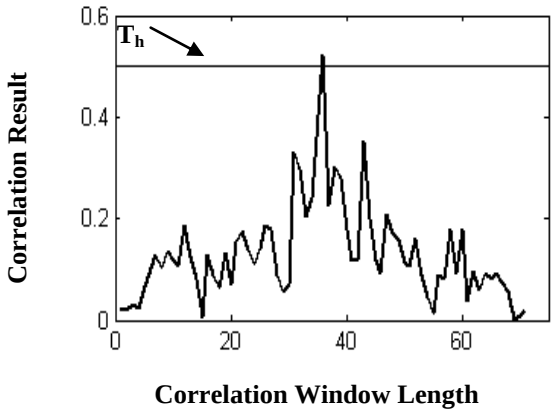


Fig (5b): The cyclic prefix (CP) correlation results for SC-FDMA based signals

Algorithm for "CP" estimation:

Step I: Correlation Test for pseudo predicted values to estimate Cyclic Prefix and useful Time:

Initialize and pick a vector of size equal to cyclic prefix according to the standard draft of WiMAX (.16e) or LTE from the transmitted vector received after passing through Rayleigh fading and Gaussian channel. Let's call it S1.

i. Now initializes another vector of the same size but the samples should be picked from the last portion of the transmitted vector affected by the channel fading and Gaussian noise. Let's call it S2.

ii. Perform normalized cross correlation of S1 and S2 making use of the energy of the signal.

Step II: Testing the condition to estimate cyclic prefix and useful time:

If

The correlation result at zero lag is above threshold Tested value of cyclic prefix and useful symbol is the closest estimate

Otherwise

Test for the other values of cyclic prefix and useful symbol

Mathematically, it can be represented in the form of Null Hypothesis H_0 :

where

H_0=Cyclic prefix length is CP_1

$$R_g = E[g(m).r^* (m+N_s\text{-}cp)] \quad (18)$$

If

$R_g > T_h$ Accept H_0

Otherwise

Discard H_0 and test for another value of CP

In above equation, T_h stands for threshold, g(m) is that part of symbol that falls in guard time and CP is the cyclic prefix.

5.3 Estimation of Useful Symbol Duration

The difference of Symbol Duration and duration of Cyclic Prefix provides the slot duration of the useful data.

6 Simulation Results

The performance evaluation of the proposed algorithm, carried out by means of simulations is discussed in this section. The baseband model of the signals is used to perform simulations. To assess the proposed algorithm, OFDMA signals and SC-FDMA signals were generated with various values of useful time, symbol time and the corresponding cyclic prefix and the proposed correlation techniques have been verified on them.

For OFDMA based WiMAX system, the useful symbol duration takes on the values in the set T_u={128,256,512,1024} and the cyclic prefix value is in the set $T_g=T_u$*{1/2, 1/4, 1/8,1/16,1/32}. Similarly, for SCFDMA based LTE system, useful symbol length is same as for OFDMA while cyclic prefix values are chosen from the set T_g ={9,10,18,20,36,40,72,80}. [13]

The baseband system model for OFDMA and SCFMA is:

Table I: Baseband system model

	OFDMA	**SCFDMA**
Useful Symbol Duration	91.4µs	66.66µs
Cyclic Prefix Duration	11.4µs	5µs
Total symbol Duration	102.8	71.66µs
Sampling Frequency	5.6MHz	7.68MHz
Subcarrier Spacing	10.94KHz	15KHz
FFT size	512	512
Modulation Scheme	64QAM	64QAM
Coding Scheme	Convolution Coding	Convolution Coding

The considered channel model is Rayleigh Fading in the presence of AWGN. Doppler polynomial co-efficients are [0.4 0.3 0.5] and the user is assumed to be moving with a speed of 65 km/hr corresponding to a doppler shift of 150 Hz. The results shown in fig (7a) and (7b) are obtained when the SNR is set 0 dB. Also, displayed in fig (6) are the results when the only perturbation is due to the addition of AWGN with the same signal model. The shape of the correlation result is different for the two signals because of the fact that the SC-FDMA signal is a single carrier signal so the side-lobes of the autocorrelation function are negligible for such signals. However, for OFDMA signal these lobes are quite notable because it is a multicarrier signal and it does not have a fix envelop.

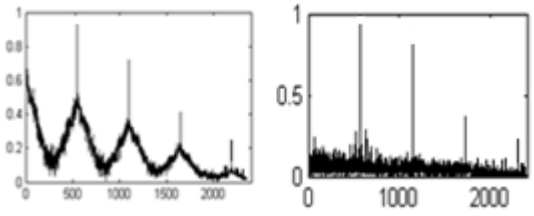


Fig. (6): Correlation results for symbol duration estimation for SC-FDMA (Left) and OFDMA (right) in the presence of Additive White Gaussian Noise

The correlation length is set to be L=2400. Midpoints of the consecutive peaks are picked and symbol duration is calculated using the distance of neighboring midpoints. For symbol duration estimation, the estimation error rates are comparably less because the sample rate is higher. So we estimate symbol duration at high sample rates for a low estimation error rate. The proposed technique then performs correlation test for cyclic prefix length verifying results against each value of the set and finally picking the one that is above the predefined threshold.

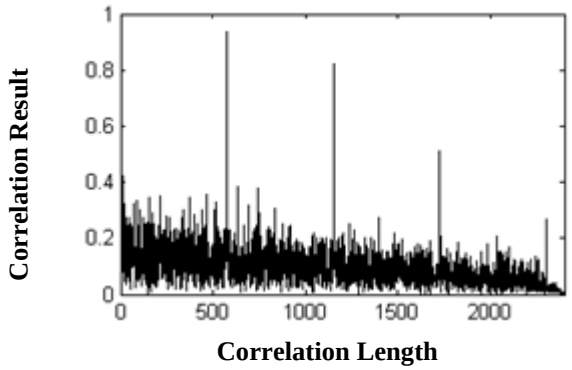


Fig (7a): Correlation results for the OFDMA based system in fading environment

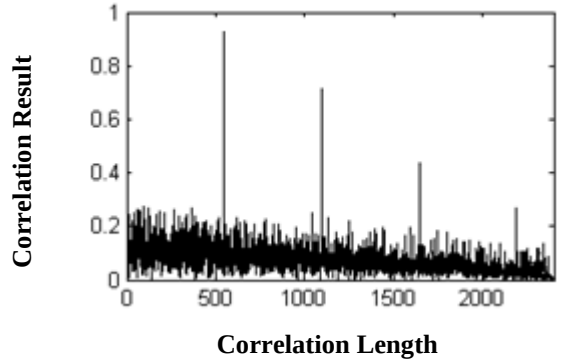


Fig (7b): Correlation results for the SC-FDMA based system in fading environment

7 Conclusions

Through this paper, we have proposed an algorithm for the automatic sensing system for LTE and WiMAX signals to achieve inter-standard reception that remains transparent for the end users. Without using the prior information, we have estimated the symbol duration and cyclic prefix of the multicarrier and single carrier systems through which useful symbol time can be find out. Performance has been assessed on Rayleigh fading channel in the presence of Additive White Gaussian Noise and results show that the algorithm is found robust even in low SNR conditions.

8 References

[1] Xiangqian Fang, Lei Huo and Tiandong Duan, "Cyclic Correlation Based Symbol Duration Estimation for OFDM Signals" in Wireless Communications, Networking and Mobile Computing WiCOM 2006

[2] Tilde Fusco, Luciano Izzo, Angelo Petrella, and Mario Tanda, " Blind symbol timing estimation for OFDM/OQAM system" , IEEE transactions on signal processing, Vol. 57, No. 12, December 2009

[3] Peng Liu, Bing-bing Li, Zhao-yang Lu, Feng-kui Gong , "A Blind Time-parameters Estimation Scheme for OFDM in Multi-path Channel" in Proceeding of Wireless Communications, Networking and Mobile Computing 2005

[4] Hiroyuki Ishii and Gregory W. Wornell, "OFDM Blind Parameter Identification in Cognitive Radios" ,IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications 2005

[5] Miao Shi, Y. Bar-Ness, Wei Su, "Blind OFDM Systems Parameters Estimation for Software Defined Radio" in New Frontiers in Dynamic Spectrum Access Networks DySPAN IEEE 2007

[6] Abdelaziz Bouzegz, Pierre Jallon and Philippe Ciblat, "A second order statistics based algorithm for blind recognition of OFDM based systems" in IEEE Global Telecommunications Conference GLOBECOM 2008

[7] Abdelaziz Bouzegz, Pierre Jallon and Philippe Ciblat, "Matched filter based algorithm for blind recognition of OFDM systems" in IEEE Vehicular Technology Conference VTC 2008

[8] Abdelaziz Bouzegz, Pierre Jallon and Philippe Ciblat, "Maximum Likelihood based methods for OFDM inter carrier spacing characterization" in IEEE Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008

[9] Tevfik Yucek, and Huseyin Arslan, "OFDM signal identification and transmission parameter estimation for cognitive Radio applications" in IEEE Global Telecommunications Conference GLOBECOM 2007

[10] Jarmo Lunden, Saleem A. Kassam and Visa Koivunen, "Non parametric Cyclic Correlation Based Detection for Cognitive Radio Systems" in

IEEE conference on Cognitive Radio Oriented Wireless Networks and Communications CrownCom 2008

[11] Lv Tiejun, Xiao Huibing and Fei Peng, "MMSE estimation of OFDM symbol timing and carrier frequency offset in time varying multipath channels" in IEEE conference on Acoustics, Speech, and Signal Processing ICASSP, 2003

[12] K. E. Baddour and N. C. Beaulieu "Autoregressive modeling for fading channel simulation", IEEE Trans. on Comm., Vol. 4, No. 4, July 2005

[13] Dr.-Ing. Carsten Ball, LTE and WiMax-Technology and Performance Comparison, Nokia Siemens Networks, April 03 2007

Multi-level Infrastructure of Interconnected Testbeds of Large-scale Wireless Sensor Networks (MI^{2}T-WSN)

Adnan M. Abu-Mahfouz[1], Leon P. Steyn[2], Sherrin J. Isaac[1], Gerhard P. Hancke[2]

[1]Advanced Sensor Networks Research Group, CSIR Meraka Institute, Pretoria, South Africa

[2]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

Abstract - *Wireless sensor networks (WSNs) have been used in different types of applications and deployed within various environments. Simulation tools are essential for studying WSNs, especially for exploring large-scale networks. However, WSN testbeds are still required for further testing before the real implementation. In this paper we propose a multi-level infrastructure of interconnected testbeds of large-scale WSNs. This testbed consists of 1000 sensor motes that will be distributed into four different testbeds. The variations of these testbeds will allow for implementing and testing algorithms and protocols that could be used for various applications and within several types of environment.*

Keywords: Testbeds, WSNs, simulation, MI^{2}T-WSN

1 Introduction

The proliferation of wireless communication technologies has enabled the development of wireless sensor networks (WSNs), which consist of a large number of small and cheap sensors with limited resources, such as computing, communication, storage and energy [1]. These sensor nodes are able to sense, measure and collect raw data from the environment, perform simple computations and then transmit only the required and partially processed data to the node responsible for fusion [2]. Sensor nodes may be deployed either manually at fixed locations or randomly into the field. After deployment, these sensor nodes start measuring various properties of the environment, such as light, humidity, temperature, barometric pressure, velocity, acceleration, acoustics and magnetic field, using the different types of sensor that may be attached to these nodes. The measured data will be transferred by a multi-hop infrastructureless architecture to a base station, where data will be manipulated and a decision can be taken.

WSNs have been deployed extensively in areas such as military operations [3], health monitoring [4], natural disaster management [5] and hazardous environments [6]. A large number of research publications have shown that WSNs are a promising approach that could provide future solutions for several problems. Although there are thousands of research publications in this field, there is still a gap between the research and the real implementation. Most of the research that has been done has used simulation tools (e.g. [7, 8]). Using a real WSN during the development phase is not an efficient method; it requires a great deal of effort to reprogram the motes, redeploy them and perform the testing again. Moreover, following the "trial and error" approach is an expensive solution and consumes too much time.

Using a small number of motes for testing purposes could not lead to reliable solutions for large-scale WSNs. On the other hand, using a large-scale WSN itself is not always possible. Running large-scale WSN experiments is difficult; deploying a large number of motes takes a considerable time, and programming each mote individually can quickly cause a significant bottleneck; moreover, the availability of a huge number of motes is a problem.

This paper proposes a multi-level infrastructure of interconnected testbeds of large-scale WSNs (Section 3). This testbed would remove the gap between the research environment and the real implementations, and it would facilitate advanced research in WSN technology. The proposed testbed is designed in such a way as to overcome the drawbacks of some of the current testbeds, which are explained briefly in Section 2.

2 Related Work

Testbeds have successfully been used to evaluate many aspects of wireless sensor networks. These testbeds have differed in many respects, such as hardware components, software architecture, network size, management system and user interface. A recent work by Steyn and Hancke [9] classifies WSN testbeds, on the basis of their features, into several categories: mainly server-based control; single-PC-based control; hybrid; multiple-site; in-band management traffic; and industrial testbeds.

Server-based control testbeds [10] are typically fixed testbed deployments with a central back-end, with various servers supporting the management and data-logging functionality. They can support large-scale experiments. However, they are costly to install and maintain. *Single-PC-*

based control testbeds [11] use a single PC as the central point, which is responsible for the management and data storage. They can easily be adapted into several scenarios, but they have limited features and cannot handle large networks. *Hybrid* testbeds [12] use several gateways in addition to the central server. A significant part of the management and control functions is done locally at the gateways. This architecture enhances the scalability of the testbed.

Multiple-site testbeds [13] are deployed over multiple sites and controlled from a central location. This deployment allows for testing in several types of environments and applications. They usually also have a large scale and can be used to realistically test WSNs for future large deployments. *In-band management traffic* testbeds [14] use additional software installed on the mote and the existing wireless communication channel to perform the management and control functions. This method enhances the scalability and flexibility of the testbed issues and reduces the cost of installation and maintenance. *Industrial* testbeds [15] take into consideration the harsh environment and several challenges that are associated with industrial applications, such as harsh wireless channels, interference and noise.

3 Proposed testbed

3.1 Overview

The proposed testbed could facilitate solutions to a number of challenges that arise in our current environment, such as pollution, health care monitoring, energy saving, pipeline leakage, situational awareness in underground mining and the like. Several solutions and products have been proposed or implemented; however, most of these solutions are complicated and expensive, not functional and difficult to maintain. Moreover, the testing facilities for these solutions are not accessible to all, may not be in real-time and may need to be run by a number of people.

To narrow the gap between the research environment and the real implementations, we need to create a multi-level infrastructure of interconnected testbeds of large-scale WSNs (MI^{2}T-WSN). As shown in Figure 1, MI^{2}T-WSN could be used as a tool to open the door for future solutions. Moreover, MI^{2}T-WSN would facilitate advanced research in WSN technology, enable the conducting of large-scale experiments that are not feasible with traditional small-scale testbeds and help other universities and institutes to build similar or smaller testbeds. Its use could improve e-skills capabilities at several levels: e-literacy, e-business skills, and skills of ICT users, ICT practitioners and R&D practitioners.

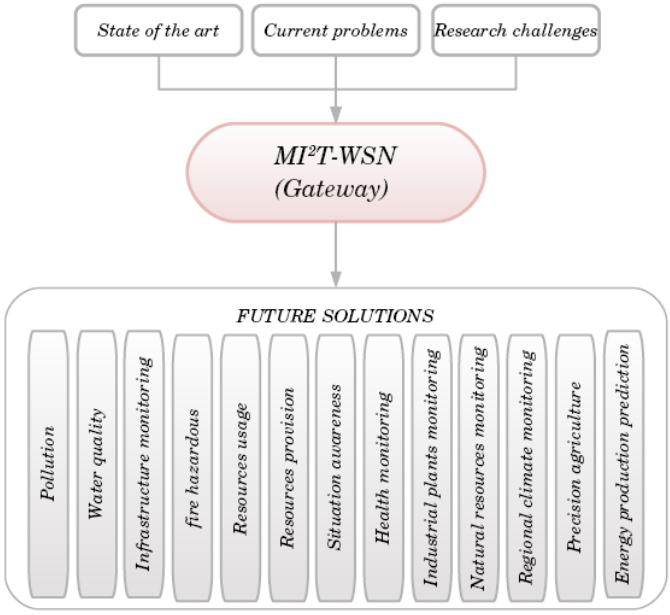


Figure 1. MI^{2}T-WSN could be the gateway for future solutions

3.2 Architecture

MI^{2}T-WSN is a large-scale wireless sensor network laboratory that consists of 1000 motes with heterogeneous sensing devices. Several types of sensor nodes (motes) have been used in the previous WSN testbeds, such as TelosB, SunSPOT, iSense and Waspmote. MI^{2}T-WSN will be built using "Waspmote", which is a new wireless sensor device, recently designed by Libelium [16]. The main characteristics of Waspmote are:

- Remote, real-time monitoring of more than 50 parameters, such as gases (CO, CO_2, CH_4), temperature, liquid level, weight, pressure, humidity, luminosity
- Robust, flexible and long-range communications (up to 40 km)
- Minimum power consumption (years of battery life)
- Easy to use: the language is C++ style
- Open source: open-source API, open-source compiler
- Many extra options: GPS, GPRS, solar panel, Bluetooth, etc.
- Several radio communication protocols: IEEE 802.15.4, ZigBee, RF (900MHz and 868MHz) and Bluetooth
- Certified by CE (Europe), FCC (EEUU), IC (Canada)

As shown in Figure 2, MI^{2}T-WSN consists of four testbeds. The indoor-lab testbed consists of 100 motes that are distributed inside a laboratory; the indoor-real testbed consists of 300 motes that are distributed in the offices, boardrooms and passages of a building; the outdoor-lab

testbed consists of 200 motes distributed above the roof; and the outdoor-real testbed consists of 600 motes distributed outside at multiple sites. Lab testbeds (indoor-lab and outdoor-lab) are distributed within a single area, which makes them easy to install and maintain. Moreover, the motes can be connected to a USB hub to provide them with the required power. Real testbeds (indoor-real and outdoor-real) are distributed in different locations, so require more effort to install and maintain. Power supply is also an issue, and they could require different sources of power such as batteries and solar panels. However, they will provide environments that are very close to those of the real-world applications.

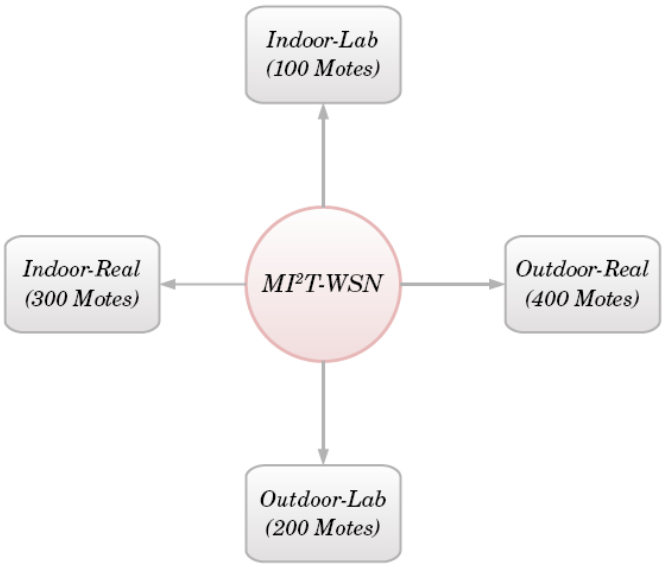


Figure 2. MI^{2}T-WSN Testbeds

3.3 Implementation

MI^{2}T-WSN will be constructed in two phases. In phase one, the indoor-lab and indoor-real testbeds will be constructed. A simple web-based interface will be created to allow researchers, developers and students within our research institute and collaborating universities to access and use these two testbeds. In the second phase, the other two testbeds, outdoor-lab and outdoor-real testbeds, will be constructed. The interface will be enhanced to allow users, developers and administrators to access MI^{2}T-WSN remotely. Different types of manuals will be created, and multi-level training courses will be conducted. In this section, the focus will be on the first phase of implementation.

A Waspmote is a sensor device specially oriented to developers. Because of the commercial availability and the modular design of these devices, we have decided to use them as the basis of our testbeds. In total, 400 Waspmotes will be deployed inside the building. Each device has modular antennas that can be interchanged to support different wireless communication protocols (ZigBee, Bluetooth, GPRS) and frequencies (2.4GHz, 868MHz, 900MHz). Although they have limited on-board sensors, they can be expanded with a variety of modular sensor boards.

To communicate with the Waspmotes we will use several Meshlium gateways, also developed by Libelium. These gateways serve as an in-band management interface to facilitate OTA programming of the Waspmote devices. They will also be used to receive the sensor data sent by the Waspmotes using the ZigBee radio. The Meshlium gateway is a Linux-based router that supports multiple programming languages and can also be customised with up to five different modular radio interfaces: Wifi 2.4GHz; Wifi 5GHz; GPRS; Bluetooth; and ZigBee. As shown in Figure 3, the received sensor data can either be stored in the Meshlium file system, its own local data base or in an external data base (MySQL). The Meshlium gateway also provides a connection to the internet so that users can connect remotely to the testbed. Depending on where the gateway is deployed it can send the information to the internet using any one of the Ethernet, Wifi, or GPRS connections.

In the indoor-lab testbed, the sensor motes will be distributed in a grid network as shown in Figure 3. For constant power to the motes, two professional, testbed-grade 49-port USB hubs will be used to connect these motes to the power supply. In the indoor-real testbed, the motes will be distributed inside the building in different locations such as offices, boardrooms and passages. Each room will typically contain a small cluster of between two and four nodes. To practically power these devices without using too much of the available power outlets in the building, smaller AC powered USB hubs can be used. Each cluster can be connected to a hub and distributed in the room using active USB cables. All the sensor data will be sent wirelessly to the Meshlium gateways to store it locally or in an external data base.

Building these two testbeds requires the following equipment: Waspmote (802.15.4 SMA 2 dbi and 5 dbi), Meshlium gateways, sensor boards (e.g. event, smart city and prototyping), sensors (e.g. temperature, force and pressure, LDR, PIR, humidity and noise sensors), USB-220 V adapters, USB hubs, Waspmote expansion boards, Waspmote antenna modules (GPS, Xbee 900 MHz and GSM/GPRS module), extra memory (32 Gb storage Meshlium and microSD 2Gb card) and extra components required for the back-end network such as routers, switches, and servers (database and application)

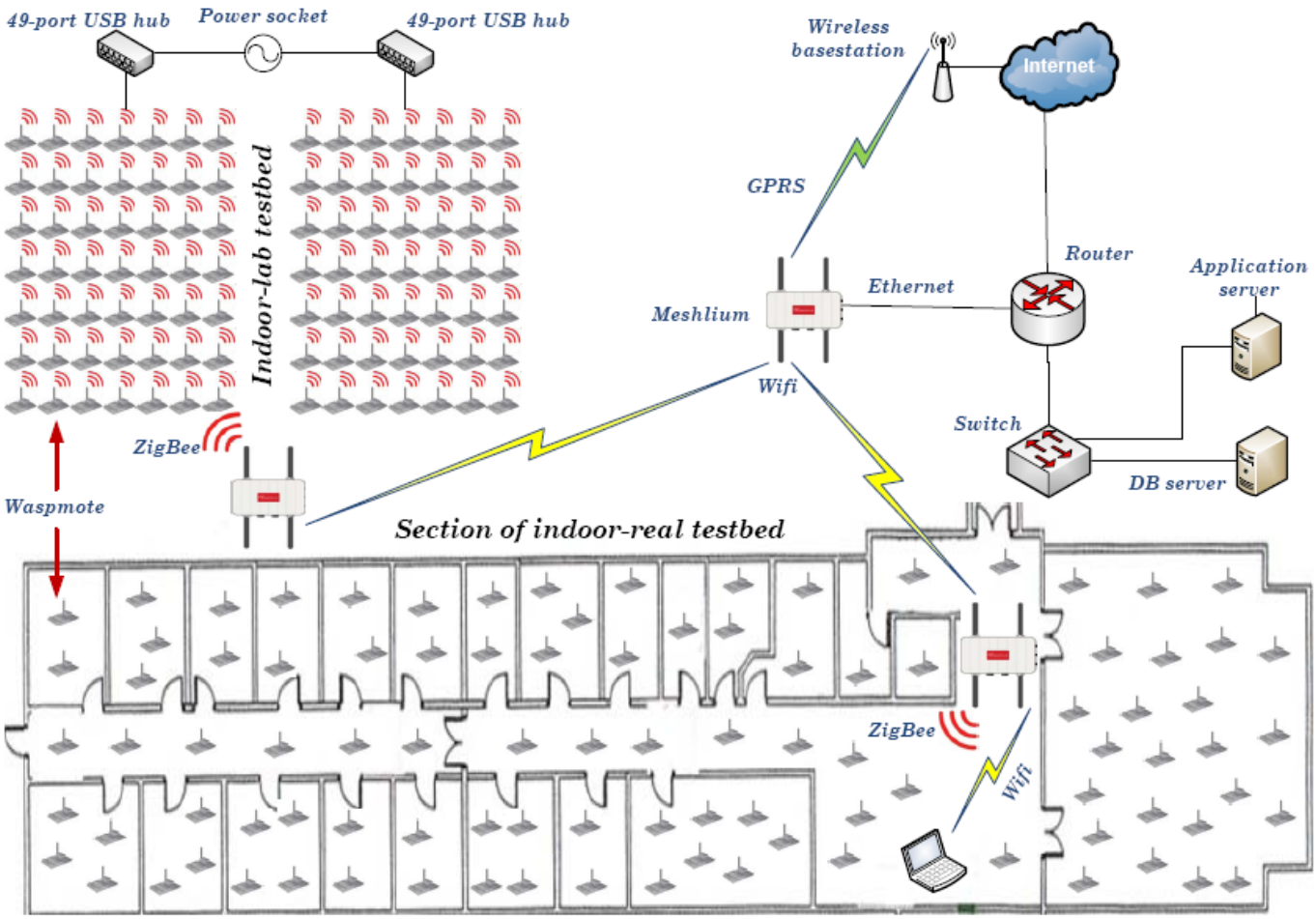


Figure 3. MI^2-WSN network architecture

3.4 Accessibility

One of the main objectives of MI^{2}T-WSN is to develop a real-time system that operates and analyses the environment continuously, and able to detect and report abnormal variations rapidly. MI^{2}T-WSN should identify end-users' needs and requirements as well as operational constraints. In order to achieve these objectives, MI^{2}T-WSN will use a web-based interface that can be accessed remotely, as shown in Figure 4. The interface provides a GUI to define and issue queries in the network and view their results, and can be customised by the end-user. The interface is able to detect and report abnormal variations and allows the end-user to report bugs and problems. The user will be able to book a specific testbed (e.g. indoor-lab) or even a certain number of motes within a specific testbed to use for a period of time.

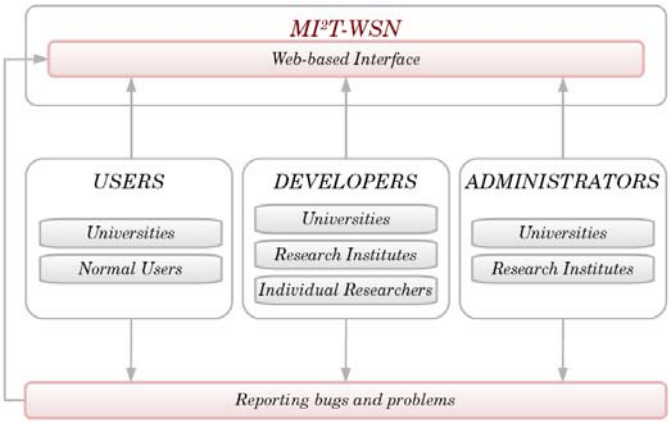


Figure 4. Remote access through a web-based interface

Three categories of users will have access to MI^{2}T-WSN: users, developers and administrators. The "user" represents the normal users who are interested in running simple experiments that consist of simple networks to test different types of applications. The "developer" represents

the developers and researchers who would like to implement and test state-of-the-art algorithms and protocols and also those who would like to develop new algorithms or protocols. The administration right is only given to those universities and research institutes that are actually participating in building these interconnected testbeds. The administrators will be able to control MI^{2}T-WSN remotely and they will be responsible for fixing any reported problems.

In order to ensure a strong end-user involvement and maximise long-term project impact, comprehensive manuals will be prepared for the users, developers and administrators. A library of open-source code and projects will be accessible through the web-based interface. Multi-level training courses will be provided for students and researchers in addition to the on-line training courses.

3.5 Research direction

Figure 5 proposes a research direction for implementing, testing or developing state-of-the-art or new protocols and algorithms. This framework encourages researchers to perform three steps before the start of real-world implementations:

- **Simulation tools:** Simulation is essential for studying WSNs, especially testing new algorithms, applications and protocols. The scalability of protocols should be validated by simulation to ensure that they will perform well even in very large networks. However, relying on simulation alone could be insufficient, and more tests should be done using testbeds.
- **Lab-testbeds:** Compared with real-testbeds, lab-testbeds provide more flexibility in terms of fast programming and testing. Therefore, it is recommended to start with lab-testbeds, especially for new algorithms and protocols.
- **Real testbeds:** The real-testbeds allow the researcher to test the new algorithms or protocols in an environment that is close to the real-world environment. The researcher will have to modify and enhance his or her algorithm or protocol to be more suitable for real-world applications.

There is a flexibility in this framework in terms of the path that the researcher can follow to reach the real-world implementation. This path depends on the type of network and applications that will be used in the real-world implementation. For example, if the researcher is planning to implement a new algorithm that will be used only in the outdoor environment, then the best path would be simulation, outdoor-lab testbed, outdoor-real testbed and finally real-world implementation. However, a researcher who is designing a general protocol that is supposed to be used in different types of environment could start with simulation and then test the new protocol using all four types of testbed.

MI^{2}T-WSN can be considered as a large-scale WSN. However, a researcher who is planning to use the implemented algorithm or protocol in a very large network that consists of tens of thousands of motes will still need to use the simulation tools again to validate the scalability issues of this algorithm or protocol before the real-world implementation.

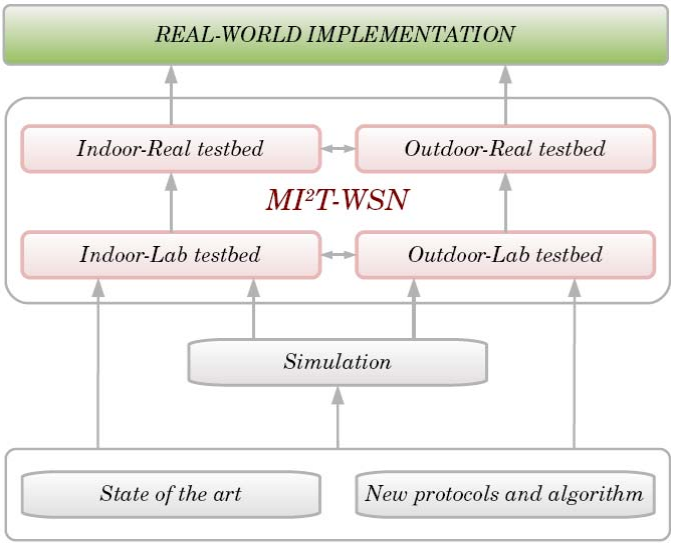


Figure 5. Research direction

4 Conclusion

In this paper, we have proposed a multi-level infrastructure of interconnected testbeds of large-scale WSNs (MI^{2}T-WSN). MI^{2}T-WSN is designed in such a way as to overcome the drawbacks of existing WSN testbeds. Moreover, it differs from them in several aspects that make it unique as a laboratory. MI^{2}T-WSN is a large-scale WSN that consists of a multi-level infrastructure of indoor and outdoor testbeds. MI^{2}T-WSN is a multi-disciplinary testbed that can be used for several applications and environments. MI^{2}T-WSN can be accessed remotely using a web-based interface. Moreover, it will help other universities and research institutes to build similar or smaller testbeds.

5 References

[1] J. Yick, B. Mukherjee and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292-2330. April 2008.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114. February 2002.

[3] M. Hussain, P. Khan and K. Sup, "WSN research activities for military application," in *Proceedings of the 11th International Conference on Advanced Communication*

Technology — ICACT '09, February 15-18, Phoenix Park, Korea, vol. 1, 2009, pp. 271-274.

[4] S. H. Toh, K. H. Do, W. Y. Chung and S. C. Lee, "Health decision support for biomedical signals monitoring system over a WSN," in *Proceedings of the 2nd International Symposium on Electronic Commerce and Security — ISECS '09*, May 22-24, Nanchang City, China, vol. 1, 2009, pp. 605-608.

[5] N. Wirawan, S. Rachman, I. Pratomo and N. Mita, "Design of low cost wireless sensor networks-based environmental monitoring system for developing country," in *Proceedings of the 14th IEEE Asia-Pacific Conference on Communications — APCC '08*, October 14-16, Tokyo, Japan, 2008, pp. 1-5.

[6] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees and M. Welsh, "Deploying a wireless sensor network on an active volcano," *IEEE Internet Computing*, vol. 10, no. 2, pp. 18-25. 2006.

[7] A. M. Abu-Mahfouz and G. P. Hancke, "Ns-2 extension to simulate localization system in wireless sensor networks," in *Proceedings of the IEEE AFRICON 2011 Conference*, 13-15 September, Livingstone, Zambia, 2011, pp. 1-7.

[8] A. M. Abu-Mahfouz and G. P. Hancke, "An efficient distributed localization algorithm for wireless sensor networks: Based on smart reference-selection method," *Submitted for Publication*, 2011.

[9] L. P. Steyn and G. P. Hancke, "A survey of wireless sensor network testbeds," in *Proceedings of the IEEE AFRICON*, September 13-15, Livingstone, Zambia, 2011, pp. 1-6.

[10] M. Sridharan, S. Bapat, R. Ramnath and A. Arora, "Implementing an autonomic architecture for fault-tolerance in a wireless sensor network testbed for at-scale experimentation," in *Proceedings of the 23rd ACM Symposium on Applied Computing*, March 16-20, Fortaleza, Ceará, Brazil, 2008, pp. 1670-1676.

[11] O. Rensfelt, F. Hermans, P. Gunningberg, L. Å. Larzon and E. Björnemo, "Repeatable experiments with mobile nodes in a relocatable wsn testbed," *The Computer Journal*, vol. 54, no. 12, pp. 1973-1986. 2011.

[12] J. P. Sheu, C. C. Chang and W. S. Yang, "A distributed wireless sensor network testbed with energy consumption estimation," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 6, no. 2, pp. 63-74. 2010.

[13] T. Ducrocq, J. Vandaële, N. Mitton and D. Simplot-Ryl, "Large scale geolocalization and routing experimentation with the SensLAB testbed," in *Proceedings of the 7th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, November 08-12, San Francisco, USA, 2010, pp. 751-753.

[14] T. Dimitriou, J. Kolokouris and N. Zarokostas, "Sensenet: A wireless sensor network testbed," in *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, October 22-26, Crete Island, Chania, 2007, pp. 143-150.

[15] J. H. Taylor and J. Slipp, "An integrated testbed for advanced wireless networked control systems technology," in *Proceedings of the 36th Annual Conference on IEEE Industrial Electronics Society (IECON)*, November 07-10, Glendale, AZ, USA, 2010, pp. 2101-2106.

[16] Libelium, http://www.libelium.com/.

Ad hoc Networks Routing : Shortest Path is Enough

Maher Heal [1] **and Marwan Fayed** [2]
[1]Department of Computing Science and Math, University of Stirling, Stirling, UK,
Email: maher.heal@cs.stir.ac.uk
[2]Department of Computing Science and Math, University of Stirling, Stirling, UK,
Email: mmf@cs.stir.ac.uk

Abstract— *It is well-known in hardwired networks that shortest path routing is optimum in regard to optimizing network performance such as maximizing network throughput for instance, but little is known in this regard in ad hoc networks routing due to the dynamic, changing topology of these networks. Via linear programming formulation of the optimum routing problem, we show shortest path is the best routing strategy as well when it comes to maximizing the mean network throughput in ad hoc networks provided that interference is neglected. However, the routing metrics in selecting routes (paths) should be dependent on links availability probabilities in these networks. Heuristic approaches were used in suggesting such metrics when the links capacities are equal and unequal.*

Keywords: Ad hoc networks, shortest path routing, linear programming, routing metric, mobility

1. Introduction

Optimum routing has been studied in the context of hardwired networks. In a hardwired network where the topology is fixed and rarely changing, optimum routing can be formulated as a linear programming problem to optimize a certain objective function. It has been shown that a routing algorithm that selects the shortest path is usually the best algorithm in optimizing many objective functions such as maximizing network throughput [1].

In Ad hoc network where the topology is changing and maybe is changing randomly, little is known whether the best routing algorithm should be shortest path algorithm or not. By generalizing the linear programming formulation in [1] to model the changing topology of the ad hoc network, neglecting interference, and by assuming links between nodes are available with a constant probability, we show the best routing algorithm to maximize the mean network throughput is shortest path as well. However, the routing metric must be dependent on these links availability probabilities. Afterwards heuristic reasoning is applied to suggest ways to calculate such metrics.

2. Related Work

Many routing protocols were suggested in the literature with different routing metrics that capture different working aspects of the ad hoc environment such as link quality, interference, mobility, and energy constraints [2]. However shortest path minimum hop count is the default metric in many popular ad hoc routing protocols, such as OLSR [3], DSR [4], AODV [5] and DSDV [6]. The first experimental work that doubted shortest path minimum hop count as the right metric to achieve high throughput is that of De Couto et al. [7]. By experimenting with two testbeds of static wireless networks running DSDV protocol, they showed the protocol may select low link quality minimum-hop counted path which leads to low throughput due to retransmissions. Although in this paper we propose better metrics for wireless networks by assuming links available with constant probabilities due to changing topology, but links availability with a certain probability could be due to link qualities. In another paper De Couto et al. [8] proposed ETX routing metric based on active probing measurements and many metrics were derived and based on that metric. However, though metrics based on active measurements with probe packets are better than minimum hop count metric in static networks, the reverse is true with mobile networks [2]. Researchers suggested mobility aware metrics, for example Mcdonald and Zanti [9] suggested a routing metric that selects more stable paths based on the availability of network paths that are subject to link failures caused by node mobility. Kamal Jain et al. [10] in their seminal paper proved that shortest path is not the optimum throughput routing strategy in multi-hop wireless networks due to interference which was modeled using conflict graphs. Moreover they proved that the optimal throughput problem is NP-hard. However their analysis was for static or infrequent changed topology networks.

In this paper however, we model topology change but we neglect interference which is a limitation of more model to be incorporated in future research and extensions. However, the contribution of this paper is in suggesting and devising the mathematical means to derive better routing metrics for mobile ad hoc networks and proving shortest path is optimum in maximizing throughput whenever the interference is low and can be neglected.

3. Background

In this section we summarize the problem of optimum routing and its linear programming formulation in hardwired networks. Details can be obtained from [1].

Let us say we have a fixed topology network where the links capacities are already selected, the traffic demand is known ahead of time, the traffic is elastic[1] and we can arbitrarily split demands on different paths. Let us say also as our objective function we want to maximize the smallest spare capacity (difference between link capacity and carried load) of all links. This objective function is reasonable because by that we guarantee there will be enough capacity in the network for more traffic and hence better throughput and less delay (see the proposition in the appendix). This problem is formulated as a linear programming problem as given below.

Let the network be represented as a directed graph $\mathbb{G}(\mathbb{N}, \mathbb{L})$ where $\mathbb{N}$ is the set of nodes (routers) and $\mathbb{L}$ is the set of directed links. This means that if a and b are two nodes in $\mathbb{N}$, then the link $a \rightarrow b$ and the link $b \rightarrow a$ are distinct. Thus, any link $l \in \mathbb{L}$ has a head node and a tail node, and as the names indicate, the link is directed from the head to the tail. In ad hoc networks the case of asymmetric links is quiet common, though some routing protocols where designed to work with symmetric links only. There are K demands that are to be routed on the network. Each demand is associated with an ordered pair of nodes (n_1,n_2), where n_1,$n_2 \in \mathbb{N}$. Note n_1 is the source of the demand, and n_2 is the destination. We number the quantities as follows: demands are numbered $1, 2, ..., k, ..., K$, nodes $1, 2, ..., i, ..., N$, so that $|\mathbb{N}| = N$, and links are numbered $1, 2, ..., l, ..., L$, so that $|\mathbb{L}| = L$.

Now denote the demands by $d(k)$, $1 \leqslant k \leqslant K$, and define a flow vector $X(k)$ of dimension $L \times 1$ corresponding to the kth demand, with $X(k)_l$ represents the amount of the kth demand carried on link l, where $1 \leqslant k \leqslant K$. The topology of the network is summarized using node-link incidence matrix A of dimension $N \times L$ where

$$A_{i,l} = \begin{cases} +1 & \text{if } i \text{ is the head of link } l \\ -1 & \text{if } i \text{ is the tail of link } l \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

And define the demand vector $V(k)$, $1 \leqslant k \leqslant K$, of dimension $N \times 1$ as

$$V(k)_i = \begin{cases} d(k) & \text{if } i \text{ is the source of demand } k \\ -d(k) & \text{if } i \text{ is the destination of demand } k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The optimization problem is given by

$$\max\ z$$

$$\begin{bmatrix} A & 0 & 0 & \dots & 0 \\ 0 & A & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & A \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(K) \end{bmatrix} = \begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(K) \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} I & I & \dots & I \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(K) \end{bmatrix} + z\mathbf{1} \leqslant C \quad (4)$$

$$X(k) \geq 0, 1 \leq k \leq K, z \geq 0 \quad (5)$$

where z is the minimum spare capacity. In equation (3), the left matrix is of dimension $KN \times KL$ and there are K block elements in each row and K block elements in each column; A is the node-link incidence matrix, of dimension $N \times L$. 0 is also a matrix of dimension $N \times L$. The middle matrix in equation (3) is of dimension $KL \times 1$ and the right matrix is of dimension $KN \times 1$.

In equation (4), the left matrix is of dimension $L \times KL$ and there are K block elements in it, where I is the $L \times L$ identity matrix. $\mathbf{1}$ is a column vector of L elements, all of which are 1. C is a vector of L elements where the l^{th} element is the capacity of link l.

A solution to the above linear programming problem exists when routing is done by selecting shortest paths where the link weights are the optimal dual variables of the dual problem. Coarse approximation is used in practice like taking the weights as the inverse of link capacities as in Cisco implementation of OSPF [11].

4. Generalization and Metrics

4.1 Generalization

The problem in ad hoc networks is exactly the same as the problem of hardwired networks, when mobility is the only factor of concern. The only difference is that the node-link incidence matrix A that defines the network topology is not static but dynamic with 1's and 0's entries changing according to the time. The entries (1's and 0's) could be assigned randomly depending on the mobility model used.

The optimum routing solution of this problem is unknown and it could be shortest path or not. We will show the optimization problem in ad hoc networks is the same as in hardwired networks when the mean throughput is maximized and accordingly shortest path is optimum.

Again the network is represented by a directed graph of fixed nodes and changing links from one time epoch to another. Let $\mathbb{N}$ be the set of nodes and $\mathbb{L}$ be the union of links sets at the different time epoches of the network life time. Thus between any two nodes a and b a link may be available for certain time epoch and not available for another. We will assume the link is available between any two nodes with a constant probability p. This probability is 0 for links that are available for a short period of time only in the network

[1]Elastic traffic has no intrinsic transfer rates or end-to-end delay requirements. It is generally the traffic generated by TCP sessions like browsing.

life time and then they disappear, although they $\in \mathbb{L}$. They have no effect on our analysis as we are concerned with the mean throughput. Assuming links are available with a constant probability is only an approximation, depending on the mobility model.

The node-link incidence matrix is a function of time now and given by

$$A_{i,l}(t) = \begin{cases} +1 & \text{if } i \text{ is the head of link } l \text{ at time } t \\ -1 & \text{if } i \text{ is the tail of link } l \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$A(t)$ is of dimension $N \times L$. The flow vector for demand k is also a function of time $X(k)(t)$; we have

$$A_i(t).X(k)(t) = \begin{cases} d(k) & \text{if } i \text{ is the source of demand } k \text{ at time } t \\ -d(k) & \text{if } i \text{ is the destination of demand } k \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$

Note that each component of $X(k)(t)$ for each link, i.e. $X(k)(t)_1$, $X(k)(t)_2$,..., $X(k)(t)_L$, may be zero when the link is not available between the nodes. Now

$$A(t).X(k)(t) = V(k)$$

where $V(k)$ is the usual demand vector in equation (2). By considering the flows of all demands, we have

$$\begin{bmatrix} A(t) & 0 & 0 & \dots & 0 \\ 0 & A(t) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & A(t) \end{bmatrix} \begin{bmatrix} X(1)(t) \\ X(2)(t) \\ \vdots \\ X(K)(t) \end{bmatrix} = \begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(K) \end{bmatrix} \quad (7)$$

$A(t)$ is the node-link incidence matrix in equation (6).

Let $C(t)$ be the capacity vector of the links. It is a vector of dimension $L \times 1$ and is a function of time. Note that $C_i(t)$ may assume the value zero when the link i is not available in certain time epoches or a constant capacity C_i, when the link is available. Then

$$X(1)(t) + X(2)(t) + \cdots + X(K)(t) \leqslant C(t)$$

It is a vector inequality, where both sides are vectors of dimension $L \times 1$.
The spare capacity vector is given by

$$Z(t) = C(t) - (X(1)(t) + X(2)(t) + \cdots + X(K)(t))$$

So far the analysis done is the same as that in [1], but time is added as a parameter to reflect the dynamic changing topology of ad hoc networks. Since we are interested in optimizing the mean performance of the network (mean throughput), we will try finding solutions that maximize the mean of the spare capacities of links. Taking the mean of both sides of the above equation, the mean spare capacity is given by

$$E(Z(t)) = E(C(t)) - (E(X(1)(t)) + E(X(2)(t)) + \cdots + E(X(K)(t))) \quad (8)$$

Note

$$E(C(t)) = \begin{bmatrix} p_1 C_1 \\ p_2 C_2 \\ \vdots \\ p_L C_L \end{bmatrix} \quad (9)$$

where p_i is the probability that link i is available.

In hardwired networks, we want to maximize the spare capacity of links and thus we maximize the minimum spare capacity since by that we guarantee all links will have a spare capacity more than that minimum which is maximized. In ad hoc networks we want to maximize the mean spare capacity of all links and hence we will maximize the minimum mean spare capacity[2] given by equation (8).

Let $z = min_{l \in \mathbb{L}} E(Z(t))$, then we have

$$E(X(1)(t)) + E(X(2)(t)) + \cdots + E(X(K)(t)) \leqslant E(C(t)) - z\mathbf{1}$$

where $\mathbf{1}$ is a column vector of L elements, all of which are 1. In matrix form

$$\begin{bmatrix} I & I & \dots & I \end{bmatrix} \begin{bmatrix} E(X(1)(t)) \\ E(X(2)(t)) \\ \vdots \\ E(X(K)(t)) \end{bmatrix} + z\mathbf{1} \leqslant \begin{bmatrix} p_1 C_1 \\ p_2 C_2 \\ \vdots \\ p_L C_L \end{bmatrix} \quad (10)$$

I is the $L \times L$ identity matrix and there are K blocks in the left matrix of the above equation and thus it is of $L \times KL$ dimension. By taking the mean of both sides of equation (7), our optimization problem is

$$\max z$$

[2]Although the mean spare capacity of each link will be guaranteed to be more than that minimum, but the minimum mean spare capacity could be due to a link that is available for a short period of time, i.e. with very low probability and hence we could maximize the minimum of mean spare capacity divided by availability probability. In that case I in left matrix of equation (10) must be the identity matrix where its diagonal 1's replaced by $1/p_i$,$1 \leqslant i \leqslant L$. Also the right matrix is replaced by links capacities vector.

$$E\left(\begin{bmatrix} A(t) & 0 & 0 & \dots & 0 \\ 0 & A(t) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & A(t) \end{bmatrix} \begin{bmatrix} X(1)(t) \\ X(2)(t) \\ \vdots \\ X(K)(t) \end{bmatrix}\right) = E\left(\begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(K) \end{bmatrix}\right) \quad (11)$$

$$\begin{bmatrix} I & I & \dots & I \end{bmatrix} \begin{bmatrix} E(X(1)(t)) \\ E(X(2)(t)) \\ \vdots \\ E(X(K)(t)) \end{bmatrix} + z\mathbf{1} \leqslant E(C(t)) \quad (12)$$

$$E(X(k)(t)) \geq 0, 1 \leq k \leq K, z \geq 0 \quad (13)$$

To find the mean of left side of equation (11), consider $A(t).X(k)(t)$; it is clear the ith entry of this multiplication is given by

$$\sum_{j=1}^{L} a_{ij}(t).X(k)(t)_j$$

where

$$a_{ij}(t) = \begin{cases} +1 & \text{if } i \text{ is the head of link } j \text{ at } t \\ -1 & \text{if } i \text{ is the tail of link } j \text{ at } t \\ 0 & \text{otherwise} \end{cases}$$

$a_{ij}(t) = 1$ or -1 with probability p_j and zero with probability $1 - p_j$, and whenever $a_{ij}(t)$ is zero $X(k)(t)_j$ is zero as well. Hence, we have:-

$$E(a_{ij}(t)X(k)(t)_j) = \begin{cases} +E(X(k)(t)_j) & \text{if } i \text{ was the head of link } j \\ -E(X(k)(t)_j) & \text{if } i \text{ was the tail of link } j \\ 0 & \text{otherwise} \end{cases}$$

Hence, equation (11) can be written as

$$\begin{bmatrix} A & 0 & 0 & \dots & 0 \\ 0 & A & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & A \end{bmatrix} \begin{bmatrix} E(X(1)(t)) \\ E(X(2)(t)) \\ \vdots \\ E(X(K)(t)) \end{bmatrix} = \begin{bmatrix} E(V(1) \\ E(V(2)) \\ \vdots \\ E(V(K)) \end{bmatrix} \quad (14)$$

$$A_{i,l} = \begin{cases} +1 & \text{if } i \text{ is the head of link } l \\ -1 & \text{if } i \text{ is the tail of link } l \\ 0 & \text{otherwise} \end{cases}$$

By this we have proved the optimization problem in ad hoc networks is the same as hardwired networks but with with variables $E(X(k)(t))$'s and z which is now the minimum mean spare capacity. Also the link capacities now are multiplied by the links availability probabilities. Since the problem is the same, then shortest path is expected to be the optimum routing in ad hoc networks but the routing metric needs to be estimated based on link availability probabilities.

4.2 Metrics

In our analysis we assumed the probability of node A being within range of node B in ad hoc network is p. Then the link between A and B will be available p of the time and unavailable $1 - p$ of the time, i.e. The channel capacity of the link between A and B is pC on average; where C is the link capacity of the link between A and B when they are within range of each other. This is similar to time division multiplexing between two connected nodes in hardwired networking where the capacity is the proportion of the time allocated for communication between A and B; though it could be better approximated by statistical multiplexing.

By applying shortest path algorithm and using the inverse of capacity as links weights as in Cisco implementation of OSPF [11], the route (path) metric will be

$$\text{Metric} = 1/p_1C_1 + \dots + 1/p_iC_i + \dots + 1/p_MC_k \quad (15)$$

where p_i is the availability probability of the link i of the path (the path has M links; 1 to M).

When all links capacities are equal which is not usually the case in ad hoc networks, then the metric is simply the addition of the inverses of links probabilities.

The case of symmetric links metric can be derived using a different approach. Rather than trying to maximize mean spare capacity to maximize mean throughput, we will try to minimize the delay and hence the throughput is maximized. Assume retransmissions between any nodes is infinite and given the propagation delay is small compared to the transmission delay, the overall end-to-end delay of the packet is directly proportional to the total number of hop transmissions (including retransmissions). Hence the mean delay is

$$1/p_1 + 1/p_2 + \dots + 1/p_M$$

on a path of M links; where p_i is the probability that the link i is available.

That can be easily seen by noting that the packet should pass link by link until reaching the destination at the end of link M and that the probability of passing link i is a geometric random variable with parameter p_i.

By comparing the above metric with the metric in equation (15) for the general case of different links capacities network, we see it is only the special case of that metric when all the links are of equal capacity.

5. Conclusion and Future Work

Shortest path routing has been proved as the optimum strategy (algorithm) in regard to optimizing mean network throughput in ad hoc networks whenever interference is neglected. That was done by showing the optimization problem has the same form as in hardwired counterpart through a generalization of the hardwired networks optimization problem to include the dynamic topology of ad hoc networks and assuming nodes are within range of each other with constant probabilities. However, the routing metrics must be dependent on these probabilities since they are parameters of the optimization problem of ad hoc networks. Heuristics were used to suggest such metrics.

More accurate metrics need to be estimated based on solutions of the optimization problem of ad hoc networks. Simulation studies to validate these metrics or the ones suggested using heuristics in optimizing network performance are required as well, by applying the new metrics to some of the popular routing protocols such as DSDV, DSR, OLSR or AODV. Ways to calculate such metrics by ad hoc hoc networks nodes and ways of realizing them in a distributed routing algorithm is an open problem.

The assumption of nodes are available with constant probabilities within each other range needs to be studied further depending on the mobility models. we modeled a network with fixed nodes and the topology is changing due to links availability only. A more general model is when nodes can join and leave randomly, and thus the topology is changing due to varying number of nodes as well. Finally, as interference is an important factor in deteriorating wireless networks performance, an extension of our model by incorporating interference as an additional constraint in the optimization problem is required.

Appendix I
Proposition: Proof of any Solution of Load (Throughput) Maximization is also a Solution of Spare Capacity Maximization

Let us name a routing strategy (algorithm) that solves our optimization problem, i.e. maximizing the minimum spare capacity γ and other strategies that don't solve our optimization problem ω.

In any network: (i) when γ is used, which means the minimum spare capacity is maximum, then we can inject more load (traffic) in the network, i.e. we can increase the demand (load), if it is not maximum. This means higher throughput. (ii) On the other hand when we have the demand (load) maximum then the routing strategy (algorithm) used must also be a solution to our optimization problem (maximizing the minimum spare capacity) and the maximum minimum spare capacity is zero.

Proof:

The first part of the proposition (i) is clear because spare capacities of all links are at least $z = min_{l \in \mathbb{L}} z_l$, z_l is the spare capacity of link l, $1 \leq l \leq L$ and that minimum is at maximum value when γ strategy (algorithm) is used. Hence we can increase the load of each link by that minimum which means we can increase the load (demand). We prove the 2nd part of the proposition (ii) as follows:

Let the load is maximum whatever the routing strategy used ω or γ, i.e. for the collection of all routing strategies (algorithms), then $min\ z_l$ should be zero for at least one value of l, $1 \leq l \leq L$. Let this is not the case, i.e. z_l is not zero for all values of l. Take now any path from a source node to a destination node. Increase the traffic (flow) on the links of this path by $min_{l \in \text{links of the path}} z_l$ and thus we were able to increase the total load (demand). This is a contradiction because the load is maximum and accordingly $min\ z_l$ should be zero for at least one value of l, $1 \leq l \leq L$. We have proved for any routing strategy γ or ω that at least one spare capacity is zero. Thus $z = min_{l \in \mathbb{L}} z_l = 0$ is independent of the routing strategy. Thus $max\ z = 0$ for the collection of all routing strategies (algorithms) ω or γ. Now when we use γ, we get z maximized and when we use any ω strategy (algorithm) we have z less than its maximum value but this means $z < 0$ which is a contradiction since $z \geq 0$. Hence only strategy γ can be used when the load is maximum.

Q.E.D

References

[1] A. Kumar, D. Manjunath, and J. Kuri, *Communication Networking-An Analytical Approach*. San Francisco, CA: Morgan Kaufann Publishers, 2004, ch. Shortest Path Routing of Elastic Aggregates, pp. 677–711.

[2] G. Parissidis, M. Karaliopoulos, R. Baumann, T. Spyropoulos, and B. Plattner, "Routing metrics for wireless mesh networks," in *Guide to Wireless Mesh Networks*, S. Misra, S. C. Misra, and I. Woungang, Eds. Springer-Verlag, London, 2009, pp. 199–230.

[3] T. Clausen and P. Jacquet, "Optimized link state routing protocol," IETF RFC 3626, October 2003.

[4] D. B. Johnson, D. A. Maltz, and J. Broch, "DSR : The dynamic source routing protocol for multihop wireless ad hoc networks," in *Ad Hoc Networking*, C. E. Perkins, Ed. Addison-Wesley, 2001, pp. 139–172.

[5] C. E. Perkins and E. M. Royer, "The ad hoc on-demand distance-vector protocol," in *Ad Hoc Networking*, C. E. Perkins, Ed. Addison-Wesley, 2001, pp. 173–219.

[6] C. E. Perkins and P. Bhagwat, "DSDV : Routing over a multihop wireless network computers," in *Ad Hoc Networking*, C. E. Perkins, Ed. Addison-Wesley, 2001, pp. 53–74.

[7] D. S. J. De Couto, D. Aguayo, B. A. Chambers, and R. Morris, "Performance of multihop wireless networks: shortest path is not enough," *Computer Communication Review*, vol. 33, no. 1, pp. 83–88, 2003.

[8] D. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. ACM Mobicom*, San Diego, CA, USA, Sept. 2003.

[9] A. Mcdonald and T. Znati, "A path availability model for wireless ad-hoc networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, New Orlean, LA, USA, Sept. 1999.

[10] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. ACM Mobicom*, San Diego, CA, USA, Sept. 2003, pp. 66–80.
[11] W. Odom, *CCNA ICND2-Official Exam Certification Guide*. Indianapolis, IN: Cisco Press, 2007, ch. OSPF, pp. 343–375.

it-RFID: an Ultra-Low Power Active RFID system with flexible Radio Triggered Wake-Up System

A.Sanchez*, Y.Boo†, S. Blanc*, and J.J. Serrano*
* ITACA - Universitat Politècnica de València. 46022 València, Spain
† Beijing University of Posts and Telecommunications (Internship in ITACA)

Abstract— *Radio Frequency IDentification (RFID) systems are becoming a very interesting solution for several applications. Although "passive" RFID systems are widely used because no energy source is needed on the tags, emergent "active" RFID technology (with an on-board battery) opens the door for new applications with challenging requirements. However, energy must be saved in "active" tags to extend its lifetime. Although RFID standards (ISO/IEC 18000-7) define some mechanisms to reduce power consumption using some wake up signal prior to any data exchange, state-of-the-art platforms do not implement this feature and waste lots of power. In this paper a new active it-RFID platform is presented that complies with ISO standard and reduces energy consumed drastically by embedding a new wake up system that combines the advantages of the highest performance systems. it-RFID performance is evaluate and compared with other solutions as well as other solutions based on Wireless Sensor Networks (802.15.4).*

Keywords: RFID, Active RFID, WSN, RT-WUp

1. Introduction and Related Work

Radio Frequency IDentification (RFID) is an automatic identification method using radio frequencies between RFID readers and tags. A reader or interrogator is the device in charge of collecting the tags' information within its RF communication range. A tag is a small transponder that contains certain information collected by the reader for different purposes.

Traditionally, the most widely approach to RFID systems is "passive" RFID technology, in which a "tag" has no power source of its own [1]. Recently, "active" RFID (with an on-board battery in the tag) has become an interesting alternative [2].

Passive tags suffers from extremely short communication range. Both the transmitter and the receiver performance of the in-built transponder are very low. Using a battery, active Radio Frequency (RF) transponders can be used with highest output power and more sensible receivers, thus range is extended. Besides sensors, memory and other elements can be embedded in a tag to flexibly obtain, process, store and transmit information. As a result, active RFID advantages enables a wide range of new applications. Nevertheless, active RFID tag lifetime is limited since energy available is finite. Specifically, wireless link consumes most of the power [3].

Several works have been done on obtaining a RFID protocols with ultra-low power consumption [3], [2]. However in most RFID real deployments, devices from various vendors can be used and they must be able to work together. For this reason, compatibility and interoperability between systems that comply with an international standard is advisable. ISO/IEC 18000 [4] is a series of standards that define the air interface for RFID devices. Although there are several parts, part 7 is the standard for active RFID systems. Although ISO/IEC 18000-7 has been proved to be inefficient [5], this paper is not focused on improving this specification, but designing a new active RFID platform that increases the active RFID application performance, and still complies with ISO 18000-7 standard.

The most remarkable research papers on active RFID platforms are based on micro-controller architectures with some configurable radio interface [5] [6] [7]. However, all these platforms have the same problem: tags are not able to efficiently detect the presence of a reader and to save power when not actually needed. For that purpose standard ISO/IEC 18000-7 contemplates the emission of a wake up signal previous to any data exchange. This signal should activate or wake up (WUp) all the tags within RF reader range. However, available platforms do not implement this feature and tags remain active listening to the channel for long time waiting for reader detection, wasting lots of power.

This wake up issue is tackled by Radio-Triggered Wake Up (RT-WUp) techniques. Using this feature, a wireless system can be activated asynchronously by a specific radio signal and turned into operative mode. State-of-the-art wake up systems [8] [9] are able to activate remote nodes within a range of 15 meters dissipating less than 9 μW when waiting for incoming wake up signals. Both systems are based in super-heterodyne receivers: incoming Radio Frequency (RF) signal is moved to a much lower Intermediate Frequency (IF). The resulting wave can be further processed and decoded with lower power consumption.

In this paper a new active RFID platform is described.

It combines a micro-controller based architecture with the best performance wake up system [9]. This system has been enhanced to extend detection range and to fulfil ISO/IEC 18000-7 specifications: a new flexible IF receiver has been investigated that combines flexible IF wake up architecture [8] with the selected wake up system [9].

The rest of the paper is organized as follows. Standard ISO/IEC 18000-7 is described in Section 2. In Section 3 the new it-RFID platform is described. Section 4 shows performance of this new platform showing the results obtained from different experiments. Finally, Section 5 concludes the paper.

2. Active RFID protocol ISO/IEC 18000-7

The tag collection algorithm defined in ISO/IEC 18000-7 uses a medium access protocol based on the framed slotted ALOHA protocol. Fig. 1 shows the tag collection sequence and timing as defined in ISO/IEC 18000-7 [4].

Reader initiates the communications in a Reader Talks First (RTF) fashion, as described in [2]. First, sends a wake up signal that is able to activate (wake up) all the tags within the RF communication range. The standard describes this wake-up signal as a sub-carrier tone of 30 kHz. Immediately afterwards, the reader sends a collection command and then all tags' data are collected by repeating a number of collection rounds. Some authors label both wake up signal and collection command together as Beacon [2].

The windows size is specified in the collection command and it defines the total time for the reader to listen for potential responses from tags. This value is defined as a multiple of 57.3 ms. Initial window size is set to 57.3 ms and then it is dynamically resized depending on collisions detected on the tag responses.

On the other side, upon receipt of a collection command, tags calculate the slot size (8 ms) and the number of slots in the current collection round using the specified window size and randomly select a slot in which to respond.

2.1 Power consumption issues and it-RFID motivation

After the window has elapsed, the reader sends point-to-pint sleep commands to all tags collected during the collection round. The tags that receive a sleep command move to sleep mode and do not participate in the subsequent collection rounds. Sleep mode is an ultra-low-power state in which the whole tag consumes a few micro-watts. The tags can remain in sleep mode until next wake up signal is detected.

However, as discussed in Section 1, currently there is no active RFID platform that is able to detect wake up with real low-power consumption. Tags listen to the channel periodically instead and, upon Beacon detection, tags response and wait for sleep command to start checking periodically the channel again [2]. This channel checking task dissipates lots of power as evaluated in Section 4.

Therefore it would be desirable to avoid this periodic checking. Moreover, energy will also be saved with an

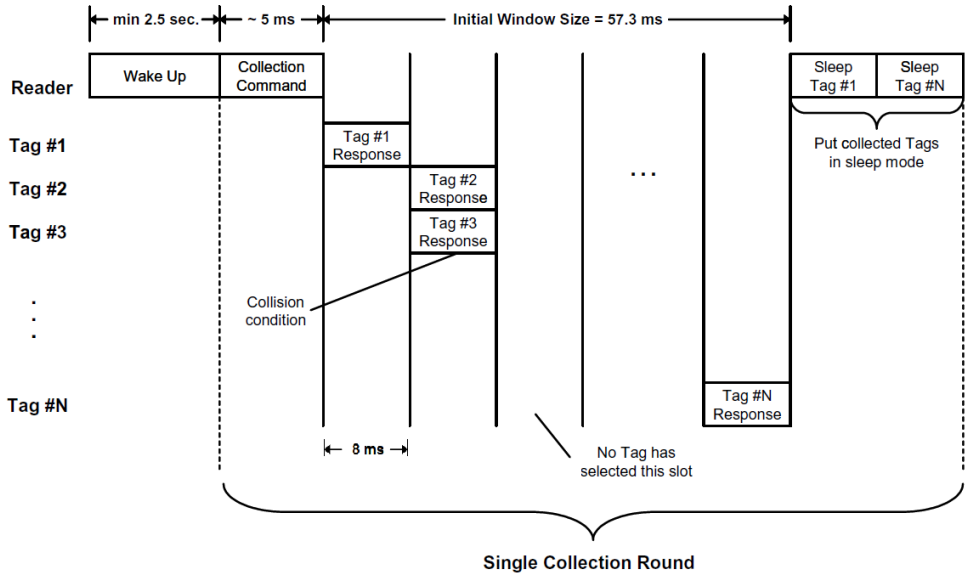


Fig. 1: Active RFID message collection protocol.

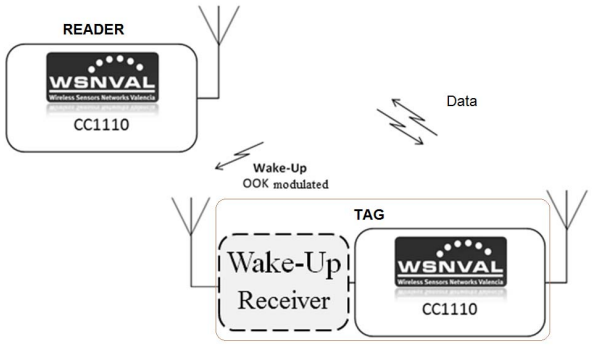


Fig. 2: Proposed Active RFID overall architecture.

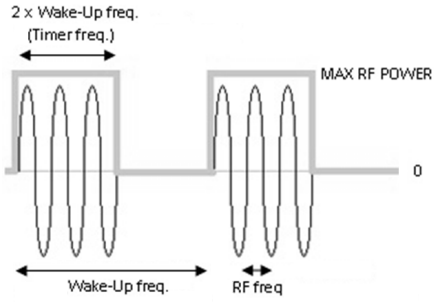


Fig. 3: Generated Wake-Up signal.

active RFID platform that a tag to allows a tag to remain in sleep mode when no RFID is a available.

3. it-RFID platform

Fig. 2 shows the it-RFID system overall architecture. As previously discussed, the main advantage of the solution presented in this paper is the attachment of a RT-WUp system that tackles both transmission and reception of ISO/IEC 18000-7 wake up signals with ultra low power consumption.

On the one hand, RFID reader is only formed by a transponder with reconfigurable radio interface to complain ISO 18000-7. Additionally it incorporates a wired interface to a remote station, but it is out of scope of this paper. On the other hand, RFID tag is composed by the same transponder and a wake up receiver circuit [9]. As explained in [9], reader radio interface is reused to transmit both RFID data and the wake up signal. This fact favours system integration and decreases final system costs [10]. The details of the different parts are explained below.

3.1 Radio Frequency Interface

The presented platform in this work has been based on the radio interface developed by the company Wireless Sensor Networks Valencia (WSNVAL) [11]. Its micro-controller core is CC1110 (Sub-1GHz System-on-Chip with 8051 MCU) from Texas Instruments (TI). It includes CC110X radio interface and can operate in 433 MHz, 868 MHz and 900 MHz bands.

Although this system was originally developed for Wireless Sensor Networks (WSN) deployment, radio parameters such as centre frequency (433.92 MHz), symbol frequency deviation (50 kHz), data-rate (27.7 kbps) and modulation (FSK) can be reconfigured by changing internal registers to fulfil ISO/IEC 18000-7 specifications.

Radio Interface power consumption transmission (TX), reception (RX) and sleep mode represents to whole it-RFID platform power consumption when running these tasks. These energy values are shown in Table 2.

3.2 Radio Triggered Wake-Up system

Radio Triggered Wake Up (RT-WUp) is an asynchronous wake up technique that keeps WSN nodes in a ultra-low power mode until a certain radio signal is detected. Wake up signal, emission and reception are critical and are accurately described and solved in [9].

3.2.1 WUp Signal

Wake up reception needs a specific wake up receiver sub-system that remains active most of the time listening to the channel. Since active RFID tags energy resources are limited, the hardware should dissipate as low power as possible to detect, decode and process high frequency wireless wake up signals. As discussed in [9], On-Off-Keying (OOK) saves much power.

Attending to the WUp receiver features, a wake up signal can be either a simple tone or a 8 to 16 bit pattern, which is compliant with 30 kHz sub-carrier defined by ISO/IEC 18000-7 standard. An example of OOK wake up signal is shown in Fig. 3.

3.2.2 WUp Transmitter

The RFID reader does not need any additional hardware apart from the in built transponder. The reader needs additional firmware to transmit suitable WUp signals as depicted in Fig. 3. WUp tones are generated by the reader micro-controller by setting an auto-reload timer at double the desired modulation frequency (60 kHz in ISO 18000-7 case) and by changing RF output power (from maximum to zero and vice-versa) upon timer overflow. Output RF carrier signal suffers On-Off switch generating a square wave of the desired frequency On-Off-Keyed.

Power dissipated to generate and emit this signal is equal to the power consumed when emitting FSK data packets.

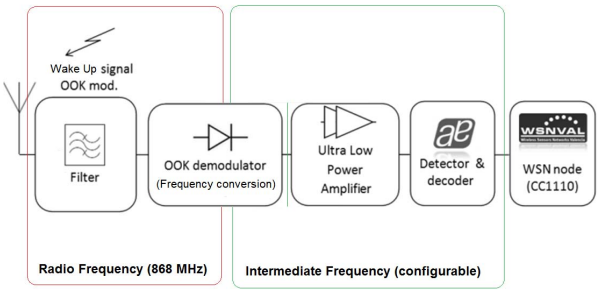


Fig. 4: Wake-Up Receiver Subsystem Block Diagram.

3.2.3 Configurable IF Super-heterodyne WUp Receiver

WUp receiver block diagram is shown in Fig. 4. This subsystem is based in a super-heterodyne structure in which an incoming signal is shifted to a lower Intermediate Frequency (IF) to be further processed. This technique has been proved to improve receivers performance [9] [8]. After filtering RF signal to avoid potential interferences, the resulting signal is mixed to a lower frequency band. Then this IF signal is processed by a commercial wake-up detector which can set a flag to activate the micro-controller.

The original wake up solution presented in [9], operates using a fixed intermediate frequency: 125 kHz. This value was set by the decoder: AS3930 [12]. As discussed by the authors, this block is critical to achieve optimal solutions for both wake up signal detection range and power consumption, reporting indeed the lowest consumption (8.7μW) and the greatest range (15 m) until now.

An interesting alternative to both improve wake up signal detection with a potential frequency deviation is reported in [8]. Using a flexible and uncertain intermediate frequency, radio-frequency blocks are simplified and the overall behaviour improves significantly. This concept is shown in Fig. 5. IF varies within a certain range. Therefore, the decoder must be able to decode wake-up signals with this uncertainty. Using this technique authors improved significantly its previous wake up receiver [13] sensitivity (50 %) and reduced its power consumption (20 %).

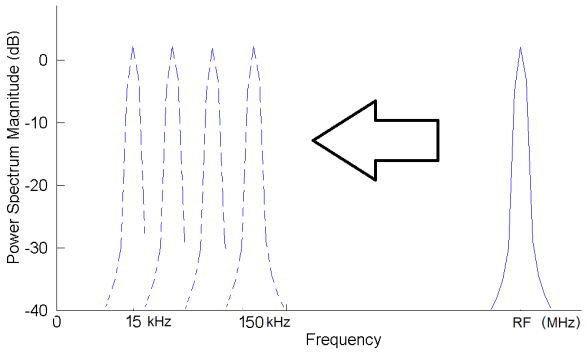


Fig. 5: RF frequency (868 MHz) and Flexible IF (15-150 kHz) signal spectrum.

it-RFID wake up receiver implements "flexible frequency range concept" to the WUp system presented in [9]. Modern low frequency wake up signal decoders have improved their performance. e.g. AS3933 [14]. This chip is compatible with the previous decoder AS3930 and implements some new interesting features. Decoded signals can not only be fixed to 125 kHz, but take a certain value from 15 kHz to 150 kHz as shown in Fig. 5. Besides, this decoder includes an internal 3 dB low noise amplifier that can be activated on demand. Decoder sensitivity increases from 100 to 80 μV_{rms} (20 %) and only extra 300 nW are dissipated.

Using the design methodology explained in [9] to embed a low-frequency decoder into a high frequency system, a new radio-triggered wake up receiver has been obtained. Since both WUp signal and transmitter can upgraded with this new enhancement by changing the WUp transmitter firmware, the result is a new radio wake up system with flexible IF.

Finally, this new wake up system has been successfully attached to the it-RFID architecture and final performance is evaluated in Section 4.

4. it-RFID evaluation

To carry out a full evaluation, a prototype of the whole proposed it-RFID platform has been developed. ISO/IEC 18000-7 message collection protocol has been implemented according to the specifications explained in Section 2.

In this paper two features are evaluated: the new flexible IF WUp system improvement compared to original WUp solution [9]; on the other hand an evaluation in a realistic scenario of the power consumption of it-RFID compared with Free2move active RFID solution [5] and with a 802.15.4 WSN platform (CC2430 from TI) [2].

4.1 Detection Distance

Maximum distance reported using the original wake up receiver was 15 meters [9] and power dissipated was 8.7 μW. In this section the proposed improvements of the new WUp system are evaluated as well as their convenience attending to the impact on other features such as power consumption, costs, etc.

Firstly, the decoder in-built 3dB amplifier performance. When active, receiver sensitivity improves and is reflected in a wake up detection range increase from 15 to 17.5 meters (16 % improvement). Since power

consumption only raises 11 % while distance raises 16 %, it is advisable to use this feature in most of the cases.

Secondly, the new flexible IF feature has been also evaluated. The decoder is capable of detecting WUp signals within 15 and 150 kHz, divided in 5 frequency rages: 15-23, 23-40, 40-65, 65-95, 95-150 kHz. WUp signal detection range is measured for each of these sub-bands.

Results are shown in Table 1. Different bands are evaluated using the following frequencies: 20 kHz, 30 kHz, 50 kHz, 80 kHz, 125 kHz. Since additional 3dB amplifier is active to obtain the best performance results, two improvement values are calculated:

- *Partial improvement* that compares the range measured with the it-RFID wake-up with 3dB gain boost activated;
- *Overall improvement* that compares the results with the original WUp system.

As can be seen in Table 1, the lowest intermediate frequency, the widest range measured. OOK demodulator block in wake up receiver (Figure 4) is based in a Dickson Charge pump composed of diodes and capacitors [9]. It has been observed that even capacitors with a few Farads are not completely charged and discharged with highest IF values. Using lower IF, OOK demodulator efficiency increases, decoder input voltage raises and range is consequently extended.

4.2 Power Consumption

Although lots of work has been done to improve ISO/IEC 18000-7 standard -specially power consumption- this work is not aware of improving the protocol itself, but evaluating the power consumption of it-RFID platform compared to other active RFID platforms used as reference.

To evaluate it-RFID power consumption, an scenario similar to the proposed in [2] has been set. Originally, authors compared Free2move (*F2m*) active RFID platform [5] implementing a Reader Talk First (RTF) protocol ISO 18000-7 compliant, with a standard WSN 802.15.4 platform based on CC2430 from TI (15.4). it-RFID with presented wake up enhancement ISO/IEC 18000-7 compliant is added to this comparative in Table 2.

RFID readers have not been considered since they are usually powered by the mains and their power consumption is not critical. However, active RFID tags energy buffers are usually limited and it is really interesting to evaluate their consumption in order to estimate their operation lifetime.

Table 1: Study of intermediate frequency in RT-WUp maximum distance

Frequency (kHz)	Range (m)	Partial Improvement (%)	Overall Improvement (%)
20	20	14.3	33.3
30	20	14.3	33.3
50	20	14.3	33.3
80	17.5	0	16.7
125	17.5	0	16.7

Power consumption values of all three platforms when running different tasks related to active RFID protocol -as proposed in [2]- are reported in Table 2. Since it-RFID and Free2move platforms implement RTF version of ISO 18000-7, different tasks duration are the same in both cases.

The main advantage of it-RFID is the exploitation of the RT-WUp to perform ISO 18000-7 WUp signal or even the whole Beacon detection. Only if the wake up signal is detected, tag switches to receive (RX) mode to decode the incoming collection command, remaining in sleep mode otherwise. Other platforms needs to periodically check the channel. This value is compared in $\hat{E}_{Beacon}$ row of Table 2.

The improvement has been evaluated in a real case scenario with some simple assumptions to avoid effects related to upper layer issues such as medium access, application, etc. Lets consider an application with one single tag and one single reader. Under this assumption we can assume that if a tag is near enough to detect reader Beacons (20 meters), tag responses are collision free and no extra collection rounds are needed in one collection round. Thus, two situations are possible: a reader is available or not.

On the one hand if a reader is available, the collection sequence explained in Section 2 is carried out. The energy consumed by a tag in one cycle can be calculated using (1). As shown in Fig. 1 two messages are sent from the reader (Collection and Sleep Command), and one is sent back from the tag (Tag Response). The rest of the time the node is sleeping.

$$E_{reader} = \bar{E}_{Beacon} + E_{TX} + E_{RX} + E_{Sleep1} \quad (1)$$

Using expression (1), it is estimated a consumption per cycle of 0.466 mJ for the Free2move platform, 0.78 mJ for the 802.15.4 platform and 0.55 mJ for the it-RFID platform. When reader is available, it-RFID is a balanced solution among the three compared platforms due to the radio chip power consumption.

On the other hand if a reader is not available, the energy calculated per cycle can be estimated using (2). While energy consumed by 802.15.4 solution is 1.25 mJ, Free2move consumes 0.47 mJ and it-RFID only consumes 9 μJ, which is three orders of magnitude below the other platforms.

$$E_{no\ reader} = \hat{E}_{Beacon} + E_{Sleep2} \quad (2)$$

Attending to the previous results, it can be concluded that power savings depend on the presence of a reader.

Table 2: Terms, power, duration time, and energy consumed by: Free2move (F2m), 802.15.4 (15.4) and it-RFID (it)

term	Power [mW]			duration [ms]		$\frac{E}{cycle}$ [mJ]			explanation
	F2m	15.4	it	F2m & it	15.4	F2m	15.4	it	
$\bar{E}_{Beacon}$	57	81	60	5	7.7	0.274	0.622	0.00004	avg. energy consumption when trying to receive a beacon signal, reader available
$\hat{E}_{Beacon}$	57	81	0.009	8	15.4	0.456	1.244	0.0006	avg. energy consumption when trying to receive a beacon signal, no reader available
E_{RX}	57	81	60	2	0.32	0.114	0.026	0.162	energy consumption when receiving a packet from a reader
E_{TX}	42	75	63	2	1.6	0.067	0.12	0.15	energy consumption when transmitting one payload packet to a reader
E_{Sleep1}	0.0011	0.003	0.003	992	990	0.011	0.003	0.009	energy consumption when sleeping after successfully delivered payload to a reader
E_{Sleep2}	0.0011	0.003	0.009	992	985	0.011	0.003	0.009	energy consumption when sleeping after listening for a beacon, no reader available

To get a wider perspective, different scenarios have been considered simulating the presence and the absence of readers with a certain probability. For each scenario, energy consumed during 24 hours operation is calculated for both platforms.

Results are shown in Fig. 6. 802.15.4 based tag consumes less power when a reader is present. Although it seems a contradiction, to check the channel waiting for Beacons dissipates actually more power than emitting and receiving packets. In Free2move platform, energy consumption remains almost constant since it consumes approximately the same energy either with a reader present or not. Finally, it-RFID draws the minimum power when a reader is absent while power consumption increases as long as radio is used to transmit useful information when a reader is present (it-RFID Energy Consumed in Fig. 6), what is much closer to the optimal case.

it-RFID improvement is also evaluated in Fig 6. it-RFID Energy Normalized curve represents the Free2move and 802.15.4 energy to it-RFID energy ratio (expressed in %) for each scenario. More than 98 % energy can be saved if there is no reader available. Nevertheless, up to 11 % can be wasted in scenarios with a reader available all the time compared to Free2move, but still saves 29 % compared to 802.15.4 solution.

Finally, another interesting energy analysis proposed in [2] is the estimation of a small CR2032 lithium cell lifetime (3V/180mAh). Authors selected the case in which no reader was available. In this paper we extend these results with it-RFID performance and also considering the case in which a reader is available all the time. These results are shown in Table 3.

Since power consumption of it-RFID with reader absent (9 μW) is even lower than the power consumption of Free2move platform in sleep mode (11 μW), battery lifetime using it-RFID is even longer than Free2move ideal case (Free2move considering that no extra power is needed to detect a Beacon). If a reader is available, differences decrease and all the solutions deplete the battery energy before two months.

Table 3: Battery lifetime when no available reader is present and reader is present all the time

Platform	3V/180mAh CR2032 lifetime [days]	
	No reader	Reader
Free2move	48	45
802.15.4	18	30
it-RFID	2586	41
Free2move ideal	2045	45

5. Conclusions and Future Work

In this paper it-RFID is presented. A new active RFID platform ISO 18000-7 standard compliant has been developed and tested using a prototype. The platform is enhanced with a Radio Triggered Wake Up system to improve RFID Beacons detection, reducing RFID tag power consumption while no RFID readers are present.

A new Radio Triggered Wake Up system has also been designed. It is based on a super-heterodyne Wake Up Receiver with flexible Intermediate Frequency and additional 3dB low power amplification. We have observed an improvement of 33 % distance compared with the original wake up system. This new system has been successfully attached to it-RFID.

it-RFID performance is compared with two significant active RFID platforms. It is proved that power consumption with this new solution saves lots of energy when implementing ISO/IEC 18000-7 standard RFID protocol. Higher abstraction layers protocols can be implemented in future using this platform to obtain optimal active RFID solutions.

6. Acknowledgements

The authors gratefully acknowledge financial support from the CICYT (research projects CTM2011-29691-C02-01 and TIN2011-28435-C03-01).

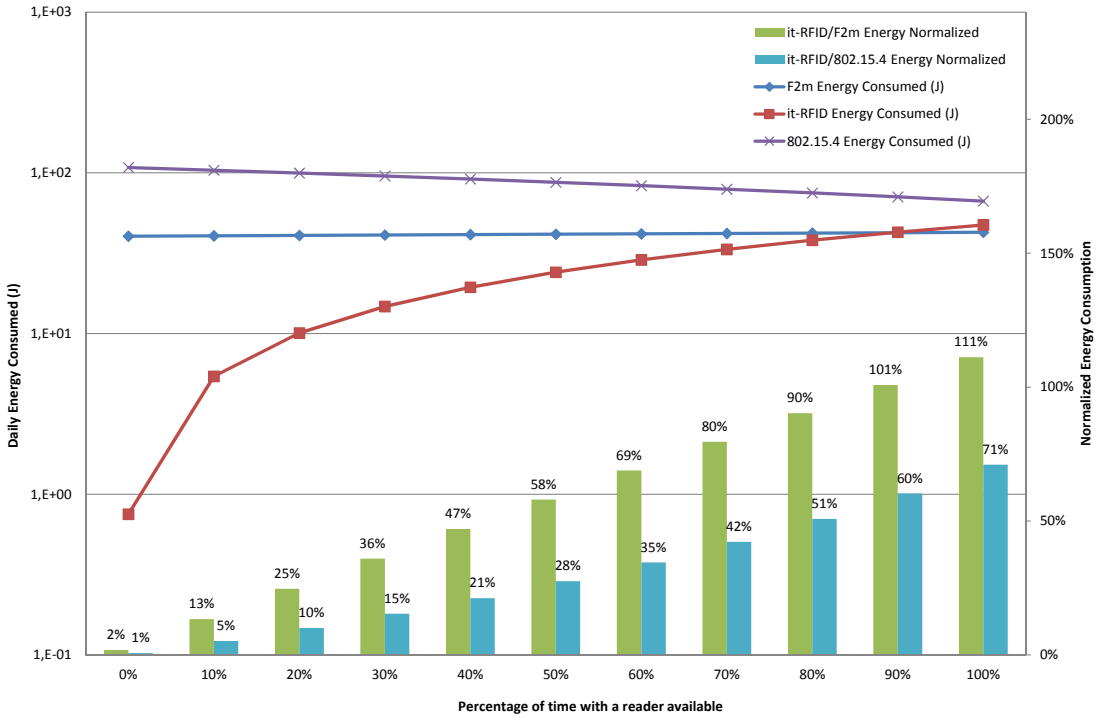


Fig. 6: Power consumption savings analysis

References

[1] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Card Identification*, 2nd ed. John Wiley & Sons, 2003.

[2] B. Nilsson, L. Bengtsson, P.-A. Wiberg, and B. Svensson, "Protocols for Active RFID - The Energy Consumption Aspect," in *2007 International Symposium on Industrial Embedded Systems*. IEEE, July 2007, pp. 41–48.

[3] W.-J. Yoon, S.-H. Chung, H.-P. Kim, and S.-J. Lee, "Implementation of a 433 MHz Active RFID System for U-Port," in *The 9th International Conference on Advanced Communication Technology*. IEEE, Feb. 2007, pp. 106–109.

[4] ISO/IEC, "18000-7," 2004. [Online]. Available: http://www.iso.org

[5] W.-J. Yoon, S.-H. Chung, S.-J. Lee, and Y.-S. Moon, "Design and Implementation of an Active RFID System for Fast Tag Collection," in *7th IEEE International Conference on Computer and Information Technology (CIT 2007)*. IEEE, Oct. 2007, pp. 961–966.

[6] U. Bilstrup and P.-A. Wiberg, "An architecture comparison between a wireless sensor network and an active RFID system," in *29th Annual IEEE International Conference on Local Computer Networks*. IEEE (Comput. Soc.), pp. 583–584.

[7] H. Cho and Y. Baek, "Design and Implementation of an Active RFID System Platform," in *International Symposium on Applications and the Internet Workshops (SAINTW'06)*. IEEE, Jan. 2006, pp. 80–83.

[8] N. Pletcher, S. Gambini, and J. Rabaey, "A 52 μW Wake-Up Receiver With -72 dBm Sensitivity Using an Uncertain-IF Architecture," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 269–280, Jan. 2009.

[9] A. Sanchez, J. Aguilar, S. Blanc, and J. J. Serrano, "RFID-based wake-up system for wireless sensor networks," in *Proceedings of SPIE*, vol. 8067, no. 1, May 2011, pp. 806 708–806 708–12.

[10] B. Van der Doorn, W. Kavelaars, and K. Langendoen, "A prototype low-cost wakeup radio for the 868 MHz band," *International Journal of Sensor Networks*, vol. 5, pp. 22—-32, 2009.

[11] WSNVAL, "Wireless Sensor Networks Valencia." [Online]. Available: www.wsnval.com

[12] Austria Microsystems, "AS3930 - Single Channel Low Frequency Wakeup Receiver."

[13] N. Pletcher, S. Gambini, and J. Rabaey, "A 65 μW, 1.9 GHz RF to digital baseband wakeup receiver for wireless sensor nodes," in *2007 IEEE Custom Integrated Circuits Conference*. IEEE, Sept. 2007, pp. 539–542.

[14] Austria Microsystems, "AS3933 – 3D Low Frequency Wakeup Receiver."

Grouped-Subcarrier Based Null-Data Switching for PAPR Reduction of OFDM with Low Computational Complexity

Sabbir Ahmed and Makoto Kawai
Graduate School of Science and Engineering,
Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu city, Japan

Abstract— *Being a multicarrier modulation scheme, Orthogonal Frequency Division Multiplexing (OFDM) systems generally produce transmit signals with high Peak to Average Power Ratio or PAPR. Amongst numerous PAPR reduction strategies found in the literature, a relatively new but promising technique is the null and data subcarrier switching method. In this paper, we propose a new approach of null-data subcarrier switching that can achieve acceptable PAPR reduction with significantly less computational overhead. Apart from offering advantageous features like side-information less detection, distortion-free signal transmission and compatibility with other existing PAPR reduction techniques, our method also overcomes the problem of high computational requirement especially when higher number of null subcarriers are used for switching. This is achieved by applying incremental searching within preformed subcarrier groups. We demonstrate the effectiveness of our method by presenting PAPR and Bit Error Ratio (BER) related simulation results.*

Keywords: OFDM, Null and data subcarrier switching, PAPR.

1. Introduction

As a multicarrier modulation scheme, Orthogonal Frequency Division Multiplexing or OFDM offers many attractive features, e.g., high data rate, robust performance in inter-symbol interference (ISI) channels, good spectrum efficiency and so on. Considering these advantages with many others, it has been selected as the physical layer standard for different contemporary communications systems [1]-[3]. However, OFDM system often contains occasional very high peaks in its transmit signal. This high peaks are generally quantified by a parameter called the Peak to Average Power Ratio or PAPR. High PAPR is a problem when the issue of amplification comes into play. High PAPR demands a highly linear amplifier so as not to cause nonlinear distortion. Because, if peak power of the signal crosses the operating range of the amplifier, it may get driven well into the saturation region causing severe Bit Error Ratio (BER) degradation. On the other hand, amplifier are generally operated near the saturation point for attaining better efficiency. This is imperative especially for power constrained hand-held mobile devices. These two scenarios combined makes PAPR reduction a crucial issue for any system that employs OFDM.

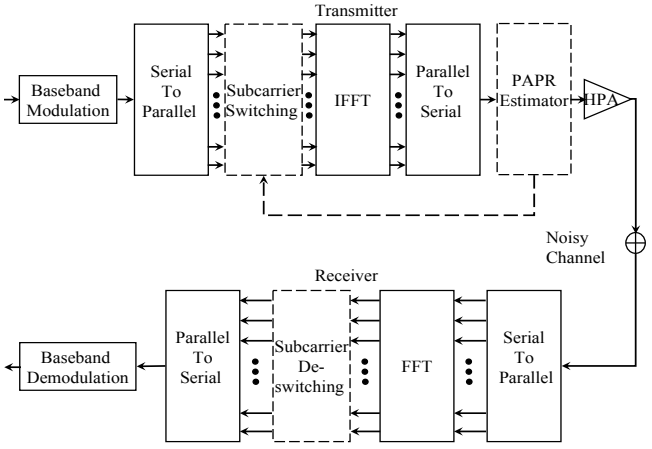


Fig. 1: System model of OFDM with subcarrier switching.

In consideration of its effect on the overall system performance, PAPR reduction remains a major research interest for quite some time. And as outcome, quite a significant number of PAPR reduction techniques can be traced in literature. Amongst them, some very well known techniques are selected mapping (SLM), partial transmit sequences (PTS) signal clipping, dummy sequence insertion etc. [4]-[7]. Signal clipping is a very simple and effective PAPR reduction technique. But it can give birth to spectral regrowth of the transmitted signal and there-by cause inter-channel interference (ICI) [8]. Filtering can solve this issue, but only at the cost of potential peak regrowth. On the other hand, SLM and PTS are distortion less techniques and hence do not suffer from the above mentioned problems. However, the transmitter must send additional information known as side-information to the receiver without which the later can not reconstruct the data sequences. This means the effective data-rate gets reduced or valuable bandwidth is lost. Moreover, the associated computational costs for both these schemes can be very high. Dummy sequence insertion method does not require side-information but data throughput needs to be sacrificed for accommodating dummy bits.

A survey of recent PAPR related research work reveals

that moving from the earlier approach of considering generalized architecture, current investigations are considering systems based on standard specifications, e.g., WiMAX (IEEE 802.16) or WLAN (IEEE 802.11a) [9]-[11]. For example, the tone reservation with null subcarriers (TRNS) scheme designed for WiMAX systems shows good PAPR reduction capability. But finding the best set of subcarriers as reserved tones within acceptable computation cost is still an open problem.

A relatively more recent scheme called the null and data subcarrier switching method [11] also shows significant PAPR reduction by switching null subcarriers with data subcarriers specified in the IEEE802.11a standard specifications [1]. Unlike signal clipping, it does not distort the transmission signal and also there is no need for side-information transmission as required by SLM or PTS techniques. But the problem of this method is that the computational complexity can be extremely high when higher number of null subcarriers for switching are used or systems with higher number of subcarriers, e.g., WiMAX is considered.

In this paper, we present new methods of subcarrier switching to reduce computational complexity of the original null and data subcarrier switching scheme without sacrificing PAPR reduction capability. For this, we at first propose incremental searching of data subcarriers for switching with null subcarriers. We show that this approach achieves similar level of PAPR reduction with significantly low computational overhead. But we also report that it may need to sacrifice the no side-information advantage especially when the number of null subcarriers for switching is increased. We overcome this problem by proposing further modification where we partition the data subcarriers into groups and search within them. We demonstrate that this method requires much less computations compared to the original method, achieves almost same level of PAPR reduction and also does not compromise on the no side-information advantage even when higher number of null subcarriers are used for switching. We argue that to achieve similar PAPR reduction, the original method with no side-information requirement becomes almost impractical due to its very high level of computational overhead.

2. System Architecture

In Fig. 1, we show the system model of our study. In this figure, blocks drawn with dashed lines represent the components required for PAPR reduction functionality. Here, at first binary random input data is baseband modulated that generates input symbols given by $x[i] = x[0], x[1], ..., x[N-1]$. They are then converted from serial to parallel and fed into the IFFT module. The IFFT module performs the task of multicarrier modulation. Output of the IFFT is parallel to serial converted and the resultant time domain signal, $X[n]$, is given by Eqn. 1.

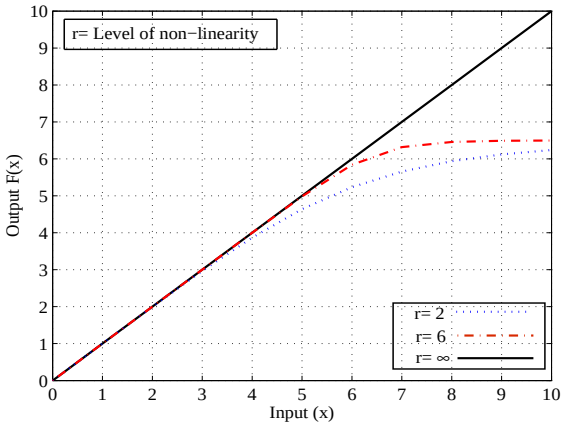


Fig. 2: Input-Output characteristics of non-linear amplifier.

$$X[n] = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} x[i] e^{j\frac{2\pi ni}{N}}, 0 \leq n \leq N-1 \qquad (1)$$

After this, the time domain signal is amplified by a High Power Amplifier (HPA) and transmitted to the noisy wireless channel. The receiver contains blocks that perform the operations of serial to parallel conversion, FFT, parallel to serial conversion and finally baseband demodulation.

Now, for PAPR reduction purpose a subcarrier switching module along with PAPR estimator block are inserted in the transmitter whereas a subcarrier de-switching module is placed in the receiver. The subcarrier switching module performs the operation of switching between null and data subcarriers and the de-switching block in the receiver executes the opposite operation. The PAPR estimator block determines and stores the PAPR value of all the switching combinations and transmits the signal with the lowest PAPR. PAPR in dB is expressed as below,

$$PAPR(dB) = 10 \log_{10} \frac{max|X[n]|^2}{E\left[|X[n]|^2\right]} \qquad (2)$$

where $E[\cdot]$ denotes expectation.

Finally, for simulating the HPA, we consider Solid State Power Amplifier (SSPA) model given in [14]. The AM-to-AM conversion characteristics of this model is depicted in Fig. 2 and the corresponding mathematical expression is given by Eqn. 3.

$$F[x] = \frac{x}{[1 + (x/A)^{2r}]^{1/2r}} \qquad (3)$$

Here, x is the amplitude of the input signal, A is the saturated output level and r is the non-linearity level. This model only considers AM-to-AM non-linearity. The parameter r can be used to tune the level of non-linearity. A large value of r turns this amplifier into a linear one where as very small values make it behave as a simple clipping amplifier.

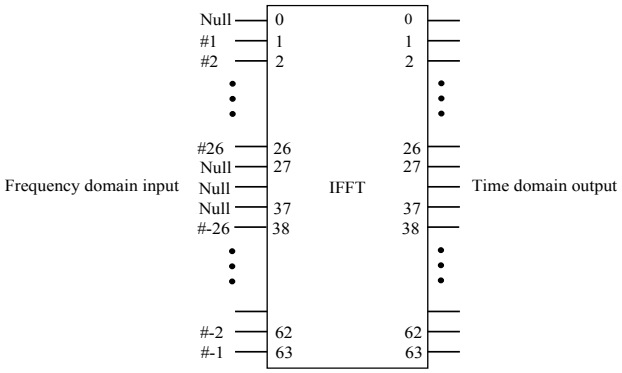


Fig. 3: IFFT layout as per IEEE 802.11a standard.

Algorithm 1 Proposed method-2.

Input: P, $\hat{g}_p, \{p = 1, ..., P \subset \mathcal{N}\}$, $\mathcal{D}$= $\{h_d, d = 1, ..., N - L\}$
Output: $\hat{h}_p$

```
Divide D into G_p groups
for p = 1 to P do
    for i = 1 to sizeof(G_p) do
        Switch ĝ_p with h_i
        Apply IFFT and Calculate PAPR_i
    end for
    ĥ_p = arg min_i PAPR_i
end for
```

3. PAPR Reduction Strategies

3.1 Null and Data Subcarrier Switching

The IEEE 802.11a WLAN standard specifies a total of 64 subcarriers that includes 12 null subcarriers. Among them, 6 null subcarriers work as the guard-band at the low-frequency edge of the spectral band while 5 null subcarriers serve as the guard-band at the high-frequency edge. The remaining one is placed at the middle to avoid direct current energy. The IFFT layout is shown in Fig. 3 [1]. The idea behind null-data subcarrier switching is to swap some of these null subcarriers with data subcarriers for PAPR reduction. [11].

Let us consider an OFDM system where the IFFT size and the total number of null subcarrier are N and L respectively. For this system we explain the null-data switching concept with the help of the following notations.

- Null subcarrier set: $\mathcal{N}$= $\{g_l, l = 1, ..., L\}$
- Data subcarrier set: $\mathcal{D}$= $\{h_d, d = 1, ..., N - L\}$
- Number of null subcarriers used for switching: P
- Indices of switching null subcarriers: $\{\hat{g}_p, p = 1, ..., P\} \subset \mathcal{N}$
- Indices of switching data subcarriers: $\{\hat{h}_p, p = 1, ..., P\} \subset \mathcal{D}$

The main concept here is to switch P number of null subcarriers with P number of data subcarriers such that if $\hat{h}_p < \hat{h}_{p+1}$ then $\hat{g}_p < \hat{g}_{p+1}$, all indices in ascending order and the task is to identify the $\hat{h}_p$ from h_d that results in lowest PAPR. As a result, $\binom{N-L}{P} = \frac{(N-L)!}{P!(N-L-P)!}$ number of different switching combinations need to be searched by the transmitter in order to choose the combination that yields the least PAPR. On the receiver side, subcarriers with low power levels are detected as null and they are de-switched with corresponding data subcarriers. This de-switching can be done without any side-information because of the constraint if $\hat{h}_p < \hat{h}_{p+1}$ then $\hat{g}_p < \hat{g}_{p+1}$. Hence, this constraint is a must for facilitating side-information free de-modulation by the receiver.

Now, as mentioned above, $\binom{N-L}{P}$ number of search operations are involved for every input data block, which in turn means that many IFFT operations are required. For example, for a system with $N = 64$ and $L = 12$, if $P = 4$ is considered, $\binom{52}{4} = 270725$ number of search operations are required. The time involved in such calculation may make this technique impractical for many delay sensitive applications. Moreover, power consumption for such exhaustive calculations can also be a hindrance to the main objective of reducing PAPR, i.e., achieving power efficiency.

3.2 Computational Complexity Reduction by Incremental Search

In order to reduce the computational burden of the original null and data subcarrier switching method without making significant sacrifice in PAPR reduction, we proposed incremental searching by considering one null-data subcarrier switching at a time [12]. For example, for $P = 2$, i.e., number of null subcarriers to be switched is two, we start by considering $\hat{g}_1$ and search $\binom{N-L}{1}$, i.e., $N - L$ times to look for the data subcarrier position, say $\hat{h}_1$ yielding lowest PAPR when switched with $\hat{g}_1$. After this, we apply the same operation for the second null subcarrier to be switched. But in order to keep the order of null subcarriers and switched data subcarriers, we first remove all the data subcarriers positions from the search space, i.e., $h_d, \{d = (> \hat{h}_1), ..., N - L\}$. Thus the size of the data subcarrier search space for $\hat{g}_p$ when

Table 1: Simulation parameters

IFFT size	64
Number of data subcarriers	52
Number of switched null subcarriers	$2, 4$
Modulation scheme	BPSK
Total number of OFDM symbols	10^4
Oversampling factor	4
HPA Model	SSPA
HPA Saturation level	$3dB$
HPA level of non-linearity, r	1
Channel model	AWGN

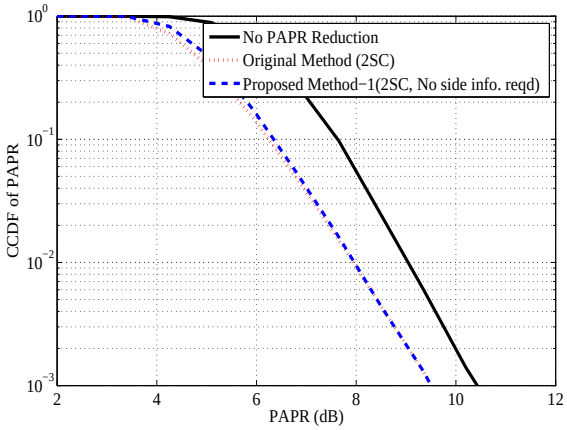


Fig. 4: Comparative PAPR for only incremental search ($P = 2$).

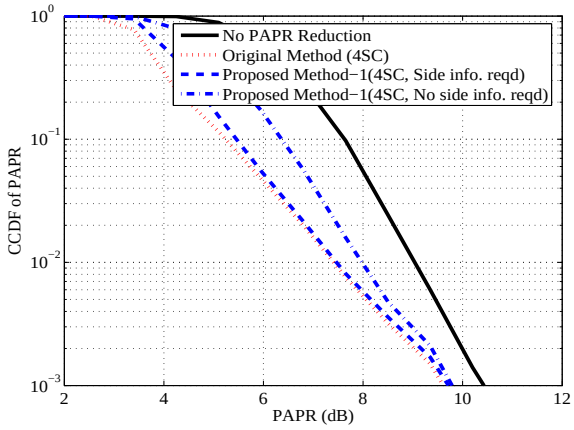


Fig. 5: Comparative PAPR for only incremental search ($P = 4$).

$p > 1$ is dependent on the outcome of search with $\hat{g}_{p-1}$. For this approach, since the search space becomes smaller for every increment in p, there remains a possibility that the search space would become very limited for $\hat{g}_p$ when $p > 1$, specially when $P > 2$, causing significant negative effect on PAPR reduction. One way of avoiding this problem, is to relax the need for keeping the order of the switched null and data subcarriers same. But it means the receiver now can not perform de-switching unless some extra information, i.e., side-information is transmitted.

3.3 Computational Complexity Reduction by Incremental Search within Subcarrier Groups

We propose segregating the data subcarriers into groups to facilitate searching with less computational burden without compromising the no side-information advantage. The algorithm behind this method is depicted in Alg. 1. Here, we at first partition total number of data subcarriers into $G_p, \{p = 1, ..., P\}$ groups and put $\frac{N-L}{P}$ number of subcarriers into each group when $N - L$ is divisible by P. Otherwise, in the first group we put slightly higher number of subcarriers compared to the other $P - 1$ groups all holding equal number of subcarriers. Then we start with G_1 and search for the data subcarrier $\hat{h}_1$ which when switched with null subcarrier $\hat{g}_1$ yields lowest PAPR. We carry on the same process until p reaches P. Unlike [13], here the length of the subcarrier groups remain constant throughout the entire search operation and hence there is no additional overhead for dynamic adjustment of subcarrier bands. This approach of searching reduces the computational burden significantly. For example, with $P = 4$, our method requires 52 number of searchings compared to 270725 of the original method. And as we show in the following section, the PAPR reduction capability remains almost the same.

4. Simulation Results and Discussion

For comparative performance analysis, we performed simulation based on the model shown in Fig. 1. We consider BPSK modulation and investigate scenarios primarily for $P = 4$. In addition, we also show results for $P = 2$ and $P = 4$ where we do not implement subcarrier grouping. For $P = 2$ and $P = 4$, we select null subcarriers for switching positioned at ± 27 and ± 27, ± 28 respectively ([1] page no.12). The rationale behind choosing the null subcarrier positions is to select them in a balanced way on either side of the data subcarrier bands so that any degradation to the guard bands can be kept to a minimum. All the pertinent simulation parameters are listed in Table 1. In Figs. 4 and 5, we show the results where we consider incremental search but do not divide data subcarriers into different groups. Figure 4 shows the comparative cumulative distribution function (CCDF) of PAPR for OFDM without any PAPR reduction scheme, the original null and data subcarrier switching and our method where number of null subcarriers for switching is 2, i.e., $P = 2$. Here, we denote the original null and data subcarrier switching as “Original Method" and our method as “Proposed Method-1", respectively. As seen here, compared to no PAPR reduction scheme, our method achieves significant PAPR reduction which is slightly inferior to the original method.

Now in Fig. 5, for $P = 4$, we show the results without and with the need of side-information transmission. As was mentioned earlier also, for $P > 2$, we can see from this figure that the no side-information advantage needs to be sacrificed in order to achieve significant PAPR reduction, otherwise PAPR reduction is drastically reduced.

In Fig. 6, we show the PAPR performance when segregation of subcarriers into different groups is considered. We referred to it as “Proposed Method-2" in In Fig. 6. Here, we can see that for PAPR values ≤ 6.5dB, our proposed

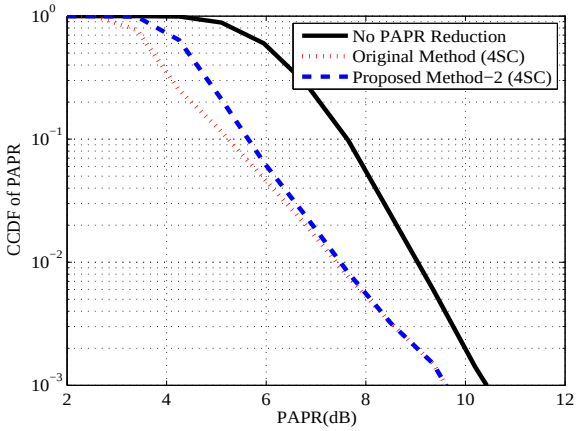


Fig. 6: Comparative PAPR for incremental search within data subcarrier groups ($P = 4$).

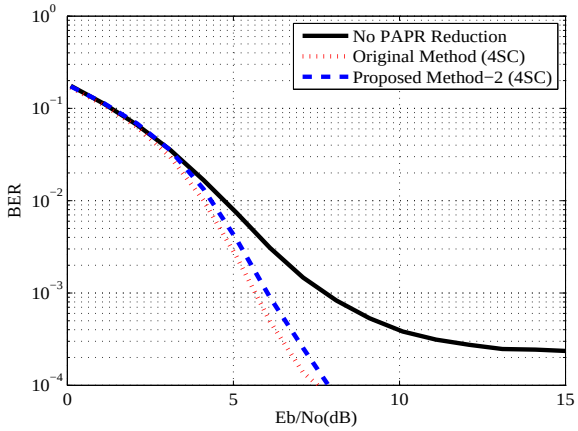


Fig. 7: Comparative BER for incremental search within of data subcarrier groups ($P = 4$).

method suffers marginal loss in PAPR reduction compared to the original scheme. But for higher PAPR, it shows almost same performance as the original one. This is accompanied by the fact that our method achieves significant eduction in computational overhead since corresponding number of searchings for the original and our proposed methods are 270725 and 52 respectively.

Finally, in order to show the effects of PAPR on the BER performance of the system, we consider a non-linear power amplifier and an Additive White Gaussian Noise (AWGN) channel. Figure 7 shows the comparative BER results of our proposed method, the original method for $P = 4$ and an OFDM system without any PAPR reduction mechanism applied. The combined effects of noise contamination of the wireless channel and the non-linear amplification due to high PAPR govern the results. As evident from Fig. 7, OFDM without any PAPR reduction mechanism shows considerable amount of error floor even at very high SNR level. And compared to it, our method shows significantly improved BER performance which is almost similar to that of the original method.

5. Conclusions

A new approach of null and data subcarrier switching scheme for PAPR reduction in OFDM systems is proposed. We show that incremental searching can reduce computational complexity of the original method but may need to sacrifice the no side-information advantage when higher number of null subcarriers are used for switching to achieve greater PAPR reduction. We then propose incremental searching within pre-formed data subcarrier groups and show that besides achieving PAPR reduction with low computational overhead it also retains the no side-information advantage. Through simulation results, we show the PAPR reduction capability and the BER performance of our method.

As a plan for future research, we are interested in investigating the robustness of our method by considering higher level modulation schemes, e.g., QPSK or QAM along with multipath propagation environment. In parallel, we are also interested in exploring the effect on the spectrum containment when switching between null and data subcarriers takes place.

References

[1] *IEEE, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, High Speed Physical Layer in the 5GHz Band*, IEEE Std 802.11a, 1999.

[2] IEEE, *Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE std. 802.16-2004 (Revision of IEEE Std. 802.16-2001), 2004.

[3] http://www.3gpp.org

[4] R. W. Bauml, R. F. H. Fischer and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping", *Electronics Letters*, vol. 32, no. 22, pp. 2056-2057, 1996.

[5] L. J. Cimini and N. R. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences", *IEEE Communications Letters*, vol. 4, no. 3, pp. 86-88, 2000.

[6] X. Li and L. J. Cimini, "Effects of clipping and filtering on the performance of OFDM", *IEEE Communications Letters*, vol. 2, no. 5, pp. 131-133, 1998.

[7] H.-G. Ryu, J.-E. Lee and J.-S. Park, "Dummy sequence insertion (DSI) for PAPR reduction in the OFDM communication system", *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 89-94, 2004.

[8] S. C. Thompson, J. G. Proakis and J. R. Zeidler, "The effectiveness of signal clipping for PAPR and total degradation reduction in OFDM systems", *in Proc. IEEE GLOBECOM '05*, St. Louise, Montana, USA, pp. 2807-2811, Januray 2006.

[9] S. Hu, G. Wu, Q. Wen and Y. Xiao, "Nonlinearity reduction by tone reservation with Null subcarriers for WiMAX system",*Wireless Personal Communications*, vol. 54, no. 2, pp. 289-305, April 2009.

[10] S. Hussain, D. Guel, Y., Louet and J. Palicot, "Performance comparison of PRC based PAPR reduction schemes for WiLAN Systems", *in Proc. European Wireless Conf., 2009*, Aalborg, Denmark, pp.167-172, 2009.

[11] K. T. Wong, B. Wang and J.-C. Chen, “OFDM PAPR reduction by switching null subcarriers and data-subcarriers", *Electronics Letters*, vol. 47, no. 1, pp. 62-63, 2011.

[12] S. Ahmed and M. Kawai, “A Reduced complexity subcarrier switching scheme for PAPR reduction in OFDM system", *in Proc. Int. Conf. on Future Generation Communication and Networking (FGCN 2011)*, Springer CCIS Part I, pp. 67-76, December 2011.

[13] S. Ahmed and M. Kawai, “Dynamic Null-Data subcarrier switching for OFDM PAPR reduction with low computational overhead", *Electronics Letters*, vol. 48, no. 9, pp. 498-499, 2012.

[14] C. Rapp, “Effects of HPA-nonlinearity on a 4-DPSK/OFDM signal for a digital sound broadcasting system", *in Proc. Second European Conference on Satellite Communications*, Liége, Belgium, pp. 179-184, October 1991.

Hierarchical Routing Approach-Based Energy Optimazation in Wireless Sensor Networks

H. Aoudia ,Y. Touati, P. Greussay and A. Ali-Cherif
Computer Science and Artificial Intelligence Lab. LIASD dept. MIME
University of Paris 8 at Saint-Denis
2, rue de la Liberté, 93526 Saint-Denis Cedex, France
Email: {hania, Touati, pg, aa}@ai.univ-paris8.fr

***Abstract*—In Wireless Sensor Network (WSN) the main challenge is to find a methodology to extend the lifetime of the network by taking into account energy considered as the main resource that should be optimized. This can make the network very effective and robust and reactive. In this paper, we are interested by the way information are transmitted and routed to the destination in an optimized manner. Thus, we propose and study a routing protocol LEACH-Mod based on clustering methodology using the concept of hierarchical standard routing protocol LEACH. The effectiveness of the proposed routing protocol is tested by conducting a comparative study with LEACH protocol.**

Keywords-*WSN; routing protocol; energy consumption.*

I. INTRODUCTION

Few years ago, WSN implementation has become increasingly important for performing potential applications for variety of fields, including medical monitoring, military operations, rescue missions, climate change, and so on [1-4]. For such applications, it is clear that routing protocols are required in order to insure an efficient communication by transmitting data between sensor nodes and the base stations. This has led to quite a number of different protocols operating in different layers of the network, with the goal of allowing the network working autonomously for as long as possible while maintaining data channels and network processing to provide the application's required quality of service QoS. Different algorithms have been proposed in this context. They can be classified according to different parameters related to a considered application and to the criteria which should be optimized. Thus, protocols can be classified as proactive, reactive and hybrid, based on their operating mode and type of applications. In this paper, we are only interested by proactive protocols. In this kind of protocols, several algorithms have been investigated in the literature. The most studied one is Low Energy Adaptive Clustering Hierarchy (LEACH), which is the first hierarchical cluster-based routing protocol for WSN classifying the nodes into clusters [5-6]. In each cluster, a coordinator node with an extra privileges called Cluster Head (CH) is responsible for creating and manipulating messages using TDMA (Time division multiple access) schedule and sending aggregated data from nodes to the base station according to CDMA (Code division multiple access). Implementing this protocol increases the network performances in term of energy consumption but suffer from many drawbacks in that CH nodes (CHs) selection is randomly and not uniformly distributed (CHs can be located at the edges of the cluster). Some variants of this protocol have also been developed [7]. Energy-LEACH is a protocol that improves the CH node selection procedure. It makes residual energy of node as the main metric which decides whether the nodes turn into CH or not after the first round. The protocol consists of several rounds. In the first one, every node has the same probability to turn into CH node, that means nodes are randomly selected as CHs, in the next rounds, the residual energy of each node is different after one round communication and taken into account for the selection of the CHs. That mean nodes have more energy will become a CHs rather than nodes with less energy. In LEACH, Each CH node communicates directly with the base station without worrying about distances. Thus, more the distances between CH node and the base station is important more energy consumption is high. In Multihop-LEACH protocol, an optimal path between CH node and the BS through other CHs is selected, and the data is transmitted via other CHs playing the role as a relay station [8]. Multihop-LEACH protocol is almost the same as LEACH protocol, only makes communication mode from single hop to multi-hop between CHs and BS. In [9], a LEACH-C protocol was proposed. It uses a centralized clustering algorithm and the same steady-state phase as LEACH. It can produce better performance by dispersing the CHs nodes throughout the network. In To determine good clusters and CHs nodes and during the set-up phase, each node sends information about its current location and residual energy level to the sink. Once the CHs nodes and associated clusters are found, the sink broadcasts a message that obtains the cluster head ID for each node. If a CH ID matches its own ID, the node becomes CH otherwise the node determines its TDMA slot for data transmission and goes sleep. In [10], Vice-LEACH protocol's developed. It consists to elect a vice-Ch in a cluster, taking the role of the CH when this latter disappear from the network. By doing this, cluster nodes data will always reach the base station and there's no need to elect a new CH each time the CH dies. This will extend the overall network life time. Other protocols have been investigated in this direction, such as HEED [11], PEGASIS [12], TEEN [13] and APTEEN [14]. These methods are quite efficient, but find

promptly their limitations when the density of the networks is large. Indeed, for a hierarchical topology, CHs nodes overloading may force the network to consume more energy and therefore, reducing its lifetime which is one of the major challenges to extend in WSN. Conceiving and developing a routing protocol for WSN requires taking into account considerations related to the problem of energy optimization. In this paper, we propose and study a hierarchical routing protocol LEACH-Mod based on clustering mechanism which considers both residual energy of sensor nodes, end-to-end delay for routing information and number of messages that have been successfully sent and received.

The remainder of this paper is organized as follows: Section 2 presents the basic principles of the standard routing protocol LEACH and the proposed routing approach LEACH-Mod. To test the effectiveness of the proposed routing approach, we have conducted in section 3, a comparative study between LEACH-Mod and a standard routing protocol LEACH. Finally, a conclusion is dressed in section 4.

II. Proposed Routing Approach Based-Energy Consumption

The proposed routing protocol LEACH-Mod combines the concept of clustering and hierarchy to achieve energy-efficiency. The network is decomposed into a set of clusters wherein nodes organize themselves to become member nodes MNs and transmit their data locally to their CH nodes. The implementation of LEACH-Mod is divided into rounds, and each round is made up of a clustering phase, a set-up phase and a steady-state phase. During the clustering phase, commonly called advertisement phase, the base station announces a new round in witch new clusters are created and each node decides whether or not to become a CH. The decision is based on a probability of election of a given node as a CH and the suggested percentage of CH nodes [5% to 15%]. This election can be expressed as follows:

$$T(n)=\begin{cases}\dfrac{P}{1-P\cdot\left(r\cdot mod\left(1/P\right)\right)} & if \quad (n\in G)\\ 0 & otherwise\end{cases} \qquad (1)$$

where P and r denote respectively the percentage of nodes wishing to be CH and the iteration of the current round. G represents the set of nodes that hasn't been a CH during the last $(1/P)$ iterations. A random number between 0 and 1 is assigned to a node n. If the number is less than a threshold $T(n)$, the node becomes a CH for the current round and notifies its neighbors of its election. Once the clusters are formed for a given round, in the setup phase, each CH associated to a cluster will play a role as a coordinator for processing and transmitting data between MNs and the base station. Thus, it broadcasts an advertisement message to the rest of the nodes. Then, each node decides to belong to a cluster by informing the associated CH that it will be a MN. It depends on the received signal strength emitted by the CH nodes. Based on the number of nodes in the cluster, the CH node creates a TDMA MAC protocol or schedule telling each node when it can transmit data by assigning a timeslot. This schedule is broadcast back to the nodes in the cluster.

The Steady-state phase [15] represents the main phase where the energy consumption should be optimized. Based on TDMA protocol, MNs send their data to their CH during their allocated transmission timeslot. The CH nodes will aggregate all the collected data and transmit them directly or indirectly to the base station. The radio of each non-CH node can be turned off until the nodes have allocated the transmission time, thus minimizing energy dissipation in these nodes. The CH node must keep its receiver on to receive all the data from the nodes in the cluster [5]. When the data are completely received, the CH node performs the signal processing functions to compress the data into a single signal. This signal will be transmitted to the base station. Since the base station is far away, this is a high-energy transmission. After all data from all nodes are transmitted to the base station, a new round will start.

A. Energy model consumption

The residual energy on the nodes is crucial for the lifetime of the WSN. The consumption of energy is mainly generated by transmitting or receiving data, processing data, and idle listening. Recently, many different energy consumption models have been studied, including the energy dissipation in transmit and receive mode, which can change the advantages of different routing protocols. In contrast to the data processing phase, the communication phase requires substantial amounts of energy [16]. The average energy consumption in each CH node can be computed as follow:

$$E_{moy}=p_r\cdot E_1+(1-p_r)\cdot E_2 \qquad (2)$$

where:

$$E_1=E_{Tx}(k,d)+E_{Rx}\left(T_{inter}/T-k\right) \qquad (3)$$

and

$$E_2=E_{Rx}\left(T_{inter}/T\right)+E_{Rx}\left(T_{intra}/T\right) \qquad (4)$$

Parameters p_r and T represent respectively the probability that each node has k bits of data to be sent and the consumed time for transmitting a byte of data. T_{inter} and T_{intra} denote respectively the communication time between CH nodes and the base station, and the communication time between CH nodes and MNs during a round.

In the first term of (4), for a given probability P corresponding to an inter-CH communication phase, all CH nodes exchange information with the base station with an energy consumption equivalent to $E_{Tx}(k,d)$. The rest of the time $\left(T_{inter}/T-k\right)$ corresponds to the listening time where energy consumption is $E_{Rx}\left(T_{inter}/T-k\right)$. The second term of (4) corresponds to a probability $(1-p_r)$ where the CH node does not transmit any data to the base station. He spends all

his time for inter-CH communication in listening mode consuming energy of $E_{Rx}\left(T_{inter}/T\right)$. During intra-CH communication phase, CH node switches to a receive mode consuming $E_{Rx}\left(T_{intra}/T\right)$.

The proposed routing algorithm (Fig.1) considers some assumptions for its implementation. The network consists of n homogeneous nodes randomly distributed, which communicate directly in a single hop with the base station. For a given round, the base station uses the same procedure for electing new CHs nodes. Collisions avoidance for inter-CHs and intra-CH radio communication are insured respectively by CSMA/CA and TDMA methods. All nodes integrate sleep and wakeup tasks for energy consumption optimization. In each cluster, the election of a new CH node is computed by the CH of the previous round according to the energy criterion.

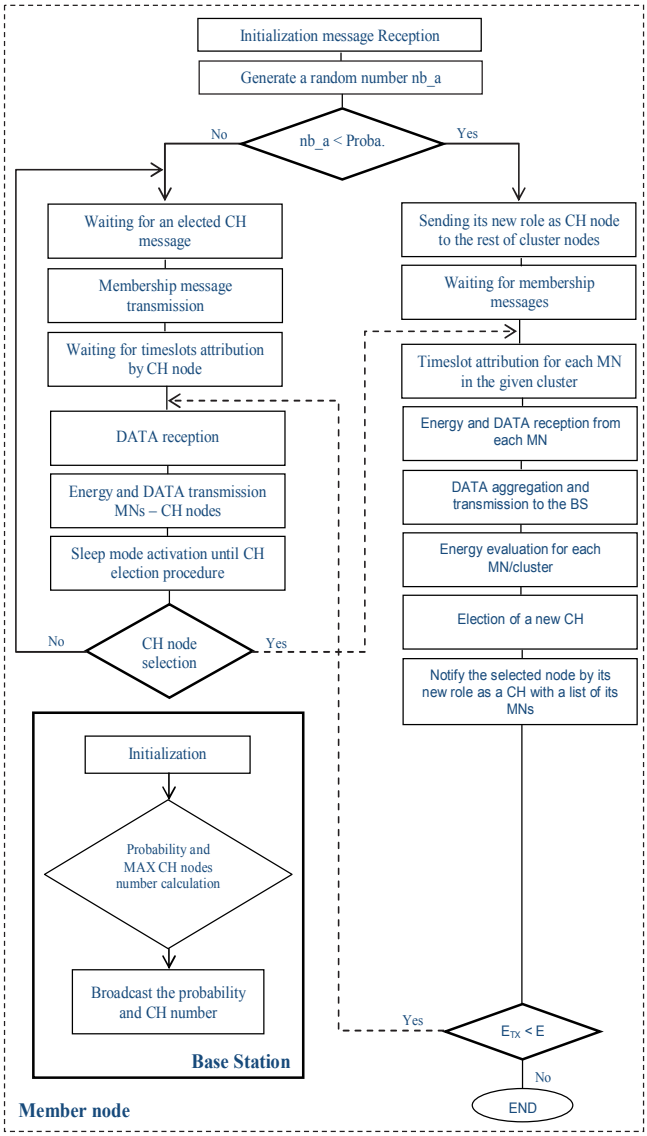


Figure 1. LEACH-Mod protocol algorithm

The energy consumption model which is implemented is described as follows [16]. For transmitting *k* bits of data over a distance d, the transmitter consumes:

$$E_{TX}(k,d) = k \cdot E_{elec}(k,d) + k \cdot \varepsilon_{amp} \cdot d^2 \qquad (5)$$

and the energy used to receive k-bits of data is:

$$E_{RX}(k,d) = E_{RX_{elec}}(k) = k \cdot E_{elec} \qquad (6)$$

With E_{elec} and E_{TX_amp} respectively transmit/receive electronic energy and amplification energy. ε_{amp} represents the amplification factor and $d_{crossover}$ the limited distance that the transmission factors change value.

In each cluster, the election of a new CH node is computed by the CH of the previous round according to the energy criterion.

At first, as in LEACH, and via a broadcast message, the base station determines the number of clusters and the probability P associated to different CH nodes. The protocol uses the formulation described in (1), and the number of CH nodes is around 10% of n. Each node receiving message from the base station must broadcast it into its neighbors. Each node must generate randomly a number nb_a that must be compared to the probability P determined by the base station. If ($P \leq nb\_a$) then the considered node can potentially be a CH, if not, it stands for other messages coming from other CH nodes. If the node is elected as CH, he informs all his neighbors of his new role by sending an ADV message containing its identifier. Based on the number of nodes in the cluster, the CH node creates a TDMA MAC protocol or schedule telling each node when it can transmit data by assigning a timeslot. This schedule is broadcast back to the nodes in the cluster. Each node decides to belong to a cluster by informing the associated CH that it will be a MN. It depends on the received signal strength emitted by the CH nodes. Once the clusters formed, each CH node will coordinate the transmission within its own group. Based on a scheduling task, it uses a TDMA method to assign for each NM a timeslot in which the node can transmit its information. Otherwise, NM activates its sleep mode to optimize its energy consumption. Different CH nodes will aggregate all the collected data and transmit them directly or indirectly to the base station according to a CSMA/CA method. Then, they proceed, in a new round, to elect a new CH node according to residual energy. New CH node informs the rest of nodes by his new role.

To evaluate performances of the proposed routing protocol LEACH-Mod, we have performed a comparative study with a standard routing protocol LEACH. The obtained results are analyzed and discussed in the next section.

III. Simulations Results and Analysis

In this section, we implement the proposed hierarchical routing protocol LEACH-Mod, and we compare the obtained results with a standard routing protocol LEACH. In this context, a simulation environment based TOSSIM platform with an object-oriented programming language NesC as a development tool [17]. The communication part is based on a scheduling policy of data via a radio propagation model called

LOSSY. Nodes forming WSN have same characteristics as MICA2 sensor model. They are randomly deployed in the operational environment. LEACH-Mod performances are evaluated using energy consumption criteria. A number of experiments were performed by taking into account the density of the network. We have proposed networks with 50, 100, 150 and 200 nodes. The simulation time is about 300 seconds.

Table 1 shows the obtained results concerning the average energy consumption in the network. We can see that the use of resources increases with the density of the network.

TABLE I. AVERAGE ENERGY CONSUMPTION

Number of nodes	Global energy consumption in the network (Joules)		Energy consumption by CH nodes and MNs (Joules)			
	LEACH	*LEACH-Mod*	*LEACH*		*LEACH-Mod*	
			CH	MNs	CH	MNs
50	17,6758163	17,2511633	22,3398	17,1458	22,3182	16,8007
100	17,964303	17,7721919	22,7770	17,3004	22,7323	17,2148
150	18,1195503	17,9554698	22,5922	17,4658	22,5361	17,2859
200	18,2804523	18,1163869	22,5605	17,5247	22,1254	17,5138

The obtained results show that the proposed routing protocol LEACH-Mod performs better than the standard routing protocol LEACH, as it's illustrated on Figure 2. With a network of 50 nodes, the average energy consumption is around 17.6758163 joules for LEACH and 17.2511633 for LEACH-Mod. With 200 nodes, it is around 18.2804523 joules for LEACH and 18.1163869 for LEACH-Mod.

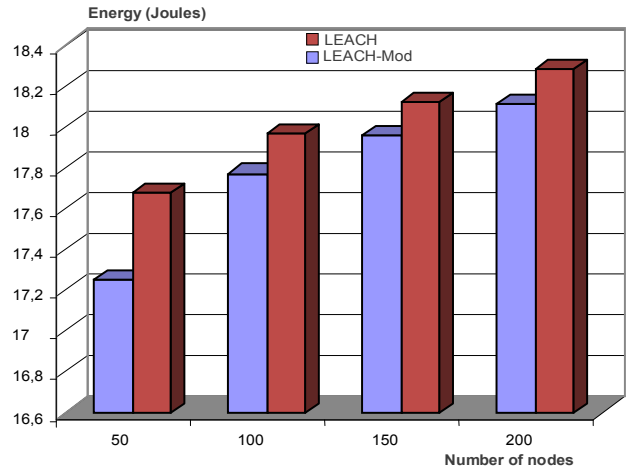


Figure 2. Average energy consumption in the network

Figures 3 and 4, illustrate respectively the quantities of energy consumption by CH nodes and MNs associated to each cluster. One can note that CH nodes activities are more important than those of MNs in the network. CH nodes provide both information exchanges in intra-CH mode, where communication is performed between CH and their MNs, and in inter-CH mode, where different CH nodes communicate with the base station. From the obtained results, considering a network with 100 nodes, in the case of the rooting protocol LEACH-Mod, the average energy consumption for CH nodes is approximately 22.7323 joules and 17.2148 joules for the MNs. Similarly, for a network with 200 nodes, the average energy consumption for CH nodes is approximately 22.1254 joules and 17.5138 joules for the MNs. Thus, the energy consumption is lower in LEACH-Mod than in standard routing protocol LEACH. This can be justified by the fact that the election of new CH nodes for each new round is directly ensured by the old CH nodes not by the base station. This procedure of election limits the number of control messages and the overloading of network.

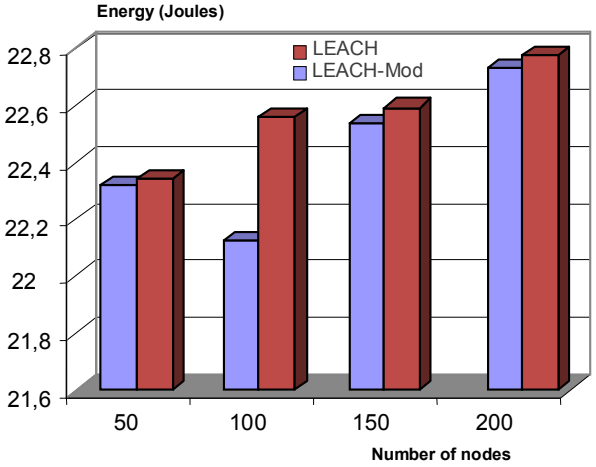


Figure 3. Average energy consumption by CHs nodes

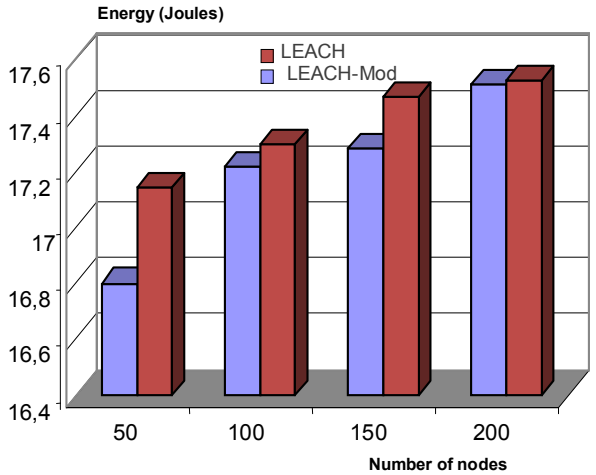


Figure 4. Average energy consumption by MNs nodes

The end-to-end delays required to route information correspond to the elapsed time for any data to reach its destination from any source. LEACH-Mod protocol includes two types of information exchange: intra-and inter-CHs. Based on some assumptions outlined above, where information exchange is insured in a single hop, data derived from MNs of each cluster moves directly to the corresponding CH-node. Thus, via an inter-CHs communication, aggregate information and transmit it to the base station. Thus, the more MNs increase, the more the time required for aggregation of data. The more the network is large the more end-to-end delay increases. As illustrated on Figure 5, the average end-to-end delay for routing information to a destination reflecting the reality. When implementing the LEACH-Mod routing protocol, the obtained results are better than those of LEACH standard routing protocol.

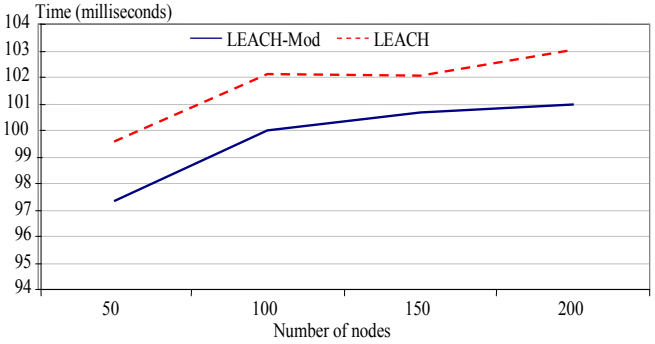


Figure 5. End-to-end delays

We can see that for a network with 100 nodes, end-to-end delays are around of 99,993 and 102,103 respectively for LEACH standard protocol and the proposed LEACH-Mod protocol. The same is true for a network with 200 nodes, where end-to-end delays are respectively 100,988 and 103,052 for LEACH standard protocol and the proposed LEACH-Mod protocol. The gain of elapsed time for any data to reach its destination varies between 2% to 5%.

To determine the network ability for processing information, we have computed the number of messages in term of DATA PACKETS exchanged between clusters and the base station. Figure 6.a shows the number of messages received by different CH-nodes during intra-CH communication phase. Those messages integrate information concerning environmental temperature. We can see that, contrary to LEACH standard routing protocol, in LEACH-Mod, CH nodes deal with a large number of messages whatever the density of the network. Indeed, for a chosen network with 100 nodes, applying LEACH and LEACH-Mod routing protocols respectively is 94 and 100. For a network with 200 nodes, it is 108 for LEACH standard protocol and 120 for LEACH-Mod. Once received and aggregated, all data are transmitted directly to the base station. Figure 6.d illustrates all messages supported by the network, including intra-CH messages and inter-CHs messages including those broadcasted by the base station during the announcement phase for selecting initial CH nodes.

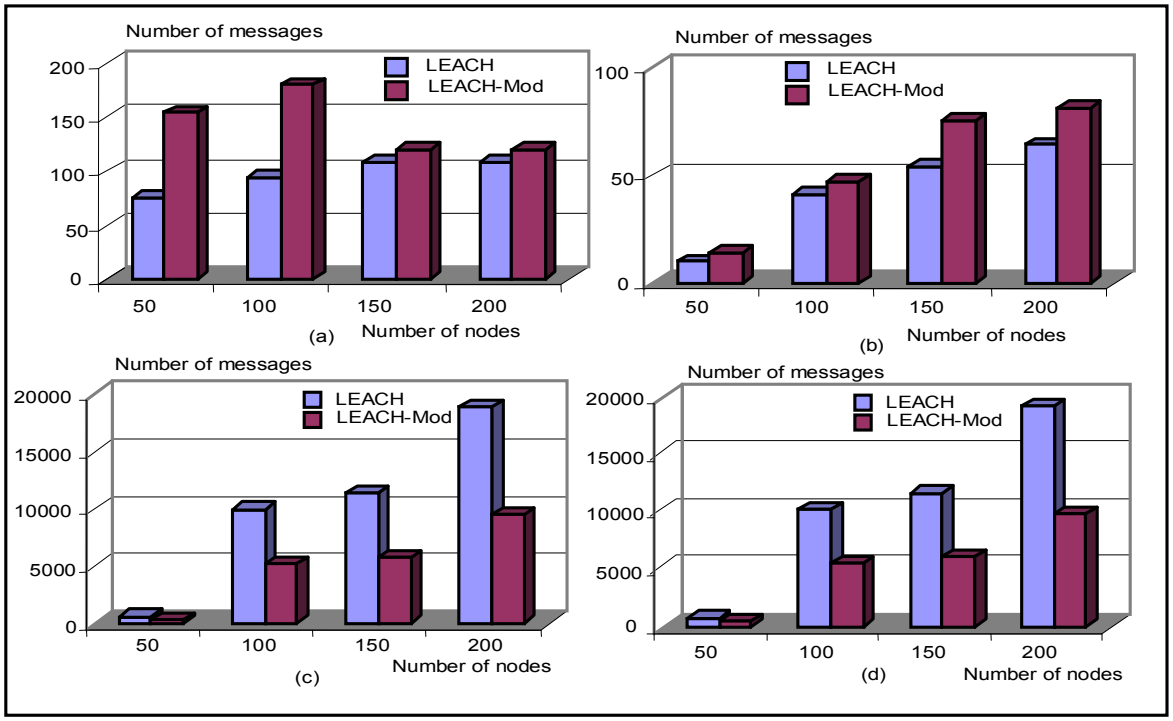


Figure 6. Traffic in the network

As shown on Figure 6.b, the number of messages sent by different CH-nodes to the base station increases with the density of the network. Indeed, unlike LEACH routing protocol, Inter-CHs messages are more important in the case of LEACH-Mod protocol. For example, using a network with 100 nodes, the number of messages for LEACH and LEACH-Mod

is respectively 41 and 57. For a network with 200 nodes, it's respectively 64 and 81 messages. This means that during inter-CHs transmission, CH-nodes have sufficient residual energy and computing time to achieve aggregation process of collected data from MNs, and then to transmit them to the base station. Unlike inter-CHs communication, intra-CH communication concerns two phases of communication: communication between CH-nodes and all nodes for establishing their membership to a cluster and communication between CH-nodes and MNs. As illustrated on Figure 6.c, in both cases, applying LEACH and LEACH-Mod routing protocols, the number of messages exchanged in intra-CH communication increases according to the network density, but less important for LEACH-Mod protocol. In a network with 100 nodes, the number of exchanging messages between CH-nodes and MNs is 10273 for LEACH protocol and 5540 for LEACH-Mod. For another network of 200 nodes, it's 19312 for LEACH protocol and 9934 for LEACH-Mod. Thus, comparing these results to those shown in Figures 3 and 4, we can conclude that this phase of communication corresponds to the phase where the network consumes more energy.

Figure 6.d illustrates all messages supported by the network, including intra-CH messages and inter-CHs messages including those broadcasted by the base station during the announcement phase for selecting initial CH nodes.

IV. Conclusion and Perspectives

In this paper, we are interested by WSN routing protocol and treating particularly the problem of resources optimization allowing the increase of network lifetime. For this purpose, we have proposed and studied hierarchical routing protocol called LEACH-Mod based on clustering methodology and set of assumptions improving network performances in terms of energy consumption, robustness, efficiency and reactivity. In contrast to standard routing protocol LEACH, where nodes select their membership to a given CHs nodes more than once, LEACH-Mod selects the CH node having the highest energy by sending a membership message to become a NM. This election procedure means that the election of new CHs nodes for each new round is directly ensured by the old CHs nodes but not by the base station. Thus, the number of control messages and overloading of network are limited. In order to test the effectiveness of the proposed routing protocol LEACH-Mod, we have conducted a comparative study with a standard routing protocol LEACH. The obtained results show that the proposed approach gives better results than LEACH.

Actually, our research focuses on fuzzy logic approach implementation integrating learning concept to improve clustering performances. We highlight another concept of inclusion surface between membership functions. The concept combines both distances between nodes and CH nodes, and energy consumption related to each considered node.

References

[1] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. Wireless sensor networks: A survey. *Computer Networks*, vol.38, n°4, 2002, pp.393-422.

[2] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. 2002. A survey on sensor networks. *IEEE Communications Magazine*. 40, 8 (2002), 102-114.

[3] Chong, C.Y. and Kumar, S.P. 2003. Sensor Networks: Evolution, Opportunities, and Challenges. *Proceedings of the IEEE*, (August 2003), 1247-1256.

[4] Culler, D., Estrin, D. and Strivastava, M. 2004. Overview of Sensor Networks. *IEEE Computer Society*, 37, 8 (2004), 41-49.

[5] Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. 2002. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. *In Proceedings of the 33rd Hawaii International Conference on System Sciences*, 2 (2000), 3005-3014.

[6] Akkaya, K. and Younis, M. 2005. A survey on routing protocols for wireless sensor networks. *Journal of Adhoc Networks*, 3, 3 (May 2005), 325-349.

[7] Tang, Y., Zhou, M.T. and Zhang X. 2006. Overview of Routing Protocols in Wireless Sensor Networks. *Journal of Software*, 7, 3 (March 2006), 410-421.

[8] Fan, X. and Son, Y. 2007. Improvement on LEACH Protocol of Wireless Sensor Network. *International Conference on Sensor Technologies and Applications*, Valencia, Spain, (2007), 260-264.

[9] Bhattacharyya, D., Kim, T. and Pal, S. 2010. A Comparative Study of Wireless Sensor Networks and Their Routing Protocols. *Sensors 2010*, 10, 12 (2010), 10506-10523.

[10] Bani-Yassein, M., Al-zou'bi, A., Khamayseh, Y. and Mardini, W. 2009. Improvement on LEACH Protocol of Wireless Sensor Network (VLEACH). *International Journal of Digital Content Technology and its Applications*, 3, 2 (2009), 132-136.

[11] Younis, O. and Fahmy, S. 2004. HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks. *Proceedings of the IEEE Transactions on Mobile Computing*, 3 (2004), 366-379.

[12] Lindsey, S. and Raghavendra, C. PEGASIS: Power-Efficient Gathering in Sensor Information Systems. *Proceedings of the IEEE Aerospace Conference*, 3 (March 2002), Big Sky, MT, USA, 1125-1130.

[13] Manjeshwar, A. and Agrawal, D.P. 2001. TEEN: A protocol for enhanced efficiency in wireless sensor networks. *In the proceedings of the 15th International Symposium on Parallel and Distributed Processing*, (April 2001), 2009-2015.

[14] Subramanian, N.V. 2004. Survey on energy-aware routing and routing protocols for sensor networks. Technical Report, Computer Science, University of North Carolina, Charlotte.

[15] Chen, H. and Megerian, S. 2006. Cluster sizing and head selection for efficient data aggregation and routing in sensor networks. *IEEE Conference on Wireless Communications and Networking Conference*, WCNC, Las Vegas, NV, (2006), 2318-2323.

[16] Dehni. L, Krief. F. and Bennani, Y. 2006. Power Control and clustering in Wireless Sensor Networks. *Challenges in Ad Hoc Networking, IFIP International Federation for Information Processing*, 197, (2006), 31-40.

[17] Levis, P., Lee, N., Welsh, M. and Culler, D. 2003. TOSSIM: Accurate and scalable simulation of entire TinyOS applications. *Proceedings of the 1st ACM Conference on Embedded Networked Sensor Systems*, New York, NY, USA, (2003), 126-137.

Performance of the Energy Aware Routing Protocol in Wireless Sensor Networks with mobile nodes

Fote Kamanyi Antonia[1], **John Ngubiri**[2], **and Julianne Sansa-Otim**[1]
[1]Department of Networks, [2]Department of Computer Science (Corresponding Author),
Makerere University College of Computing and Information Sciences
antoniafote@gmail.com, {ngubiri, sansa}@cit.mak.ac.ug

Abstract— *Wireless Sensor Networks (WSNs) are increasing in popularity. This is due to several applications (like car tracking, acute patient monitoring and forest fire detection). Energy in WSNs is a scarce resource and therefore has to be optimized. Several studies on energy aware routing schemes have been made. However, most of them cater for fixed nodes yet in some cases, some WSN nodes are mobile. In this paper, we extend the Energy Aware Routing Protocol (EARP) [1] to cater for WSNs with some mobile nodes. We propose EARP with Mobility Support (EARP-MS) and evaluate its performance. We show that (i) the energy consumed increases as the distance between the source and recipient nodes increase, (ii) providing for node mobility prolongs the WSN lifespan, (iii) mobile nodes have higher residual energy than the static nodes and (iv) the average transmission time is lower when some nodes are mobile..*

Keywords: WSN, mobile nodes, routing, EARP, EARP-NS

1. Introduction

WSNs are used in a variety of applications. These include forest fire detection, acute patient monitoring and target tracking. In general, when a node senses a change in environment, it conveys the message through the network to the base station. Along the way, some processing may take place on the nodes. Nodes, therefore, incur energy as they route the data through the WSN. It is rare that after some time of operation, the energy in all nodes in a WSN is the same. This is because the amount of the energy consumed differs among the nodes. This is dictated by the role and how active a node is. Energy consumed by a node gathering data differ from that consumed by a node transmitting data. Likewise, energy consumed by the cluster head differs from the rest. Even among nodes of similar roles, energy consumption is bound to be different. For example transmitting nodes far away from the cluster heads tend to lose more energy. Further more, redeployment of nodes for prolonging network lifetime also causes residual energy not to be equal among sensor nodes. Overall, Routing of the data from the source to the cluster head comes at a cost. It is the main determinant of the energy levels in a WSN. It is of paramount importance that all nodes have sufficient energy as total draining of some nodes' energy my mean that it is impossible for data to take a certain route. A lot of research on improving WSN lifespan has been carried out [1], [2], [3], [4].

Ming et al [1] analyzed three algorithms, LEACH [2], HEED [3] and EECS [1]. They proposed the EARP that addresses most of the weaknesses of LEACH, HEED and EECS. They showed that indeed EARP is superior compared to the three. Much as EARP is effective, it assumes (i) all nodes are static and (ii) each node captures data from environment. This is not always true. There are cases where some nodes are mobile and data comes from mobile nodes (e.g. car surveillance) which cannot be addressed by EARP.

In this paper, we extend the EARP to cater for mobile nodes as well as cases where not all the nodes can collect data. We improve the algorithm to cater for the above scenario and also test the effectiveness of the improved algorithm.

The rest of the paper is organized as follows, In Section 2 we discuss earlier research related to this work which also form a basis for it. In Section 3, we discuss the improved EARP that caters for mobile nodes. We then present and discuss our experimental results in Section 4 and conclude the paper and propose future research in Section 5.

2. Related work

Substantial work has been done on energy aware protocols in WSNs. The most recent and popular work include LEACH [2], HEED [3], EECS [1], PEGASIS [4], M-LEACH [5] and EARP [1].

2.1 LEACH

The Low-Energy Adaptive Clustering Hierarchy (LEACH) was proposed by Heinzelman et al [2]. It assumes a unique base station outside the sensor network and that all the sensor nodes are able to directly communicate with the base station. In order to save energy, LEACH chooses a fraction p of the sensor nodes to serve as cluster heads.

p is determined before deployment. The rest of the sensor nodes join the appropriate clusters determined by the signal strength from cluster heads. LEACH divides its operation into rounds which guarantee cluster head rotation in each round. In each round, after cluster formation phase, the cluster heads aggregates the data received from their cluster members and send the aggregated data to the base station by single hop communication. This reduces the data that needs to be transmitted to the base station and save energy.

The typical many-to-one traffic pattern in sensor networks causes uneven energy consumption (and hence levels) among nodes. Nodes near a sink or cluster head have much heavier traffic and hence more power consumption compared to other nodes. In order to deal with the heterogeneous energy distribution among the nodes, Heinzelman et al [2] proposed that the node with the higher energy should have a higher probability of becoming a cluster head.

2.2 HEED

Younis et al [3] proposed the Hybrid Energy-Efficient Distributed clustering algorithm(HEED). HEED periodically selects the cluster head using a hybrid of node residual energy and a secondary parameter (like node degree or proximity to neighbors). It incurs low message overhead and achieves fairly uniform cluster head distribution across the network.

2.3 EECS

The Energy Efficient Clustering Scheme (EECS) chooses cluster heads based on nodes with more residual energy. Ming et al [1] argued that setting the residual energy as the primary parameter for cluster heads election does not help balance the energy load for the proper nodes. This is mostly in heavy energy heterogeneous circumstances. In most local clustering algorithms in WSNs, as a means of prolonging the sensor network lifetime, the probability of a sensor node being selected as a cluster head primarily only depended on its own residual energy. EECS does not help balance the energy load for the proper nodes in some special cases and this could lead to the problem that some nodes will be exhausted quickly.

2.4 EARP

Ming et al in [1] analyzed weaknesses in LEACH, HEED and EECS and proposed the Energy Aware Routing Protocol (EARP) to address them. In order to minimize energy consumption in each round, Ming et al [1] argued that the nodes should join the nearest head. The cluster heads always rotate through out the lifespan of the network in order to balance the energy among nodes. EARP bases cluster head selection on the residual energy of the candidate and average residual energy of all the nodes in the cluster.

Algorithm 1 EARP Algorithm

```
if state ← candidate then
  broadcast and receive Node_ResidualMsg from all neighbor nodes
  update neighbourhood table NT[] and t ← computational result of the broadcast delay time for a competing cluster head
  while timer for cluster head election not expired do
    if CurrentTime ≪ t then
      Compete_Msg is overheard from a neighbor NT[i] and state ← plain
      nt[I].state = head
      continue
      if state = candidate then
        state ← head
        broadcast Comepete_Msg
        wait (2Δt)
        if have not received any Compete_Msg then
          continue
        else
          if the weight of the head election is the largest one then
            continue
          else
            plain
            if the value in weight broadcasted by NT[i] is the largest one then
              NT[i].state = head
            end if
          end if
        end if
      end if
    end if
  end while
end if
```

Overall, EARP addresses the weaknesses of its predecessors. LEACH [2] is not efficient because the cluster head broadcast to the whole network in cluster formation. HEED [3] is not efficient because it produces multiple broadcasts during cluster formation. EECS [1] is not efficient because of its unbalanced energy load in heavy energy heterogeneous circumstances. EARP addresses these weaknesses. However, it also assumes static nodes each of which captures data from environment. Cases of mobile nodes are not addressed by EARP.

Table 1 gives a summary of a comparison of LEACH, HEED, EECS and EARP with regard to the requirements for clustering protocols.

Table 1
A SUMMARISED COMPARISON OF LEACH, HEED, EECS AND EARP

Requirement	LEACH	HEED	EECS	EARP
Distributed clustering algorithm	Yes	Yes	Yes	Yes
CH distribution	Yes (optimal CH predetermined)	Yes	Yes	Yes
CH election procedure	A fraction p of the SN is chosen to server as CHs	Selected based on residual energy & a secondary parameter	Selection based on residual energy	Selection based on residual energy & average energy
clustering algorithm efficiency	No (CH broadcast to the whole SN in cluster formation)	No (Multiple broadcasts during cluster formation)	No (Unbalanced energy load in energy heterogenous circumstances)	Yes
Energy Heterogeneity	No	No	No	Yes

3. EARP for mobile nodes

We now present the generation of the EARP-MS algorithm and the changes made to EARP to cater for mobile nodes in WSN. We also present the extended radio communication and residual energy expressions for EARP-MS. The generic assumption made is that some nodes move while others are static. The nodes that move sense the data from the environment and transmit to the base station though the static nodes.

3.1 Modifications in EARP

In order to make the EARP-MS, we make some modifications in the EARP. The following are the modifications made

1) States:
   In EARP a node is either a plain cluster member or a CH. In EARP-MS plain cluster members can be mobile or static. The static nodes can have some time intervals δt when they are not active.
2) Cluster head role:
   In EARP the CH role can be rotated among all nodes. However, in EARP-MS It can only be rotated among static nodes. This is so because mobile CHs would need to continuously update the clusters.
3) Neighborhood table (NT[i]):
   In EARP NT[i] is constantly being updated with messages of residual energy of neighbors from which the CH is to be chosen. In EARP-MS the NT[i] is only updated with residual energy of static nodes.
4) Sending of aggregated data:
   In EARP the CH sends the aggregated data to the next hop or base station. In EARP-MS, the CH sends the aggregated data directly to the base station.
5) Generation of data:
   In EARP any node generates data. In EARP-MS only mobile nodes generate data.

The changes made to the EARP to obtain EARP-MS caters for the mobile nodes and also contribute to energy efficiency in the network. If a node is moving towards the CH, the transmission range is reduced and therefore there is less transmition energy. Since there many CHs, it is possible that in most of the directions it moves, it is moving closer to a certain CH. Moving nodes, therefore, in many cases lead to energy conservation.

3.2 EARP-MS algorithm

We now present the way the improved algorithm address the operations in the WSN.

3.2.1 Identification of CH

In cluster head identification, only static nodes are considered. The nodes compete from among themselves and the most appropriate node is selected. The process resumes when the time for a node to be a cluster head expires and it has to be done again.

The cluster head got performs this role for a predetermined time t when the process is doen again and the new cluster head is got.

Algorithm 2 GetClusterHead()

C = set of all static nodes in cluster
$\forall n_i \in C$, broadcast residual energy to all members
update NT [i]
head = null
$Res_{max} = 0$
for $i == 0; i \leq n(C); i++$ **do**
  res = Residual Energy of n_i
  if $res \geq Res_{max}$ **then**
    head = n_i
  end if
end for
cluster head = head

3.2.2 Data Transmission

In data transmission, EARP-MS takes into consideration the fact that some of the nodes are moving while others are static. Therefore, the neigbouriness of the nodes are not fixed throught the lifetime of the WSN. A transmitting node can be able to sleep for a short time as it wait for a moving node. At the same time, it can be able to route through the static nodes.

Algorithm 3 TransmitData()

C' = set of all mobile nodes in a cluster
while C' $\neq \phi$ **do**
  for all $n_i \in C'$ with data to transmit **do**
    check if the CH is near
    if cluster head is near **then**
      send data directly to cluster head
    else
      check if there is a neighbouring static node
      if there is a neignbouting static node **then**
        transmit to the neighbouring node
      else
        sleep(δt)
        break
      end if
    end if
  end for
end while

3.2.3 Optimization

EARP-MS also makes an effort to optimize the energy in the nodes by ensuring that the same set of data is not sent several times (hence consuming energy unnecesarily).

Algorithm 4 Optimize()

C'' = set of mobile nodes ready to transmit
if subset of C'' has the same data **then**
  identify one node close to CH
  identified node send data to CH
  others sleep for duration of transmission
end if

3.3 Radio communication model for EARP-MS

Heinzelman et al [6] proposed a radio model to calculate the energy consumption in a WSN. During transmission, the energy consumed is given by $E_{TX}(k,d) = E_{elec}k + E_{amp}kd^2$ and during reception the energy used is given by $E_{RX}(k) = E_{elec}k$ where $E_{elec} = 50nj/bit$ and $E_{amp} = 100pj/bit/m^2$. This model has been widely adapted in several studies [1], [7], [8], [9]. We use the same model but extend it to cater for WSNs with mobile nodes.

During transmission, the energy consumed is

$$E_{TX}(k, d(t)) = E_{elec}k + E_{amp}kd(t)^2 \quad (1)$$

where $d(t)$ is the mobile sensor's distance from the cluster head at time t.

During target detection, the energy consumed is given by

$$E_i^{Tg}(k, dist(i,j)) = E_{amp}k(l^2 + d(t)^2) + E_{elec}k \quad (2)$$

where $A(X_j, Y_j)$ = the detected position of the target on the field A and $B(X_i, Y_i)$ is the position of the mobile node at a particular time and $l^2 = (x_i - x_j)^2 + (y_i - y_j)^2$.

During idling, the energy consumed is given by

$$E_{IX}(k) = R_x k \quad (3)$$

where R_x is the receiver which is on and idle.

During Processing, the energy consumed is given by

$$E_{PX}(k) = (E_{elec}k) + (E_{elec}k^{ag}) \quad (4)$$

where $E_{elec}k^{ag}$ is the energy consumed by the CH during data (k) aggregation and $E_{elec}k$ is the energy consumed when the CH receive data for aggregation.

The residual energy is therefore given by

$$E_i^t = E_0 - (E_{TX}(k,d) + E_{RX}(k) + E_{IX}(k)) \quad (5)$$

where E_0 is the initial energy level of the sensor, $E_{TX}(k,d)$ is energy consumption during transmission, $E_{RX}(k)$ is energy consumption during reception and $E_{IX}(k)$ is energy consumption during idling. In case a certain node is dynamic, then d is a function of time.

4. Experimental Results

We simulated the WSN model described and investigated the performance of the proposed algorithm.

4.1 Simulation setup

We adopted the setup values that Ming et al [1] used except for the values of speed and the distance range since their network consisted of static nodes. Our WSN has some mobile nodes and therefore the mobile nodes distance changes with time and are moving (at a constant speed) which is not the case in a static network. The initial instances of the parameters used are summarized in Table 2.

parameter	value
initial energy (E_0)	$2 * 10^9$ nJ
data packet size (k)	4200bits
broadcast packet	200bits
distances (d)	$1m \sim 70m$
Constant speed of mobile node	10m/s
transm Elecs=recv Elec (E_{elec})	50nj/bit
transmitter amplifier (E_{amp})	100pj/bit/m^2

Table 2
INITIAL PARAMETER INSTANCES

4.2 Residual energy during data transmission

We simulated the WSN considering only the transmission state of the node to calculate the residual energy. We studied the variation of the residual energy with different cluster sizes and summarize the trends of the residual energy in Figure 1.

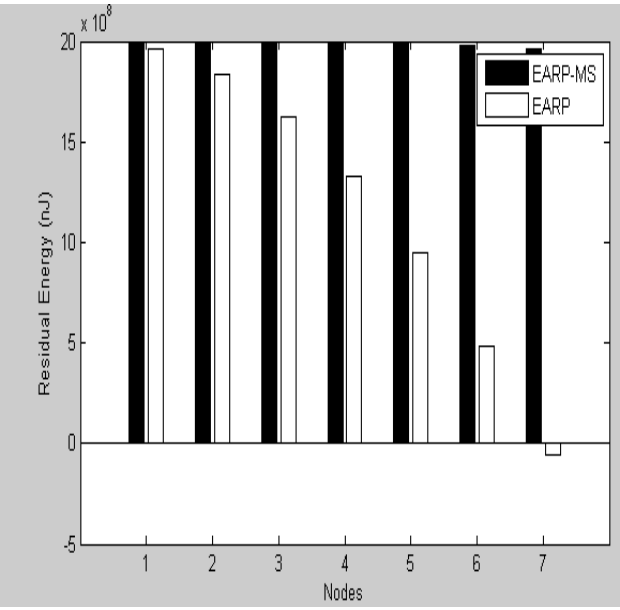


Fig. 1
DIFFERENCE IN RESIDUAL ENERGY

From Figure 1, we observe that there is (i) a general reduction in energy with both algorithms as the network size is increased (ii) for the same network size, residual energy of EARP is less than that of EARP-MS and (iii) the rate of reduction of residual energy for EARP is higher than that of EAPR-MS.

This trend can be explained by the fact that since in EARP-MS the nodes move, the overall (average) distance between them and the CH is reduced. This is due to the fact that since CHs are distributed in the network fairly evenly, a moving node will, in most cases, be moving closer to a certain CH. This reduces the distance data has to move to the CH and hence the transmission energy. This in effect keeps the residual energy high. EARP and EARP-MS are both energy efficient clustering protocols for WSN, but EARP-MS uses the mobile nodes to its advantage which makes it more efficient.

The residual energy trends imply that a network with mobile nodes is bound to be more long lasting since the residual energy is lost at a lower rate. Likewise, it is sustainable in case the network has to be big since the residual energy does not reduce considerably with increasing network size. This implies that EARP-MS is suitable for big WSNs which are typically employed in real life.

4.3 The effect of the distance of static nodes

We also investigated the effect of the distance of a static node from the CH. we varied the distance between $10m$ and $70m$ and the results are summarized in Figure 2.

From Figure 2, we observe that an increment in distance leads to a reduction in residual energy. We observe that at a distance of about $70m$, the residual energy falls to zero. This implies that the WSN would die out.

This can be explained by the fact that the further the transmitting node is to the receiving node (CH), more energy is consumed and therefore less residual energy for the transmitting node.

This therefore implies that to ensure longevity of the WSN, it is neccesary for the distance between nodes to be kept low enough. This implies that if the geographical area occupied by the WSN has to be maintained, the number of nodes in the network has to be increased.

4.4 Average energy consumption comparison between EARP and EARP-MS

We compute the average energy in the network. We used initial parameters in table 2. The average energy consumption for EARP was found to be $8.4021 * 10^8 nJ$ while that of EARP-MS was $0.8742 * 10^8 nJ$. This shows that there is more energy consumed in EARP compared to a case of EARP-MS. This shows that EARP-MS is more energy efficient compared to EARP.

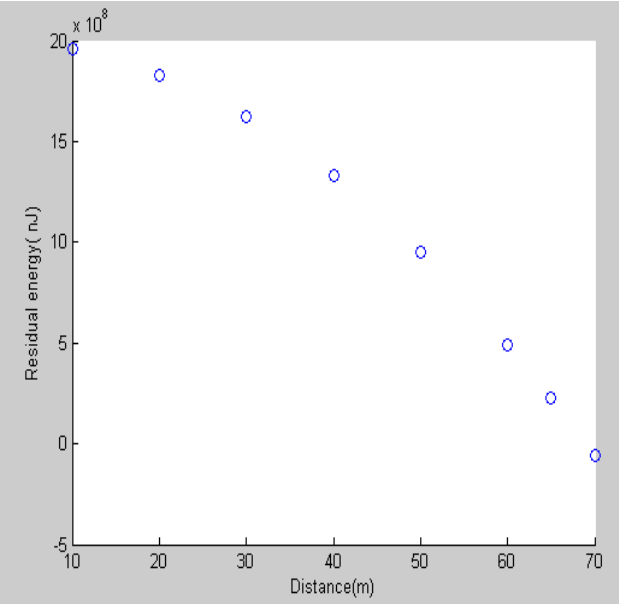


Fig. 2
RESIDUAL ENERGY VERSES DIFFERENT DISTANCES TO THE CH

5. Conclusion and Future work

In this paper we have modified EARP into EARP-MS and extend the functionalities of EARP so that it can be able to cater for mobile nodes. We have made preliminary tests on the performance of EARP-MS and showed that it comes with improved WSN lifespan. The increase in lifespan is largely caused by the reduction in overall distances of data transmission created by moving collector nodes.

Further work can be done by looking at other performance metrics (like average delay and data delivery ratio), making a deeper investigation in heterogeinity of the nodes as well as making analytical solutions that can give indicative mathematical bounds.

References

[1] M. Liu, J. Cao, G. Chen and X. Wang, "An Energy-Aware Routing Protocol in Wireless Sensor Networks",*in Proceedings from Sensors*, vol 9, p.445-462, 2009.

[2] Heinzelman, W.R., "An Application -Specific Protocol Architecture for Wireless Microsensor Networks", *IEEE Trans. Wireless Commun*, vol 1, p.600-670, 2002

[3] Younis, O.; Fahmy and S. Heed, " A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks", *IEEE Trans. Mob. Comput*, vol 3, p.366-379, 2004.

[4] Lindsey, S.; Raghavendra and C. PEGASIS, "Power-Efficient gathering in sensor information systems", *In Proceeding of IEEE Aerospace Conference*, vol 3, 1125-1130,2002.

[5] Mhatre, V., Rosenberg, C., " Homogeneous vs. Heterogeneous Clustered Networks: A Comparative Study", *In Proceedings of IEEE ICC 2004*, June 2004.

[6] Heinzelman, W.R. Chandrakasan, A.;, Balakrishnan and "Energy-efficient communication protocol for wireless microsensor networks", *In Proceedings of the Hawaii International Conference on System Sciences, Maui* , Hawaii, Jan 2000

[7] Lee, S.H., Yoo, J.J., Chung and T.C., "Distance-based Energy Efficient Clustering for Wireless Sensor Networks", *In Proceedings of the 29th Annual IEEE international Conference on Local Computer Networks (LCNâĂŹ04)*, 2004.

[8] Chan, H.; Perrig and A. ACE, " An Emergent Algorithm for Highly Uniform Cluster Formation", *In Proceedings of the First European Workshop on Sensor Networks (EWSN)* 2004.

[9] Ye, M.; Li, C.F.; Chen, G.; and Wu, J. EECS " An Energy Efficient Clustering Scheme in Wireless Sensor Networks", *Int. J. Ad Hoc Sensor. Network*, vol 3, p.99-119, 2007.

Operators Intervention Strategies with New MCDM for Load Balancing in Heterogeneous Mobile Networks

Bassem El Zant

Department of Electricity and Mechanics, Saint-Joseph University, Faculty of Engineering – ESIB
Campus of Sciences and Technology, Mar Roukos, Mkalles, Beirut, Lebanon
bassem.elzant@ usj.edu.lb

Abstract. *Wireless networks are designed to operate independently. With the many existing inconveniences of 3G networks, operators are persuaded to switch to heterogeneous access networks. This will mainly reduce the costs that operators should pay in order to move to 4G; and thus make use of the existing networks infrastructures, to come up with heterogeneous environment where multiple wireless technologies coexist. In this paper, three strategies along with the operators' intervention are programmed: binary, binary with return to initial condition, and fuzzy. The reason behind this is to select the best strategy (least drop probability, least delay, best performance parameters, etc), which is the fuzzy one; whereby regular filling of the different networks is provided, and thus network fluctuation is well illustrated. As for the worst strategy, it is the binary with return to initial condition; whereby operators will have less resources to exploit, and so less throughput and higher delay.*

Keywords: Multi-criteria decision method (MCDM), Heterogeneous Mobile Networks, Satisfaction-based decision method, Binary and Fuzzy logic.

1. INTRODUCTION

Wireless networks are designed to operate independently. With the explosion of traffic, telecom operators are confronted with the problem of mobile infrastructure saturation. The use of new technologies (Wireless MAN-Advanced, development of WiMAX), and LTE-Advanced) enables the operators to integrate different radio technologies already deployed; such as pooling of resources of WiFi, mobile WiMAX, the LTE and HSPA +. Besides, taking into consideration the 3 G disadvantages (whereby the deployment has proven to be costly), and the higher costs to be paid in order to migrate to 4G networks or to increase the density of the existing 3G networks, operators are increasingly convinced by deploying heterogeneous access networks. In the heterogeneous networks, multiple wireless technologies coexist, and the radio resource management is coordinated. In such networks, mobile users can connect to different radio access technologies. To optimize the system performance, network operators aim to balance loads, as much as possible, in its various radio access networks.

In heterogeneous wireless networks, the main challenge is to keep connections among the different networks like WiFi, WiMax, WLAN etc. The 4th generation of wireless (NGWN/4 G) networks is expected to present heterogeneity in terms of wireless technologies and services. The mainly advantage of the mobile networks 3G (UMTS and 1xEV-/ DV) is their global coverage. However, the weaknesses of 3G lie in their bandwidth capacity and operating costs. On the other hand, the WLAN technology such as IEEE 802.11 offers higher bandwidth with low operating costs, although it covers a relatively short range. In addition, technological advances in the evolution of mobile devices made possible the support of different Radio Access Technologies. This raised much interest for integration and interoperability of 3 G wireless networks and wireless local networks, to take advantage of their respective potentials. In this paper, we will start by defining the heterogeneous mobile networks and their benefits before presenting different intervention strategies.

2. HETEROGENEOUS MOBILE NETWORKS

Heterogeneous networks integrate different radio technologies that have been already deployed to combat the problem of network saturation, and to share various resources of the operators. Two methods could be used to access the heterogeneous networks: the loose coupling, and the tight coupling. Besides, there are two entities responsible for the management of the radio resources: the RRM for the local processing of resources, and the CRRM for common resource management.

3. PROBLEM

Network selection is made by the user as per the provided criteria: the throughput and the cost. However, to optimize the overall performance, networks operators should intervene. They play on guarantees of QoS (the offered throughput), and economic incentives (the cost). This will guide the final network selection of the user the way system performance will be the best. To alleviate networks, the operators offer a lower throughput or a higher cost. However, to attract new arrivals, they provide higher throughput, or lower cost. Different strategies are described to change the throughput and the cost.

4. EXISTING METHODS AND RELATED WORK

Different methods exist in the literature to improve the quality of services and try to provide solution to the above problems.

In [1] we can see how the handover technique is used to redirect the mobile user's service network from current network to a new network or one Base Station (BS) to another one or from one Access Point (AP) to another one using the same or different technologies in order to reduce the processing delay in the overlapping area. Handover network type [11] has horizontal and vertical handover. The homogenous wireless network performs horizontal handover, if there are two BSs using the same access technology. This type of mechanism use signal strength measurements for surrounding BSs to trigger and to perform the handover decision. [2] describes the concept of being always best connected describe the user experience and business relationship in an ABC environment and outline different aspects of an ABC solution that broad the technology and business base of 3G. It shows how in heterogeneous wireless networks environment, Always Best Connected (ABC) requires dynamic selection of the best network and access technologies when multiple options are available simultaneously. The Mobile Station (MS) or BS should be equipped with multiple network interfaces to reach different wireless network. The authors in [4] have done a comparison among SAW, Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Grey Relational Analysis (GRA) and Multiplicative Exponent Weighting (MEW) for vertical handoff decision. The vertical handoff decision algorithm for heterogeneous wireless network has been discussed in [3]. The author formulated the problem as Markov decision process. And the vertical handoff decision is formulated as fuzzy multiple attribute decision making (MADM) in [5]. A vertical handoff decision scheme DVHD uses the MADM method to avoid the processing delay in [6], when the goal in [7] of the authors was to reduce the overload and the processing delay in the mobile terminal. A novel vertical handoff decision scheme to avoid the processing delay and power consumption has been proposed in this paper. A novel distributed vertical handoff decision scheme using the SAW method with a distributed manner to avoid the drawbacks has been suggested in [9] when the authors in [9] define using the emerging IEEE 802.21 standard a Media Independent Handover (MIH) functions as transport service in order to offer a vertical handoff decision with a minimum of processing delay. A four-step integrated strategy for MADM-based network selection has been proposed in [10]. In [12] a comparative analysis of MADM methods including SAW, MEW, TOPSIS, ELECTRE, VIKOR, GRA, and WMC is illustrated with a numerical simulation, showing their performance for different applications such as: voice and data connections, in a 4G wireless system.

5. KEY PERFORMANCE INDICATORS (KPIS)

Different KPIs are defined to assess the user satisfaction and operators gain. To assess the satisfaction of the user, we take into consideration the applications' types. So, for the streaming and inelastic applications, we consider the throughput, the delay and the drop probability. And for the elastic applications, we consider the throughput and performance test d/ d_c where d is the effective flow received per the user and d_c represents the flow of comfort. It then sets the user's satisfaction by:

$$S = w_{QoS} * S_{QoS} + w_{cost} * S_{cost} \qquad (1)$$

Where w_{QoS} and w_{cost} represent the respective weights assigned to the QoS criteria and monetary cost, and S_{QoS} and S_{cost} the respective satisfaction of the user. The functions of satisfaction are identical to those defined for the calculation of the best alternative and satisfaction will be measured at the actual rate received by the user. To assess the operator gain we calculate the average revenue per user for consumption per Kbyte.

6. OPERATOR STRATEGIES

The operator strategies are implicitly embedded in the system's information. This will influence the decision process. The operator will inform the user of the minimum and maximum throughput of each network access with the corresponding monetary cost and the user is who choose the most suitable access technologies among different ones. We can say that the operator will affect and guide the user when making the final decision. In our work, the monetary cost will be fixed when the QoS incentives will vary dynamically, and vice versa to optimize short and long term goals. In both cases the monetary cost and the different incentives for QoS will reflect the operator strategies and contribute in the final decision of the user. The operator will hide the total capacity of the system and will only provide to the users information about the incentives of QoS and presents the different guaranteed throughputs offered with the cost. On the other hand, when the operator is ready to reserve, in a RAT j, a band for the session of the service class i, excellent throughput will be suggested, and then the class i will attract new coming sessions to the RAT j. So to avoid new sessions, the operator can offer excellent throughput or low cost, which will push the user to pass to one RAT. In conclusion, the dynamic incentives of QoS variation or cost variation will allow the operator to more or less attract new users to a class in a specific RAT and then the operator will contribute in the final decision of the user.

7. NETWORK SELECTION PROCESS

Before opening a new session or to make a handover, the mobile must evaluate different alternatives and select the best network and the best class of service. For this multi-criteria decision, we should consider the QoS requirements, radio conditions, the cost and the preferences of the user. The different decisions attributes are as following: the minimum rate guarantees (d_{min}), the maximum rate (d_{max}) and cost (C). Figure 1 shows the hierarchical representation of the criteria.

During the first stage, the general criteria must take into consideration the user's preferences. The user may prefer

paying more for a better quality of service or he prefers to save money and get low quality. The user therefore assigns weights to QoS incentives and the cost according to his/her preference. However the secondary criteria depend on the type of application, for example the minimum throughput is critical for a CBR application (constant bit rate) when this did not make sense for a Best Effort application. As following distinct types of applications:

a) Inelastic applications: work with constant throughput applications, therefore the maximum throughput will not be considered and has no importance so his weight will be null. d_{min} and C will be taken into consideration in this case.
b) Streaming applications: used for real time services with variable throughput (ex: Video service using Mpeg4). Three parameters will be considered for those applications: d_{max}, d_{min} and C.
c) Elastic applications: used for data transfer services such as the transfer of files, email, and web traffic services. For these applications just d_{max} and C will be considered, when d_{min} will be ignored because these applications don't require guarantees in QoS.

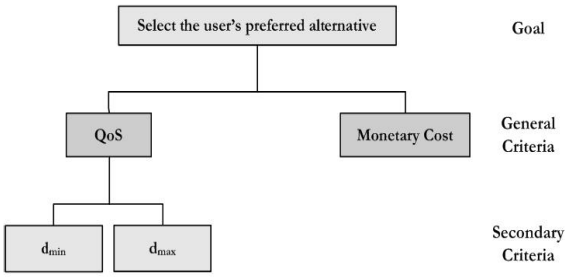


Figure 1. Hierarchical representation of the decision criteria

To evaluate the different alternatives of a session and select the best choice, a method for multi-criteria (RMD) decision will be adopted. Indeed, it defines a utility function which depends on the standard weight of the values of the various criteria. Several utility functions have been proposed, the simple additive weighting (SAW) and the multiplicative exponent weighting (MEW), but these methods do not consider the current needs of the session. For example, when a user using CBR application and is willing to pay, the best alternative will be severed without taking into account the throughput required for the session, and thus an overqualified alternative will be severed, while there is another less expensive alternative meeting the requirements of the session.

8. SELECTION OF NETWORK ARCHITECTURE

We suggest a hybrid approach taking into account the preferences of the operator and the user. Indeed, the policies of the operator are implicitly integrated in the system information. This information will be sent to all mobile terminals. The mobile terminal will decode the information, assess different options and then choose the best network. The operator offers three classes of services (Premium, Regular, and Economic). Financial and QoS considerations will determine the selection of the best alternative. An alternative is defined by a combination of available access technology (WiMax, WiFi, 3G ...) and classes of services such as Premium, Regular, and Economic. When a new session will be opened or where a hand over, the mobile terminal will receive information from the system, decodes the incentives of the operator, evaluates the different alternatives and then chooses the best network with the service class. The information of the available alternatives includes:

The monetary cost: we suggested a cost according to the QoS. As there are different classes of services (Premium, Regular, and Economic) in different networks, the network operators provide different flows for each. Indeed, a Premium session will be more priority, will get a better quality of service and will be more expensive than the Regular, and the Economic. Mobile users will be charged according to their priority (class of service). The prices will change dynamically in real time according to the conditions of the networks, the radio resources used efficiently, and the performance of the system will be improved. As it's more complex to implement a dynamic price, we consider a static price with the different classes of service which will be fixed and does not change with the network load conditions. The flat pricing strategies are not used because they will cause a waste of resources and will force light users to subsidize heavy users and prevent the deployment of internet service quality. In addition, the use of flat price will result in the congestion of a computer network and will degrade the performance of the system.

Accordingly, as we will provide a guaranteed QoS, we propose a model based on the used volume, and then the sessions will be billed according to the amount of traffic that they transmit. The price of traffic per unit will depend on the radio resource access and the class service. In conclusion, the monetary cost is for each unit of traffic and in our case is defined per Kbyte.

QoS incentives: different QoS parameters to consider are based on the application's requests. They specify the minimum number (n_{min}) and the maximum (n_{max}) units of radio resource localized for a session. These sub-parameters depend on radio resources and the service class. The total traffic for a specific RAT is hidden. Different nmin and nmax are generated for the different classes of services reflecting the strategies of the operator. These sub-parameters don't necessarily reflect the conditions of the network but rather the operator's wishes to serve the different sessions of different classes. Based on the SNR report, the mobile terminals will adopt the modulation type and the FEC for the encoding. This is why the number of bits per RRU and the minimum and maximum flows depend on radio's conditions. Indeed, during the evaluation of different alternatives, the mobile terminal combines its radio conditions (which differed from rat to another) and with the reported QoS Sub-parameters then determines the minimum and maximum expected flow. E.g.

for OFDM-based technologies the minimum flow expected will be:

$$d_{min} = \frac{n_{min} \times N_u \times K \times R_c}{SI} \quad (2)$$

With n_{min}: the minimum guaranteed of OFDM reserved symbol, N_u: the number of carrier used for data transmission, K: the number of bits per symbol module (vary with the modulation) , Rc: FEC report, and SI: scheduling interval .

To make it more simple and homogeneous the QoS incentives for different radio technologies, we will express the Qos sub-parameters in terms of minimum and maximum guaranteed flow d_{min} and d_{max} (instead of n_{min} and n_{max}). To evaluate the different alternatives, the mobile terminal will determine the expected flow that represent the result of multiplication of minimum flow (resp. maximum) of the class of service in the alternative with the gain of modulation g_M and g_C coding gain.

$$d_{\min oumax}^{userk} = d_{\min oumax}^{service\ classe\ i} * g_M * g_C \quad (3)$$

9. New Multi-Criteria Decision Method

As shown in the previous paragraph, there are different existing methods used to decide the selection of an alternative. The method used in this paper is a multi-criteria decision method (MCDM). It defines the alternative as a combination between a network and one of the classes of service Premium, Regular or Economic. The alternatives will be evaluated according to their monetary cost, minimum flow that they guarantee and the maximum rate they offer based on the satisfaction of the user. We define then a function of satisfaction for each type of session (inelastic, streaming and elastic) and user profile. The HWAN (or NGWAN) allow the efficient use of available radio resources and then they can serve more customers that will generate more profit. Mobile users can connect simultaneously or not, the different access technologies that meet their needs in terms of QoS or cost. As cited before, our method is a hybrid method (shared between network and users). We define an environment that integrates the operator's objectives and the user's preferences. So our method of decision is based on the user's satisfactions. Our method selects the best alternative based on the expected user satisfaction. The utility function is defined as the weighted sum of partial satisfaction functions. The function of partial satisfaction ($s_{c,p}$) depends on the decision criterion (c) and the profile of the user (p). There are two types of users: those who are willing to pay for best performance and those who prefer to save. As mentioned before, three are three different types of applications, so we will have six users' profiles:

a) Profile 1: The user Initializes an inelastic session and is willing to pay.
b) Profile 2: The user Initializes a streaming session and is willing to pay.
c) Profile 3: The user Initializes an elastic session and is willing to pay.
d) Profile 4: The user Initializes an inelastic session and prefers to save money.
e) Profile 5: The user Initializes a streaming session and prefers to save money.
f) Profile 6: The user Initializes an elastic session and prefers to save money.

The function of satisfaction expected S (a_i) for the alternative a_i is given by:

$$S(a_i) = w_{dmin,\,p} * s_{dmin,\,p} + w_{dmax,\,p} * s_{dmax,\,p} + w_{cost,\,p} * s_{cost,\,p} \quad (4)$$

Where ($w_{dmin,\,p}$, $w_{dmax,\,p}$, $w_{cost,\,p}$) represents the static weight vector of profile p, and $s_{dmin,\,p}$, $s_{dmax,\,p}$, $s_{cost,\,p}$ represent the functions of partial satisfaction of profile p.

The function of satisfaction of flow depends on the QoS needs of the session. Inelastic applications are characterized by a fixed flow, R_f, and the QoS requirements for these applications are strict and inflexible and therefore the function of satisfaction of the minimum flow ensures is defined by:

$$S_{d_{min,p}} = \begin{cases} 0 & si\ d_{min}(a_i) < R_f \\ 1 & si\ d_{min}(a_i) \geq R_f \end{cases} \quad (5)$$

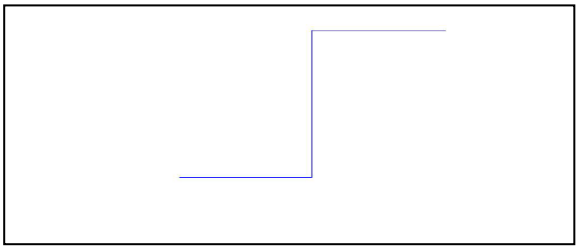


Figure 2. Form of the function of satisfaction for an inelastic session

Where d_{min} (a_i) represents the minimum rate guaranteed by the alternative a_i. In this case the function of maximum satisfaction will not be considered.

Streaming sessions require a minimum rate but also a maximum flow as it is real-time applications. Their function of satisfaction is in the form of sigmoid and is defined by:

$$S_{d_{min,p}} = 1\text{-}\exp\frac{-\alpha\left(\frac{d_{min}(a_i)}{R_{av}}\right)^2}{\beta + \frac{d_{min}(a_i)}{R_{av}}} \quad (6)$$

$$S_{d_{max,p}} = 1\text{-}\exp\frac{-\alpha\left(\frac{d_{max}(a_i)}{R_{av}}\right)^2}{\beta + \frac{d_{max}(a_i)}{R_{av}}} \quad (7)$$

Where α, β are positive constants that determine the shape of the sigmoid, d_{max} (a_i) and d_{min} (a_i) are the maximum and minimum flow guaranteed by the alternative a_i and average flow R_{av}.

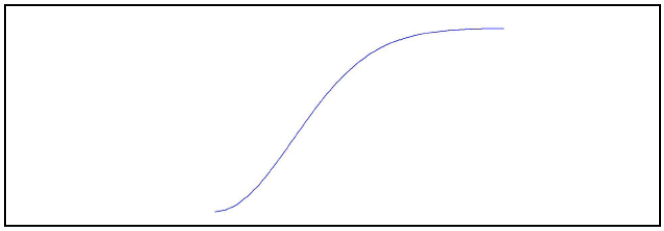


Figure 3. Form of the function of satisfaction for the streaming sessions.

For elastic sessions, the function of satisfaction is a concave function defined by:

$$S_{d_{max,p}} = 1\text{-exp} - \frac{d_{max}(a_i)}{R_c} \quad (8)$$

Where R_c is the flow of comfort.

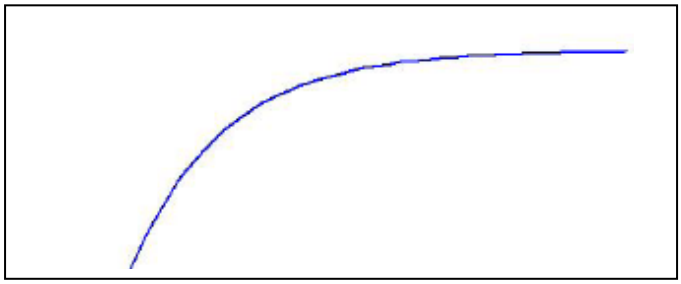


Figure 4. Form of the function of satisfaction of the elastic sessions

Satisfaction of the monetary cost function:

This function depends on the user's tolerance, it is modeled by: $S_{cost,p} = \exp - \frac{cost(a_i)^2}{\lambda_p}$ (9)

Where cost (a_i) is the monetary cost of the alternative a_i and λ_p a positive constant that depends on the profile p. More the user is tolerated in terms of cost, more λ_p is greater ($\lambda_1 = \lambda_2 = \lambda_3$ and $\lambda_4 = \lambda_5 = \lambda_6$).

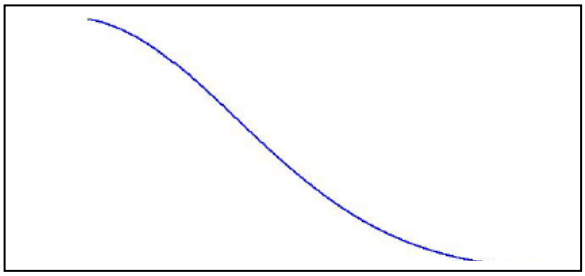


Figure 5. Form of the function of satisfaction of the monetary cost

10. SIMULATION

For the simulation, three similar Radio Access Technologies (R1, R2 and R3) are considered. They all have the same service classes (Same Cost and QoS incentives for all classes: Premium, Regular and Economic). Our goal behind this simulation is to let the user make a choice in terms of Cost and QoS incentives; as he won't be concerned with the technology behind this.

As a result, there will be possibility that there is no load balancing: R1 can be filled first, then R2 and R3. To avoid this case, it is necessary that the operators intervene in the network. The intervention will occur at special moments that will be determined later for load balancing.

10.1. Discrete Events System

A discrete event system is a system described by discrete state variables, i.e. changes occur on the occurrence of a set of states.

We have different types of possible events during the lifetime of the system; thus, we must describe the operating logic between events (determine state changes for each event and the events that result).In our system, we define three main events: session Arrival (A), session departure (D) and the end of a frame (FT).

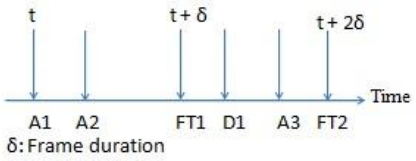


Figure 6: Discrete Events System

If the event is an arrival we must expect the departure, mark the session as active, and expect the new arrival. If it's a departure, we must free the resources and mark the session as terminated. And if it is an end of frame, there must resources allocation.

Matlab has been used to make the simulation, as well as the programming of the intervention of the operator under different strategies. The operator can intervene either by adjusting the Cost or the QoS incentives.

The following approaches have been simulated for the adjustment of QoS incentives:

10.1.1 Binary logic

The QoS incentives are decreased in each thereshold.

$$D = \begin{cases} D_{init} \text{ if Network Occupancy} < S_1 \\ D_{init}\text{-X} \quad \text{if S1<Network Occupancy<S2} \\ D_{init}\text{-Y} \quad \text{if Network Occupancy>S2} \end{cases}$$

10.1.2 Binary logic with return to initial conditions

It's similar to the binary logic. But in this scenario, the value of the QoS incentives are returned to their initial values after decreasing the QoS incentives in the three networks in each threshold.

10.1.3 Fuzzy logic

The following function is defined:

$$D = \begin{cases} D_{init} \text{ if Network Occupancy} < S_1 \\ D_{init}\text{-}\dfrac{(D_{init}-D_f)*(\text{dminTotal}-S1)}{S2-S1} \\ \quad \text{if S1<Network Occupancy<S2} \\ D_f \text{ if Network Occupancy>S2} \end{cases}$$

All the above scenarios could be programmed by increasing costs instead of decreasing the QoS incentives.

11 Results

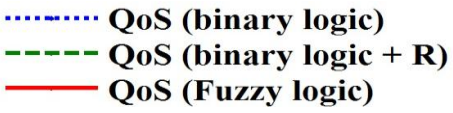


Figure 7: Figures legend

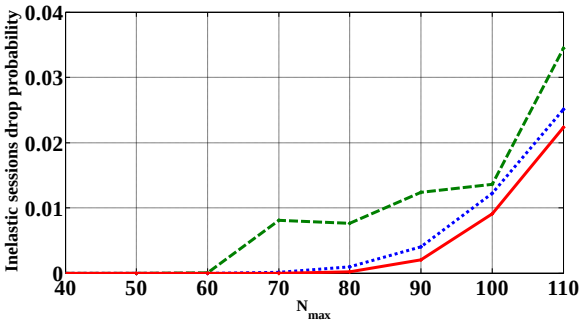


Figure 8: Inelastic session drop probability

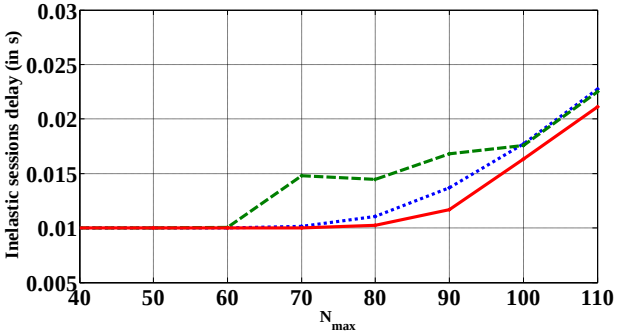


Figure 9: Inelastic session delay

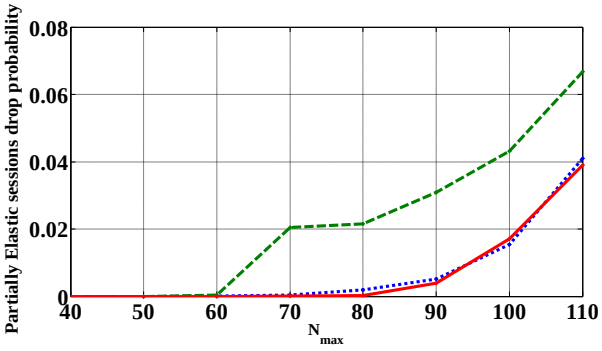


Figure 10: Partially elastic session drop probability

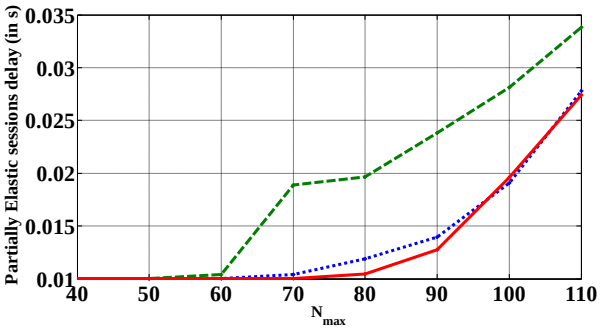


Figure 11: Partially elastic session delay

As per the above figures, it is obvious that the fuzzy logic intervention offers the best performances for both the inelastic and the partially elastic sessions. These performances are shown by the delay (Figure 8, 10), and by the drop probability (Figure 7, 9). The main reason is that the different networks are filled in a regularly and continuous way with the fuzzy logic. This will give the sessions a higher amount of resources to exploit in each network. Therefore, sessions will have a higher throughput and thus a lower delay and drop probability. On the other hand, it's clear that the binary logic with return to initial conditions provides the worst performance. Whereby, the operators are having fewer resources to exploit by offering higher throughputs with initial conditions.

As per the binary logic, it has approximately the same results as the fuzzy logic. The fuzzy logic could be considered as binary logic with multiple thresholds.

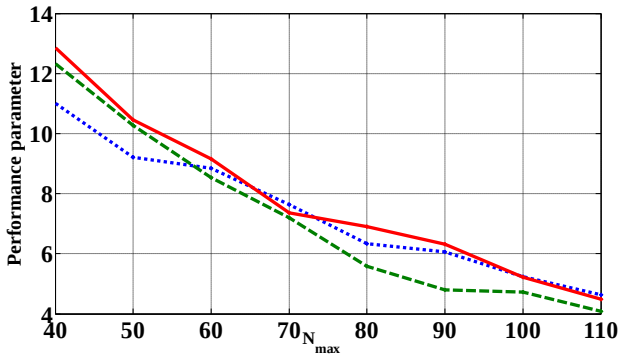


Figure 12: Performance parameter for elastic session

Same results are shown for the performance parameter of elastic sessions (Figure 12) with best performance for fuzzy logic and worst for binary logic with return to initial conditions.

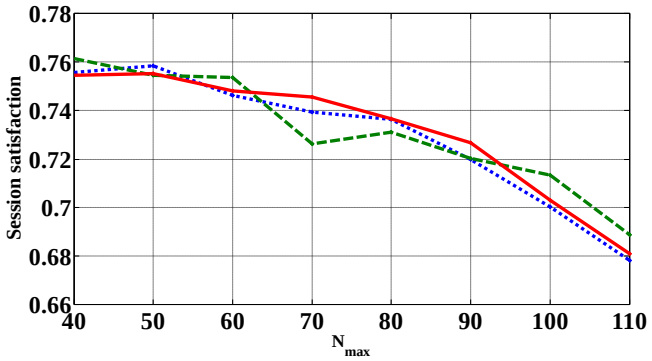


Figure 13: Session satisfaction

The satisfaction (Figure 13) is higher with fuzzy logic for medium load. It decreases quickly for medium load using the binary logic with return to initial conditions.

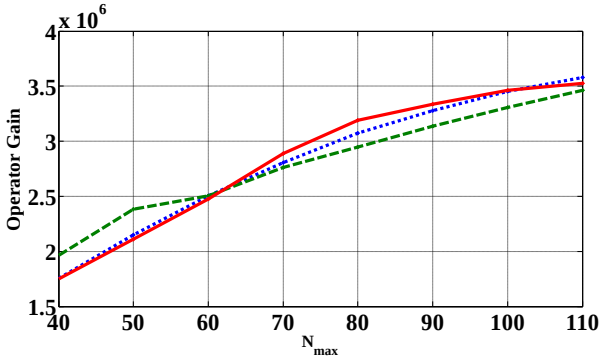


Figure 14: Operator gain

The operator gain (Figure 14) is a bit higher using the fuzzy logic and has the lowest values using the binary logic with return to initial conditions.

12 Conclusion and Future Work

Different operators' intervention strategies were presented. Each strategy is based on a specific logic: binary, binary with return to initial conditions, and fuzzy logic. After the implementation of these strategies, it was obvious that fuzzy logic presents the best performances; whereby the lowest delays, drop probability and the highest performance parameters are met. Besides, the operators benefit from a higher financial gain. The results also show that the operators should reduce the QoS incentives.

For future work, these methods could be implemented with SON (Self Organization Networks), which is a new technology presented in the LTE standard for auto-configuration, auto-optimization and auto-exploitation of cellular networks equipments for mobile telephony.

13 Acknowledgment

I would like to thank my classmate Rita Maria Salameh and Mr. Melhem El Helou for their great support.

14 REFERENCES

[1] Tripathi, N.D.J., Reed, H..VanLandinoham, H.F.: Handoff In Cellular System. IEEE Personal Communications, 49, 2276--2285 (2000)

[2] Gustafsson E., Jonsson, A., Research, E.: Always Best Connected. IEEE Wireless Communications, 10, No. 1, 49--55 (2003)

[3] Steven-Navarro, E., Wong, V.W.S., Lin, Y.: A Vertical Handoff Decision Algorithm For Heterogeneous Wireless Networks. Wireless Communications and Networking Conference, IEEE, Kowloon 2007, pp. 3199--3204.

[4] steven-Navarro, E., Wong, V.W.S.: Comparison between vertical handoff decision algorithms for heterogeneous wireless network. Vehicular Technology Conference, IEEE 63rd ,Melbourne, Vic. , 947--951 (2006)

[5] Zhang, W.: Handover Decision Using Fuzzy MADM In Heterogeneous Networks. Wireless Communications and Networking Conference, *IEEE*, Vol.2, 653-658 (2004)

[6] Tawil, R., Pujolle, G., Salazar, O.: A Vertical Handoff Decision Schemes In Heterogeneous Wireless Systems. Vehicular Technology Conference, VTC Spring 2008 IEEE, Singapore, 2626--2630 (2008)

[7] Tawil, R., Demerjain, J., Pujolle, G., Salazar, O.: Processing-Delay Reduction During The Vertical Handoff Decision In Heterogeneous Wireless System. International Conference on Computer Systems and Applications, AICCSA IEEE/ACS,381--385 (2008)

[8] Savitha, K., Chandrasekar, C.: Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks. Global Journal of Computer Science and Technology Volume 11 Issue 9 Version 1.0 May (2011)

[9] Tawil, R., Pujolle, G., Salazar, O.: Vertical Handoff Decision Schemes For The Next Generation Wireless Networks. Wireless Communications and Networking Conference, 2789--2792 (2008)

[10] Tawil, R., Pujolle, G., Demerjain, J.: Distributed Handoff Decision Scheme Using MIH Function For The Fourth Generation Wireless Networks, 3rd International Conference on Information and Communication Technologies: From Theory to Applications, 1--6 (2008)

[11] Wang, L., Binet, D.: MADM- Based Network Selection In Heterogeneous Wireless Networks: A Simulation Study. 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronics Systems Technology, 559--564 (2009)

[12] Nasser, N., Hasswa, A., Hassanein, H.: Handoff In Fourth Generation Heterogeneous Networks. Communications Magazine, IEEE , 44, 96--103 (2006)

Approach to partition heterogeneous mobile ad hoc network into homogeneous clusters

D. Karimou[1], **H. Saliah-Hassane**[2]
[1]LICEF Research Center Montreal Canada
[2] TELUQ | University of Quebec & LICEF Research Center Montréal Canada

Abstract - *In heterogeneous system nodes have different architectures, each of which produced different Quality of Service (*QoS*). Since nodes have different characteristics, communication is a crucial problem. This difference is huge and can sometimes influence the quality of broadcasts and communications. Applying proprietary protocols are difficult in heterogeneous environment because qualities of services offered vary from one node to another. It would make it far easier to manage quality of service in a network formed by homogeneous nodes (homogeneous architectures) than manage quality of service throughout the network of heterogeneous nodes. This paper draws this problem and addresses approach to partition nodes by proposing a protocol that allows splitting a heterogeneous mobile ad hoc environment into several homogeneous clusters. We show how nodes can be partitioned into several groups formed by members who have similar characteristics, a same degree of similarity, a same patterns and offer the same quality of service.*

Keywords: classification, Mobile ad hoc network, QoS Networks, Clustering.

1 Introduction

Mobile wireless networks have received much attention in recent years. This is due to their potential in many areas such as military applications, control systems, security, target tracking, rescuing disaster areas… Nowadays, mobile networks are becoming an essential component of our daily activities.

Unlike wired networks, mobile ad hoc network is a distributed system formed by entities connected through mobile wireless technologies. Network is a temporary system, without any existing infrastructure or centralized administration. Network nodes are deployed in environments without using any fixed support. Nodes must be able to self-organize and cooperate to provide services. Mobile architectures offer several possibilities of communication through media such as laptops, *PDAs* etc... These equipments are equipped with different interfaces such as *UMTS*, *WiFi*, *WiMax*, Bluetooth etc. with different levels of service. Communication systems are composed of several components such as physical devices, medium access, and applications.

Depending of systems, architectures and protocols, network is heterogeneous. Indeed, either the bit rate, increased radio bandwidth, and terms of energy conservation, heterogeneous mobiles networks nodes present different capacities to communicate and broadcast messages. This difference significantly affects the quality of service that depends on services produced by different nodes in the system.

Heterogeneous networks are applied in several areas. Imagine a particular deployment of a network formed by several fixed nodes, mobile infrastructures with sensor nodes and mobile ad hoc nodes. These nodes can self operate to form a single network with high QoS performance. Mobile nodes can be used for public safety on people involved in an environment. Mobile ad hoc nodes can also be deployed on inaccessible areas where there are no infrastructures. Sensors are also used in inaccessible areas where they can reassemble data from environment by using possibilities of mobile nodes to cooperate and send data to fixed nodes that are gateways to a command post or treating data. For example, such facilities can be used in areas of claims, to measure physical quantities, biological, environmental and traffic.

Techniques of "clustering" are unsupervised learning methods that involve using mathematical and statistical techniques to group several similar objects into groups or classes. They are used in many fields such as "e-business", biology, "web-mining", image analysis. Several partitions algorithms have been proposed in the past, with multiple steps and producing different results. In our approach, we propose an efficient method for grouping nodes into clusters formed by degree of similarity. These identifiers are defined during deployment by using identity or color named *idR*. The idea in this paper is to propose an alternative way to partition heterogeneous mobile ad hoc network into homogeneous clusters. This solution can manage quality of transmissions. Indeed, as the nodes of network consist of different architectures with different characteristics, communications and cooperative initiatives vary with the Quality of Service of each node. Resource requests, ability to coordinate transmission, bandwidth, and storage capacity of node also vary according platforms. Indeed, in heterogeneous network, some nodes have qualities of services (transport, storage, bandwidth ...) limited capacity and interact less with those with more capacity around them. Similarly, transmission powers vary depending on technical architecture of nodes. Some nodes may be equipped

with batteries with large energy potential. For example, in many multimedia technology flows need most network bandwidth. But with actual miniaturization of electronic components, devices such as *PDAs*, laptops and portable audio/video are diverse en terms to need resources. Some of these devices have better quality transmission than others. It would make it far easier to manage quality of service in a network formed by homogeneous nodes than manage quality of service throughout the network of heterogeneous nodes. In this article, we build from heterogeneous mobile ad hoc network, several groups of nodes consisting of architectures and similar characteristics. The advantage of this architecture is to allow nodes with the same characteristics to communicate and cooperate. This will improve quality of local service group. Subsequently, protocols and bridge links can be established to ensure quality of communications between clusters. This article does not deal with these protocols, but how the heterogeneous network is partitioned. This protocol is based on traditional methods of clustering. For this, we study partitioning techniques identified in literature and apply an appropriate method in case of mobile ad hoc networks composed of a set of heterogeneous nodes operating in a dynamic environment.

1.1 Previous works

Research on heterogeneous environment becomes crucial. Recently, such research is more valuable because, with the miniaturization of electronic components, networks are formed with heterogeneous technologies with various qualities of services.
Clustering is a method used in many fields. Grouping nodes in mobile network into homogeneous clusters provides good quality of cooperation and information transmission. Qualities of Service depend on network and nodes technology. Therefore, as nodes are heterogeneous, system will consist of different quality levels of services related to network components. It is necessary to define an architecture less influenced by heterogeneity of system components. To achieve this, we use the way to group mobile network nodes into several similar groups. Indeed, existing clustering methods in literature are not cases of this similarity between nodes for partition criteria.

Clustering methods [1] [3] [5] are based on choosing a group leader called clusterhead reflecting identifiers of nodes, neighbors and degrees of nodes mobility to partition network . These methods are based on a synchronization of different neighboring nodes to build a system formed by single hop clusters. In [2] and [3], authors proposed a clustering method which is based on the weight of nodes to select a clusterhead. Performance factors such as maximum number of neighbors of a clusterhead or factors such as weight are important criteria for connecting a node to a clusterhead. These methods in turn form a set of single hop clusters. Other methods of clustering have been presented in [4] and [7] that provide clusters formed by links to *k* hops. But it should be noted that these methods of partition does not take into account the heterogeneity of nodes to ensure a better quality of service during broadcast and cooperation between nodes.

1.2 Our results

We propose in this paper an approach to partition heterogeneous environment of mobile ad hoc network into several parts formed by homogeneous nodes. This protocol is based on different methods for grouping nodes with similar characteristics. This similarity is announced before the deployment of nodes in the network identification numbers by using degree of similarity noted *idR*. This approach differs from existing methods. In effect, following the heterogeneity of nodes, we must ensure that the quality of service provided by any node is same for the other nodes in the same group because when one node has failed qualities then qualities of broadcast in the same group become dysfunctional. This model allows partitioning nodes according to their degree of similarity technology and architecture. Advantage of such partition is to have a cluster system consisting of nodes with the same levels of quality to function. It is possible to apply uniform management policies and quality of service consistent within each cluster. Proprietary protocols may be sometimes applied by group of homogeneous nodes. Also, other protocols and techniques may be applied between the clusters as bridges bindings for broadcasts and external collaborations. These protocols can be specific depending on architecture of cluster nodes to unite for communications. Certainly, we will build on the existing clustering methods but we propose a new protocol that allows partitioning nodes of a heterogeneous mobile ad hoc network into a set of clusters formed by homogeneous elements with similar characteristics.

The remainder of this paper is organized as follows: some definitions and working environment are presented in Section 2. In Section 3, we present the clustering protocol which essentially comprises three parts. The first step of section deals about clustering of nodes in network without considering heterogeneity of system components. The second step is to gather the information necessary to execute the next step 3. Then in the third step, we see clustering method which allows having clusters formed by nodes with the same degrees of similarity. A conclusion ends this paper.

2 Basic definitions

Our environment of work is a mobile ad hoc network formed by a set of nodes that can transmit or receive information from their neighborhood through *k* channels of broadcast frequencies. It is an autonomous system formed by heterogeneous mobile nodes that are interconnected by wireless links without any infrastructure or centralized administrator. Network is dynamic and nodes can move on transmission channels. We consider that stations have already identities by using the identification protocol as in [6]. Thus

each station has an identity between *1* and *p*. Each node also has an *idR* (ID degree of similarity).

Assumption 1

1. Communications are made via mobile links. A link exists between two nodes in the network only if they are on the same transmission frequency. Two nodes that are connected to one hop are called neighbors and all nodes that can transmit directly to any node are the immediate neighborhood of this node.
2. Residents of a cluster: each station belonging to a cluster is a resident of this cluster. Therefore, this node may transmit information to all its neighbors.

Assumption 2

Consider a set of nodes that communicate through a multi-hop ad hoc network. We assume that we have *k* transmission channels.

3 Partition nodes into homogeneous clusters

Our approach of partitioning heterogeneous mobile ad hoc environment is formed by three steps:

Step1: Partition nodes into ordinary clusters. First, we present this approach that allows to partition nodes into several heterogeneous clusters regardless of their degree of similarity. This method of clustering is based on method of Basagni [2] but with some modifications to reflect our environment of work.

Step 2: broadcast reconstitution information. In this step, information on identities of degree of similarity between clusterheads is broadcasted into clusters. This step broadcasts information that is required for each node to move into a cluster with the same degree of similarity.

Step 3: clustering nodes with similar degrees of similarity. In this step, heterogeneous nodes are grouped into several clusters according to their common characteristic defined by an identity. To this end, we introduce the classical methods of partition. At the end of this phase, all nodes are grouped into *k* clusters. Each cluster can contain nodes with similar degrees of similarity and share the same common characteristics.

3.1 Step 1: Procedure of clustering

The main purpose of this step is to partition network into cluster. This process produces a system composed by a set of clusters with arbitrary elements. The goal of clustering is to partition entire network into different parts. Each part named cluster is formed by a set of connected nodes (clusterhead, ordinary nodes) that provide broadcasting information on the network or on a cluster without all the nodes participate in this operation.
Consider that we have a graph *G (V, E)* represents indirect communications network where *V* is the set of mobile nodes *V = {v1, v2, ... vp} {| V |}* and *E* is the set of edges. We consider that each node *vi* has a positive weight *(wi)*. For two nodes *vi, vj, wi # wj*. Similarly, we consider for each pair of nodes *vi, vj,* edge *eij = eji*. In other words, the connections between two edges are bidirectional (symmetric graph). Clusters are notes *Ci* for $1 \leq i \leq k$ and the clusterhead *CHi for* $1 \leq i \leq k$. The clustering process divides *V* into a set of *k* subsets *{V1, V2, ... Vk}* where *V = Uki = 1Vi*, so that each subset *Vi* form a connected graph *G*. And each subset represents a cluster.

Our clustering approach is based on the protocol of Basagni [2] in which he proposed the use of the weight of the nodes in the choice of clusterhead. We assume in this approach the constraints below:

Each node knows its degree, its weight, and the degrees of the nodes in its neighborhood.

- At the end of clustering process, each clusterhead knows the degree of all others clusterheads.

The clustering process consists of two phases:

- Phase 1: choosing the clusterhead. This first phase selects a clusterhead for each partition. Each cluster is composed of a clusterhead node leader and a set of ordinary nodes. The choice of clusterhead is based on the degree associated with each node. In case of equality, we refer to the weights of nodes to decide between. In this particular case, the node with the highest weight becomes clusterhead. In this way, nodes with larger degrees initialize the clustering process by broadcasting a message to their neighbors who announced their wish to be clusterhead.

If a node *u* receives a message from a node *v* and the degree of *v* is greater than *u*, then *u* is resigned and considers *v* as clusterhead.
One node *i* becomes a clusterhead if at least one of the following conditions is met:

- *i* is the largest degree among its *1*-neighborhood (close to a jump). Or

- *i* is not the node with the highest degree in its *1*-neighborhood, but all its neighbors belong to *1*-jump to other clusters

Phase 2: choice of ordinary nodes. For a node *v* with degree less than all its *1*-neighborhood, two situations arise:

- If *u* is a node in its 1-neighborhood as clusterhead, then the node v is located in the cluster formed by *u*.

- If his neighbor *u* greater degree belongs to another cluster, then two cases are possible. If it exists in its 1-neighborhood, another clusterhead then *v* moves in this cluster. If not, then *v* creates its own cluster.

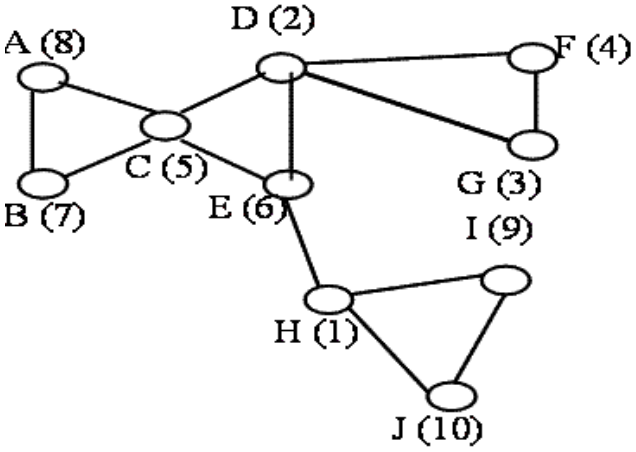


Figure 1: Initial Mobile Ad Hoc Network

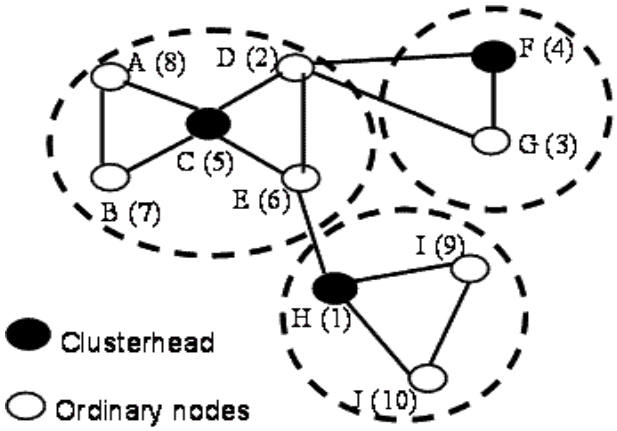


Figure 2 : Clustering Algorithm

For example, the algorithm *(see Figure 1 and 2)* creates two no-covering clusters in which each node selects a clusterhead of connection among its neighbors. At the end of the process, the following conditions are met:

- The entire network is partitioned into clusters. Each node belongs to exactly one cluster. A node is either a clusterhead or an ordinary node directly connected to a clusterhead with the higher level in its *1*-neighborhood;

- Each cluster has a degree equal to 2 and each node is located at a clusterhead or breaks its clusterhead;

- At the end of the process, each clusterhead knows the number of nodes that are part of the cluster. Also, device nodes which connect the cluster to its neighboring nodes called edges are known by the two clusters they belong.
Indeed, edge nodes are the stations that make the junction between two clusters and probably communicate with the other edge nodes in other clusters. These nodes provide translation of internal and external information to be sent or received clusters.

- Each node usually has at least one clusterhead in its *1*-neighborhood and belongs to the cluster formed by its neighboring node with the highest degree;

- Two clusterheads can never be ordinary neighbors and each node is one hop from its clusterhead.

3.2 Step 2: information collection procedure

This step allows each node to have information necessary to enable him to join the cluster formed by nodes with the same characteristics by their *id*R. At the end of process of step 1, system is partitioned into different clusters *CL1, CL2.... CLK.* Clusterheads nodes and edge nodes are known. We consider that each clusterhead is responsible for broadcasting information to its neighborhood or into the other cluster using its boundaries nodes.

3.2.1 Order for broadcasting *idR*

It is important to know what time each cluster can start broadcasting its *idR* information. Order is based on the weight of clusterheads. For that, clusterhead broadcast their weight each other. Since the weights are different, when each cluster receives the weight of its neighbors, it can directly know its broadcasting time.
This allows knowing the cluster that starts process of broadcasting.
After that, we consider that each cluster is reduced to its clusterhead. Each clusterhead represents its cluster. The entire network is reduced to a network formed by clusterheads such us in figure 3. Broadcasts are operated successively in the order.

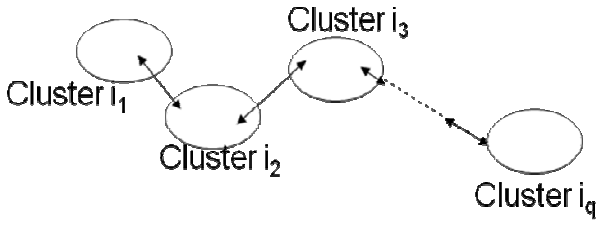


Figure 3

To allow one node to find cluster which regroups node with the same degree of similarity, it is necessary to this node to know clusterhead with which it have the same *idR*.
Then, each node must know *idR* information of clusterheads. For this, each clusterhead broadcasts its *idR* into the others clusters. To send this information, a clusterhead uses its boundaries nodes to joint when it is its turn to broadcast.
Locally, when extern *idR* is received by a clusterhead, it broadcasts to its residents this new *idR*.
At the end of this broadcast, each clusterhead and each node know *idR* identities for all clusterheads which form partitions.

3.3 Step 3: Grouping nodes into a cluster with the same degree of similarity.

3.3.1 Computing similarity criteria

To group objects that are alike, it is necessary to choose a similarity criterion called the similarity criteria. Several algorithms exist to calculate the similarity between objects, clusters or object and cluster. This similarity is expressed by the distance. However, the definition of distance is different depending on the nature of variables (Booleans, objects, etc. vectors). For this, we will rely on the information we have on the nodes such us identity *idR*.

In our approach, computing measures of similarities are based on the assessment of metric indices between variables *idR*. These variables are integers, so we consider that a node is identical to its *idR*. Consequently, we calculate a similarity index of putting *idR* effects (importance between a cluster center *Ci* and a node *j*). This index represents the euclidean function between *Ci* and the clusterhead node *j*.

$$d(Ci, j) = \arg\min \sqrt{(idR_{ci} - idR_j)} \qquad (1)$$

$j \in 1,2,.....k$ and idR_{ci}, idR_j are respectively *idR* for clusterhead *CLi* and *idR* for node *j*.

This distance (equation 1) is all the more small than the node and the cluster are similar.

After this computation, this node leaves the cluster to migrate to its new cluster.

But, in some case, we have two clusterheads or more which have a same degree of similarity (same *idR*). If this case occurs, node computes an operation (equation 2) to find the nearest clusterhead. This step solves the duplicates of clusters composed by nodes with similar degrees of similarity. Duplicate is occurred when two clusterheads have the same *idR*.

For, this operation, we refer to an approach which like to a classical method of classification of *K-means*. [8][9][10][11]

In this method, objects are grouped around the centers of classes determined iteratively. The number of clusters is fixed and each cluster is represented by its center c_i.

Assigning an object x_k to a cluster is a process and winning class is one whose center is closest according to a measure of distance:

$$f(x_k, c_i) = \arg\min \left\| c_i - x_k \right\|^2 \qquad (2)$$

Where x_k and c_i represent respectively the positions of node *xk* and the center of the cluster.

This algorithm seeks the minimum between the object and each center.

We can refer to this method because after step 2, we consider that clusterheads broadcast also their positions. Therefore each node usually performs the operation (3).

$$f(x_k, c_i) = \arg\min \left\| c_i - x_k \right\|^2 \qquad (3)$$

1) Consider cluster with the same *idR*

2) Assign node x_k to cluster which has center c_i such that *f* is minimal.

When minimum of this function (3) is obtained, according *k-means* method and the above information, node and its clusterhead have the same *idR* identity, namely configuration, architecture offering the same qualities of services etc…

Once the minimum is obtained then node migrates into cluster which has center c_i. Operation is repeated for each ordinary node to choose a nearest cluster when two clusterheads or more have the same *idR*. At the end of this process, all clusters are composed by nodes with similar degrees; namely a same identity of degree of similarity.

3.3.2 Grouping cluster with the same idR

Other problem that may occur is to have clusters formed by clusterhead which have the same identity *idR*. If this case occurs, we can refer to agglomerative method for grouping several clusters into one cluster.

1) Indeed, techniques of *k-means* [8] can just group a cluster with an object, but in agglomerative method [12][13], we can group two clusters together to form one cluster. This method is based on the distance between two clusters according to the process.

2) Each object is stored in a cluster, then we look for the closest two clusters by the function:

$$M(i) = \min[\, d_{jk} \,] \qquad (4)$$

Such as $idR_j = idR_k$ with $j, k = 1 \ldots$ and $j < k$.

d_{jk} is a distance between a pair of clusters:

$$d_{jk} = \sqrt{\frac{n_j \times n_k}{n_j + n_k}} \times \left| m_j - m_k \right| \qquad (5)$$

n_j, and n_k represent the numbers nodes (density) into clusters j and k. m_j, and m_k represent the gravity center of each cluster j, k.

At the end of this process (equation 5), we have just clusters formed by nodes which are close *idR*. This means that the entire network is partitioned and produces some clusters formed by nodes of the same degree of similarity.

4 Conclusion

In this paper, approach show that it is possible to partition heterogeneous mobile ad hoc network into several homogeneous clusters. Main motivation of this paper is to make contribution to give a better quality of service for broadcast in a heterogeneous mobile ad hoc network. We proposed partitioning method based on method of Basagni [2] and conventional methods of clustering. Our Approach demonstrates that olds methods of clustering do not take into account heterogeneity of nodes for the partition criteria. However, architectural and physical characteristics of different levels can influence the quality of the entire network broadcasts. Our approach provides a contribution to this problem.

It would also be interesting to study the protocols that can be applied between two partitions to ensure consistent quality of broadcasts and services in the joining of clusters which are created. Our future work will be based on the study of such protocols.

5 References

[1] D.J Baker and A. Ephremides, ''the architectural organization of mobile Radio Network via a distributed algorithm", *IEEE Transactions on Communications, COM-29*, 1981

[2] S. Basagni. Distributed clustering for multihop wireless networks. In A. Annamalai and C. Tellambura, *Proceedings of the IEEE International Symposium on Wireless Communications (ISWC'99)*, pp. 41-42, Victoria, BC, Canada, June 3-4 1999.

[3] P. Basu, N. Khan, D Thomas and C. Little, '' A mobility based Metric for Clustering in Mobile Ad-Hoc Networks'' *21 st international Conference on Distributed Computing Systems Workshops (ICDCSW'01)*, 2001.

[4] G. Chen, F. Nocetti, J.S. Gonzalez, and I.Stojmenovic, ''Connectivity-based K_hop Clustering in wireless networks'', *in Proc. Of the 35th Hawaii International Conference on System Sciences (HICSS-35)*, January 2002.

[5] M. Gerla, and J. Tsai, ''Multicluster, Mobile,multimedia radio network'', *ACM-Baltzer Journal of Wireless Networks*, Vol.1, No.3, pp.255-265,1995

[6] D. Karimou and J. F. Myoupo. A Markov Chain Based Protocols for Dynamic Initialization of Single Hop Mobile Ad Hoc Networks, *International Conference on wireless Networks (PDCN' 05)*. Innsbruck, Austria, June 21-24, 2004, pp. 38-43.

[7] S. Sivavakeesar, G. Pavlou, C. Bohoris and A. Liotta, ''Effective Management through Prediction Based-Clustering Approach in the Next-Generation Ad hoc Networks'', *In the Proc. Of the IEEE International Conference on Communications (ICC '04)*, France, June 2004.

[8] Forgy, E. W. (1965) Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics 21*, 768–769.

[9] Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. *Wiley*, New York.

[10] DIDAY E.1971. Une nouvelle méthode en classification automatique et reconnaissance des formes. La méthode des nuées dynamiques. *Revue de Statistiques Appliquées*, vol.XX, no2, pp.19-33.

[11] Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics* 28, 100–108.

[12] Theodoridis, S. and Koutroubas, K. (1999). *Pattern Recognition. Academic Press.*

[13] Zhang, T., Ramakrishnman, R., and Linvy, M. (1996). BIRCH: An Efficient Method for Very Large Databases. *ACM SIGMOD*, Montreal, Canada.

Understanding the effect of network-coding and video-encoding on multimedia streaming for peer-to-peer (P2P) systems in wireless networks

Makbule Ozturk and Victor Clincy
Computer Science Dept, Kennesaw State University, Kennesaw, GA, USA

Abstract - *Although great advances have been realized in wireless networking over the last ten (10) years, the wireless transport medium is not ideal for multimedia video applications. This paper proposes the basis for a study in quantitatively understanding the effect in deploying various combinations of video encoding and network coding approaches on peer-to-peer streaming systems in a wireless environment. A simulation model will be implemented and used for analyzing the video traffic and its flow through the network. The PSNR (peak signal-to-noise ratio) will be used in understanding the quality of video for the various cases. The PSNR is a common quality metric used that is based on the MSE (mean-squared-error) of two images. The study will also use the coding-bit-rate, packet-loss-ratio and video-characteristics as key factors in quantitatively comparing various coding schemes [19]. In all cases, the video packets will be multicasted and a retransmission-based error-control method will be used.*

Keywords: Wireless networks, P2P, Multimedia streaming, compression, network coding

1 Introduction

Streaming video traffic over a wireless peer-to-peer system is becoming more and more in demand. Applications of video streaming over wireless are ranging from multimedia applications in mobile devices to high data-rate video applications in home networks. Although great advances have been realized in wireless networking over the last ten (10) years, the wireless transport medium is not ideal for multimedia video applications due to fragile and dynamically changing links, and contention issues with limited resources. There is also an effect of interference due to devices operating in the same frequency range.

Video and multimedia streaming is a continuous process that is delay sensitive such that, any packets not received by a certain timeframe are dropped. Traditional approaches to wireless transmission are not sufficient for good quality-of-service of wireless video streaming.

2 Key Issues

Some of the key issues that must be considered and included in the study pertain to resource allocation, video coding or compression, transport layer reliability, and network coding.

2.1 Resource Allocation

Resource allocation is one of the main challenges in wireless video streaming due to the erratic behavior of the medium. Implementation of a suitable video coding algorithm and reliable transporting of coded packets to receivers are the other aspects that need to be taken into account in wireless video streaming. Novel cross-layer design frameworks propose approaches to resource allocation, routing and rate allocation. [2, 21-22]. Due to continuously changing link qualities, routing is a significant concern in wireless mesh networks. Routing solutions are needed to prevent low throughput in multi hop links which occurs with the contention among adjacent links [1]. In cross-layer design frameworks, two or more layers are jointly optimized for increasing the streaming performance. In the cross-layer design study of Wang et al., network coding was also used to adopt reliability of end-to-end transmission at the application layer [2].

2.2 Video Coding

Video encoding or compression is essential before transmission because of the large bit rates of a digital video and limited bandwidth of a wireless channel. In addition, compression methods are very important in terms of achieving optimal QoS by aiming to send only relevant data [18]. There are three sources of redundancies which are reduced by encoding or compression. One such redundancy is spatial redundancy. Spatial redundancy is having neighboring pixels similar in a single frame. The other redundancy is called temporal redundancy, where adjacent frames are correlated. And as a result of reducing these redundancies, a third one occurs in the stream of output symbols. Intra-frame coding, inter-frame coding and variable length coding are techniques used for eliminating the redundancies [16, 17].

H.26X is the family of video coding standards. Some of the mainstreamed standards are Advanced Video Coding

(H.264/AVC), Scalable Video Coding (H.264/SVC), and Multiview Video Coding (H.264/MVC). A next generation coding standard is H.265. Another emerging concept is Distributed Video Coding (DVC) which is popular for video surveillance applications where the decoder is more complex than the encoder. On the other hand, in traditional H.26X coding systems, the encoder has more complexity with motion estimation. Traditional coding systems are more suitable for wireless video streaming applications like Video-on-Demand (VoD) and applications where the video is compressed once and decoded more than once [14]. Source coding techniques and channel coding solutions are widely proposed for successful video streaming in wireless networks.

Scalable video coding (SVC) is the scalable extension of H.264/AVC standard. SVC offers scalability by having video which is coded into more than one layer. With a base layer and enhancement layers, SVC provides a wide range of video quality, resolution, and picture rate in a heterogeneous network. Bit errors and channel errors are the main causes of packet losses in video transmission channels. Using the user datagram protocol which discards packets with errors is another issue which decreases the video quality. It is likely to have a transmission error propagate temporally and spatially, unless error control methods are applied. Error-resilient coding is an error-control approach which adds redundancy to data to get minimum distortion. The redundancy can provide concealment of errors, detection of data losses and preventing of error propagation. [20].

2.3 Transport Layer

TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) are currently the major protocols used for video transmission over the Internet. Stream Control Transmission Protocol (SCTP) is an up-and-coming transport protocol that is message-oriented like UDP and ensures reliable, in-sequence transport of messages with congestion control like TCP. TCP is not preferred for video streaming applications due to stringent delay requirement which is hard to satisfy because TCP uses a back-off procedure and retransmission process [4]. A traditional TCP approach can be better used by having multiple TCP connections for one streaming application. With the multi-TCP design, it overcomes the short-term limited bandwidth concern and rate control insufficiency of TCP. This approach is implemented at the application layer by giving control of the sending rate to the receiver during congestion [5]. Another study that uses multipath TCP-based streaming, proposes a design named Dynamic MPath Streaming (DMP). DMP model shows that it is possible to achieve satisfactory performance with TCP in more stable networks like wired networks [4].

In regards to UDP, Real Time Protocol (RTP) and Real Time Control Protocol (RTCP) run on top of the UDP. These protocols are suitable for media streaming and IPTV with the properties provided like security, media codec control and error protection. These protocols also provide basic RTP media encapsulation and synchronization, quality-of-service signaling and multi-party communication coordination [24].

Bandwidth limitations and packet losses are the main issues in multicasting video packets over the Internet. Aiming for more reliable data communications by reducing the number of packet losses, the retransmission process at the link-layer is provided by IEEE 802.11 standard MAC protocol. However, link-layer retransmission is not effective in the presence of heavy network traffic loads. As the retransmission limit increases, end-to-end delay increases and as a result, it reduces the quality of streaming video over multi-hop wireless mesh networks. If there is TCP traffic along with video streaming, both video and TCP performances are not satisfactory with retransmissions [6]. Retransmission based error control methods are widely studied for video transmission over wireless networks and will be implemented in the study [7-10].

Some parameter settings need to be adjusted in reducing TCP delays, such as; disabling Nagle's algorithm which causes transmission delays by limiting the transmission of small segments; enabling SACK; and using larger receiver buffers for delay sensitive applications. It is also possible to reduce the delay by using packet splitting and having multiple parallel connections with the video flow [23].

2.4 Network Coding

Network coding makes sending more information in one transmission possible by combining packets from different flows into a single packet. In the study of Seferoglu and Markopoulou, video-aware opportunistic network coding scheme increased the throughput and as a result enhanced the video quality [15].

In lossy networks, network coding reduced the number of retransmission of the packets thus increasing the reliability of the network. When network coding was compared to link-by-link ARQ and end-to-end FEC error-control techniques, the results showed that network coding outperformed the other techniques both in lossy links and links where loss-packet probability was lower [3].

Rate adaptation at the application layer avoids link congestion, maintains continuous stream of video by meeting the deadlines of the packets and on the other hand reduces video quality [12].

As it is mentioned before, video streaming requires a particular amount of bandwidth in satisfying the consistency condition. Video servers need to provide high quality video to receivers even in peak traffic times. There are also other concerns like cost and storage in such client-server and P2P systems. Multicasting proposes a solution to these concerns by replicating the packets in intermediate nodes in network [12].

Multicasting reduces the cost by proposing a multicast tree alternative to the client-server systems; however, routers which can manage multicasting, are not widely deployed. Peer-to-peer (P2P) systems with overlay structures do not need a central server like conventional client-server systems [11, 12]. Network coding can also be used with peer-to-peer (P2P) streaming systems. One of the main issues in using P2P streaming is the degree of heterogeneity of the receivers in terms of bandwidth, CPU capacity and screen resolution; and limited upload capacity of the peers.

3 Conclusion

It is expected that the simulation study findings will indicate improved performance with appropriate combinations of video encoding and network coding approaches deployed on peer-to-peer streaming systems in a wireless environment. One expects the scalable video coding approach to effectively deal with the heterogeneous receiver problem for P2P systems. One also expects network coding to provide higher throughput to the system [13]. OpNET's Modeler networking simulation environment will be used in modeling the problem and conducting the study.

4 References

[1] Xiaoqing Zhu and Bernd Girod. Video Streaming Over Wireless Networks, Proc. European Signal Processing Conference (EUSIPCO'07), pp. 1462-1466, Poznan, Poland, September 2007. Invite Tutorial Paper.

[2] Zheng Wang, Huifang Chen, Lei Xie and Kuang Wang. Cross-layer Design for Wireless Video Stream Transmission, Wireless Communications and Networking Conference (WCNC), pp. 1-6, Sydney, NSW, April 2010.

[3] M. Ghaderi, D. Towsley and J. Kurose. Reliability benefit of network coding, Tech. Report 07-08, Computer Science Department, University of Massachusetts Amherst, February 2007.

[4] Wang, B., Wei, W., Guo, Z., and Towsley, D.F. Multipath live streaming via TCP: Scheme, performance and benefits. In Proceedings of TOMCCAP. 2009.

[5] Sunand Tullimas, Thinh Nguyen, Rich Edgecomb, and Sen-ching Cheung. 2008. Multimedia streaming using multiple TCP connections. ACM Trans. Multimedia Comput. Commun. Appl. 4, 2, Article 12 (May 2008), 20 pages. DOI=10.1145/1352012.1352016 http://doi.acm.org/10.1145/1352012.1352016

[6] An Chan, Sung-Ju Lee, Xiaolin Cheng, Sujata Banerjee, and Prasant Mohapatra. 2008. The impact of link-layer retransmissions on video streaming in wireless mesh networks. InProceedings of the 4th Annual International Conference on Wireless Internet (WICON '08). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, , Article 55 , 9 pages.

[7] Leonardo Badia, Nicola Baldo, Marco Levorato and Michele Zorzi. A Markov Framework for Error Control Techniques Based on Selective Retransmission in Video Transmission over Wireless Channels. IEEE Journal on Selected Areas in Communications, pp. 488-500, April 2010.

[8] Supavadee Aramvith, Chia-Wen Lin and Ming-Ting Sun. Wireless Video Transport Using Conditional Retransmission and Low-Delay Interleaving. IEEE Transactions on Circuits and Systems for Video Technology, Volume: 12 Issue: 6, pp. 558 – 565, Jun 2002.

[9] Árpád Huszák, Sándor Imre. Content-Aware Selective Retransmission Scheme in Heavy Loaded Wireless Networks IFIP International Federation for Information Processing, 2008, Volume 284/2008, 123-134, DOI: 10.1007/978-0-387-84839-6_10

[10] Yu, H. , Yu, S. , & Wang, C. (2004). A Highly Efficient, Low Delay Architecture for Transporting H.264 Video over Wireless Channel. Signal Processing: Image Communication, 19(4), 369-385.

[11] Hulya Seferoglu, Ozgur Gurbuz, Ozgur Ercetin and Yucel Altunbasak. Video Streaming to Multiple Clients Over Wireless Local Area Networks. IEEE International Conference on Image Processing, pp.1681-1684, October 2006.

[12] M. U. Demircin, "Robust Video Streaming Over Time-Varying Wireless Networks", Thesis to Georgia Institute of Technology, 2008

[13] Shabnam Mirshokraie and Mohamed Hefeeda. 2010. Live peer-to-peer streaming with scalable video coding and networking coding. In Proceedings of the first annual ACM SIGMM conference on Multimedia systems (MMSys '10). ACM, New York, NY, USA, 123-132.

[14] Christos Grecos and Qi Wang. 2010. Video networking: trends and challenges. In Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia (MoMM '10). ACM, New York, NY, USA, 17-24. DOI=10.1145/1971519.1971524

[15] Hulya Seferoglu and Athina Markopoulou. 2009. Video-aware opportunistic network coding over wireless networks. IEEE J.Sel. A. Commun. 27, 5 (June 2009), 713-728.

[16] F.H.P. Fitzek and S. Hendrata and P. Seeling and M. Reisslein. 2004. Chapter in Wireless Internet -- Video Streaming in Wireless Internet. CRC Press.

[17] John G. Apostolopoulos. Video Compression. 2001.

http://www.hpl.hp.com/personal/John_Apostolopoulos/MITSpring2001/lecture1_video_compression.pdf

[18] PREDICTION METHODS FOR MPEG-4 AND H.264 VIDEO TRANSMISSION Journal of Electrical Engineering [1335-3632] Pilka yr:2011 vol:62 iss:2 pg:57 -64

[19] Ron Shmueli, Ofer Hadar, Revital Huber, Masha Maltz, and Merav Huber. Effects of an Encoding Scheme on Perceived Video Quality Transmitted Over Lossy Internet Protocol Networks. IEEE Transactions on Broadcasting, vol. 54, issue: 3, pp.628-640, September 2008.

[20] Error Resilient Coding and Error Concealment in Scalable Video Coding IEEE transactions on circuits and systems for video technology [1051-8215] Guo yr:2009 vol:19 iss:6 pg:781 -795

[21] Nicholas Mastronarde, Deepak S. Turaga, and Mihaela van der Schaar. Collaborative Resource Exchanges for Peer-to-Peer Video Streaming Over Wireless Mesh Networks.

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 25, NO. 1, JANUARY 2007

[22] Cheng-Hsin Hsu and Mohamed Hefeeda. 2011. A framework for cross-layer optimization of video streaming in wireless networks. ACM Trans. Multimedia Comput. Commun. Appl. 7, 1, Article 5. DOI=10.1145/1870121.1870126

[23] Eli Brosh, Salman Abdul Baset, Dan Rubenstein, and Henning Schulzrinne. 2008. The delay-friendliness of TCP. SIGMETRICS Perform. Eval. Rev. 36, 1 (June 2008), 49-60. DOI=10.1145/1384529.1375464

[24] Schierl, T.; Gruneberg, K.; Wiegand, T.; , "Scalable video coding over RTP and MPEG-2 transport stream in broadcast and IPTV channels," Wireless Communications, IEEE , vol.16, no.5, pp.64-71, October 2009 doi: 10.1109/MWC.2009.5300304

TA-MAC: A Traffic-Adaptive MAC Protocol for Asynchronous Wireless Sensor Networks

Hao Wang[1], Vyacheslav Zalyubovskiy[2], Vladimir Shakhov[3], and Hyunseung Choo[1*]
[1]School of Information and Communication Engineering
Sungkyunkwan University, Suwon, South Korea
wanghao@skku.edu, choo@ece.skku.ac.kr
[2]Sobolev Institute of Mathematics RAS Novosibirsk, Russia
slava@math.nsc.ru
[3]Informatics Systems Section ICMMG SB RAS Novosibirsk, Russia
shakhov@rav.sscc.ru

Abstract— *Wireless Sensor Network MAC protocols usually employ periodic sleep and wakeup, achieving low duty-cycle to save energy and to increase the lifetime of battery-powered sensor devices. Previously proposed PW-MAC (Predictive Wakeup MAC) protocol is based on asynchronous duty cycling. PW-MAC minimizes sensor node energy consumption by enabling senders to predict receiver′s wakeup times. However, PW-MAC protocol performs poorly in the drastically changed traffic wireless sensor networks. In such environment, the duty cycles of sensor nodes should be adjusted appropriately to ensure energy efficiency while meeting traffic demands and keeping low latency. In this paper, we propose a TA-MAC (Traffic-Adaptive MAC) protocol that dynamically adjusts node′s duty cycle based on the traffic to reduce packet transmission delay and save the sensor node′s energy.*

Keywords: Dynamical duty cycle; Traffic load; MAC Protocol; WSNs

1. Introduction

Modern technology has significantly improved the development of wireless sensor networks (WSNs). In wireless sensor networks (WSNs), energy consumption is one of the most important factors because it is difficult to recharge or replace the battery of each sensor node. Several duty cycle-based MAC protocols have been proposed [1]-[3] to maximize the energy conservation. These protocols focus on reducing the idle listening as much as possible. Each sensor node turns its radio on only periodically, alternating between active and sleeping states. The average duty cycle measures the ratio of the time a node is awake to the total time. For example, with a 5% duty cycle, a node has its radio on only 5% of the time, resulting in substantial energy savings. When being active, a node is able to transmit or receive data, whereas when sleeping, the node completely turns off its radio to save energy, duty cycles of 1-10% are typical in order to maximize energy saving.

There are two types of duty cycle MAC protocols: synchronous and asynchronous. Synchronous approaches [4] synchronize neighboring nodes in order to align their active or sleeping periods. Neighbor nodes start exchanging a packet only within the common active time, enabling a node to sleep for most of the time in an operational cycle without missing any incoming packet. This approach greatly reduces idle listening time, but the required synchronization introduces extra overhead and complexity, and a node may need to wake up multiple times if its neighbors are on different schedules. Existing asynchronous approaches [3], [5], [6], on the other hand, allow nodes to operate independently, each on its own duty cycle schedule, by employing low power listening (LPL). In LPL, prior to data transmission, a sender transmits a short preamble lasting at least as long as the sleep period of the receiver. When the receiver wakes up and detects the preamble, it stays awake to receive the data.

Existing duty cycle MAC protocols, both synchronous and asynchronous ones, are mainly optimized for light and stable traffic loads. A WSN, however, could often experience bursty and high traffic loads. For example, either broadcast [7] or convergecast traffic [8]-[10] could suddenly increase channel contention in a local neighborhood. In WSNs, broadcast is widely used for various network wide queries and updates [11], and convergecast is often observed when multiple sensors that have detected the same event to send their reports to the sink node or to a node that does data aggregation. As existing approaches are mainly optimized for stable or light traffic loads, we found that they become less efficient in latency, power consumption, and packet delivery ratio as traffic load increases. As traffic in a WSN can be quite dynamic, depending on the events being sensed and the sensing application and protocols being used, an idea of WSN MAC protocol should perform well under a wide range of traffic loads, including high loads and bursty traffic.

*Corresponding author.

In this paper, we propose a Traffic-Adaptive MAC Protocol (TA-MAC) to achieve low energy consumption in WSNs without sacrificing performance, such as transmission latency or throughput. Like duty cycle asynchronization-based protocols, TA-MAC allows nodes to set their own sleep/wake schedules independently. TA-MAC also tries to increase the throughput and to decrease the transmission delay by adjusting the sensor′s duty cycle based on the traffic load. The transmission delay can therefore be shortened dramatically. If traffic is light, the sensor′s duty cycle is decreased, so the energy consumption is decreased; otherwise, the sensor′s duty cycle is increased, and the transmission delay is reduced. In this way, channel utilization and throughput are improved. Our contributions include the following:

- We present the TA-MAC (Traffic-Adaptive MAC Protocol) which is designed to minimize sensor node energy consumption and packet transmission delay by enabling sensor to dynamically adjust its duty cycle based on the traffic load. TA-MAC reduces the packet transmission delay when the traffic load is high.
- We show that TA-MAC also achieves significant improvement in energy savings during very low traffic load, compared to the current duty cycle protocols, as well as to reduce energy consumption at high traffic load.
- The performance of TA-MAC is evaluated through simulations and compared in detail with the existing MAC protocols.

The remainder of this paper is organized as follows. In Section 2, we discuss related work. Section 3 presents the preliminaries while Section 4 presents TA-MAC Protocol. Section 5 demonstrates the performances through simulation. Finally, we conclude our work in section 6.

2. Related Work

The MAC protocols that use a duty-cycle in a wireless sensor network can be classified into 2 basic types: the synchronous method and the asynchronous method. A synchronous MAC protocol adjusts the awake period by using a visual synchronous method between groups where data communication takes place. Examples include S-MAC [2], T-MAC [4], and DMAC [12]. However, these protocols require time synchronization, which causes control message overhead and makes sensor nodes more complex and expensive. On the other hand, in asynchronous duty cycle MAC protocols, each sensor node wakes up and sleeps independently. Thus, time synchronization is not necessary.

Most asynchronous duty cycle MAC protocols adopt a random wake-up interval in order to avoid repeated collisions. Given that sensor nodes wake up at different times with random wakeup intervals, it is necessary to ensure that a sender and its intended receiver are active at the same time period to transmit data. To do this, preamble-based protocols were proposed as B-MAC [13], Wise-MAC [5], and X-MAC [6]. A receiver-initiated duty cycle MAC protocol, RI-MAC, was proposed in [14]. PW-MAC [15] is an another MAC protocol based on asynchronous duty cycling. PW-MAC minimizes sensor energy consumption by enabling senders to predict receiver wakeup times.

Wise-MAC [5] is also based on preamble transmission. However, after data transmission, a receiver sends an ACK message including the time remaining until its next wake-up time. Thus, the sender can wake up and start to send a preamble just before the receiver wakes up. As the time duration of the preamble transmission is short, Wise-MAC reduces energy consumption and improves channel utilization. However, repeated collisions can occur because this protocol assumes a periodic wakeup time. In addition, simultaneous preamble transmissions from hidden nodes degrade its performance.

In RI-MAC [14], when a packet arrives at a sender, it wakes up and simply waits for a base beacon from its intended receiver. When the receiver wakes up, it sends a base beacon as an invitation for data transmission. If the sender receives the base beacon, data transmission is started. This protocol uses a short beacon message instead of preambles which waste the wireless medium. Thus, it decreases energy consumption and delay significantly. However, the idle listening time of senders is long. In other words, a sender should wake up and remain active until the intended receiver sends a beacon message.

PW-MAC [15] is designed to minimize sensor nodes′ energy consumption by enabling senders to predict receiver′s wakeup times. PW-MAC achieves very high energy efficiency by minimizing idle listening and overhearing. And the PW-MAC reduces both the senders′ and receiver′ duty cycle compare to the previous MAC protocols. We will detailedly present the predictive wakeup process in Section 3.

3. Preliminaries

In this paper, we use the same predictive-wakeup model as in [15]. A node does not explicitly send its wakeup times to other nodes. Instead, a sender independently deduces future wakeup times of a receiver based on the sender′s knowledge of the receiver′s pseudo-random wakeup-schedule generator. Different nodes use different parameters for their pseudo-random number generators to avoid nodes persistently generating the same numbers. If the parameters of the pseudo-random number generator of a node R are learned by another node S, S can deduce the values of all future pseudo-random numbers generated by R.

In PW-MAC, to enable a sender to accurately predict the wakeup times of a receiver, they require every node to compute its wakeup times using its pseudo-random wakeup-schedule generator. For the sake of simplicity, they use

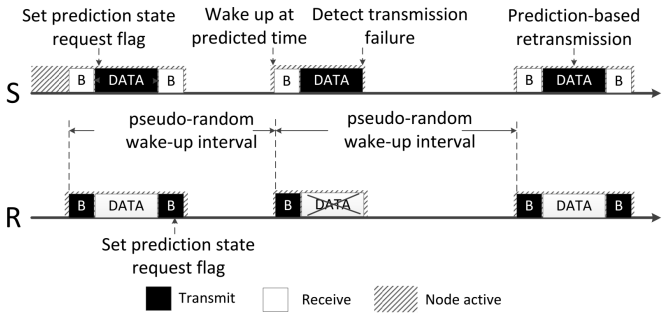


Fig. 1: The predictive-wakeup mechanism of PW-MAC

the following linear congruential generator to generate the pseudo-random numbers as the sensor′s wakeup intervals:

$$X_{n+1} = (aX_n + c)\ mod\ m, \tag{1}$$

where $m > 0$ is the modulus, a $(0 < a < m)$ is the multiplier, c $(0 < c < m)$ is the increment, and X_n $(0 < X_n < m)$ is the current seed. Each X_{n+1} generated can be used as a pseudo-random number and becomes the new seed.

The prediction state of R learned by S comprises the parameters and current seed of the pseudo-random number generator of R (6 bytes in total), as well as the current time difference between S and R (4 bytes).

Figure 1 illustrates the predictive-wakeup mechanism of PW-MAC. Each node periodically wakes up and broadcasts a beacon, denoted as B in the Figure 1, to announce that it is awake and ready to receive DATA packets. In PW-MAC, as shown for node R in this figure, the interval between wakeups is calculated by using equation (1). When S has a packet to send to R, S does not have the prediction state of R, S turns on its radio and waits for a beacon from R. After receiving R's beacon, when S transmits the DATA packet, S sets a special flag in the DATA packet header to request R's prediction state. Once receiving this DATA packet, R sends another beacon that serves both to acknowledge the DATA packet reception (i.e., an ACK beacon) and to allow additional DATA packets to be sent to R. In response to the prediction state request from S, R also embeds its current time and prediction state in the beacon. The current time of R is used by S to compute the time difference between S's and R's clocks. With the prediction information received from the ACK beacon, S can predict future wakeup times of R. In the future, if S has another DATA packet for R, S wakes up shortly before the predicted wakeup time of R, as illustrated in Figure 1.

In sensor networks, latency has been a key factor affecting the applicability of sensor networks to some delay-sensitive applications, such as those used in health and military areas. With PW-MAC protocol, the sensors use the pseudo-random numbers as the wakeup interval. Since they are random numbers, it is inadaptable to bursty or high traffic load.

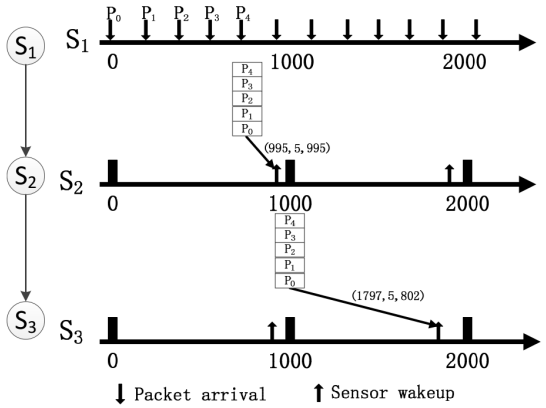


Fig. 2: The problem of PW-MAC in high traffic load

Figure 2 shows an example of data transmission with PW-MAC in high traffic load. We use a triple to denote the wakeup time (T_w), the number of received packets (R_n) and the maximum packet delay (D_{max}) for each sensor wakeup, (T_w, R_n, D_{max}). In Figure 2, node S_2 needs to receive data from S_1 and then transmit the data to node S_3. By using the different parameters to generate the pseudo random number as the wakeup interval, these three nodes have the different wakeup schedule. In the high traffic load, node S_1 generates many packets and put the packets in its buffer. By using PW-MAC, node S_1 predictes node S_2 wakeup time. When S_2 wakes up at time 995, S_1 sent the 5 buffered packets to S_2 at once time. So for node S_2 at time 995, the wakeup time (T_w) is 995, the number of received packets (R_n) is 5 and the maximum packet delay (D_{max}) is 995. For the same reason, for node S_3 at time 1797, the wakeup time (T_w) is 1979, the number of received packets (R_n) is 5 and the maximum packet delay (D_{max}) is 802. By observing this example, we know that PW-MAC has a large packet transmission delay. Moreover, the sender first put the packets in the buffer, so in the high traffic load, the number of packets in the buffer will increase significantly, then the probability of leading to collision will significantly increase. Obviously, by using PW-MAC, the energy efficiency is very low in the low traffic load. Because in the low traffic load, there are not many packets for transceiving, the sensors just follow their schedule and wake up. After doing the idle listening, the sensors go to sleep. The main energy consumption is idle listening and the energy efficiency is low.

4. Proposed Scheme

Reducing the duty cycle of sensors can save energy to prolong sensors′ life when sensors are in low traffic loads. Extending the duty cycle of sensors can improve data collection efficiency and reduce transmission delay when sensors are in high traffic loads. To overcome the aforementioned issues, we propose a dynamic traffic-aware duty cycle adjustment MAC protocol, TA-MAC, that can adjust

duty cycle adaptively according to status of sensor′s traffic load. TA-MAC is based on the PW-MAC protocol, with more functions are designed by us to enhance it.

4.1 Receiver-Initiated MAC Protocol

TA-MAC uses the receiver-initiated approach. A receiver-initiated asynchronous approach has better performance during high traffic load due to the avoidance of unnecessary channel occupancy and provision of back-to-back data transmission. Each node periodically wakes up according to its own wakeup interval to receive packets from the upstream nodes. When radio turns on, a node broadcasts a beacon, after performing a short backoff to avoid the collision from simultaneous beacon packet transmissions by the neighboring nodes. It then waits to receive packets from any of its upstream senders until a time-out occurs. The receiver goes to sleep if it detects the channel busy during beacon transmission or detects collision during waiting for packets or no packets arrived after a time-out occurs. If a node finds the channel busy for a consecutive number of times, it might shift its wakeup time to avoid the ongoing transmission, which coincides with the wake-up time of that node.

4.2 Determination of wakeup interval

For adapting the duty cycle based on the traffic load, we modify the linear congruential generator formula to generate the pseudo random as the sensor wakeup interval, as shown below:

$$X_{n+1} = (aX_n + c)\ mod\ m_1 \tag{2}$$

$$X'_{n+1} = j_i + X_{n+1} \tag{3}$$

where a, c and m_1 have the same meaning with the parameters in the equation (1), $m_1 < m$. In the equation (3), $j_i \in \{j_{min}, j_{nor}, j_{max}\}$, and $j_{min} < j_{nor} < j_{max}$. To avoid repeated collisions, different nodes use different parameters for their pseudo random generator. By using different values of j_i, we can dynamically change the sensor node wakeup interval and adjust the node′s duty cycle based on its traffic load.

4.3 Operation of TA-MAC Protocol

In TA-MAC, we have estimated multi-level traffic load-status by taking the radio capacity usage into account. We have included a simplified traffic load estimation technique in which each sensor node measures the load based on its capacity usage. Furthermore, we classify the traffic load status into the following categories: low traffic load (L), normal traffic load (N) and high traffic load (H). Figure 3 shows that the traffic load state transition diagrams.

We use the same duration T as the measurement period. Using equation (4), we calculate the maximum traffic load TL_{max} during the period T.

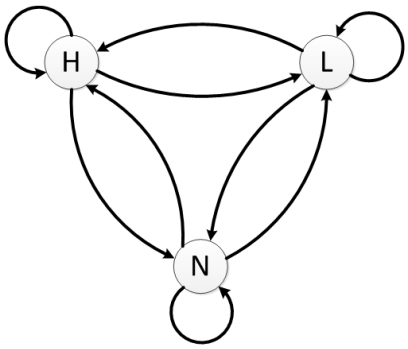


Fig. 3: Traffic load state transition diagrams

$$TL_{max} = \frac{T \times B}{l} \tag{4}$$

where B is the bandwidth of the channel and l is the length of a packet, they are constant. Each sensor node calculates its traffic load TL_t during the period T. Based on the comparison of TL_{max} and TL_t, the sensor node dynamically adjusts its duty cycle according to the following rules:

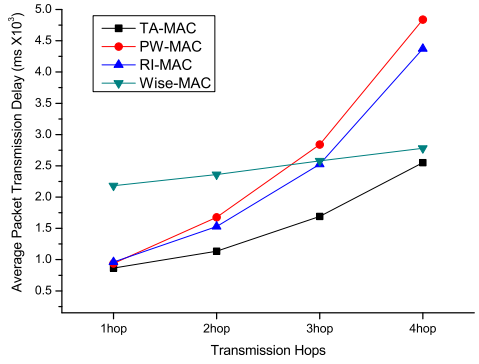
- If $TL_t > TL_{max} \times \gamma$, the current traffic load is high, the sensor node should use j_{min} in the equation (2) and increase its duty cycle to reduce the collision and packet transmission delay.
- If $TL_t < TL_{max} \times \beta$, the current traffic is low, the sensor node should use j_{max} in the equation (2) and decrease the duty cycle to save energy.
- Otherwise $TL_{max} \times \beta \leq TL_t \leq TL_{max} \times \gamma$, the current traffic load is normal, the sensor node should use j_{nor} in the equation (2) and keeps its duty cycle.

The categorization of the traffic load status based on the parameter value of β and γ is a critical task because the choice of load transition values requires a trade-off between capacity utilization and congestion/load awareness. A small value of β keeps the network in a low congestion and contention state, in which the capacity utilization would be at a minimum. Conversely, although a large transition value of γ can ensure better utilization of the capacity, it may quickly overload the network, resulting in more congestion and collision losses. Therefore, we have to choose suitable load transition values such that optimal capacity utilization and congestion avoidance can be guaranteed.

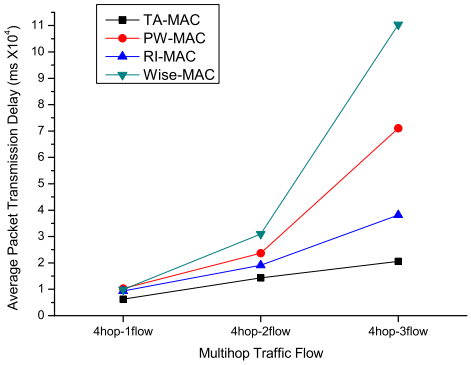
5. Performance Evaluation

In this section, we evaluate TA-MAC by comparing its performance with that of Wise-MAC [5], RI-MAC [14] and PW-MAC [15]. And the following metrics are measured in our evaluation:

- Average data packet delivery latency: the average time cost by each delivered DATA packet from the source to the destination.

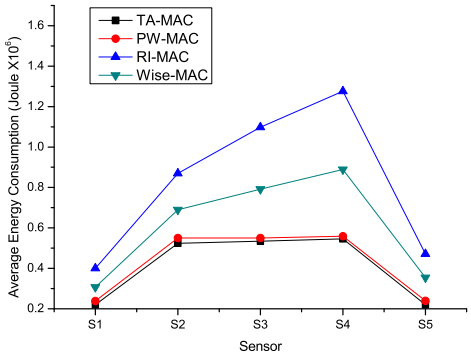


(a)

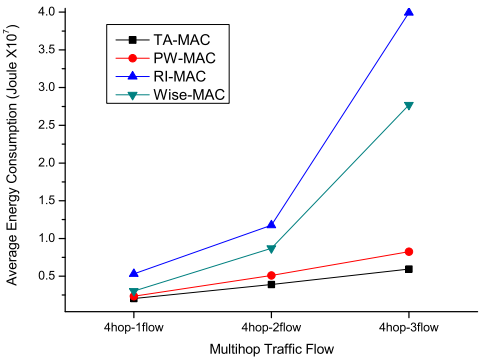


(b)

Fig. 4: Average packet transmission delay in scenario 1



(a)



(b)

Fig. 5: Average energy consumption in scenario 1

- Average energy consumption: the energy consumption cost by each sensor node in the simulation period.

5.1 Simulation Environment

To evaluate the effectiveness of the proposed protocol, we implement a simulator with Java. The parameters a, c, m and m_1 of a node's pseudo-random number generator were configured as node ID $\times$ 20, 7, 1000 and 100, respectively. The parameters j_{min}, j_{nor} and j_{max} are 400, 900 and 1400, respectively. Based on the analysis in [16], we select 0.25 and 0.75 as the low and high traffic load threshold. Table 1 shows the energy model parameters. These simulations were conducted on 15 sensor nodes in a 3$\times$ 5 grid topology. There are from 1 to 3 concurrent multihop traffic flows, with hop count ranging from 1 to 4. Each traffic flow traverses a distinct multihop path. To evaluate the influences of wireless collisions on packet transmissions, neighboring sensors were placed within the radio interference range of each other. We use the energy cost model as in [17]. In this model, the total energy consumption of each node is calculated by the following formula:

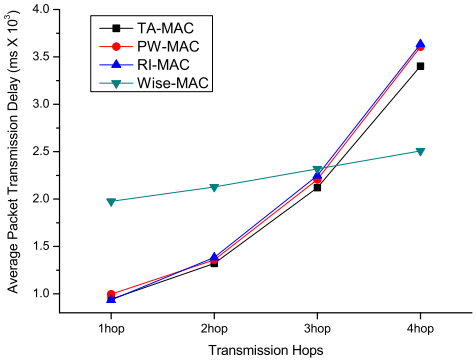
$$E = N_{send} \times T_{send} \times P_{send} + N_{rec} \times T_{rec} \times P_{rec} + T_{sleep} \times P_{sleep} + T_{idle} \times P_{idle} \quad (5)$$

where N_{send} and N_{rec} denote the number of packets sending and receiving by the sensor node. Table 1 shows the parameter values which are used in the simulation.

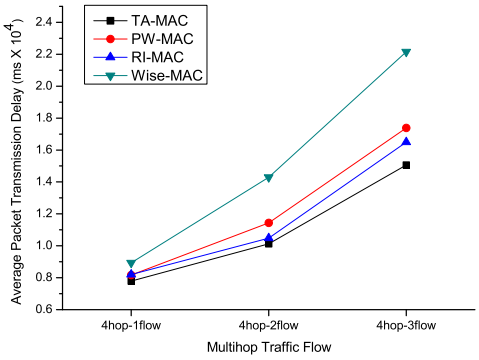
Table 1: Energy model parameters

Symbol	Meaning	Value
P_{send}	Energy consumption for sending packet per 1 ms	60mw
P_{rec}	Energy consumption for receiving packet per 1 ms	45mw
P_{sleep}	Energy consumption in sleeping state per 1 ms	90μmw
P_{idle}	Energy consumption in idle listening state per 1 ms	45mw
T_{send}	The time required for sending one packet	2.5 ms
T_{rec}	The time required for receiving one packet	2.5 ms

In the simulation, packet arrivals follow a Poisson process. The average inter-arrival time between two consecutive packets generated by a given sensor is 200ms. Traffic arrival is assumed to be Poisson with a rate being λ packets per

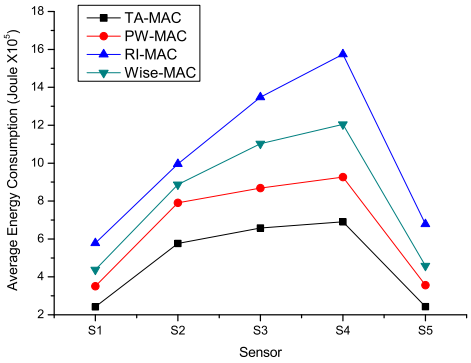


(a)

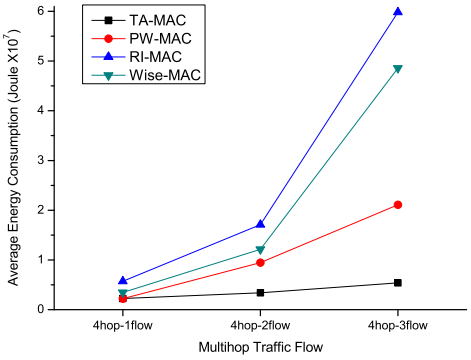


(b)

Fig. 6: Average packet transmission delay in scenario 2



(a)



(b)

Fig. 7: Average energy consumption in scenario 2

node per 200ms. The results are computed as the average of 100 rounds. Each round lasts 120s. In scenario 1, we set the value of λ to 100 for generating the low traffic from 0s to 30s and from 90s to 120s. Between 31s and 89s, we set the value of λ to 300 for generating the high traffic. In scenario 2, we set the value of λ to 100 for generating the low traffic from 0s to 45s and from 75s to 120s. Between 46s and 74s, we set the value of λ to 300 for generating the high traffic.

5.2 Simulation Results

The first part of each graph in Figures 4-7 show the performance of the four protocols when there is one multihop traffic flow, with lengths ranging from 1 to 4 hops. The second part of each graph in Figures 4-7 show the performance of these protocols when there are one to three concurrent 4-hop traffic flows. A multihop traffic flow traverses multiple intermediate forwarders to reach the destination. The destination nodes only receive packets and do not generate or forward packets.

Figure 4 shows the average packet transmission delay and average energy consumption in the scenario 1 which is mainly high traffic load. With the traffic hop and traffic flow increasing, the performance of all four MAC protocols decreases. Figure 4(a) shows the performance when there is only one multihop traffic flow, TA-MAC achieves a smaller packet transmission delay than Wise-MAC, RI-MAC and PW-MAC. This is because that TA-MAC increases the sensors duty cycle when the traffic load increase, then the sensor′s wakesup interval reduces and receives the buffered packets earlier. Moreover, with the hops increasing, the packet transmission delay of Wise-MAC just has a small increase due to that the sensor nodes use the fixed wakeup interval in Wise-MAC. In Figure 4(b), when the number of concurrent multihop traffic flows increased, the performance of Wise-MAC, PW-MAC and TA-MAC degraded significantly. This performance degradation is due to the following reason: as the number of senders increases from 4 to 12, so does the probability of having node wakeup schedule collisions.

Figure 5 shows the average energy consumption in the scenario 1. For energy consumption, S_1 just sends the packet and S_5 as the destination just receives packets. However,

S_2 - S_4 as the intermediate packet forwards receive and transmit the packets, so S_2 - S_4 spend more energy than the S_1 and S_5. In Figure 5(a) and (b), with the traffic hops and flows increasing, all four MAC protocols consume more energy. The RI-MAC has the highest energy consumption. The reason is that when a sender in RI-MAC has a packet to send, it immediately wakes up to wait for the receiver, leading to a large sender duty cycle due to its idle listening until the receiver wakes up. TA-MAC and PW-MAC have the similar performance of energy consumption. Because of the collision, PW-MAC has to spend more energy on retransmitting. However, in the scenario 1, TA-MAC has to increase the wakeup times and duty cycle, so it also costs some more energy to reduce the packet transmission delay.

Figure 6 shows the average packet transmission delay in the scenario 2. In scenario 2, the main traffic load is low traffic load and the probability of leading to collision is low. So in Figure 6(a), the performance of packet transmission delay is almost same among RI-MAC, PW-MAC and TA-MAC. For Wise-MAC, the sensor nodes use the fixed wakeup interval, in one traffic flow, average packet transmission delay for each sensor are almost same. Moreover, in Figure 6(b), with the number of flows increasing, Wise-MAC has the highest average packet transmission delay compare to other MAC protocols due to the fixed sensor wakeup interval. TA-MAC achieves the lowest packet transmission delay in one and multiple flows.

Figure 7 shows the average energy consumption in the scenario 2. In the low traffic load, the main part of energy consumption is the short idle listening after the sensor wakes up. According the Figure 7, TA-MAC achieves smaller energy consumption than Wise-MAC, RI-MAC and PW-MAC. This performance is due to the following reason: in the low traffic load, TA-MAC increases the sensors′ duty cycle, then let the sensors reduce the number of wakeup times and sleep long to reduce the energy consumption. As the same reason as in the scenario 1, RI-MAC has the highest energy consumption in the scenario 2.

6. Acknowledgement

This research was supported in part by MKE(NIPA,KEIT) and MEST(NRF), Korean government, under ITRC NIPA-2012-(H0301-12-3001), IT R&D Program[10041244, SmartTV 2.0 Software Platform], and PRCP(2011-0018397), respectively.

7. Conclusion

In order to reduce the energy consumption when nodes are in idle listening, duty-cycle-based MAC protocols are introduced to let node go into sleep mode periodically. The existing duty cycle protocols (synchronous or asynchronous) do not maintain their duty cycle based on the availability of traffic, which cause unnecessary wake up and energy wastage in case of changing traffic. In this paper, a dynamical traffic-aware MAC protocol for reducing packet transmission delay and increasing energy efficiency in wireless sensor networks is proposed. The proposed TA-MAC protocol provides better data transmission delay when sensors are in high traffic load. Meanwhile, the TA-MAC protocol saves energy when sensors are in low traffic load. Simulation results show that the TA-MAC has better performance in terms of packet transmission delay and energy consumption than previously known MAC protocols.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, pp. 102–114, 2002.

[2] W. Ye, S. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc.IEEE INFOCOM'02*, pp. 1567–1576, 2002.

[3] J. Polastre, J. Hill, and D. Culler, "Versatile Low Power Media Access for Wireless Sensor Networks," in *Proc.ACM SenSys'04*, pp. 95–107, 2004.

[4] T. Van. Dam, and K. Langendoen, "An adaptive energy efficient mac protocol for wireless sensor networks," in *Proc.ACM SenSys'03*, pp. 171–180, 2003.

[5] A. El-Hoiydi, and J. D. Decotignie, "Low Power Downlink MAC Protocols for Infrastructure Wireless Sensor Networks," *Mobile Networks and Applications*, vol. 10, pp. 675–690, 2005.

[6] M. Buettner, V. Yee. Gary, E. Anderson, and R. Han, "X-MAC: A Short Preamble MAC Protocol for Duty-Cycled Wireless Sensor Networks," in *Proc.ACM SenSys'06*, pp. 307–320, 2006.

[7] S. Ni, Y. Tseng, Y. Chen, and J. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," in *Proc.ACM MobiCom'99*, pp. 151–162, 1999.

[8] H. Zhang, A. Arora, Y. Choi, and M. G. Gouda, "Reliable Bursty Convergecast in Wireless Sensor Networks," in *Proc.ACM MobiCom'05*, pp. 266–276, 2005.

[9] A. S. Rodionov, H. Choo, H. Youn, and V. V. Shakhov, "Discord Detection for a Process with a Predefined Interval of Observations," *International Journal of Computer Mathematics*, vol. 80, pp. 181–191, 2003.

[10] V. V. Shakhov, H. Choo, H. Youn, and Y. Bang, "Discord model for detecting unexpected demands in mobile networks," *Future Generation Computer System*, vol. 20, pp. 181–188, 2004.

[11] F. Stann, J. Heidemann, R. Shroff, and M. Z. Murtaza, "RBP: Robust Broadcast Propagation in Wireless Networks," in *Proc.ACM SenSys'06*, pp. 85–98, 2006.

[12] G. Lu, B. Krishnamachari, and C. S. Raghavendra, "An adaptive energy efficient and low-latency MAC for data gathering in wireless sensor networks," in *Proc.IEEE IPDPS'04*, pp. 26–30, 2004.

[13] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proc.ACM SenSys'04*, pp. 95–107, 2004.

[14] Y. Sun, O. Gurewitz, and D. Johnson, "RI-MAC: a receiver-initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks," in *Proc.ACM SenSys'08*, pp. 1–14, 2008.

[15] L. Tang, Y. Sun, O. Gurewitz, and D. Johnson, "PW-MAC: An Energy- Efficient Predictive-Wakeup MAC Protocol for Wireless Sensor Networks," in *Proc.IEEE INFOCOM'11*, pp. 1305–1313, 2011.

[16] C. Y. Wan, S. B. Eisenman, and A. T. Campbell, "Overload Traffic Management for Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 3, pp. 18–38, 2007.

[17] Y. W. WU, X. Y. Li, and Y. H. Lin, "Energy-Efficient Wake-Up Scheduling for Data Collection and Aggregation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, pp. 275–287, 2010.

Performance Improvement of Optimized Link State Routing (OLSR) Protocol

Navaid Akhter[1], Ammar Masood[2], Irfan Laone[3]
Institute of Avionics and Aeronautics/Department of Avionics, Air University, Islamabad, Pakistan [1,2,3]

Abstract- *OLSR, a leading proactive protocol of MANET maintains consistent and up to date network topology at all the times and has emerged as the choice for MANETs due to low latency for route determination. Hence, OLSR generate a large amount of control overhead in order to maintain an up-to-date routing table which consumes bandwidth that should have been employed by user data traffic instead. This paper addresses this issue by optimizing OLSR under specific network and mobility conditions which are actually more of practical interest and thereby, our work does have a valuable contribution to provide guidelines for large number of cases of general interest. The proposal has shown to consistently outperform the default implementation by reducing the routing overhead under specific network and mobility conditions considered at no extra cost. Other parameters like data traffic and end-to-end delay also improved with the approach presented in this study which shows the efficiency of the scheme selected.*

Keywords: MANETs, Routing Protocols, OLSR, Improvement in Control Messages' Intervals, Optimality in performance

1 Introduction

Research concerning MANETs is currently of great interest. The performance of MANET is related to the efficiency of the routing protocols in adapting to frequently changing network topology and link status [1]. Because of the importance of routing protocols in the dynamic multi hop networks, a number of routing protocols have been proposed in the last few years; concurrently, a great deal of research work is being undertaken by researchers to improve their performances. In OLSR (a leading MANET routing protocol), maintaining an up-to-date routing table for the entire network calls for excessive communication between the nodes as periodic control messages updates are flooded throughout the network. Hence OLSR generate a large amount of control overhead which consumes valuable bandwidth that should have been employed by user data traffic instead. Therefore, excessive control overhead in OLSR is detrimental to its overall performance in data forwarding, which has been analyzed for improvement in our research work. Other parameters like data traffic and end-to-end delay also improved with the approach presented in this study.

2 MANET Routing Protocols

Mobile Adhoc Network (MANET) is an autonomous system of mobile nodes connected by wireless links [2]. MANET routing protocols are based on how routing information is acquired and maintained by the mobile nodes and thus, can be divided into proactive, reactive and hybrid routing protocols [3]. With proactive routing protocol, nodes in a MANET continuously evaluate routes to all the reachable nodes and attempt to maintain consistent, up-to-date routing information. On the other hand, in reactive routing protocols for MANETs (also called "on-demand" routing protocols), routing paths are explored only when needed. Hybrid routing protocols are proposed to combine the merits of both proactive and reactive routing protocols and to overcome their shortcomings.

3 Optimized Link State Routing (OLSR) Protocol

The Optimized Link State Routing Protocol (OLSR) [4] is developed for MANETs and does not need central administrative system to handle its routing process. Because of its proactive characteristic, the protocol provides all the routing information to all participating hosts in the network at all times. However, as a drawback, OLSR protocol needs that each host periodically send the updated topology information throughout the entire network by flooding. This increases the protocol's bandwidth usage as the routing overhead is high. Although, flooding in OLSR is minimized by the Multi Point Relays (MPRs), which are the only nodes allowed to forward the topological messages [5,6], still the routing overhead is high as compared to reactive routing protocols.

3.1 Control Messages Intervals

OLSR employs two types of control messages: Hello messages and TC messages.

3.1.1 Hello Interval

This parameter represents the frequency of generating a Hello message. Hello Interval determines the time between successive Hello messages, which is set to 2 seconds by default. Hello messages are never forwarded.

3.1.2 TC Interval

This parameter represents the frequency of generating a TC message. In OLSR, the rate of the topological state updates is the sending rate of TC messages. TC messages are broadcasted periodically within the TC interval, to other MPRs, which can further relay the information to further MPRs. TC messages are broadcasted once per refreshing period and the default value is 5 seconds. TC messages are one of the major sources of overhead in OLSR, as they are flooded throughout the network, but they are essential to maintain consistent connectivity knowledge of complete network.

3.2 Problem in OLSR

One advantage of OLSR is that it provides lower route discovery latency than on-demand protocols because of its proactive nature. But the flip side is that OLSR generates a large amount of control overhead which consumes precious bandwidth. Since the resources in wireless networks are severely constrained, the increased channel contention could lead to network congestion resulting in significant lowering of network performance. Further, scalability issues arise in OLSR due to the excessive routing message overhead caused by the increased network population. The size of routing table grows non-linearly with the increase in number of nodes and the control messages can block the actual data packets. Hence, excessive control overhead in OLSR is detrimental to its overall performance in data forwarding and poses a research challenge that need to be addressed.

3.3 Proposed solution

Optimization of local and global topology dissemination intervals (i.e. Hello and TC intervals respectively) is proposed under specific network and mobility conditions which are actually more of practical interest and thereby, our work does have a valuable contribution to provide guidelines for large number of cases of general interest that result in low routing overhead (as compared to the default settings) and thus beneficial for OLSR performance. This study targets on reduction in control overhead with improvement in performance of OLSR by optimizing control messages intervals.

3.3.1 Logical Reasoning of proposed solution

Hello messages are broadcasted periodically for link sensing and neighbor detection. This is also required to complete the MPR selection process. After the MPR selection process is completed, TC messages are generated and are disseminated throughout the network. Subsequent to the receipt of these TC messages, the nodes calculate the routing table and the links are available for data communication. Further, MPRs broadcast the TC messages in the network to maintain a consistent and up-to-date view of complete network topology. In case of any topology changes, the MPR selection process is re-initiated and the routing table is re-computed by the nodes [8]. MANETs require minimum control overhead to reduce channel contention and battery consumption problems. TC messages share a large amount of overhead in OLSR because of its global dissemination nature. Decreasing the broadcast frequency of TC messages reduces the overall routing traffic sent while not incurring any degradation in throughput / end-to-end delay under specific network and mobility conditions as shown in this study. Further, due to frequent topology changes caused by high mobility, the routing information needs to be updated more frequently so as to update the topology changes and guarantee the correctness of route selection. This requires that nodes of OLSR employed MANET detect link changes more quickly and broadcast topology updates with lesser delay. This can be achieved by increasing the Hello messages sending rate for faster response to the link and neighbor changes (especially in case of high mobility scenarios); hence, providing better throughput as compared to the default Hello interval. The increase in routing overhead because of the increase in Hello sending rate above is compensated by the reduction in routing overhead due to the decrease in TC messages sending rate as mentioned earlier.

4 Related Work

The authors of OLSR in RFC 3626 (OLSR) [4] pointed out that the nodes may send control messages at different rates, if beneficial for specific deployment. Many strategies have been proposed by OLSR researchers using different performance metrics to improve the performance of OLSR by varying control messages intervals [7, 8, 9, 10, 11, 12, and 13]. However, these works usually target to reduce control overhead while having certain deficiencies and implementation complexities. Our work, however, does not include any added complexity or depends upon any measurement of network parameters and provides improved performance of OLSR under specific network and mobility conditions by just modifying the OLSR control messages intervals. With our approach, now network is able to achieve an increase in data traffic

received (vis-à-vis the payload with default control messages intervals) while routing traffic and end-to-end delay are both reduced. We have constructed fairly robust scenarios for experiments to investigate the effect of control messages intervals on the routing overhead of OLSR.

5 Performance Evaluation

Because of the unavailability of wide range of real MANETs, the performance analysis of wireless applications or protocols in the context of MANETs often require to be evaluated through simulation studies [14]. The performance analysis on a real network (if available) can be rather tedious if large networks are considered (typically hundreds of nodes). This is why simulation is an important tool in the sense that it can often help to improve or validate protocols [15]. OPNet Modeler 14.5 network simulator was used for analysing the performance of OLSR in this study.

5.1 Choice of Network and Mobility Conditions

The MANET routing protocols perform differently under different network & environmental factors like node mobility, number of nodes, number of source-destination pairs, traffic type, traffic intensity, propagation models etc. For the purpose of performance analysis in this study, we selected two factors: Node mobility and Number of nodes, because of their major impact on the mechanics of the protocol vis-à-vis routing overhead. We started with a carefully designed network scenario for all the experiments and varied one parameter at a time and thus stressed the network in different axis as shown in Table 1.

Table 1
Network & Mobility Factors

S No	Network & Mobility factors	Scenarios
1	Mobility (Speed)	5, 10, 20, 30, 40 mps (18, 36, 72, 108, 144 kph)
2	Number of nodes	50 nodes

As mobility has the most significant impact on MANET routing protocols [16], scenarios have been constructed to evaluate the proposed solution against varying nodes speed (from 5 mps to 40 mps) while keeping the number of nodes fixed. The speed range has been selected with keeping in view the speed selected for extreme practical scenarios (low to high) and most of the other research work in this area. Table 2 depicts the parameters selected for the Scenarios.

Table 2
SIMULATION PARAMETERS

S No	Parameters	Values
1	Traffic	33% of total number of nodes
2	Simulation Area	1500 m x 1500 m
3	Area of movement	Within Network
4	Initial node placement	Random
5	WLAN standard	IEEE 802.11b
6	Transmission Power	0.001W (for 250 m range)
7	Packet Reception Power Threshold (Receiver Sensitivity)	-90 dbm (for 250m range)
8	Data Rate	2 Mbps
9	Mobility Model	Random Way Point (RWP)
10	Propagation Model	Free Space
11	Traffic Type	CBR
12	Packet Size	512 bytes
13	Traffic Intensity / Packet Rate	15 packets / second
14	Simulation Time	30 minutes

5.2 Performance Metrics

The performance metrics investigated during this study were the data traffic received and routing overhead in OLSR protocol vis-à-vis its improvement by optimizing the OLSR control messages interval under various network and mobility conditions. However, end-to-end packets delay was also kept under-check so as to ensure that the optimization of control messages for improvement in routing overhead do not degrade this parameter. The definition of improved performance is that the routing protocol must provide applications with high data traffic received, minimal routing overhead and low end-to-end delay.

6 Results

6.1 Experiments

The simulation study adopts a step-by-step performance optimization approach. Firstly, the impact of each control messages' interval of OLSR has been analyzed distinctly on the data traffic received, routing traffic overhead and the end-to-end packets delay by stressing the mobility factor. The outcome of the results has been analyzed further to see how these control messages' interval can be optimized simultaneously so as to efficiently maximize the data traffic received (payload) while minimizing the routing overhead and end-to-end packets delay. The steps are depicted in Figure 1.

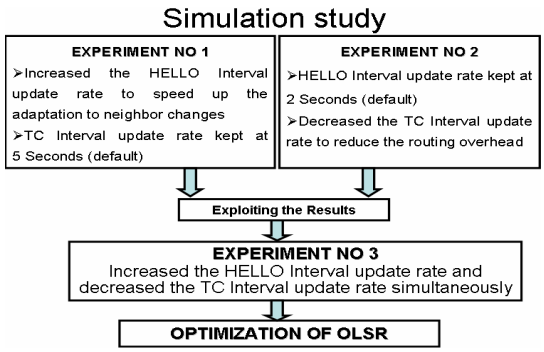


Figure 1: Simulation study steps

6.2 Experiment No 1

In experiment No 1, the Hello message update rate has been increased from default value (2 seconds) to various values like 1.9, 1.8, 1.7, 1.6, 1.5 seconds etc in order to facilitate the routing protocol to speed up the adaptation to neighbor changes while keeping the TC interval at 5 seconds (default value). The value of the state holding timer interval (neighbor hold time) was adjusted accordingly. After various iterations, it was found that Hello interval of 1.8 seconds (TC at default value) provides the best balance between data traffic and routing traffic and the results are shown in Figure 2.

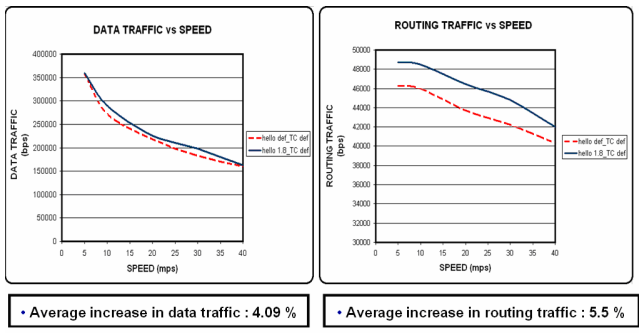


Figure 2: Results of data traffic and routing traffic vs speed with change in Hello interval

In Figure 2, the data traffic and routing traffic are plotted on Y-axis and variation in speed is plotted on X-axis. In all the simulations across specific range of nodes speed while changing the Hello interval and keeping the TC interval fixed, it can be appreciated that increase in Hello sending rate (i.e. Hello 1.8 secs) from the default value (Hello 2 secs) improves the data traffic received as it helps the routing protocol to quickly adapts the changes in neighbors and update the routing tables accordingly. On the other hand, the routing traffic overhead also increases with the increase in Hello sending rate which clearly depict that although fast Hello messages improves the protocol reactivity to link failures; however this is at the cost of increased routing overhead.

Hence, from Experiment No 1, it is concluded that an improvement in data traffic received by increasing the Hello messages sending rate is at the expense of increased routing traffic overhead.

6.3 Experiment No 2

In experiment No 2, the TC interval has been decreased from default value (5 seconds) to various values like 6, 7, 7.5 seconds etc in order to reduce the routing overhead while keeping the Hello interval at 2 seconds (default value). The value of state holding timer interval (topology hold time) was adjusted accordingly. Since the Hello interval is kept constant, the reduction in overall routing overhead is the result of decrease in TC messages overhead. The TC interval of 7.5 seconds was found to be the optimized value under the considered mobility conditions that provide a decrease in routing traffic overhead while having almost no effect on data traffic and the results are as shown in Figure 3.

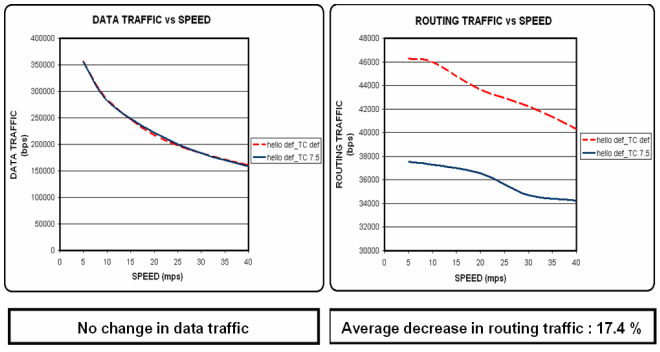


Figure 3: Results of data traffic and routing traffic vs Speed with change in TC interval

In Figure 3, the routing traffic and data traffic are plotted on Y-axis and variation in speed is plotted on X-axis. Firstly observing the routing traffic behavior with default values of OLSR, it is revealed that as the speed increases, the routing traffic decreases. This is because of the reason that as mobility increases, link breakages increases and therefore TC messages are either not generated or if they are generated than they are not forwarded to the entire network. Now with the modified settings, it is evident that as TC messages sending rate is reduced from 5 seconds (default) to 7.5 seconds, the routing traffic is less as compared to the routing traffic at default TC interval. Further, it is observed that decreasing the TC sending rates from default value to 7.5 secs although reduces large routing overhead; brings no significant change in data traffic received. This is because of the fact that repetitive TC messages are broadcasted throughout the network to maintain the network topology. Lowering down the sending rate of these TC messages

under specific mobility conditions although reduces large routing overhead, brings no significant change in data traffic received at the nodes which is more sensitive to the change in Hello interval than the TC interval.

6.4 Outcome of Experiment No 1 and Experiment No 2

Through simulations it has been explored that TC messages generate more overhead than Hello messages because TC messages are forwarded globally to each node in the network while Hello messages are only exchanged locally between neighboring nodes. Increasing the Hello messages sending rate helps the routing protocol to quickly adapt the changes in neighbors and update the routing tables accordingly. Hello interval rate has been increased from default value (2 secs) to 1.8 secs in order to speed up the adaptation to neighbor changes and thus achieving higher data traffic received than what it is achieved at the default value. This is particularly to cater the declined performance of OLSR under high mobility scenarios. Decreasing TC messages sending rate leads to significant reduction in control overhead but do not downgrade the data traffic received under specific mobility conditions considered in the experiments.

6.5 Experiment No 3

Experiments 1 and 2 provides a comprehensive understanding of OLSR's control timers' behavior vis-à-vis performance metrics and gives insightful guidance in optimizing these timers for an improved performance in data traffic received while introducing low routing overhead as compared to the default values. The results have been exploited further in experiment 3 to formulate that how these two timers can be optimized simultaneously under considered network and mobility conditions so as to efficiently minimize the routing overhead while achieving maximum data traffic received without compromising on end-to-end delay. The OLSR control messages intervals Hello 1.8 secs and TC 7.5 secs (as discussed in experiments 1 and 2) were selected and compared against the default intervals to see if there is any improvement as stated above.

6.5.1 Routing traffic vs Speed:

In Figure 4, the routing traffic is plotted on Y-axis and variation in speed is plotted on X-axis.

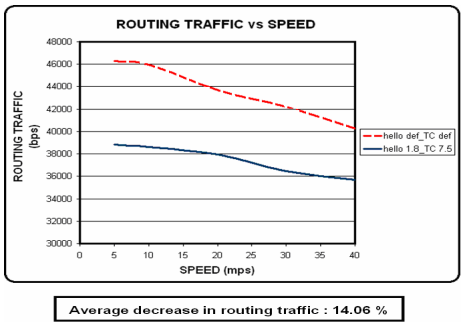


Figure 4: Result of Routing traffic vs Speed with change in HELLO and TC intervals

First observing the routing traffic behavior with default values of OLSR (i.e. Hello def_TC def), it is revealed that as speed increases the routing traffic decreases. This is because of the reason that as mobility increases, link breakages increases and therefore TC messages are either not generated or if they are generated than they are not forwarded to the entire network. This results into decrease in routing traffic as the speed increases and vice versa. Now with the modified intervals (i.e. Hello 1.8_TC 7.5), the similar behavior of decrease in routing overhead with increase in mobility is observed as stated above. Further, as the TC messages sending rate is reduced from 5 seconds (default) to 7.5 seconds, the routing traffic is noticeably reduced as compared to the routing traffic at default TC interval. Also due to the increase in Hello sending rate from 2 seconds (default) to 1.8 seconds, the routing traffic would have increased (as observed in experiment 1). However, this has been compensated with the reduction of large routing overhead due to the decrease in TC interval. Hence, the overall result is the reduction of routing overhead as compared to the routing overhead with default Hello and TC values. The average reduction in routing overhead achieved with the modified Hello and TC intervals is 14.06 %.

6.5.2 Data traffic vs Speed

In Figure 5, the data traffic is plotted on Y-axis and variation in speed is plotted on X-axis.

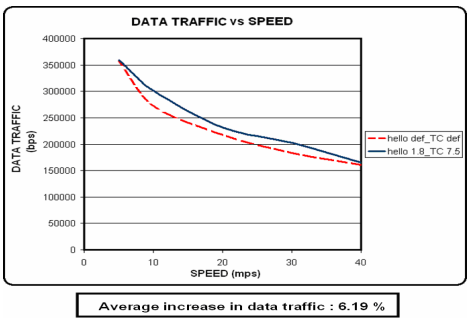


Figure 5: Result of Data traffic vs Speed with change in HELLO and TC intervals

Firstly, observing the data traffic behavior with default values of OLSR (i.e. Hello def_TC def); it is revealed that as speed increases the data traffic decreases. This is because of the reason that as mobility increases, link breakages increases and therefore the nodes are unable to forward the data to the required destination which resulted into dropping of data packets before reaching to the destinations. Now with the modified interval i.e. Hello 1.8_TC 7.5, similar behavior of decrease in data traffic is observed with the increase in node's mobility (due to the same reason as mentioned above). However the data traffic is now improved than what it is achieved with the default OLSR intervals (i.e. Hello def_TC def). This is because of the increase in Hello messages sending rate which speed up the routing protocol's adaptation to neighbor changes and route maintenance and thus resulting into less data drop and increase in data traffic received at the destinations. Further, it is observed that both the curves are tending to converge at very low speeds and at very high speeds. It is because of the reason that at very low speeds, there are no significant changes in neighbors so the default interval as well as modified Hello interval works almost the same manner and the change in Hello interval does not make any difference. Similarly at very high speeds, the topology changes might be too dynamic to be captured by the periodic updates of OLSR with default as well as with the modified settings so the change in Hello interval does not make any significant impact in this regime also. The average increase in data traffic achieved with the modified values of Hello and TC intervals is 6.19%

6.5.3 End to End packets delay vs Speed:

In Figure 6, the delay is plotted on Y-axis and variation in speed is plotted on X-axis. Firstly, observing the end to end packets delay behavior with default values of OLSR (i.e. Hello def_TC def); it is revealed that as speed increases, the end to end packets delay decreases.

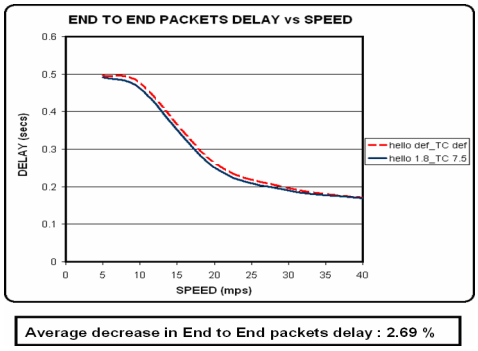


Figure 6: Result of End to End packets delay vs Speed with change in HELLO and TC intervals

This is because of the reason that as mobility increases, link breakages increases and therefore less number of source-destination pairs are now available at high speeds as compared to the scenarios at low speeds. This results into increase in channel capacity because of the occupation of same number of available channels now with less number of source destination pairs. Hence, the packets reach to the destination with lesser problems of channel contention and therefore end to end packets delay decreases. The similar behavior is observed with the modified intervals of OLSR because of the same reason as mentioned above. However, now with the modified intervals, the end to end packets delay is less as compared to the default settings. This is because of the increase in hello sending rate which increases the routing protocol's adaptation to neighbor changes and route maintenance that decreases the overall end to end packets delay.

6.6 Summary of Experiment No 3

The simulation results demonstrated that by optimizing the Hello and TC intervals, optimality in the routing protocol performance is achieved under specific mobility factors considered. Through simulations it has been explored that increasing rate of hello update leads to improvement in link establishment and node status maintenance. Further, decreasing rate of TC updates leads to significant reduction in control overhead but do not downgrade the data traffic received under specific mobility conditions considered. Hello interval has been slightly decreased from default value (2 secs) to 1.8 secs in order to alleviate the degraded performance of OLSR under high mobility scenarios thus achieving higher data traffic received than the default value. Increase in routing traffic due to increase in Hello interval has been compensated by decreasing the TC sending rate from 5 to 7.5 secs which drastically reduced the overall routing overhead while not posing any significant impact on data traffic received. This also resolves the problem of high routing overhead of OLSR (generated due to its proactive nature) under the specific mobility conditions considered. In the proposed Hello and TC intervals, OLSR is now able to sustain an increased data traffic received compared to the default values of Hello and TC intervals and at the same time, both the routing traffic and end-to-end packet delay are also reduced.

7. Conclusion

MANET is an autonomous system of mobile nodes connected by wireless links. The performance of MANET is related to the efficiency of the routing protocols. OLSR, a well known proactive protocol has emerged as the choice for MANETs (especially for delay sensitive applications) due to low latency for route determination. But at the same, time associated high routing overhead (due to proactive nature) has emerged as

a major performance issue in OLSR. In this study, we have addressed this issue by optimizing the OLSR under specific mobility conditions which are actually more of practical interest and thereby, our work does have a valuable contribution to provide guidelines for large number of cases of general interest. The default parameters of Hello and TC intervals of OLSR are selected such that the network performance is improved. The behavior of the routing protocol is tested based on the influence of node mobility using various performance metrics. From the results of simulations, it is concluded that the optimization of OLSR control messages intervals has shown to consistently outperform the default implementation of OLSR under specific mobility conditions considered during this study. We envisage undertaking research to analyze the scalability of OLSR protocol vis-à-vis control messages intervals with number of nodes and to set the boundary limits through detailed simulation studies. Furthermore, the performance analysis of OLSR protocol must be analyzed with realistic mobility models [17] so as to finalize realistic protocol performance.

8. References

[1] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, A review of routing protocols for mobile adhoc networks, Elsevier Journal of Ad Hoc Networks, 2004.

[2] The Book of Visions 2000, Visions of the Wireless World, IST – WSI Project, November 2000.

[3] S. Corson, J. Macker, MANET RFC 2501, MANET: Routing Protocol Performance Issues and Evaluation Characteristics, January 1999.

[4] C Adjih, T. Clausen, P. Jacquet, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, L. Viennot: Optimized Link State Routing Protocol, RFC3626, IETF, October 2003.

[5] A. Qayyum, L. Viennot, A. Laouiti: Multipoint relaying: An efficient technique for flooding in mobile wireless networks, INRIA research report N 3898, March 2000, INRIA Rocquencourt, France. http://www.inria.fr/rrrt/rr-3898.html.

[6] Jerome Harri, Christian Bonnet and Fethi Filali, OLSR and MPR: Mutual Dependences and Performances, EURECOM research report RR-05-138, 2005.

[7] Yang Cheng Huang, Saleem Bhatti, Daryl Parker, TUNING OLSR, The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'06.

[8] C. Gomez, D. Garcia, J. Paradells Improving Performance of a Real Ad-hoc Network by Tuning OLSR Parameters, 10th IEEE Symposium on Computers and Communications, ISCC 2005.

[9] Carlos Miguel Tavares Calafate, Roman Garcia, Pietro Manzoni, Optimizing the implementation of a MANET routing protocol in a heterogeneous environment, Proceedings of Eighth IEEE International Symposium on Computers and Communications, ISCC '03.

[10] Mounir Benzaid, Pascale Minet, Khaldoun Al Agha, Integrating fast mobility in the OLSR routing protocol, Mobile and Wireless Communication Networks, MWCN 2002.

[11] P. Samar and Z. Haas, Strategies for Broadcasting Updates by Proactive Routing Protocols in Mobile Ad hoc Networks, Proceedings of the IEEE Military Communications Conference (MILCOM), Anaheim, California, USA, October 2002.

[12] Pedro E. Villanueva-Peña, Thomas Kunz, Pramod Dhakal, Extending Network Knowledge: Making OLSR a Quality of Service Conducive Protocol, IWCMC 2006.

[13] F. Bai, N. Sadagopan, A. Helmy, IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of Routing protocols for Adhoc NeTworks, IEEE INFOCOM, 2003.

[14] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET Simulation Studies: The Incredibles," ACM SIGMOBILE Mobile Computing and Communications Review (MC2R), pp. 50-61, October, 2005.

[15] OPNet tutorial Modeling concepts reference manual.

[16] S Gowrishankar , T G Basavaraju, S. K. Sarka, Effect of Random Mobility Models Pattern in Mobile Ad hoc Networks, International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007.

[17] Jungkeun Yoon, Mingyan Liu, Brian Noble, Random Waypoint Considered Harmful, IEEE INFOCOM 2003.

Interconnection of Mobile WiMAX with UMTS

Mohamed Ould El Boukhary[1], Peng Manman[1], Li Renfa[1]
[1] College of Information Science and Engineering, Hunan University, 410082 China
Mohamed.bkh@gmail.com, pmmbysj@hnu.cn, scc_lrf@hnu.cn.

Abstract *- The WiMAX communications future`s will mainly focus on user Centricity and personalization. Multiple networks will be overlaid to provide supplement connectivity to a user. Competitive options in the current communication technologies need to be carefully studied for designing an interworking solution. Extensive amount of work is done on the interworking of UMTS with WiMAX. Currently, the interruptions of mobile WIMAX with UMTS (Universal Mobile Telecommunications System) during the vertical handover, pose a real problem to the end users especially when the service is sensitive to delay. In order to reduce the latency during the vertical handover; in this paper we propose a combination between the two mechanisms, the IEEE 802.21 architecture, that allows obtaining information on networks before the Handover; and the MSCTP protocol that uses during the vertical handover, To test our proposition, we will design an interconnection and mobility model between two WMAN radio technologies, which one belongs to the network family based on IP protocol: IEEE Mobile WiMAX, and the second belongs to the telecommunication network family; UMTS. To evaluate the performance of the proposition, we will apply first each mechanism only on the model; then our proposed solution; and we will make a comparison between the results obtained, and the simulations will be derived under the NS2 simulator; based on the tracking of a mobile station moving between the Mobile WiMAX and the UMTS networks.*

Keywords: *802.16e 802.21, Interconnection, Mobility, MSCTP, UMTS, VoIP.*

1. Introduction

Several research studies [1], [3], [4], [5], [6], [7], [8] were presented in the context of interoperability between mobile WiMAX and UMTS , but the majority of this work deals only the problem of interconnection, and other work dealing with the Handover propose a single mechanism of mobility. The main purpose of this study is to propose a model of interconnection between Mobile WiMAX and UMTS. The proposed Model based on a combination of two vertical handover mechanisms, to improve the QoS at the vertical handover between mobile WiMAX and UMTS. The proposed combination is the result of the integration of the protocol MSCTP known for his technique of Multi-homing in the protocol stack architecture of IEEE 802.21 can provide information On neighboring networks during handover To assess the effectiveness of our proposal, we will compare through simulations of some QoS parameters, the results obtained with the combination of both mechanisms, and results of each mechanism alone.

2. Protocol stacks based on MIH and MSCTP

In this section we will present the resulting protocols stack of the implementation of mobile WiMAX and UMTS networks in the IEEE 802.21 architecture, and integration of the protocol MSCTP within this architecture. The protocols architecture of 802.16e composed of a physical layer and three MAC layer [20] [21].

To establish the link between the WiMAX network and MIH, IEEE 802.21 [9], [10], [11], it defines new SAP (Service Access Point): The M_SAP (Management SAP) for the management plane; and C_SAP (Control SAP) for the control plane. The M_SAP specifies the interface between the MIHF (MIH Function) and the management plan. It allows MIHF payload to be encapsulated in managing messages. The primitives specified by M_SAP are used by MS to transfer packets to a BS, both before and after it has completed the procedures for entry into the network The C_SAP specifies the interface between the MIHF and the control plane. It provides before entering the network, while CS_SAP provides the data plan after the entry into the network. This SAP is used for MIH

exchanges between the MIHF and lower layers of the management plan. M_SAP and C_SAP are common between MIHF and the NCMS (Network Control and Management System); NCMS is linked with MIH_SME_SAP (MIH Station Management Entity SAP) to the MIHF (Media Independent Handover Function) which acts as an intermediary between the networks and upper layers, such as IP. [21], [22], [23], the protocol architecture of UMTS is composed of a physical layer, a MAC layer, a layer RLC and RRC layer. As the upper layers, UMTS layer is composed of a GMM layer (GPRS Mobility Management) and a SM layer (Session Management). [24]

The IEEE 802.21 defines a new SAP [9], [10], [11]: MIH-3GLINK-SAP to establish the link between the MIHF and the UMTS network. MIH-3GLINK-SAP is divided into two SAPs: the MIH-MGMT-SAP (Management SAP) to establish the link between the MIHF and the GMM, and RRC-SAP to establish the link between the MIHF and the RRC layer. Then, the MIH_SAP was defined to establish the link between the MIHF layer and the upper layers. In this paper we propose the IP layer up to the MIHF layer, and the MSCTP protocol layer up to IP layer. On the application layer in this case the VoIP application, and finally the MIH user layer. This protocols stack proposition allows us to use the vertical handover between WiMAX and UMTS, using IEEE 802.21 architecture and MSCTP Protocol.

The protocols stack is presented in the figure below:

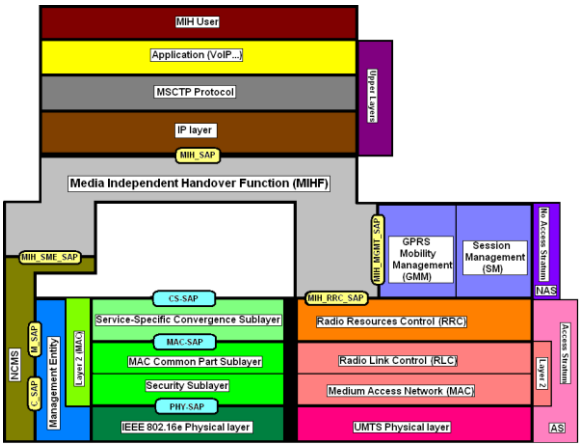


Figure 1. Protocols stack for MIH and MSCTP

3. Model interconnection and mobility scenarios

3.1 Interconnection model and mobility scenarios

In this section we describe the model of interconnection between mobile WiMAX and UMTS networks using IEEE 802.21 and MSCTP, and MS (mobility scenarios) between the two networks. We propose a WiMAX cell with coverage of 3 km radius, and UMTS cell with the same coverage radius. And the two cells have the same common area of Handover, with surface of 3 km² .with this interconnection; the mobile subscriber must operate with the two networks. Regarding the mobile WiMAX, the BS is linked to an ASN-GW; the ASN-GW is linked via the network IP to a CSN. (WiMAX ISP) which allow to access to the internet and external network [26], In the UMTS network, the Node B is linked to the RNC, and then SGSN is finally GGSN to access to the PDN network (Packet Data Network) [25].

The interconnection between the two networks is provided by two gateways installed between the two networks: WAG (WiMAX Access Gateway) and PDG (Packet Data Gateway) [12] [13], [14]. Through the WAG, the data from / to the WiMAX network is routed to provide the MS with UMTS services. The functions of WAG include strengthening the routing of packets through the PDG, performing accounting information, and filtering out packet. The main functions of the PDG are to route the packets received from/ sent to the PDN to / from MS, and to perform the foreign agent functions [27], [28], [29]. To use the services of MIH in the interconnection model, the mobile subscriber must implements the MIH module, an MIH server must be installed between the two networks, and the CSN WiMAX, must include an MIH module and the UMTS core network as well. In this case, the exchange of information on networks will be made between the MIH modules and the MIH server.

To use the MSCTP protocol, the end users: the mobile subscriber and its correspondent should implement the MSCTP protocol as upper layers. In the first scenario the MS using VoIP application, and localized first in the position 1 in the UMTS cell, will be moved to the second position in the handover area, between two networks. Then it will leave the Handover area to the third position in the beginning of the WiMAX coverage area. And finally, it

will reach the fourth position in the WiMAX cell. In the second scenario, the MS will traverse the same path, but in the opposite side, from WiMAX cell to UMTS cell.

Then finally, in the both scenarios, we propose two mobility speeds: as 50 and 100 km / h for showing the impact of the speed increasing in the case of vertical handover. The model of interconnection and the second mobility scenario are illustrated in the figure below:

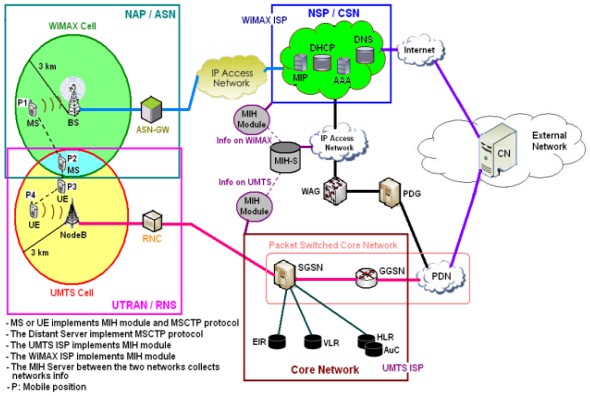


Figure 2. Model of interconnection and second mobility scenario

3.2 Handover scheme

This section describes the exchange of information between all entities during the mobility scenario for WiMAX to UMTS [15], [16], [17], [18].

First when the MS is localized at the first position in the WiMAX cell, it communicates with its distant server (DS) in an external network via the WiMAX network

(BS - ASN- GW - CSN - Internet - CN) Then, when it begins to move to the position 2 in the handover area, the MIHF entity in the MS begins the communication with MIH module to detect the new networks via MIH reports messages [30]. When one of the communication reports with MIH models is positive, and the MS detects a new network at position 2, in this case the initiation and preparation of the Handover start, the exchange of messages between MIH entities focuses on measurement, Mapping and comparison of the level of QoS between the two networks. If the ratio of measures is positive, the resource reservation of the UMTS network will be established.[30]

At this stage, the MS continues to communicate with the DS through the WiMAX network. Then, when it starts to leave the position 2 to the position 3, in the beginning of the UMTS cell, it proceeds at the execution of handover, and for this it must first obtain an IP address of UMTS networks, via the PDP context [31].

A PDP context is a set of information that characterizes a transmission service base. It includes parameters that allow a subscriber to communicate with a PDP addresses defined, according to a specific protocol (IP or X.25), according to a QoS profile determined (throughput, delay, priority ...). After obtaining the IP address UMTS, the MS now has two IP addresses. And with MSCTP protocol, the MS can classify its IP addresses and select one as primary address. It will therefore use DAR extension MSCTP protocol to prevent its DS to add the UMTS IP as second IP address obtained recently. And DS will response with an ACK [32], [33], in this level, at the position 3; the MS will choice the UMTS IP addresses as primary address. So it will still use the DAR extension via the UMTS to prevent the DS to set the UMTS IP address as primary address, and the DS will response with an ACK. The MS will continue its communication with DS via the UMTS network. Finally, when the MS leaves the third position toward the fourth position in the UMTS cell, it will send a message to prevent the DS to delete the old IP address; because it no longer needed, the DS response with an ACK, and the MS finally will continue to communicate with the DS via UMTS network (NodeB - RNC - SGSN - GGSN - PDN - DS).

4. Simulations

4.1 Parameters of simulations

The parameters adopted under NS2 simulator for both types of networks are illustrated in the following tables. The propagation model used to simulate communication in the WiMAX network is the Two- Ray Ground. This is a useful propagation model is based on the optical geometric, and considers both the direct and indirect path (reflection model) between the transmitter and receiver.

	IEEE 802.16e
Transmission Power (Pt_)	15 W
Receiving Threshold (RXThresh_)	9.375e-13 W
Carrier Sending Threshold (CSThresh_)	1.259e-14 W
Coverage Radius (Distance D)	3 Km
Radio Propagation Model	Two-Ray Ground [19]
Transmit Antenna Gain (Gt_)	1 dB
Receive Antenna Gain (Gr_)	1 dB
System Loss (L_)	1dB
Transmit Antenna Height (ht_)	1.5 m
Receive Antenna Height (hr_)	1.5 m
Modulation	OFDMA
Fréquence (Freq_)	3.5 GHz

Table 1. NS2 Parameters for WiMAX

	UMTS
(Pt_ UE)	0.25 W
(Pt_ NodeB)	3 W
Power Consumption (Pt_consume UE)	0.125 W
(Pt_consume NodeB)	1 W
Power at Idle State (Pt_idle UE)	0.005 W
(Pt_idle NodeB)	0.5 W
Data rate (Bandwidth_)	384 Kb/s
(Freq_)	2 Ghz
Coverage Radius	3 Km
Modulation	WCDMA
(Gt_ UE)	2 dB
(Gt_ NodeB)	18 dB
(Gr_ UE)	2 dB
(Gr_ NodeB)	18 dB
Radio Propagation Model	Okumura-Hata

Table 2. NS2 Parameters for UMTS

The propagation model used to simulate communication in the UMTS network is the Okumura–Hata model [34], [35], it is an empirical propagation model that decomposes the weakening of the radio signal into different functions, and is based on the use of a large amount of measurements to determine the statistical general properties . In the simulations, the type of traffic exchanged is the VoIP (real-time traffic chosen), and in NS2 we use CBR traffic / SCTP with a fixed size of a packet equivalent to 160 bytes.

The mobile station will traverse 2 km in WiMAX cell or UMTS cell, and 1 km in Handover zone. The duration of a simulation is set at 360 seconds, and the results will calculated every 20 seconds out of the handover area, and every 10 seconds in the Handover area.

4.2 Performance Criteria

To compare the performance of the MSCTP protocol only, 802.21 standards only (MIH + PMIP), and the two mechanisms together, (MSCTP + MIH) in the case of vertical handover between Mobile WiMAX and UMTS are: End-to-end, delay, packets loss ratio, and throughput.

5. Results

5.1 The Delays

In this section we will calculate the delays of the packets during the simulation period for the two mobile speeds 50 and 100 km / h, with the two handover techniques: MIH architecture only and MSCTP protocol only, and with the two techniques together. Let's start with the results of delays in the case of handover from UMTS cell to Mobile WiMAX cell.

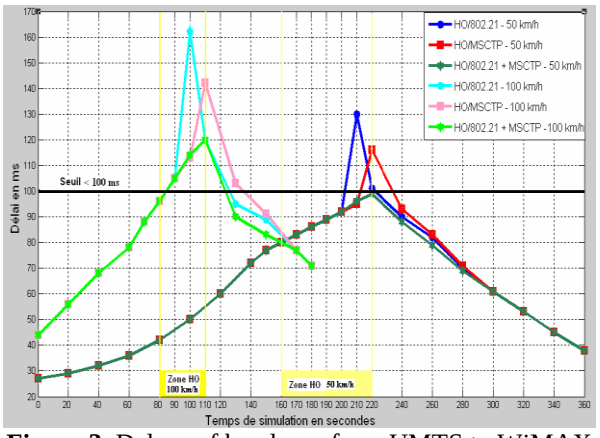


Figure 3. Delays of handover from UMTS to WiMAX

In this figure, we set a threshold of 100 ms [2] to evaluate the level of QoS during the simulation to reflect an acceptable level of QoS. The values of the delays should not exceed this threshold. We preceed to the comments of the curves: With a speed of 50 km / h, we obtain a single curve that does not exceed the threshold: it is the curve of MSCTP + MIH. The MSCTP curve exceeds slightly 100 ms during the handover, and reached a maximum of 116

ms. the curve of MIH only exceeds the threshold as to achieve a maximum of 130 ms during the handover. Then, with a speed of 100 km / h, the values of delay increase for all curves. In the case of using MSCTP + MIH handover , the values of the delay during the Handover no longer meet the threshold as in the previous case with an average speed 50 km/h , and reach a maximum of 120 ms during the handover . It is the same for other curves with a speed of 100 km / h; the results are superior to those obtained with an average speed of 50 km / h. We now plot to the results of delays in the case of handover from WiMAX mobile cell to UMTS cell:

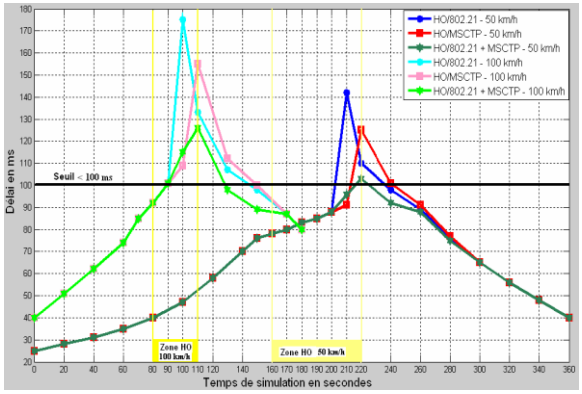


Figure 4 . Delays of handover from WiMAX to UMTS

First we note in this figure that in the case of handover from WiMAX to UMTS, the values of the delays for all the curves are slightly higher than those obtained in the case of handover from UMTS to WiMAX. For example, the scenario MSCTP + MIH, with a speed of 50 km / h, the maximum value reached 103 ms versus value of 99 ms reached in the first scenario. This is due to bound the

difference duration between the resource reservations during the handover process to obtain a new IP address. Finally, the best technique observed after simulations of delays and ensures a higher level of QoS for VoIP traffic, is the combination MSCTP + MIH. Furthermore the increase in speed causes the degradation delays. And finally, the Handover Mobile from WiMAX to UMTS produces better results slightly than those obtained in the opposite direction

5.2 Packets Loss Ratio

In this section we calculate the percentage of packets loss with the same conditions as in the previous section. We set a threshold of 1% [2] to evaluate the QoS level obtained after the simulations. The results obtained in the case of handover from UMTS to WiMAX mobile are shown in the figure below:

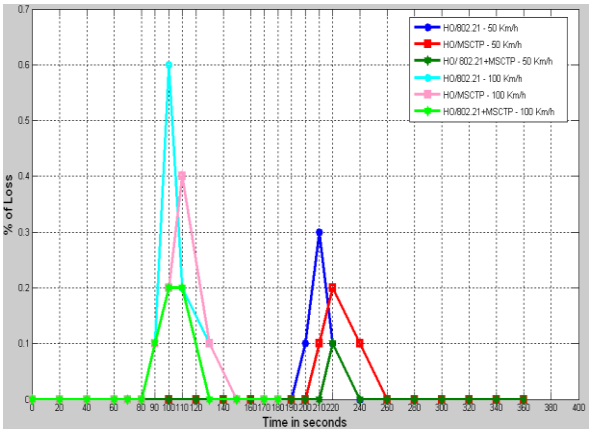


Figure 5 . Packet loss of the handover from UMTS to WiMAX

In this figure, all results obtained meet the threshold fixed as 1% [2] for traffic VoIP. The QoS level is acceptable to all the curves. The results with a speed of 50 km / h , however, are better than those obtained with the speed of 100 km / h , and the results of packet loss during handover obtained with MSCTP + MIH reflect a higher level compared to the use of each mechanism only. Using MSCTP only, the results are slightly better than those obtained with MIH only.

We present now the results of packet loss obtained in the case of handover from WiMAX to UMTS:

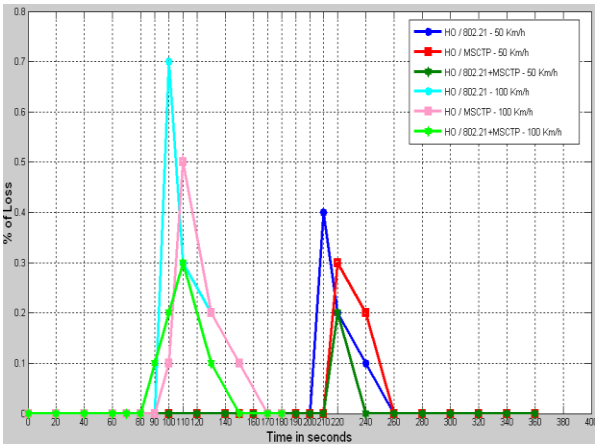


Figure 6. Packet loss of the Handover from WiMAX to UMTS

As for the delay, the degree of QoS during the handover from UMTS to WiMAX is slightly better than that obtained in the opposite direction. For example, the maximum value obtained during the handover from UMTS to WiMAX using MSCTP + MIH, with a speed fixed as 50 km / h, and is equal to 0.1 % versus 0.2% obtained in the case of Handover from WiMAX to UMTS. The best results obtained in this figure are in the case of handover using MIH+MSCTP.

5.3 Throughputs

Finally, in this section, we calculate the values of the throughput over a mobile simulation. We will start as the previous two sections by the case of handover from UMTS to WiMAX:

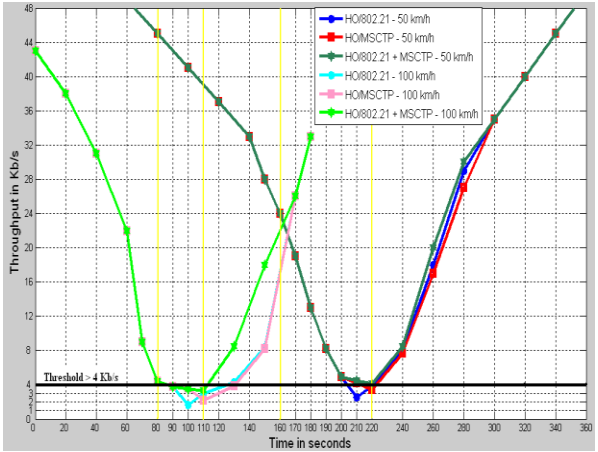


Figure 7. Throughputs of Handover from UMTS to WiMAX

Looking at this figure, the minimum value obtained with the speed of 50 km / h, and applying MSCTP + MIH during the handover, is 4.1 Kb / s. The values obtained with MSCTP only and MIH only during the handover is less than 4 kb /, and the curve of MSCTP produced slightly better results than those obtained with MIH. With a speed of 100 km / h, the results increase. For example, the minimum obtained MSCTP + MIH curve is equal to 3.3 kb / s. below is shown the results of throughput in the case of handover from WiMAX to UMTS:

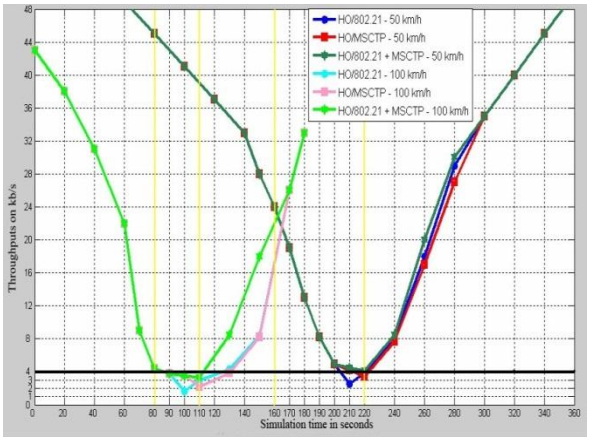


Figure 8. Throughputs of the Handover from WiMAX to UMTS

The throughput values decreased slightly in this scenario by comparing them with the previous scenario. For the case of Handover using MSCTP + MIH with a speed equal to 50 km / h , the minimum value obtained was 3.8 kb / s, versus 4.1 KB / s obtained in the previous scenario.

6. Conclusion

This study focused on the interconnection of two networks of different families, and we focused on the aspect of QoS for VoIP traffic in the case of Vertical handover process. For doing that, we have proposed an interconnection model based on two mechanisms for handover (MIH and MSCTP), and we have simulated this model with two speed classes, and in both directions from one network to another. from the results obtained , we can conclude that with an average mobility and use of MIH + MSCTP during the handover , especially when the mobile moves from UMTS to WiMAX, we obtained good results with minimal service interruption , and also obtained QoS level better

than that obtained with MSCTP only or MIH only. The handover from UMTS to WiMAX generate slightly better results than those obtained in the opposite direction with a simple or full mobility.

Finally, the results obtained in the case of full mobility require a significant review because they are not acceptable in comparison with the QoS level needed for VoIP traffic.

7. References

[1] Tarek Bchini , Nabil Tabbane , Emmanuel Chaput , Sami Tabbane & Andre-Luc Beylot , "Interconnection & Handover between IEEE 802.16e & UMTS ", IJACT: International Journal of Advancements in Computing Technology, Vol. 1, No. 2, pp. 99 ~ 09, 2009

[2] WiMAX Community, « WiMAX Fundamentals, 1.7.3 Quality of Service », Juin 2007.

[3] Quoc-Thinh. Nguyen-Vuong, Lionel. Fiat and Nazim. Agoulmine, « An Architecture for UMTS-WIMAX Interworking », Proc of 1st IEEE Symp. on Broadband Convergence Networks, pp. 1-10. Avril 2006.

[4] Shahbaz Khan, Shoaib Khan, Sahibzada Ali Mahmud, Hamed Al-Raweshidy, « Supplementary Interworking Architecture for Hybrid Data Networks (UMTS-WiMAX) », Proc. Int. Multi-Conf. on Computing in the Global Info, pp 57–61. Août 2006.

[5] S.Vijary Anand, « Qos based handover layer for a multi-RAT mobile terminal in UMTS and Wi-MAX networks », Communication Systems Software and Middleware and Workshops, pp 472-479. Janvier 2008.

[6] B. Liu, P. Martins, A. E. Samhat and P. Bertin, « A Layer 2 Scheme for Inter-RAT Handover between
UMTS and WiMAX in Tight Coupling Architecture », IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, pp 1-5. Septembre 2008.

[7] F. Xu, et al., « Interworking of WiMAX and 3GPP networks based on IMS », IEEE Commun. Mag., vol. 45, pp. 144-150, 2007.

[8] Joo-Young Baek, Deok-Jin Kim, Young-Joo Suh, Eui-Seok Hwang and Young-Don Chung, « Network Initiated Handover Based on IEEE 802.21 Frameworks for QoS Service Continuity in UMTS/802.16e Networks », in Proc. IEEE Vehicular Technology Conference, VTC Spring, pp. 2157-2161, Mai 2008

[9] E. Borcoci , « Wimax Technologies: Architecture, protocols, Resource Management & Applications »,Conférence CTRQ, Juillet 2008.

[10] V. Gupta, « IEEE 802.21 MIHS », IEEE 802.21 Working Group, Mai, 2008.

[11] U. Olvera, « IEEE 802.21 MIHO », IEEE 802.21 session 12, Hawaii, Janvier, 2006.

[12] K-H. Li, « Wimax Solution Division », Intel Mobility Group, Juin 2006.

[13] Q.Thinh, N.Vuong, L. Fiat, N. Agoulmine, « An Architecture for UMTS-WiMAX Interworking » Broadband Convergence Networks, pp. 1-10. Avril 2006.

[14] Y. Ching, « A Gateway to Integrate Heterogenous Networks », Industriel technology Research Institute, 2007.

[15] J-Y Baek, D-J Kim, Y-J Suh, E-S Hwang, Y-D Chung, « Network-Initiated Handover Based on IEEE 802.21 Framework for QoS Service Continuity in UMTS/802.16e Networks », VTC, pp. 2157-2161. Mai 2008.

[16] N-H. Thanh, N-T. Hung, T-N. Lan, T-Q. Thanh, D. Hanh, T. Magedanz, « mSCTP-based proxy in Support of multimedia session continutity and QoS for IMS-based networks », ICCE, pp. 162-168. Juin 2008.

[17] D. Kessens, J. Soininen, « UMTS/GPRS system overview from an IP addressing perspective », Nokia, technical white paper. 2006

[18] D. Feng, « Seamless Handover between CDMA2000 and, 802.11 WLAN using mSCTP », Thèse, 2006.

[19] Information Sciences Institute (ISI), « NSNAM Web Pages, 18.2 Two-Ray Ground reflection Model », Janvier 2009.

[20] IEEE Std, "Air Interface for Fixed and Mobile Broadband Wireless Access Systems", IEEE 802.16e, Part 16, February 2006.

[21] Eugen Borcoci , "WiMAX Technologies: Architecture, protocols, Resource Management and Applications", CTRQ Conference, July 2008.

[22] Vivek Gupta, "IEEE 802.21 MIHS", May, 2008.

[23] Ulises Olvera, "IEEE 802.21 MIHO", January, 2006.
[24] Qualcomm University, "Understand HSPA: High-Speed Packet Access for UMTS", November 2006.
[25] Esteban Zimanyi, "Performance analysis of vertical Handover between UMTS and 802.11 networks", Memory graduation, 2005.
[26] Kuo-Hui Li, "WiMAX Solutions Division", Intel Mobility Group, June 2006.
[27] Quoc-Thinh Nguyen-Vuong, Lionel Fiat, Nazim Agoulmine, "An Architecture for UMTS-WIMAX Interworking", Broadband Convergence Networks, 2006.
[28] YuChing, "A Gateway to Integrate Heterogenous Networks", Industriel technology Research Intitute, 2007.
[29] Arkadiusz Sitek, "Mobility Management in FMC", Multi-Service Access Everywhere (MUSE) Summer School, June 2007.
[30] Joo-Young Baek, Deok-Jin Kim, Young-Joo Suh, Eui-Seok Hwang, Young-Don Chung, "Network-Initiated Handover Based on IEEE 802.21 Framework for QoS Service Continuity in UMTS/802.16e Networks", Vehicular Technology Conference (VTC), May 2008.
[31] David Kessens, Jonne Soininen, "UMTS/GPRS system overview from an IP addressing perspective", Nokia, March 2006.
[32] Nguyen Huu Thanh, Nguyen Tai Hung, Tran Ngoc Lan, Tran Quang Thanh, Do Hanh, T. Magedanz, "mSCTP-based proxy in support of multimedia session continutity and QoS for IMS-based networks", International Conference on Communications and Electronics (ICCE), June 2008.
[33] Thesis, Deng Feng, "Seamless Handover between CDMA2000 and 802.11 WLAN using mSCTP", 2006.
[34] Oktay Akcakaya, Eda Kocaman, Osman Kaldirim, "Propagation Models", 2007
[35] Jyri Hamalainen "S72.3210 Channel Modeling for Radio Communication Systems, 3cr", October 2008.

Platform development for medical system in u-Health care environment

Minwoo Jung and Jeonghun Cho
School of Electronic Engineering, Kyungpook National University, Daegu, Republic of Korea

Abstract - *u-Health care has to measure various vital sign. Currently, Personal Health Device (PHD) communicates with a gateway. Then, vital sign transmitted from Gateway to monitoring system. Recently, auto diagnosis is trend in the medical industry. The auto diagnosis has to measure various vital sign to a patient. In this research, we facilitate communication among various PHD in order to reduce transmission of redundant data. Communication among the PHD complies with the Bluetooth HDP that the Bluetooth Special Interest Group (SIG) establishes. The PHD observes the IEEE 11073 PHD standard in order to assure interoperability among the PHD. We propose new u-Health system platform in order to auto diagnosis. The auto diagnosis provides reduction of medical expenses and medical diagnosis efficiency.*

Keywords - u-Health care; PHD; Bluetooth ;platform; auto diagnosis; M2M;

1 Introduction

As the number of the aged is growing increasingly, seniors who suffer from chronic disease grow rapidly. Then, u-Health care that patient can manage consistently one's health in their daily life is being studied. u-Health care is a medical system that it measures consistently a vital sign with Personal Health Device (PHD) in their daily life and transmit to medical center. u-Health care facilitate early diagnosis and decreasing medical expenses through efficient health care, it will be solved to lack medical professional. ICT (Information & Communication Technology) which used in u-Health care is measuring technology, transmitting technology, collecting technology, analyzing technology, feedback technology, etc. The most important technology in u-Health care environment is communication and interface among PHD and gateway. Recently, a number of manufacturer produce PHD in the market. Such proprietary protocol raises the interoperability problem among PHD. Then, the standardization of PHD required. The ISO/IEEE 11073 PHD is an internationally harmonized family of standards, produced by a grouping of manufacturers, institutions and IEEE Institute [1]. It consolidates previous IEEE 11073 Medical Information Bus and CEN standards, to cover different levels of the ISO Model, with model for access to the data and with services and communication protocols for interoperability between medical devices [2]. Lack of interoperability among PHD and third-party hospital information system solutions introduces communication overhead, imposes the installation and provision of complex networks designs, and limits the monitoring capabilities while moving inpatients for diagnosis examinations [3]. Wireless communication such as Bluetooth for u-Health care is essential technology. Bluetooth, as a short range wireless technology, is very suitable for many medical applications. Wireless sensors in hospitals, homes and applications using a GSM//3GPP networked infrastructure for forwarding medical data to a central server are just few examples. Particular applications using the mobile phone as kind of gateway are very interesting. Bluetooth systems for medical applications use proprietary implementation and data formats. In most cases applications that run on top of the Serial Port Profile (SPP) are used. Such systems that they don't come from one supplier suffer from severe interoperability problems. Since the implementation is customized for just one vendor and device, data exchange between such systems is often difficult. Even Bluetooth interoperability with PC's using different Bluetooth stack versions from different vendors is hard to achieve. Specific SPP solutions depend on virtual COM ports and specific stack APIs. Such an approach creates dependencies on a specific stack and in some cases on the operating systems used. To solve those issues the Bluetooth Special interest Group (SIG) started a program several years ago to define a new medical application. In June 2008 the Bluetooth SIG released the Bluetooth Health Device Profile [4]. Recently, Machine-to-Machine (M2M) is studied in order to improve on u-Health care system. M2M refers to technologies that allow both wireless and wired systems to communicate with other devices of the same ability. M2M uses a device to capture an event, which is relayed through a network to an application that translates the captured event into meaningful information. This is accomplished through the use of telemetry, the language machines use when in communication with each other. Such communication was originally accomplished by having a remote network of machines relay information back to a central hub for analysis, which would then be rerouted into a system like a personal computer [5]. u-Health care system, which is based M2M, facilitates auto communication among PHD. M2M is required sensor network technology, embedded system technology, signal processing technology in order to implement u-health care system. In this context, we propose new u-Health system platform that it facilitate M2M that communicate among PHD. Stack of the platform is followed IEEE 11073 PHD and Bluetooth HDP. M2M is based Bluetooth HDP that facilitate

communication among PHD. The designed u-Health system platform is porting in Microprocessor for embedded system that facilitates continuous patient′s health management. The second section describes the background for u-Health system platform. Section Ⅲ presents the implementation of new u-Health system platform. Finally, last section presents conclusion and future work.

2 BACKGROUND

2.1 IEEE 11073 PHD

The IEEE 11073 standards define medical devices using the conceptual model, where system application processes use services to establish associations with other devices and to access managed objects in the Medical Data Information Base (MDIB), which resides locally or on a remote device. Within this conceptual model, the IEEE 11073 standards define a family of sub-standards that map to the full seven-layer ISO/OSI model. IEEE 11073 standards allow communication between medical devices and external medical systems. They provide automatic capture of data of the patient's vital signs and of information associated with operation of the device. Figure 1 shows the correspondence between the names of the standard documents with the levels of communication and also with the parts that were absorbed at the beginnings of the Pilot Project. After that the set is renamed 11073-x and 1073.x. The correspondence between the ISO and IEEE nomenclatures is ISO 11073, IEEE 1073. IEEE took the editorial decision of naming the standards in the same way as ISO and, therefore, they are now named ISO/IEEE 11073, and can also be found with the short nomenclature X73. The overall ISO/IEEE 11073 system model is divided into three principal components. The three components consist of the DIM, the service model, and the communication model. These three models work together to represent data, define data access and command methodologies, and communicate the data from an agent to a manager. The DIM characterizes information from an agent as a set of objects. Each object has one or more attributes. Attributes describe measurement data that are communicated to a manager as well as elements that control behavior and report on the status of the agent. The service model provides data access primitives that are sent between the agent and manager to exchange data from the DIM. These primitives include commands such as Get, Set, Action, and Event Report. The communication model supports the topology of one or more agents communicating over point-to-point connections to a single manager. For each point-to-point connection, the dynamic system behavior is defined by a connection state machine. The connection state machine defines the states and sub-states an agent and manager pair passes through, including states related to connection, association, and operation. The communication model also defines in detail the entry, exit, and error conditions for the respective states including various operating procedures for measurement data transmission. The communication model also includes assumptions regarding the underlying communication layers' behavior. Another function of the communication model is to convert the abstract data modeling used in the DIM into transfer syntax, for example, to binary messages using medical device encoding rules (MDER), is sent using the communication model [6].

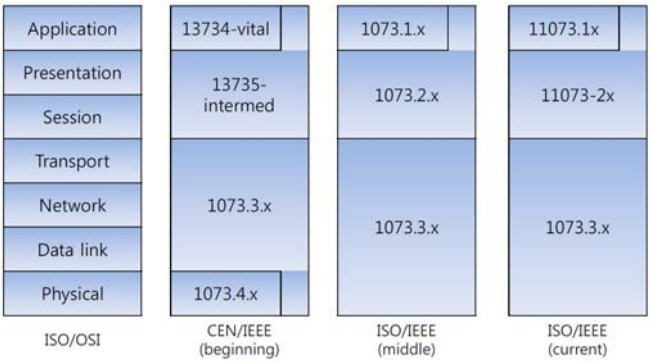


Figure 1 Reference Models for PHD

2.2 Bluetooth HDP

The Medical Working Group of the Bluetooth SIG began defining a specification addressing the needs of the medical community. Under Bluetooth, a profile defines the characteristics and features including function of a Bluetooth system. The end result of this work was the HDP specification that included the MCAP (Multi-Channel Adaptation Protocol) and made use of the Device ID Profile (DI). Figure 2 describes the interaction between a Bluetooth Protocol and a HDP in an overall Medical Device Application.

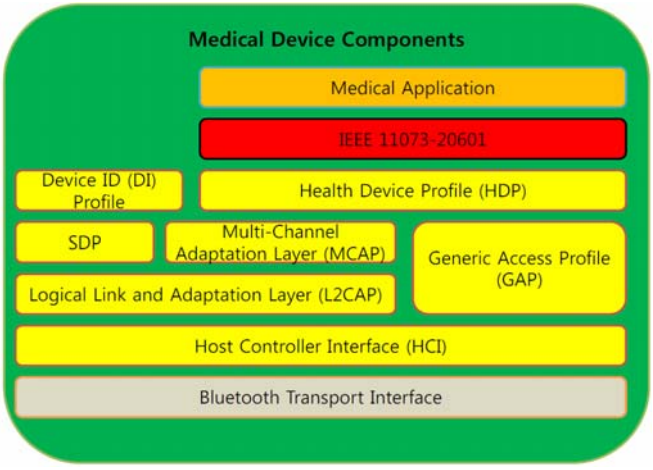


Figure 2 architecture of a Bluetooth HDP

Medical Application describes the actual device application, including its user interface, application behavior, and integration layer to IEEE 11073-20601 stack implementation.

IEEE 11073-20601 stack performs building, transmission, reception, and parsing of IEEE PDU packets for the associated PHD being developed. This component will directly link to the HDP. Device ID (DI) Profile is a Bluetooth profile designed to provide device specific information through use of the Service Discovery Protocol (SDP). If vendor specific information is required as part of a particular Medical Device, this profile provides specific behavior to acquire this information. A good HDP implementation offers API's to register and query for such vendor specific information. These API's can then be integrated directly into the Medical Application. Health Device Profile (HDP) is the core Bluetooth profile designed to facilitate transmission and reception of Medical Device data. The API's of this layer interact with the lower level MCAP layer, but also perform SDP behavior to connect to remote HDP devices. SDP is the Service Discovery Protocol used by all Bluetooth profiles to register and discover available services on remote devices so that connections over L2CAP can be established. Multi-Channel Adaptation Layer (MCAP) is used by HDP and facilitates the creation of a Communications Link (MCL) for exchanging generic commands, and also one or more Data Links (MDL) to transfer actual Medical Device data. MCAP is specific for the HDP and guarantees reliable transmission of data. Generic Access Profile (GAP) describes the required features of all core Bluetooth profiles including inquiry, connection, and authentication procedures. Logical Link and Adaptation Layer (L2CAP) supports protocol multiplexing, packet segmentation and reassembly, quality of service, retransmission, and flow control for the Bluetooth packets transmitted through MCAP. Host Controller interface (HCI) describes the commands and events that all Bluetooth hardware implementations can understand. Bluetooth Transport Interface describes the UART, USB, SDIO, 3-wire, ABCSP, etc. transport interface to the actual Bluetooth hardware components being used. Typically, UART and USB are the most widely used transports [7].

2.3 Machine to Machine

Currently, various international organizations for standardization are using the term such as machine to machine (M2M). M2M is service that provided user data that transmits collecting data by object such as a machine and a device. Figure 3 represent a simple construction of M2M that is defined in European Telecommunications Standards Institute (ETSI). Access network facilitates application of various access technologies such as xDSL, HFC, PLC, Satellite, GERAN, UTRAN, eUTRAN, WLAN, WiMAX. The gateway of M2M provides function that is able to connect M2M device domain with network domain. M2M area network defines network that connect the M2M gateway with M2M devices [8].

M2M technology facilitates various application of industry. M2M facilitate to use many applications of Existing Ubiquitous Sensor Network (USN). M2M and USN have something in common in many ways. Then M2M can adopt in u-Health care.

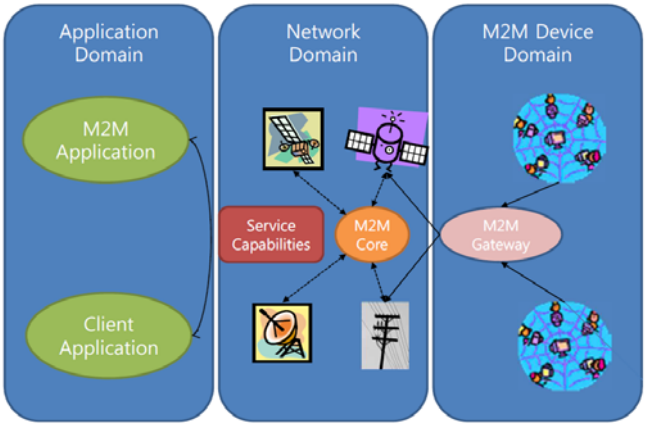


Figure 3 construction of M2M

3 IMPLEMENTATION

3.1 System Architecture

We design new u-Health system platform that facilitate communication among PHD. The platform consists of PHD, Gateway, Monitoring System. The PHD consists of bio-sensor, signal processing module, communication module. The signal processing module has mainly functionality of ADC. The communication module implements the Bluetooth HDP. The PHD observes IEEE 11073 PHD that facilitates interoperability among one. PHDs do not operate concurrently. When a PHD is generated event, another PHD that receive event message operates measurement of vital sign in order to make a health diagnosis of patient. For example, pulse oximeter that extracts SpO2, heart rate operates consistently in order to monitoring of patient's health. When pulse oximeter detects abnormal vital sign, it transmits ECG device event message. ECG device that receive event message operates in order to diagnose condition of heart. PHDs build network system for this situation. The gateway, communicate with PHDs, collect vital sign of patient. The gateway consists of communication module, storage. The communication module implements RS232 module that facilitates communication with Monitoring system and Bluetooth HDP module that facilitates communication with module of PHD. The storage of gateway collects vital sign of patent. The monitoring system communicates with the gateway. The monitoring system presents vital sign of patient that collect from the gateway. The monitoring system implements Personal Computer (PC). The gateway and the monitoring system is communicated RS-232. The platform usually operates minimum the PHD. So the platform can protect transmission of redundant data. The communication among PHDs controls each PHD behavior. The network of PHDs utilizes M2M technology.

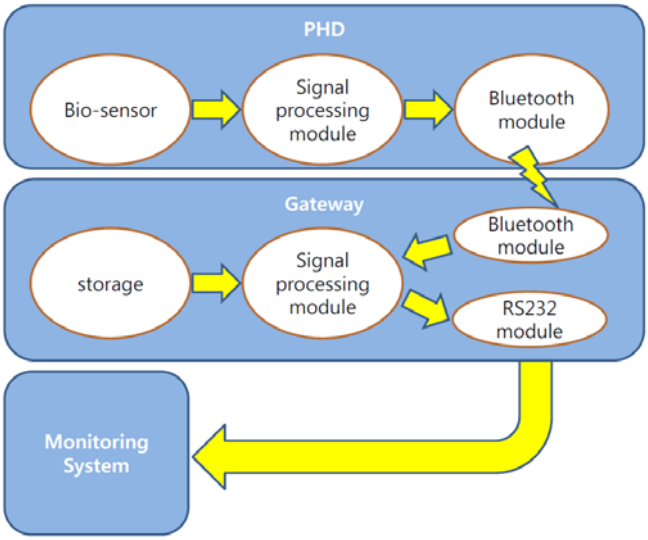


Figure 4 block diagram of proposed system

3.2 Implementation of standardizaiton

Proposed platform observes IEEE 11073 and Bluetooth HDL that communicate among PHDs. The programming framework for IEEE 11073 basically consists of Java, C and Abstract Syntax Notation One (ASN.1) as a language for the exchange of messages coded with Medical Device Encoding Rules (MDER). The application layer is defined by several protocols.

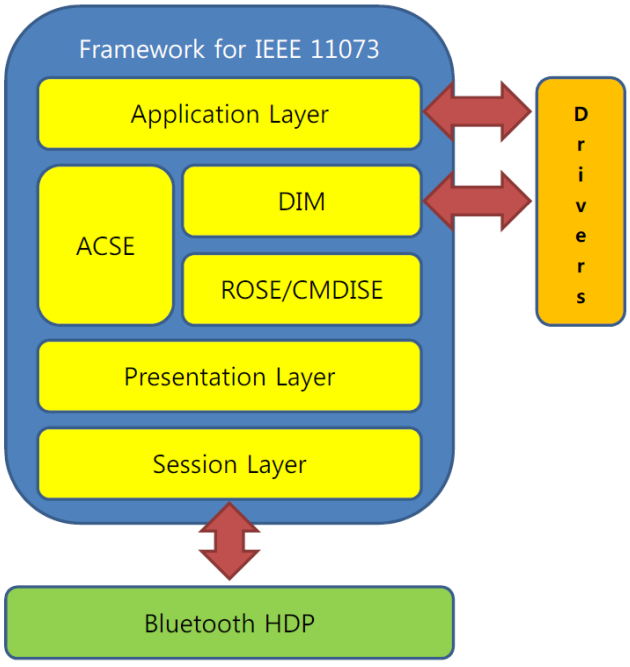


Figure 5 Architecture of the framework for IEEE 11073

ACSE is in charge of association control, CMISE is in charge of the basic services defined in VITAL and ROSE is in charge of the link between call requests and responses. Both ROSE and CMISE are merged into CMDISE. The presentation layer is mainly a negotiation mechanism for the syntaxes used by higher layers. The abstract syntax to use is specified by MDDL, and the transfer syntax is described by the MDER. The session layer provides support to the ACSE. The implementation of the transport layer varies regarding RS-232 or Bluetooth being in our case Bluetooth HDP.

Figure 6 presents stack of Bluetooth HDP. The PHD is the small device that will act as the transmitter of the medical data. The gateway is the feature rich device that will act as the receiver of the medical data. Bluetooth HDP devices that are categorized as a PHD include weight scales, blood pressure meters, thermometers, glucose meters, transmit application data over a reliable data channel to a gateway. Other PHD such as pulse oximeters, EEG, ECG transmits application data over a streaming data channel to a gateway. Multiple PHDs transmit application data over both reliable and streaming data channels to a gateway. This data can then be routed on to a physician through an alternate transport to a medical server at a hospital. The PHDs may be a combination device utilizing multiple data channels. Proposed system implements network of M2M environment which is based Bluetooth HDP. As proposed system implements M2M environment, it facilitates communication between PHDs.

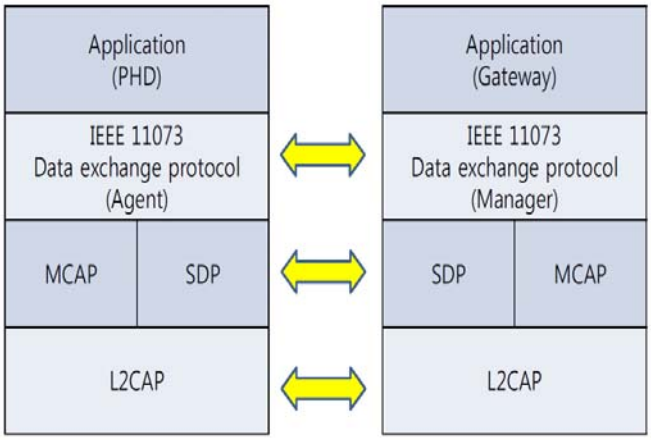


Figure 6 Stack of Bluetooth HDP

4 conclusion

We have proposed new platform that operates in machine to machine environment. The new platform is able to protect transmission of redundant data and communication among PHDs. So the PHDs compose network around the gateway. PHDs observe IEEE 11073 and Bluetooth HDP. Then it can assure interoperability of PHDs. The platform facilitates development of auto diagnosis easily. The auto diagnosis provides effective diagnosis and reduction of medical expenses. It also facilitates auto control about treatment of patients. In future, we need to research about standardization of machine to machine. As the platform guarantee interoperability of medical device, the more medical device is able to use in machine to machine environment. The system solves lack of medical specialist.

5 Acknowledge

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation.

6 References

[1] Jianchy Yao, Steve Warren, "Applying The ISO/IEEE 11073 Standards to Wearable Home Health Monitoring Systetms," in Journal of Clinical Monitoring and Computing, vol. 19, No. 6, 2005.

[2] M. Martinez-Espronceda, L. Serrano, "Implementing ISO/IEEE 11073: Proposal of two different strategic approaches," in EMBS 2008, August 2008.

[3] Juan M. Corchado, Javier Bajo, Dante I. Tapia, Ajith Abraham, "Using Heterogeneous Wireless Sensor Networks in a Telemonitoring System for Healthcare," in IEEE Transactions on Information Technology in Biomedicine, vol. 14, No. 2, March 2010.

[4] Bluuetooth SIG, www.Bluetooth.com.

[5] Sam Lucero, "Maximizing Mobile Operator Opportunities in M2M", ABIresearch, 2010.

[6] L. Schmitt, T. Falck, F. Wartena, D. Simon, "Novel IEEE 11073 Standards for Personal Telehealth Systems Interoperability", Joint Workshop on High Confidence Medical Devices, 2007

[7] David Kammer, Gordon McNutt, Brian Senese, Jennifer Bray, "Bluetooth", SYNGRESS, 2002

[8] Neyre Tekbiyik, Elif Uysal-Biyikolu, "Energy efficient wireless unicast routing alternatives for machine-to-machine networks," Journal of Network and Computer Applications, vol. 34, issue 5, September 2011.

The Influence of Network on Digital Forensic

Inikpi O. Ademu[1], Chris O. Imafidon[2]
[1]School of Architecture, Computing and Engineering, University of East London, Docklands Campus, London, United Kingdom
[2]*School* of Architecture, Computing and Engineering, University of East London, Docklands Campus, London, United Kingdom,
[2]Former Head of Management Unit, Queen Mary, University of London, London, United Kingdom

Abstract *– Previously, it was just enough to identify a standalone computer as detached objects containing digital evidence. Only collecting a computer and media storage devices would guarantee collection of all relevant digital evidence. Nowadays, computing has become networked. Individuals and businesses rely on e-mail, e-commerce and other network resources. Many computers are connected together using various network technologies. It is important for digital forensic investigators to be skilled and knowledgeable at pursuing the cyber dealings to find related digital evidence on the private network, public internet etc. A good understanding of the technology entailed will allow digital investigators to recognize, collect, preserve, examine and analyze evidence related to crimes involving networks. The aim of the research is to identify the fundamental network technologies that enable tracking down unknown offenders through networks and related criminal activities*.

Keywords – Access Point, Digital Forensic, Traffic, Internet Protocol, Transport Control Protocol

1. Introduction

The majority of organizations rely deeply on digital devices and the internet to operate and improve their business, and these businesses depend on the digital devices to process, store and recover data. A large amount of information is produced, accumulated, and distributed via electronic means. It is necessary for forensic experts to increase their abilities to gather evidence from digital devices (Ademu et al, 2012). A recent study demonstrates that in 2008, 98% of all documents created in organizations were created electronically (Sommer, 2009). Digital forensics plays an important part in the investigation of crimes involving digital devices. Digital forensic techniques are used primarily by private organisations and law enforcement agencies to capture, preserve and analyze evidence on digital devices. Digital evidence collected at a crime scene has to be analyzed and connections between the recovered information need to be made and proven. The search for digital evidence is thus a tedious task that consumes time. An extremely large amount of evidence needs to be processed in a very limited time frame which leads to delay in processing schedules.

Digital forensic science provides tools, techniques and scientifically proven methods that can be used to acquire and analyze digital evidence. Digital forensic investigators interact with digital evidence and digital forensic tools. The digital evidence can be used to backtrack or reconstruct illegal event. Digital forensic investigators are constantly trying to find better and more efficient ways of uncovering evidence from digital sources. An understanding of the technology entailed will allow digital investigators to recognize, collect, preserve, examine and analyze evidence related to crimes involving digital devices and also the networks. In a situation where a crime only involves email, an understanding of the network protocols is useful but not important. The digital forensic investigator may only need a basic understanding of the email to perform an effective investigation. Majority of the crimes involving networks require digital forensic investigators to be familiar with the fundamental technology. The goal of the research is to identify the fundamental network technologies that enable tracking down unknown offenders through networks and related criminal activities.

2. Sources of Digital Evidence on the Network

Digital forensic investigators need a basic understanding of network to interpret digital evidence found on PC such as e-mail, web browser and file transfer. An understanding of the fundamental network technologies is essential to track down unknown offenders through networks and related criminal activities to the network. For instance when digital investigators do not have access to a computer used to commit crime, evidence can be reconstructed using only evidence on the networks. Victims may be persuaded to bring their hard drive, sources of evidence on the internet that may reveal whom the victim was communicating with involve e-mail, log files on the victim's Internet Service Provider's system and backup tapes (Casey, 2004). Mobile telephone records may help verify whom the victim was communicating with and where the victim went. The sources of digital evidence on the network include the following:

- **Server Logs**

Server log file can be used to obtain a more complete picture of what occurred on a system. Log files used time stamps to indicate when an entry was added, and these must be synchronized to make sense (Sunday, 2003).

- **Contents of Network Devices**

Routers can direct data from one network to another filter unwanted traffic and keep logs that can be a brilliant source of digital evidence (Casey, 2004). The Firewalls can keep detailed logs of successful and unsuccessful attempts to reach the hosts that it protects. Intrusion Detection Systems (IDS) are used to collect information from a variety of system and network sources then analyze the information for signs of intrusion and misuse. In the network-based intrusion detection architecture, the system is used to analyze network packets (Sunday, 2003). Network-based architectures are used to detect access attempts and denial of service attempts originating outside the network. This architecture consists of sensors deployed throughout a network. These sensors then report to a central command console.

- **Traffic on both wired and wireless networks.**

Peer-to-peer networking has been advanced by wireless technology that uses radio frequency, infrared, lasers and microwaves such as Bluetooth enabled computers, personal digital assistants, mobile phones etc. when a Bluetooth enabled devices is on, it attempt to communicate with other devices around. Majority of the component of networked systems contain information about the activities of people who use them. Routers are at high risk of attack and computer intruders target routers to eavesdrop on traffic and disrupt or gain access to network (Casey, 2002).

3. Network Communication Technologies

Computer connected to a network is generally known as a host, and uses a modem or network interface card (NIC) to send and receive information over wires or through the air (Casey, 2004). In the past tap was used to connect a host to the network because this approach was difficult to maintain, devices called hubs were developed to reproduce the single network cable configuration. In order to increase network security and efficiency, switches replaced hubs. Techniques have been developed to enable eavesdropping on switched networks, undermining the security provided by these devices (Casey, 2002). The Transport Control Protocol/Internet Protocol (TCP/IP) is used as protocols to communicate with computers connected to the internet. Routers are an important component of computer network, essentially directing data to the correct place, they are used to connect host to two or more networks and direct traffic between them. Firewalls are similar to routers in a way that they direct traffic from one network to another. This security device is designed to block traffic by default and must be configured to permit traffic that meets certain criteria. The services that network enables, such as sending and receiving e-mails rely on the client/server model. Telnet provides client/server communication, enabling remote users to log into a server and execute commands.

4. Network Technologies

Network technologies enable multiple hosts to share a single transmission medium such as wire or the air. In the process of the host sharing a transmission medium, only one host can use the medium at a particular time, but there would be an interference with each other if two host were allowed to use the transmission medium at the same time. There are different network

technologies but for the purpose of the research three of them will be discussed.

- **Ethernet**

Ethernet is one of the most used network technologies because it's fast and inexpensive. One of the most recent forms of Ethernet uses wires similar to the telephone code. Ethernet uses Carrier Sense Multiple Access with Collision Detection (CSMA/CD) to synchronize communication.

- **Asynchronous Transfer Mode (ATM)**

ATM uses fiber optic cables and ATM switches to enable computers to communicate at very high rates. ATM uses technology similar to telephone system to establish a connection between two hosts. Host are connected to a main ATM switch and these switches can be connected to form a larger network. One of the hosts contacts the main switch when it wants to communicate with the other host. The switch contacts the other host and then establishes a connection between them.

- **Wireless**

Host connected using one of the IEEE 802.11 standards do not require wires, and they transmit data through the air using radio signals. The commonly used standard and their spectrums are IEEE 802.11a (2.4) and IEEE 802.11b (5GHz). Computers, personal digital assistants and other devices with a compatible wireless NIC uses Access point to communicate with each other. Access point are also generally connected to the wired network like an Ethernet network to enable communication with wired devices and the internet. The main limitation of 802.11 networks is that a computer must be within a certain distance of an access point to achieve reliable connectivity and even then, data are only transmitted at a limited speed.

- **Internet Protocols**

Two host using different network technologies cannot communicate directly. There are two methods of enabling communication between hosts using different network technologies, translators and common languages. When connecting different network, it is more efficient to join them using devices with the necessary network interface cards and then use a common internet protocol like TCP/IP that every host can understand. The most widely used internet protocols are the Transport Control Protocol (TCP), the User Datagram Protocol (UDP), and the Internet Protocol (IP). These protocols including few supporting protocols are collectively referred to as the TCP/IP internet protocol suite. In order to identify and deal with digital evidence on the internet, digital investigators need a concrete understanding of TCP/IP.

5. Applying Digital Forensic to Network

The computer forensic investigation paradigm is laborious and requires significant expertise on the part of the investigator. Computational intelligence is expected to offer more assistance in the investigation procedures and better knowledge reuse and sharing in computer forensics explains (Ruibin and Gaertner, 2005). Digital forensics is the science of digital crime investigation. The main purpose of digital investigation is to collect digital evidences, without altering or damaging it (Kruse and Heiser, 2001). The use of computer system and other electronic devices has been widely used in the last two decades. The large amount of information is produced, accumulated, and distributed via electronic means. The majority of organizations interact with electronic devices every day for this purpose, there is a need for finding digital evidences in computer systems and other electronic devices.

Since digital devices such as computers are vulnerable to attack by some criminals, digital forensics is very important. Understanding digital forensic procedures will help to capture vital information which can be used to prosecute an intruder that compromises a digital devices or network. Also, deciding on the specific tools for computers or other digital devices that is needed to correctly analyze evidence is crucial (Ademu et al, 2011).

Network contain digital evidence that can be establish that a crime has been committed, determine how a crime was committed, can provide investigative guides, reveal links between an offender and the victim, disprove or support witness statements and identify likely suspects. For instance child pornography on the internet has led digital investigators to victim, an e-mail of a missing person has created link between the victim and the offender. Processing a hard drive for digital evidence is a well defined procedure.

In dealing with evidence on the internet though, digital investigators are faced with some challenges. One major problem is that data on the networked systems

are dynamic and volatile, making it difficult to take a snapshot of a network. On like in dealing with standalone computer system, shutting down a network will result in the destruction of most of the digital evidence it contains. An offender can be at many places on the network at a particular time. This distribution of criminal activity and associated digital evidence makes it difficult to isolate a crime scene. But having evidence distributed on numerous computers can be an advantage in an investigation. The distribution of information makes it difficult to destroy digital evidence because if digital evidence is destroyed on a particular computer, a copy can be found on various computers on the network or on backup tapes. It is a good practice for organisations to backup their information regularly and can store copy of all backups in a different location (Ademu and Imafidon, 2012).

In some situations, digital evidence exists on networks that were not directly involved in a crime. Most times system administrators help digital investigators obtain evidence. There are always more sources of digital evidence on the network than even the system administrators realize. Therefore, to ensure that all relevant data is located, digital investigators must use their understanding of networks in general. Collecting digital evidence from a large network requires significant planning. Planning is important in cases that involve digital devices. If possible when generating a search warrant, care needs to be taken in researching the search site in order to determine what digital device to expect, what the system are used for and if or not is a network environment. If the digital equipments are used for business purposes this will influence the planning and preparation process. It is also important to know that without this information, it is difficult to know what expertise and evidence collection tools are required. At this stage proper preparation need to be done for tools and storage capacities that will be used. In order for plans and procedures for investigation be successful, it is important to provide adequate acquisition process. Before conducting an online investigation, corporate security professionals and law enforcement officers alike should obtain permission to proceed.

Digital evidence must be preserved as soon as it has been identified and collected. Digital evidence must be preserved in a way that it can be authenticated. The empirical law of digital collection and preservation states that if only one copy of the digital evidence is made, that evidence will be damaged or completely lost. Therefore, at least two copies of the evidence are taken. One of these is sealed and then placed in secure storage. This is the original copy and will only be opened for examination under instruction from the court in the event of a challenge to the evidence presented after forensic analysis on the second copy (Sunday, 2003). The main aspect of preserving digital evidence is collecting it in a way that does not alter it. It is sometimes desirable to preserve digital evidence on a networked system by gaining physical access to the associated computer and making a bit-stream copy of the contents. But the difference when working with networked system comes when digital investigators cannot make a bit-stream copy of digital evidence. There are situations where a bit-stream copy cannot be made such as when the system cannot be shut down, the hard drive may be too large to copy or the digital investigators often rely on large Internet Service Providers to collect evidence from their own systems such as subscriber information. Digital investigators collect digital evidence remotely when there is a strong chance that it will be destroyed before they reach the system (Casey, 2004).

Digital evidence is a rich and often unexplored source of information. It can enable an investigator to create an incredibly detailed picture of events surrounding a crime. Pollitt (2007) explains that one of the foundational ways in which researchers try to understand the scientific basis of discipline is to construct models which reflect their observation. When events happened it is very important computers, for instance notes the time an event occurred, such as the time an individual logged on using a password, digital devices can be useful for reconstructing the sequence of events. The position of digital evidence in relation to other objects can be very informative. Determining where a computer intruder hides files can help reconstruct a crime and also help investigators of the same crimes discover similar hiding places. Determining where an object or individual was in relation to other objects or individual is useful when investigating digital crimes mainly in a networked environment Casey (2004). In a large digital fraud case, thousands of people and computers can be involved, making it difficult to keep track of the lots of relationships between objects. Creating a diagram visualizing the associations between the objects and individuals can clarify what has happened. Analyzing digital evidence from networks frequently needs specialised knowledge of tools and the underlying network technology. As a rule digital devices used to store and analyze digital evidence should not be connected to the public internet. This could be risky because individuals on the internet could gain unauthorized access to evidence. There should be forensic duplicates of compromised systems, ensure strict use of write-protectiontechnology, positive legal chain of custody must be maintained. Hashes and checksums should also be maintained on forensic

duplicates and forensic working copies should be made and used during analysis. Presenting the findings to non technical individuals can be challenging but remains one of the most important stages in a forensic examination because a digital forensic examiners findings will not be used if they not understood (Ademu et al, 2011).

Conclusion

Digital investigators are assured to consume an important amount of time and are likely to lose valuable digital evidence without an understanding of where information can be found on the network. It is important for digital forensic investigators to have a good knowledge of network technologies and how the network functions. Not to be surprise, connecting computers together is risky, and an individual including criminals can gain unauthorized access to a distant network. With the appropriate authority and precautions, digital investigators can gain access to and collect evidence from distant networks. Digital investigators can capture digital evidence as it travels over a network, and computer networks enable digital investigators to communicate with each other and observe criminal activity and communication effectively. The main challenge for digital investigators is to follow cyber traces swiftly to find pockets of evidence before they are lost. By learning how computer networks function and how forensic science can be applied to networks, digital evidence can be used to address the growing problem of cybercrime.

Acknowledgement

The authors would like to thank Dr David Preston, and the University of Cambridge Computer laboratory for providing support during this research.

References

[1] Ademu, I. Imafidon, C. (2012) The need for digital forensic investigative framework Vol. 2 (3) Available at: http://www.ijesat.org/Volumes/2012_Vol_02_Iss_03/IJESAT_2012_02_03_01.pdf (Accessed on 25 May 2012)

[2] Ademu, I. Imafidon, C. Preston, D., (2012) Intelligent Software Agent applied to Digital Forensic and its Usefulness Vol. 2 (1) Available at: http://interscience.in/IJCSI_Vol2Iss1/IJCSI_Paper_21.pdf (Accessed 10 April 2012)

[3] Ademu, I. Imafidon, C. I. Preston, D. (2011) A New Approach of Digital Forensic Model for Digital Forensic Investigation Vol. 2, (12) Available at: http://thesai.org/Downloads/Volume2No12/Paper%206-A%20New%20Approach%20of%20Digital%20Forensic%20Model%20for%20Digital%20Forensic%20Investigation.pdf (Accessed 28 April 2012)

[4] Casey, E (2004) Digital evidence and computer crime forensic science, computers and the internet 2nd Edition P 101 London: Academic Press

[5] Casey, E. (2002) Handbook of computer crime and investigation P 116 London: Academic Press

[6] Kruse, W. Heiser, J. (2001) Computer Forensics Incident Response Essentials P 170 Indianapolis: Addison

[7] Pollitt, M. (2007) An Ad Hoc Review of Digital Forensic Models, Vol. 10(12)

Available at: http://www.ieeexplore.ieee.org/ie15/4155337/4155338/04155349.pdf? (Accessed 17 June 2012)

[8] Ruibin, G. Garrtner, M. (2005) Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework. Vol. 4(1) Available at: http://www.utica.edu/academic/institutes/ecii/publications/articles/B4A6A102-A93D-85B1-95C575D5E35F3764.pdf (Accessed 20 May 2012)

[9] Sommer, P. (2009) Directors' and corporate advisors' guide to digital investigations and evidence: Information Assurance Advisory Council 2nd edition. Available at:

www.iaac.org.uk/Portals/0/DigitalInvestigationsGuide.pdf (Accessed 3 April 2012).

[10] Sunday, H. (2003) Digital Evidence Available at: http://infohost.nmt.edu/~sfs/Students/HarleyKozushko/Papers/DigitalEvidencePaper.pdf

(Accessed 24 May 2012)

High Coding Opportunity and Low Interference Channel Assignment in Multi-channel Wireless Networks

Yunlong Zhao[1,*], Shilong Kang[2], Hao Wang[1], Guangjun Wu[1], Boshi Wang[1]
[1]School of Computer Science and Technology, Harbin Engineering University, 150001, Harbin, China
[*]zhaoyunlong@hrbeu.edu.cn, halinw@hrbeu.edu.cn, junbryant@163.com, boshwong@gmail.com
[2]Automation Company of TISCO, 030003, Tai Yuan, China
kingksl@126.com

Abstract - *Multi-channel has become an important technology to improve the performance of wireless network, and network coding has changed the conventional routing transmission model. But most of the existing research works on network coding do not consider the multi-channel networks scenarios. This paper proposes a novel multi-channel assignment algorithm based on network coding, which is named as HLCA (High coding opportunity Low interference Channel Assignment), which is to assign channels according to the level of coding opportunity and the channel interference. The simulation results show that HLCA works better than conventional multi-channel assignment and conventional network coding technology.*

Keywords: Multi-channel; Network Coding; Coding Opportunity;

1 Introduction

Network coding (NC) [1] theory was proposed since 2000, and has been proved to be a valid method that can improve the wireless network capacity compared with the conventional routing technology. Fig.1 shows a general network coding structure, node A and node B have packets P1 and P2 separately to send to each other, and node M can encode them together as P1 ⊕ P2. The nodes use same communication channels, i.e. a=b=c=d in Fig.1, then it will totally cost three time-units to finish the transmission. However, the interference between channels restricts the transmission, such as node A cannot communicate with node M when node B send packet P2 to node M because a 、b 、c and d are the same channels. So how to decrease the channel interference in the course of network coding has been regarded as a key issue in the study of network coding.

Multi-channel technology has become one of the key methods to improve the wireless network performance. The interference between channels restricts the communication of nodes, so that network coding performance is limited. We assume that node *A*、*B* and *M* own two interfaces individually. Supposing $a \neq b$ in Fig.1, so node *A* and node *B* can communicate with node *M* during one time-unit. It costs only two time-units to finish the process of transmission in this wireless network topology.

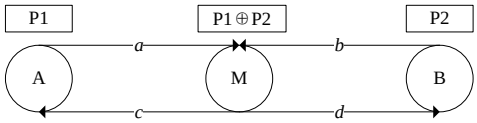


Fig.1 Network coding topology structure

The essential problem about multi-channel research is to design a rational channel assignment algorithm, and several multi-channel assignment algorithms have been proposed so far. Naveed [2] proposed a novel multi-channel assignment technology, which is based on cluster. Using cluster improves the efficiency of the multi-channel assignment and the normal cluster of allocation is based on the star structure. Marina and Das [3] proposed a method, which realizes topology control by using multi-channel technology, and the basic demand of the multi-channel technology is to ensure the connection of network topology. Ramachandran et al [4] researched on multi-channel based on the cognitive interference between channels.Obviously, the main idea for multi-channel assignment is to decrease the channel interference.

The COPE [5] is the first network coding protocol, which is used in the wireless network, and COPE has been certified that can improves wireless network performance. Zhang [6] and Su [7] have proved the availability of network coding which is communicated through multi-channel. Ref. [6] is based on the anneal algorithm, and has proposed a routing protocol called CDR (Coding Directed Routing). Kwon and Frkri [8] proposed a novel network coding model, and this method is based on three source nodes, three destination nodes and one coding node. In Ref. [8], the source nodes can collect other source nodes' information, and destination nodes enhance the decoding probability. A multi-channel assignment algorithm is also proposed in Ref. [8], but the assignment of different channels for source nodes and destination nodes is so idealized.

In this paper, we propose a novel multi-channel assignment algorithm, which is called High-Coding opportunity Low-Interference Channel assignment (HCLA). HCLA is a multi-channel assignment strategy based on the high coding opportunity. The remainder of this paper is organized as follows: In Sec. 2, we describe the multi-channel network coding determination conditions through formulations. In Sec. 3, we design the HLCA details, and in

Sec. 4, a simulation to evaluate the performance of HLCA is conducted based on NS-2, and the simulation results are analyzed.

2 Formulation Judge

In Ref. [8], the author proposed a formulation to explain the network coding, but it doesn't consider the communication channel number between pair of nodes. In this section, we present a novel formulation to judge the multi-channel network coding, and we consider the same scenario in Fig.1 and nodes deploy with two interfaces. In this model, we assume source nodes *A* and *B* have *n* continuous equal-size packets to send to node *M*. We do not consider packet losses in this paper. The novel definitional operation is shown as '$\cap^c$'. Such as, '$a \cap^c b$' means that packet *a* is sent through channel *b*. The source nodes *A* and *B* have packets sequences $P_a(x)$ and $P_b(x)$, which are transmitted to node *M*. Let

$$P_a(x)=\sum_{i=0}^{n-1}a_i x^i \qquad P_b(x)=\sum_{i=0}^{n-1}b_i x^i \tag{1}$$

Where a_i and b_i are packets corresponding to the packet number *i* and the exponent of x denotes the time index. We define XOR operation in the polynomial domain as:

$$P_a(x)\oplus P_b(x)=\left(\sum_{i=0}^{n-1}a_i x^i\right)\oplus\left(\sum_{i=0}^{n-1}b_i x^i\right)\oplus\sum_{i=0}^{n-1}(a_i\oplus b_i)x^i \tag{2}$$

Further, we define $T_k(C_k x)$ to express packet is transmitted by channel C_k, let

$$T_k(C_k x)=\sum_{i=0}^{n-1}(C_k^i x^i) \tag{3}$$

Where C_k^i denotes the packet is transmitted through channel C_k corresponding to the time index *i*. So when node *M* sends packet through channel would be shown as follows:

$$\begin{aligned}P_{m,send}(C_m x)&=P_{m,send}(x)\cap^c T_m(C_m x)\\&=\sum_{i=0}^{n-1}\{(m_i\cap^c C_m^i)x^i\}\end{aligned} \tag{4}$$

Using (1)-(4), the packets sent from node A and B to node M are expressed as follows:

$$\begin{aligned}P_{a,send}(C_a x)&=P_{a,send}(x)\cap^c T_a(C_a x)\\&=\sum_{i=0}^{n-1}a_i x^i\cap^c\sum_{i=0}^{n-1}(C_a^i x^i)=\sum_{i=0}^{n-1}\{(a_i\cap^c C_a^i)x^i\}\end{aligned} \tag{5}$$

$$\begin{aligned}P_{b,send}(C_b x)&=P_{b,send}(x)\cap^c T_b(C_b x)\\&=\sum_{i=0}^{n-1}b_i x^i\cap^c\sum_{i=0}^{n-1}(C_b^i x^i)=\sum_{i=0}^{n-1}\{(b_i\cap^c C_b^i)x^i\}\end{aligned} \tag{6}$$

Where Pa,send(x) and Pb,send(x) are native packets. Node A and node B can transmit packets to node M simultaneously, because node A and node B communicate with node M through different channels. Node M receives Pa,send(x) and Pb,send(x), and transmits the XOR of two packets with appropriate delays to the destination nodes B and A, that is:

$$\begin{aligned}P_{m,send}(C_m x)&=\{P_{a,send}(C_a x)\cap^c T_m(C_m x)\}x\oplus\{P_{b,send}(C_b x)\cap^c T_m(C_m x)\}x\\&=\left\{\sum_{i=0}^{n-1}\{(C_a^i\cap^c a_i)x^i\}\oplus\sum_{i=0}^{n-1}\{(C_b^i\cap^c b_i)x^i\}\right\}x\cap^c T_m(C_m x)x\\&=\sum_{i=0}^{n-1}\{(C_a^i\oplus C_b^i)\cap^c(C_a^i\oplus b_i)\cap^c(a_i\oplus C_b^i)\cap^c(a_i\oplus b_i)\cap^c C_m^i\}x^{i+1}\end{aligned} \tag{7}$$

$$P_{m,send}(C_m x)=\sum_{i=0}^{n-1}\{(C_a^i\oplus C_b^i)\cap^c(a_i\oplus b_i)\cap^c C_m^i\}x^{i+1} \tag{8}$$

Where node *M* needs one time-unit delay to send XOR-packet. As the XOR operation rule, the (7) could be sampled to (8). However, the source nodes communicate with destination nodes though different channels, so $C_a \neq C_b$. (8) can simple to (9).

$$P_{m,send}(C_m x)=\sum_{i=0}^{n-1}\{(a_i\oplus b_i)\cap^c C_m^i\}x^{i+1} \tag{9}$$

In this network coding model, the destination node A and node B receive XOR packet, and decode as follows:

$$\begin{aligned}P_{a,decode}(C_m x)&=P_{m,send}(C_m x)x\oplus P_a(x)x^2\\&=\sum_{i=0}^{n-1}b_i x^{i+2}=P_b(x)x^2\end{aligned} \tag{10}$$

$$\begin{aligned}P_{b,decode}(C_m x)&=P_{m,send}(C_m x)x\oplus P_b(x)x^2\\&=\sum_{i=0}^{n-1}a_i x^{i+2}=P_a(x)x^2\end{aligned} \tag{11}$$

Where (10) and (11) can judge that node A and node B can decode the XOR packet successfully, and there is only a two time-units delay. From (5) to (11), we assume there are *n* packets to exchange, it needs *2n+1* time-units delay (as shown in Table 1). Table 1 shows that the network coding gain for multi-channel network compared with single channel network coding is *(3n-1)/(2n+1)*, and the value is *3/2* when *n* inclines to infinitely great.

Table1. Send time and Receive time for Single/Multiple channel Network Coding

Packet Number Index	a_0	a_1	a_2	... ...	a_{n-1}
Send Time for Single Channel Network Coding	x^0	x^3	x^6	... ...	$x^{3(n-1)}$
Receive Time for Single Channel Network Coding	x^2	x^5	x^8	... ...	x^{3n-1}
Send Time for Multiple Channel Network Coding	x^0	x^2	x^4	... ...	$x^{2(n-1)}$
Receive Time for Multiple Channel Network Coding	x^1	x^3	x^5	... ...	x^{2n+1}

Equation (5) to (11) can judge whether the multi-channel network coding conditions are satisfied, so we would design a novel multiple channel assignment algorithm—High-Coding opportunity Low-Interference Channel Assignment (HLCA).

3 Design of the HLCA

3.1 Relationship between node degree & coding opportunity

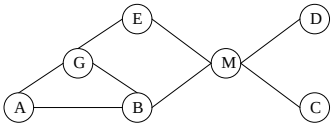


Fig.2 A random wireless network topology

Fig.2 shows a random wireless network topology, there are sever wireless nodes and the straight line between two nodes denotes they are one-hop neighbors. It is obvious that node degree reflects the connectivity level of wireless network topology. The degree of node *M* in Fig.2 is four and node *G* and node *B* are both three. As the network coding conditions, Fig.3 shows the whole coding opportunity for node *M* and node *G*, so the coding opportunity of node *M* is higher than node *G*. it can be inferred that the coding opportunity is direct ratio with node degree.

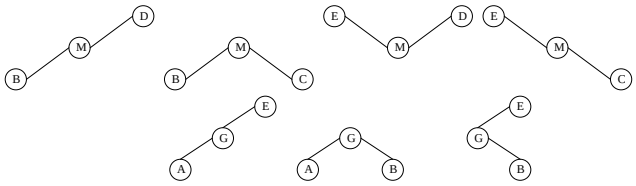


Fig.3 Coding opportunity for node *M* and node *G*

One important issue about multi-channel assignment is ensuring the connection of the wireless network topology. So according to the relationship between node degree and coding opportunity, we give some definitions as follows:

Definition 1: *Potential coding nodes*, which may have the coding opportunity.

Definition 2: *Potential coding structure*, which includes a potential coding node and its one-hop neighbor nodes.

Definition 3: *Edge nodes*, which are at the outermost layer of the network topology.

The definitions of potential coding node and the potential coding structure are proposed to determine the priority of nodes. The premise of HLCA strategy is to ensure the connectivity of network topology. So as this rule, we classify node priority as follows:

Situation 1: When nodes belong to two or more potential coding structures at the same time, they enjoy the highest priority. These nodes are defined as high priority nodes;

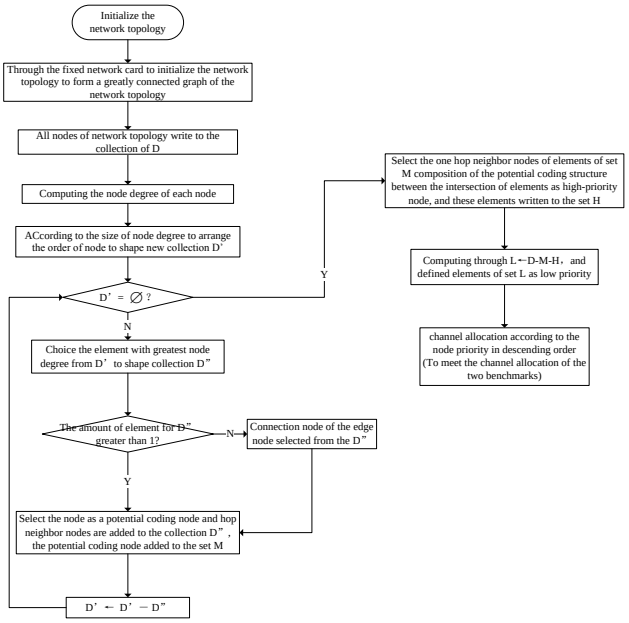
Situation 2: The potential coding nodes are defined as middle priority nodes;

Situation 3: The edge nodes and nodes do not belong to the situation 1 and 2 are defined as low priority nodes.

The connectivity degree of node determines the nodes' impact in network topology. So the definition of node priority is based on the influence degree in the network topology. The node which belongs to two or more potential coding structures guarantees the connectivity between each potential coding structure, so it makes highest effect to current network topology. So as above rules, it could obtain the condition for node priority assignment.

3.2 The particular design of HLCA

The particular design of HLCA can be described as following flow chart:



According to the flow chart, the design of HLCA can be described as following steps:

1) Initialize network topology: The first step of HLCA is to initialize the current network topology, and the nodes assign a fixed channel No. to one of their interfaces. This ensures a maximum connectivity of the network topology.

2) Computing the node degree for each node: After initializing the current network topology, we set up a set to store all the nodes in current network topology, and compute the node degree one by one. Pick out the greatest degree node as the first potential coding node, and then remove this node and the nodes in the same potential coding structure with it. According to this rule, we will get all the potential coding nodes from the node set till it is empty.

3) Set the priority for each node: According to the 3 situations of node priority, set the node priority for every node in the current network topology.

4) Assigning channel between nodes: Build the set H、M and L to store the nodes which own high、middle and low priority respectively. Then assign channels according to the node priority order which shown as Fig.4. The channel assignment must follow the basic principle of multi-channel allocation: the adjacent channel can not be assigned with the same channel.

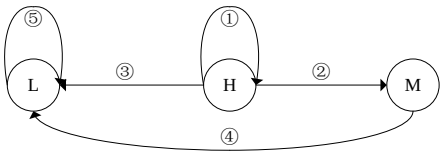


Fig.4 Priority order of the nodes for channel assignment

4 Performance Evaluation

We evaluate HLCA via simulations on NS-2. The throughput and coding gain brought by using multiple channels and multiple radios are the key metrics to check. We compare HLCA with the following schemes:

a) Compared with single-channel and single-radio network coding in throughput: The COPE is a typical single-channel network coding technology using in the wireless network, but the channel interference decreases the superiority of network coding. However, HLCA initializes the network topology and increases the coding opportunity. The simulation results are shown as follows.

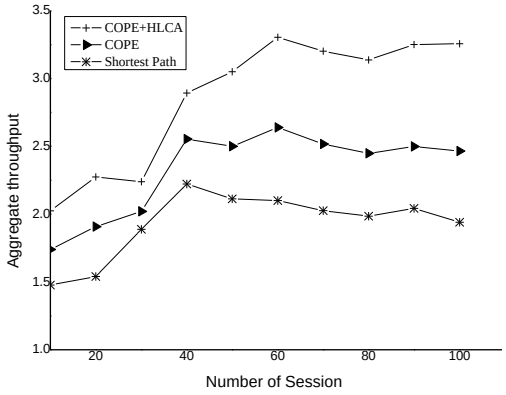


Fig.5 The relationship between different time slots and throughput

Fig.5 indicates that HLCA+COPE improve approximately 32% on throughput than conventional COPE protocol. This is due to the increasing coding opportunities and decreasing channel interference and the simulation results content the theoretical analysis.

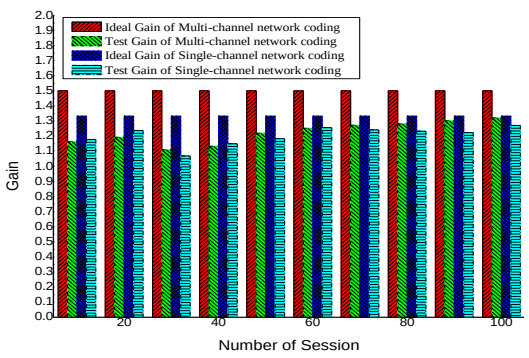


Fig.6 The contrast of coding gain between ideal and simulation

b) Compared with single-channel and single-radio network coding in coding gain: Coding gain reflects the transmission time, as shown in Fig.6, ideal gains of different models have been analyzed, and the result indicates that the multiple channel network coding performs better.

5 Conclusions

In this paper, we provided a new strategy, HLCA, based on network coding technology to optimize the process of multi-channel assignment. We classify all the nodes into various sets with different priority levels according to the node degree, and the nodes with higher degree have higher coding opportunity. Then, we assign channels for the nodes following the priority order. Compared with the conventional scheme, HLCA, with high coding opportunity and low interference, performs better on the transmission efficiency of entire network, and evaluation results indicate that the network throughput and coding gain based on HLCA have been improved.

6 Acknowledgment

This research was supported by the National Natural Science Foundation of China under Grant No.61003235, the Fundamental Research Funds for the Central Universities under Grant No. HEUCFZ1105 & No. HEUCF100607, the Natural Science Foundation of Heilongjiang Province of China under Grant No.F200902, and Harbin Scientific & Technological Innovation Research Funds under Grant No.2011RFQXG012.

7 References

[1] Ahlswede R, Cai N, Li S, et a1. Network information flow. IEEE Transactions on Information Theory,2000, 46(4), 1204—1216

[2] Anjum Naveed, Salil S. Kanhere. Cluster-based Channel Assignment in Multi-radio Multi-channel Wireless Mesh Networks, IEEE 34th Conference on Local Computer Networks (LCN 2009), 53-60

[3] M. Marina and S. R. Das. A topology control approach for utilizing multiple channels in multi-radio wireless mesh network. In Proc. IEEE Broadnets'05, 381-390

[4] K. Ramachandran, K. Almeroth, E. Belding-Royer, and M. Buddhukot. Interference aware channel assignment in multi-radio wireless mesh network. In Proc. IEEE INFOCOM, 2006

[5] Katti S, Hu W, Medard M. XORs in the air: Practical network coding[C]. Proc of ACM SIGCOMM, Pisa, Italy, 2006, 36(4):243-254

[6] Xinyu Zhang, Baochun Li. On the Benefits of Network Coding in Multi-Channel Wireless Networks. Skin, 2008

[7] Hang Su, Xi Zhang. Throughput-Gain Analysis of Network Coding in Multi-Channel Multi-Radio Weariless Network. IEEE ICC 2009

[8] Seok-Chul Kwon, Faramarz Fekri. A Novel Collaboration Scheme for Multi-Channel/Interface Network Coding. IEEE Transactions on Wireless Communications, 1(10), 2011, 188-198

Real Time Incubator Monitoring System Using Wireless Sensor Network

Karan Kolla, Rakesha R, Tejus S, Narendra Kumar G and Alice Abraham
Dept. of Electronics & Communication, UVCE, Bangalore University, INDIA
Email: tejus261990@gmail.com, gnarenk@yahoo.com

Abstract *- Although the Infant Mortality Rate of preterm babies has improved dramatically, they still remain vulnerable to many complications. The high rate of pathogenic infections in preterm babies addresses the need for a better prevention, detection and treatment facility. Modern Pediatrics Hospitals are equipped with incubators for post-delivery infants by providing them ambient environmental conditions which are essential and crucial for the healthy growth of the neonatal which was available only from experimental techniques owing to the complex human physiological phenomena till now. Pediatricians need real time data to monitor various parameters of the infants in these incubators for periodic medical requirements. Performance evaluation of this MANET based real time incubator monitoring system is simulated on NS2 considering each incubator as an access point equipped with various sensors which is analyzed with the Enhanced-Adhoc Ondemand Distance Vector(E-AODV) protocol. Tests have been carried out in the pediatrics hospitals to monitor new born infants during post-delivery complications. This paper proposes miniaturized sophisticated intelligent wireless sensors networks to provide real time monitoring of parameters like ECG, Blood Pressure, Oximetry, Pulse, and Temperature of the infant/patient and the humidity of the incubator. The simulation results are encouraging enough to warrant their use as a real time information delivery system that would greatly help Pediatrics consultants and particularly infants and can rely on this extremely potential Intelligent Sensors Network that provides accurate pathological parameters.*

Keywords: Access points, Wireless Sensor Networks, E-AODV, Incubators

1 Introduction

Infant primary defense against the pathogens is its immune systems which are underdeveloped in its first 6 months of gestational period. The only defense mechanism it has are the immunoglobulin antibodies being passed by mother through the placenta in the blood stream recognizing the pathogenic bacteria, virus, fungi entering the body, helps in preventing infections to the fetus which has a considerable amount of antibodies in its blood stream and they continue to receive them through breastfeeding and are passively immunized. Neonates tend to produce antibodies at a slow rates compared to adults for the first 6 months as new-borns are highly vulnerable of catching unwanted infections if not given proper care. It is advocated for the use of a monitoring system wherein real time physiological parameters of the new-borns are kept track of and facilitated with required conditions. New-borns are usually housed in an incubator system which requires a nurse or a physician to overlook it constantly. Babies that are born under less than 37 weeks of gestation period are generally considered to be premature. Neonates are susceptible to many diseases owing to the fact that their organs and immune systems are not matured. It has been studied that neonates tend to become ill severely mainly during and after birth. Such neonates who are vulnerable to attacks from foreign agents or in case of a self-system collapse are usually housed in a Neonatal Intensive Care Unit (NICU). Early detection and prevention of any complications arising reduces the Infant Mortality Rate. A better health can hence be provided to neonates by reliable and friendly monitoring systems and have found success in the health sector.

Premature babies are usually prone to physiological diseases apart from congenital diseases. Common problems are the ones related to Respiratory system like the Apnea, Anemia, Chronic Lung Disease, Cardiovascular system, the Bradycardia, and many others including Hernia, Hemorrhages, Jaundice, and Sepsis and Feeding intolerance. Thus it is critical to keep a real time tab on the baby's physiological parameters so that doctors and physicians can initiate the necessary procedures to save the baby.

A real time incubator system wherein the new-borns are constantly monitored is proposed. This real time monitoring of incubators and emergency notification is very useful and is relied upon by the Neonatologists as well as Pediatricians. It is of prime importance that the baby' vital signs are monitored continuously. Wide ranges of available physiological monitoring hardware help us get the biometric values in analog format and the system we propose transmits this data to the main base station in digital format. Each proposed incubator will have a set of six sensors which will provide the ECG, Blood Pressure, Oximetry, Pulse and Temperature of the infant in the incubator. Data collected from all the sensors is stored temporarily in the node provided in the Incubator itself and transmits the data to the Base Station over a wireless channel for analysis. The doctor or the nurse can access the data of a particular incubator on request for analyzing an irregularity found in the readings of any of

the parameters, and then an emergency notification is sent to the concerned pediatrician/consultant and family members.

2 SENSING PRINCIPLES

2.1 Blood Pressure

The Blood Pressure of the baby is measured using a CMOS Blood Pressure Sensor kept on the skin of the patient to monitor continuously. It measures the displacement of a surface caused by the movement of the blood vessel walls, due to its over pressure inside known as tonometry, Fig. 1.

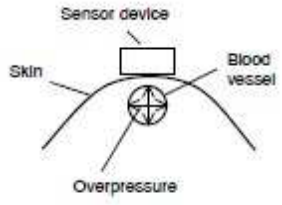


Fig. 1 Blood Pressure Sensor Schematic

Array of such force sensors are used to increase the accuracy applied on to the skin surface, the array which gives the highest value is considered. Each of these force sensors has suspended elastic membrane with a top electrode for capacitive read out of deflection of the surface made of CMOS technology where there will be reference element with respect to which the sensors will calculate the displacement, Fig. 2.

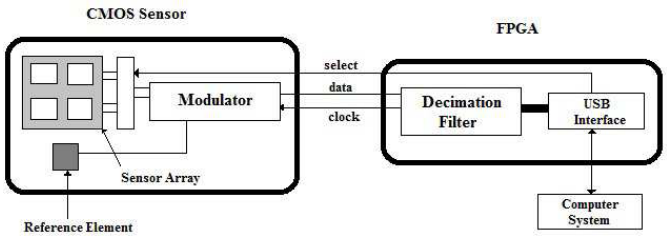


Fig. 2 Blood Pressure Sensor

2.2 Pulse Oximetry

Pulse Oximetry uses light wavelengths to measure oxygen saturation. Light is emitted from light sources/emitters which goes across the pulse oximeter probe and reaches the light detector. A finger is placed in between the light source/emitter and the light detector; the light will have to pass through the finger to reach the detector. Part of the light will be absorbed by the finger (the blood within) and the part not absorbed reaches the light detector, Fig.3.

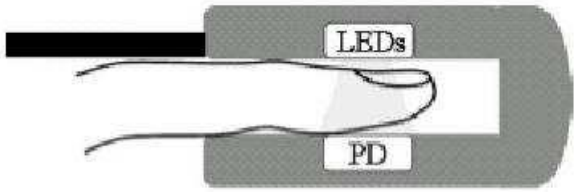


Fig. 3 Pulse Oximetry Sensor

The amount of light absorbed by the finger depends on the following factors:

1) The light absorbing substances (Hb in Blood) concentration.

2) The light paths length in the absorbing substance.

3) Oxyhemoglobin and Deoxyhemoglobins differential absorption of red and infrared wavelengths. Hence we will be using red light and IR light LEDs in the sensor probe.

It is observed from experiments that Oxy Hb absorbs more infrared wavelength than red wavelength and Deoxy Hb absorbs more red wavelength than infrared wavelength. In the finger, its not just the blood that absorbs the light, skin and other tissues also do absorb some light. Blood is pulsating and hence any variation in absorption must be due to blood in the finger being the properties of the Human Blood, the red and infrared wavelengths in measuring the Blood Oxygen Saturation.

A basic Oximetry Sensor uses 2 LEDs; one being a red LED and another IR LED and these are lit up in a particular timed sequence. A detector is used at the other end inner wall of the finger probe where the varying light is received. The readings from these sensors are continuously fed to a computer and are available in the form of a graph called the plethysmographic trace (pleth) using complex calculations by the computer.

2.3 Temperature

Sensors widely used in the medical applications to measure the temperature of the infant are generally of the Contact Sensors type. While using these sensors, one infers the readings by considering the sensors and the body in contact are in thermal equilibrium.

2.4 ECG

ECG, electrocardiogram is a recording of the electrical activity of the heart (cardiac) muscle as obtained from the surface of the skin. Electrocardiograph (ECG) is one of the most widely used biomedical sensing procedures to date. The heartbeat is the definitive indicator for a wide range of physiological conditions. Current flows in the form of ions in

the body and signals contraction of cardiac muscle leading to the heart's pumping action. Quasar's sensor is a compact and most commonly used ECG sensor that does not require skin preparation, gels, or adhesives. It includes not only a sensing device, but also signal conditioning circuitry such as low noise amplifiers and voltage reference chips.

2.5 Humidity

Humidity sensors are used to find the amount of water vapor content in its surroundings. The capacitive humidity sensors consist of a substrate on which a thin film of polymer or metal oxide is deposited between two conductive electrodes. The sensing surface is coated with a porous metal electrode to protect it from contamination and exposure to condensation. The substrate is typically glass, ceramic, or silicon. The Humidity is measured by virtue of the dielectric constant of this material varying according to the variations of the environment.

3 Communication Channel

This paper focuses on establishing Enhanced AODV protocol based communication between the Doctors/consultants and the incubators. A communication channel is created between each of the nodes housed on each incubator and the base station where all the data is received and stored. 3 routing tables are used by Enhanced-AODV. The 3 tables being Routing Table, Distance vector table and Path memory table for each node. This protocol uses 4 types of messages, the ping message (PNG) will be broadcasted by the nodes to know the direction and distance of the neighboring nodes, the router request message (RREQ) will be sent from each node once the direction and distance of the neighboring nodes are determined, the RREQ message will decide the final path for the packets and the Route Reply Message (RREP) will confirm the path and ensures that the connection is established. The Acknowledgement (ACK) message is transmitted by the receiving node after the reception of each packet. The Distance Vector Table contains the information about the direction and distance of the neighboring nodes. The Path memory table deals with the number of nodes per sector and help is creating the path for the packet. The Enhanced-AODV ensures optimized communication between the nodes and the base Station. The Data received by the Base Station is stored, analyzed and compared with the standard values. Each concerned authority is provided with a PDA using which they can access the readings as and when required. When the authority wants to access the data, they can send a request using their PDA. The request message will contain the node (Incubator) IP address. Depending on this request, the Base Station will send the reading to the PDA of the authority. The actual biometric parameters will be available on the PDA after this.

4 Implementation

A simplified network environment is created wherein the sensors on the infant are feeding the real time physiological parameters to the Physicians on their PDAs routed via a common base station housed at the hospital's data center. Each Incubator is fitted with an on-board module referred to here as node, whose only job is to collect the data from the various sensors, convert them into digital signals and transmit wirelessly using E-AODV routing protocol to the next immediate node or to the base station directly. The sensors are wired to the incubators module. These modules are equipped with a small memory storage capacity and are used to temporarily store the collected parametric data. The module communicates with the base station and upon successful establishment of communication channel begins the transfer of data. The Base Station is a server set up on the hospital premises which handles all the incoming information about the infants and maintains a database which is equipped with a state of the art application that does the actual monitoring of the infants parameters. It is pre-fed with standard levels of a healthy infant which it uses to compare with the incoming data from all the incubators and generates alerts the concerned doctor and the nurse station notifying immediately about the emergency generated to the Doctors cell phone and family members from the Base Station.

The proposed system will have the doctor PDAs be installed with an application which assists them in the treatment process. The application stalls all the current processes running on the PDA and receives the messages from the Base Station seeking attention from the doctor. The doctor can as well access the database maintained by the Base Station whenever in need to monitor real time readings of the infant parameters and pleth.

5 Simulation and Results

Network Simulator-2 was used to simulate the communication network in the proposed system. The environment is simulated in a 2-D topology system using E-AODV Protocol considering 6 nodes as constituted by 6 different sensors, the data from each of these nodes is sent over wired channel to the module housed on the incubator. The data from each of these nodes is sent to the Base Station over a wireless channel with a packet size of 5kB and an interval arrival time of 10 packets per second. The data traffic generation is performed using File Transfer Protocol (FTP) Model and is sent over a multiplexed wireless channel of 1ms duration per node.

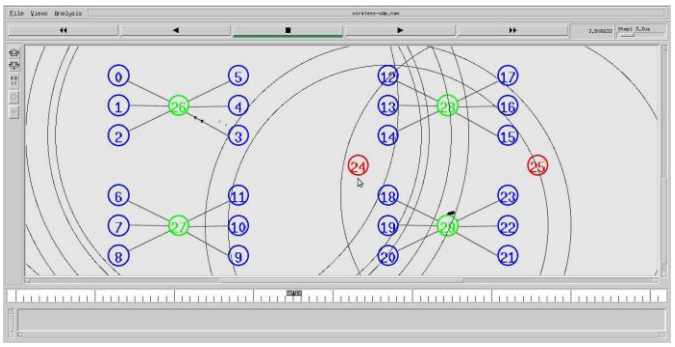


Fig. 4 Simulation showing the data transfer from individual sensors to the incubator node

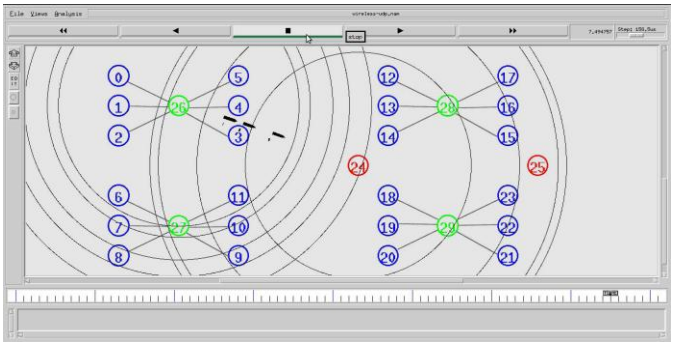


Fig. 5 Simulation showing the data transfer from incubator node to the nearest access point.

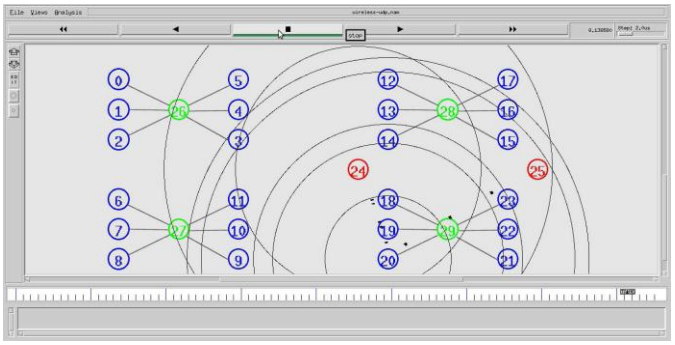


Fig. 6 Simulation showing the data transfer from access point to the main base station.

6 Conclusion and Future Work

The proposed Real time incubator monitoring system collects various biometric parameters of the infant in the incubator and sends this over a wireless network to the Base Station. The proposed system ensures proper monitoring of the infants and helps in case of emergencies as well using enhanced-AODV protocol.

The future work is focused on the implementation of the monitoring system to monitor all the pathological parameters of infants in all the hospitals with common physician/consultant.

7 References

[1] Victor Shnayder, Bor-rong Chen,Konrad Lornicz, Thaddeus R. F. Fulford- Jones and Matt Welsh. "Sensor Networks for Medical Care". 8th-11th June 2011, LISS, Bejing, CHINA.

[2] Emil Jovanov, Dejan Raskovic, John Chapman, John Price, Anthony Moore, Abhishek Krishnamurthy. "Patient Monitoring Using Personal Area Networks of Wireless Intelligent Sensors".

[3] Chulsung Park, Pai H., Chou Ying Bai, Robert Matthews, and Andrew Hibbs. "An Ultra Wearable, Wireless, Low Power ECG Monitoring System".

[4] Tia Gao, Christopher Pesto, Leo Selavo, Yin Chen. "Wireless Medical Sensor Networks in Emergency Response: Implementation and Pilot Results".

[5] Wei Chen, Sibrecht Bouwstra1, Sidarto Bambang Oetomo1, and Loe Feijs. "Intelligent Design for Neonatal Monitoring with Wearable Sensors".

[6] Shreyas S Vidyarthi, Vijayalakshmi S, Vijaeendra Simha G A, G Narendra Kumar. "Energy Efficient AODV Protocol using Directional Antennas for Multi-Rover Systems". ICUMT, Budapest, Hungary.

[7] G. Narendra Kumar, G. K Rama Murali, Shreyas S Vidyarthi, Vijayalakshmi. S, Aparajitha Murali. "Manet and Spread Spectrum based communication for Multi-Rover System".

The Model Analysis of VANET at Intersections Based on Markov Process

Hengliang Tang, Hao Wu, and Jia Meng
State Key Laboratory of Rail Traffic Control And Safety, Beijing Jiaotong University, Beijing, China

Abstract - *VANET (Vehicular Ad Hoc Network) is a special case about MANET (Mobile Ad Hoc Network), has also received extensive attention as a part of ITS(Intelligent Transportation System). Mobility model is the basic research in VANET, our objective is set up a vehicular mobility model reflecting the real-world scene. In this paper, we propose a mobility model at intersection and present mathematical analysis by the use of Markov analysis method, finally provide the validation and simulation. We compare the predicted TFO (Traffic flow occupancy) probability distribution to the observed results, and the errors between them are analyzed. Both simulation and real observation data are used to demonstrate the effectiveness of the method. The prediction results can help decide the red-time controlling strategies. The results show the proposed model is approximate to the intersection in actual urban environment and can be used as a practical one.*

Keywords: Vehicular Ad Hoc Network; Mobility model; Markov process; Traffic flow occupancy

1 Introduction

Vehicular communication is seen as a key technology for improving road safety and comfort through Intelligent Transportation Systems (ITS). The growing interest toward the possible applications of wireless technologies to a vehicular environment has recently led consortia (US VII [1], EU C2C-CC [2]) and standardization bodies (IEEE [3]) to develop technologies and protocols for data transmission between vehicles and between vehicles and road infrastructures. Vehicular ad hoc networks have been envisioned to be useful in road safety and many commercial applications.

VANETs are self-organizing communication networks built up from moving vehicles, and are thus characterized by very high speeds and limited degrees of freedom in nodes movement patterns. Such particular features often make standard networking protocols inefficient or unusable in VANETs. The current major objective of protocols developed for VANET is to benefit from the messages exchanged between cars to alter traffic, either for safety issues with advanced safety messages, or for traffic management in order to avoid traffic jams.

A critical aspect in a simulation study of VANETs is the need for a mobility model reflecting the real behavior of vehicular traffic. Traffic condition has become more and more serious in most cities of our country. Solving transportation problems has been very important part of city transportation design. But city transportation is rather complex, so developing relative mathematical models and make use of computer to process numerical simulation is important method of solving the complex problem.

The rest of the paper is organized as follows: Section II shows the related work of mobility model in VANET. Scenario model at intersection including traffic lights and other obstacles in urban environment is described in Section III. Section IV proposes a mobility model at intersection and classifies vehicles various of groups according the direction of vehicles. Section V presents numerical analysis aim to the mobility model. The validation and simulation are discussed in Section VI, and Section VII concludes the paper.

2 Related Work

In literatures, vehicular mobility models are usually classified as either macroscopic or microscopic[4].The macroscopic description models gross quantities of interest, such as vehicular density or mean velocity, treating vehicular traffic according to fluid dynamics, while the microscopic description considers each vehicle as a distinct entity, modeling its behavior in a more precise, but computationally more expensive way.

However, a micro-macro approach may be seen more as a broad classification schema than a formal description of the models'functionalities in each class. A more precise way that we suggest for looking at mobility models consists in identifying functional blocks: motion constraints, traffic generator, time and external influences. On the one hand, motion constraints describe the relative degree of freedom of each vehicle. Macroscopically, motion constraints are streets or buildings, but microscopically, constraints are modeled by neighboring cars, pedestrians, or by modelization's diversities either due to the type of car or to the driver's habits. On the other hand the traffic generator defines different kinds of cars and deals with their interactions according to the environment under study. Macroscopically, it models traffic densities, speeds and flows, while microscopically it deals with properties like the inter-distance between cars, acceleration, braking and overtaking. Another important aspect of realistic motion modeling is time, which can be seen as the third functional block that describes different mobility configurations for a specific time of the day or day of the week. Finally, we also have to add a fourth fundamental block, the External Influence, modeling the impact of a communication protocol or any other source of information on the motion patterns.

It is important to use a realistic mobility model so that results from the simulation correctly reflect the real-world

performance of a VANET. A realistic mobility model should consist of a realistic topological map which reflects different densities of roads and different categories of streets with various speed limits. Another important parameter should be modeled is the obstacles. In the real world, a vehicle node is typically constrained to streets which are separated by buildings, trees or other objects. Such obstructions often increase the average distance between nodes as compared to an open-field environment. In addition, each vehicle needs to decide a turning direction at the intersection (e.g. turn left, turn right or go straight). Such a turning model could have an effect on the congestion of the road as well as the clustering of the vehicles. Furthermore, a smooth deceleration and acceleration model should be considered since vehicles do not abruptly break and move. Many prior studies [5], [6] have shown that a realistic mobility model with sufficient level of details is critical for accurate network simulation results.

3 Scenario Model

As is known to all, the most prominent advantage of Ad Hoc network is nodes can communication randomly in free movement, but vehicular Ad Hoc network is different from mobile Ad Hoc Network, especially in the urban traffic environment, the nodes in VANET have their own characteristics of the motion, and also often cause signal attenuation even communication interrupt because of the building block in the process of communication, the characteristics of movement also led to the different network topology changes, the real-world network scene and mobility model need to consider many factors such as road conditions, urban condition, traffic speed, vehicle density and obstacles, etc.

Owing to the activities of vehicles on the highway can be regarded as a kind of one-dimensional restricted movement, the model is relatively simple, the speed can be limited in a range and the direction is also single. We can use a random way point model, and the movement direction from [0, 2π] Angle to zero or π interval, which can realize the highway mobile model.

Vehicular mobility model is more complex in urban environment than on the highway. Traffic situation in cities is very complex, especially the motion of vehicles suffers from a lot of restriction, which mainly is caused by the vehicle driving rules and fixed urban road topology in actual urban environment. If we still use random waypoint model in VANET simulation, it will be to produce all sorts of freak, such as vehicles through the wall or rebound from the wall, the vehicles drive into river away from the streets, the vehicles change the direction randomly and even cause collision, etc. These phenomena are obviously not consistent with the reality. So in urban environment, the nodes do not choose road topology and traffic rules as a reference, their movement will be in the distortion state, and influence the accuracy of the simulation results seriously.

Through the analysis of real-time traffic information, automatically select the best road driving route in order to alleviate traffic jam. By the use of the sensors and camera system, we can perceive about the Surrounding environment and make a quick adjustment when meeting obstacles or driving conditions change.

4 Mobility Model in Intersection

Intersection is made up of two parts: horizontal street and vertical street. Horizontal street is the East-West direction and vertical street is the South-North direction. Mobile nodes can move along the horizontal and vertical grid of the topology map. At the intersections of the streets, the mobile nodes choose to turn right or left or go forward (remain unchanged) according to a certain probability. The speed of mobile nodes in a certain moment is related with the previous moment. In addition, the speed of vehicles in the same lane also has relation each other.

According to vehicular mobility model, considering typical case in actual urban traffic, we establishes a mobility model at intersection in urban road traffic as follows: vertical street controlling the north and south direction and s horizontal street controlling the east and west direction, each direction has two lanes, a total of four lanes. In order to facilitate description, the streets are assumed as grids that have regular shape, meanwhile Rectangular blocks instead of buildings. Figure 2 shows the model.

In order to simplify the modeling, assuming there are N mobile nodes in the model area, distributed randomly in the lanes of model area, each node has the same signal transmission range and the same speed. The model complies with limited steps execution in discrete time and the transition of limited state in the system according to the different probability, so modeling as a Markov (DTMCs) model. The roads are bidirectional and the intersections are regulated by means of traffic lights. At each intersection, the vehicle chooses its next direction following a Markov chain whose probabilities are calculated based on road segments attraction weights. Two coefficients that take into account roads congestion and vehicle previous movements are defined in the model; they are applied to the Markov chain in order to make displacements more realistic.

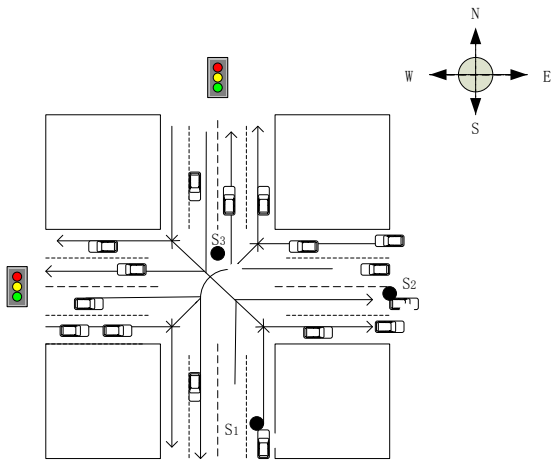


Fig.1 vehicular mobility model at intersection

As shown in figure 1, intersection is divided into four directions, east and west direction in horizontal street, north and south direction in vertical street. Vehicles in every direction are divided into three groups, set to D =

{left, right, straight}, on behalf of turn left, turn right and go straight respectively, then a intersection is divided into twelve group, set to M={ N→E , N→S , N→W , S→W , S→N , S→E , W→N , W→E , W→S , E→S , E→W , E→N}, every direction also have corresponding transition probability P_{ij}, which represents transition probability of direction a group belongs to in the course of driving. For example, P_{NS} represents the radio between the vehicles from north to south and the total vehicles from north .Therefore, $P_{NE}+ P_{NS} +P_{NW}=1$. At the same time, the vehicles with the same destination are called groups with a high social relationship.

5 Numerical Analysis

5.1 Some Definitions

Definition 1: Any vector μ=(μ1, μ2,……,μn), each element is non-negative, and the sum of all elements is 1, is defined as probability vector.

Definition 2: In one matrix $P=(Pij)_{n\times n}$, if every row vector is a probability vector, we call this matrix probability matrix or stochastic matrix.

Definition 3: If non-zero vector μ=(μ1, μ2,……,μn) multiplies matrix P, the result is also μ , i.e. μP =μ, we call μ the fixed point or balance point of matrix P.

Definition 4: As any probability matrix P is concerned, matrix P multiplies P m times, the result is Pm, if all elements of Pm (m is a natural number and m>1) are positive numbers, we call the Markov chain the irreducible ergodic Markov chain.

5.2 Markov Process Analysis

A discrete Markov chain model can be defined as (S, P, X(0)). S corresponds to the state space, P is a matrix representing transition probabilities from one state to another, and X(0) is the initial probability distribution of the states in S . Provided that we study a simple Markov Process, and there are n states in the system: S_1, S_2, S_3, ……, S_n, i. e. state space is a limited set of (S_1, S_2, S_3, ……, S_n). If the result of the ith test is S_i, we say that the system keeps state S_i in ith step, as shown in Fig.5. If the system keeps state S_i at some moment, and transit to S_j the next moment, and the transition probability is P_{ij}, (in Fig.2, P_{ii} and P_{jj} are the transition probabilities to keep the same states, S_i and S_j), after state S_i occurred, the transfer probability P_{ij} that state S_j occurring arrayed as transfer matrix P as follows:

$$p=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (1)$$

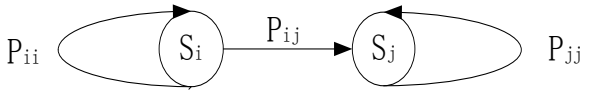


Fig.2 Probability transmit diagram

where P_{ij} = transition probability of state S_i transmit to S_j; i,j=l,2, ... n

If the system transits from state Si to state Sj through k steps, the transition probability is $P_{ij}^{(k)}$, and $P_{ij}^{(k)}$ are arrayed as a matrix $P^{(k)}$, we call it k-step transition matrix, and k-step transition matrix equal to P^k; i.e. $P^{(k)}=P^k$, expressed as follows:

$$P^{(2)} = P \bullet P = P^2 \quad (2)$$

$$P^{(n)}=P\bullet P \ \ldots\ldots P= P\bullet P^{n-1}= P^{n-1}\bullet P = P^n \quad (3)$$

Suppose $X^{(0)} = (X_1^{(0)}, X_2^{(0)}, \ldots X_n^{(0)})$ is the initial probability distribution of the states in S, and after k th step's state transition, the state vector is $X^{(k)}=(X_1^{(k)}, X_2^{(k)}, \ldots X_n^{(k)})$, $P^{(k)}$ illustrates the k-step probability transition matrix. Then:

$$X^{(k)} = X^{(0)} \bullet P^{(k)} \quad (4)$$

According to equations, the Markov prediction models:

$$\begin{bmatrix} X_1^{(k+1)} \\ X_2^{(k+1)} \\ \vdots \\ X_n^{(k+1)} \end{bmatrix}^T = \begin{bmatrix} X_1^{(k)} \\ X_2^{(k)} \\ \vdots \\ X_n^{(k)} \end{bmatrix}^T \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (5)$$

Where k=1, 2,…n.

5.3 Steady State Probability

The limit of k-step probability transition matrix Pk of.the irreducible ergodic Markov chain is existent, i.e. the system approach steady state gradually, which has nothing to do with the initial condition of the system. We can also determine the future state of the system, and its probability in equilibrium state, even we don't know the initial state of the system. Markov Chain procedure can help us deduce the system state after k steps transition, so that we can apply Markov analysis method in traffic situation forecasting and controlling policymaking.

According to Definition 3, the equilibrium equation is as follows:

$$X\bullet P = X \quad (6)$$

Where: X = equilibrium state vector; P = matrix of state transition probability.

The solution vector X of this equation indicates the probability distribution of the equilibrium state.

Markov analysis method is widely applied to many fields, such as: in market prediction [13]-[15], and in traffic transition prediction [16]. We will use this method in traffic situation prediction at signalized intersection in the following.

Suppose P_{ij} represents the transition probability from I to J (I，J=N，S，E，W), then the transition probability matrix P obtained as follows:

$$P=\begin{bmatrix} 0 & P_{NS} & P_{NE} & P_{NW} \\ P_{SN} & 0 & P_{SE} & P_{SW} \\ P_{EN} & P_{ES} & 0 & P_{EW} \\ P_{WN} & P_{WS} & P_{WE} & 0 \end{bmatrix} \quad (7)$$

Where P_{ij}=transition probability of state I transit to state J; I，J=N，S，E，W.

$$\begin{bmatrix} \delta_{NN_1} \\ \delta_{SN_1} \\ \delta_{EN_1} \\ \delta_{WN_1} \end{bmatrix}^T = \begin{bmatrix} \delta_{NC} \\ \delta_{SC} \\ \delta_{EC} \\ \delta_{WC} \end{bmatrix}^T P = \begin{bmatrix} \delta_{NC} \\ \delta_{SC} \\ \delta_{EC} \\ \delta_{WC} \end{bmatrix}^T \begin{bmatrix} 0 & P_{NS} & P_{NE} & P_{NW} \\ P_{SN} & 0 & P_{SE} & P_{SW} \\ P_{EN} & P_{ES} & 0 & P_{EW} \\ P_{WN} & P_{WS} & P_{WE} & 0 \end{bmatrix} \quad (8)$$

Where δ_{IC} represents probability of TFO of driving to direction I of current cycle; δ_{IN_1} represents probability of TFO of driving to direction I of the next one cycle; I= N，S，E，W

$$\begin{bmatrix} \delta_{NN_k} \\ \delta_{SN_k} \\ \delta_{EN_k} \\ \delta_{WN_k} \end{bmatrix}^T = \begin{bmatrix} \delta_{NN_1} \\ \delta_{SN_1} \\ \delta_{EN_1} \\ \delta_{WN_1} \end{bmatrix}^T P^{k-1} = \begin{bmatrix} \delta_{NC} \\ \delta_{SC} \\ \delta_{EC} \\ \delta_{WC} \end{bmatrix}^T P^k = \begin{bmatrix} \delta_{NC} \\ \delta_{SC} \\ \delta_{EC} \\ \delta_{WC} \end{bmatrix}^T \begin{bmatrix} 0 & P_{NS} & P_{NE} & P_{NW} \\ P_{SN} & 0 & P_{SE} & P_{SW} \\ P_{EN} & P_{ES} & 0 & P_{EW} \\ P_{WN} & P_{WS} & P_{WE} & 0 \end{bmatrix}^k \quad (9)$$

Where k=1, 2,…n.

δ_{IN_k} represents TFO of driving to direction I of the next k cycle in the future; I= N, S, E, W; k=1,2,…n

$E_{\delta I}$ represents errors between observed values and forecasted values of traffic flow occupancy to direction I. $E_{\delta I}$ are calculated by equation:

$$E_{\delta I}=\frac{\delta_{IO}-\delta_{IF}}{\delta_{IO}}\times 100\%, \quad I=N,S,E,W \quad (10)$$

δ_{IO} represents observed TFO of driving to direction I, I=N, S, E, W

δ_{IF} represents forecasted TFO of driving to direction I, I=N, S, E, W

6 Validation and Simulation

Table I. The Observation Numbers of Different Direction

Travel direction	N	S	E	W
N	0	1064	223	232
S	901	0	145	455
E	383	60	0	530
W	522	461	503	0

We can obtain P matrix through the TABLE I.

$$P=\begin{bmatrix} 0 & P_{NS} & P_{NE} & P_{NW} \\ P_{SN} & 0 & P_{SE} & P_{SW} \\ P_{EN} & P_{ES} & 0 & P_{EW} \\ P_{WN} & P_{WS} & P_{WE} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.70 & 0.15 & 0.15 \\ 0.60 & 0 & 0.10 & 0.30 \\ 0.25 & 0.40 & 0 & 0.35 \\ 0.35 & 0.30 & 0.35 & 0 \end{bmatrix} \quad (11)$$

The traffic detector at an intersection of the forth crossing in Beijing collected the observation data. Table I presents the number of vehicles form one direction to another at intersection within ten minutes. Predicted values and observed values of TFO of ten cycles in the future are illustrated in Table II, one cycle is one minutes.

Table II. Predicted values and observed values of TFO

cycle	δ_{NO}	δ_{NF}	δ_{SO}	δ_{SF}	δ_{EO}	δ_{EF}	δ_{WO}	δ_{WF}
1	0.2998	0.3114	0.3506	0.3569	0.1504	0.1483	0.1992	0.1834
2	0.3023	0.3154	0.3417	0.3323	0.1446	0.1466	0.2114	0.2057
3	0.2905	0.3080	0.3512	0.3411	0.1449	0.1525	0.2134	0.1983
4	0.3081	0.3122	0.3578	0.3361	0.1406	0.1497	0.1935	0.2019
5	0.3075	0.3098	0.3447	0.3390	0.1474	0.1511	0.2004	0.2001
6	0.3039	0.3112	0.3532	0.3373	0.1481	0.1504	0.1948	0.2011
7	0.2966	0.3104	0.3553	0.3383	0.1473	0.1508	0.2008	0.2005
8	0.3014	0.3109	0.3524	0.3377	0.1474	0.1506	0.1988	0.2008
9	0.3069	0.3106	0.3509	0.3381	0.1464	0.1507	0.1958	0.2006
10	0.2982	0.3108	0.3482	0.3379	0.1491	0.1506	0.2045	0.2008

Table III. Errors Between Observed Values and Forecasted Values of TFO

Cycle	$E_{\delta N}$ (%)	$E_{\delta S}$ (%)	$E_{\delta E}$ (%)	$E_{\delta W}$ (%)
1	-4.87	-1.80	1.40	7.93
2	-4.33	2.75	0	2.70
3	-6.02	2.88	-5.24	7.08
4	-1.33	6.06	-6.47	-4.34
5	-0.75	1.65	-2.51	0.15
6	-2.40	4.50	-1.55	-3.23
7	-4.65	4.78	-2.38	0.15
8	-3.15	4.17	-2.17	-1.00
9	-1.21	3.65	-2.94	-2.45
10	-4.23	2.96	-1.01	1.81

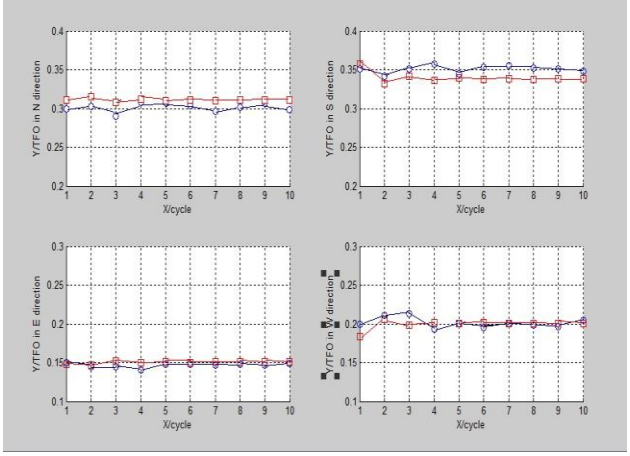


Fig.3 Simulation about TABLE II

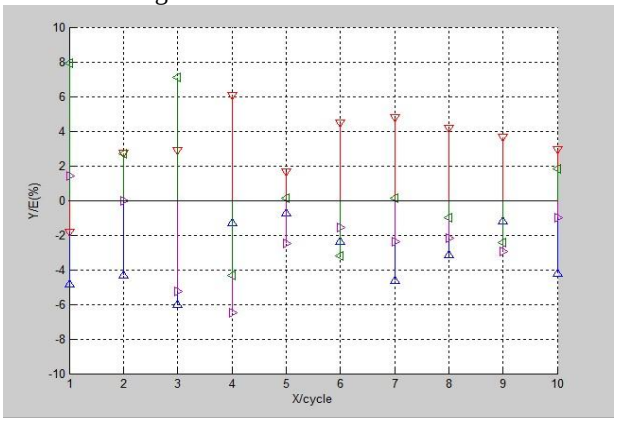


Fig.4 Simulation about TABLE III

We deal with the data in TABLE II and III respectively, obtain the simulation results, presented by figure 3 and figure 4 respectively.

From the comparison we can see that the maximum error is not greater than 10%; So Markov analysis method is feasible for the traffic situation at intersection in urban environment.

7 Conclusion

Mobility model is an important tool for simulation and analysis in VANET. In this paper we imitate a Scenario model at intersection including traffic lights and other obstacles in urban environment, and then design a mobility model based on the scene according to the different direction of vehicles. Finally we analysis the mobility model through the Markov Process and validate the rationality. According to the realistic issue of signalized intersection traffic flow, we put forward an estimate model of different traveling directions based on Markov process. Its assumption is that transfer probability keeps fixed. Markov analysis method is employed to predict the possibility of vehicles from different directions entering into intersection, and the prediction results do help the real-time control strategy making. Both prediction and real observation data are used to demonstrate the effectiveness of the method. The prediction results can help decide the real-time controlling strategies. The results show the proposed model is approximate to the intersection in actual urban environment and can be used as a practical one. The future work is that the researchers ought to build more realistic mobility models based on mobility characteristics of vehicles and satisfy scalability of scenario model in intersection.

8 Acknowledgement

This paper is supported by the State Key laboratory of Rail Traffic Control and Safety (Contract No.RCS2010ZZ004), Beijing Jiaotong University. This paper is also supported by program for Changjiang Scholars and Innovative Research Team in University under Grant No. IRT0949 and the Joint State Key Program of the National Natural Science Foundation of China and the National Railway Ministry of China (Grant No. 60830001).

9 References

[1] Vehicle Infrastructure Integration (VII), http://www.its.dot.gov/vii/.

[2] Car-2-Car Communication Consortium (C2C-CC), http://www.car-to-car.org/.

[3] Wireless Access for the Vehicular Environment (WAVE),http://grouper.ieee.org/groups/802/11/Reports/tgp update.htm.

[4] D. Helbing, "Traffic and Related Self-driven Many-particles Systems", Rev. Modern Physics, Vol. 73, pp. 1067-1141, 2001.

[5] Amit Kumar Saha and David B. Johnson. Modeling mobility for vehicular ad hoc networks. In Proc. of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks (VANET), October 2004.

[6] John Heidemann, Nirupama Bulusu, Jeremy Elson, Chalermek Intanagonwiwat, Kun chan Lan, Ya Xu, Wei Ye, Deborah Estrin, and Ramesh Govindan. Effects of detail in wireless network simulation. In Proc. of Communication Networks and Distributed Systems Modeling and Simulation Conference, January 2001.

[7] Qiaoru Li; Lianyu Wei; Shoufeng Ma; The model analysis of vehicles situation and distribution in intersections based on Markov process. Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE ,Vol.2,pp.1076-1080,2003.

[8] Jiang Lianfu, Wang Yanzhang, Manlie. Application of PRfmcated Concrete Based en Markov Chain in Marketing Rcdiction. Joumal of Dalian University of technology Vo1.42. Na.5.2002.9.

[9] M.E.Faulad van. M.NemaroIlahi. Optimization of Green-times at an Isolated Urban Crossroads. The European Physical Joumal B Condensed Matter. Vo1.22.No.3.2001.8.

[10] A. Mahajan, N. Potnis, K. Gopalan, and A. A. Wang, "Urban mobility models for vanets," in Proc. of the IEEE Workshop on Next Generation Wireless Networks (WoNGeN), Dec. 2006.

[11] R. Baumann, S. Heimlicher, and M. May, "Towards realistic mobility models for vehicular ad-hoc networks," in 2007 Mobile Networking for Vehicular Environments, May 2007, pp. 73–78.

[12] I. Stepanov, P.-J. Marron, and K. Rothermel, "Mobility modeling of outdoor scenarios for manets," in Proc. of the 38th annual Symposium on Simulation ANSS '05, Apr. 2005, pp. 312–322.

[13] Chief Editor: Hu Xucheng. Market Occupancy (Markov) Forecasting Methods, 80 Practical Market Forecasting Methods . Peking Economic Institute Press, 1993.

[14] Meng Zhaopeng, Liu Li, Luo Peng, Study on Computer Sales Market Model Based on Markov Analysts, Decision and Decision Support Synem.Vo1.7 No.1,pp. 88-93.

[15] Jiang Lianfu, Wang Yanzhang, Manlie. Application of Prefbrecated Concrete Based on Markov Chain in Marketing precdiction. Journal of Dalian University of technology Vol.42, No.5, 2002.9.

[16] Cheng Zude, A study on Shifting Traffic Volume of the Vehicular Ferry and Yangpu Bridge by Grey Theory and Operations Research Method, Joumal of Maritime University, Vol.15, No.3, 1994.9.