ISSN: 1945-3078 Volume 2, Number 6, June 2010

Scientific Research

Wireless Sensor Network

Chief Editor : Kosai Raoof

ISSN: 1945-3078

www.scirp.org/journal/wsn/

Journal Editorial Board

ISSN 1945-3078 (Print) ISSN 1945-3086 (Online)

http://www.scirp.org/journal/wsn/

Wireless Sensor Network, **2010, 2, 411-491**
Published Online June 2010 in SciRes (http://www.SciRP.org/journal/wsn/)

TABLE OF CONTENTS

Volume 2 Number 6 June 2010

Wireless Sensor Network (WSN)

Journal Information

SUBSCRIPTIONS

The *Wireless Sensor Network* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:

Print: $50 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: wsn@scirp.org

Wireless Sensor Network, **2010, 2, 411-418**
doi:10.4236/wsn.2010.26053 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Heuristic Spectrum Assignment Algorithm in Distributed Cognitive Networks

Li Yu, Cong Liu, Zuhao Liu, Wenyu Hu
Huazhong University of Science and Technology, Department of Electronic Information Engineering, Wuhan National Laboratory for Optoelectronics, Div Commun and Intelligent Networks, Wuhan, China
E-mail: *hustlyu@mail.hust.edu.cn*, {*liuconggg, liuzuhao, babywinnerhu*}*@gmail.com*
Received December 14, 2009; *revised February* 5, 2010; *accepted February* 10, 2010

Abstract

Cognitive radio is an exciting emerging technology that has the potential of dealing with the urgent requirement and scarcity of the radio spectrum. Although having multiple radio interfaces and available spectrum bands can generally increase the effective throughput, a problem arises as to what the best strategy to dynamically assign available bands to secondary users for maximizing throughput by minimizing the interference, and what the best scheme to allocate the spectrum holes to unlicensed users to maximize the fairness. This paper presents a distributed and heuristic spectrum assignment algorithm for multi-radio wireless cognitive networks in a cognitive network environment. The proposed algorithm (Fairness Bargaining with Maximum throughput, FBMT) considers the problems including system throughput and the fairness. Extensive simulation studies in 802.11 based multi-radio cognitive networks have been performed. The results indicate that the proposed algorithm can facilitate a large increase in network throughput and acquire a good fairness performance in comparison with a common spectrum assignment mechanism that is used as a benchmark in the literature.

Keywords: Cognitive Network, Distributed Spectrum Assignment, Throughput, Fairness, FBMT

1. Introduction

As wireless technologies continue to grow, more and more spectrum resources will be needed. Within the current spectrum regulatory framework, however, all of the frequency bands are exclusively allocated to specific services, and no violation from unlicensed users is allowed. A recent survey of spectrum utilization made by the Federal Communications Commission (FCC) has indicated that the actual licensed spectrum is largely underutilized in vast temporal and geographic dimensions [1]. For instance, a field spectrum measurement taken in New York City has shown that the maximum total spectrum occupancy is only 13.1% from 30 MHz to 3 GHz [2,3]. Similar results, obtained in the most crowded area of downtown Washington, D.C., indicated occupancy of less than 35% of the radio spectrum below 3 GHz. Moreover, the spectrum usage varies significantly in various time, frequency, and geographic locations.

1.1. Spectrum Opportunity

Spectrum segments in a licensed band that are currently unused by its primary users (PU) are referred to as the spectrum opportunity/holes. Spectrum holes appear as useable spectrum bands to a secondary user (SU) with respect to the licensed band in question. For a secondary user, given its locations, a set of spectrum holes can be available at a given time. Such a set of available bands are referred to as the spectrum opportunity of the secondary user.

Spectrum utilization can be improved significantly by allowing a secondary user to utilize this spectrum opportunity when the primary user is absent. Cognitive radio (CR), as an agile radio technology, has been proposed to promote the efficient use of the spectrum [4]. By sensing and adapting to the environment, a CR is able to fill in spectrum holes and serve its users without causing harmful interference to the licensed user.

Amid these usage trends, it is desirable to introduce dynamic spectrum allocation strategies so that the unused segments of a licensed band which owned by its primary users can be used by unlicensed or secondary users. A fixed spectrum policy was sufficient in the past but with the increasing disparity in utilization rate between the licensed and unlicensed bands, dynamic spectrum alloca-

tion policies are needed.

1.2. Related Work

Dynamic spectrum allocation strategies enable secondary user to sense local usable bands (including the bands without license), and to use them by some certain rules/ algorithms without interference to primary users. According to the difference of structures of networks, the algorithms can be divided into two categories: the centralized and the distributed. In centralized spectrum allocation algorithms, due to higher computing complexity, the centralized control devices become a bottle-neck because they are not competent for such complex computing work. So the distributed algorithms are much better than the centralized algorithm in spectrum assignment. Most distributed algorithms adopt heuristic assignment methods, in which algorithm's astringency is a very important target, and the higher the astringency of algorithms is, the better the algorithm can adapt to the time-varying systems. Beside astringency, in [5-7], Zheng *et al.* put forward other two basic targets, Max-Sum-Bandwidth (MSB) and Max-Min-Bandwidth (MMB), these targets are used to describe the improvement of throughput that the algorithm brings to system. In [8], Cao *et al.* proposed a FBFP (fairness bargaining with feed poverty) algorithm to Ad Hoc networks, in this algorithm, secondary users are divided into different groups based on their regions. Simulation result present that this algorithm can decrease the computing complexity effectively but increase the system cost because it needs extra packet to set up and maintain the groups. In [9,10], Wang and Liu proposed several algorithms which based on the max-spectrum-employ rate and max- fairness, but the throughput of system is absent in their consideration. In [11], by using the time-continuous Markov chain to model the spectrum access, Xing *et al.* proposed a random spectrum access algorithm, but this algorithm is only applicable in simple systems which consist of fewer primary users and secondary users.

Motivated by this, considering both the fairness and throughput, we bring forward a heuristic algorithm called Fairness Bargaining with Maximum Throughput (FBMT). The simulation will demonstrate that FBMT not only contribute to promote the systems fairness, but also improve the system throughput.

The rest of this paper is organized as follows. In Section 2, a completely distributed system model will be briefly reviewed, and a mathematic model will be designed for the spectrum assignment and coordination process. In Section 3, we give the detailed description of FBMT. We evaluate the performance of FBMT by our simulation in Section 4. Finally, we conclude the paper.

2. Preliminary

2.1. Wireless Cognitive Network Architecture

First, we characterize a distributed system model: M primary users and N secondary users are randomly distributed in an $X \times Y$ area (**Figure 1(a)**), secondary can using a certain signal detection techniques such as matched filter detection, energy detection or cyclostationary detection to get the usable spectrum opportunity. In the system, available spectrums are divided into K completely orthogonal bands. We assume that these bands are symmetric, which means that these bands have the same bandwidth, meanwhile, have the same spectrum quality and the spectrum availability. We also assume that the interference between two nodes only rest with the relative distance of theirs, all of the primary users and

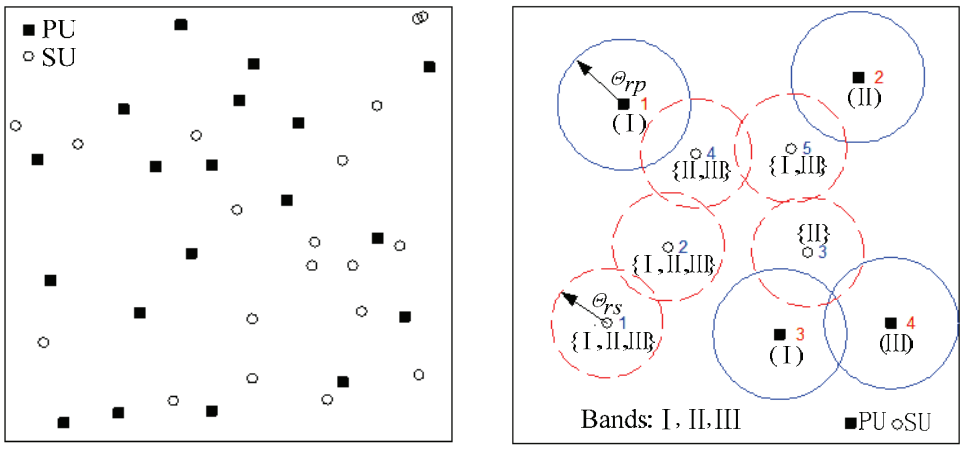


Figure 1. (a) Sketch map of the cognitive radio system; (b) An example of cognitive radio system.

secondary users utilize the omni-antenna, we define the transmission radius of primary users and secondary users are $\Theta_{rp}(i,k)$ and $\Theta_{rs}(j,k)$ respectively, where i, j and k denote the serial number of the primary user, secondary user and spectrum bands respectively. If the distance between the primary user and secondary user is greater than $\Theta_{rp}(i,k)+\Theta_{rs}(j,k)$, the secondary user will not conflict with primary user when they select the same band to propagate data simultaneously. In order to analyze simply, we assume that all of the primary users have the same transmission range (express as Θ_{rp}), and all of the secondary users have the same transmission range (express as Θ_{rs}). Moreover, we define that the node i and j are neighbours if their transmission coverage area is overlapped with each other.

Figure 1(b) depicts generic wireless cognitive network architecture. The network consists of four primary users and five secondary users; there are three orthogonal wireless bands which are denoted by Ⅰ, Ⅱ and Ⅲ. From the figure we can see that the transmission range of SU_3 conflicts with the transmission of PU_3 and PU_4, which occupy the band Ⅰ and Ⅲ respectively. Definitely, SU_3 can only utilize the spectrum Ⅱ to transmit data. As for SU_1, it can use the spectrum Ⅰ, Ⅱ and Ⅲ without interference to any PUs; As for SU_2, it also can use the spectrum Ⅰ, Ⅱ and Ⅲ, meanwhile, SU_2 and SU_1 is neighbour node mutually, in order to avoid the conflict between them, they cannot use the same spectrum to propagate data simultaneously. We need to establish some rules to assign the available spectrums to each secondary user. In this paper, we aim to maximize the system throughput by maximize the total spectrum utilization rate.

2.2. Definitions

1) In a network waiting for spectrum assignment, there are N secondary users indexed from 1 to N competing for K spectrum bands indexed 1 to K.

2) $S=\{s_{i,k}\,|\,i=1,...,N;k=1,...,K\}$ characterize the per user available spectrum, *i.e.*, spectrum band k is available for user i if $s_{i,k}=1$. Due to differences in user locations, technology employed in different spectrums and user requirements, different users will perceive different available spectrum. In **Figure 1(b)**, we can obtain

$$S=\begin{bmatrix}1&1&0&0&1\\1&1&1&1&1\\1&1&0&1&1\end{bmatrix}^T$$

3) We also consider that the throughput achieved by different bands is different, depending on the user's environment and the attribute of spectrums. Let

$$B=\{b_{i,k}\,|\,i=1,...,N;k=1,...,K\}$$

describe the reward that a user i gets by successfully acquiring available spectrum band k, *i.e.*, $b_{i,k}$ represents the maximum throughput that can be acquired (assuming no interference from other neighbours). Let

$$S_B=\{s_{i,k}\cdot b_{i,k}\}_{N\times K}$$

denote the bandwidth weighted available spectrum.

4) We characterize interference between two competing users by a constraint set. Let

$$C=\{c_{i,j,k}\,|\,c_{i,j,k}\in\{0,1\}\}_{N\times M\times K}$$

represent the constraint, where if $c_{i,j,k}=1$, users i and j can cause interference if they use the spectrum band k simultaneously.

5) We define a valid spectrum assignment

$$A=\{a_{i,k}\,|\,a_{i,k}\in\{0,1\}\}_{N\times K}$$

where $a_{i,k}=1$ denotes that spectrum band k is assigned to user i, A satisfies all the constraints defined by C, that is:

$$a_{i,k}\cdot a_{j,k}=0, if\ c_{i,j,k}=1; i,j\le N; k\le K \quad (1)$$

Finally, we use $\Lambda_{N,K}$ to denote the set of valid spectrum assignments for a given set of N users and K spectrum bands.

6) We must pay attention that different secondary users have different neighbors in different bands, which means that different secondary users have various influence to system throughput if they are assigned to identical band. We must take this heterogeneity into consideration in spectrum allocation process, we define the gain matrix

$$R=\{r_{i,k}\,|\,i=1,...,N;k=1,...,K\}$$

to describe this impact, where $r_{i,k}=b_{i,k}/\varphi_{i,k}$, $\varphi_{i,k}$ is the neighbor numbers of user i in spectrum band k.

2.3. Optimization Problems

The objective of a general resource allocation problem can be defined in terms of a utility function. In spectrum related resource allocation problems, we usually need to solve the optimization problem expressed as follows:

1) **Max-Sum-Bandwidth (MSB):** This aims to maxi-

mize the total spectrum utilization in the system. The optimization problem is expressed as:

$$\max_{A\in\Lambda_{N,K}} \sum_{i=1}^{N}\sum_{k=1}^{K} s_{i,k}\cdot b_{i,k} \quad (2)$$

2) **Max-Min-Bandwidth (MMB):** This aims to maximize the bottleneck user's spectrum utilization. The optimization problem can be expressed by:

$$\max_{A\in\Lambda_{N,K}} \min_{i<N} \sum_{k=1}^{K} s_{i,k}\cdot b_{i,k} \quad (3)$$

2.4. Color-Sensitive Graph Coloring Problem

By mapping each spectrum into a color, the spectrum allocation problem can be simplified as a GMC (Graph Multi-Coloring) problem. First, we define a bidirectional graph $G=\{V,E_C,S_B\}$, where V is a set of nodes denoting the users that share the spectrums, S_B represents the bandwidth weighted available spectrum as defined in Section 2.2, or the color list at each node, and E_C is a set of undirected edges between nodes representing interference constraints between two nodes defined by C. For any two distinct nodes $i, j\in V$, an $k-color$ edge between i and j, is in E_C if and only if $c_{i,j,k}=1$. Hence, any two distinct nodes can have multiple colored edges between them. We define the color k specific degree of a node i, *i.e.*, $\varphi_{i,k}$ to represent the number of neighbours that are color k mutually constrained with i (those who can not use k if i uses color k). It is also a relatively good measure of the impact (to neighbours) when assigning a color to a node. The equivalent graph coloring problem is to color each node using a number of colors from its color list, such that if a color k edge exists between any two distinct nodes, they can't be colored with k simultaneously. We name this the Color Sensitive Graph Coloring (CSGC) problem.

3. Proposed Spectrum Assignment Algorithm

The optimal coloring problem is known to be NP-hard. In this section, we discuss a set of heuristic based approaches that produce good coloring solutions. In particularly, we extend some of the well-known graph coloring solutions toward our problem settings and optimization goals.

3.1. The Analysis of Existing Algorithms

3.1.1. Collaborative-Max-Sum-Bandwidth Rule

This rule aims to maximize the sum of bandwidth weighted color usage, corresponding to MSB optimization defined in (2). When a node i is assigned with a color k, his contribution to the sum bandwidth in a local neighborhood can be computed as $b_{i,k}/\varphi_{i,k}$ since his neighbors can not use the color. Here $\varphi_{i,k}$ represents the number of k color constrained neighbour of a node i in the current graph. We propose to label the node according to

$$label_i = \max_{k\in s_i} b_{i,k}/\varphi_{i,k} \quad (4)$$

$$color_i = \arg\max_{k\in s_i} b_{i,k}/\varphi_{i,k} \quad (5)$$

where s_i represents the color list available at node *i* at this stage. This rule considers the tradeoff between spectrum utilization (in terms of selecting the color with the largest bandwidth) and interference to neighbours. This rule enables collaboration by taking into account the impact to neighbours. If two nodes have the same label, then the node with lower assigned bandwidth weighted colors will get a higher label.

3.1.2. Random (RAND) Rule

Each node generates a random label which is between $[0, window_i]$, where $window_i$ denote the threshold of the node *i*. At the beginning of the allocation, all of the nodes have the same threshold. In an allocation process, if node i get a color, $window_i = window_i/2$, else $window_i = window_i \times 2$. In each stage, each node labels itself according to the above labeling rules, and broadcasts the label to his neighbors. A node with the maximum label within his neighborhood gets to grab the color associated with his label and broadcasts the color assignment to his neighbors. After collecting assignment information from surrounding neighbors, each node updates his color list and recalculates the label. This process is repeated until the color list at each node is exhausted or all the nodes are satisfied.

Although RAND can optimize fairness performance in spectrum allocation process, however, it cannot guarantee the network throughput. This will be demonstrated in next section.

3.2. Fairness Bargaining with Maximum Throughput (FBMT)

In this section, we describe the Fairness Bargaining with Maximum Throughput algorithm-a distributed heuristic spectrum assignment strategy.

The algorithm is based on the following assumptions:

1) Distributed wireless network architectures.

2) All the primary users have the same transmission range, the same to secondary users.

3) All of the secondary users can detect and utilize all

of the spectrum holes.

4) Primary users have the preferential right of spectrum bands.

5) Users can communicate with each other by using some mechanism.

In FBMT algorithm, we aim to maximize the sum of bandwidth, corresponding to MSB optimization defined in (2). When a node i is assigned with a color k, its contribution to the sum bandwidth in a local neighbourhood can be computed as $r_{i,k} = b_{i,k} / \varphi_{i,k}$. We also search for the fairness in spectrum distribution, defined a fairness factor, which can be expressed by

$$f_{i,k}(t) = \frac{\sum_{j=1}^{N} 1\left\{d_{i,j} \le 2 \cdot \Theta_{rs}\right\}}{\sum_{k=1}^{K} s_{i,k}(t-1)}. \quad (6)$$

where t denotes current allocation times, numerator denotes the neighbour numbers of node i in spectrum band k, denominator means the number of bands which is assigned to node i before the *t-th* assignment, $d_{i,j}$ denotes the distance between the node i and node j. We define

$$f_{i,k}(t) = \sum_{j=1}^{N} 1\left\{d_{i,j} \le 2 \cdot \Theta_{rs}\right\} \quad (7)$$

if $\sum_{k=1}^{K} s_{i,k}(t-1) = 0$.

In FBMT, We propose to label the node according to

$$label_i = \max_{k \in s_i} r_{i,k} \cdot f_{i,k}(t) \quad (8)$$

$$color_i = \arg\max_{k \in s_i} r_{i,k} \cdot f_{i,k}(t) \quad (9)$$

where $i = 1,...,N$, $k = 1,...,K$. This rule not only considers the tradeoff between spectrum utilization and interference to neighbours, but also considers the fairness in spectrum allocation by introduction of the fairness factor. The Pseudo code of HBTM algorithm is shown in **Figure 2**.

```
A = 0
Initiating f_{i,k}(t) according to Equation (6)
While there's channel available to secondary users do
//one iteration procedure
  for i = 1 to N do
    for k = 1 to K do
      if s_{i,k}(t) == 1
        RwdWgh(i,k) = max(r_{i,k}(t) · f_{i,k}(t)),
                      1 ≤ j ≤ N, j ≠ i, c_{i,j,k=1}
        if r_{i,k}(t) > Rwd(i,k) or r_{i,k}(t) == Rwd(i,k)
          (and i got the less bandwidth) then
          a_{i,k} = 1
        end if
      end if
    end for
  end for
  Broadcasting channel assignments and available info
  to neighbors.
  All secondary users update matrix S, R, Λ
  All secondary users update f_{i,k}(t-1)
end while
```

Figure 2. Pseudo code of HBTM algorithm.

In order to analysis the fairness of each spectrum algorithms, we also define the fairness

$$\Gamma_{fairness}(A) = \frac{\left(\sum_{i=1}^{N}\sum_{k=1}^{K} a_{i,k} \cdot b_{i,k}\right)^2}{\left(N \cdot \sum_{i=1}^{N}\left(\sum_{k=1}^{K} a_{i,k} \cdot b_{i,k}\right)^2\right)}, \quad A \in \Lambda_{N \times K} \quad (11)$$

In (11), we can conclude that $\Gamma_{fairness} \in [0,1]$, the more closely the $\Gamma_{fairness}$ approach to 1, the more impartial the process of spectrum allocation has. We also define

$$\Gamma_{fairness}(A) = \left(\sum_{i=1}^{N}\sum_{k=1}^{K} a_{i,k} \cdot b_{i,k}\right)^2, \quad A \in \Lambda_{N \times K} \quad (12)$$

if $\sum_{i=1}^{N}\left(\sum_{k=1}^{K} a_{i,k} \cdot b_{i,k}\right)^2 = 0$.

4. Performance

4.1. Performance Index

We often evaluate an algorithm by the throughput and fairness it can achieve in spectrum allocation. The throughput, as defined herein, is the sum of bands each user is assigned, which can be expressed by

$$\Gamma_{Throughput}(A) = \sum_{i=1}^{N}\sum_{k=1}^{K} a_{i,k} \cdot b_{i,k}, A \in \Lambda_{N,K} \quad (10)$$

4.2. Simulations

4.2.1. Simulation Parameters

In this section we evaluate the performance of FBMT, We design and implement a prototype system which consists of some primary users and some secondary users, these users are deployed at random in our simulation. Our experiments apply following evaluation metrics:

- The throughput $\Gamma_{Throughput}$;
- The fairness $\Gamma_{fairness}$;

We compared FBMT with a recent CMSB algorithm and RAND algorithm, CMSB represents the Max-bandwidth-based spectrum allocation algorithm, and RAND algorithm is the Max-fairness-based.

In our simulations, we have considered values of N and M ranging from 10 to 35. The values of K ranging from 10 to 30, for each situation, we have assigned $b_{i,k}$ with different value (constant or variable). The parameters of simulation circumstance are shown in **Table 1** in detail.

Table 1. The simulation circumstance.

Parameter	Value
Area	1.0 × 1.0
Transmission range of Pus	0.25
Transmission range of Sus	0.15
The number of Pus	10, 15, 20, 25, 30, 35
The number of Sus	10, 15, 20, 25, 30, 35
The number of spectrums	10, 12, 14, 16, 18, 20, 25, 30
The simulation time	600 s

4.2.2. Numerical Result

Figures 3, **5** and **7** depict the throughput of three algorithms on condition of the various numbers of primary users, secondary users and spectrums, the same with **Figures 4**, **6** and **8** with the fairness. All the comparisons have different $b_{i,k}$. In **(a)**, $b_{i,k}$ is a constant, while in **(b)** the value of $b_{i,k}$ is a random distributed in [0, 1]. In **Figures 3**, **5** and **7**, we can find out if $b_{i,k}$ is identical, the throughput of FBMT is almost equal to CMSB, but larger than RAND; if $b_{i,k}$ is a random number, the throughput of FBMT has 10% gain compared with CMSB, and has a more gain compared with RAND. In **Figures 4**, **6** and **8**, we can find the FBMT is all square to RAND but better than CMSB in fairness, but the former has the lower throughput relatively.

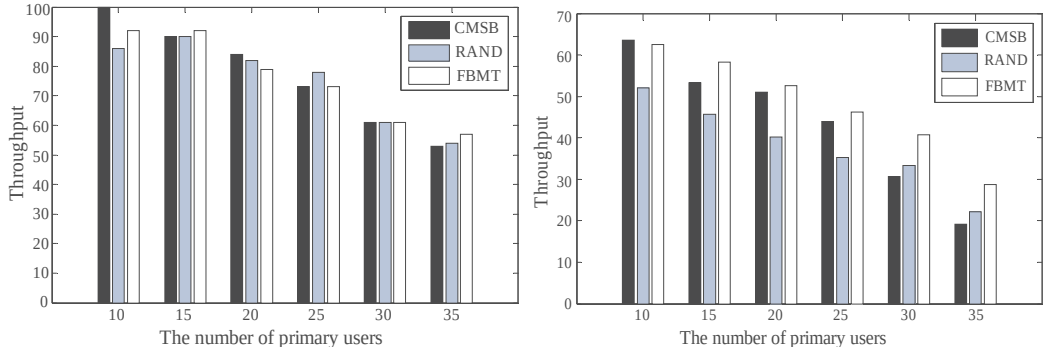
The simulation result shows that the CMSB, although has a high throughput, but with a low fairness performance; the RAND, has an agreeable fairness performance, but not well in enhancing the throughput. The FBMT, not only succeeds in throughput achieving, but also makes a good fairness performance in spectrum assignment. Relatively FBMT is the best spectrum allocation algorithm in these algorithms.

Figure 3 shows the system throughput of different primary users with CMSB, RAND and FBMT algorithms respectively on the premise of 30 SUs competing 15 channels. We can find that if $b_{i,k}$ is identical such like $b_{i,k}$ = 1, the three algorithms almost have the same throughput, while if random, the throughput of FBMT is maximum, CMSB less if primary users are less than 30, and RAND least. **Figure 4** shows the fairness of PUs with three algorithms. We can find that CMSB has much lower performance compared with FBMT and RAND. Moreover, if $b_{i,k}$ = 1, FBMT has the best fairness performance.

Figures 5 and **6** depict the throughput and fairness of secondary users. When there are 20 primary users and 15 channels for secondary users, when $b_{i,k}$ is a random number, the FBMT present the most preferable performance, CMSB less and RAND least. However when $b_{i,k}$ is set 1, CMSB has better throughput gain. But whether $b_{i,k}$ is confirmed or not, CMSB has the lowest fairness performance.

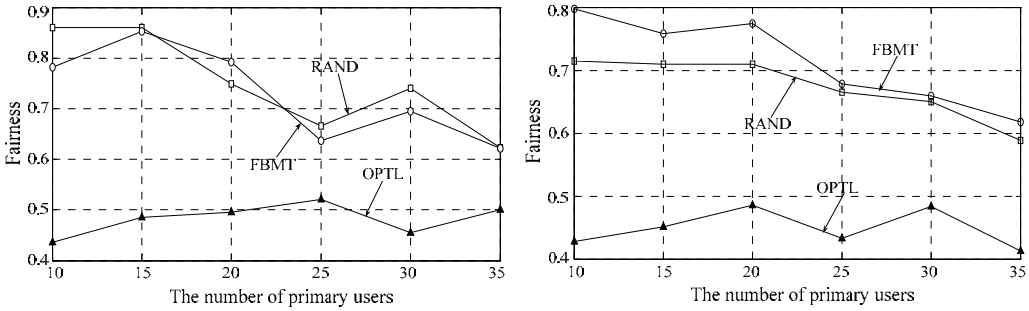
Figures 7 and **8** show the results in which the channel number is fixed. If $b_{i,k}$ is a constant, we can find that RAND has the lowest throughput and FBMT the highest. However if $b_{i,k}$ is not fixed, the throughput of three algorithms is almost the same. Also we can find that CMSB still has the lowest fairness performance.

The simulation result reflects that although CMSB has a high throughput, the fairness performance is not satisfying. RAND has an agreeable fairness characteristic but not much well in throughput. However, FBMT can achieve both performances of throughput and fairness.



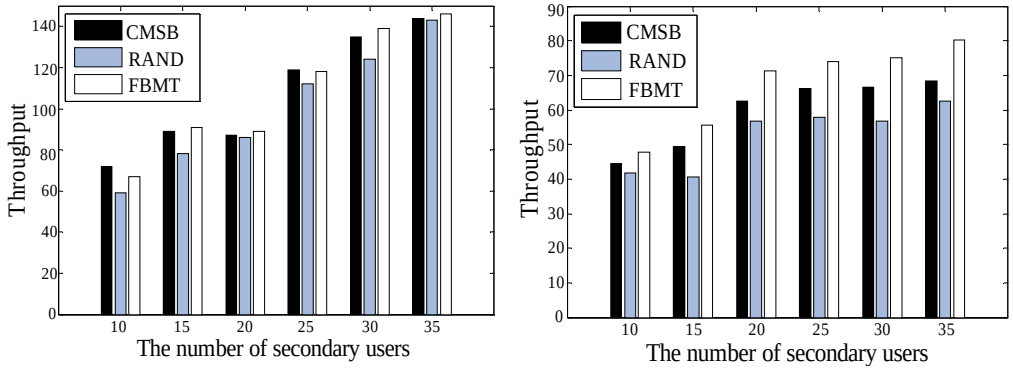
(a) N = 30, K = 15, $b_{i,k}$ = 1 (b) N = 30, K = 15, $b_{i,k} \in [0, 1]$

Figure 3. The number of primary users vs. system throughput.



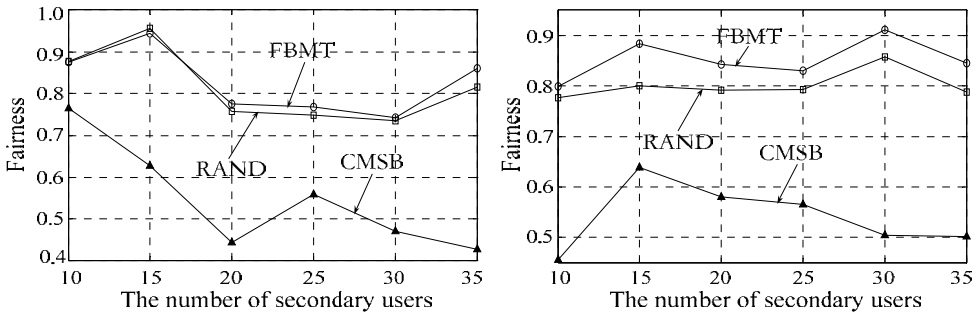
(a) $N = 30$, $K = 15$, $b_{i,k} = 1$ (b) $N = 30$, $K = 15$, $b_{i,k} \in [0, 1]$

Figure 4. The number of primary users vs. fairness.



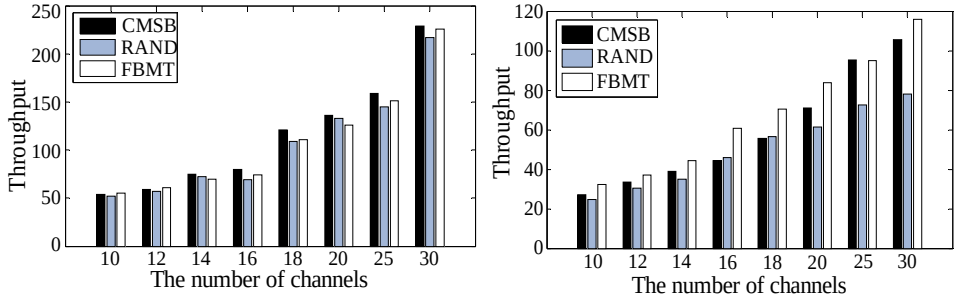
(a) $M = 20$, $K = 15$, $b_{i,k} = 1$ (b) $M = 20$, $K = 15$, $b_{i,k} \in [0, 1]$

Figure 5. The number of secondary users vs. system throughput.



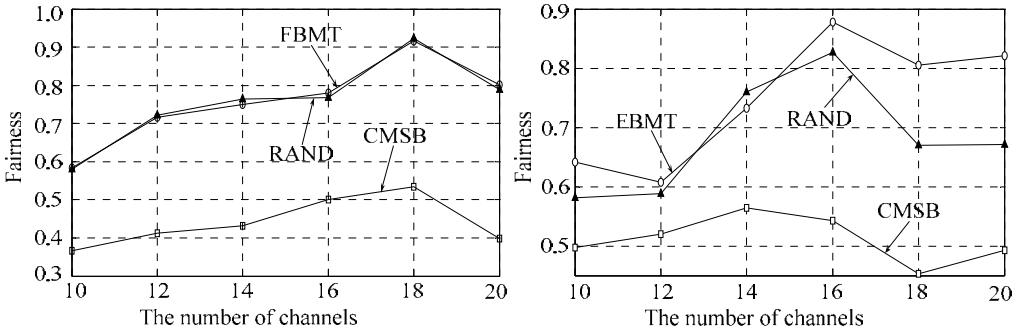
(a) $M = 20$, $K = 15$, $b_{i,k} = 1$ (b) $M = 20$, $K = 15$, $b_{i,k} \in [0, 1]$

Figure 6. The number of secondary users vs. fairness.



(a) $M = 15$, $K = 20$, $b_{i,k} = 1$ (b) $M = 15$, $K = 20$, $b_{i,k} \in [0, 1]$

Figure 7. The number of spectrums vs. throughput.



(a) $M = 15$, $K = 20$, $b_{i,k} = 1$ (b) $M = 15$, $K = 20$, $b_{i,k} \in [0, 1]$

Figure 8. The number of spectrums vs. fairness.

5. Conclusions

In this paper, we explore the tradeoff in spectrum utilization and interference mitigation in open spectrum system. We develop a new graph-theoretical model to characterize the spectrum access problem under a number of different optimization functions, taking into account heterogeneity in both spectrum availability, reward and interference constraint.

We then propose an algorithm where each user can opportunistically utilize its available spectrum. Experimental results confirm that our algorithm significant benefits most in system throughput and also have a good performance in fairness.

6. Acknowledgements

This work was supported in part by National 863 Projects of China (2009AA01Z205), Fund of National Laboratory (P080010) and Program for New Century Excellent Talents in University (NCET070339).

7. References

[1] S. W. Ellingson, "Spectrum Occupancy at VHF: Implications for Frequency-Agile Cognitive Radios," *Proceedings of IEEE Vehicle Technologies Conference*, Vol. 9, No. 2, 2005, pp. 1379-1382.

[2] M. A. McHenry, "NSF Spectrum Occupancy Measurements Project Summary," Shared Spectrum Company Report, Vienna, 2005.

[3] M. McHenry, E. Livsics, T. Nguyen and N. Majumdar, "XG Dynamic Spectrum Access Field Test Results," *IEEE Communications Magazine*, Vol. 6, No. 45, 2007, pp. 51-57.

[4] J. Mitola and G. Q. Maguire, "Cognitive Radio: Making Software Radios More Personal," *IEEE Personal Communications*, Vol. 8, No. 6, 1999, pp. 13-18.

[5] H. Zheng and C. Peng, "Collaboration and Fairness in Opportunistic Spectrum Access," *Proceedings of the* 2005 *IEEE International Conference on Communications* (*ICC*'05), Seoul, Korea, 2005, pp. 3132-3136.

[6] C. Peng, H. Zheng and B. Y. Zhao, "Utilization and Fairness in Spectrum Assignment for Opportunistic Spectrum Access," *Mobile Networks and Applications*, Vol. 11, No. 4, 2006, pp. 555-576.

[7] J. Zhao, H. Zheng and G. Yang, "Distributed Coordination in Dynamic Spectrum Allocation Networks," *Proceedings of the* 2005 1*st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks* (*DySPAN*'05), Baltimore, 2005, pp. 259-268.

[8] L. Cao and H. Zheng, "Distributed Spectrum Allocation Via Local Bargaining," *Proceedings of the* 2*nd Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, Santa Clara, 2005, pp. 475-486.

[9] X. Liu and W. Wang, "On the Characteristics of Spectrum-Agile Communication Networks," *Proceedings of the* 2005 1*st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks* (*DySPAN*'05), Baltimore, 2005, pp. 214-223.

[10] W. Wang and X. Liu, "List-Coloring Based Spectrum Allocation for Open-Spectrum Wireless Networks," *Proceedings of the IEEE International Symposium on Vehicular Technology* (*VTC*'05-*Fall*), Dallas, 2005, pp. 690-694.

[11] Y. Xing, R. Chandramouli, S. Mangold and S. N. Shankar, "Analysis and Performance Evaluation of a Fair Spectrum Access Protocol for Open Spectrum Wireless Networks," *Proceedings of the* 2005 *IEEE International Symposium on Communications* (*ICC*'05), Seoul, Korea, 2005, pp. 1179-1183.

Wireless Sensor Network, **2010, 2, 419-440**
doi:10.4236/wsn.2010.26054 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Classification and Review of Security Schemes in Mobile Computing

Sathish Alampalayam Kumar
Department of Computer Science and Information Technology PSG Institute of Advanced Studies, Coimbatore, India
E-mail: *sathish.ap@gmail.com*
Received October 13, 2009; *revised November* 20, 2009; *accepted December* 25, 2009

Abstract

In this paper, we present the classification and review of security schemes in mobile computing system. We classify these schemes based on types the infrastructure used in the mobile computing system-Mobile Ad Hoc Networks (MANET) and Mobile Agent model. Mobile Ad Hoc Networks are pervasive, ubiquitous and without any centralized authority. These unique characteristics, combined with ever-increasing security threats, demand solutions in securing ad hoc networks prior to their deployment in commercial and military applications. This paper reviews the prevailing mobile ad hoc network security threats, the existing solution schemes, their limitations and open research issues. We also explain the Intrusion detection and response technique as an alternate method to protect the MANET based mobile computing systems and their approaches. A literature review of important existing Intrusion Detection approaches and Intrusion Response Approaches for MANET is also presented. This paper also presents the limitations of existing Intrusion Detection and Response Approaches for MANET and open research issues in providing MANET security. With respect to Mobile Agent based mobile computing system, we have presented the classification of various types of security attacks in Mobile Agent based model and presented the security solutions for those type of attacks proposed by the various schemes and the open research issues in providing security for Mobile Agent based mobile computing system. Such classification enhances the understanding of the proposed security schemes in the mobile computing system, assists in the development and enhancement of schemes in the future and helps in choosing an appropriate scheme while implementing a mobile computing system.

Keywords: Wireless Sensor Networks, Beta Trust Model, Trust Routing Protocol, Network Security, Trust Evaluation

1. Introduction

Although the wonderful invention of Internet offers access to information sources worldwide, we do not expect to benefit from that access until we arrive at some familiar point-whether home, office, or school. However, the increasing variety of wireless devices offering IP connectivity, such as PDA's, handhelds, and digital cellular phones, is beginning to change our perceptions of the Internet.

Mobile computing and networking should not be confused with the portable computing and networking we have today. In mobile networking, computing activities are not disrupted when the user changes the computer's point of attachment to the Internet. Instead, all the needed reconnections occur automatically and none interactively. Mobile Internet implies changing the point of attachment as the host (mobile station) roams between cells.

Truly, mobile computing offers many advantages. Confident access to the Internet anytime, anywhere will help free us from the ties that bind us to our desktops. Having the Internet available to us as we move will give us the tools to build new computing environments wherever we go. This is especially convenient in a wireless LAN office environment, where the boundaries between attachment points are not sharp and are often invisible.

However, there are still some technical obstacles that must be overcome before mobile networking can become widespread. The most fundamental is the security management, which is almost an afterthought until the recent years. Providing security services in the mobile computing environment is challenging because it is

more vulnerable for intrusion and eavesdropping. Authentication mechanisms are designed to protect a system from unauthorized access to its resources and data. However, at present, completely preventing breaches of security seems unrealistic, especially in mobile computing systems [1,2]. A Personal Area Network (PAN) level firewall as envisioned for the next generation wireless networks can protect only if the users are at home and not when the users are roaming [3]. Even if such a firewall is provided, the communication would get fragmented by these 'check points' on the network, as each firewall needs maintenance of activities like log control, software update etc., creating unnecessary overhead. Thus existing technologies like firewalls and Virtual Private Network (VPN) sandboxes cannot be directly applied to the wireless mobile world. Even if the firewall concept were achieved by creating a private extranet (VPN) which extends the firewall protected domain to wherever the user moves, this would still lead to inefficient routing. Security is a fundamental concern for mobile network based system. Harrison *et al.* [4] identify security as a "severe concern" and regard it as the primary obstacle to adopting mobile systems.

2. Mobile Computing Systems Security

2.1. Mobile Computing Systems Security Classification

The security approaches for mobile computing systems can be classified as shown in the following **Figure 1**.

2.2. MANET and Security Attacks in MANET

2.2.1. MANET Background

A Mobile Ad hoc Network (MANET) is a collection of wireless mobile nodes forming a temporary network without any centralized authority. In a MANET, each wireless mobile node operates not only as an end-system, but also as a router to forward packets. The nodes are free to move about and organize themselves into a network. MANET does not require any fixed infrastructure such as base stations; therefore, it is an attractive networking option for connecting mobile devices quickly and spontaneously. For instance, first responders at a disaster site or soldiers in a battlefield must provide their own communications. A MANET is a possible solution for this need to quickly establish communications in a mobile, transient and infrastructure-less environment. This is one of many applications where MANET's can be used. Mobile ad-hoc networks are the future of wireless networks. Nodes in these networks will generate both user and application traffic and perform various network functions.

In the last decade, wired and wireless computer network revolution has changed the computing scenario. The possibilities and opportunities due to this revolution are limitless; unfortunately, so too are the risks and chances of attacks due to intrusion by malicious nodes [4]. Intrusion is defined as an attack or a deliberate unauthorized attempt

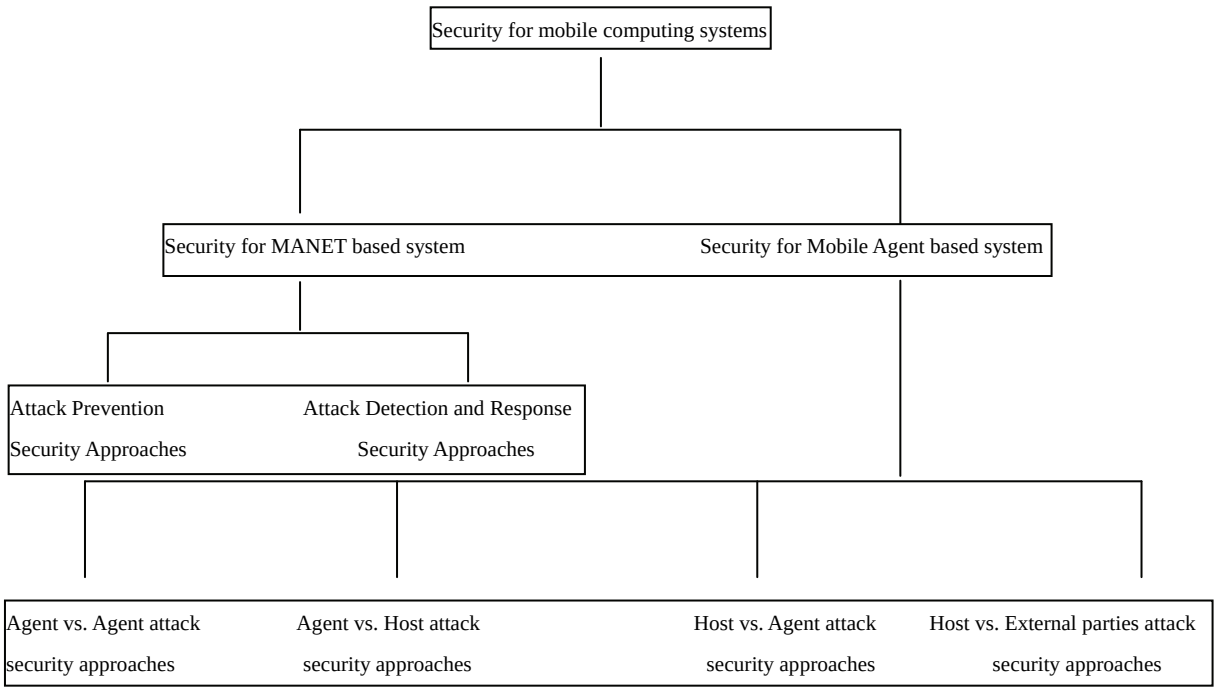


Figure 1. Taxonomy of security for mobile computing systems.

to access information, manipulate information, or render a system unreliable or unusable [5]. According to [6], threat can be defined as "the potential possibility of a deliberate unauthorized attempt to 1) access information, 2) manipulate information and 3) render a system unreliable or unusable. By security we mean protecting nodes from damages due to either voluntary or accidental attacks [7]. This protection is provided by predicting an attack by monitoring a set of metrics measured from the ad hoc network, and then responding and modifying the security of the network based on the vulnerability level at a given time.

Security in mobile ad hoc network is essential even for basic network functions like routing which are carried out by the nodes themselves rather than specialized routers. The intruder in the ad hoc network can come from anywhere, along any direction, and target any communication channel in the network. Compare this with a wired network where the intruder gains physical access to the wired link or can pass through security holes at firewalls and routers. Since the infrastructure-free mobile ad hoc network does not have a clear line of defense, every node must be prepared for the adversary. The centralized or hierarchical network security solution for the existing wired and infrastructure-based cellular wireless networks will not work properly for Mobile Ad Hoc Networks [8]. Securing the ad hoc networks, like any other field of computers, is based on the principle of confidentiality and integrity. These principles exist in every field, but the presence of malicious nodes, selfish nodes, covert channels and eavesdroppers in the mobile ad hoc network makes this an extremely important and challenging problem [9]. In the past several years, there has been a surge of network security research in the field of information assurance that has focused on protecting the network using techniques such as authentication and encryption. These techniques are applicable in the wired and infrastructure-based cellular network. In the case of infrastructure-free Mobile Ad Hoc Networks these techniques are not applicable [8]. In the infrastructure-free networks, the nodes themselves perform basic network functions like routing and packet forwarding. Therefore, mobile ad hoc network security is a pressing issue, which needs immediate research attention [10-13]. Providing security services in the mobile computing environment is challenging because it is more vulnerable for intrusion and eavesdropping. The challenge of mobile ad hoc network security has attracted several researchers with the aim of securing mobile ad hoc computer networks.

2.2.2. Security Attacks in MANET

A MANET can be subjected to active attacks and passive attacks. Active attacks refer to the direct attacks by a hostile entity during execution or transmission phase. Some of the major types of active attacks are routing attacks and active DoS attacks. Passive attacks refer to the indirect attacks by an entity in the network during collaboration. Some of the major types of passive attacks include actions like selfishness, eavesdropping, traffic analysis and passive DoS attacks.

1) Active Attack in MANET:

a) Routing Attacks:

Routing attack is a significant problem because nodes within the ad hoc network themselves performs routing functions and the security concepts are not incorporated in most of the routing protocols. Also, routing tables form the basis of network operations and any corruption to the routing table may lead to significant adverse consequences.

Designing a secure ad hoc network routing protocol is a challenge for the following reasons: Firstly, routing relies on the trustworthiness of all the nodes involved and it is difficult to distinguish selfish nodes from normal nodes. Secondly, rapid mobility of nodes that perform the role of routing and network topology makes the design of a secure routing protocol more difficult. Active routing attacks differ in their behavior depending on the nature of the routing protocol. In the case of link-state routing protocol, a router sends information about its neighbors. Hence a malicious router can send incorrect updates about its neighbors, or remain silent if the link state of the neighbor has actually changed. However, in the case of distance-vector protocols, routers can send wrong and potentially dangerous updates regarding any nodes in the network, since the nodes do not have the full network topology. These attacks in case of both link-state and distance-vector protocols are very difficult to prevent if the routers exhibit Byzantine faults [14].

In the MANET shown in **Figure 2**, let us assume that packets are supposed to traverse from source node A to destination node C. However, the intruder updates the routing table so that the packets traverse from B to D instead of C, and hence the packets from A never reach C. This also causes congestion on domains served by nodes A, D and E, due to the bombardment of packets whose actual destination was C. Thus the attack can lead to network performance degradation.

Some of the important and common methods of routing attacks are:

i) Router Protocol Poisoning: In this attack an intruder causes the disruption by poisoning the routing protocol. Securing these attacks is important because the routing protocol forms the basis of network operations, and any corruption of the protocol may lead to significant consequences. These attacks on the Mobile Ad Hoc Networks can lead to looping, congestion, sub optimal routing and partitioning [15]. Thus, they can ultimately affect the performance of an ad hoc network.

ii) Injecting incorrect information in the routing table: In this type of routing attack, malicious nodes or an intruder would inject incorrect routing information, which

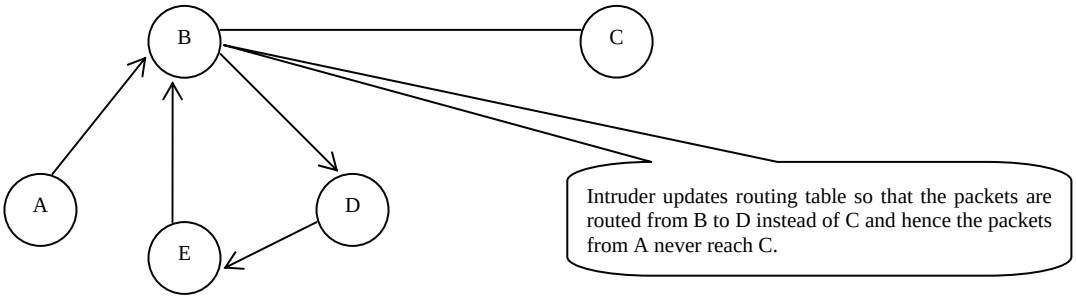


Figure 2. Routing loop attack.

in turn would poison the routing tables. These attacks would result in the artificial partitioning of the network, and the hosts residing in one partition would not be able to communicate with hosts residing in the other partition.

iii) Routing Loop Attacks: In this attack, intruder or malicious nodes update the routing table to create a loop, so that packets can traverse in the network without reaching the destination, thereby conserving energy and bandwidth.

b) Active DoS Attacks:

These attacks can be defined as the direct denial of service attacks on a node by another hostile node through packet flooding, packet modification, deletion or forging of packets or routing table. Following are some of the common types of active DoS attacks by selfish nodes or adversaries: replay of expired routing information, bogus nodes create traffic by bombarding the neighboring nodes with the packets, radio jamming, flooding centralized resource with the requests, ability to change routing protocol to operate as the user wants, Byzantine failure, sleep deprivation torture (Battery Exhaustion) and injecting incorrect routing information.

Active DoS attack is depicted in **Figure 3**, where node B is a host node and C is the intruder. The intruder node C creates a huge traffic resulting in the exhaustion of the node B's resources. This results in the inability of node B to serve genuine nodes A, D, E and F fairly. Thus, DoS attacks on the mobile ad hoc networks can lead to network performance degradation.

2) Passive Attack in MANET

a) Selfish Attacks:

Passive attacks could be caused by selfishness, eavesdropping and traffic analysis. In this section we explain selfishness attacks to give an idea of passive attacks. In the selfishness attacks, the selfish node abuses constrained resources, such as battery power, for its own benefit [16]. They do not intend to directly damage other nodes in the network. Attackers may also get hold of a node and modify its behavior to make it malicious, so the node would perform selfish attacks in need of resources. These attacks have limited effectiveness compared to the routing-table "poisoning" and DoS attacks [17]. This is because, the attacks are limited to a part of the network rather than the whole network as in the case of routing protocol attacks.

Some of the common types of selfish node attacks in mobile ad hoc network are packet mistreatment and energy consumption attacks. In this kind of attack, a node in mobile ad hoc network does not perform the expected network functions, like packet forwarding or routing, and later claims that the transaction or communication never took place [17]. It could be deliberate or accidental, due to false repudiation of a transaction or due to scarce resources in the mobile ad hoc networks.

As shown in **Figure 4**, the packets are supposed to traverse from source node A to destination node C. However, selfish node B discards the packets from A and hence the packets from A never reach C. This results in 'black hole' attacks. This in turn may result in deadlock issues which result in performance degradation. Some of the important and common methods of selfish attacks are:

i) Packet mistreatment or interception: In this kind of attack, a selfish node does not perform the function of packet forwarding. As mentioned earlier, interruption

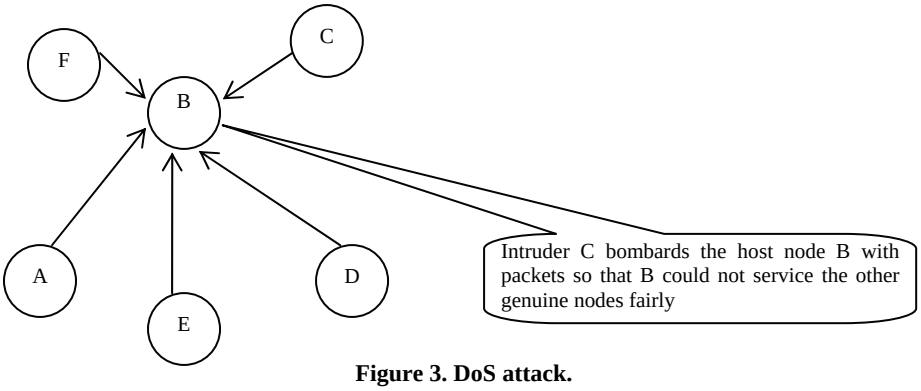


Figure 3. DoS attack.

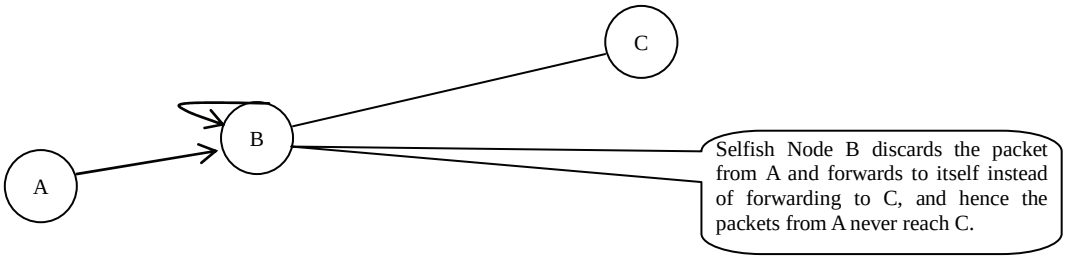


Figure 4. Packet mistreatment attack.

of packets may reduce the overall throughput of the network. In a specialized form of packet discarding, selfish nodes do not forward the packets to host destination, but to itself. This result in black hole and DoS attacks.

ii) Energy consumption: In this kind of attacks, nodes try to save significant battery power by not performing networking functions such as routing. This is due to the fact that in ad hoc network most of the energy is consumed by routing of packets. For instance, experiments have shown that if the average hop from source to destination is 5, approximately 80% of the available energy is spent in sending packets from source to destination by packet forwarding [17].

2.3. Mobile Agent Model and Security Threats in Mobile Agent Model

2.3.1. Mobile Agent Model Background

A distributed mobile agent system model for a wireless internet host environment involves the following parties, mobile agents and fixed base stations as shown in **Figure 5**. Some of the wireless models [18] applied for special applications like mobile military networks assumes mobile base stations. However, in this discussion we assume the base station is fixed.

Mobile Agent:

The Mobile Agent (MA) is a software component [19] like

- A thread as in Telescript, that can migrate among different nodes carrying its execution state (*i.e.*, program counter, call stack etc.) Here the run-time image of the component is transferred as a whole, including its execution state.

The task to rebuild the execution state is carried out by the run-time support of the Mobile Code System.

- Or just a code fragment as in TACOMA [20] associated with initialization data that can be shipped to a remote host. They don't have the ability to migrate once they have started their execution. These systems claim to be able to move the state of a component along with its code. This assertion is justified by the availability of mechanisms that allow the programmer to pack some portion of the data space of an executing component before the component's code is sent to a remote destination.

It is the programmer's task to rebuild the execution state of a component after its migration, using the data transferred with the code.

Thus a mobile agent (with respect to design paradigm) contains.

- Code component-Executing Unit (EU) (Sequential flows of computation), which encapsulate the know-how to perform a particular computation.
- Resource component-(entities that can be shared among multiple EUs such as a file in a file system, an object shared by threads in a multi-threaded object-oriented language, or an operating system variable) that represents data or devices used during the computation.

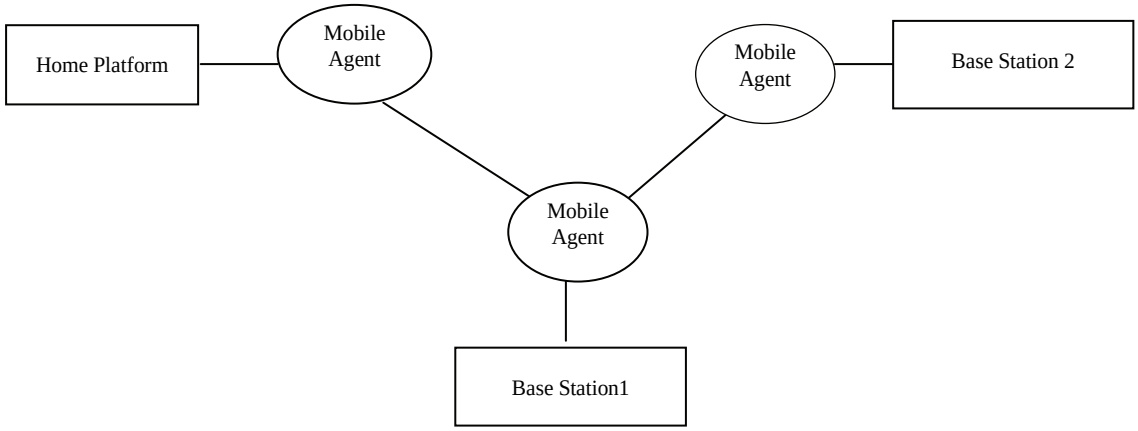


Figure 5. Mobile agent model in mobile computing.

- Computational components that are active executors capable to carry out a computation.

Mobility allows an agent to move or hop among base station. The base station provides a computational environment in which an agent operates. The purpose of Mobile Agent in terms of Artificial Intelligence (AI) research paradigm is a software component that is able to achieve a goal by performing actions and reacting to events in a dynamic environment. The behavior of this component is determined by the knowledge of the relationships among events, actions and goals. However, in terms of Distributed Systems research paradigm, the purpose of the mobile agent is to allow the migration of the whole computational component to a remote site, along the code it needs, some resources required to perform the task along with its execution state of an EU to a different CE (Computation Environment or Host).

Mobile Agents are increasingly becoming popular with the ubiquitous and widespread deployment of wireless and internet technologies. With the help of mobile agents it is possible to create distributed applications where the programs can autonomously traverse from one computer to another and get executed. They are more powerful than an ordinary applets [21] due to the AI component, they decide themselves where and when to traverse and execute. They are prominently applied in mobile computing systems. Connection management for mobile computing requires continuous reconfiguration of the data links. If connectivity fails, the mobile computing system requires applications to handle extended off-line periods. "Mobile software agents are very useful in this context, since they could encapsulate long-lasting transactions. They could carry a request to server, cause its execution and bring back the result as soon as the connectivity is reestablished [21]." Due its ability to preprocess the results, it makes use of the slow communication link between the mobile device and the network.

2.3.2. Security Threats in Mobile Agent Based Model: In the mobile agent-host model the security attacks or threats could be classified into four categories:

- mobile agent attacked by another mobile agent
- mobile agent attacking by the host
- host attacked by a mobile agent
- host attacked by external unauthorized party like an agent or host

For the ease of understanding, any agent or host attack could be further classified into active or passive attacks. Before further classification, it is essential to define active and passive attacks.

Active attacks can be defined as the direct attacks on an entity by another hostile entity during its execution or transmission like code/message modification, deletion or forging.

Passive attacks can be defined as the indirect attacks on an entity by another hostile entity during its execution or transmission like eavesdropping and traffic analysis.

Mobile Agent Attacked by another Agent:

Different types of attacks by a MA against another MA can be classified as shown in the following taxonomy.

1) Active Attacks:

Denial of service: In these attacks agent could spam other agents causing resource constraints by repeatedly sending messages to another agent, may place undue burden on the message handling routines of the recipient. Agents can also intentionally distribute false or useless information to prevent other agents from completing their tasks correctly or in a timely manner.

Unauthorized Access: In these attacks agent would invoke other agent's public methods by accessing or modifying agent's code or data, which could change the behavior of agent from trusted to harmful one.

2) Passive Attacks:

Repudiation: Agent participating in a transaction or communication later claims that the transaction or communication never took place—could be deliberate or accidental, due to false repudiation of a transaction or due to imperfect business transactions within an organization.

Masquerade: In this category an agent posing as host could deceive other agents and it harms both the agent that is being deceived and the agent whose identity has been assumed, especially in agent societies where reputation is valued and used as a means to establish trust.

Mobile agent attacked by the host:

Different types of attacks by a host against MA can be classified as shown in the following taxonomy.

3) Active Attacks

Denial of Service: In these attacks host would ignore agent service request by not executing the agent or turning away the request. This would introduce unacceptable delays for critical tasks like handoff in the mobile computing world. Agents on other platforms waiting for the results from a non-responsive agent in the malicious host platform could cause deadlock or livelock problems.

Alteration: Since agent visits various base stations or hosts during its life time, it could be altered by any of the hosts an agent passes through its lifetime. Thus a mobile agent is exposed to a new risk each time it is in transit and each time it is instantiated on a new platform.

Copy and Replay: In these attacks an agent or its message could be copied and replayed several times by the host.

4) Passive Attacks

Masquerade: In these attacks host deceives a mobile agent as to its true destination and corresponding security domain. Thus it harms both the agent and the host or platform it assumes. This is a more serious problem than an agent masquerading as other agent.

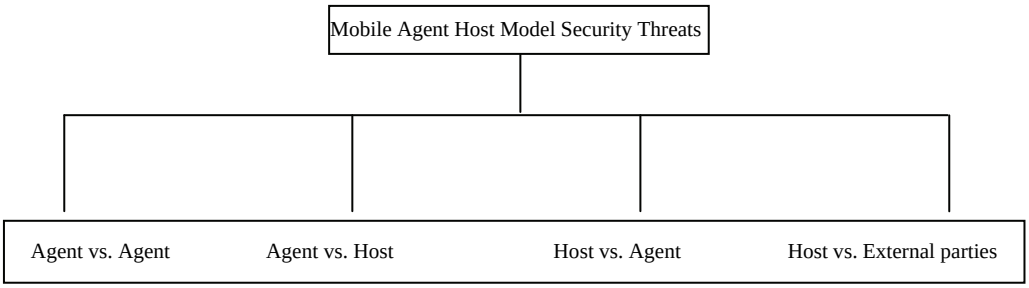


Figure 6. Taxonomy of mobile agent model security threats.

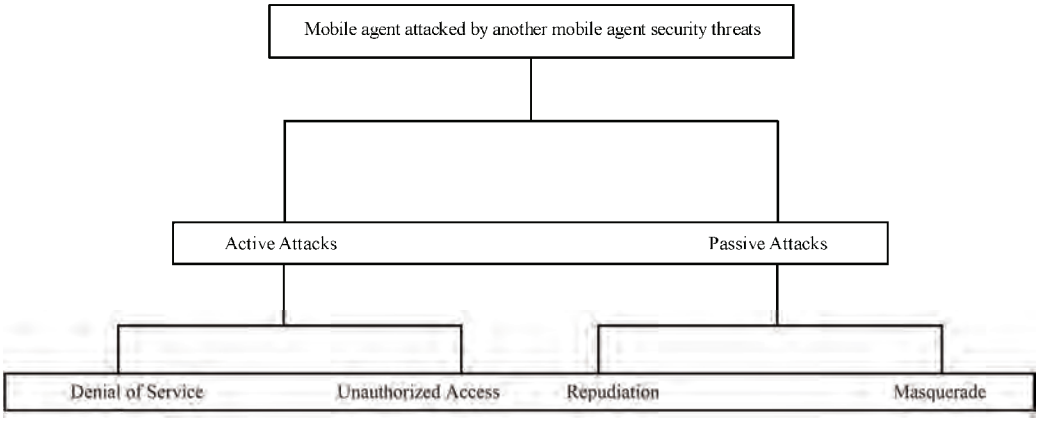


Figure 7. Taxonomy of mobile agent attacked by another agent attacks.

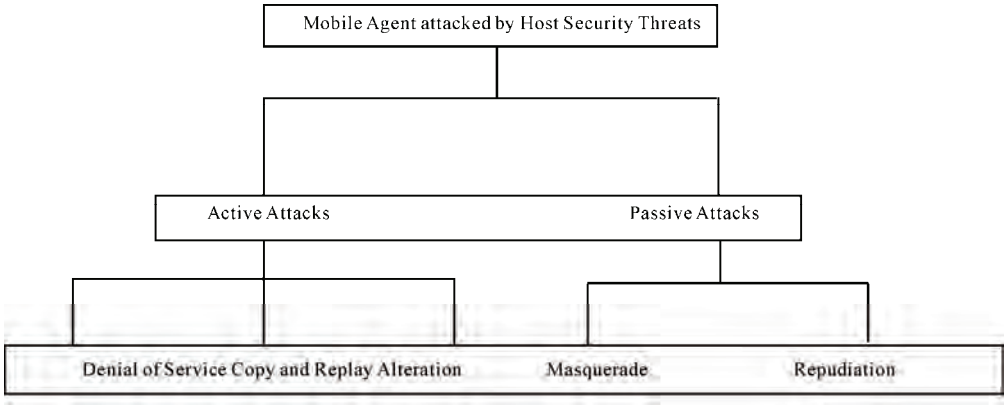


Figure 8. Taxonomy of mobile agent attacked by host security threats.

Repudiation: Host participating in a transaction or communication with an agent later claims that the transaction or communication never took place-could be deliberate or accidental, due to false repudiation of a transaction or due to imperfect business transactions within an organization.

Host attacked by mobile agents

Different types of attacks by a MA against host can be classified as shown in the following taxonomy.

5) Active Attacks:

Denial of Service: In these attacks agent consume excess amount of host resources so that the host can not service other agents properly.

Unauthorized access: In these attacks, agent without proper authorization could harm the host.

6) Passive Attacks

Masquerading: In these attacks agent may pose as an authorized agent to gain access to services and resources to which it is not entitled, to shift the blame for any actions for which it does not want to be held accountable and to damage the trust the legitimate agent has established in an agent community and its associated reputation.

Host attacked by other unauthorized external parties including host and agents:

Different types of attacks by an external party like an

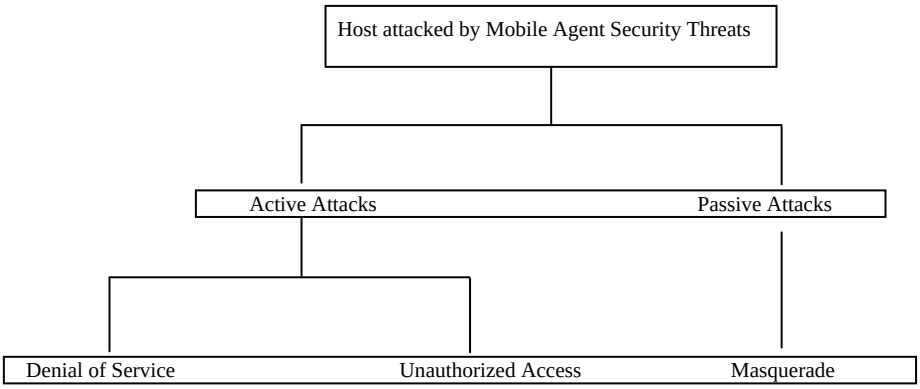


Figure 9. Taxonomy of host attacked by mobile agent security threats.

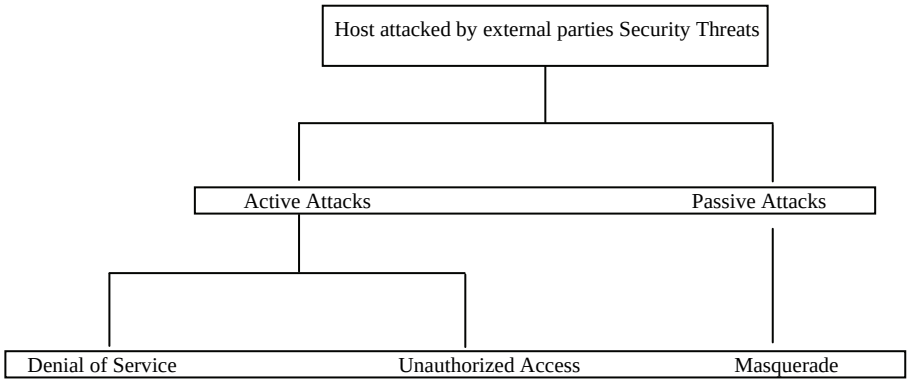


Figure 10. Taxonomy of host attacked by external parties security threats.

external MA or an external host against the host can be classified as shown in the following taxonomy.

7) Active Attacks

Unauthorized Access: In these attacks, remote users, processes, and agents may request resources from the host, for which they are not authorized.

Denial of service: In these attacks, the agent services offered by the host or base station can be disrupted by common denial of service attacks.

8) Passive Attacks

Masquerade: An agent on a remote base station can masquerade as another agent and request services and resources for which it is not authorized. They may act in conjunction with its platform (base station) to deceive the host.

3. MANET Security Approaches

3.1. MANET Attack Prevention Approaches

In this section, we classify the MANET security work into two broad categories based on the type of attack: active attack or passive attack.

3.1.1. Review of MANET Attack Prevention Security Schemes for Active attacks

In ad hoc networks, a mobile node or host may depend on other node(s) to route or forward a packet to its destination. The security of these nodes could be compromised by an external attacker or due to the selfish nature of other nodes. This would create a severe threat of Denial of Service (DoS) and routing attacks where malicious nodes combine and deny the services to legitimate nodes. Unlike nodes in a wired network, the nodes of MANET may have less processing power as well as battery life and consequently would try to conserve resources. In this scenario, the usual authentication and encryption methods would not apply to a MANET the same way they would in a wired network [22]. However, both authentication and encryption are even more important in a MANET [23,24]. Steiner *et al.* have developed a Group key Diffie-Hellman (GDH) model that provides a flexible solution to group key management. Yi *et al.* [25] have developed the MOCA (MObile Certification Authority) protocol that helps manage heterogeneous mobile nodes as part of a MANET. MOCA uses Public Key Infrastructure (PKI) technology.

The impact of authentication attacks is quite widespread and it includes unauthorized access, denial of service, masquerading, information leakage, and domain hijacking. Capkun *et al.* [26] have developed some solutions using a concept that they introduce, called Maximum Degree Algorithm (MDA), for preventing denial of

service due to poor key management.

Routing is an important aspect of moving packets around in a network. It is a challenging problem because nodes within the ad hoc network themselves performs routing function and the security concepts were not incorporated into the routing protocols when they were designed. It is important because the routing table forms the basis of the network operations and any corruption of routing table may lead to significant consequences. Routing attacks in mobile ad hoc network are more challenging since routing relies on the trustworthiness of all the nodes involved and it is difficult to distinguish selfish nodes from normal nodes. Basically there are two methods used for routing: AODV (Ad hoc On-demand Distance Vector) routing and DSDV (Destination Sequenced Distance Vector) routing. These two methods can be classified as reactive and proactive respectively since AODV method discovers a route only when needed whereas the DSDV method maintains a dynamic routing table at all times.

A reactive routing method was proposed by Yang *et al.* [27]. In this method, a unified network layer prevention method known as Self Organized Security (SOS) scheme that uses AODV routing is used. This scheme takes a self-organized approach by exploiting full localized design, without assuming any apriori trust or secret association between nodes. In this model, each node has a token in order to participate in the network operations, and its local neighbors collaboratively monitor it to detect any misbehavior in routing or packet forwarding services. Upon expiration of the token, each node renews its token via its multiple neighbors. The period of the validity of a node's token is dependent on how long it has stayed and behaved well in the network. A well-behaving node accumulates its credit and renews its token less frequently as time evolves. In essence, this security solution exploits collaboration among local nodes to protect the network layer without completely trusting any individual node.

Another reactive scheme, called Techniques for Intrusion-Resistant Ad Hoc Routing Algorithms (TIARA) was proposed by Ramanujam *et al.* to detect and eliminate DoS [28]. This model presents a new approach for building intrusion resistant ad hoc networks in the wake of DoS attacks using wireless router extensions. This approach relies on extending the capabilities of existing ad hoc routing algorithms to handle intruders without modifying the existing routing algorithms. This scheme proposes a new network layer mechanism for detecting and recovering from intruder induced malicious faults that work in concert with existing ad hoc routing algorithms and augment their capabilities.

Hu *et al.* [29] have developed a DSDV-based secure routing method called SEAD (Secure Efficient Ad hoc Distance vector). This method uses efficient one-way hash functions and does not use symmetric cryptographic operations in the protocol in order to support the nodes of limited CPU processing capability and to guard against Denial-of-Service (DoS) attacks. The primary reason for this is due to the fact that the nodes in an ad hoc network are unable to verify asymmetric signatures quickly enough for routing protocols to decide on the routing path.

Routing attacks differ in their execution depending on the nature of the routing protocol. In the case of link state routing protocol such as AODV, a router sends information about its neighbors. Hence, a malicious router can send incorrect updates about its neighbors or remain silent if the link state of the neighbor has actually changed. However, in case of distance vector protocols such as DSDV, routers can send wrong and potentially dangerous updates regarding any nodes in the network since the nodes do not have the full network topology. Awerbuch *et al.* [30] studies the behavior of routers in the presence of Byzantine faults. They use an On-demand Secure Routing Protocol (OSRP) that defines a reliability metric based on past records and use it to select the secure path. Reliability metric is represented by a list of link weights where high weights correspond to low reliability. Each node in the network maintains its own list, referred to as a weight list, and dynamically updates that list when it detects faults. Faulty links are identified using a secure adaptive probing technique that is embedded in the normal packet stream. These links are avoided using a secure route discovery protocol that incorporates the reliability metric. This protocol achieves these functionality by three successive phases: Route discovery with fault avoidance phase whose input is source node's weight list and output is the full least weight path from the source node to the destination node, Byzantine fault detection phase whose input is the full weight path and output is a faulty link and link weight management phase which takes a faulty link as an input and whose output is the weight list which in turn is used by the route discovery phase to avoid faulty paths. This is a very efficient approach to detect secure routes. In a related paper, Awerbuch [30] discusses a method for secure ad hoc routing.

Zhou *et al.* [31] have an alternative solution for the problems with AODV and DSDV routing methods. They have developed a hybrid approach using both AODV and DSDV methods. This method, known as the Key Management Service (KMS), defends routing from denial of service attacks in ad hoc networks by taking advantage of multiple routes between nodes. Due to the dynamic changes in topology, the routing protocols of ad hoc network need to handle outdated routing information, which is similar to that of the compromised routing attacks. The principle here is that as long as there are enough proper nodes, the routing protocol would be able to find the routes working around the compromised nodes. Thus, if the nodes can find multiple routes, nodes can switch to an alternate route when a fault has been

detected in the primary route. This method also uses replication and new cryptographic schemes, such as threshold cryptography, to build a highly secure and highly available key management service, which forms the core of the security framework.

In addition to the methods discussed above, there are some additional methods proposed in the literature to handle various forms of attacks. For example, the Secure Routing Protocol (SRP) by Papadimitratos *et al.* [31] guarantees correct route discovery, so that fabricated, compromised, or replayed route replies are rejected or never reach the route requester. SRP assumes a security association between the end-points of a path only and so intermediate nodes do not have to be trusted for the route discovery. This is achieved by requiring that the request along with a unique random query identifier reach the destination, where a route reply is constructed and a message authentication code is computed over the path and returned to the source. The authors prove the correctness of the protocol analytically.

Another preventive solution for DoS attacks in ad hoc wireless networks is proposed by Luo *et al.* [32]. In this solution they distribute the functionality of authentication servers, thus enabling each node in the network to collaboratively self-secure themselves. This is achieved by using the certificate-based approach. This scheme supports ubiquitous security for mobile nodes, scales to network size, and is robust against adversary break-ins. In this method centralized management is minimized and the nodes in the network collaboratively self-secure themselves. This scheme proposes a suite of fully distributed and localized protocols that facilitate practical deployment. It also features communication efficiency to conserve the wireless channel bandwidth and independency from both the underlying transport layer protocols and the network layer routing protocols.

The ARIADNE method developed in Europe is another important secure on-demand routing protocol. Developed by Hu *et al.* [33], ARIADNE (Alliance of Remote Instructional Authoring and Distributed Networks for Europe) prevents attackers from tampering with uncompromised routes consisting of uncompromised nodes. It is based on Dynamic Source Routing (DSR) approach and relies on symmetric cryptography only. ARIADNE protocol is designed in three stages: The first stage presents a mechanism that enables the target to verify the authenticity of the Route Request. Second stage presents a key management protocol that relies on synchronized clocks, digital signatures, and standard MAC (Message Authentication Code) for authenticating data in Route Requests and Route Reply. The final stage presents an efficient per-hop hashing technique to verify that no node is missing from the node list in the Request. Hu *et al.* present simulations that show that the performance is close to DSR without optimizations.

Marti *et al.* [34] have taken another variation on the DSR method. This method shows increased throughput in Mobile Ad Hoc Networks by complementing DSR with a watchdog for detection of denied packet forwarding and a pathrater for trust management and routing policy rating that every path uses, thus enabling nodes to avoid malicious nodes in their routes as a detective and reactive protection measure. This reaction does not punish malicious nodes that do not cooperate, but actually relieves them of the burden of forwarding for others while having their messages forwarded, and it allows nodes to use better paths and thus increase their throughput.

The traditional Secure Routing Protocol (SRP) is well suited for a wired network. In developing a similar protocol for MANETs, Yi *et al.* [35] propose a new routing technique called Security-Aware ad hoc Routing (SAR) that incorporates security attributes as parameters into ad hoc route discovery. SAR enables the use of security as a negotiable metric to improve the relevance of the routes discovered by ad hoc routing protocols. Ad hoc routing protocols enable nodes in ad hoc networks communicate with their neighbors through Route REQuest (RREQ) packets and Route REPly (RREP) packets. In SAR, the security metrics are embedded into RREQ packets. Intermediate nodes receive these packets with particular security level and process these packets or forward the packets depending on the security level of the intermediate node. If it cannot provide required security level, RREQ packets are dropped. Otherwise RREP packets are sent back to the source from destination or intermediate nodes. This approach, though resource intensive is a useful alternative for preventing attacks.

So far we have looked at research that addresses authentication, denial of service, selfish node and routing protocol attacks in a MANET. One of the main requirements in a MANET is for each node to let other nodes know of their presence and readiness to participate in the MANET. In a wireless local area network, an Access Point (AP) is used to let the mobile nodes communicate with other nodes on the network. In a MANET, there is no Access Point and so each node must know the other nodes that participate in the MANET. One way to let the other nodes know of their presence, a mobile node sends out beacon signals. Binkley *et al.* [36] propose an authenticated routing protocol to address link security issues in this regard. This proposal also reduces the DoS threats like replay attacks caused by an Address Resolution Protocol (ARP) or ad hoc routing protocol spoof, which would destroy a link-layer route to a host. This protocol transmits beacons similar to that of mobile IP agents. When a host node or agent receives the transmitted beacons, they authenticate them and if it is authentic, they add the MAC-to-IP address binding contained in the beacon into their table of authentic bindings.

Another security scheme proposed by Kong *et al.* [37] and Luo *et al.* [32] supports ubiquitous security services

for mobile hosts through threshold secret sharing mechanism where they distribute certificate authority functions. These methods are based on RSA cryptography and provide distributed localized certificate services like certificate issuing, renewal and revocation. These methods employ localized certification schemes to enable ubiquitous services. This model uses RSA system key pair denoted by {Sk, Pk} where Sk is the system secret/private key and is used to sign certificates for all entities in the network. Pk is the system public key which verifies the certificate signed by Sk. In this scheme, Sk is shared among network entities but not visible or known by any component in the network, except at the boot strapping phase. Each entity Vi also maintains a secret share Pvi and a RSA personal public and private key pair {Ski, Pki} besides the system key pair. Thus, it uses the concept of threshold secret sharing and updating each entity's secret share periodically to further enhance robustness against break-ins. This scheme scales to network size and is robust against break-ins. In the threshold secret sharing mechanism each entity holds a secret share and multiple entities in a local neighborhood jointly provide complete services.

There are several open issues in the models that were reviewed. The important among them are explained as follows: The GDH method needs further study for the detection and resolution of inconsistent certificates, improvement of certificate graph models and enhancing the use of existing PKI infrastructure. The MOCA method uses a unicast approach that only exploits information in the local routing cache. One useful extension would be to devise a way for a node to browse neighboring nodes' routing tables. This would help in avoiding flooding. The CORE method considers only attacks from selfish nodes but not from active intruders. Hence one has to extend this method for intruder attacks as well. The solution for attack by selfish nodes presented in the nuglets method is focused just on packet forwarding attacks. Application-level issues like mutual provision of information services in an ad hoc network have to be addressed in order to better utilize the nuglet counter. The CONFIDANT method assumes that nodes are authenticated and that no node can pretend to be another in order to get rid of a bad reputation. This assumption could lead to misplaced trust in systems. The Guardian Angel method is not a comprehensive security scheme since it does not take into account the attacks like packet forwarding and denial of service or routing attacks, which are commonplace today.

3.1.2. Review of MANET Attack Prevention Security Schemes for Passive Attacks

We noted earlier some of the problems due to selfish nodes not performing their role properly in a MANET. Actions of a selfish node could lead to congestion, lower throughput and denial of service. Buttyan *et al.* [38] have shown by simulation that a selfish node does not participate actively in packet forwarding in order to conserve electrical energy. This study shows that typically every node spends 80% of the energy in forwarding packets. This work also introduces a special counter called nuglet counter that is used to keep track of selfish behavior of nodes. In trying to solve the selfish node problem, Michiardi *et al.* [39] have developed a model called CORE (Collaborative REputation). Under CORE's approach, every node monitors the behavior of the neighboring nodes for a particular requested function and collects data about the execution of that function. If the observed result of the function matches with the expected result, then the observation takes a positive value. This mechanism allows a node to detect if any of its neighbors are selfish nodes and gradually isolate them.

As seen above, the problem of selfish behavior by nodes in a MANET is something significant that needs to be addressed. In a MANET, many nodes try to conserve battery life and consequently resort to selfish behavior by dropping packets rather than forwarding them as they are supposed to do in a network. Buchegger *et al.* [40] study the vulnerabilities exposed by selfish nodes in a MANET. Buchegger *et al.* [40] introduce a new protocol called CONFIDANT (Cooperation of Nodes-Fairness In Distributed Ad hoc NETworks) to address this problem. Each node maintains reputation indexes about each of its neighbors based on their behavior and use these indexes to isolate misbehaving nodes. Avoine *et al.* [41] have developed a cryptography-based fair key exchange model called Guardian Angel. This model uses a probabilistic approach without involving a trusted third party in key exchange.

3.1.3. Limitations of Existing MANET Attack Prevention Schemes and Open Research Issues

1) Active Attack Security Approaches

The scheme GDH needs further exploration of mechanisms for the detection and resolution of inconsistent certificates, improvement of certificate graph models and making use of existing PKI infrastructure [26]. Scheme MDA does not provide authentication of the participants. In addition, more formal arguments need to be developed to support optimality claims [41]. Unicast approaches by the scheme MOCA only exploit information in the local routing cache. One potential extension is to let a node browse into neighboring nodes' routing tables. For example, a node may be short of one or two cached routes and that would lead to flooding. If the node has a way to peek into the neighbors' routing tables and find a couple of new cached routes, it can avoid flooding. Potential overhead for this approach would be the extra communication required between neighbors to exchange the information in routing tables. Whether the benefit would surpass the overhead is an interesting question to investigate [25]. All the unicast based approaches in the

MOCA protocol do not take into account the direction of Certification REQuests (CREQs). At a worst case, all the MOCAs picked by its unicast approach could reside on one side of the network from the requesting node. Then it is possible that all the CREQs are sent into one direction sharing the same next hop nodes, potentially causing unnecessary contention. This leads to a failure or at least delayed responses. One possible solution for such a scenario is to utilize the next hop field in the cached routing table entries. For example, by selecting a set of MOCAs with all the different next hops, one can expect to have a spatial load balancing effect in that each CREQ will go out in different directions [25].

The SEAD approach does not incorporate mechanisms to detect and expose nodes that advertise routes but do not forward packets [29]. In the Beacon scheme, scalability is an issue if there are large numbers of nodes compared to the available bandwidth. The proposed model assumes all nodes in a network share a symmetric key used only for beacon authentication. In addition to problems with scalability, every agent and mobile node at the site has to know the network authentication key. The symmetric keys might be replaced with public key cryptography. Public-key signature and verification of beacons and Mobile-IP registration messages is feasible, even though transmitting such a signature requires more link bandwidth. Every node can possess its own key and simply sign its beacons and registrations. The distribution of certificates such that mobile nodes and agents can verify a beacon is again a higher-level problem [36]. SOS model provides fully localized design, easy support of dynamic node membership, limited intrusion tolerance capacity and decreasing overhead over time. While these characteristics are appealing, this scheme also has limitations as this is achieved at the increased computational overhead (associated with asymmetric cryptography primitives) compared with other hash function based designs [27]. In the TRUST model when a new node enters the system, it assumes that the node already has an initial certificate. This results in the problem of registering users. Also when two ad hoc networks merge, this model does not provide mechanisms for nodes originated from different networks to certify and authenticate each other [32]. In SRP model, fair utilization of network resources is an issue. Possible ways to dismay nodes from broadcasting at the highest possible rate is still an issue [36]. Since the ARIADNE model does not possess the optimizations of DSR, the resulting protocol is less efficient than the highly optimized version of DSR that runs in a trusted environment [33]. An important aspect of OSRP scheme is that the algorithm can be used to detect a fault. However, it is difficult to design such a scheme that is resistant to a large number of adversaries. The method suggested in this paper uses a fixed threshold scheme. This scheme does not explore other methods, such as adaptive threshold or probabilistic schemes which may provide superior performance and extensibility. Also this scheme does not provide means of protecting routing against traditional denial of service attacks [30]. The Watchdog and Pathrater model assumes that there are no apriori trust relationships. Performance of model is bound to suffer when trusted node lists in ad hoc networks are also taken into account. Also, in this model, all the simulations are based on Constant Bit Rate (CBR) data with no reliability requirements. The analysis should be extended to explain how the routing extensions perform with TCP flows common to network applications [34].

2) Passive Attack Security Approaches

The scheme CORE considers only attacks from selfish nodes but not from active intruders. Hence the scheme needs to be extended and tested for intruder attacks as well. Also there is no definition of formal method to analytically prove robustness of CORE [39]. The solution for attack by selfish nodes, presented in Nuglets model is focused just on packet forwarding attacks. This model also does not address application-level issues like mutual provision of information services in an ad hoc network [38]. The CONFIDANT protocol assumes that nodes are authenticated and that no node can pretend to be another in order to get rid of a bad reputation [40]. The Guardian Angel model is not a comprehensive security scheme and does not take into account the attacks like packet forwarding and denial of service or routing attacks [41].

3.2. MANET Intrusion Detection and Response Security Approaches

3.2.1. Review of MANET Intrusion Detection Security Approaches

The following are some of the popular IDA models that we studied in our literature survey. Kachirski and Guha proposed an IDS model which is efficient and bandwidth-conscious [42]. It targets intrusion at multiple levels and fits the distributed nature of IDA for Mobile Networks. The method has clusters and the IDA on cluster head employs independent detection decision-making after gathering information from other nodes. It utilizes mobile agent for communication among various nodes. This model provides a framework to work with multiple types of audit data. It is expandable, meaning, if the IDA needs to work with new types of audit data, it can do so by just incorporating extra agents that can monitor the new type of audit data. Unfortunately, its performance is not verified by any implementation. Once its performance is proved to be on an acceptable level, this framework can serve as a generic and expandable architecture for commercial products, since having a possibility to add in more functionality is an important property for successful products. Because it utilizes the cluster heads, it is supposed to make the network more efficient by

limiting the resources usage for IDA purposes to only a few nodes. Such a framework can be applied in an environment where the security requirement is medium and efficiency requirement is high. Also, it may easily be expanded for multi-layered mobile networks.

IDS model for wireless Mobile Ad Hoc Networks proposed by Zhang and Lee implements local and collaborative decision making with anomaly detection [43]. In this approach, individual IDA agents can work by themselves and also collaborate in decision making. Each IDA agent runs on a node and monitors local activities. If a node detects local intrusion with strong evidence, then the node concludes that intrusion has happened and initiates an alarm response. However, if the evidence is not strong enough but needs investigation in a wider area in the network, then the IDA agent can start collaboration procedure which is a distributed consensus algorithm. This model provides a framework that fits the distributed nature of mobile networks as well. It also works with multiple types of audit data. If the IDA needs to work with new types of data, it can add in more data collection module in the IDA agent. It uses data mining as the local intrusion detection mechanism. The data mining is supposed to be superior in terms of both detection rate and false alarm rate. Also, because this IDA does not use mobile agents for communication, it can be designed for high security need, if it can find an effective way to protect from Byzantine nodes.

Huang and Lee have proposed a cluster-based scheme in which a cluster head is elected by a group of nodes in a neighborhood (citizen nodes) and the head node monitor the citizen nodes [44]. Once the cluster head is elected, the other nodes need to transmit the features they obtain locally to the cluster head. This IDA uses anomaly detection implemented with data mining as its detection technique [44]. This model improves the efficiency of mobile networks by limiting the resources usage for IDA purposes to only a few nodes. The implementation proves it can also achieve satisfactory level of detection rate. Such a framework can be applied in environments where the security requirement is medium but efficiency requirement is high. Also, it may easily be expanded for multi-layered mobile networks [45].

Patrick and Camp have designed architecture for ad hoc networks, where each node runs a local IDA [46]. Each node detects intrusion locally and uses external data to confirm the detection. The nodes use mobile agents to communicate and collaborate. This model provides a scalable architecture by using mobile agents. If the IDA needs more functionality, it can just incorporate more mobile agents with new tasks. It is supposed to reduce network traffic for intrusion detection purpose. However, since this architecture relies heavily on the use of mobile agents, it incurs computational complexity in creating and managing all the agents. This architecture needs an implementation to verify its performance.

Bo, Wu and Pooch have proposed an IDA model which uses collaboration mechanism with anomaly detection [47]. In this model, a network is divided into logical zones. Each zone has a gateway node and individual nodes. Individual nodes have an IDA agent to detect intrusion activities individually. Once an individual node detects intrusion, it generates an alert message. Gateway node aggregates and correlates the alerts generated by the nodes in its zone. An algorithm is used to aggregate the alerts based on the similarities in the attributes of the alert [45]. Only gateway nodes utilize the alert to initiate an alarm [46]. This method does not use mobile agents but has gateway nodes, which work just like a cluster head. This architecture can be applied in environment where the requirement for IDA performance and security is high.

Huang *et al.* have proposed a detection algorithm scheme that uses the statistics of packets, namely, the relation between different features such as the correlation between the number of packets dropped and the percentage of change in routing table [48]. This algorithm can be used as an intrusion detection engine in other IDA architectures. This model has low overhead, but was designed only for one routing protocol-OLSR and needs modification for other protocols.

Tseng *et al.* have proposed an IDS system where the normal behavior of critical objects in the network is constructed with the normal specification first. Then the actual behavior is compared to the normal specification [49]. It uses distributed network monitor to trace the request-reply flow in the routing protocol. The network monitor runs a specification based detection algorithm to make decisions [50,51]. This model is novel with no conventional local detection mechanism, but has low efficiency since packet is checked at each hop.

Neighborhood Watch, an IDS protocol proposed by Sowjanya and Shah has two neighboring nodes of which one node is used to ensure that the packets are not modified while traveling in the network [52]. This is done by comparing the information in each packet at each hop. It has two modes: passive mode-to protect a single host and active mode-to collaboratively protect the nodes in a cluster. In active mode, a cluster head starts a voting algorithm to determine whether intrusion really happens.

Puttini *et al.* have proposed an IDS architecture where information in the management information base (MIB) is used as input data [53]. It also uses mobile agent and a collaborative decision making mechanism. This model is distributed and efficient in use, with high scalability and can detect attack at multiple levels, but has security, computational cost and management problems related to mobile agents.

IDS Model proposed by Brutch and Ko is a statistical anomaly detection algorithm [54]. It works by first assuming that the audit trail generated from a host has been converted to a canonical audit trail (CAT) format. It then

uses a CAT file to generate session vectors representing the activities of the users' sessions. These session vectors are then analyzed against specific types of intrusive activities to calculate "anomaly scores". If the scores cross some thresholds, warnings reports are generated. The algorithm analyzes a session vector in three steps:

1) it calculates a Bernoulli vector,

2) it calculates the weighted intrusion score, and

3) it calculates the suspicion quotient. The Bernoulli vector is generated from the session vectors as well as some threshold vectors. It is a simple binary vector in which the values in the vector are set to one if the corresponding arbitrary counts fall outside the threshold for a particular user group. The weighted intrusion score is generated for a particular session and for a particular intrusion type. It can be used to assign a suspicion value to the session. This suspicion value, or suspicion quotient, for a session is determined by what percentage of random sessions have a weighted intrusion score less than or equal to the weighted intrusion score of the current session. It describes how closely a session resembles the intrusion type as compared to all other sessions. The Haystack algorithm gets its name by being the algorithm implemented in the IDA called Haystack. Haystack is a host-based system, which attempts to detect several types of intrusions: attempted break-ins, masquerade attacks, penetration of the security system, leakage of information, denial of service, and malicious use. It was initially developed for use in the US military network. This algorithm is designed for use in a secured wired military network. If in a wireless ad hoc environment, it requires a designated node to act as a central administrator and all the other nodes to allow the central administrator to retrieve audit trails from them. The central administrator can be pre-designated by the human initiator of the ad hoc network or can be assigned by programming. The audit trails requested can be submitted by the nodes themselves or by the mobile agents allowed to run on the nodes.

An IDS approach, Indra, proposed by Janakiraman *et al.* is a distributed intrusion detection scheme based on sharing information between trusted peers in a network to guard the network as a whole against intrusion attempts [55]. It is a detection tool that takes a proactive and P2P approach to network security. The basic idea behind this model is cross monitoring or simply called "neighborhood watch," and is very simple. In this method, the hosts on the P2P network join together to form some sort of an immune system where each host distributes information on attempted attacks among the interested peers in the network. Such information is usually gathered by the intended victim of an attack and by notifying its adjacent hosts, an alarm can be sounded. This allows the system to react proactively or retroactively. When an alarm is sounded, subsequent attacks to other hosts are repelled straightaway as the adjacent hosts would have forewarned other hosts.

Most of the surveyed models use packets and network traffic related information such as updates in routing table or request-reply flow in the network. Among the ones that use packets related information, IDS approach proposed in [50,51] uses the information inside the packets header directly, such as network address or port number. Other models using packet or network traffic related information mainly use statistical data processes from packet information, such as the statistics of the number of packets received and sent or the statistics of change in routing table. IDS Model as described in [48] utilizes the statistics derived from packet or traffic related statistics, for instance, the correlation between the number of packets dropped and the percentages of updates in routing table. Intrusion Detection approaches illustrated in [43] allow the IDA to work on different types of audit data or the possibility to adapt to different types of audit data. This property is valuable and should be an important consideration for the future design of IDA. Most of the architectures detect only the fact that an intrusion happens. Some models go further to obtain more information, such as the type of attack and the location of the intruder. For instance, Zone based IDA can detect both the type and location of the attack [46].

Some of the intrusion detection models utilize cluster head or gateway nodes [42]. The advantage of cluster head is that some of the resource consuming computation, such as intrusion detection, can be carried out only on some nodes of the network. Therefore, most other nodes can focus on the real work of network traffic. The cluster head usually collects information from cluster member to make the detection decision. In some methods, the original input data is further processed or formatted before it is sent to the cluster head. By doing this, the network traffic for transferring such data is reduced. The computation on the cluster head can also be reduced because the incoming data from member nodes is already formatted for the IDA use. The security communication between the cluster head and its member nodes should receive attention of research.

Most of the methods in our review, except the model proposed in [49], utilize anomaly detection. The anomaly detection is more suitable than misuse detection in Mobile Ad Hoc Networks. In Mobile Ad Hoc Networks, the anomaly detection has a weakness: the profile of normal behavior needs to be updated periodically. This places a heavy burden on the limited network resources.

3.2.2. Review of MANET Intrusion Response Security Approaches

Although intrusion response component is related and coexist with the intrusion detection framework, it receives considerably less attention than detection framework owing to the inherent complexity in developing and deploying response in an automated fashion [56]. Most

of the security models generate an alarm informing the administrator, who then decides the response. However, it is desirable that the response consists of an automated corrective action to protect the network from an identical future attack.

There are few IDA models that provide the integrated detection and response feature. Zhang *et al.* in their framework have explained that local response module triggers action local to the mobile node and the global response module coordinates actions among neighboring nodes, such as the IDS agents in the network electing a remedy work [43]. They have also explained that the type of response depends on the type of intrusion, the type of protocols, applications and the confidence in the evidence with examples. However, they have not provided any implementation details regarding the intrusion response aspect of the model. Similarly, there is no documentation on the simulation or experimental results on the response aspect of the model. However, there is a detailed explanation on the experimental results of the detection framework of the model. Thus, even though the idea of integrated detection and response model seems feasible, it appears that the implementation and simulation have not been conducted. Similarly, few related IDA models propose response actions/frameworks for responding to the attacks once it is detected [57-65]. However the response system incorporating all those actions is not implemented.

There are a few intrusion prevention approaches described in the literature for mobile ad hoc network security as well. Puttini *et al.* have proposed a secure routing protocol that combines a certificate based authentication service with intrusion detection model to provide preventive and corrective protections for Mobile Ad Hoc Networks [53]. Bhargava *et al.* have proposed a security model for AODV routing protocol to prevent attacks in mobile networks [66].

3.2.3. Limitations of existing MANET Intrusion Detection and Response Security Approaches and Open Research Issues

The misuse detection systems use patterns of known attacks to match and identify those intrusions [67]. Although it can accurately and efficiently detect instances of known attacks, it lacks the ability to adapt in detecting new type of attacks. The anomaly detection systems on other hand detect intrusions by finding deviations from the established user profiles. Anomaly detection should detect new types of intrusions but it could have higher false positive rate [68]. Traditionally, IDA are developed using expert knowledge of the system and attack methods [48]. Due to the complexity of modern network system and sophistication of attackers, expert knowledge engineering is often very limited and unreliable [43]. Some IDA schemes are very sensitive to the data representation. For instance, these schemes may fail to generalize an unseen data if the representation contains irrelevant information. In some instance, it has been observed that training of IDA requires a noise-free data (the data that is labeled 'normal') [42]. It has been observed that the existing IDA performs poorly in detection as well as the false positive rates at higher mobility rates [46]. It has recently been observed that Denial of Service (DoS) attacks are targeted even against the IDA [18]. Thus, IDA themselves needs to be protected. An IDA should also be able to distinguish an attack from an internal system fault.

The identification of intruder and appropriate response techniques to protect Mobile Ad Hoc Networks still represents a challenging issue. The need to coordinate intrusion detection and response techniques and the need to respond and control the identified attacks effectively, require further research. It can be noted that though the response concepts are explained in the existing intrusion detection models, implementation details and results for the response framework are not provided to demonstrate and validate their response techniques. Also according to our literature review, we observe that none of the existing models has proposed an intrusion control approach for mobile and senor networks, such that detection and response are done continuously to protect the mobile ad hoc networks.

To summarize, the related existing intrusion detection and intrusion response approaches suffer from one or more of the following limitations specifically with respect to mobile ad hoc networks:

- Lower detection rate when mobility is used as a pa rameter.
- Higher false positive rate when mobility is used as a parameter.
- Appropriate response techniques to protect Mobile Ad Hoc Networks after threat detection.

4. Review of Mobile Agent Model Security Approaches

In the following sections, we present the security approaches for the different attack scenarios explained earlier in Section 2.

4.1. Security Approaches for Mobile Agent Attacked by Another Agent

Location privacy through user smart card is proposed by [69]. This scheme takes care of the unauthorized access, masquerade attacks, which is achieved through secret keys for secure communication with network and the other users. It has some advantages like location and identification privacy in addition to just content privacy. This proposal uses digital mix proposed by Chaum [70]. A digital mix enables two parties to communicate with-

out unauthorized parties being able to determine either the message content or the source and destinations of the messages. In addition, the sender of a message can remain anonymous to the recipient. This is achieved through an intermediate computer called a 'mix' processes messages so that header information is hidden from following communications. The main idea is new authentication, digital mix, information leak and billing services. The architecture new security features for mobile networks with existing infrastructure be provided through additional intelligent network services.

Profiling mobile users by Bayesian decision algorithm [71] proposes to provide detection and response solution for an agent attacked by agent privacy problems like masquerade and unauthorized access. This proposal focuses on the application of anomaly detection techniques to mobile networks and generation of user profiles within GSM mobile networks.

Enhanced privacy and authentication for GSM by C. H. Lee *et al.* [72] proposes three improved methods to enhance the security, to reduce the storage space, to eliminate the sensitive information stored in VLR, and consequently to improve the performance of the system. It includes an improved authentication protocol for the mobile station, a data confidentiality protocol, and a location privacy protocol. This proposal tends to improve but not to alter the existing architecture of the system, which is a very useful feature for the practical reasons. This scheme attempts to provide a solution for unauthorized access and masquerading by means of improved authentication protocol which eliminated the redundant sensitive information stored in Virtual Location Register (VLR), data confidentiality protocol (with/without session key table in Home Location Register (HLR/SC) and location privacy protocol with/without conference key shared by HLR's.

4.2. Security Approaches for Mobile Agent Attacked by the Host

Mobile code cryptography [21] provides solution through encrypted functions and digital signing. This proposal uses cryptographic primitives and homomorphic encryption schemes (public key) and function composition schemes. This solution tries to prove that mobile code holds the key to uncouple the secure execution of programs from the trustworthiness of the underlying execution support. This solution tries to prove that one can obtain a system where a host can execute an encrypted function without having to decrypt it. The functions would be encrypted such that the resulting transformation can be implemented as a (mobile) program that will be executed on a remote host. The executing computer will see the program's clear text instructions but will not be able to understand the function that the program implements. This scheme attempts to provide a solution for masquerade and eavesdropping attacks by host on agent. This is achieved with the help of cryptography and encrypting agent functions that are executed by the host. This is realized via homomorphic functions and homomorphic encryption scheme.

Secure and open mobile agents (SOMA) [73] provide secrecy and integrity to the mobile agents by means of encryption and authenticated channels. Here agent is encrypted and digitally signed. This model has no overhead as in Trusted Third Party (TTP) solutions. The solution is an efficient, scalable and robust than multiple host (MH) protocols. However this proposal does not discusses about secrecy during the agent execution and secure delegation. This scheme attempts to provide a solution for eavesdropping, masquerade and alteration attacks on agent by host. This is achieved through a security infrastructure and layered security policies that imposes authorizations and authentications. The security infrastructure consists of a policy server, a domain server for domain management, a role server for role management, a certification authority for issuing and the lifecycle management of certificates, an authentication server, an authorization server.

Another proposal, AJANATA [74] provides secure access to system resources and supports isolated protection domains for agents by using supported thread groups and class loaders. This security architecture provides a solution for providing denial of service, alteration, eavesdropping and masquerade attacks by host on agent. This is achieved by authentication protocol, by generic Agent-Server class, Ajanta security manager. Authentication protocol's name services enforce its security policies. The architecture also provides protected name spaces for different users. This model uses proxy concept and protects the information of agent. The proposed architecture is built upon Java's security model and address problems related to protecting agent servers, agents and the name service information.

A solution through smartcards [9] by multifunctional trusted smart cards uses Java card for authentication and signing device, when user sends an agent and for trusted computing base attached to host environment. This scheme attempts to protect agent from alteration, denial of service and masquerade attacks by host on agent. This is achieved by allowing agents to carry encrypted code parts and protecting an agent's itinerary by means of security store. The decrypted form will be visible to smartcard only. This is achieved by using public key encryption with certified public keys. This approach protects specific parts of mobile agent better than just using Java Card alone. This proposal which uses trusted computing base claims better protection for the agents than the mobile code cryptography, encrypted functions, code obfuscation and cryptographic trace etc.

A public key based secure Mobile IP was proposed by Zao *et al.* [75] in their Mobile IP Security System (Mo-

IPS) was based on a DNS based X.509 PKI and the innovation in cross certification and zero-message key generation. This proposal attempts to provide solution for alteration, masquerading and eavesdropping attacks by means of key management and cryptographic keys for authentication, access control and using secure tunneling. The system supplies cryptographic keys for authenticating Mobile IP v4 location management messages and establishing IPSec tunnels for Mobile IP redirected packets. It was developed to support three services that are essential to the safe operation of Mobile IP: 1) authentication of Mobile IP control messages for location update, 2) access control of Mobile Nodes to resources in the foreign networks, and 3) secure tunneling to redirected IP datagram. Public key technology is used for the scalability reasons. A DNS based PKI has clear advantage over a distributed system of key distribution centers (KDC) since PKI solves the potentially complicated server discovery problem, and it eliminates the need for real-time key dispatches by the KDC.

Sufatrio and Lam [76] proposed a solution for the security aspect of the registration protocol, an extension in Mobile IP. This scheme provides solution for the masquerade, alteration, non-repudiation and eavesdropping attacks, through the public-key based authentication with a minimal use of public key cryptography. This scheme also attempts to provide solution for a replay attack on mobile agent's registration. It provides a scalable solution for authentication and non-repudiation and also strives for minimal computing and administration cost on the mobile agent.

Detecting malicious changes to an agent's state during its execution or data does not yet have a general solution yet.

4.3. Security Approaches for Host Attacked by Mobile Agents

Authentication protects host [3] by preventing agent pretending as host. This is achieved through shared key for encryption messages or privacy.

The issues that face this model are the authentication is needed whenever the agent traverses each new cell, especially with network partitions. This model tries to address the following security goals.

1) Walkstation (mobile agent or computer) and the basestation must be able to authenticate each other. It prevents a malicious station from pretending to be a base station and also it permits the walkstation to choose the services of a particular base station in the presence of collocated networks.

2) Once authenticated walkstation and basestation should be able to communicate securely. Privacy has two dimensions: data privacy and location privacy.

3) Walkstations should be provided location privacy. Some applications will require location privacy, while others may exploit the knowledge of walkstations. The goal is to provide location privacy at the lowest layer. Higher layers may disseminate location information according to the needs of the applications.

4) The security should be optional (due to the tradeoff in the limited resources and the security) and efficient. This scheme attempts to provide secured solution for unauthorized access and masquerade attacks. This is achieved by mutual authentication of base station and walk station and thereby generating a shared key for encryption of messages. This scheme relies on private/public key mechanism to achieve the solution.

The proposal of SOMA architecture provides authentication and authorization for the host security from mobile agents. This model addresses the issue of balanced trade off between several requirements, often contrasting security, flexibility, usability and efficiency. This scheme proposes a scheme for the protection of agents from malicious hosts (sites), which is fundamental for agent-based applications in untrusted environments and are still an active research area. This scheme attempts to provide a solution for masquerade and unauthorized access attacks by agents on host.

The solution through Proof Carrying Code (PCC) [77] provides a security for hosts in the masquerade and unauthorized attacks via proof checker ensured by code producer which is "tamper proof" and "self certifying code/agent". Necula suggests that the theory of programming languages, including formal semantics, type theory and applications of logic, are critical to solving the untrusted-code security problem essentially through the exploitation of static checking for achieving a high level of security in mobile-code applications. The advantages of PCC are that the burden of providing security is shifted to code producer; they are tamperproof and self certifying.

PCC is a technique by which host establishes a set of safety rules that guarantee safe behavior of programs, and the code producer creates a formal safety proof that proves, for the untrusted code, adherence to the safety rules. Then, the host is able to use a simple and fast proof validator to check, with certainty that the proof is valid and hence the foreign code is safe to execute.

Lu *et al.* [23] proposed an algorithm for fair service in error-prone wireless channels this algorithm provides short term fairness among flows which perceive a clean channel, long term throughput and fairness bounds for all flows with bounded channel error, an expanded schedulable region by decoupling delay/bandwidth weights, and supports both delay sensitive and error sensitive data flows. This wireless fair service algorithm attempts to provide solution for denial of service attacks, by providing a performance effective fair service in error-prone communication channels.

Trost and Binkely proposed [24] an authenticated link-level ad hoc routing protocol for Mobile IP, which ad-

dresses link security issues. This scheme attempts to provide solution for unauthorized access and masquerade attacks. It addresses the issues of attacker stealing host's packets. The protocol also eliminates denial of service attacks caused by an ARP spoof destroying data link layer towards a host. The protocol also tries to limit the eavesdropping, copy and replay, alteration attacks an unwanted visitor to do for a host. This is achieved by not only correct implementation of sound protocols but also by proper maintenance methodologies. In this protocol, mobile agent's and node's packets are authenticated and security problems associated with ARP spoofing are also reduced by this scheme. The authentication is provided through network authentication key and adhoc key. This scheme also attempts to provide a solution for the replay attacks by agent.

Perkins proposed a Mobile IP/AAA trust model [78] which relies on the existence of servers that are capable of performing accounting, authentication, and authorization (AAA) services. This new infrastructure is designed to meet the emerging needs of cellular telephony [79] for mobile data service to a large population of mobile telephone users, and eventually over VoIP. Several schemes like security infrastructure in CDMA networks [80] uses the Mobile IP/AAA trust model for their solution. This model attempts to provide a solution for alteration, eavesdropping and masquerade attacks by satisfying the AAA security requirements and protocols.

4.4. Security Approaches for Host Attacked by Other Unauthorized External Parties Including Host and Agents

Protection of dumb host by a scheme for authenticating host in a secure mobile network [81] attempts to provide solution for masquerade and unauthorized attacks. This is achieved using a hierarchy of mobile agents and relies upon the computation priorities to determine which agent is to be active in each authentication request. The scheme attempts to solve the This scheme proposes a hierarchical simulation model and analyzes several factors involved in the computation of priorities, to determine the optimal weightings of each factor involved and the dependence, if any, of these weightings on the factors of the hierarchy itself.

Protection for host by fault tolerant authentication [13] has some positive aspects like fault tolerance and scalability issues taken care, clusters of a node than single over the other proposals like Virtual Router Redundancy Protocol (VRRP), which are not scalable. This proposal attempts to solve the masquerade and unauthorized access attacks on hosts by using hierarchical authentication and a flat model as in a LAN environment. These techniques make use of backup servers. However, the performance issues that affect performance are still the issues that are to be taken care by partitioning the secret key database and through analysis to discover the parameters that affect the performance of the system and study how the priorities depend on these factors.

MACKMAN [82] propose a solution motivated by the deficiencies found in the registration and authentication service of the existing protocols such as GSM, CDPD, and IS-41. This solution employs mutual authentication and digital signatures to provide a more secure registration and authentication service for mobile computing by using Elliptic Curve RSA (ECRSA) for the efficiency reasons. This scheme provides solution for unauthorized access, denial of service and masquerade attacks by addressing the following issues:

- Trustworthiness of Intermediate Network.
- Mutual Authentication between a mobile agent and mobile host.
- Data Confidentiality against both active and passive intrusion by malicious agents.
- Untraceability requires protection of registered users from unauthorized entities. A mobile host should be able to request network services without divulging any access control information to eavesdroppers. The degree of untraceability availability to mobile host depends upon the policies enforced by the underlying system and the tradeoffs between cost and benefits.
- Time Synchronization, since the mobile agents travel across various time zones and administrative authorities and hence the time synchronization in security systems for mobile environments is not recommended.
- Optional Security and Modes of Security: Due to the scarce mobile environment resources likes bandwidth and power and hence various modes of security should be made optional.
- Flexibility: The security system for mobile environments should provide enough flexibility to incorporate future advances in shared-key cryptographic techniques.
- Interoperability: The security system for mobile environments should provide for interoperability between numerous variations and versions of cryptographic products.

Multicast security proposed by LiGong and N. Shahchum [83] tries to provide security in a group-oriented secure data exchange in a multicast environment which could be extended to a mobile environment, where it attempts to solve identity of the originator of a message and group-oriented authentication. These mechanisms are incorporated into session, presentation, and network layers of the network architecture, where they consist of authentication, encryption, and physical access to the tree, respectively. This scheme attempts a solution for masquerade, unauthorized access and denial of service attacks in a multicast environment. Masquerade attack is solved through authentication (using pair wise authenti-

cation model) and secure session membership policies, registration, deregistration, secure session communication (using a common encryption key) and secure broadcast using polynomial interpolation. The problem of message eavesdropping and masquerading is achieved through encryption and decryption. The problem of unauthorized access attacks is solved through pair wise authentication model.

Joseph and Kaashoek proposed [84] proposed building reliable mobile computing applications using the Rover toolkit, to add server-side support for reliable operation, in addition to the existing client-side support. In this scheme they attempted to provide solution for denial of service attacks by implementing server failure recovery procedures and server failure detection.

4.5. Limitations of the Existing Schemes and the Open Research Issues

Since security in mobile computing is an after thought until the recent years, there are many open issues that need to be addressed. Many proposals address the issue of site protection against malicious agents. The complementary problem of protecting agents while executing in potentially malicious sites (host or base station) is specific to MA technology. The secrecy and integrity during agent execution need to be preserved in order to leverage the MA exploitation in wide application contexts. The agency secrecy of both code and state parts represents a challenging issue [85]. It seems rather impossible to hide the agent code from the site responsible for its execution. The same applies to the state part if the code has to work on it.

So far a little research was done on protecting a mobile agent from malicious hosts: the main focus was on making the execution of mobile code efficient and safe for the host. This is due to the assumption that mobile code is impossible to protect without resort to special hardware, simply because the code has to be executed by the hosting system.

However protecting a mobile agent against malicious hosts is not a “nice-to-have” feature but is essential for an agent system’s usefulness [21]. The security research issues could be summarized as follows:

- Can a mobile agent protect itself against tampering by a mobile host? (code and execution integrity)
- Can a mobile agent remotely sign a document without disclosing the user’s private key? (computing with secrets in public).
- Can a mobile agent conceal the program it wants to have executed? (code privacy)
- Secure routing or denial of service attacks protection.
- Can a host (computer) execute a cipher program without understanding it?

Other relevant issues include

- The protection of the executing host from malicious actions of mobile code.
- The protection of the network as a whole (e.g., from spamming agents or hosts).
- The secure routing of mobile code.
- The detection of tampering by and the identification of a malicious host.
- The protection of mobile code against input/output analysis.

In a dynamic system, mobile agents entering remote domains need to have the ability to inherit permissions from their home agents while maintaining information and location security. The security mechanism should be designed so that the provision of security does not add significant delays during call setup and communication and does not waste the scarce resources like wireless link bandwidth and the battery power [10]. Proposed security schemes should be efficient in the number and size of messages exchanged and should not cause the channel bandwidth to increase or cause propagation of errors nor should it result in increased error rates.

Another issue typical to mobile computing environment is the issue of time synchronization, since the mobile agents travel across various time zones and administrative authorities and hence the time synchronization in security systems for mobile environments is not recommended. Also, any security system for mobile environments should provide enough flexibility to incorporate future advances in shared-key cryptographic techniques and numerous variations of cryptographic products.

5. Conclusions

In this paper we have presented the taxonomy of security schemes for mobile computing systems. We have classified them based upon the infrastructure that makes up the mobile computing system and then by the type of attacks. The classification helps increasing our understanding of the security issues and requirements of the mobile computing and the schemes that could solve these issues and requirements. In general, there are tradeoff between the resource constraints, performance, scalability and the provision of security features. Also, there is a no single scheme that provides a general solution for the different kind of security threats in the mobile computing environment. With respect to the MANET based mobile computing system, our analysis shows that the potential threats faced by MANETs come in the form of denial of service, selfish node behavior, or routing attack. Also majority of the recent effort is spent to secure active MANET attacks rather than passive MANET attacks. With respect to the mobile agent model based mobile computing system, providing security for the mobile agent from the fixed host seems to be more challenging than providing the security for fixed host from mobile

agent. The taxonomy developed in this paper highlights the contributions for different types of attacks and shows the different types of approaches taken to provide security. This taxonomy should help researchers focus on underlying methods, limitations of the existing schemes and open research issues needed to secure MANETs.

6. References

[1] H. Reiser and G. Vogt, "Security Requirements for Management Systems Using Mobile Agents," *Proceedings of the* 5*th IEEE Symposium on Computers and Communications*, Antibes-Juan Les Pins, 2000, pp. 160-165.

[2] J. E. Canavan, "Fundamentals of Network Security," Artech House, Boston, 2001.

[3] S. Funfrocken, "Protecting Mobile Web-Commerce Agents with Smartcards," *Proceedings of the* 1*st International Symposium on Agent Systems and Applications*, Palm Springs, California, 1999, pp. 90-102.

[4] H. Deng, Q. Zeng and D. P. Agrawal, "Network Intrusion Detection System Using Random Projection Technique," *Proceedings of the International Conference on Security and Management*, Las Vegas, 2003, pp. 10-16.

[5] A. Sundaram, "An Introduction to Intrusion Detection," *Crossroads*: *The ACM Student Magazine*, Vol. 2, No. 4, 1996, pp. 3-7.

[6] J. P. Anderson, "Computer Security Threat Monitoring and Surveillance," James P. Anderson Co., Fort Washington, 1980.

[7] A. Mitrokotsa, N. Komninos and C. Douligeris, "Intrusion Detection with Neural Networks and Watermarking Techniques for MANET," *Proceedings of IEEE International Conference on Pervasive Services*, Los Alamitos, CA, USA, 2007, pp. 118-127.

[8] J. Hubaux, L. Buttyan and S. Capkun, "The Quest for Security in Mobile Ad Hoc Networks," *Proceedings* of *the MobiHoc Conference*, California, 2001, pp. 146-155.

[9] F. Stajano and R. Anderson, "The Resurrecting Duckling: Security Issues for Ad Hoc Wireless Networks," *Proceedings of International workshop on Security Protocols*, Berlin, 1999, pp. 172-194.

[10] P. Vinayakray-Jani, "Security within Ad Hoc Networks," *Presented at First PAMPAS Workshop*, London, 2002, pp. 66-67.

[11] K. Wrona, "Distributed Security: Ad Hoc Networks and Beyond," *Presented at First PAMPAS Workshop*, London, 2002, pp. 70-71.

[12] L. Buttyan and J. Hubaux, "Report on a Working Session on Security," *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 7, No. 1, 2003, pp. 74-94.

[13] P. Michiardi and R. Molva, "Simulation-Based Analysis of Security Exposures in Mobile Ad Hoc Networks," *Proceedings of European Wireless Conference*, Florence, 2002, pp. 287-292.

[14] B. Awerbuch, D. Holmer, C. Nita-Rotaru and H. Rubens, "An On-Demand Secure Routing Protocol Resilient to Byzantine Failures," *Proceedings of ACM Workshop on Wireless Security*, Atlanta, 2002, pp. 21-30.

[15] P. Papadimitratos and Z. Haas, "Secure Routing for Mobile Ad Hoc Networks," *Proceedings of SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*, San Antonio, 2002, pp. 27-31.

[16] S. Buchegger and J. Boudec, "Nodes Bearing Grudges: Towards Routing Security, Fairness and Robustness in Mobile Ad Hoc Networks," *Proceedings of* 10*th Euromicro Workshop on Parallel*, *Distributed and Network-based Processing*, Canary Islands, 2002, pp. 403-410.

[17] P. Michiardi and R. Molva, "Prevention of Denial of Service Attacks and Selfishness in Mobile Ad Hoc Networks," Research Report RR-02-063, Institute Eurecom, 2002.

[18] B. K. Bhargava, S. B. Kamisetty and S. K. Madria, "Fault Tolerant Authentication in Mobile Computing," *Proceedings of International Conference on Internet Computing*, Las Vegas, Nevada, USA, June 2000, pp. 395-402.

[19] A. Fugetto, G. P. Pivvo and G. Vigna, "Understanding Code Mobility," *IEEE Transactions on Software Engineering*, Vol. 24, No. 5, 1998, pp. 342-361.

[20] D. Johansen, R. V. Renessee and F. B. Schneider, "An Introduction to the TACOMA Distributed System-Version 1.0," Technical Report, Department of Computer Science, University of Tromso and Cornell University, 1995.

[21] T. Sander and C. Tschud, "Towards Mobile Code Cryptography," *Proceedings of IEEE Symposium on Security and Privacy*, California, 1998, pp. 215-224.

[22] B. Askwith, M. Merabti, Q. Shi and K. Whiteley, "Achieving User Privacy in Mobile Networks," *Proceedings of* 13*th Annual Computer Security Applications Conference*, USA, 1997, pp. 108-116.

[23] T. G. Brutch and P. C. Brutch, "Mutual Authentication, Confidentiality and Key Management (MACKMAN) System for Mobile Computing and Wireless Communication," *Proceedings of* 14*th Annual Computer Security Applications Conference*, Scottsdale, Arizona, 1998, pp. 308-317.

[24] B. K. Bhargava, S. B. Kamisetty and S. K. Madria, "Fault Tolerant Authentication in Mobile Computing," *Proceedings of International Conference on Internet Computing*, Las Vegas, Nevada, USA, 2000, pp. 395-402.

[25] S. Yi and R. Kravets, "Key Management for Heterogeneous Ad Hoc Wireless Networks," Technical Report UIUCDCS-R-2002-2290, Department of Computer Science, University of Illinois, 2002.

[26] S. Capkun, L. Buttyan and J. P. Hubaux, "Self Organized Public-Key Management for Mobile Ad Hoc Networks," *Transactions on Mobile Computing*, Vol. 2, No. 1, 2003, pp. 52-64.

[27] H. Yang, X. Meng and S. Lu, "Self-Organized Network Layer Security in Mobile Ad Hoc Networks," *Proceedings of ACM MOBICOM Wireless Security Workshop*, Atlanta, 2002, pp. 11-20.

[28] A. A. Ramanujam, J. Bonney, R. Hagelstrom and K. Thurber, "Techniques for Intrusion-Resistant Ad Hoc

Routing Algorithms (TIARA)," *Proceedings of MILCOM Conference*, Los Angeles, 2000, pp. 660-664.

[29] Y. Hu, D. B. Johnson and A. Perrig, "SEAD: Secure Efficient Distance Vector Routing for Mobile Wireless Ad Hoc Networks," *Proceedings of 4th IEEE Workshop on Mobile Computing Systems & Applications*, New York, 2002, pp. 3-13.

[30] B. Awerbuch, D. Holmer, C. Nita-Rotaru and H. Rubens, "An On-Demand Secure Routing Protocol Resilient to Byzantine Failures," *Proceedings of ACM Workshop on Wireless Security*, Atlanta, 2002, pp. 21-30.

[31] P. Papadimitratos and Z. Haas, "Secure Routing for Mobile Ad Hoc Networks," *Proceedings of SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*, San Antonio, 2002, pp. 27-31.

[32] H. Luo and S. Lu, "Ubiquitous and Robust Authentication Services for Ad Hoc Wireless Networks," Technical Report, Department of Computer Science, 2000.

[33] Y. Hu, A. Perrig and D. B. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks," *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking*, Atlanta, 2002, pp. 12-23.

[34] S. Marti, T. J. Giuli, K. Lai and M. Baker, "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," *Proceedings of 6th Annual Conference on Mobile Computing and Networking*, Boston, 2000, pp. 255-265.

[35] S. Yi, P. Naldurg and R. Kravets, "Security-Aware Ad Hoc Routing for Wireless Networks," *Proceedings of Second ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Urbana, 2001, pp. 299-302.

[36] J. Brinkley and W. Trost, "Authenticated Ad Hoc Routing at the Link Layer for Mobile Systems," *Wireless Networks*, Vol. 7, No. 2, 2001, pp. 139-145.

[37] J. Kong, H. Lou, K. Xu, D. Gu, M. Gerla and S. Lu, "Adaptive Security for Multi-Layer Ad Hoc Networks," *Special Issue of Wireless Communication and Mobile Computing*, Vol. 2, No. 5, 2002, pp. 533-547.

[38] L. Buttyán and J. P. Hubaux, "Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks," *ACM Journal for Mobile Networks* (*MONET*), Vol. 8, No. 5, 2003, pp. 579-592.

[39] P. Michiardi and R. Molva, "Core: A Collaborative Reputation Mechanism to Enforce Node Cooperation in Mobile Ad Hoc Networks," *Proceedings of Communication and Multimedia Security Conference*, Portoroz, 2002, pp. 107-121.

[40] S. Buchegger and J. Boudec, "Performance Analysis of the CONFIDANT Protocol: Cooperation of Nodes-Fairness in Distributed Ad Hoc NeTworks," *Proceedings of Mobi-Hoc Conference*, Switzerland, 2002, pp. 226-236.

[41] G. Avoine and S. Vaudenay, "Cryptography with Guardian Angels: Bringing Civilization to Pirates," *ACM Mobile Computing and Communications Review* (*MC2R*), Vol. 7, No. 1, 2003, pp. 74-94.

[42] O. Kachirski and R. Guha, "Effective Intrusion Detection Using Multiple Sensors in Wireless Ad Hoc Networks," *Proceedings of 36th International Conference on System Sciences*, Hawaii, 2003, pp. 57-64.

[43] Y. Zhang, W. Lee and Y. Huang, "Intrusion Detection Techniques for Mobile Wireless Networks," *Wireless Networks*, Vol. 9, No. 5, 2003, pp. 545-556.

[44] Y. Huang and W. Lee, "A Cooperative Intrusion Detection System for Ad Hoc Networks," *Proceedings of ACM Workshop on Security of Ad Hoc and Sensor Networks*, Fairfax, Virginia, 2003, pp. 135-147.

[45] H. Debar and A. Wespi, "Aggregation and Correlation of Intrusion Detection Alerts," *Proceedings of 4th International Symposium on Recent Advances in Intrusion Detection*, Davis, CA, USA, 2001, pp. 85-103.

[46] P. Albers and O. Camp, "Security in Ad Hoc Networks: A General Intrusion Detection Architecture Enhancing Trust Based Approaches," *Proceedings of 1st International Workshop on Wireless Information Systems*, Ciudad Real, Spain, 2002, pp. 1-12.

[47] B. Sun, K. Wu and U. W. Pooch, "Integration of Mobility and Intrusion Detection for Wireless Ad Hoc Networks," *International Journal of Communication Systems*, Vol. 20, No. 6, 2006, pp. 695-721.

[48] Y. Huang, W. Fan, W. Lee and P. S. Yu, "Cross-Feature Analysis for Detecting Ad Hoc Routing Anomalies," *Proceedings of 23rd International Conference on Distributed Computing Systems*, Providence, 2003, pp. 478-487.

[49] C. Tseng and P. Balasubramanyam, "A Specification-Based Intrusion Detection System for AODV," *Proceedings of ACM Workshop on Security of Ad Hoc and Sensor Networks*, Fairfax, 2003, pp. 125-134.

[50] R. Sekar, "Specification-Based Anomaly Detection: A New Approach for Detecting Network Intrusions," *Proceedings of 9th ACM Conference on Computer and Communications Security*, Washington, DC, USA, 2002, pp. 265-274.

[51] Y. Okazaki, I. Sato and S. Goto, "A New Intrusion Detection Method Based on Process Profiling," *Proceedings of Symposium on Applications and the Internet*, Nara City, Japan, 2002, pp. 82-91.

[52] R. Sowjanya and H. Shah, "Neighborhood Watch: An Intrusion Detection and Response Protocol for Mobile Ad Hoc Networks," UMBC Technical Report, 2002.

[53] R. Puttini, J. Percher, L. Me, O. Camp and R. De Souza, "A Modular Architecture for Distributed IDS in MANET Structures," *Lecture Notes in Computer Science*, Vol. 2669, 2003, pp. 91-113.

[54] P. Brutch and C. Ko, "Challenges in Intrusion Detection for Wireless Ad Hoc Networks," *Proceedings of Symposium on Applications and the Internet Workshop*, Orlando, Florida, 2003, pp. 368-373.

[55] R. Janakiraman, M. Waldvogel and Q. Zhang, "Indra: A Peer-to-Peer Approach to Network Intrusion Detection and Prevention," *Proceedings of 12th IEEE International Workshops*, Linz, 2003, pp. 226-231.

[56] N. Stakhanova, S. Basu and J. Wong, "Taxonomy of Intrusion Response Systems," Technical Report 06-05,

Computer Science, Iowa State University, 2006.

[57] M. M. Islam, R. Pose and C. Kopp, "An Intrusion Detection System for Suburban Ad-Hoc Networks," *Proceedings of IEEE Tencon Conference*, Melbourne, 2005, pp. 41-46.

[58] G. Vigna, S. Gwalani, K. Srinivasan, E. Belding-Royer and R. Kemmerer, "An Intrusion Detection Tool for AODV-Based Ad Hoc Wireless Networks," *Proceedings of the* 20*th ACSA Conference*, Tucson, 2004, pp. 16-27.

[59] R. Puttini, J. Percher, L. Me and R. Sousa, "A Fully Distributed IDS for MANET," *Proceedings of IEEE Symposium on Computers and Communications*, Brasilia, 2004, pp. 331-338.

[60] B. Lu and U. W. Pooch, "Cooperative Security-Enforcement Routing in Mobile Ad Hoc Networks," *Proceedings of the* 4*th IEEE International Conference on Mobile and Wireless Communications Network*, 2002, pp. 157-161.

[61] D. Sterne, P. Balasubramanyam, D. Carman, B. Wilson, R. Talpade, C. Ko, R. Balupari, C. Y. Tseng, T. Bowen, K. Levitt and J. Rowe, "A General Cooperative Intrusion Detection Architecture for MANETs," *Proceedings of the* 3*rd IEEE International Workshop on Information Assurance*, College Park, MD, USA, 2005, pp. 57-70.

[62] A. Patwardhan, J. Parker, A. Joshi, M. Iorga and T. Karygiannis, "Secure Routing and Intrusion Detection in Ad hoc Networks," *Proceedings of the* 3*rd International Conference on Pervasive Computing and Communications*, Hawaii, 2005, pp. 191-199.

[63] Y. Fu, J. He and G. Li, "A Distributed Intrusion Detection Scheme for Mobile Ad Hoc Networks," *Proceedings of Computer Software and Applications Conference*, 2007, pp. 75-80.

[64] N. Komninos, D. Vergados and C. Douligeris, "Detecting Unau-thorized and Compromised Nodes in Mobile Ad Hoc Networks," *Ad Hoc Networks*, Vol. 5, No. 3, 2007, pp. 289-298.

[65] A. Mitrokotsa, M. Tsagkaris and C. Douligeris, "Intrusion Detection in Mobile Ad Hoc Networks Using Classification Algorithms," *IFIP International Federation for Information Processing*, Palma de Mallorca, 2008, pp. 133-144.

[66] S. Bhargava and D. P. Agrawal, "Security Enhancements in AODV Protocol for Wireless Ad Hoc Networks," *Proceedings of IEEE Vehicular Technology Conference*, Atlantic City, 2001, pp. 2143-2147.

[67] B. Sun, K. Wu and U. Pooch, "Routing Anomaly Detection in Mobile Ad Hoc Networks," *Proceedings of* 12*th International Conference on Computer Communications and Networks*, Dallas, 2003, pp. 20-23.

[68] R. Guha, O. Kachirski, D. G. Schwartz, S. Stoecklin and E. Yilmaz, "Case-Based Agents for Packet-Level Intrusion Detection in Ad Hoc Networks," *Proceedings of* 17*th International Symposium on Computer and Information Sciences*, Florida, 2002, pp. 315-320.

[69] B. Askwith, M. Merabti, Q. Shi and K. Whiteley, "Achieving User Privacy in Mobile Networks," *Proceedings of the* 13*th Annual Computer Security Applications Conference*, San Diego, 1997, pp. 108-116.

[70] D. Chaum, "Security without Identification: Transaction Systems to Make Big Brother Obsolete" *Communications of the ACM*, Vol. 28, No. 10, 1985, pp. 1030-1044.

[71] B. Roland, K. Dogan and R. Peter, "How to Increase Security in Mobile Networks by Anamoly Detection," *Proceedings of the* 14*th Annual Computer Security Applications Conference*, Phoenix,1998, pp. 3-12.

[72] C. H. Lee, M. S. Hwang and W. P. Yang, "Enhanced Privacy and Authentication for the Global System for Mobile Communications," *Wireless Networks*, Vol. 5, No. 4, 1999, pp. 231-243.

[73] P. Bellavista, A. Corrradi and C. Stefanelli, "SOMA Secure and Open Mobile Agent Programming Environment," *Proceedings of the* 4*th International Symposium on the Autonomous Decentralized Systems*, 1999, pp. 238-245.

[74] N. M. Karnik and A. R. Tripathi, "A Security Architecture for Mobile Agents in Ajanta," *Proceedings of the International Conference on Distributed Computing Systems*, Taipei, Taiwan, 2000, pp. 402-409.

[75] J. Zao, S. Kent, J. Gahm, G. Troxel, M. Condell, P. Heliek, N. Yuan and I. Castineyra, "A Public-Key Based Secure Mobile IP Wireless Networks," *Wireless Networks*, Vol. 5, No. 5, 1999, pp. 373-390.

[76] S. K. Y. Lam, "Mobile IP Registration Protocol: A Security Attack and New Secure Minimal Public-Key Based Authentication," *Proceedings of the* 4*th International Symposium on Parallel Architectures, Algorithms and Networks*, Singapore, 1998, pp. 364-369.

[77] G. Necula and P. Lee, "Research on Proof-Carrying Code for Untrusted-Code Security," *Proceedings of IEEE Symposium on Security and Privacy*, 1997, p. 204.

[78] C. E. Perkins, "Mobile IP Joins Forces with AAA," *IEEE Personal Communications*, Vol. 1, No. 4, 2000, pp. 59-61.

[79] T. Hiller *et al.*, "3G Wireless Data Provider Architecture Using Mobile IP and AAA," IETF Internet Draft, 1999.

[80] P. J. McCann and T. Hiller, "An Internet Infrastructure for Cellular CDMA Networks Using Mobile IP," *IEEE Personal Communications*, 2000, pp. 26-30.

[81] D. McClure and B. Bhargava, "On Assigning Priorities of Keying Parameters in a Secure Mobile Network," *Proceedings of IEEE Workshop on Reliable and Secure Application in Mobile Environment*, New Orleans, 2001.

[82] T. G. Brutch and P. C. Brutch, "Mutual Authentication, Confidentiality and Key Management (MACKMAN) System for Mobile Computing and Wireless Communication," *Proceedings of the* 14*th Annual Computer Security Applications Conference*, Phoenix, 1998, pp. 308-317.

[83] L. Gong and N. Shacham, "Multicast Security and its Extension to a Mobile Environment," *Wireless Networks*, Vol. 1, No. 3, 1995, pp. 281-296.

[84] A. D. Joseph and M. F. Kaashoek, "Building Reliable Mobile-Aware Applications Using the Rover Toolkit Wireless Networks," Vol. 3, No. 5, 1997, pp. 405-420.

[85] A. Corradi, R. Montanari and C. Stefanelli, "Mobile Agent Protection in the Internet Environment," *Proceedings of* 23*rd Annual International Computer Software and Applications Conference*, Phoenix, 1999, pp. 20-25.

Wireless Sensor Network, **2010, 2, 441-446**
doi:10.4236/wsn.2010.26055 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Practical Considerations for Wireless Sensor Network Algorithms

Gertjan Halkes, Koen Langendoen
Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
E-mail: {*g.p.halke, k.g.langendoen*}*@tudelft.nl*
Received April 1, 2010; *revised April* 28, 2010; *accepted May* 6, 2010

Abstract

Many researchers from different backgrounds have found interesting research challenges that arise from the physical constraints and envisaged applications in Wireless Sensor Networks (WSNs). The WSN community that has formed over the years is divided into two sub-communities: the systems sub-community and the theory sub-community. However, there seems to be no connection between the two. Algorithms developed from a theoretic perspective are rarely implemented on real hardwares. In this paper we identify the most important reasons why these algorithms are disregarded by the systems sub-community, and provide pointers to remedy the lack of connection.

Keywords: Theory, Practice, Algorithms

1. Introduction

Wireless Sensor Networks (WSNs) exploit the possibilities that miniaturization provide by creating small and cheap devices that can communicate wirelessly and provide a way to bring the real world into the realm of computing. Using WSNs many new applications come within reach, for example monitoring of real-world events in remote and poorly accessible places [1,2].

The distributed nature of WSNs and research challenges arising from the constraints dictated by economics and physics have attracted many researchers from different backgrounds such as distributed systems, networking and signal processing. These different backgrounds also have their impact on the WSN community. The WSN community that has formed over the years is divided into two sub-communities: the systems sub-community and the theory sub-community. However, there seems to be no connection between the two sub-communities. Algorithms developed from a theory perspective are rarely implemented on real hardware. In this context we are reminded of the following quote that summarizes the essence of our observations:

In theory, there is no difference between theory and practice. But, in practice, there is.

Jan L. A. van de Snepscheut

With this paper we intend to make a start at removing the difference between theory and practice.

We identify the most important reasons why algorithms created from a theory perspective are disregarded by the systems community. By studying the papers from major theory and systems conferences we conclude that there are three main issues that make implementation of many algorithms developed from a theory perspective infeasible or undesirable. Firstly, the unreliability of the underlying wireless network is often ignored. Secondly, energy consumption is not always taken into account when designing and evaluating an algorithm. Lastly, algorithms are sometimes organised in rounds which hampers implementation on real hardware. We also provide recommendations on how to design algorithms such that the systems community will more often use them.

The rest of the paper is organised as follows: in Section 2 we give an overview of the different perspectives on WSNs that influence the design of algorithms and protocols from the two sub-communities. Then, in Section 3 we provide the results of our study of a representative collection of papers from WSN conferences. In Sections 4 through 6 we analyse the main issues we identified in greater detail and provide recommendations for each of them. Finally, in Section 7 we conclude the paper.

2. Main Perspectives on Wireless Sensor Networks

To understand the origin of the gap between the theory and systems sub-communities, we start by looking at the

characteristics that the sub-communities use when describing WSNs.

From the theory perspective a WSN is very much like a classical distributed system. What sets WSNs apart is that the network connectivity is dictated by physical proximity instead of having a fully connected network. Of course the subject of the algorithms is different as well because of the different application areas. Algorithms for WSNs mostly focus on subjects like localization, information dissemination, distributed calculation of statistics and metrics on measured data, and distributed consensus. Naturally this is not an exhaustive list of all the characteristics taken into account by researchers within the theory sub-community. However, these are the common characteristics that can be found in virtually all research papers, either explicitly stated or left implicit.

The systems perspective is dominated by the harshness of reality and the laws of physics. An important characteristic is that WSN nodes are running of batteries or ambient energy sources. Furthermore, it is intended that sensor networks can be left to operate for years without any human intervention. This means that energy is a scarce resource and energy efficiency is key in designing protocols and algorithms. As the radio is and will remain the largest consumer of energy in the node it is important to communicate only when necessary. A second important characteristic is that wireless communication is inherently unreliable and unstable. However unfortunate, this is a result of the laws of physics and therefore has to be dealt with. This does mean that every protocol and algorithm needs to be designed to cope with unreliable communication.

3. Problem Analysis

To assess the impact of the two different perspectives on WSNs, we analyzed papers from major theory sub-community conferences and from important systems conferences for comparison. From this analysis we compiled a list of issues that complicate the implementation of the presented algorithms in the real world. Below is an extract from the list we compiled:

- Communication is assumed to be reliable.
- Energy consumption in the form of communication is not taken into account, or not analyzed precisely enough.
- Algorithms are organized in synchronous rounds.
- Each node has a known and stable set of neighbors.
- The propagation model is assumed to be a Unit Disk Graph (UDG) and calculations are made based on this assumption.
- Algorithms include manipulations of large matrices, which is infeasible on sensor node processors.
- Large messages are exchanged between nodes, which require several packets in WSNs.

Table 1. Percentages of papers in studied theory and systems conferences per implementation issue[a].

	Theory	Systems
Communication assumed reliable	47 (61%)	8 (17%)
Energy cost insufficiently accounted	47 (61%)	14 (30%)
Rounds-based organisation	19 (25%)	1 (2%)
Number of papers	77	47

([a]The papers were taken from the IPSN 2005-2007 and DCOSS 2005-2006 conferences (theory), and from the SenSys 2005-2006 and EWSN 2006-2007 conferences (systems).)

Some of these issues can be dealt with without touching the workings of an algorithm, while other issues touch on fundamental assumptions underlying the design of that algorithm. The first three items of the above list fall into the latter category and those are the issues we focus on in this paper.

In **Table 1** we have listed the occurrences of the three issues in the theory and systems conference proceedings we studied. It should be noted that the separation of theory and systems papers does not exactly follow the conference foci. In theory conferences there are usually a few systems-type papers, and vice versa. Although this cross pollination somewhat dilutes the numbers, it is clear that the issues complicating implementation are far more prevalent in theory papers than in systems papers.

Adapting algorithms that have been designed without considering these issues can be done, but is exceedingly difficult [3]. Therefore, we now provide a more detailed description of the issues and recommendations on how they can be avoided.

4. Communication Reliability

To implement any algorithm that involves more than a single node, nodes will have to communicate. At the lowest layer the only primitive that is available is a local broadcast to a node's "neighbors". That is, the radio can be used to send bits to other nodes that are sufficiently close by to receive them.

However, physical proximity is far from the only factor that determines whether the bits will actually reach the intended recipients. Obstacles of many kinds can distort and reflect the signals and prevent proper reception (see **Figure 1**). Examples of such objects are of course walls and other built structures, trees and plants, but also humans, animals and vehicles. Note that the obstacles may be mobile, which will make for changing signal conditions and therefore changing channel reliability. So, even though two nodes may have no trouble

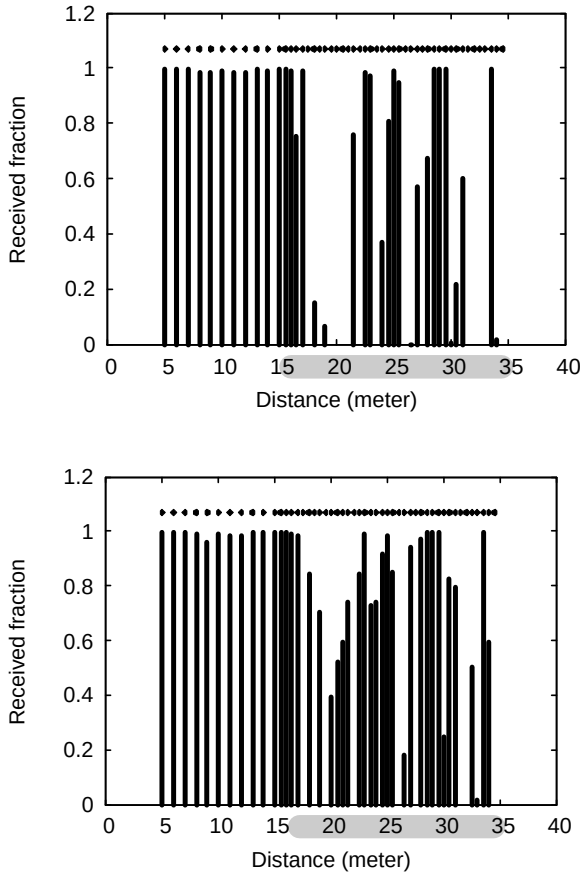


Figure 1. Reception rates in an office corridor at different distances from a sending node on two consecutive days (taken from [5]).

communicating one moment, communication could be completely impossible the next moment.

Other factors that influence reception include temperature, humidity, and other transmitters, all of which vary over time (see **Figure 1** for an example). It is clear that all these factors taken together make it impossible to predict whether a given signal will reach the intended recipient. Indeed, several experimental studies have shown the unreliability of the channel [4-7].

Given the unreliable channel, one of the MAC protocol's tasks is to increase the reliability of the channel as a means of communication. Paradoxically, although the radio channel is a broadcast medium it is easier to increase the reliability of unicast radio transmissions than it is to increase the reliability of broadcast transmissions (see below). But even for unicast communication it is not possible to achieve 100% reliability. This means that protocols and algorithms have to take into account that any communication step may fail, especially when using broadcast messages. Also, it is worth noting that any increase in reliability comes at the cost of increased energy consumption, for example in the form of retransmissions.

4.1. Increasing Reliability

Although 100% reliability is impossible to achieve, there are possibilities to increase the reliability of the links in use. Topology control algorithms are a good example [3,4,6]. By only using the links that meet some quality criterion it is possible to remove much of the unreliability. However, link instability caused by quickly changing channel conditions such as collisions, interference from other devices and moving obstacles can not be eliminated this way.

An important thing to realize is that topology control is not free. A topology control algorithm needs to assess the quality of links, and the only way to do that is to observe messages sent over the link. In many cases the algorithm will send its own messages to do so. These costs need to be taken into account when assuming the presence of a topology control algorithm.

Almost all of the papers that assume reliable communication use broadcast messages in the algorithm. However, increasing the reliability of broadcast communications is a very difficult problem. First of all, in order to determine whether a message arrived, a node needs to know all the nodes a message is to arrive at. Therefore, each node has to know its neighbors. Again a topology control algorithm can provide this information; at a price. Second, all nodes that have received a message need to acknowledge the reception. Each one of these messages costs energy, and because all receiving nodes will want to acknowledge the message at the same time channel contention will be high. Third, if not all nodes have received the broadcast, some form of retransmission scheme has to be employed, which is not trivial and will also cost even more energy.

Another option for increasing the reliability of broadcasts is to send a unicast message to each neighbor separately (repeated send). This is of course a very costly solution, which again requires knowledge of a node's neighbors. Furthermore, if a node has many neighbors it will take a long time to reach all of them.

4.2. Handling Unreliability

Recommendation 1: *Design algorithms such that unreliable communication is not disruptive.*

From the previous discussion it is clear that 100% reliable communication is not achievable. **Table 2** summarizes what types of communication are available in WSNs. Many research papers from the theory subcommunity assume, either explicitly or implicitly, that reliable communication is provided by the communication subsystem. Worse, most of these papers assume re-

Table 2. Available communication types in WSNs.

	Reliable	Unreliable
Unicast	+/–[a]	+
Broadcast	–	+

([a]Reliability vs. energy trade-off.)

liable broadcast. Clearly, this mismatch between assumed and provided level of service makes straightforward implementation of the algorithms impossible.

Recommendation 2: *Analyze the impact of unreliable communication on the algorithm's performance.*

Although it may be difficult to take unreliable communication into account from the start of algorithm design, an analysis of the algorithm's performance with message loss is essential. This can include an analysis of performance degradation, or an analysis of the increased cost to obtain the same quality of result. Before any implementation on real hardware it is vital to know if and how well the algorithm will be able to cope with message loss.

5. Energy Efficiency

From a systems perspective an important aspect of WSNs is the limited amount of energy. As the radio is the most important source of energy consumption, a lot of research is focused on limiting the time the radio is turned on [8,9]. Part of this problem is solved by the MAC layer orchestrating the communication in such a way that the radio can be turned off most of the time. However, this only reduces the overhead associated with radio communication. The sending of messages itself still costs energy. It is up to the higher layers to limit the number of messages sent as much as possible.

The focus on energy efficiency is not always found in algorithm papers. Five categories of papers can be distinguished in this context, ordered by level of detail:

- papers that ignore communication cost all together (Theory: 39% vs. Systems: 26%),
- papers that provide an order estimate of the number of messages sent (22% vs. 4%),
- papers that provide an analysis of the number of messages sent (12% vs. 6%),
- papers that provide simulation or real-implementation results on the number of messages sent (17% vs. 38%), and
- papers that provide energy consumption figures from a simulator or a real implementation (10% vs. 26%).

For compiling **Table 1** we have considered the first two categories insufficient for accounting energy consumption.

At first sight the last category may seem the most desirable, but this is deceptive. The exact energy consumption resulting from the use of an algorithm is very dependant on the underlying MAC protocol and radio hardware. WSN-specific MAC-protocols are highly optimized for a particular scenario and the associated traffic pattern. Using a different MAC protocol can easily increase or reduce the energy consumption with a significant factor. For example, for data gathering traffic an optimized MAC protocol like DMAC [10] uses significantly less energy than most general-purpose MAC protocols. Measuring the energy consumption in Joules is however useful for low-level protocols.

If we want to find the best way to analyze energy efficiency, we first have to look at what we want to do with this information. In the end, energy efficiency is simply a metric to compare and rank different algorithms. To compare different algorithms with respect to energy efficiency, what is required is a precise analysis of the number of messages sent.

Recommendation 3: *Specify an algorithm's energy efficiency by analyzing the number of messages sent, differentiating between unicast and broadcast messages.*

As mentioned previously, different MAC protocols have different energy consumption profiles. The most important differences stem from the distinction between broadcast and unicast messages. Some protocols are optimized for energy-efficient unicast traffic and a broadcast message can consume as much energy as several unicast messages, e.g., WiseMAC [11]. Other protocols are more geared towards broadcast, which makes unicast approximately as expensive as broadcast, e.g., B-MAC [12]. **Figure 2** shows the energy consumption profiles for several state of the art MAC protocols. Note that the graphs do not indicate energy efficiency, only the ratio between unicast and broadcast energy consumption.

Another reason to separate unicast messages from broadcast messages is that many MAC protocols include retransmissions for unacknowledged unicast messages, while broadcast messages are only sent once. These MAC level retransmission schemes make it even harder to compare the cost of unicast messages with the cost of broadcast messages.

These considerations lead to the conclusion that a thorough analysis of the communication cost should separate broadcast and unicast messages.

6. Algorithm Organization

Even when an algorithm designer has taken message loss and energy consumption into account in the design of his algorithm, it may still be next to impossible to implement the algorithm on real hardware. The single most common cause is that the algorithm is designed based on the communication-round paradigm. There are synchronous and

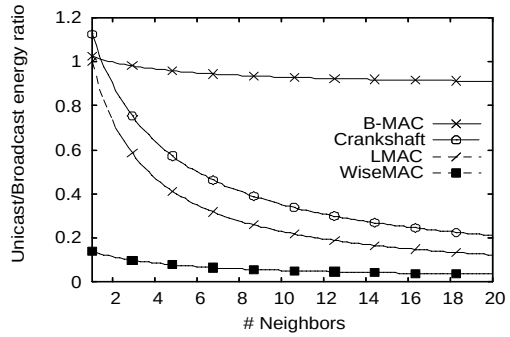


Figure 2. Energy consumption profiles for different MAC protocols, expressed as the ratio between the energy spent for sending one unicast message and the energy spent for one broadcast message. The use of overhearing-avoidance techniques makes unicast messages cheaper than broadcast messages when there are many neighbors.

asynchronous versions of this paradigm, but both are problematic in the context of WSNs. In this section we will detail the problems with both versions.

6.1. Synchronous Communication Rounds

The use of synchronous communication rounds is very common in classical distributed systems. Each calculation and communication round is executed by each node simultaneously. This synchrony is easily achievable on stable and reliable networks such as the Internet and local networks of grid computers. However, implementing an algorithm based on synchronous communication rounds on WSNs is very difficult for several reasons.

First of all, an algorithm has to be started. This may seem like a trivial operation, but certainly is not. To start an algorithm, all nodes will have to agree to start at the same time. Reaching agreement can only be done through communication, which, as we have seen in Section 4, is unreliable. Hence, starting a synchronous algorithm can not be done reliably. The same problem arises in establishing when a round has ended. Using a timer will not help either because it is impossible to bound the time required to finish the communication for a single round.

A second problem in using an algorithm based on communication rounds is that each round will cause a peak load in the network. This peak load will decrease reliability of network communication and disrupt traffic from concurrently running algorithms and applications.

A final problem is what to do when nodes join the network. Although for many of the algorithms this simply means that a node will only be able to participate in the next run (not round) of the algorithm. For algorithms that manage the network structure such as clustering algorithms or topology control algorithms it may mean that a node will not be able to participate in the network for a long time.

6.2. Asynchronous Communication Rounds

The communication-rounds paradigm can also be implemented asynchronously. This can be done by including the round number in each message. Once a node has received all messages from its neighbors, it can proceed to the next round. To start the algorithm a single, perhaps designated, node can simply start its first round. When a node receives a message that indicates that it belongs to the first round, the receiving node will also send its message for the first round. However, implementing this asynchronous approach is not without problems in WSNs.

The most important issue is that a node will have to know all its neighbors to be able to determine when to start the next round. Furthermore, the neighbor set is not necessarily stable. As a result, the algorithm may never terminate because a node will not receive messages from all the nodes it expects messages from. Instability of the neighbor set may be caused by communication unreliability, node failure and node mobility.

6.3. Reactive Organization

Recommendation 4: *Design algorithms that react to other nodes' messages, rather than using the concept of rounds.*

Given the problems in implementing a rounds-based algorithm detailed above, it is obvious that a different organization is preferable. One simple organization that does not suffer from the round starting problem is an organization where a node simply reacts asynchronously to messages from neighboring nodes. When a round should be started, a single node can simply decide that a new run of the algorithm is required and send the first message.

This organization does introduce several new problems for naive implementations, but these are easily solved. For example, it is not a good idea to react to a received message by immediately sending a message as well. This will make for high contention conditions during the execution of an algorithm, which needs to be avoided. Adding a random delay before sending a response to a neighbor's message will also allow reacting to other neighbors with a single message. To prevent deadlock and starvation, the delay should not be reset on receiving another message.

A second problem is the termination of an algorithm. The number of executed rounds can no longer be used to terminate an algorithm. Convergence is a good option for termination in asynchronous algorithms, and is actually also often used in rounds-based algorithms. Another option is to bound the maximum number of messages sent for one algorithm run.

It is important to keep in mind that communication is

not 100% reliable. Although in synchronous round-based algorithms this is necessarily a problem, this need not be a problem in asynchronous algorithms because further iterations of an algorithm should compensate for the missing values.

We are aware that an asynchronous algorithm is harder to analyze with respect to both the number of messages sent per algorithm run, as well as convergence and stability properties. This is the price to be paid for creating an implementable algorithm.

7. Conclusions

In this paper we have identified three common causes for the lack of interest from the WSN systems sub-community for the algorithms developed by the theory sub-community. Firstly, the unreliability of the underlying network is often ignored. Secondly, energy consumption of the algorithm is not always taken into account when designing and evaluating the algorithm. Lastly, algorithms are sometimes organized in rounds which hamper implementation on real hardware.

To close the gap between the theory and systems sub-communities, we provide the following recommendations: 1) Design algorithms such that unreliable communication is not disruptive. 2) Analyze the impact of unreliable communication on the algorithm's performance. 3) Specify an algorithm's energy efficiency by analyzing the number of messages sent, differentiating between unicast and broadcast messages. 4) Design algorithms that react to other nodes' messages, rather than using the concept of rounds.

Although we realize it is sometimes more feasible to analyze algorithms in an abstracted environment, this does mean that the results are not directly applicable to real-life sensor-networks. The results for the abstracted environment can be used as a first step to a complete working algorithm, but unfortunately the step to reality is often overlooked.

8. Acknowledgements

We would like to thank Andreas Meier for his comments and proof-reading.

9. References

[1] A. Mainwaring, J. Polastre, R. Szewcyk, D. Culler and J. Anderson, "Wireless Sensor Networks For Habitat Monitoring," *Proceedings of* 1*st ACM International Workshop on Wireless Sensor Networks and Application* (*WSNA*), Atlanta, September 2002, pp. 88-97.

[2] K. Martinez, J. K. Hart and R. Ong, "Environmental sensor networks," *IEEE Computer*, Vol. 37, No. 8, 2004, pp. 50-56.

[3] M. Dyer, J. Beutel and L. Thiele, "S-XTC: A Signal-Strenght Based Topology Control Algorithm for Sensor Networks," *Proceedings of* 8*th International Symposium on Autonomous Decentralized Systems* (*ISADS*'07), Sedona, March 2007, pp. 508-518.

[4] S. Lin, J. Zhang, G. Zhou, L. Gu, T. He and J. A. Stankovic, "ATPC: Adaptive Transmission Power Control for Wireless Sensor Networks," *Proceedings of* 4*th ACM Conference on Embedded Networked Sensor Systems* (*SenSys*'06), Boulder, November 2006, pp. 223-236.

[5] N. Reijers, G. Halkes and K. Langendoen, "Link Layer Measurements in Sensor Networks," *Proceedings of* 1*st IEEE Conference on Mobile Ad-Hoc and Sensor Systems (MASS*'04*)*, Fort Lauderdale, October 2004, pp. 24-27.

[6] A. Woo, T. Tong and D. Culler, "Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks," *Proceedings of* 1*st ACM Conference on Embedded Networked Sensor Systems* (*SenSys*'03), Los Angeles, November 2003, pp. 14-27.

[7] J. Zhao and R. Govindan, "Understanding Packet Delivery Performance in Dense Wireless Sensor Networks," *Proceedings of* 1*st ACM Conference on Embedded Networked Sensor Systems* (*SenSys*'03), Los Angeles, November 2003, pp. 1-13.

[8] A. Cerpa and D. Estrin, "ASCENT: Adaptive Self-Configuring Sensor Networks Topologies," *IEEE Transactions on Mobile Computing*, Vol. 3, No. 3, 2004, pp. 272-285.

[9] K. Langendoen and G. Halkes, "Energy-Efficient Medium Access Control," In: R. Zurawski, Ed., *Embedded Systems Handbook*, CRC Press, 2005, pp. 1-34.

[10] G. Lu, B. Krishnamachari and C. Raghavendra, "An Adaptive Energy-Efficient and Low-Latency MAC for Data Gathering in Sensor Networks," *International Workshop on Algorithms for Wireless*, *Mobile*, *Ad Hoc and Sensor Networks* (*WMAN*), Santa Fe, April 2004, pp. 224-231.

[11] A. El-Hoiydi and J.-D. Decotignie, "WiseMAC: An Ultra Low Power MAC Protocol for Multi-Hop Wireless Sensor Networks," *Proceedings of* 1*st International Workshop on Algorithmic Aspects of Wireless Sensor Networks* (*ALGOSENSORS*'04), Turku, July 2004, pp. 18-31.

[12] J. Polastre, J. Hill and D. Culler, "Versatile Low Power Media Access for Wireless Sensor Networks," *Proceedings of* 2*nd ACM Conference on Embedded Networked Sensor Systems* (*SensSys*'04), Baltimore, November 2004, pp. 95-107.

Wireless Sensor Network, **2010, 2, 447-461**
doi:10.4236/wsn.2010.26056 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Web Services Invocation over Bluetooth

Auletta Vincenzo, Blundo Carlo, De Cristofaro Emiliano, Raimato Guerriero
Dipartimento di Informatica ed Applicazioni, Università di Salerno, Salerno, Italy
E-mail: {*auletta, carblu, emidec, raimato*}*@dia.unisa.it*
Received December 17, 2009; *revised January* 20, 2010; *accepted January* 26, 2010

Abstract

Over the last years, technology evolution is leading the way towards autonomous, ubiquitous and widespread interactions among small computing devices. To this aim, communication technologies that support dynamicity and mobility and work on inexpensive small devices have attracted much attention. The Bluetooth specification particularly fits this idea, providing a free, versatile, and flexible wireless network technology with low power consumption. On the other hand, as the degree of penetration of computational services has increased in everyday life, users' habits have deeply changed, resulting into an increasing request for mobile and ubiquitous services. In a few years, most of the devices accessing the web services will be mobile. Therefore, we need solutions that encompass networking and application issues involved in realizing mobile and ubiquitous access to the services. In this paper, we analyze how Bluetooth can be used to design, develop, and deploy Web Services-based applications that run on mobile devices. We propose and evaluate a framework that allows the interaction with Web Services from mobile devices using Bluetooth as communication channel.

Keywords: Bluetooth, Web Service, Communication Channel

1. Introduction

The technology evolution. Advances in communication technologies drove a deep transformation of users habits, in particular with an increasingly requirement of support to mobility and connectivity. Up to a few years ago mobile devices were very simple and resource limited. As a result, applications produced for these devices were bounded to the device environment. Nowadays, mobile devices such as smartphones or PDAs have enhanced their range of action, turning into fundamental working instruments. Modern applications, however, require connectivity and thus a critical issue for the penetration of mobile devices is the capacity to run network applications, especially Web applications. In the last years, several new protocols have been presented for wireless communications, such as IRDA, WLAN, and GPRS/UMTS. However, IRDA connections are limited to two devices with a direct line of sight, and thus IRDA is not practically useful for a real intercommunication scheme. WLAN instead has been designed as a powerful technology to support multipoint connections, but penetration of WLAN on mobile devices and particularly on smartphones is still low. GPRS/UMTS are widely supported but they provide connectivity at modest speed and require a personal account with a phone company. At the same time, we witnessed the growth of Bluetooth [1], that is a low-cost, robust, powerful, and flexible short-range wireless link layer technology with low power consumption. It operates in a license-free frequency range, so that user is not charged for accessing the network nor needs an account with any company, thus allowing a relevant decrease of communication costs. Nowadays, the evolution of Bluetooth technology is driven by the Bluetooth SIG, that consists of over 7000 member companies that guarantee a large support to this technology. In fact, Bluetooth technology is used in many wide-spread different devices, such as handhelds, mobiles, smart-phones, laptops, PDAs. A thorough overview on Bluetooth is given in [2] and [3].

Recent work. Lately, a research study [4] has forecasted that, within a few years, most of the devices accessing the Web and Web Services will be mobile and presumably most of them will be Bluetooth-enabled. Therefore, we need solutions that encompass networking and application issues involved in realizing mobile and ubiquitous access to the services. Several research groups are proposing frameworks for developing applications over Bluetooth-based networks (for instance, [5] and [6]) and evaluate the possibility of using this technology for building ad-hoc networks suitable for dedicated applications, such as voice transmission [7], audio streaming [8], context-aware applications [9], and Internet access point [10].

Research goals. In this paper, we analyze how Bluetooth can be used to design, develop, and deploy Web Services-based applications that run on mobile devices. In fact, we propose and evaluate a solution that allows the interaction with Web Services from mobile devices using Bluetooth as communication channel. We consider a scenario where mobile devices consume Web Services but do not offer them. Some preliminary results on the proposed solution can be found in [11] and [12]. Also, in [13], the authors address the problem of deploying Web Services on mobile devices, providing a solution that however relies on the expensive Bluetooth's PAN profile, which is available only on PDAs and requires a preliminary pairing of the devices.

Our solution instead relies on the standard and inexpensive JSR-82 API [14] and on a tunneling mechanism realized by an intermediate software layer to incapsulate HTTP packets into Bluetooth ones. In this way, the connections are state-less and without any preliminary pairing (details are given in Section 5).

Main Contributions. Our work achieves a twofold goal. First, our solution provides ubiquitous Bluetooth-based access to Web Services and it is completely transparent to both users and application programmers. Also, our solution is to be widely supported at no extra cost by mobile devices. To this aim, we have devoted our attention to the free Bluetooth technology as opposed to other wireless technologies.

Typical applications that we have in mind are: information retrieving (e.g., accessing train timetables in a station) or micropayment applications (e.g., buying ticket in a cinema or on a bus), but our solution puts no restrictiton on what one user can require. We designed a framework in such a way that, for a programmer, it will be very simple to port a Web Services-client application from a desktop environment (Axis Client API-based) to a mobile Bluetooth-enabled device. Indeed, we developed a J2ME package (named wsbt) exposing the entire Web-service stack to the client devices. For the programmer, it will be enough to change the package to import from org.apache.axis.client to wsbt.

Windows and Linux Implementation. In this work, we present two different implementations of our solution. We can summarize differences between the two implementations as follows: The first one is Windows-based, works on top of a third-party implementation of the Java API for using Bluetooth connections, and operates at a high level; the second one is Linux-based and works on top of our implementation of a Java package for exploiting Bluetooth features giving to the programmer control over several low level parameters of the Bluetooth channel. Several motivations suggested us to provide both implementations. First of all we would like to have our software available on both Windows and Linux, but, to our knowledge, there is no high level implementation of the Java API for Bluetooth for Linux, and, on the other hand, we have no direct access to the Bluetooth stack in Windows. The second motivation is to evaluate whether the low level control of the Bluetooth channel makes the programmers able to tune the communication parameters in order to significantly improve communication performances, such as latency and throughput.

Hence, we remark that the implementation of the Java API for Bluetooth in the Linux environment (that we named JBlueZen) is a programming effort of independently interest.

Efficiency. Our performance evaluations confirm the real applicability and lightness of the framework showing that Bluetooth is well suited to be the transport layer for Web Services accessing from wireless devices. It is worthwhile to notice that our proposed framework has a small footprint. Indeed, the Java code to be put onto the client, to get a web services client application running, needs just 50 KB of memory (including any external library).

Paper Organization. The rest of this paper is structured as follows. In Section 2, we highlight some general concepts about the pertinent technologies, such as Bluetooth, J2ME, and SOAP. In Section 3, we present an overview of our solution, and in Section 4, we show our design choices. The framework implementing our solution is presented and described in Section 5. In Section 6, we present the implementation of the framework client-side; while, in Section 7, two different implementations of the server-side are illustrated. Finally, in Section 8 we present and comment the results of our performance evaluations of the two proposed solutions.

2. Endorsed Technologies

The goal of this paper is to describe the design of a framework that allows Java programmers to easily and directly invoke Web Services from mobile devices over a Bluetooth connection. Hence, the basis for our work are Java 2 Micro Edition (J2ME) [15], the Standard Bluetooth.

Technology [16], and SOAP [17]. J2ME describes how to write Java applications on mobile devices and defines details for the communication between devices. Bluetooth is a low-cost, flexible, robust short-range wireless networking technology with low-power consumption. SOAP is a protocol for exchanging XML-based messages over computer networks. In this section we describe all the technologies that will be used in our framework.

2.1. The Bluetooth Wireless Technology

The Bluetooth specification was introduced in 1994 by Ericsson to provide radio communications between mobile phones, headsets and keyboards. The specifications were then released by the Bluetooth Special Interest Group (SIG) [16] in September 1999. Within this tech-

nology, radio communications can take place by mean of integrated and cheap devices with small energy consumption. This technology achieves its goal by embedding tiny, inexpensive, short-range transceivers into electronic devices that are available today. Bluetooth devices operate in a license-free frequency range (starting from 2.4 GHz).

Bluetooth-enabled devices can dynamically *discove*r other devices in their range and their supported services, through an inquiry process.

An overview of the Bluetooth stack is presented in **Figure 1**. The *radi*o level is the lowest one and it defines the technical details of the communication.

The *baseban*d layer handles channels and physical links, providing services such as error correction and security. It supports multipoint communications through FH/TDMA (Frequency Hopping/Time Division Multiple Access). The master device is in charge of defining the hopping sequence to all the slave devices. A physical channel is shared between the master and a slave using a time division scheme in which data are transmitted in one direction at time, with transmissions alternating between the two directions.

Up in the stack we find: the *Lin*k *Managemen*t *Protoco*l (LMP) handling link setup, authentication, and link configuration; the *Hos*t *Controlle*r *Interfac*e (HCI) which provides a uniform method of accessing the Bluetooth baseband capabilities; the *Logica*l *Lin*k *Contro*l *an*d *Adaptatio*n *Protoco*l (L2CAP) which deals with data multiplexing and segmentation. Finally, on top of L2CAP, we find several data communication protocols. The main protocols are:

1) SDP (Service Discovery Protocol), which handles the discovery of devices and services within the device's transmission range.

2) RFCOMM, which implements emulation of serial connections, setting up point-to-point connections. It supports framing and multiplexing and achieves all the required functions for serial data exchange.

3) OBEX (Object Exchange), which is built on the top of RFCOMM to implement exchange of objects, such as files and vCards. Originally, it was developed by IrDA (Infrared Data Association) for IR-enabled devices.

4) TCS (Telephony Control protocol Specification), which defines ways to send audio calls between Bluetooth devices.

The Bluetooth technology is also composed by a set of profiles. Bluetooth profiles describe several scenarios where Bluetooth technology is responsible of transmission. Each scenario is described by a user model and the corresponding profile gives a standard interface that applications can use to interact with the Bluetooth protocols. The profile concept is used to decrease the risk of interoperability problems between different manufacturers' products.

In order to interface applications to the physical layer,

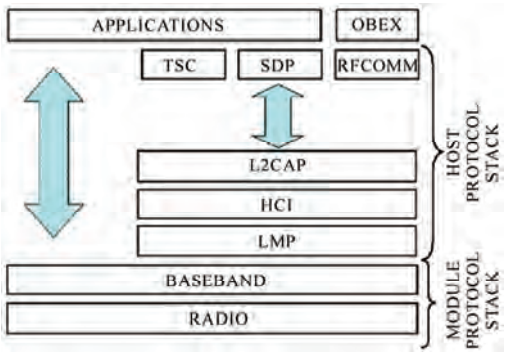


Figure 1. The bluetooth stack.

a Bluetooth Stack implementation is necessary. The stack provides a standard interface between the application layer and the Bluetooth specification. This interface is used to overcome the compatibility problems between the application and different Bluetooth devices. Indeed, Bluetooth stacks are responsible of implementing the Bluetooth wireless specifications. There are several different stacks targeted to different devices, applications, and operating systems. Currently available Bluetooth stack implementations are:

1) Mobile devices vendors' embedded stacks. Vendors providing Bluetooth-enabled devices have to build their own Bluetooth stack; for smartphones stack implementations obviously depend on the OS (e.g., Symbian).

2) Broadcom BTW (not free) [18]. It is addressed to PC OEMs and accessory manufactures to quickly and easily add Bluetooth technology to desktop PC and notebooks running Windows.

3) Microsoft BT Stack [19]. It is the Microsoft version of the Bluetooth stack and it is embedded in Windows XP SP 2. It provides the support for most of Bluetooth profiles, essentially the ones based on the RFCOMM protocol.

4) BlueZ (free and open-source) [20]. It is the Linux Bluetooth Stack. The code is licensed under the GNU General Public License and is included in the Linux 2.4 and Linux 2.6 kernel series. It provides a direct access to the transmission layer and allows developers to set several parameters of the communication.

2.2. J2ME

The J2ME (Java Platform Micro Edition) is a collection of Java APIs for developing applications targeted to resource-constrained devices such as PDAs and smartphones. Formally, J2ME is an abstract specification, but the term is frequently used also to refer to runtime implementations. The advantages of using Java as programming language are code portability and an increase of mobile devices flexibility. In particular, it provides support for deploying dedicated applications, named MIDlets, on the mobile device. They allow programmers to increase

available features and capabilities of mobile devices. Since the range of micro devices is so diversified and wide, J2ME was designed as a collection of configurations, where each configuration is tailored to a class of devices. Each configuration consists of a Java Virtual Machine and a collection of classes that provide a programming environment for the applications. Configurations are completed by profiles, which add classes to provide additional features suitable to a particular set of devices. J2ME defines two configurations: the *Connected Device Configuration* (CDC) [21] and the *Connected Limited Device Configuration* (CLDC) [22].

CDC is addressed to small, resource-constrained devices such as TV set-top boxes, auto telematics. It can add a graphical user interface and other functionalities; CLDC, instead, is addressed to devices with limited memory capacity. In this paper, we restrict our attention to the CLDC configuration. CLDC is a low level specification that includes a set of APIs providing basic features for resource-constrained devices, such as smartphones and PDAs. Producers should add features to CLDC by providing new libraries and thus creating a Profile. The first profile proposed for CLDC was the MIDP (Mobile Information Device Profile) [23]. MIDP is a set of Java libraries that permits to create an application environment for mobile devices with limited resources. Here, limitations include: amount of available memory, computational power, network communications with strong latency, and low bandwidth. MIDP 1.0 specification was produced by MIDPEG (MIDP Expert Group), as part of the JSR-37 [24] standardization effort; while, the MIDP 2.0 specification was released with the JSR-118 [25] standardization effort. MIDP 2.0 devices have to meet the following requirements:

1) *Memory*, 250 KB of non volatile memory for MIDP components, 8 KB for user data.

2) *Display*, 96 × 54 resolution, 1-bit color depth, 1:1 aspect ratio.

3) *Networking, bidirectional and wireless communication, limited bandwidth.*

2.3. JSR-82

Although the synergy between MIDP and J2ME technologies supplies a large number of communication schemes, it does not provide support for the Bluetooth technology. Therefore, the Java Expert Group JSR-82 [14] introduced the *Java APIs for Bluetooth Wireless Technology* (JABWT) that provides a standard and high-level support for handling Bluetooth communications in Java applications. These APIs operate on top of CLDC to extend MIDP functionalities. Their development is still in progress, but about twenty mobile vendors have adopted them in their devices. The last released version (Version 1.1) provides support for:

1) Data transmission on the Bluetooth channel (audio and video are not supported).

2) Protocols: L2CAP, RFCOMM, SDP, OBEX.

3) Profiles: GAP, SDAP, SPP, GOEP

The Generic Access Profile (GAP) defines the generic procedures related to discovery of Bluetooth devices and link management aspects of connecting to Bluetooth devices. The Service Discovery Application Profile (SDAP) defines the features and procedures for an application in a Bluetooth device to discover services registered in other Bluetooth devices and retrieve any desired available information pertinent to these services. The Serial Port Profile (SPP) defines the requirements for Bluetooth devices' necessary for setting up emulated serial cable connections using RFCOMM between two peer devices. The Generic Object Exchange Profile (GOEP) defines the requirements for Bluetooth devices necessary for the support of the object exchange usage models.

The interaction between the J2ME environment and the Bluetooth API is shown in **Figure 2**. Using JABWT, it is possible to interact with the Bluetooth stack in a Java application. In particular, it is possible to call services such as device and service discovery, establishment of RFCOMM, L2CAP, and OBEX connections.

In order to use the Java APIs for Bluetooth, a real implementation of the JSR-82 specification is necessary on the device. The current JSR-82 implementations are:

1) Mobile devices vendors' embedded JSR-82 implementations.

2) Atinav aveLink suite (not free) [26]. It offers both an implementation of the Bluetooth stack and the implementation of all the standard profiles for ANSI C, JSR-82 for J2SE Java, JSR-82 for J2ME, Windows and Windows CE.

3) Impronto Rococo (not free) [27]. It is a complete product that provides the Bluetooth Stack and the integration layer, the JVM and the JSR-82 implementation layer both for J2SE and J2ME.

4) Avetana (not free) [28]. It enables writing J2SE applications to access the Bluetooth layer; it is available for Windows, MacOS X, and Linux platforms.

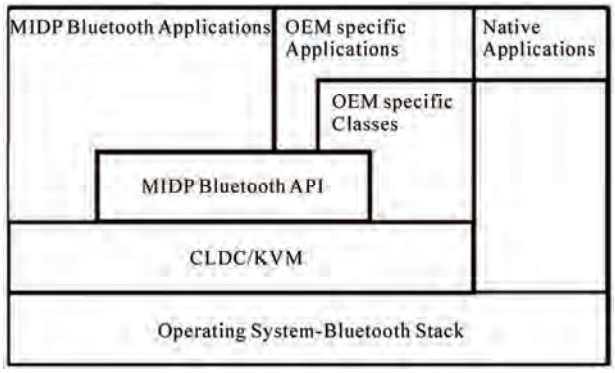


Figure 2. J2ME-Bluetooth API interaction architecture.

5) BlueCove (free) [29]. It provides the Java JSR-82 support for J2SE applications over the Windows XP SP2 Bluetooth stack.

2.4. SOAP

SOAP is a lightweight protocol for exchanging information in a distributed environment. It is an XML based protocol consisting of three parts: an envelope that defines a framework for describing what is in a message and how to process it, a set of encoding rules for ex-pressing instances of application-defined datatypes, and a convention for representing remote procedure calls and responses [30].

In our framework the client application runs on a mobile device; it is implemented as a MIDlet and communicates with the Web Services Container through SOAP messages. Thus, we need a J2ME-compatible library that can be used in a MIDlet to serialize/deserialize messages according to the SOAP specification. Our framework uses the kSOAP library [31] to serialize/deserialize the messages on the client side. The kSOAP library provides highlevel classes to construct the envelope, the body, and the header of the SOAP message and to specify their elements. Such a library supports a subset of SOAP 1.1 features. Indeed, the library supports only simple data types such as strings, integers, etc. However, it is possible to extend these basic functionalities so that any complex datatype can be included in the SOAP message. One has to simply provide custom classes for these data types. Such classes have to implement the Java KvmSerializable interface.

3. Architecture Overview

In this section we will present a lightweight framework that allows an application programmer to design client applications running on mobile Bluetooth-enabled devices that invoke remote Web Services. We assume a client-server interaction, where client and server communicate over a Bluetooth channel using SOAP [17] as messaging system. We refer to the common scenario shown in **Figure 3**.

We refer to the usual architecture, where a client exchanges SOAP messages with a server using HTTP as the transport protocol to invoke a Web Service. We observe that wireless communication by means of HTTP over GPRS and wLAN is widely supported. On these channels it is easy to create an HTTP connection and invoke remote Web Services using HTTP as a transport protocol. However, to the best of our knowledge, there is no implemented support for executing an HTTP POST operation over a Bluetooth channel within a J2ME MIDlet. To overcome this limitation, we introduced in our framework a new entity (*i.e.*, a distributed proxy) that takes care of binding SOAP messages to the Bluetooth transport protocol. Such entity interfaces clients to Web Services Containers. In this way, we can maintain the same server-side architecture and guarantee the interoperability of the application running on the mobile device with any Web Service (see **Figure 4**). Therefore, the architecture of our framework is based on three different entities:

1) CLIENT. It runs on a Bluetooth-enabled mobile device and it invokes the Web Service on a Bluetooth channel.

2) PROXY. It interfaces clients with the Web Services Container.

3) WSC. The Web Services Container replies to clients' requests communicating through the PROXY.

4. Design Choices

Our design choices descend from our prerequisites of having a framework to invoke web services over a Bluetooth connection that is:

1) *Transparent*: in the sense that it should allow programmers to develop device-independent applications.

2) *Data independent*: in the sense that client applications could exchange any kind of data, even user-defined.

Moreover, as implementation constraint, we restricted our attention to license free implementation of the Bluetooth stack.

In order to obtain transparency to programmers, we have to use the JSR-82 API standard [14], allowing applications to run both on mobile devices with limited computation power using J2ME as environment and on powerful computer machines using J2SE environment. Then, to our knowledge, we have only two alternatives that are license free.

1) To use the BlueCove JSR-82 implementation that interfaces with the Microsoft Bluetooth stack found in Windows XP starting from the SP2 version.

2) To use BlueZ, the Linux Bluetooth stack, and provide a JSR-82 implementation for BlueZ.

Obviously, the first choice is an off *the shelf* solution; while, the second one requires an *in-house* development.

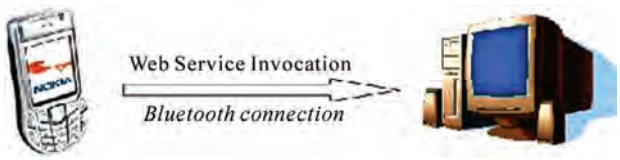


Figure 3. Our scenario.

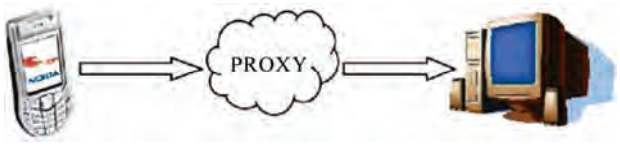


Figure 4. Proxy invocation scenario.

But the latter allows us to operate in a completely free and open environment. Besides, BlueCove, when used on winsock, supports only RFCOMM as communication protocol and it does not allow setting low level parameters on the connection as BlueZ does. Since our goal is to perform remote procedure calls using the Web Services technology, we have considered two solutions for the message exchanging protocol: XML-RPC [32] and SOAP [17]. XML-RPC is a way to perform procedure calling using HTTP as transport protocol and XML as encoding protocol; while, SOAP is a lightweight protocol for exchange of information in a decentralized and distributed environment. To have a complete data independent infrastructure, we have chosen JAX-RPC [33]. In fact, even if XML-RPC is lighter, easier to use for developing purposes and more suitable to the bandwidth limitation of the Bluetooth channel, the JAX-RPC solution provides more powerful features and advantages:

1) SOAP passes parameters by name while XML-RPC passes parameters by position, resulting in a dependance on the order of parameters.

2) SOAP allows user-defined record types by extending the XML document using XML Schemas; while, XML-RPC only allows the base types defined in the specification.

3) Both SOAP and XML-RPC support passing binary data in an XML document using Base-64 encoding, but XML-RPC defines string parameters as being ASCII text. Some XML-RPC servers will enforce this, forcing the user to pass internationalized text as Base-64 encoded data.

4) XML-RPC is defined as operating over an HTTP connection, while SOAP describes the envelope format for an RPC request which may be sent over HTTP, SMTP, or any other protocol.

In particular, we consider the *Axis Client* API [34] as the model for carrying out client-side interactions. Axis-Client is part of the *Axis* API [35]. This API is license free and it is the most diffused free SOAP engine. Since in our scenario client applications are mobile and run on resource-constrained devices, we consider the kSOAP [31] library, a SOAP API suitable for the Java 2 MicroEdition, based on kXML [36]. The feature set of kSOAP is a subset of SOAP 1.1 features. It provides classes and methods to construct the envelope, the body, and the header of the SOAP message and to specify their elements.

5. The WSBT Framework

The goal of our work is to investigate the possibility of using Bluetooth as the communication channel for Web Service invocation from a mobile client. J2ME and JSR-82 give an appropriate programming environment for achieving our goal, but they do not provide the required support for directly accessing a Web Services Container over a Bluetooth connection from a mobile client. This is due to the impossibility of addressing IP-based transport protocols (e.g., HTTP, FTP, ...) using Bluetooth as the physical layer. Bluetooth SIG does define appropriate profiles and protocols, (e.g., PAN profile) but their use requires a preliminary pairing of the devices. Thus, a middleware needs to be created on top of the operating system level to incapsulate HTTP packets in the Bluetooth ones. However, Bluetooth communications established within the JSR-82 API are state-less and without any preliminary pairing. They only allow sending and receiving bytes over L2CAP or RFCOMM connections. For this reason, we implemented a tunneling mechanism by introducing an intermediate software layer which acts as a distributed proxy to achieve Web Services invocation over Bluetooth. Our PROXY allows application developers to deploy Web Services client applications with no extra work required to use Bluetooth as the physical layer. The PROXY is a double sided software entity:

1) MASTER (server-side of the PROXY). It extends the Web Service Container (WSC) features to accept requests coming over the Bluetooth channel. It acts as a master device, accepting incoming requests and sending back responses.

2) SLAVE (client-side of the PROXY). It lies on the smartphone and essentially it binds SOAP messages to the Bluetooth physical layer.

The result is a lightweight framework, which we named WSBT (Web Service over Bluetooth). Our proposed framework has a small footprint. Indeed, the code of the jars that need to be put onto the client to get a web services client application running needs just 50 KB of memory (this figure also include the kSOAP library whose size is 41 KB). The entities of WSBT are shown in **Figure 5**, where a CLIENT addresses Web Services on the Container through the use of the two-sided proxy.

The invocation of a service on WSC is executed by following the next steps which are summarized in **Figure 6**:

1) The CLIENT uses the classical mechanism to invoke a Web Service, ignoring the presence of the Proxy and all the details about the Bluetooth communication.

2) The SLAVE is in charge of serializing data to prepare SOAP requests for a remote Web Service.

3) The SLAVE discovers the MASTER and establishes a Bluetooth connection (either L2CAP or RFCOMM) to send the SOAP request message as a raw byte stream.

4) The MASTER receives the byte stream representing the SOAP message and posts it to the WSC.

5) The WSC elaborates the SOAP request message and returns a SOAP response message that is bound in the HTTP POST response packet.

6) The MASTER forwards back responses over the Bluetooth channel as rough byte stream, without interpreting them.

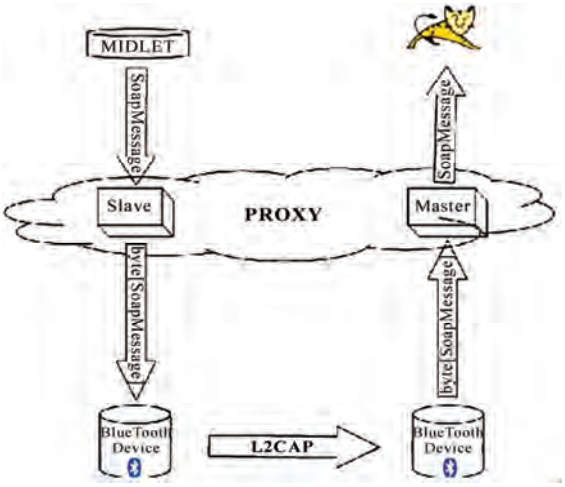


Figure 5. The entities of the WSBT framework.

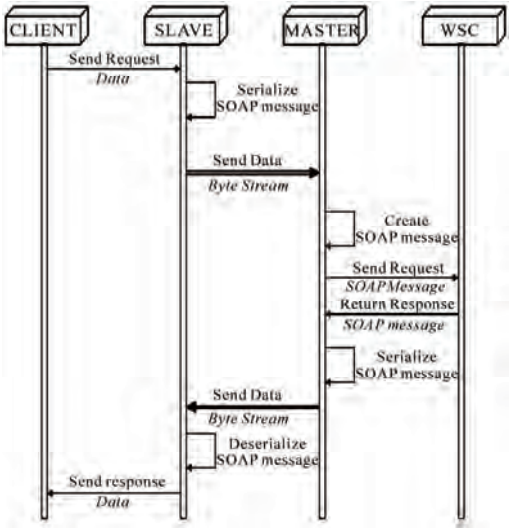


Figure 6. Sequence diagram of the WSBT framework.

7) The SLAVE receives the byte stream from the Bluetooth channel, reconstructs the SOAP return message, deserializes it, and returns results to the CLIENT.

6. Proxy's Client-Side (SLAVE)

In order to guarantee transparency to the programmers, we want to design APIs such that all details related to the use of Bluetooth as the communication channel are hidden. Our APIs are modeled on the AXIS Client API and implement Web Services invocations in a JAX-RPC [33] compliant way. To this aim, we developed a J2ME package (wsbt) whose structure is given in **Figure 7**. The Call class is the core of our framework and it is in charge of implementing the invocation mechanism. It provides methods of the Axis Client's Call interface [37]. Therefore, porting a Web Services-client application from a desktop environment (Axis Client API-based) to a mobile Bluetooth-enabled device is very simple. Indeed, it is enough to change the package to import form org. apache.axis.client to our wsbt.

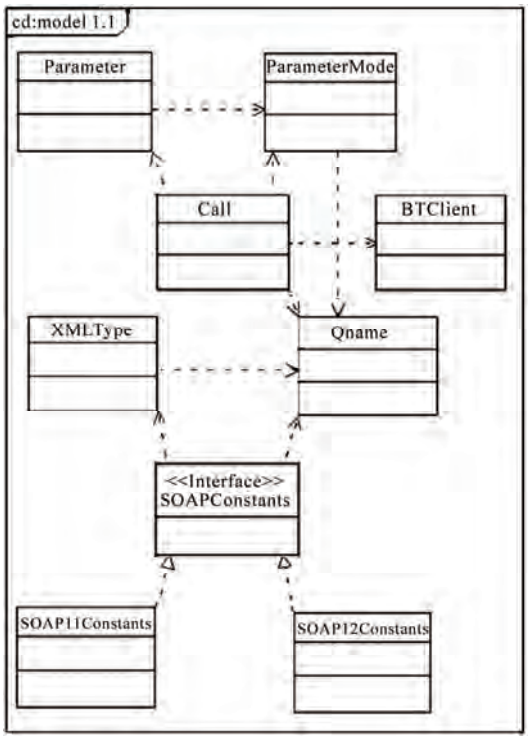


Figure 7. Structure of the wsbt Java Package.

Figure 8 depicts the fragment of a Java code for the invocation of a simple EchoString service. Invocations involving user-defined classes are not much harder to code: they only require the registration of the type mapping.

The actual invocation is carried out by the invoke method. This method, whose code is given in **Figure 9**, takes as input the list of parameters and constructs a SOAP message. This SOAP message is serialized and then sent over the Bluetooth channel, as a raw array of bytes. Finally, the method waits for the SOAP response message, deserializes it and returns it. More precisely, the invoke method performs the following operations:

1) Create a SoapObject through the use of the kSOAP library. We need to specify the namespace for the soap object and the operation name associated to the Call object.

2-5) Add to the SoapObject, through the method addProperty, all the input parameters for the operation associated with the Call object.

6) Serialize the SoapObject, *i.e.*, create an XML message representing the SOAP request. Details are hidden into the private method serialize.

7-8) Create an instance of the BTClient class and use the performInvocation method to send the SoapRequest message over the Bluetooth channel and to get the Soap-

Response.

9) Deserialize the response by invoking the private method deserialize and return the result.

We remark that the Bluetooth communication is handled by the *BtClient class*. When the *performInvocation* method of the *BtClient* object is invoked the following operations are performed:

1) Discovery of the Bluetooth Master;

2) Instauration of a L2CAP or RFCOMM channel;

3) Transmission of the request message over the channel and reception of the response message.

The code of performInvocation is given in **Figure 10**.

7. Proxy's Server-Side (MASTER)

The server-side of the proxy will run as a J2SE application on a desktop computer. Design of this part has been driven by the requirements described in Section 4.

As previously discussed, the MASTER accepts client Bluetooth connections, gets SOAP requests, posts them to the WSC, and sends back obtained SOAP responses. It consists of two main classes: the BTServer and the Poster.

The BTServer class takes care of:

1) setting the device in *discoverable* mode

2) activating a listening connection

3) accepting incoming connections

4) performing I/O on the Bluetooth channel

5) instantiate a Poster object, passing to it the URL of the service to invoke and the SOAP request to send.

The Poster class is in charge of:

1) performing an HTTP post operation on the WSC, using the Apache *Commons Http-Client* package

2) giving back the SOAP request to the BTServer object

```
(1) Call call = new Call();
(2) String address = "http://localhost:8080/axis/EchoService";
(3) call.setTargetEndpointAddress(address);
(4) call.setOperationName("echo");
(5) call.addParameter("toEcho", XMLType.XSD_STRING, ParameterMode.IN);
(6) call.setReturnType(XMLType.XSD_STRING);
(7) String result = (String)
(8) call.invoke(new Object[] [] { new String("Hello World")});
```

Figure 8. Invocation of the EchoString service.

```
public Object invoke(Object[] args) {
(1) SoapObject method = new SoapObject("urn:method", operation);
(2) for (int i = 0; i < params.size(); i++) {
(3)     Parameter tmp = (Parameter) params.elementAt(i);
(4)     QName qname = tmp.getXmlType();
(5)     method.addProperty(tmp.getType(), args[i]); }
(6) String soapRequest = serialize(method);
(7) BTClient bt = new BTClient();
(8) String soapResponse = bt.performInvocation(address, soapRequest);
(9) return deserialize(soapResponse);
}
```

Figure 9. Java code for invoke.

```
public String performInvocation(String address, String soapRequest) {

        /* Check if the address is already in cache */
        String BT_url = queryCache(address);

        /* If not in cache, perform discovery and update the cache */
        if (BT_url == null)
           BT_url = discoverMaster(address);

        /* If the discovery has failed, throw an Exception
        if (BT_url == null)
           throw UnavailableMasterException;

        /* Once the BT url has been recovered perform the connection */
        /* Get a L2CAP or RFCOMM connection */
        _connection = Connector.open(BT_url);

        /* Send the SOAP request and get the SOAP response over the BT channel */
        String soapResponse = null;
        try {
                send(address, soapRequest);
                soapResponse = receive();
                _connection.close();
        } catch (IOException ioe) { }

        return soapResponse;
}
```

Figure 10. Java code for the performInvocation method of BTClient class.

In order to have a licence-free and JSR-82 compliant implementation of our Proxy's server-side, we have considered two alternatives:

To use BlueCove [29], the free implementation of the JSR-82 API is within the Microsoft Windows XP SP2.

To use BlueZ [20], the Linux Bluetooth stack, and provide a JSR-82 implementation for BlueZ.

7.1. BlueCove-Based Implementation

BlueCove [29] is a free implementation of the JSR-82 API that runs over the Windows XP SP2 Bluetooth stack. When used on winsock, it only provides the support for RFCOMM protocol, which is the emulation of a serial port communication enabling programmers to open inbound DataInputStream and outbound DataOutputStream. According to the JSR-82 API, the Bluetooth listener is implemented by a Notifier object, which handles a StreamConnection. **Figure 11** shows main steps of the proxy's server-side.

The Java-code in **Figure 11** executes the following operations:

1) Set the device in discoverable mode.

2-3) Activate a listening connection on localhost, on the channel 1, named "rfcomm test".

4) Accept incoming connections.

5) Open an InputStream on the connection.

6-7) Read data on the stream.

8-9) Post the SOAP request at the specified address, using the Poster class, to get the Soap response.

10) Open an OutputStream on the connection.

11) Write data, *i.e.*, the Soap response.

7.2. JBlueZen: BlueZ-based JSR-82 API Implementation

BlueZ is the implementation of the Bluetooth stack included in the Linux kernels 2.4 and 2.6. It gives to programmers direct access to the transmission layers and allows to set up several communication parameters to tune the transmission to the application characteristics. To our knowledge, there is no JSR-82 implementation that runs over BlueZ and thus we have written our own. As presented in **Figure 12**, the BlueZ stack offers to programmers a Berkeley socket interface (C-based) to handle L2CAP, RFCOMM, and SDP features. As a result, we had to implement a set of C functions which access Berkeley sockets in order to accomplish data exchange and service discovery. These functions will be then interfaced with Java applications through the use of the Java Native Interface (JNI) [38]. The resulting scheme is the Java-JNI package presented in **Figure 13**, which we named JBlueZen. Programming on RFCOMM sockets to build a RFCOMM server is very similar to programming on TCP/IP sockets, with some relevant differences like the maximum number of allowed ports (65536 for TCP/IP, 30 for RFCOMM) and different functions for the byte-ordering (big-endian, littleendian). The main steps performed by our implementation to open a listening RFCOMM socket are shown in **Figure 14**.

1) Create a socket: AF BLUETOOTH indicates that we are using the Bluetooth communication channel, SOCK STREAM that it is stream-oriented service, and BTPROTO RFCOMM that we are using RFCOMM.

2-4) Populate the loc addr structure used to set information over the Bluetooth adapter in the bind system call: use Bluetooth as the communication channel, accept connection from any device and on the specified channel (*i.e.*, the RFCOMM mechanism for implementing a port).

5) Bind the socket on a listen port.

6) Listen on the socket for incoming connections.

7) Accept incoming connections and get a communication socket through.

We remark that using the *setsockop*t system call, it is also possible to set some socket options, such as the authentication and encryption to use on the Bluetooth transmission, or the device role (MASTER or SLAVE). For L2CAP, it can also be set the maximum amount of consecutive bytes that can be sent/received on the connection (MTU).

The mechanism for L2CAP is similar to RFCOMM and it is shown in **Figure 15**. The only difference here is the type of socket to be used: SOCK SEQPACKET instead of SOCK STREAM.

The last protocol to implement is the Service Discovery, which requires a bit more coding effort. BlueZ provides a set of C functions to address service discovery

```
(1)  (LocalDevice.getLocalDevice).setDiscoverable(DiscoveryAgent.GIAC);
(2)  (StreamConnection) notifier = (StreamConnection)
(3)  Connector.open("btspp://localhost:1;name=rfcommtest;master=true;encrypt=false;
                     authorize=false;authentication=false;receiveMTU=512;transmitMTU=512");
(4)  notifier.acceptAndOpen();
(5)  InputStream input = notifier.openInputStream();
(6)  /* Perform buffered readings to get the sopaRequest */
(7)  String soapRequest = input.read();
(8)  Poster poster = new Poster(address);
(9)  String soapResponse = poster.doPost(soapRequest);
(10) OutputStream output = notifier.openOutputStream();
(11) output.write(soapResponse.getBytes());
```

Figure 11. Java code for RFCOMM implementation proxy's server-side.

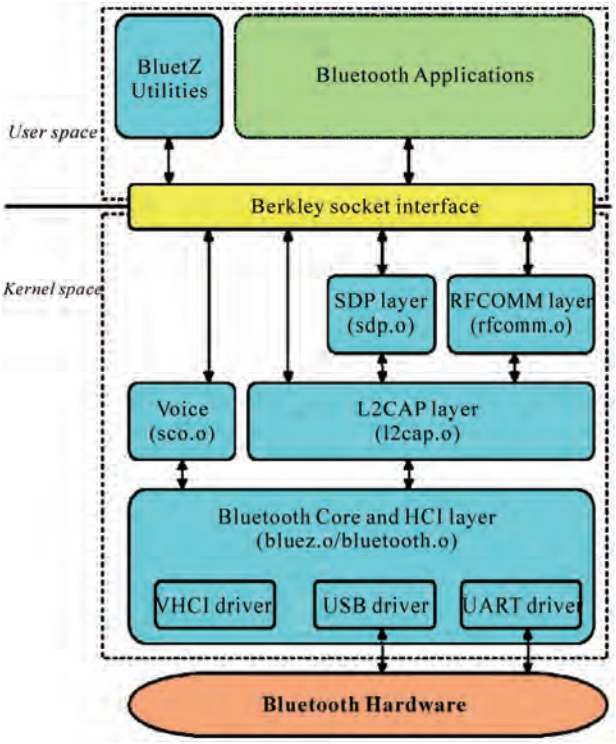


Figure 12. The BlueZ Bluetooth stack.

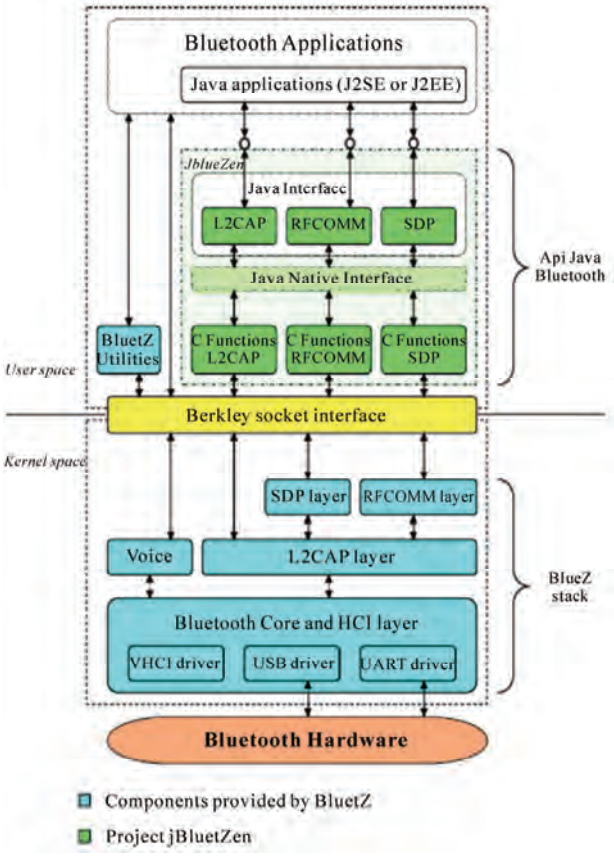


Figure 13. The JBlueZen package.

features both on the 'server' side and on the 'client' side. To implement the discovery in the JSR-82 standard way, on the server-side, we had to handle the service record and register services on it; on the client-side we had to access the HCI to perform inquiry and connect to the SDP server running on the remote device to get all the info. The last step was to provide a Java interface (JBlueZen) to the described functions we implemented in C. This was made through the use of the JNI to implement all the communication (L2CAP and RFCOMM) and the discovery features in the JSR-82 standard way.

7.3. JBlueZen-Based Implementation

Using the Java interface of our package JBlueZen, we are able to implement the MASTER (proxy's server-side) on Linux. The mechanism is similar to the Windows implementation, but in this case we can choose among two different communication protocols: L2CAP and RFCOMM and we can set many parameters on the Bluetooth connection which were fixed in the Windows implementation, such as the device role (MASTER or SLAVE) or the types of packet to use (DH-DM 1-3-5).

Obviously, the code fragment that uses the RFCOMM implementation is identical to the code given in **Figure 14**. The L2CAP implementation is instead implemented as shown in **Figure 16**.

The Java code in **Figure 16** executes the following operations:

1) Set the device in discoverable mode.

2) Activate a listening connection on the localhost, on the psm 1001, named "testl2cap", and with the desired settings. Remark that psm is the Protocol Service Multiplexer and is the mechanism used by L2CAP to implement multiplexing on connections.

3) Accept incoming connections.

4-5) Get the in/outbound maximum transfer unit.

6) Read data on the stream.

7-8) Post the SOAP request at the specified address, using the Poster class, to get the Soap response.

9) Write data on the connection.

8. Performance Evaluation

In this section, we present the results of several experiments that we ran to evaluate the real applicability and the lightness of the deployed framework. We also compare the performances of the two different implementations that we described in the previous sections. For our experiments, we set up the following small test-bed:

1) (Linux) Server-side: Workstation HP XW6000, Xeon 2.8 GHz Dual-Processor, 2 GB RAM, with Trust BT 180 Bluetooth USB dongle, running Fedora Core 4 with 2.6.11 Linux Kernel.


```
(1) int sock = socket(AF_BLUETOOTH, SOCK_STREAM, BTPROTO_RFCOMM);
(2) loc_addr.rc_family = AF_BLUETOOTH;
(3) loc_addr.rc_bdaddr = *BDADDR_ANY;
(4) loc_addr.rc_channel = channel;
(5) bind(sock, (struct sockaddr *)&loc_addr, sizeof(loc_addr));
(6) listen(sock, MAX_NO_CONNECTIONS);
(7) int sockfd = accept(sock, (struct sockaddr *)&rem_addr, &opt);
```

Figure 14. C-code to open a RFCOMM listening socket on BlueZ.

```
(1) int sock = socket(AF_BLUETOOTH, SOCK_SEQPACKET, BTPROTO_L2CAP);
(2) loc_addr.l2_family = AF_BLUETOOTH;
(3) loc_addr.l2_bdaddr = *BDADDR_ANY;
(4) loc_addr.l2_psm = psm;
(5) bind(sock, (struct sockaddr *)&loc_addr, sizeof(loc_addr));
(6) listen(sock, MAX_NO_CONNECTIONS);
(7) int sockfd = accept(sock, (struct sockaddr *)&rem_addr, &opt);
```

Figure 15. C-code to open a L2CAP listening socket on BlueZ.

```
(1) (LocalDevice.getLocalDevice).setDiscoverable(DiscoveryAgent.GIAC);
(2) (L2CAPNotifier) notifier = (L2CAPNotifier)
          Connector.open("btl2cap://localhost:1001; name=testl2cap;
                         master=true;encrypt=false;authorize=false;
                         authentication=false;receiveMTU=512;
                         transmitMTU=512");
(3)  L2CAPConnecction conn = (L2CAPConnection) notifier.acceptAndOpen();
(4)  int inMTU = conn.getReceiveMtu();
(5)  int outMTU = conn.getTransmitMtu();
     /* Perform cycling readings over the inMTU to get the soapRequest */
(6)  String soapRequest <- conn.read();
(7)  Poster poster = new Poster(address);
(8)  String soapResponse = poster.doPost(soapRequest);
(9)  conn.write(soapResponse.getBytes());
```

Figura 16. Java code for L2CAP implementatio of Proxy's server-side.

2) (Windows) Server-side: Workstation HP XW6000, Xeon 2.8 GHz Dual-Processor, 2 GB RAM, with Trust BT 180 Bluetooth USB dongle, running Windows XP SP2.

3) Client-side: Nokia 6630 Smartphone with Symbian OS, MIDP 2.0 and JSR-82 APIs support.

Our tests were aimed to evaluate the effciency of our framework in terms of used bandwidth and of serialization/deserialization performances. Thus, we had to measure times for serializing and deserializing messages and transmission times over the Bluetooth channel. Noticing that on a Bluetooth channel transmission and receiving times are significantly different. In fact, usually Bluetooth stack implementations assign a much larger bandwidth to master-slave communications than to slave-master ones. Thus, the same message will require more time to be transmitted from the mobile device to the container than to be transmitted from the container to the mobile. In our experiments, we invoked a test Web Service implementing just an echo service and measured only the transmission time from the mobile device to the container. Clearly, these times are an upper bound to the times we could measure for transmissions in the other direction. We also measured the serialization/deserialization times on the mobile device. Noticing that these operations are local to the mobile device and their execution times do not depend on our framework but only on the implementation of the Java Virtual Machine and of the KSOAP library that are used on the mobile device. However, we have measured these times to evaluate the lightness and effectiveness of our framework for developing applications invoking web services from mobile devices with limited resources.

To have a comprehensive analysis of the performances of our framework several experiments were run. In the first experiment, we measured times necessary to complete a client request. We repeated the same experiment using both the Blue-Cove-based implementation and the JBlueZen-based implementation in order to compare the effciency of the two implementations. The second experiment was aimed to isolate and evaluate the impact of serialization and deserialization on the performance of the framework. The third experiment was aimed to measure the discovery delay. In fact, our framework allows clients to dynamically discover Bluetooth servers by handling Bluetooth specification policies for devices and services discovery. In our last experiment, we tested framework's performances with different Bluetooth communication modalities, taking advantage of the capacity of the JBlueZenbased implementation to modify some

low-levels Bluetooth communication parameters.

8.1. Transmission Times

Our first experiment was aimed to evaluate the time required to transmit a message on the mobile-container channel with respect to the size of the message. We measured the time taken to complete the invocation of the send method in the performInvocation method (see **Figure 10**). Notice that, since the IO is blocking, the send method returns only when all input data is sent and an ACK is received from the container for the correct reception of the last data byte. To compare our two implementations we repeated the experiment in three different cases: using the JBlueZen-based implementation both with a RFCOMM or L2CAP connection and using the BlueCove-based implementation with a RFCOMM connection. We remark that BlueCove, when used on winsock, does not allow setting a L2CAP connection.

We distinguished two cases: unstructured and structured messages. In the first case, we assumed that the message consists of a string (array of bytes) and we ran tests for strings of size ranging from 0.5 KB to 30 KB (the size is increased by 0.5 KB in each test). We were not able to perform tests for larger strings since in the JVM deployed on the mobile device used in our experiments (*i.e.*, Nokia 6630) it is not possible to instantiate String objects of size greater than 30 KB. In the second case, we assumed that messages contain complex data types. Our framework represents complex data types as JavaBeans. We assumed that the request message contains an array of Address objects, which consist of six String objects of fixed length and two Integer objects. We measured transmission times with respect to the length of the array and we ran tests for array lengths ranging from 1 to 65. **Figure 17** shows message lengths with respect to the length of the array. For each test, we repeated the invocation 50 times and computed the average times. To guarantee that for each iteration the device were in the same initial conditions, we used dedicated threads. In fact, for each invocation a new thread was created and destroyed after the operation.

Figures 18 and **19** show transmission times for unstructured messages with respect to the size of the string. Notice that the size of the string is not equal to the size of the message sent on the Bluetooth channel (we have to consider additional bytes inserted by KSOAP and by the serializer). However, we have observed that this overhead is constant (approx. 500 bytes) and it does not depend on the string's size. It can be seen that our two implementations are equivalent when using the RFCOMM communication modality, while L2CAP communication modality is 20% more effcient.

Figure 20 shows transmission times for structured messages with respect to the length of the array using a

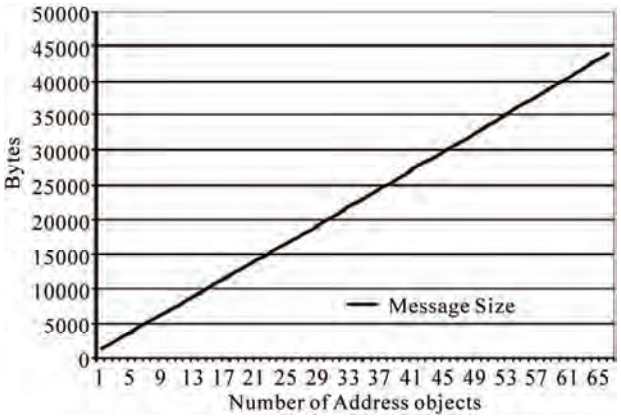


Figure 17. Array of address objects: message length (in bytes) vs. array length.

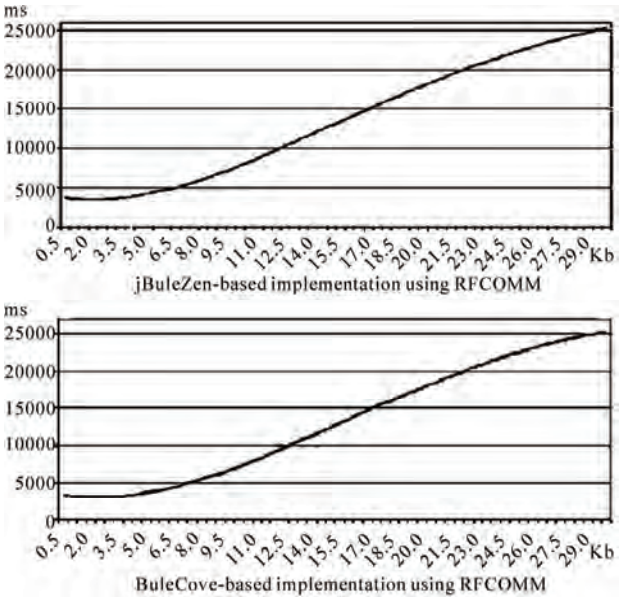


Figure 18. Times to send unstructured data using the RFCOMM communication modality.

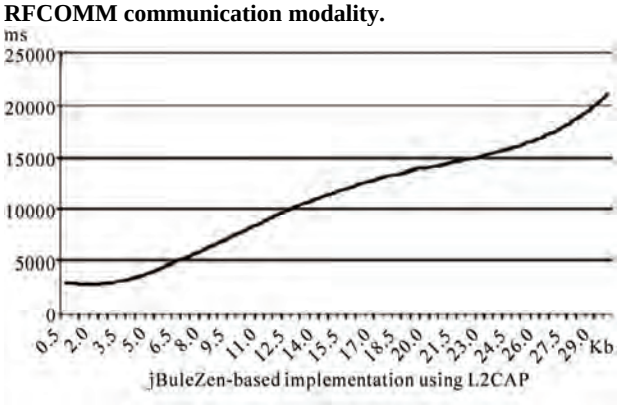


Figure 19. Times to send unstructured data using the L2 CAP communication modality.

RFCOMM connection. Our experiments show that the two implementations are equivalent but the BlueCove-

based implementation is a little bit more effcient. Moreover, the average throughput is around 10 KBps. Finally, **Figure 21** shows transmission times for L2CAP connections (in the BlueZen-based implementation). As expected, performances are slightly more effcient using L2 CAP than usign RFCOMM.

8.2. Serialization-Deserialization

The second experiment was aimed to measure the processing times for serialization and deserialization to evaluate lightness of our framework. **Figure 22** shows times to serialize and deserialize SOAP messages containing arrays of Address objects with respect to the length of the array. Notice that serialization/deseraliziation times are equal for both our implementations since they depend only on the implementation of the Java Virtual Machine and of KSOAP library. We can observe that, when the array length is large enough, serialization is heavier than deserialization: differences between serialization and deserialization times are due to the different behavior of the kSOAP when parsing and serializing XML documents.

It can be seen that the serialization time increases linearly with the length of the message and it weights less than 1/3 of the transmission time. Thus, we can state that our framework is suffciently light and it has no dramatic impact on the effciency of the framework.

8.3. Discovery

This experiment was aimed to measure the discovery delay. In fact, our framework allows the client to dynamically discover Bluetooth servers and not to be bounded to hard-coded settings. For this reason, the framework handles Bluetooth specification policies for devices and services discovery.

We ran 50 discovery operations, getting only 1 timeout error (see **Figure 23**). In all the other cases, we measured a discovery delay that is around 14.5 seconds. However, introducing a cache mechanism to store addresses of recently used Bluetooth devices we were able to reduce delays to a few hundreds milliseconds (see **Figure 24**).

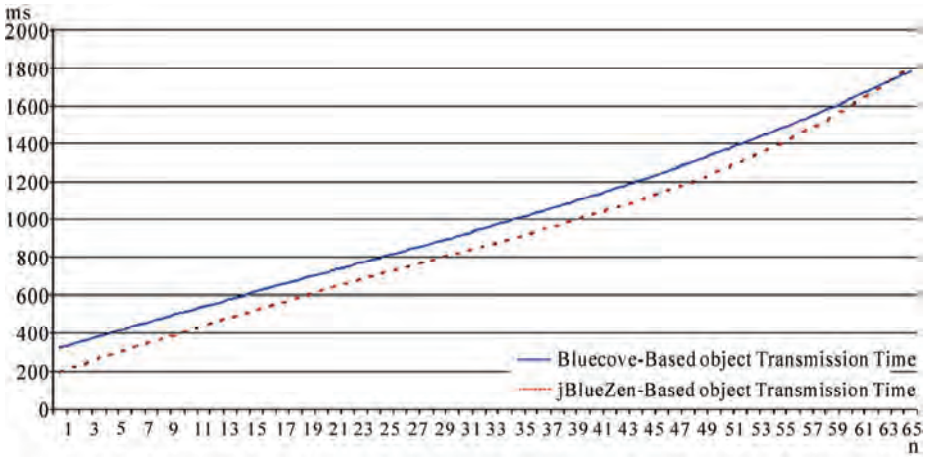


Figure 20. Total times to send structured data using RFCOMM communication modality.

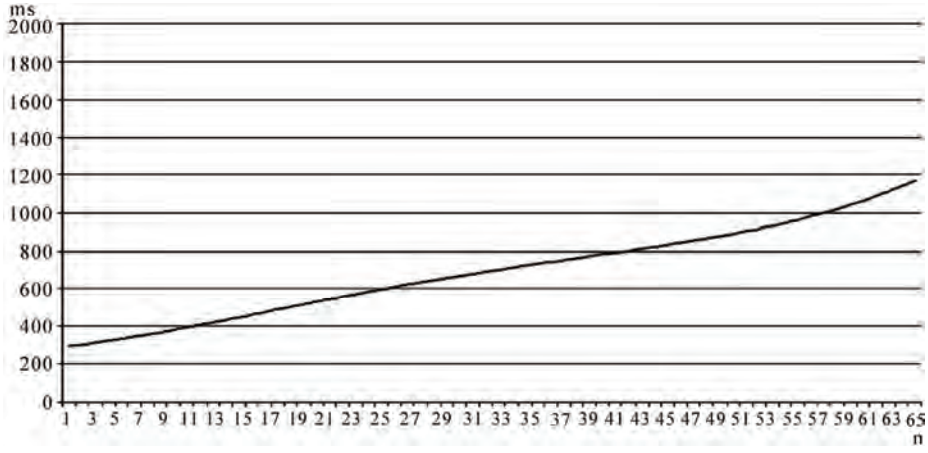


Figure 21. Total times to send structured data using L2CAP communication modality.

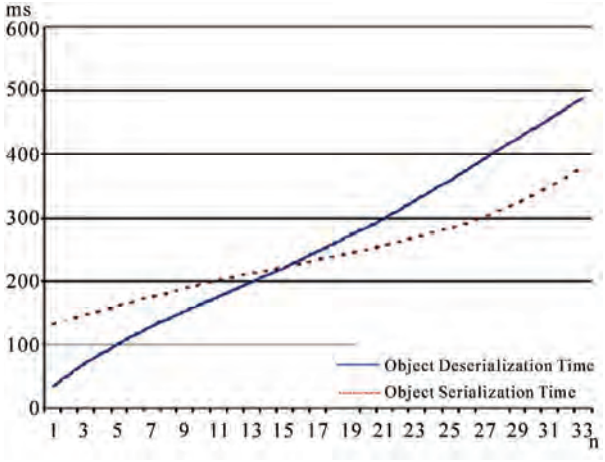


Figure 22. Serialization and deserialization times for an array of Address objects.

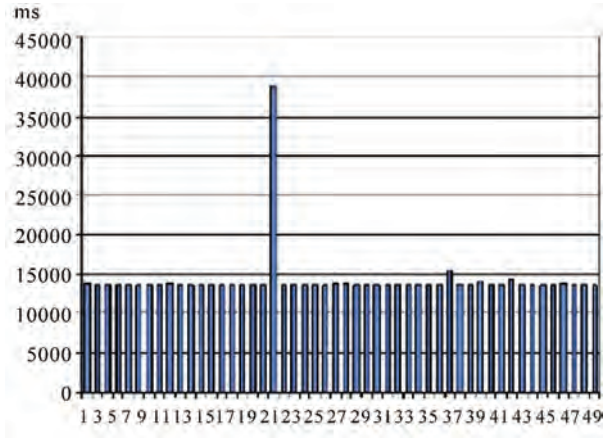


Figure 23. Discovery delay.

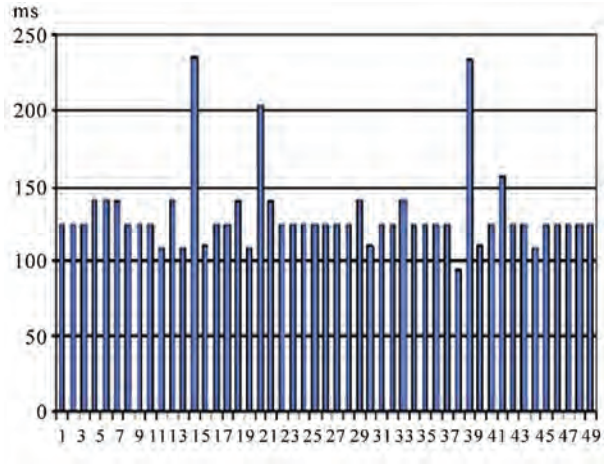


Figure 24. Discovery delay using cache.

9. Conclusions

In this paper, we presented a framework to allow applications running on mobile devices to invoke Web-Services over a Bluetooth-connection. We gave two different implementations of our framework, and give an extensive analysis of their performances. We can summarize differences between the two implementations as follows: the first implementation is Windowsbased and works on top of BlueCove, that is a third-party free implementation of the JSR-82 API that runs over the Windows XP SP2 Bluetooth Stack; the second one is Linux-based and works on top of our implementation of a Java package for exploiting Bluetooth features giving to the programmer control over several low level parameters of the Bluetooth channel.

Our experiments confirm the real applicability and lightness of the framework showing that Bluetooth is well suited to be the transport layer for Web Services accessing from wireless devices. Moreover, our tests show that the Windows-based implementation is a little bit more effcient when using the RFCOMM communication modality, but the Linuxbased implementation obtains the best performances when using the L2CAP communication modality.

We think that results presented in this paper show that Bluetooth is a good candidate to be the leading communication technology to provide access to the Web from mobile, low cost devices.

10. References

[1] "The Bluetooth Technology," March 2008. http://www.bluetooth.com

[2] "Bluetooth Wireless Technology," March 2008. http://www.ericson.com/technology/techarticles/Bluetooth.shtml

[3] C. Bisdikian, "An Overview of the Bluetooth Wireless Tecnology," *IEEE Communication Magazine*, Vol. 39, No. 12, 2001, pp. 86-94.

[4] O. P. Association, "Going Mobile: An International Study of Content Use and Advertising on the Mobile Web," March 8, 2007. http://www.onlinepublishers.org /media/176W opa going mobile report mar07.pdf

[5] J. Beutel and O. Kasten, "A Minimal Bluetooth-Based Computing and Communication Platform," Technical Report, Engineering and Networks Lab, Swiss Federal Institute of Technology, 2001.

[6] J. Misic, K. L. Chan and V. B. Misic, "Tcp Traffic in Bluetooth 1.2: Performance and Dimensioning of Flow-Control," *Proceedings of the* 2005 *IEEE Wireless Communications and Networking Conference*, New Orleans, 2005, pp. 1798-1804.

[7] F. Kargl, S. Ribhegge, S. Schlott and M. Weber, "Bluetooth-Based Ad-Hoc Networks for Voice Transmission," *Proceedings of the* 36*th Annual Hawaii International Conference on System Sciences*, Hawaii, 2003, pp. 314-322.

[8] S. Zeadally and A. Kumar, "Protocol Support for Audio

Streaming between Bluetooth Devices," *Proceedings of the* 2004 *IEEE Radio and Wireless Conference*, Atlanta, 2004, pp. 303-306.

[9] J. Cano, D. Ferrandez-Bell and P. Manzoni, "Evaluating Bluetooth Performace as the Support for Context-Aware Applications," *Proceedings of the* 12*th IEEE International Conference on Computer Communications and Networks*, Dallas, 2003, pp. 333-347.

[10] Y. Lim, J. Kim, S. L. Min and J. S. Ma, "Performance Evaluation of the Bluetoothbased Public Internet Access Point," *Proceedings of the* 15*th International Conference on Information Networking*, Beppu City, Oita, 2001, pp. 643-648.

[11] V. Auletta, C. Blundo, E. D. Cristofaro and G. Raimato, "A Lightweight Framework for Web Services Invocation over Bluetooth," *Proceedings of the* 2006 *IEEE International Conference on Web Services* (*ICWS*'06), Chicago, 2006, pp. 331-338.

[12] V. Auletta, C. Blundo, E. D. Cristofaro and G. Raimato, "Performance Evaluation of Web Services Invocation over Bluetooth," *Proceedings of the ACM International Workshop on Performance Monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks*, Terromolinos, Spain, 2006, pp. 1-8.

[13] S. Berger, S. McFaddin, C. Narayanaswami and M. Raghunath, "Web Services on Mobile Devices-Implementation and Experience," *Proceedings of the* 5*th IEEE Workshop on Mobile Computing Systems and Applications*, Monterey, California, 2003, pp. 100-109.

[14] "JSR 82: Java APIs for Bluetooth," March 2008. http://www.jcp.org/en/jsr/detail?id=82

[15] "Java 2 Platform, Micro Edition (J2ME)," March 2008. http://java.sun.com/j2me/

[16] "The Official Bluetooth Membership Site," March 2008. http://www.bluetooth.org

[17] "SOAP Version 1.2," March 2008. http://www.w3.org/TR/soap/

[18] "Broadcom Bluetooth Solutions," March 2008. http://www.broadcom.com/products/Bluetooth/Bluetooth-RF-Silicon-and-Software-Solutions

[19] "Windows Support for Bluetooth," March 2008. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/bluetooth/bluetooth/about bluetooth.asp

[20] "Bluez: Official Linux Bluetooth Protocol Stack," March 2008. http://www.bluez.org/

[21] "JSR 36, JSR 218: Connected Device Configuration (CDC)," March 2008. http://java.sun.com/products/cdc/

[22] "JSR 30, JSR 139: Connected Limited Device Configuration (CLDC)," March 2008. http://java.sun.com/products/cldc/

[23] "Mobile Information Device Profile (MIDP): JSR 37, JSR 118," March 2008. http://java.sun.com/products/midp/

[24] "Mobile Information Device Profile (Midp): JSR 37," March 2008. http://jcp.org/aboutJava/communityprocess/final/jsr037/index.html

[25] "Mobile Information Device Profile 2.0 (Midp 2.0): Html, JSR118," March 2008. http://jcp.org/aboutJava/communityprocess/final/jsr118/index

[26] "Bluetooth Solutions by Atinav Avelink," March 2008. http://www.avelink.com/bluetooth/index.htm

[27] "Impronto Rococo Software," March 2008. http://www.rococosoft.com/

[28] "Avetana Jsr-82 Implementation," March 2008. http://www.avetanagmbh.de/avetanagmbh/produkte/jsr82.eng.xml

[29] "Blue Cove Jsr-82 Implementation," March 2008. http://code.google.com/p/bluecove/

[30] "Webservices-Soap," March 2008. http://ws.apache.org/soap/

[31] "kSOAP 2," March 2008. http://kobjects.org/

[32] "XML-RPC," March 2008. http://www.xmlrpc.com/

[33] "Java API for xML-Based RPC," March 2008. http://java.sun.com/webservices/jaxrpc/

[34] "The axis Client Api," March 2008. http://ws.apache.org/axis/java/apiDocs/org/apache/axis/client/package-summary.html

[35] "Web Service Axis," March 2008. http://ws.apache.org/axis/

[36] "kXML," March 2008. http://kxml.sourceforge.net/

[37] "The Call Class JavaDoc," March 2008. http://ws.apache.org/axis/java/apiDocs/index.html

[38] "Java Native Interface," March 2008. http://java.sun.com/j2se/1.4.2/docs/guide/jni/index.html

Wireless Sensor Network, **2010, 2, 462-466**
doi:10.4236/wsn.2010.26057 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Reconstruction of Wireless UWB Pulses by Exponential Sampling Filter

Juuso T. Olkkonen[1], Hannu Olkkonen[2]
[1]*VTT Technical Research Centre of Finland, Espoo, Finland*
[2]*Department of Physics and Mathematics, University of Eastern Finland, Kuopio, Finland*
E-mail: *juuso.olkkonen@vtt.fi, hannu.olkkonen@uef.fi*
Received November 16, 2009; *revised December* 25, 2009; *accepted January* 20, 2010

Abstract

Measurement and reconstruction of wireless pulses is an important scheme in wireless ultra wide band (UWB) technology. In contrary to the band-limited analog signals, which can be recovered from evenly spaced samples, the reconstruction of the UWB pulses is a more demanding task. In this work we describe an exponential sampling filter (ESF) for measurement and reconstruction of UWB pulses. The ESF is constructed from parallel filters, which has exponentially descending impulse response. A pole cancellation filter was used to extract the amplitudes and time locations of the UWB pulses from sequentially measured samples of the ESF output. We show that the amplitudes and time locations of p sequential UWB pulses can be recovered from the measurement of at least $2p$ samples from the ESF output. For perfect reconstruction the number of parallel filters in ESP should be $2p$. We study the robustness of the method against noise and discuss the applications of the method.

Keywords: Wireless Sensor Networks, UWB, Network Security, Finite Rate of Innovation

1. Introduction

The measurement and reconstruction of some classes of signals containing discontinuities such as impulses and edges is difficult [1,2]. Sampling methods are historically relied on Shannon's theorem [3]. The perfect reconstruction (PR) of the continuous signal from the sampled version requires that the signal is band limited, *i.e.*, its frequency spectrum has a maximum frequency f_M. The PR is possible only if the sampling frequency $f_s \geq 2 f_M$. For example, signals in optical devices and radiation detectors are not band limited and classical sampling approaches are not relevant for extracting the information.

Sampling scheme with finite rate of innovation (FRI) [4,5] has recently got vast interest in signal processing society, since the FRI methods are not restricted to the recovery of the band-limited signals. The key idea in FRI methods is that the signal (e.g., the Dirac impulse stream) is measured with a sampling filter, which is constructed using analog circuits. The output of the sampling filter is measured and the original signal is reconstructed from the discrete samples. Excellent articles and reviews consider the FRI sampling and reconstruction techniques [4-9] and many feasible applications have been published. However, the experimental work on the verification of the underlying theoretical considerations is lacking, especially the effect of noise on the reconstruction accuracy.

The information in wireless ultra wideband (UWB) devices is usually carried out by monocycle Gaussian pulses. In year 2002 the FCC restricted the allowed frequency band between 3.1-10.6 GHz for unlicensed UWB transmission [10]. The monocycle Gaussian pulse stream does not strictly meet this constraint and other pulse shapes have been introduced, e.g., the family of the orthogonal UWB pulse waveforms [11]. In low-range wireless UWB communication devices, which transmit sequential pulses, the information is coded to the time locations of the pulses. The UWB pulse generators are relatively easy to implement in VLSI [12]. The pulse stream is designed so that its power spectral density coincides with the FCC criteria.

In this work we study the FRI-like method aimed at sampling and reconstruction of the UWB pulses. As a sampling device we apply an exponential sampling filter (ESF), whose output is measured sequentially. We study the robustness of the method against noise and discuss the applications of the method.

2. Theoretical Considerations

2.1. Sampling of the UWB Pulse Train

We consider the UWB pulse train, which is approximated by

$$I(t) = \sum_{i=1}^{p} A_i \, \delta(t - t_i) \tag{1}$$

where the Dirac distribution $\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ for $x \neq 0$. A_i denotes the amplitude and t_i time location. The pulse train is fed to the exponential sampling filter (ESF) consisting of *N* parallel RC-filters (**Figure 1**). The ESF has the causal impulse response

$$h(t) = \sum_{k=1}^{N} B_k \exp[-\alpha k t] \quad t \geq 0 \tag{2}$$

The output signal of the ESF is

$$\begin{aligned} x(t) &= I(t) * h(t) = \int_{-\infty}^{\infty} I(\tau) h(t-\tau)\, d\tau \\ &= \sum_{i=1}^{p} A_i \sum_{k=1}^{N} B_k \exp[-\alpha k (t - t_i)] \quad t \geq 0 \end{aligned} \tag{3}$$

The key idea is to start the measurement of the ESF output at time $t_0 \geq t_i$ $(i = 1, 2, ..., p)$, simultaneously closing the input of the ESF (**Figure 1**). We define the discrete time variable *n* as

$$t = t_0 + nT \ (n = 0, 1, 2, ...) \Rightarrow nT = t - t_0$$

where *T* is a sampling period. For the samples of the ESF output $x_n = x[nT]$, $n = 0, 1, 2, ...$ the following is valid

$$x_n = \sum_{i=1}^{p} A_i \sum_{k=1}^{N} B_k \exp[-\alpha k (t_0 + nT - t_i)] \tag{4}$$

By rearranging and denoting $\Delta_i = t_o - t_i$

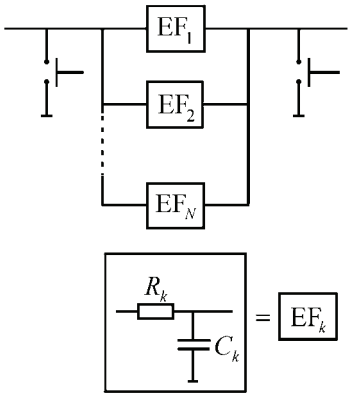


Figure 1. The construction of the ESF from the parallel exponential filters (EF), which are built using RC-circuits. The ESF parameter α is related as $\alpha kT = 1/(R_k C_k)$, where R_k and C_k are the component values.

$$x_n = \sum_{k=1}^{N} B_k \sum_{i=1}^{p} A_i \, e^{-\alpha k \Delta_i} \, (e^{-\alpha k T})^n = \sum_{k=1}^{N} b_k \lambda_k^n \tag{5}$$

where

$$b_k = B_k \sum_{i=1}^{p} A_i \, e^{-\alpha k \Delta_i} \text{ and } \lambda_k = e^{-\alpha k T} \tag{6}$$

2.2. Reconstruction of the Pulse Train

Our task is to recover the amplitudes and time locations of the pulses from the measured output samples $x_n = x[nT]$, $n = 0, 1, 2, ...$ We may write (5) in the form of matrix/vector equation

$$\begin{bmatrix} x_0 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_N \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \cdots & \lambda_N^{N-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \tag{7}$$

$$\Leftrightarrow \mathbf{x} = \boldsymbol{\lambda} \mathbf{b} \Rightarrow \mathbf{b} = \boldsymbol{\lambda}^{-1} \mathbf{x}$$

The Vandermonde matrix structure $\boldsymbol{\lambda}$ in (7) is non-singular having rank *N* and the solution of the vector $\mathbf{b}$ requires knowledge of the *N* measurement values. From (6) we obtain

$$c_k = b_k / B_k = \sum_{i=1}^{p} A_i \, e^{-\alpha k \Delta_i} = \sum_{i=1}^{p} A_i \, r_i^k \tag{8}$$

where

$$r_i = e^{-\alpha \Delta_i} \tag{9}$$

The *z* transform of the c_k sequence gives

$$\begin{aligned} Z\{c_k\} &= \sum_{k=1}^{N} \sum_{i=1}^{p} A_i \, r_i^k \, z^{-k} = \sum_{i=1}^{p} A_i \sum_{k=0}^{N-1} (r_i z^{-1})^{k+1} \\ &= \sum_{i=1}^{p} \frac{A_i r_i z^{-1} (1 - r_i^N z^{-N})}{1 - r_i z^{-1}} \end{aligned} \tag{10}$$

We define the pole cancellation filter (PCF) as

$$H_{pc}(z) = 1 + \sum_{n=1}^{p} h_n \, z^{-n} = \prod_{j=1}^{p} (1 - r_j z^{-1}) \tag{11}$$

where r_j is defined by (9). Next we consider the product filter $P(z)$

$$\begin{aligned} P(z) &= Z\{c_k\} H_{pc}(z) \\ &= \sum_{i=1}^{p} A_i r_i z^{-1} (1 - r_i^N z^{-N}) \prod_{\substack{j=1 \\ j \neq i}}^{p} (1 - r_j z^{-1}) \end{aligned} \tag{12}$$

We may note that the roots of the $P(z)$ equal the roots of the PCF. The impulse response of the $P(z)$ is

$$p_n = \sum_{k=0}^{p} c_{n-k} h_k \tag{13}$$

For solution of the roots of the $P(z)$ we set $p_n = 0$ for $n \geq 0$. This yields the matrix/vector equation

$$\begin{bmatrix} c_{2p} \\ c_{2p-1} \\ \vdots \\ c_{p+1} \end{bmatrix} + \begin{bmatrix} c_{2p-1} & c_{2p-2} & \cdots & c_p \\ c_{2p-2} & c_{2p-3} & \cdots & c_{p-1} \\ \vdots & \vdots & \vdots & \vdots \\ c_p & c_{p-1} & \cdots & c_1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix} = 0 \quad (14)$$

$$\Leftrightarrow \mathbf{c} = -\mathbf{C}\,\mathbf{h} \Rightarrow \mathbf{h} = -\mathbf{C}^{-1}\mathbf{c}$$

We may note that the solution of the p coefficients of the PCF requires the knowledge of the $2p$ values of the c_k sequence. The polynomial $[1\ h_1\ h_2 \ldots h_p]$ has the roots $r_i = e^{-\alpha \Delta_i}$ $(i = 1, 2, ..., p)$, which give $\Delta_i = -\log(r_i)/\alpha$ and the time locations as $t_i = t_0 + \log(r_i)/\alpha$. For solution of the amplitudes we may write (8) in matrix/vector form

$$\begin{bmatrix} c_1 \\ c_1 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & \cdots & r_p \\ r_1^2 & r_2^2 & \cdots & r_p^2 \\ \vdots & \vdots & \vdots & \vdots \\ r_1^p & r_2^p & \cdots & r_p^p \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix} \quad (15)$$

$$\Leftrightarrow \mathbf{c} = \mathbf{R}\,\mathbf{a} \Rightarrow \mathbf{a} = \mathbf{R}^{-1}\,\mathbf{c}$$

To summarize, the reconstruction of p sequential pulses needs the measurement of at least $2p$ samples from the ESF output. The impulse response of the ESF must obey (2), where $N = 2p$. It should be pointed out that the solution of the matrix Equations (7) and (15) require the inversion of the Vandermonde matrix, which can be obtained by an analytical formula [13]. The inversion method yields more stable results compared with the general matrix inversion.

2.3. Reduction of Noise

In practical measurements noise arising in electronic circuits interferes the results. We apply the SVD based subspace method for reducing the noise in measurement signal. Let us construct the Hankel matrix containing the measurement values $x_n = x[nT]$, $(n = 0, ..., M \geq 2p)$

$$\mathbf{H} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{M/2} \\ x_1 & x_2 & \cdots & x_{M/2+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{M/2} & x_{M/2+1} & \cdots & x_M \end{bmatrix} \quad (16)$$

where the antidiagonal elements are identical. The singular value decomposition (SVD) of the matrix $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathbf{T}} \quad (17)$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices. $\mathbf{\Sigma}$ is a diagonal matrix consisting of the singular values in descending order. The decomposition (17) can be separated as

$$\mathbf{H} = \begin{bmatrix} \mathbf{U_s} & \mathbf{U_n} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_n} \end{bmatrix} \begin{bmatrix} \mathbf{V_s} & \mathbf{V_n} \end{bmatrix}^{\mathbf{T}} \quad (18)$$

$$= \mathbf{U_s}\mathbf{\Sigma_s}\mathbf{V_s^T} + \mathbf{U_n}\mathbf{\Sigma_n}\mathbf{V_n^T} = \mathbf{H_s} + \mathbf{H_n}$$

where $\mathbf{\Sigma_n}$ contains the smallest singular values. $\mathbf{H_n}$ matrix belongs to noise subspace. The matrix $\mathbf{H_s}$ is then related to the noise free signal subspace [14]. The signal matrix $\mathbf{H_s}$ is not precisely Hankel matrix, but some variation occurs in the antidiagonal elements. We reconstructed the noise free Hankel matrix by replacing the antidiagonal elements by their mean values. This enables the computation of the noise cancelled x_n $(n = 0, 1, ..., M)$ sequence. The dimension of the signal subspace is *N*, *i.e.*, the number of parallel RC-circuits in ESF. It should be pointed out that the Hankel matrix (16) must be a full matrix. Therefore *M* should be even.

3. Experimental

Extensive simulations were performed to validate the theoretical results. The αT parameter in (5) varied between 0.1-0.3. The number of the impulses in one burst varied between 3-7 and the number *N* in ESF (2) in the range 3-15. The amplitudes of the pulses were randomly distributed between the limits 0.1-0.9. The simulations warranted the condition that for the recovery of *p* pulses at least $2p$ samples are needed in the descending part of the ESF output. A typical simulation study is illustrated in **Figure 2**. In every case the ESF method recovered the amplitudes and the time locations of the pulses with a machine precision.

A prototype ESF was constructed with the aid of six parallel RC-circuits (**Figure 1**). The output was measured by a 16 bit analog-to-digital converter (ADC), which was equipped with a sample and hold amplifier (S/H). The input of the ESF was closed using the analog CMOS switch. The ESF was reset by grounding the output by an analog CMOS switch. Using a commercial UWB pulse generator two sequential pulses were fed to the input of the ESF and the descending part was measured with a 40 MHz sampling frequency. Due to the noise interference in practical measurements the use of the SVD based noise suppression method was essential. The prototype ESF recovered the amplitudes and the appearance times with a good accuracy. The mean error (standard deviation divided by the mean value) in amplitudes was 2.7% and in time locations 0.2%.

4. Discussion

The present work proposes a new approach for measuring of the impulse train using the ESF. The ESF yields exponentially descending pulses, which are sampled sequentially. In the beginning of the measurement period the input is closed and after measurement period the ESF is reset. We showed that using $2p$ parallel RC filters it is possible to reconstruct *p* impulses. The required number

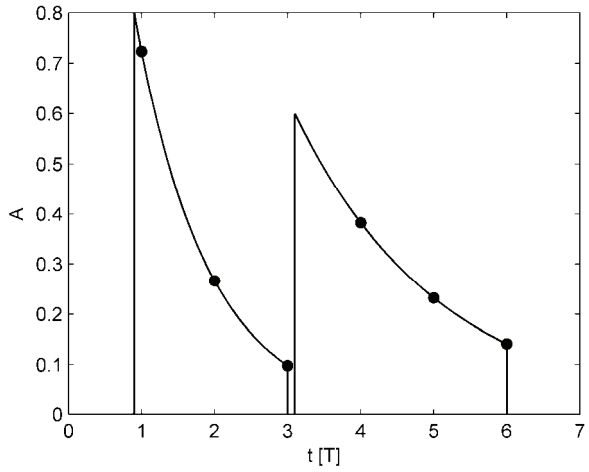


Figure 2. The minimum time interval $dt_i = t_i - t_{i-1}$ must be $dt_i > 2T$ for perfect reconstruction using ESF with two parallel RC circuits and three sequential samples per impulse. The dashed circles denote the measurement values and the solid line reconstructed ESF output via Equation (3).

of the samples was proved to be at least 2*p*. The simulations using noise free signal indicated an accurate and precise reconstruction property, in practice the ESF method showed a high sensitivity to the noise. To obtain a tolerable reconstruction error, the ESF circuit requires Faraday gage-type shielding and careful consideration of grounding and signal cables. The experimental verification of the effect of different types of noise sources (50 Hz + harmonics pick-up, 1/f-noise and random noise) needs further study.

In this work the noise was cancelled using the SVD based subspace method (16-18), which eliminates the noise interference if the number of samples exceeds 2*p*. The key idea is to compute the noise free signal as a mean of the antidiagonal elements in the signal subspace matrix $\mathbf{H}_s$. The error in the noise cancelled sequence $x_n(n = 0, 1, ..., M)$ attains a minimum in the middle of the sequence, where *M* data points are used for computation of the mean value. Evidently the higher error in both ends of the sequence is due to the lower number of points.

The ESF method has a close relationship with the FRI methods, which are based on the sequential measurement of the output of the analog circuit network [5]. The reconstruction properties of the Gaussian, sinc-type and triangle-wave sampling filters have been described [4-6, 8], but to the best of the authors knowledge, ESF-type circuits have not been previously used for the sampling and recovery of the UWB pulse sequences. A clear difference between the FRI sampling filters and the ESF comes from causality. An exponential impulse response $\exp[-\alpha k(t - t_i)]$ is causal and can be separated in (5) only if $t \geq t_i$. On the contrary, FRI sampling filters are not causal and this restriction is not needed in the deduction of the reconstruction algorithms. The FRI reconstruction algorithms usually require the solution of two matrix/vector equations, which are Yale-Walker and Vandermonde structures. In our approach three matrix/vector solutions (7, 14, 15) are required.

The present ESF measurement scheme can be considered as a general framework for measurement of the UWB pulse amplitudes and time locations. The experiments showed that the reconstruction error is mainly due to the additive random noise affecting on the amplitudes of the pulses. Hence, as an example of the practical construction of the wireless sensor network the ratio of the two sequential UWB pulse amplitudes can be related to the device address. The ratio of the UWB pulse amplitudes is not affected by the attenuation and distance variations. The time difference between the UWB pulses is reconstructed more accurately and it can be used to code the transmitted information.

The ESF method can be applied in many areas of wireless sensor technology and instrumentation. For example, in optical instruments the time difference between laser pulses can be used in testing light transmission in optical fibres. Many other sensors have pulse-type outputs.

5. Acknowledgements

This work was supported by the National Technology Agency of Finland (TEKES).

6. References

[1] J. Haupt and R. Nowak, "Signal Reconstruction from Noisy Random Projections," *IEEE Transactions Information Theory*, Vol. 52, No. 9, 2006, pp. 4036-4048.

[2] Y. C. Eldar and M. Unser, "Nonideal Sampling and Interpolation from Noisy Observations in Shift-Invariant Spaces," *IEEE Transactions Signal Processing*, Vol. 54, No. 7, 2006, pp. 2636-2651.

[3] M. Unser, "Sampling-50 Years after Shannon," *Proceedings of IEEE*, Vol. 88, No. 4, 2000, pp. 569-587.

[4] P. Marziliano, "Sampling Innovations," Ph.D. Dissertation, Communications laboratory, Lausanne, Switzerland, 2001.

[5] M. Vetterli, P. Marziliano and T. Blu, "Sampling Signals with Finite Rate of Innovation," *IEEE Transactions Signal Processing*, Vol. 50, No. 6, 2002, pp. 1417-1428.

[6] I. Maravic and M. Vetterli, "Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise," *IEEE Transactions Signal Processing*, Vol. 53, No. 8, 2005, pp. 2788-2805.

[7] P. Marziliano, M. Vetterli and T. Blu, "Sampling and Exact Reconstruction of Bandlimited Signals with Additive

Shot Noise," *IEEE Transactions Information Theory*, Vol. 52, No. 5, 2006, pp. 2230-2233.

[8] P. L. Dragotti, M. Vetterli and T. Blu, "Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang-Fix," *IEEE Transactions Signal Processing*, Vol. 55, No. 5, 2007, pp. 1741-1757.

[9] I. Jovanovic and B. Beferull-Lozano, "Oversampled A/D Conversion and Error-Rate Dependence of Nonband Limited Signals with Finite Rate of Innovation," *IEEE Transactions Signal Processing*, Vol. 54, No. 6, 2006, pp. 2140-2154.

[10] Y. P. Nakache and A. F. Molisch, "Spectral Shaping of UWB Signals for Time-Hopping Impulse Radio," *IEEE Journal of Selected Areas of Communications*, Vol. 24, No. 4, 2006, pp. 738-744.

[11] B. Parr, B. Cho, K. Wallace and Z. Ding, "A Novel Ultra-Wideband Pulse Design Algorithm," *IEEE Communications Letters*, Vol. 7, No. 5, 2003, pp. 219-221.

[12] M. Miao and C. Nguyen, "On the Development of an Integrated CMOS-Based UWB Tunable-Pulse Transmit Module," *IEEE Transactions Microwave Theory and Techniques*, Vol. 54, No. 10, 2006, pp. 3681-3687.

[13] V. E. Neagoe, "Inversion of the Van Der Monde Matrix," *IEEE Signal Processing Letters*, Vol. 3, No. 4, 1996, pp. 119-120.

[14] E. Biglieri and K. Yao, "Some Properties of Singular Value Decomposition and their Applications to Digital Signal Processing," *Signal Processing*, Vol. 18, No. 3, 1989, pp. 277-289.

Wireless Sensor Network, **2010, 2, 467-471**
doi:10.4236/wsn.2010.26058 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Research on Application of ZigBee Technology in Flammable and Explosive Environment

Yang Li[1,2], Ke Zhang[2]
[1]*Beijing Jiaotong University, Beijing, China*
[2]*Beijing institute of Petro-chemical Techmology, Beijing, China*
E-mail: *liyang@bipt.edu.cn*
Received April 4, 2010; *revised May* 6, 2010; *accepted May* 15, 2010

Abstract

Wireless Sensor Network based on ZigBee technology is a wireless network which is composed of many nodes of ZigBee RF chips, sensors and MCUs, especially suitable for application of the remote monitoring system in flammable and explosive environments. This paper presents the characteristics and advantages of ZigBee technology, also discusses the system for hardware and software design. This system effectively fulfills the remote monitoring in flammable and explosive environments and posses high practical values.

Keywords: ZigBee, Wireless Sensor Network, Remote Monitoring, Flammable and Explosive

1. Introduction

In the oil, chemicals and other inflammable and explosive production environment, it is essential to the production site, some information (such as pressure, temperature, gas concentration, etc.) for data collection and transmission network in order to achieve remote monitoring and control. At present, widely used by cable way through all kinds of information sent to the monitoring center, however, most point of the scene to monitor geographical dispersion, environment, complex terrain, which is encountered in many practical applications: 1) To set up a qualified cable transmission network, there is the complexity of the construction and implementation of the difficulty; 2) In order to achieve all-round production of effective environmental monitoring, dispersed layout must be many types, the large number of nodes to monitor the data acquisition, cable monitoring is often difficult to achieve; 3) The cable monitoring system has its own limitations, such as the laying of a high fixed cost of communication lines. In recent years, the emergence of ZigBee technology has provided some very good ideas to solve these problems mentioned above.

In this paper, the ZigBee technology wireless sensor network system for remote monitoring the production environment of the explosive have been put forward and designed. The system layout of the scene in the production of the sensor node to all kinds of information can be wirelessly sent to the central node, the center node of data through GPRS or RS232 interface module to the monitoring host (PC machine) for remote monitoring of the production environment. The monitoring system has the advantages of low cost, low power consumption, wireless transmission, and reliable communications.

1.1. ZigBee Technology

ZigBee is a relatively recent emergence of wireless network communication technology, more than 100 well-known by the global coalition of hardware and software companies committed to the development of a short-range, low rate, low-power wireless network standards, the main development direction for the wireless sensor networks, home automation, remote control, industrial automation, agricultural automation, and medical care and other applications.

1.1.1. ZigBee Technology Features

Construction of ZigBee-based wireless network technology has the following features:

1) Data transfer rate low: Only 20 k bytes/s to 250 k bytes/sec, to focus on low-delivery applications;

2) Low power consumption: Due to the use of DSSS technology ZigBee replace FHSS technology, and use of hibernation wake-up mechanism for the work of machines, two on the 5th ordinary dry-cell batteries can be used for 6 months to 2 years, which eliminates the frequent replacement of the battery charge or trouble;

3) Low cost: because of low data rate ZigBee, the agreement is simple and royalty-free, so greatly reduced cost;

4) Network capacity: ZigBee Network Node Manager

may be a number of sub-node, a node can manage up to 254 sub-node. At the same time, the node can be from one network node management, can be composed of 65,536 large-scale network nodes;

5) The short time delay: delay-sensitive applications for optimized, communication delay and activated from hibernation is very short delay, typically 15 ms latency to 30 milliseconds;

6) Safety: ZigBee provides the data integrity checks and authentication functions, the use of a common encryption algorithm AES2128, while the flexibility to determine their security attributes;

7) Reliable: the mechanism used to avoid collisions, as well as the need for a fixed bandwidth communications business set aside a dedicated time slot, when sending data to avoid competition and conflict; node module automatically between the functions of dynamic network information in the entire ZigBee networks the way through the automatic route for transmission, thus ensuring the reliability of information transmission;

8) Flexible working band: the use of the Channel 2.4 GHz, 868 MHz (Europe) and 915 MHz (USA), are license-free band.

1.1.2. ZigBee Architecture

ZigBee protocol stack structure as shown in **Figure 1**, IEEE802.15.4 responsible for the physical layer and MAC layer protocol, ZigBee Alliance to develop the network layer, application layer security services interface and protocol, can be user-defined application layer.

1) Physical Layer: the bottom is the agreement to assume a direct role in working with the task of the outside world, is responsible for the modulation data, send and receive work. IEEE 802.15.4 in the physical layer (PHY) defined two criteria, namely 2.4 GHz and 868/915 MHz PHY physical layer. The band can be used in the respective channel 16, 10 or 1, each provision of 250 kb/s, 40 kb/s and 20 kb/s transfer rate, the physical layer are used in various band direct sequence spread spectrum technology.

2) MAC Layer: The IEEE 802.15.4 protocol definition, the use of the CSMA-CA mechanism to avoid the collision, and its features include wireless links between devices to establish, maintain and disconnect, to confirm the frame mode to send and receive, channel access and control, and rapid automatic Frame Check request reissued, reserved slot management and information management, such as broadcasting. The definition of a radio frame layer, data frames, frames and MAC command confirmation frame, such as four kinds of frame types.

3) Network/Security Layer: The main mechanism responsible for the establishment of networks and management, self-configuration and self-repair function, the realization of the node to join or leave the network, receive or discard the other nodes, as well as to locate and

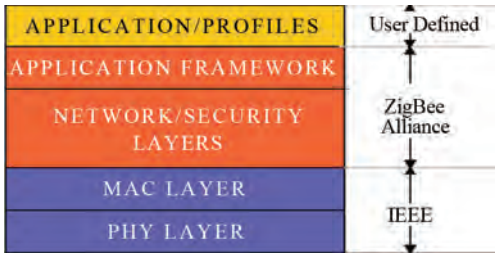


Figure 1. IEEE 802.15.4 stack.

transmit data routing functions, supports a wide range of routing algorithm and a variety of network topology.

4) Application Interface Layer: is responsible for the different applications mapped to the ZigBee network, including security attributes to set up and a number of business data flow and other functions together.

5) Application Layer: The principal objective is the realization of the network communication between different devices, applications and settings for access to information services, calls for the application layer protocol to provide continuous and discrete control applications and other support.

1.1.3. ZigBee Wireless Network Type

ZigBee network topology has three: star, cluster tree and mesh type as shown in **Figure 2**. Among them, the star-shaped network is a network-based control center, by a coordinator node and a number of terminal nodes, terminal nodes realized through co-ordination among the communication; cluster tree network has increased the concept of routing, terminal node coordinator can not only access node can access any node has a routing function, it has routing node can not be direct communication between each other only through the co-ordination of the routing nodes between the completion of the communication; mesh network to provide a more flexible mechanism, through the self-organizing routing and wireless data communications to provide multiple paths, when the optimal communication path failure, a redundant mesh network in the other path to choose the most appropriate the path for data communication, therefore, effectively reduce the net-work structure of information transmission delay and improve the reliability of the network communications.

Know from the above, ZigBee network node from the logic function can be divided into co-ordination, routers and terminal equipment, and physical properties from the ZigBee network nodes can be divided into full-featured and simple equipment FFD function device RFD. Among them, the full-function device can act as a coordinator, router or terminal equipment, and can only serve as a simple function of terminal equipment devices.

Each ZigBee network must include a co-ordinator, ZigBee is the network coordinator of the center responsible for the organization and maintenance of the network by

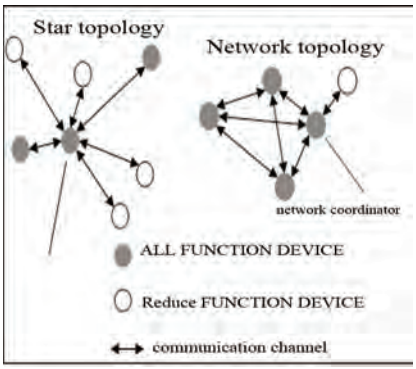


Figure 2. ZigBee topology.

adding a new node and the allocation of l6-bit short address; Zigbee routers responsible for routing node, and used to expand the scope of the network; and the terminal equipment is to achieve a specific functional unit.

2. Explosive Production Environment Remote Monitoring System Architecture

System structure shown in **Figure 3**, the entire system by monitoring the host, GPRS module (or, a ZigBee coordinator node, a number of ZigBee routers, ZigBee nodes and a number of nodes of terminal equipment. This is a cluster tree network structure is conducive to the number of network nodes and the physical expansion of the scope, complex, multi-node wireless network communication system is also an important reference value.

The co-ordination of the network nodes, network management functions, the receiving terminal device node for the data upload, and transfer through the GPRS network to the monitoring center. Router nodes for routing of information transmitted, allowing other nodes join the network. Node device to the network coordinator from time to time collect information to send and receive commands from the monitoring host. ZigBee module used for GPRS networks and Internet networks, the Internet (also available in other ways), the realization of ZigBee network datas to monitor the upload and download the host commands. Host real-time monitoring of the collection, storage, monitoring and processing equipment from a remote terminal nodes of information, and can overrun the police at any time, such as setting parameters for the production environment to achieve effective monitoring and management, its functions are divided into two major parts, 1) Data Monitoring: to receive from the ZigBee network information collected, the corresponding data into the database; to receive instructions from the managers, and command frame format in accordance with the configuration commands, GPRS module through the command issued to the ZigBee network and do the action. 2) Data Management: The database can be found, query data from the current

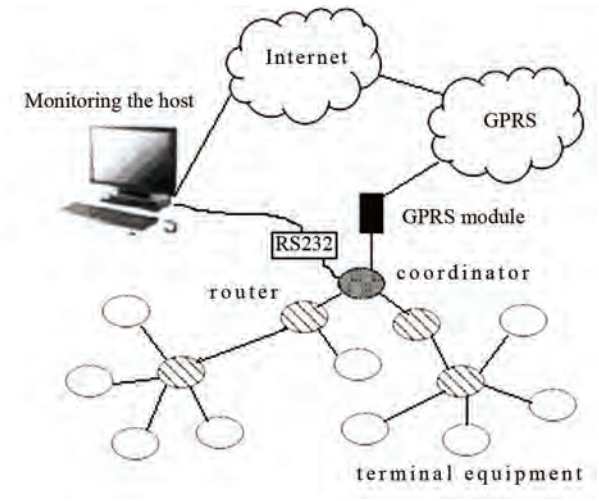


Figure 3. Structure of the framework of remote monitoring system.

ZigBee network information, such as: the production of the ambient temperature, pressure, overrun alarm, such as the peak period.

ZigBee end-node using the occasional wake-up call from time to time work, time to wake up from hibernation to start data acquisition, ZigBee routing node to send a message, send completed and then enter hibernation. ZigBee routing nodes will collect the data sent to the ZigBee coordinator node, gateway GPRS module through the data uploaded to the remote monitoring center.

3. System Hardware Design

Explosive production environment is a remote monitoring system by a number of ZigBee network node. Each node is basically the same hardware structure, but not the same network layer. ZigBee hardware in each node has two basic components: micro-controller and wireless receiver send some. Hardware-specific features into the single chip to achieve by the burning process to decide.

To facilitate the design and cost savings, the system uses a wireless transceiver and controller integrated with the CC2430 single-chip solution module, the module from a Norwegian company Chipcon is in line with standard IEEE 802.15.4-chip ZigBee products. It incorporates a single chip ZigBee radio frequency (RF) front-end, memory and microcontroller. It uses an 8-bit MCU (8051), and 128 kb with a programmable flash memory and 8 kb of RAM, also includes analog-to-digital converter (ADC), 4 timers (Timer), watchdog timer, 32 kHz crystal oscillator of the sleep mode timer, power-on reset circuit, power-fail detection circuit, as well as 21 programmable I/O pin. Its block diagram is shown in **Figure 4**.

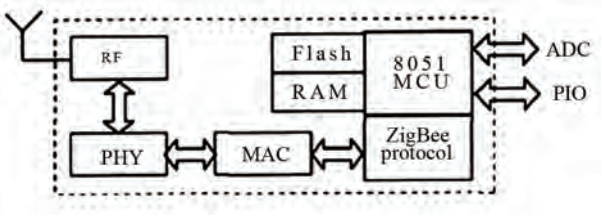


Figure 4. CC2430 structure.

CC2430 network, send and receive data is built on the ZigBee protocol stack based on. If we are to use the CC2430 ZigBee chip, first of all be familiar with the ZigBee protocol stack (see **Figure 1**), based on the actual needs of the application layer programming. At present, TI has fully disclosed the ZigBee protocol stack CC2430 proceedings; it is relatively easy program development process to achieve. FLASH memory in the CC2430 in the internal processor to run the application, when the system is activated, the chip will be stored in FLASH in the program in the implementation of the RAM.

CC2430 chip with only a small number of external components to send and receive signals will be able to achieve functional, very simple and practical. ZigBee nodes of the system hardware structure diagram as shown.

According to actual needs of each node can choose a number of different sensors (such as temperature, pressure, etc.) to meet the requirements of the monitoring points. Coordinator node GPRS module also need to host and monitor connected to RS232 interface can also be used to connect the host and monitor (if communication from the close), chip-level conversion MX3232. Node power circuit terminals using two alkaline batteries on the 5th, as the coordinator node transceiver has been in a state, so the use of external power supply.

4. System Software Design

ZigBee protocol stack to provide a number of Application Programming Interface, such as aplFormNetwork (), aplJoinNetwork (), aplSendMSG() function and so on, the user can call these functions to prepare their own applications.

System Software primarily to achieve two basic functions: 1) to achieve a co-ordinator node and other network nodes and communications; 2) cycle to achieve the various nodes and send sensor data acquisition.

4.1. Node Design Coordinator

Network as the network coordinator of the center is responsible for the establishment of a network, information reception, aggregation, processing and sending control instructions and implementation. Coordinator power to start the procedure after the initialization, by calling the function aplFormNetwork ()create a network, select a Coordinator PANID as a network logo, create a routing table, and then released to inform the other routers broadcast frame or a terminal node The addition of node equipment. Start by sending a GPRS module AT command set serial communication rate, the establishment of data communication socket connections ready to begin to send and receive data and instructions to implement various operations. Process flow chart is shown in **Figure 6**.

4.2. Terminal Node Programming

Terminal node network is mainly responsible for a variety of information (such as temperature, pressure, etc.) to collect data to send and receive commands to control the implementation. End-node power-start initialization process

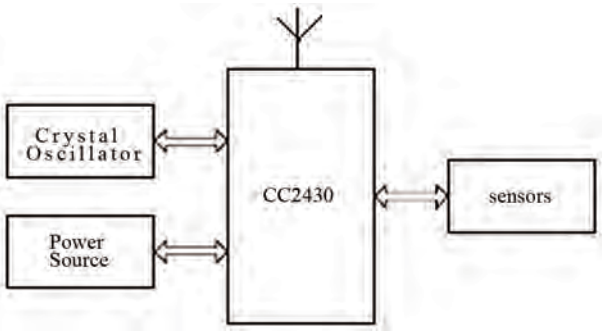


Figure 5. ZigBee node hardware structure.

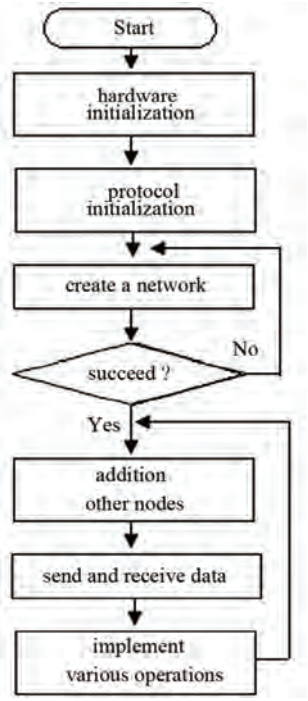


Figure 6. Flow chart of the program coordinator node.

by calling the function aplJoinNetwork () to join the network, active and effective network of channel scanning, the nearest coordinator to find a suitable node or router node apply to join the network, access is approved, the beginning of information gathering send and receive instruction implementation. Terminal node to take a cyclical mode, do not work in a dormant, down, low-power to achieve energy-saving effect.

Part of the node device code is as follows:

```
void main (void){
  hallnit（）；  // hardware initialization
  apllnit（）；  // protocol stack initialization
  ......
  do{
  aplJoinNetwork（）；  // join the network
  while（apsBusy（））{ apsFSM（）；}
  }
  while （aplGetStatus（）＝WXLPAN_STATUS_
SUCCESS）；
  while（1） {apsFSM（）；}
  ......
}
```

5. Conclusions

In this paper, the design of ZigBee technology production environment explosive remote monitoring system, the degree of coverage cluster tree network structure, node chip CC2430, has flexible, economic practicability, design easy. If the target needs to be identified in accordance with the actual sensor types and routers and the number of terminal nodes, the system can be applied to the production and living more occasions to address the practical application of the cost of wired network cabling is too high, not the arrival of the regional environment kinds of information to monitor the problem.

6. References

[1] J. A. Gutierrez, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs), (IEEE Standarted for Information Technology802.15.4.)," Institute of Electrical & Electronics, 2003.

[2] P. S. Neelakanta and H. Dighe, "Robust Factory Wireless Communicat Ions: A Performance Appraisal of the Bluetooth and the ZigBee Collocated on an Indust Rial Floor," *IEEE Computer Society*, Vol. 3, 2003, pp. 2381-2386.

[3] Chipcon, "CC2430 Preliminary Data Sheet (Rev. 1.03) SWRS036A," Chipcon, 2005.

[4] R. Reese, "A ZigBee TM-Subset/IEEE 802.15.4 TM Multiplatform Protocol Stack," Electrical/Computer Engr MSU, 2006.

[5] "ZigBee Protocol Specification," http://www.ZigBee.org

Wireless Sensor Network, **2010, 2, 472-482**
doi:10.4236/wsn.2010.26059 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Interference Management for DS-CDMA Systems through Closed-Loop Power Control, Base Station Assignment, and Beamforming

Mohamad Dosaranian Moghadam[1], Hamidreza Bakhshi[2], Gholamreza Dadashzadeh[2]
[1]*Department of Electrical Engineering, Science and Research Branch of Islamic Azad University, Tehran, Iran*
[2]*Department of Electrical Engineering, Shahed University, Tehran, Iran*
E-mail: *m_dmoghadam@qiau.ac.ir*, {*bakhshi*, *gdadashzadeh*}*@shahed.ac.ir*
Received April 8, 2010; *revised May* 15, 2010; *accepted May* 23, 2010

Abstract

In this paper, we propose a smart step closed-loop power control (SSPC) algorithm and a base station assignment method based on minimizing the transmitter power (BSA-MTP) technique in a direct sequence-code division multiple access (DS-CDMA) receiver with frequency-selective Rayleigh fading. This receiver consists of three stages. In the first stage, with constrained least mean squared (CLMS) algorithm, the desired users' signal in an arbitrary path is passed and the inter-path interference (IPI) is reduced in other paths in each RAKE finger. Also in this stage, the multiple access interference (MAI) from other users is reduced. Thus, the matched filter (MF) can use for more reduction of the IPI and MAI in each RAKE finger in the second stage. Also in the third stage, the output signals from the matched filters are combined according to the conventional maximal ratio combining (MRC) principle and then are fed into the decision circuit of the desired user. The simulation results indicate that the SSPC algorithm and the BSA-MTP technique can significantly reduce the network bit error rate (BER) compared to the other methods. Also, we observe that significant savings in total transmit power (TTP) are possible with our methods.

Keywords: Adaptive Beamforming, Antenna Array, Base Station Assignment, Closed-Loop Power Control, Constrained LMS, DS-CDMA

1. Introduction

Code-Division Multiple Access (CDMA) for cellular communication networks requires the implementation of some forms of adaptive power control. In uplink of CDMA systems, the maximum number of supportable users per cell is limited by multipath fading, shadowing, and near-far effects that cause fluctuations of the received power at the base station (BS). Two types of power control are often considered: closed-loop power control and open-loop power control [1,2]. In a closed-loop power control, according to the received signal power at a base station, the base station sends a command to a mobile set to adjust the transmit power of the mobile. Also, closed-loop power control is employed to combat fast channel fluctuations due to fading. Closed-loop algorithms can effectively compensate fading variations when the power control updating time is smaller than the correlation time of the channel. However, in an open-loop power control, a mobile user adjusts its transmit power according to its received power in downlink [1-5]. In this paper, an adaptive closed-loop power control algorithm is proposed to compensate for near-far effects.

Diversity and power control are two effective techniques for enhancing the signal-to-interference-plus-noise ratio (SINR) for wireless networks. Diversity exploits the random nature of radio propagation by finding independent (or, at least, highly uncorrelated) signal paths for communication. If one radio path undergoes a deep fade, another independent path may have a strong signal. By having more than one path to select from, the SINR at the receiver can be improved. The diversity scheme can be divided into three methods: 1) The space diversity; 2) The time diversity; 3) The frequency diversity. In these schemes, the same information is first received (or transmitted) at different locations (or time slots/frequency bands). After that, these signals are combined to increase the received SINR. The antenna array is an example of the space diversity, which uses a beamformer to increase the SINR for a particular direction [6-8].

The first goal of this paper is to extend the works in [9] and [10] by considering multiple-cell system and closed-loop power control. In these works, a RAKE receiver in single-cell system with conjugate gradient adaptive beamforming was proposed in the presence of frequency-selective Rayleigh fading channel, and perfect power control (PPC) was considered.

In this work, the performance analysis of direct sequence (DS)-CDMA system in frequency-selective Rayleigh fading channel has been studied. If the delay spread in a multipath channel is larger than a fraction of a symbol, the delayed components will cause inter-symbol interference (ISI). Adaptive receiver beamforming schemes have been widely used to reduce both co-channel interference (CCI) and ISI and to decrease the bit error rate (BER) by adjusting the beam pattern such that the effective SINR at the output of the beamformer is optimally increased [11].

In this paper a RAKE receiver in DS-CDMA system is analyzed in three stages according to **Figure 1** [9]. In the first stage, this receiver uses constrained least mean squared (CLMS) adaptive beamforming algorithm to find optimum antenna weights assuming perfect estimation of the channel parameters (direction, delay, and power) for the desired user. The desired user resolvable paths' directions are fed to the beamformer to reduce the inter-path interference (IPI) from other directions. Also, the RAKE receiver uses conventional demodulation in the second stage and conventional maximal ratio combining (MRC) in the third stage to reduce multiple access interference (MAI) and the other interferences. Reducing the MAI and CCI will further decrease the system BER.

To improve the performance of cellular systems, base station assignment (BSA) technique can be used. In the joint power control and base station assignment, a number of base stations are potential receivers of a mobile transmitter. Here, the objective is to determine the assignment of users to base stations which minimizes the allocated mobile powers [12-15]. In simple mode and in multiple-cell systems, the user is connected to the nearest base station. This way is not optimal in cellular systems under the shadowing and multipath fading channels and can increase the system BER.

Accordingly, the second goal of this paper is to use base station assignment technique. In [14], the combined the base station assignment and power control was used to increase uplink capacity in cellular communication networks. In that work, it was shown that if there exists at least one feasible base station assignment, the proposed algorithm will find the jointly optimal base station assignment and minimal transmitter power level for all users. In this paper, we present the base station assignment method based on minimizing the transmitter power (BSA-MTP) for decreasing the BER in all cells.

The organization of the remainder of this paper is as follows. The system model is presented in Section 2. The RAKE receiver structure is described in Section 3. In Section 4, we propose smart step closed-loop power control (SSPC) algorithm. In Section 5, the BSA-MTP technique is presented. Section 6 describes switched-beam (SB) technique and equal sectoring (ES) method. Finally, simulation results and conclusions are given in Section 7 and Section 8, respectively.

2. System Model

In this paper, we focus on the uplink communication paths in a DS-CDMA cellular system. L replicas of the signal, due to both some form of diversity reception (for instance antenna diversity) and channel frequency selectivity, are assumed Rayleigh distributed and optimally combined through a RAKE receiver according to **Figure 1**.

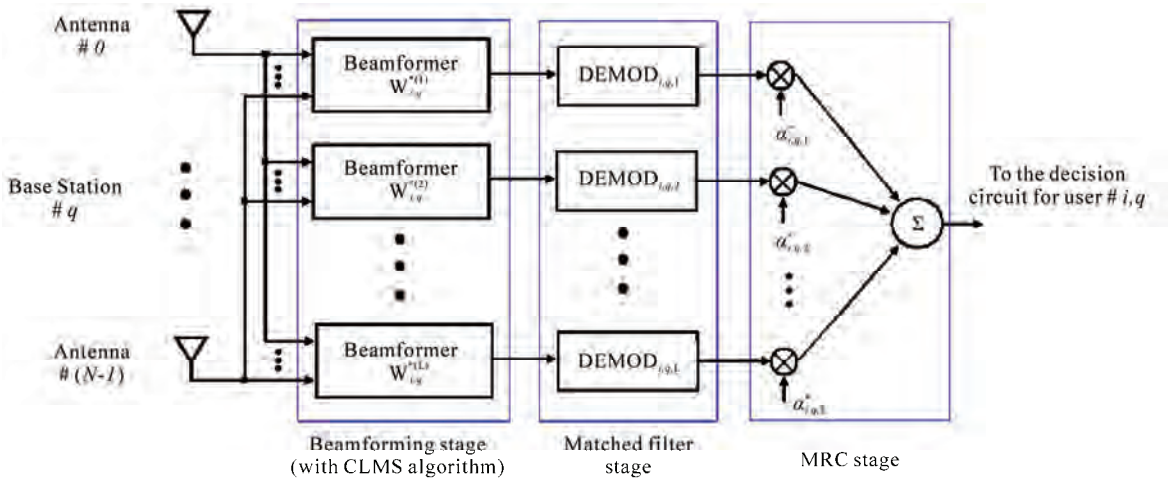


Figure 1. Block diagram of a three-stage RAKE receiver in DS-CDMA system [9].

Also assume that there are M active base stations in the network, with K_m users connected to m th base station, where $1 \le m \le M$. Also assume that each base station uses an antenna array of S sensors and N weights, where $S = N$, to receive signals from all users. Also, for simplicity we assume a synchronous DS-CDMA scheme and BPSK modulation in order to simplify the analysis of proposed methods. Additionally, in this paper we assume a slow fading channel. Hence, the received signal in the base station q and sensor s from all users can be written as [9,16]

$$r_{q,s}(t) = \sum_{k} \sqrt{p'_{k,m}\Gamma_k(x,y)} \sum_{l=1}^{L} \alpha_{k,m,l} b_{k,m}\left(t-\tau_{k,m,l}\right) \times c_{k,m}\left(t-\tau_{k,m,l}\right) \exp\left(-j\, s\, k_d \sin\theta_{k,m,l}\right) + n(t) \quad (1)$$

where $c_{k,m}(t)$ is the pseudo noise (PN) chips of user k in cell m (user k,m) with a chip period of T_c; $b_{k,m}(t)$ is the information bit sequence of user k,m with a bit period of $T_b = GT_c$ where G is processing gain; $\tau_{k,m,l}$ is the l th path time delay for user k,m; $\theta_{k,m,l}$ is the direction of arrival (DoA) in the l th path for user k,m; $\alpha_{k,m,l}$ is the complex Gaussian fading channel coefficient from the lth path of user k,m; $k_d = 2\pi d/\lambda$ where λ is signal wavelength and d is the distance between the antenna elements that for avoid the spatial aliasing should be defined as $d = 0.5\lambda$ and $n(t)$ is an additive white Gaussian noise (AWGN) process with a two-sided power spectral density (PSD) of $N_0/2$. Also for conventional BSA technique, $\Gamma_k(x,y)$ is defined as

$$\Gamma_k(x,y) = \begin{cases} 1 & ;k \in S_{\mathrm{BS}q} \\ \dfrac{\min\limits_{m\in\Theta_k}\left\{ d_{k,m}^{L_\alpha}(x,y) 10^{\xi_{k,m}/10} \right\}}{d_{k,q}^{L_\alpha}(x,y) 10^{\xi_{k,q}/10}} & ;k \in S_o \end{cases} \quad (2)$$

where L_α is path-loss exponent; $d_{k,m}(x,y)$ and $d_{k,q}(x,y)$ are the distance between user k and BS m and BS q, respectively (see **Figure 2**). Also the variable Θ_k defined the set of the nearest BS*s* to user k; $\xi_{k,m}$ is a random variable modeling the shadowing between user k and BS m; $S_{\mathrm{BS}q}$ is the set of users that connected to BS q and S_o is the set of users that not connected to BS q [2]. Also in (1)

$$p'_{k,m} = d_{k,m}^{-L_\alpha}(x,y) 10^{-\xi_{k,m}/10} \times p_{k,m} \quad (3)$$

is the received power in the BS m of user k,m in the presence of closed-loop power control where $p_{k,m}$ is the transmitted power of user k,m that in the case of the PPC, $p'_{k,m}$ is fixed for all users within cell m ($p'_{k,m} = E_b/T_b$ where E_b is the energy per bit for all users) [2,9].

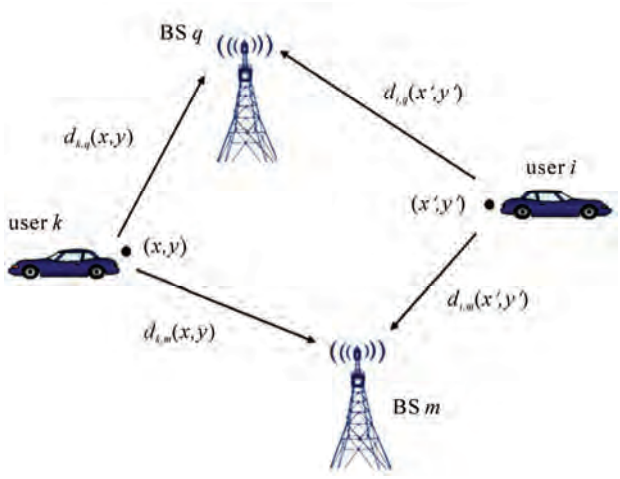


Figure 2. The distance between two pairs of mobile transmitters and base station receivers [12].

The received signal in the base station q in sensor s for user i,q is given by [9]

$$r'_{i,q,s}(t) = \sum_{l=1}^{L} \sqrt{p'_{i,q}} b_{i,q}\left(t-\tau_{i,q,l}\right) c_{i,q}\left(t-\tau_{i,q,l}\right) \alpha_{i,q,l} \times \exp\left(-j\, s\, k_d \sin\theta_{i,q,l}\right) + I_{i,q,s}(t) + n(t) \quad (4)$$

where $I_{i,q,s}(t)$ is the interference for user i,q in sensor s and can be shown to be

$$I_{i,q,s}(t) = \sum_{m=1}^{M} \sum_{\substack{k=1 \\ k,m\neq i,q}}^{K_m} \sum_{l=1}^{L} \sqrt{p'_{k,m}\Gamma_k(x,y)} b_{k,m}\left(t-\tau_{k,m,l}\right) \times c_{k,m}\left(t-\tau_{k,m,l}\right) \alpha_{k,m,l} \exp\left(-j\, sk_d \sin\theta_{k,m,l}\right) \quad (5)$$

where K_m is the number of users in cell m and M is the number of base stations/cells.

3. RAKE Receiver Performance Analysis

The RAKE receiver structure in the DS-CDMA system is shown in **Figure 1**. The received signal is spatially processed by a beamforming circuit with CLMS algorithm, one for each resolvable path (L beamformers). The resultant signal is then passed on to a set of parallel matched filters (MFs), on a finger-by-finger basis. Also, the output signals from the L matched filters are combined according to the conventional MRC principle and then are fed into the decision circuit of the desired user [9].

3.1. Constrained LMS Algorithm

It is well known that an array of N weights has $N-1$ degree of freedom for adaptive beamforming [9,16]. This means that with an array of N weights, one can generates $N-1$ pattern nulls and a beam maximum in desired directions. From (5), it is clear that the number of users is $K_u = \sum_{m=1}^{M} K_m$ and the number of interferes is $LK_u - 1$. To null all of these interferes; one would have to have LK_u weights, which is not practical. So, we focus only on the L paths of the desired user. Thus, the minimum number of the antenna array weights is L where, typically, L varies from 2 to 6 [9].

In this paper, we use the CLMS adaptive beamforming algorithm. This algorithm is a gradient based algorithm to minimize the total processor output power, based on the look direction constraint. The adaptive algorithm is designed to adapt efficiently in agreement with the environment and able to permanently preserve the desired frequency response in the look direction while minimizing the output power of the array. The combined form of the constraint is called constraint for narrowband beamforming [12,17].

This form consider a narrowband beamformer where the narrowband signal from each element of smart antenna are multiplied by the complex weight calculated by using narrowband adaptive beamforming algorithm, and then summed to produce the output of the array. The definition of the complex weights of this beamformer in the n th iteration for user i,q in the j th path is as follows [16,18].

$$\mathbf{w}_{i,q}^{(j)}(n) = \left[w_{i,q,0}^{(j)}(n)\ w_{i,q,1}^{(j)}(n) \ldots w_{i,q,N-1}^{(j)}(n) \right]^T \quad (6)$$

Accordingly, the output of the array in the n th iteration in the j th path for user i,q is given by

$$y_{i,q}^{(j)}(n) = \mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{r}'_{i,q}(n) \quad (7)$$

where $\mathbf{r}'_{i,q} = \left[r'_{i,q,0}\ \ r'_{i,q,1}\ \ \ldots\ \ r'_{i,q,N-1} \right]^T$.

The expected output power of the array in the n th iteration is given by

$$\begin{aligned} \mathrm{E}\left(\left| y_{i,q}^{(j)}(n) \right|^2 \right) &= \mathrm{E}\left(y_{i,q}^{(j)}(n) y_{i,q}^{(j)}(n)^* \right) \\ &= \mathrm{E}\left(\mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{r}'_{i,q}(n) \mathbf{r}'_{i,q}(n)^H \mathbf{w}_{i,q}^{(j)}(n) \right) \\ &= \mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{R}_{r'r'} \mathbf{w}_{i,q}^{(j)}(n) \end{aligned} \quad (8)$$

where $\mathrm{E}(.)$ is denoted the expectation and $\mathbf{R}_{r'r'}$ is the correlation matrix of the received vector $\mathbf{r}'_{i,q}(n)$.

A real-time CLMS algorithm for determining the optimal weight vector for user i,q in the j th path is [17,18]:

$$\begin{cases} \mathbf{w}_{i,q}^{(j)}(n+1) = \mathbf{w}_{i,q}^{(j)}(n) + \mu\, g\left(\mathbf{w}_{i,q}^{(j)}(n) \right) \\ \mathbf{w}_{i,q}^{(j)H} \mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right) = 1 \end{cases} \quad (9)$$

where

$$\begin{aligned} \mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right) = [1\ \ &\exp\left(-j\, k_d \sin\theta_{i,q,j} \right) \ \ldots \\ &\ldots \exp\left(-j\, k_d (N-1) \sin\theta_{i,q,j} \right)]^T \end{aligned} \quad (10)$$

denotes spatial response of the array for user i,q in the j th path. Also in (9), $\mathbf{w}_{i,q}^{(j)}(n+1)$ is the new weight computed at the $(n+1)$ th iteration for user i,q in the j th path. Also, the variable scalar μ denotes a positive scalar (gradient step size) that controls the convergence characteristic of the algorithm, that is, how fast and how close the estimated weights approach the optimal weights, and $g\left(\mathbf{w}_{i,q}^{(j)}(n) \right)$ denotes an unbiased estimate of the gradient of the power surface ($\mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{R}_{r'r'} \mathbf{w}_{i,q}^{(j)}(n)$) which is the expected output power of the array) with respect to $\mathbf{w}_{i,q}^{(j)}(n)$ after the *n*th iteration. The algorithm is "constrained" because the weight vector satisfies the constraint at each iteration, that is $\mathbf{w}_{i,q}^{(j)H} \mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right) = 1$. Rewrite the CLMS algorithm as follows [17].

$$\mathbf{w}_{i,q}^{(j)}(n+1) = \boldsymbol{\beta}_{i,q}^{(j)} \left(\mathbf{w}_{i,q}^{(j)}(n) + \mu\, g\left(\mathbf{w}_{i,q}^{(j)}(n) \right) \right) + \frac{\mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right)}{N} \quad (11)$$

where

$$\boldsymbol{\beta}_{i,q}^{(j)} = \mathbf{I} - \frac{\mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right) \mathbf{a}_{i,q}^{(j)}\left(\theta_{i,q,j} \right)^H}{N} \quad (12)$$

The gradient of $\mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{R}_{r'r'} \mathbf{w}_{i,q}^{(j)}(n)$ with respect to $\mathbf{w}_{i,q}^{(j)}(n)$ is given by [17]

$$\begin{aligned} g\left(\mathbf{w}_{i,q}^{(j)}(n) \right) &\stackrel{\Delta}{=} -\frac{\partial}{\partial \mathbf{w}_{i,q}^{(j)*}} \left(\mathbf{w}_{i,q}^{(j)}(n)^H \mathbf{R}_{r'r'} \mathbf{w}_{i,q}^{(j)}(n) \right) \\ &= -2\mathbf{R}_{r'r'} \mathbf{w}_{i,q}^{(j)}(n) \end{aligned} \quad (13)$$

and its computation using this expression requires knowledge of $\mathbf{R}_{r'r'}$, which normally is not available in practice. For a standard LMS algorithm, an estimate of the gradient at each iteration is made by replacing $\mathbf{R}_{r'r'}$ by its noise sample $\mathbf{r}'_{i,q}(n+1)\mathbf{r}'_{i,q}(n+1)^H$ available at time instant $(n+1)$, leading to

$$g\left(\mathbf{w}_{i,q}^{(j)}(n) \right) = -2\mathbf{r}'_{i,q}(n+1)\, y_{i,q}^{(j)*}(n) \quad (14)$$

The CLMS is a fast convergence algorithm. However, it is drastically sensitive to the mismatch in the direction of arrival. Meanwhile, the weights estimated by the standard algorithm are sensitive to the signal power, requireing a lower step size in the presence of a strong signal for the algorithm to converge, which in turn regarding the decrease of mis-adjustment error, the convergence time is increased [17,19].

It should be mentioned that for the antenna arrays weight vector in the CLMS algorithm and for big μ, will converge after a few iteration (is approximately equal to the number of beamformer weights, *i.e.*, $n = N$) [19].

Accordingly, the output signal from the j th beamformer ($j = 1,...,L$) can be written as [9]

$$y_{i,q}^{(j)}(t) = \sqrt{p'_{i,q}}\, b_{i,q}\left(t-\tau_{i,q,j}\right)c_{i,q}\left(t-\tau_{i,q,j}\right)\alpha_{i,q,j} + \tilde{I}_{i,q}^{(j)}(t) + I_{i,q}^{(j)}(t) + n^{(j)}(t) \quad (15)$$

where $n^{(j)}(t)$ is a zero mean Gaussian noise of variance σ_n^2 and $\tilde{I}_{i,q}^{(j)}(t)$, the IPI, is defined as

$$\tilde{I}_{i,q}^{(j)}(t) = \sum_{\substack{l=1\\ l\neq j}}^{L}\sqrt{p'_{i,q}}\; g_{i,q}^{(j)}\left(\theta_{i,q,l}\right)\alpha_{i,q,l}\, b_{i,q}\left(t-\tau_{i,q,l}\right) \times c_{i,q}\left(t-\tau_{i,q,l}\right) \quad (16)$$

and $I_{i,q}^{(j)}(t)$, the MAI, is defined as

$$I_{i,q}^{(j)}(t) = \sum_{m=1}^{M}\sum_{\substack{k=1\\ k,m\neq i,q}}^{K_m}\sum_{l=1}^{L}\sqrt{p'_{k,m}\Gamma_k(x,y)}\, g_{i,q}^{(j)}\left(\theta_{k,m,l}\right) \times \alpha_{k,m,l}\, b_{k,m}\left(t-\tau_{k,m,l}\right) c_{k,m}\left(t-\tau_{k,m,l}\right) \quad (17)$$

where $g_{i,q}^{(j)}(\theta)$ is the magnitude response of the j th beamformer for user i,q toward the DoA θ [9].

3.2. Matched Filter

Using beamforming in the first stage, will reduce the IPI for the desired user and the MAI from the other users whose signals arrive at different angles from the desired user signal (out-beam interference). Now, in the second stage of the RAKE receiver, the output signal from the j th beamformer is directly passes on to a filter matched to the desired user's signature sequence. The j th matched filter output corresponding to the n th bit is [9]:

$$z_{i,q}^{(j)}(n) = \sqrt{p'_{i,q}}\, b_{i,q}(n)\,\alpha_{i,q,j} + \tilde{I}_{i,q}^{\prime(j)}(n) + I_{i,q}^{\prime(j)}(n) + n^{\prime(j)}(n) \quad (18)$$

where

$$\tilde{I}_{i,q}^{\prime(j)}(n) = \frac{1}{T_b}\int_{(n-1)T_b+\tau_{i,q,j}}^{nT_b+\tau_{i,q,j}} \tilde{I}_{i,q}^{(j)}(t)\, c_{i,q}\left(t-\tau_{i,q,j}\right)dt \quad (19)$$

$$I_{i,q}^{\prime(j)}(n) = \frac{1}{T_b}\int_{(n-1)T_b+\tau_{i,q,j}}^{nT_b+\tau_{i,q,j}} I_{i,q}^{(j)}(t)\, c_{i,q}\left(t-\tau_{i,q,j}\right)dt \quad (20)$$

and

$$n^{\prime(j)}(n) = \frac{1}{T_b}\int_{(n-1)T_b+\tau_{i,q,j}}^{nT_b+\tau_{i,q,j}} n^{(j)}(t)\, c_{i,q}\left(t-\tau_{i,q,j}\right)dt \quad (21)$$

If we assume that the paths' delays from all users are less than the symbol duration $\left(\tau_{k,m,l} < T_b\right)$ for all users' signals on all paths, the n th bit IPI and MAI at the output of the j th matched filter are expressed as [9]

$$\tilde{I}_{i,q}^{\prime(j)}(n) = \sum_{\substack{l=1\\ l\neq j}}^{L}\sqrt{p'_{i,q}}\, g_{i,q}^{(j)}\left(\theta_{i,q,l}\right)\alpha_{i,q,l}\, b_{i,q}(n) \times R_{i,i}\left(\tau_{i,q,j}-\tau_{i,q,l}\right) \quad (22)$$

and

$$I_{i,q}^{\prime(j)}(n) = \sum_{m=1}^{M}\sum_{\substack{k=1\\ k,m\neq i,q}}^{K_m}\sum_{l=1}^{L}\sqrt{p'_{k,m}\Gamma_k(x,y)}\, g_{i,q}^{(j)}\left(\theta_{k,m,l}\right) \times \alpha_{k,m,l}\, b_{k,m}(n) R_{i,k}\left(\tau_{i,q,j}-\tau_{k,m,l}\right) \quad (23)$$

where the autocorrelation function $R_{i,k}(\tau)$ is [9,20]:

$$R_{i,k}(\tau) = \frac{1}{T_b}\int_{T_b} c_{i,q}(t)\, c_{k,m}(t+\tau)\,dt \quad (24)$$

If all users' delays are multiples of the chip period (T_c), then

$$R_{i,k}(\tau) = \frac{1}{G}\sum_{l_1=0}^{G-1}\sum_{l_2=0}^{G-1} c_{i,q}(l_1)\, c_{k,m}(l_2)\, R_c\left(\tau-(l_1-l_2)T_c\right) \quad (25)$$

where the autocorrelation function $R_c(\tau)$ is:

$$R_c(\tau) = \frac{1}{T_b}\int_{T_b} c(t)\, c(t+\tau)\,dt \quad (26)$$

In the case of a maximal-length sequence (m-sequence) and for $0 \le \tau \le T_b$, we have [20]:

$$R_c(\tau) = \begin{cases} 1-\dfrac{|\tau|}{T_c}(1+1/G) & ;|\tau| \le T_c \\ -1/G & ;|\tau| \ge T_c \end{cases} \quad (27)$$

3.3. Maximal Ratio Combining

Diversity combining has been considered as an efficient way to combat multipath fading because the combined

SINR is increased compared with the SINR of each diversity branch. The optimum combiner is the MRC whose SINR is the sum of the SINR's of each individual diversity branch [20,21].

After the finger-matched filter, the fingers' signals are combined according to the MRC principle in the third stage of the RAKE receiver. In this paper, we use the conventional MRC that the signal of user i,q in the j th path is combined using multiplying by the complex conjugate of $\alpha_{i,q,j}$.

The SINR in output of the RAKE receiver for user i,q is [9,21]:

$$\mathrm{SINR}_{i,q}(\alpha)=\sum_{j=1}^{L}\mathrm{SINR}_{i,q}^{(j)}(\alpha) \tag{28}$$

where

$$\mathrm{SINR}_{i,q}^{(j)}(\alpha)=\frac{p'_{i,q}\left|\alpha_{i,q,j}\right|^2}{\mathrm{E}\left(\tilde{I}_{i,q}^{\prime(j)}\right)^2+\mathrm{E}\left(I_{i,q}^{\prime(j)}\right)^2+\mathrm{E}\left(n'^{(j)}\right)^2} \tag{29}$$

is the SINR in output of the RAKE receiver in path j for user i,q.

Also, we can be rewritten the SINR in (29) by (30), that shown at the bottom of the page, where $\bar{\Gamma}_k(x,y)$ $=\mathrm{E}\left(\Gamma_k(x,y)\right)$ and $\bar{\alpha}_{k,m,j}^2=\mathrm{E}\left(\left|\alpha_{k,m,j}\right|^2\right)$ [9,22].

In order to perform the BER, we assume Gaussian approximation for the probability density function of interference plus noise. The conditional BER for a BPSK modulation is [9,20]:

$$\mathrm{BER}_{i,q}(\alpha)=Q\left(\sqrt{2\times\mathrm{SINR}_{i,q}(\alpha)}\right) \tag{31}$$

where

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp\left(-u^2/2\right)du \tag{32}$$

4. Smart Step Closed-Loop Power Control Algorithm

A major limiting factor for the satisfactory performance of CDMA systems is the near-far effect. Power control is an intelligent way of adjusting the transmitted powers in cellular systems so that the total transmit power (TTP) is minimized, but at the same time, the user SINRs satisfies the system quality of service (QoS) requirements [23].

Depending on the location where the decision on how to adjust the transmitted powers is made, the power control algorithm can be divided into two groups: centralized power control and distributed power control [1-6, 12]. In centralized power control, a network center can simultaneously compute the optimal power levels for all users. However, it requires measurement of all the link gains and the communication overhead between a network center and base stations. Thus, it is difficult to realize in a large system [24]. Distributed power control, on the other hand, uses only local information to determine transmitter power levels. It is much more scalable than centralized power control. However, transmitter power levels may not be optimal, resulting in performance degradation [25].

The distributed closed-loop power control problem has been investigated by many researchers from many perspectives during recent years [4,23,26]. For instance, the conventional fast closed-loop power control strategy used in practice in CDMA systems is a fixed step controller based on SINR measurements. The fixed step closed-loop power control (FSPC) algorithm is defined by [4]

$$p_{i,q}^{n'+1}=p_{i,q}^{n'}+\delta\,\mathrm{sign}\left(\gamma_{i,q}^{*}-\gamma_{i,q}^{n'}\right) \tag{33}$$

where $p_{i,q}^{n'}$, $\gamma_{i,q}^{*}$, and $\gamma_{i,q}^{n'}$ are the transmitter power, SINR target, and measured SINR of user i,q at time n', respectively, and δ is the fixed step size. Also $p_{i,q}^{n'+1}$ is transmitter power control (TPC) command in the feedback link of the base station to user i,q at time $n'+1$ (all signals are in decibels).

Also, the distributed traditional closed-loop power control (DTPC) is defined by [23] the variance of

$$p_{i,q}^{n'+1}=\frac{\gamma_{i,q}^{*}}{\gamma_{i,q}^{n'}}\,p_{i,q}^{n'} \tag{34}$$

In both algorithms, the simple intuition behind this iteration is that if the current SINR $\gamma_{i,q}^{n'}$ of user i,q is less than the target SINR $\gamma_{i,q}^{*}$, then the power of that user is increased; otherwise, it is decreased. It should be mentioned that convergence speed of DTPC algorithm is higher than FSPC algorithm. Also, the SINR mis-

$$\mathrm{SINR}_{i,q}^{(j)}(\alpha)=\frac{p'_{i,q}\left|\alpha_{i,q,j}\right|^2}{p'_{i,q}\bar{\alpha}_{i,q,j}^2\sum\limits_{\substack{l=1\\l\neq j}}^{L}\left|g_{i,q}^{(j)}\left(\theta_{i,q,l}\right)\right|^2R_{i,i}^2\left(\tau_{i,q,j}-\tau_{i,q,l}\right)+\sum\limits_{m=1}^{M}\sum\limits_{\substack{k=1\\k,m\neq i,q}}^{K_m}p'_{k,m}\bar{\Gamma}_k(x,y)\bar{\alpha}_{k,m,j}^2\sum\limits_{l=1}^{L}\left|g_{i,q}^{(j)}\left(\theta_{k,m,l}\right)\right|^2R_{i,k}^2\left(\tau_{i,q,j}-\tau_{k,m,l}\right)+\frac{N_0}{2T_b}} \tag{30}$$

adjustment in FSPC algorithm is higher than DTPC algorithm. But, it has been shown that the FSPC algorithm converge to a bound region $\left|\gamma_{i,q}^{*}-\gamma_{i,q}^{n'}\right| \le 2\delta k_{ld}$, where k_{ld} is the loop delay [4].

Also in [26], variable step closed-loop power control (VSPC) algorithm has been proposed. In this algorithm, variable step size is discrete with mode q_v. It is shown that the performance of VSPC algorithm with mode $q_v = 4$ is found to be worse than that of a fixed step algorithm ($q_v = 1$) under practical situations with loop delay of two power control intervals, but the convergence speed of VSPC algorithm is higher than FSPC algorithm. Also in this algorithm, the variance of the SINR mis-adjustment is reduced in compared to FSPC algorithm.

Practical implementations of power control in CDMA systems utilize closed-loop control, where the transmitter adjusts its power based on commands received from the receiver in a feedback channel. To minimize signaling overhead, typically one bit is used for the power control command. In practice, the command must be derived based on measurements made at the receiver, transmitted over the feedback channel to the transmitter, and finally processed and applied at the transmitter. All these operations constitute a loop delay, which can cause problems if it is not properly taken care of in the design of the power control algorithm. In many cases the loop delay is known due to a specific frame structure inherent in the system. A typical loop delay situation encountered in WCDMA systems is shown in **Figure 3**. The slot at time $n't$ is transmitted using power $p^{n'}$. The receiver measures the SINR $\gamma^{n'}$ over a number of pilot and/or data symbols and derives a TPC command. The command is transmitted to the transmitter in the feedback link and the transmitter adjusts its power at time $(n'+1)t$ according to the command. It should be mentioned that since the power control signaling is standardized, the loop delays are known exactly [4].

In this paper, we propose the smart step closed-loop power control algorithm. The SSPC algorithm defines as follows.

$$p_{i,q}^{n'+1} = p_{i,q}^{n'} + \delta\left|\gamma_{i,q}^{*}-\gamma_{i,q}^{n'}\right|\operatorname{sign}\left(\gamma_{i,q}^{*}-\gamma_{i,q}^{n'}\right) \qquad (35)$$

The SSPC algorithm is implemented as follows.

1) Select the initial transmitted power vector ($n'=0$) for all users within cell m as

$$\mathbf{p}_m^0 = \left[\, p_{1,m}^0 \;\; p_{2,m}^0 \ldots p_{K_m,m}^0 \,\right], \quad m = 1,2,\ldots,M \;.$$

2) Estimate the weight vector for all users with the CLMS algorithm using (11).

3) Calculate the SINR for all users using (28).

4) If $\left|\gamma_{k,m}^{*}-\gamma_{k,m}^{n'}\right| > \varepsilon_0$ for each user then set $n' = n'+1$ and calculate the TPC for all users at time $n'+1$ using (35) and go back to 2), where ε_0 is threshold value.

Finally, if $\left|\gamma_{k,m}^{*}-\gamma_{k,m}^{n'}\right| < \varepsilon_0$ for all users then algorithm ends.

As will be seen from simulation results, because of variable coefficient in the sign function, the convergence speed of our algorithm is higher than VSPC and FSPC algorithms.

5. BSA-MTP Technique

The system capacity might be improved, if the users are

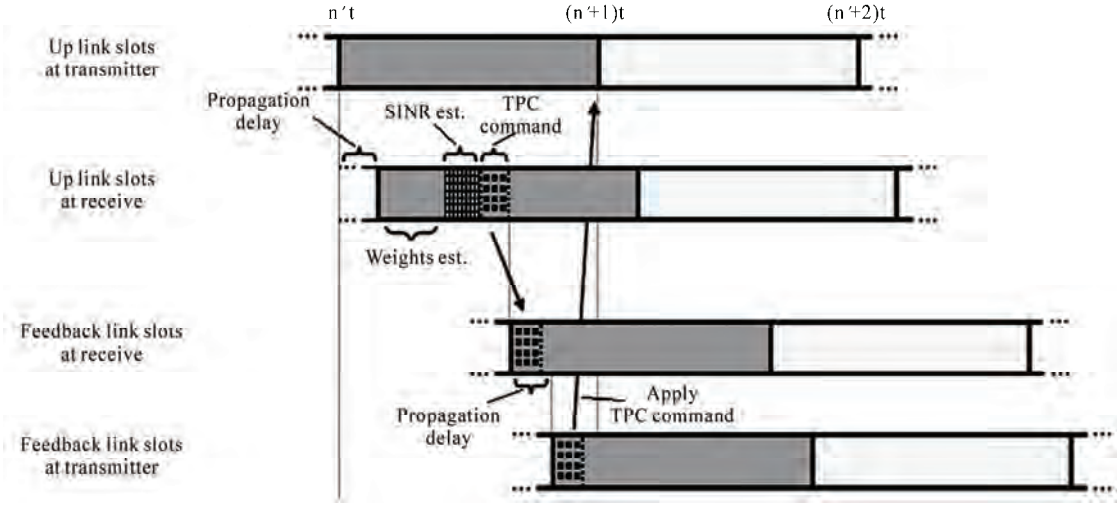


Figure 3. Example of power control timing in WCDMA systems [4].

allowed to switch to alternative base stations, especially when there are congested areas in the network. Obviously, when uplink performance is of concern, the switching should happen based on the total interferences seen by the base stations [15].

So far, we have considered the power control problem for a number of transmitter-receiver pairs with fixed assignments, which can be used in uplink or downlink in mobile communication systems. In an uplink scenario where base stations are equipped with antenna arrays, the problem of joint power control and beamforming, as well as base station assignment, naturally arises [12].

In this paper, we modify the BSA-MTP technique to support base station assignment as well. The modified technique can be summarized as follows.

1) Initially by the conventional BSA technique, each mobile connects to its base station, according to (2).

2) Estimate the weight vector for all users with the CLMS algorithm.

3) Each mobile updates its transmitted power based on the SSPC algorithm using (35).

4) Finally, $K_r = \lfloor K_u / M \rfloor$ users that their transmitted power is higher than the other users to be transferred to other base stations according to the following equation, where the function $\lfloor x \rfloor$ returns the integer portion of a number x.

$$\Gamma_k(x,y) = \begin{cases} 1 & ;k \in S_{\mathrm{BS}q} \\ \dfrac{\min\limits_{\substack{m\in\Theta_k \\ m\neq q}} \left\{ d_{k,m}^{L_\alpha}(x,y) 10^{\xi_{k,m}/10} \right\}}{d_{k,q}^{L_\alpha}(x,y) 10^{\xi_{k,q}/10}} & ;k \in S_{\overline{\mathrm{BS}q}} \\ \dfrac{\min\limits_{m\in\Theta_k} \left\{ d_{k,m}^{L_\alpha}(x,y) 10^{\xi_{k,m}/10} \right\}}{d_{k,q}^{L_\alpha}(x,y) 10^{\xi_{k,q}/10}} & ;k \in S_o \end{cases} \quad (36)$$

where $S_{\overline{\mathrm{BS}q}}$ is the set of users that are in cell q but not connected to BS q [2].

It should be mentioned that the technique for users that are present in the border of cells, the BER can be effecttively reduced.

6. Switched-Beam Technique and Equal Sectoring Method

One simple alternative to the fully adaptive antenna is the switched-beam architecture in which the best beam is chosen from a number of fixed steered beams. Switched-beam systems are technologically the simplest and can be implemented by using a number of fixed, independent, or directional antennas [27]. We list the SB technique conditions for this paper as follows.

1) Coverage angle for all beams is 30° and overlap between consecutive beams is 20°. Thus each base station has 36 beams.

2) Each user can be use one beam for its each path to communicate with a base station at any time

Also, one of simple methods to sectorize a cell is equal sectoring, in which all sectors have the same coverage angle. In this paper, we assume three sectors for each base station with sector angle 120° for the ES method.

7. Simulation Results

We consider $M = 4$ base stations for a four-cell CDMA system on a 2×2 grid as **Figure 4**. We assume a uniform linear array of S omni-directional antennas in each base station with antenna spacing $d = \lambda / 2$. Also, we assume BPSK m-sequence code spreading with processing gain $G = 64$; resolution $R = 1$; the input data rate $T_b = 9.6$ kbps; the number of antenna weights $N = 3$; the number of antenna sensors $S = 3$; threshold value $\varepsilon_0 = 0.1$ dB; frequency-selective fading channel with $L = 2$ resolvable propagation paths; variance of the complex Gaussian fading channel coefficient $\sigma_\alpha^2 = 4$ dB; fixed step size for SSPC, FSPC, and VSPC algorithms $\delta = 0.01$; mode $q_v = 4$ for VSPC algorithm [26]; variance of the log-normal shadow fading $\sigma_\xi^2 = 8$ dB; path-loss component $L_\alpha = 4$; initial value for weight vectors in the CLMS algorithm $\mathbf{w}(0) = \mathbf{0}$; initial value for transmitted power vectors $\mathbf{p}_m^0 = \mathbf{0}$. The SINR target value is the same for all users and is set to $\gamma^* = 5(7\text{ dB})$. It also is assumed that the distribution of users in all cells is uniform.

First, in order to compare the BSA-MTP and conventional BSA techniques, we assume the PPC, and the BER

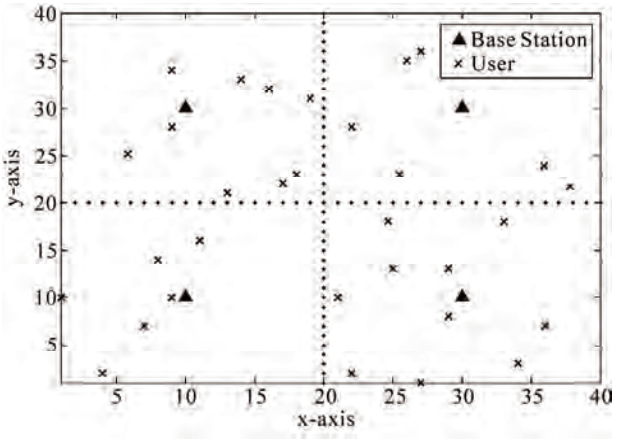


Figure 4. Location plot of base stations and users in four cells.

has been calculated from (31). Finally, we compare the TTP with the joint SSPC algorithm and BSA-MTP technique in comparison with other methods.

Figure 5 shows the average BER versus the signal-to-noise ratio (SNR) for different receivers (one, two, and three-stage receivers) in the case of $K_u = 32$ active users and the PPC case. It should be mentioned that in this simulation, $K_r = 8$ users can to be transferred to other base stations with the BSA-MTP technique. Also in this simulation we use CLMS algorithm or SB technique in the first stage of the RAKE receiver. It is clear that, in MF only receiver (one-stage receiver) and in the case of the conventional BSA technique, we still have the error floor at high SNR. Using CLMS and MRC receiver (two-stage RAKE receiver) or CLMS, MF, and MRC receiver (the three-stage RAKE receiver as **Figure 1**) has a better performance than using MF only. Also observe that using the BAS-MTP technique in the case of three-stage RAKE receiver (CLMS method in the first stage), the average BER is lower than the conventional BSA technique. For example, at a SNR of 20 dB, the average BER is 0.0096 for the three-stage RAKE receiver with the conventional BSA technique, while for the BSA-MTP technique, the average BER is 0.0031. Also it can be seen that the average BER in the SB technique is less than the CLMS algorithm. Also, it is clear that the MAI is not removed totally and the performance is still worse than the single user per cell bound.

Figure 6 shows the average BER versus the number of active users (K_u) for different receivers as **Figure 5**, in the case of the PPC and $\mathrm{SNR} = 10$ dB. At a BER of 0.005, the three-stage RAKE receiver (CLMS method in the first stage) with the BSA-MTP technique support $K_u = 29$ users, while for the three-stage RAKE receiver and the conventional BSA technique support $K_u = 18$ users. We also observe that the three-stage RAKE receiver can achieve lower BER than the one and two-stage receivers. Also at a BER of 0.002, the three-stage RAKE receiver for the SB technique in the first stage and for the BSA-MTP technique support $K_u = 49$ users, while the CLMS, MF, and MRC receiver support $K_u = 18$ users. It should be mentioned that increasing the number of active users in the SB technique, will lead more complexity in receiver in comparison with the CLMS algorithm. Also increasing the number of active users (K_u), will increase the number of users that can to be transferred to other base stations (K_r) in the BSA-MTP technique.

Figure 7 shows the comparison of the average SINR achieved over $K_u = 32$ users versus the power control iteration index (n') for SSPC, VSPC, and FSPC algorithms and for BSA-MTP and conventional BSA tech

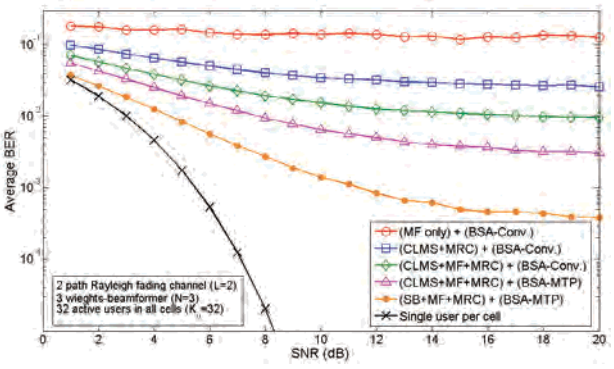


Figure 5. Average BER of all users versus the SNR for the PPC case.

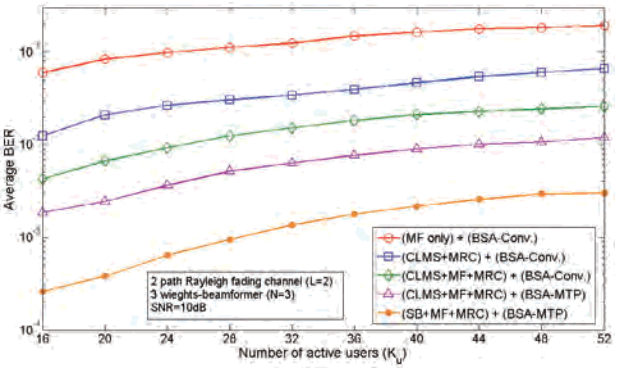


Figure 6. Average BER for all users versus the number of active users for the PPC case and SNR = 10 dB.

niques. In this simulation, the three-stage RAKE receiver uses CLMS, SB, or ES methods in the first stage. Also, we assume that each user to have a maximum power constraint of 1 watt. Accordingly, we observe that the convergence speed of the SSPC algorithm is faster than the VSPC and FSPC algorithms. The figure also shows that the SSPC algorithm with the BSA-MTP technique converges faster than the SSPC algorithm for the conventional BSA technique. In addition, we see that the convergence speed of the SSPC algorithm for the SB technique is faster than the CLMS and ES methods. Also observe that the average SINR level achieved is below the target SINR value for the ES method, because in this method, the MAI is much higher than SB technique and CLMS algorithm.

Figure 8 shows the comparison of TTP usage versus the power control iteration index (n') when there are $K_u = 32$ users in all cells according to **Figure 7**. But in this simulation, we assume that users have no maximum power constraints. Similar to **Figure 7**, we observe that the ES method never can achieve the target SINR value. Also this figure shows that the SSPC algorithm offers more savings in the TTP as compared to the VSPC and FSPC algorithms. In addition, the figure shows that the

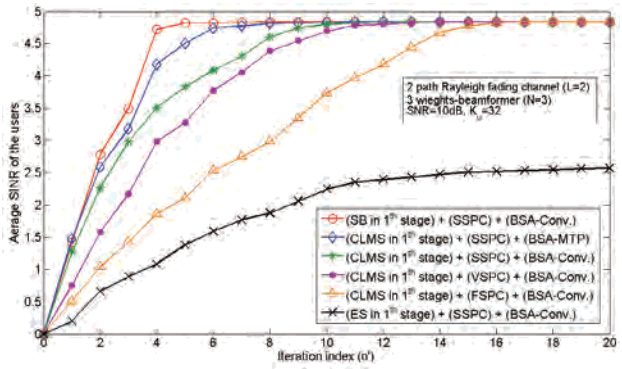


Figure 7. Average SINR of all users versus power control iteration index with maximum power constraint of 1 watt.

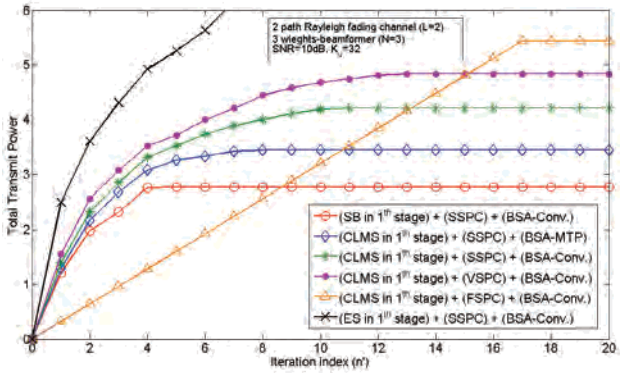


Figure 8. Total transmit power of all users versus power control iteration index. No power constraints.

TTP in BSA-MTP technique is less than conventional BSA technique. Also it can be seen that the TTP for the SB technique is lower than the CLMS algorithm, because in SB technique, the MAI is lower than CLMS algorithm.

8. Conclusions

In this paper, we studied the RAKE receiver performance of multiple-cell DS-CDMA system with the space diversity processing, Rayleigh frequency-selective channel model, closed-loop power control, and base station assignment. This receiver consists of CLMS, MF, and MRC in three stages.

Accordingly, we proposed the SSPC algorithm and the BSA-MTP technique to reduce the CCI and the MAI. It has been shown that, by using antenna arrays at the base stations, the SSPC algorithm and the BSA-MTP technique will decrease the interference in all cells. In addition, it can be seen that the TTP in the SSPC algorithm is less than the VSPC and FSPC algorithms. Also our results show that the TTP for BSA-MTP technique is lower than conventional case. Thus, it decreases the BER by allowing the SINR targets for the users to be higher, or by increasing the number of users supportable at a fixed SINR target level. On the other hand, it has been shown that the convergence speed of the SSPC algorithm is increased in comparison with the VSPC and FSPC algorithms. It has also observed that using the BSA-MTP technique will decrease the average BER of the system to support a significantly larger number of users.

9. References

[1] A. Abrardo and D. Sennati, "On the Analytical Evaluation of Closed-Loop Power-Control Error Statistics in DS-CDMA Cellular Systems," *IEEE Transactions on Vehicular Technology*, Vol. 49, No. 6, 2000, pp. 2071-2080.

[2] L. Carrasco and G. Femenias, "Reverse Link Performance of a DS-CDMA System with Both Fast and Slow Power Controlled Users," *IEEE Transactions on Wireless Communications*, Vol. 7, No. 4, 2008, pp. 1255-1263.

[3] L. Qian and Z. Gajic, "Variance Minimization Stochastic Power Control in CDMA System," *IEEE Transactions on Wireless Communications*, Vol. 5, No. 1, 2006, pp. 193-202.

[4] M. Rintamaki, H. Koivo and I. Hartimo, "Adaptive Closed-Loop Power Control Algorithms for CDMA Cellular Communication Systems," *IEEE Transactions on Vehicular Technology*, Vol. 53, No. 6, 2004, pp. 1756-1768.

[5] J. Wang and A. Yu, "Open-Loop Power Control Error in Cellular CDMA Overlay Systems," *IEEE Journal on Selected Areas in Communications*, Vol. 19, No. 7, 2001, pp. 1246-1254.

[6] J. T. Wang, "Admission Control with Distributed Joint Diversity and Power Control for Wireless Networks," *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 1, 2009, pp. 409-419.

[7] M. Dosaranian-Moghadam, H. Bakhshi, G. Dadashzadeh and P. Rahmati, "Adaptive Beamforming Method Based on Constrained LMS Algorithm for Tracking Mobile User," *IEEE Global Mobile Congress*, Shanghai, October 2009, pp. 1-6.

[8] J. Chang, L. Tassiulas and F. Rashid-Farrokhi, "Joint Transmitter Receiver Diversity for Efficient Space Division Multiaccess," *IEEE Transactions on Wireless Communications*, Vol. 1, No. 1, 2002, pp. 16-27.

[9] N. A. Mohamed and J. G. Dunham, "A Low-Complexity Combined Antenna Array and Interference Cancellation DS-CDMA Receiver in Multipath Fading Channels," *IEEE Journal on Selected Areas in Communications*, Vol. 20, No. 2, 2002, pp. 248-256.

[10] N. A. Mohamed and J. G. Dunham, "Adaptive Beamforming for DS-CDMA Using Conjugate Gradient Algorithm in a Multipath Fading Channel," *Proceedings of* 1999 *IEEE Emerging Technologies Symposium*, Dallas, April 1999, pp. 859-863.

[11] F. Rashid-Farrokhi, K. J. Ray-Liu and L. Tassiulas, "Transmit Beamforming and Power Control for Cellular Systems," *IEEE Journal on Selected Areas in Communications*, Vol.

16, No. 8, 1998, pp. 1437-1450.

[12] F. Rashid-Farrokhi, L. Tassiulas and K. J. Ray-Liu, "Joint Optimal Power Control and Beamforming in Wireless Networks Using Antenna Arrays," *IEEE Transactions on Communications*, Vol. 46, No. 10, 1998, pp. 1313-1324.

[13] R. D. Yates and C. Huang, "Integrated Power Control and Base Station Assignment," *IEEE Transactions on Vehicular Technology*, Vol. 44, No. 3, 1995, pp. 638-644.

[14] S. V. Hanly, "An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular Spread Spectrum Capacity," *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 7, 1995, pp. 1332-1340.

[15] M. Mahmoudi and E. S. Sousa, "Joint Power Control, Base Station Assignment and Sectorization for CDMA Cellular Systems," *Proceedings of* 2000 *IEEE Vehicular Technology Conference*, Boston, September 2000, pp. 573-580.

[16] J. Litva and T. Kwok-Yeung, "Digital Beamforming in Wireless Communications," Artech House, Boston, 1996.

[17] X. Y. Sun, X. H. Lian and J. J. Zhou, "Robust Adaptive Beamforming Based on Maximum Likelihood Estimation," *International Conference on Microwave and Millimeter Wave Technology*, Nanjing, April 21-24, 2008, pp. 1137-1140.

[18] M. Z. Shakir and T. S. Durrani, "Narrowband Beamforming Algorithm for Smart Antennas," *International Bhurban Conference on Applied Sciences & Technology*, Islamabad, Pakistan, January 2007, pp. 49-54.

[19] S. Haykin, "Adaptive Filter Theory," 3rd Edition, Prentice Hall, New Jersey, 1996.

[20] R. L. Peterson, R. E. Ziemer and D. E. Borth, "Spread-Spectrum Communications," Prentice-Hall, New Jersey, 1995.

[21] N. Kong and L. B. Milstein, "Average SNR of a GeneralIzed Diversity Selection Combining Scheme," *IEEE Communications Letters*, Vol. 3, No. 3, 1999, pp. 57-59.

[22] J. C. Liberti and T. S. Rappaport, "Smart Antennas for Wireless Communications IS-95 and Third Generation CDMA Applications," Prentice-Hall, New Jersey, 1999.

[23] A. Yener, R. D. Yates and S. Ulukus, "Interference Management for CDMA Systems through Power Control, Multiuser Detection, and Beamforming," *IEEE Transactions on Communications*, Vol. 49, No. 9, 2001, pp. 1227-1239.

[24] S. Grandhi, R. Vijayan and D. Goodman, "Centralized Power Control in Cellular Radio Systems," *IEEE Transactions on Vehicular Technology*, Vol. 42, No. 4, 1993, pp. 466-468.

[25] S. Grandhi, R. Vijayan and D. Goodman, "Distributed Power Control in Cellular Radio Systems," *IEEE Transactions on Communications*, Vol. 42, No. 2-4, 1994, pp. 226-228.

[26] A. Kurniawan, "Effect of Feedback Delay on Fixed Step and Variable Step Power Control Algorithm in CDMA Systems," *The* 8*th International Conference on Communication Systems*, Indonesia, 2002, pp. 1096-1100.

[27] B. Allen and M. Beach, "On the Analysis of Switched-beam Antennas for the W-CDMA Downlink," *IEEE Transactions on Vehicular Technology*, Vol. 53, No. 3, 2004, pp. 569-578.

Wireless Sensor Network, **2010, 2, 483-491**
doi:10.4236/wsn.2010.26060 Published Online June 2010 (http://www.SciRP.org/journal/wsn)

Energy Conservation Challenges in Wireless Sensor Networks: A Comprehensive Study

Suraiya Tarannum
Department of Telecommunication Engineering AMC Engineering College, Bangalore, India
E-mail: *ssuraiya@gmail.com*
Received December 31, 2009; *revised February* 5, 2010; *accepted February* 10, 2010

Abstract

A Wireless Sensor Network (WSN) consists of a large number of randomly deployed sensor nodes. These sensor nodes organize themselves into a cooperative network and perform the three basic functions of sensing, computations and communications. Research in WSNs has become an extensive explorative area during the last few years, especially due to the challenges offered, energy constraints of the sensors being one of them. In this paper, a thorough comprehensive study of the energy conservation challenges in wireless sensor networks is carried out. The need for effective utilization of limited power resources is also emphasized, which becomes pre-eminent to the Wireless Sensor Networks.

Keywords: Wireless Sensor Network, Sensor Node, Communication Protocols Architecture, Energy Consumption of Sensor Node, Energy Conservation, Communication Protocols

1. Introduction

Wireless Sensor Networks (WSNs) are a spatially distributed autonomous system which is a collection of power-conscious wireless sensors without the support of pre-existing infrastructure. A co-operative system is created, formed by a group of specialized transducers with communication infrastructure intended to monitor and record conditions at diverse locations. A WSN is used for information gathering, performing data-intensive tasks such as habitat monitoring, seismic monitoring, terrain, surveillance etc. Sensor Networks are a gaint leap toward "Proactive Computing", a paradigm where computers anticipate human needs and if necessary, act on their behalf. Sensor Networks and proactive computing has the potential to improve our productivity and enhance safety, awareness and efficiency at the societal scale [1].

Building sensors has been made possible by the recent advances in Micro-Electro-Mechanical System (MEMS) technology and wireless communications technology making it a pragmatic vision to deploy a large-scale, low power, inexpensive wireless sensor network [2]. Such an approach promises advantages over the traditional sensing methods in many ways: large-scale, dense deployment not only extends the spatial coverage and achieves higher resolution, but also increases the fault-tolerance and robustness of the system.

The recent advances in MEMS (Micro Electro Mechanical Systems) [3], Digital Signal Processing and Wireless Communications have led to the production of new class of wireless, battery operated smart sensor nodes [4]. These nodes organize themselves to form active, full-fledged processing elements, capable of measuring the real world phenomena, filtering, sharing and combining these measurements. In such networks, the devices identify themselves and each other, to route data without possessing any prior knowledge of or assumptions of the network topology, which may change, run out of power or experience shifting waves of interference.

A network formed by a web of such sensors is deployed in remote areas or hostile terrains, without the infrastructure support from the outside world. This exerts serious physical constraints on the application of single sensor, and thus, all the sensor nodes can form an autonomous and robust data computing and communication distributed system for automated information gathering and distributed sensing. Sensor networks are highly distributed networks of small, lightweight nodes termed motes, deployed in large numbers to monitor the environment or a system by measuring physical parameters such as temperature, pressure or relative humidity.

Each sensor node consists of three subsystems: the sensor subsystem which senses the environment, the processing subsystem which performs local computations on the sensed data, and the communication subsystem which is responsible for message exchange with neighbouring sensor nodes. While individual sensors have

limited sensing region, processing power and energy, networking a large number of sensors gives rise to a robust, reliable and accurate sensor network covering a wider region.

The network so formed is fault-tolerant since many nodes participate in sensing the same events. Furthermore, the nodes cooperate and collaborate on their data, which leads to accurate sensing of events in the environment. The two most important operations in a sensor network are data dissemination, which is the propagation of data/queries throughout the network, and data gathering where the collection of observed data takes place from each of the sensor nodes. Finally, the aggregated data is sent to the sink/basestation. A typical scenario of the WSN is depicted in **Figure 1**.

Efficient management of energy deserves much of the attention in the WSNs. Routing protocols designed for WSNs must therefore effectively tackle these issues in order to enhance the lifetime of the network. Hierarchical routing techniques are preferable in this direction. The arrangement of the nodes in the form of a load balanced hierarchy proves beneficial.

In this paper, a state-of-art study of the energy conservation challenges in wireless sensor networks is described. The rest of the paper is organized as follows. In section 2, the sensor node is described. The applications and the issues and challenges are described in sections 3 and 4 respectively. Section 5 throws light on the energy consumption details and the communication protocol architecture is described in section 6. Section 7 enlightens the energy conservation challenges in communication protocols and related design issues in wireless sensor networks. The usual performance evaluation metrics employed in WSNs are described in sections 8 and 9 contains the conclusions.

2. The Sensor Node

The sensor node is an atomic element of the wireless sensor network, which gathers data from it's surroundings, and transmits them to the *base station/sink* enroute the radio transmission medium. Every node is provided with a unique ID number and has an input queue as a buffer. At any point of time, a sensor can behave as a transmitter node, relay node, and sink node or all them. In many application scenarios, a myriad of sensor nodes are spread across a large geographical area, which collaborate and organize themselves in order to carry out the desired task. This implies that a sensor node forms an integral and the most important unit of the wireless sensor network and deserves understanding of its internal architecture.

A sensor node is typically made up of four basic components as shown in **Figure 2**: A sensing/actuating unit, a processing unit, transceiver section and power supply unit. In addition to this, the sensor node may also be equipped with location detection unit such as a Global Positioning System (GPS), a mobilizer etc. The sensor networks consist of different types of sensors such as seismic, thermal, visual, and infrared and are used to monitor a variety of ambient conditions such as temperature, humidity, pressure and characteristics of objects and their motion. The sensors give these nodes their eyes and ears. Sensor nodes can be used in military, health, chemical processing and disaster relief scenarios. The sensor node architecture is described in **Figures 2** and **3**.

2.1. Sensing Unit

The sensing unit is usually made up of two subunits, the sensors themselves and analog-to-digital converters (ADCs). The signals generated by the sensors, based on the phenomenon to be sensed, are analog in nature and hence need to be converted to a digital to aid further processing. These signals are then fed to the processing unit.

2.2. Processing Unit

The processing unit forms the core of the sensor node.

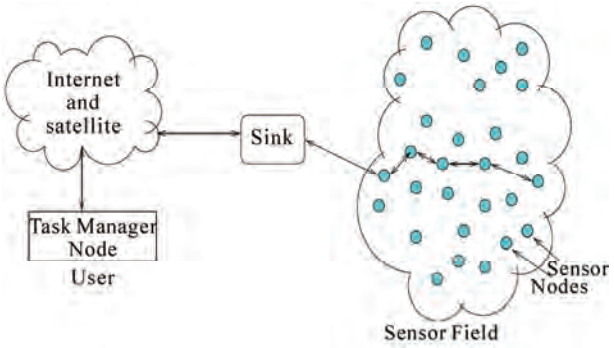


Figure 1. A typical scenario of a Wireless Sensor Network (WSN).

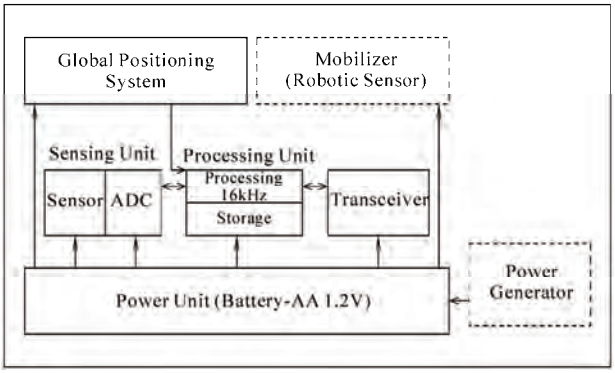


Figure 2. Sensor node architecture.

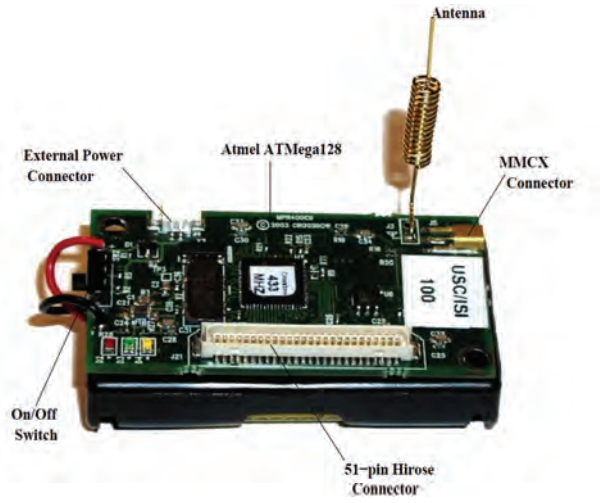


Figure 3. A typical sensor node.

This unit in association with a small storage unit, manages the procedures that make the sensor node collaborate with the other nodes to carry the sensing tasks. The processors employed in the sensor nodes include, the *Atmel AtMega Microcontroller, MSP*430, *Intel Strong ARM* [5], to name a few.

2.3. Communication Unit

The transceiver unit connects wirelessly through the RF channel and is linked to an omni-directional antenna that allows for communications in all directions. The main task of a transceiver is to convert a bit stream arriving from the processing unit into electromagnetic radio waves. Some of commonly used transceivers in sensor nodes are *RFM TR family, Chipcoa CC*10000 *family, The Infineon TDA* 525*x family* etc.

The transceiver unit may be passive or an active optical device or a RF device. RF communications require modulation, band pass filtering, demodulation and multiplexing circuitry, which make them more complex and expensive. Moreover, the path loss of the transmitted signal between two sensor nodes may be as high as the fourth order exponent of the distance between them, because the antennas of the sensor nodes are close to the ground. Nevertheless, RF communications is preferred in most of the ongoing sensor network research, because the packets conveyed in sensor networks are small, data rates are low and frequency reuse is high due to short communication distances. These characteristics also make it possible to use low duty cycle radio electronics for sensor networks.

Of the three domains, a sensor node expends maximum energy in data communications. This involves both data transmission and reception. It is found that for short-range communications with low radiation power, the transmission and reception energy costs are nearly the same. Mixers, frequency synthesizers, voltage controlled oscillators, phase locked loops (PLLs) and power amplifiers, all consume valuable power in the transceiver circuitry.

2.4. Power Supply Unit

The sensor nodes can be powered from energy storage devices or by energy scavenging. The former technique employs a variety of tiny batteries made up of thin films of vanadium oxide and molybdenum oxide [6]. These are fabricated using micro-machined cavities containing an electrolyte, in addition to chemical energy storage. The latter technique employs energy scavenging from the environment in order that the sensor node can operate uninterrupted. The most widely used energy scavenging technique is the solar radiation. There is a possibility of energy-harnessing from body heat in bio-medical applications.

The battery forms the heart of the sensor system as it decides the lifetime of the system. The battery lifetime needs to be prolonged to maximize the network lifetime. Network Lifetime is defined as the maximum number of times a certain data collection function or task can be carried out without any node running out of energy. It is also defined as the time elapsed until the first node in the network is completely depleted of its energy and is determined by the ability to conserve energy in the network. The requirement is that the size of the battery should be as small as possible, the same time being energy efficient. Batteries with energy scavenging capabilities are being designed to increase the lifetime of the sensor system. Two AA sized batteries of 1.2 V each are employed in the battery subsection [1].

Most of the sensor network routing techniques require the knowledge of precise location of nodes that are deployed in the sensor field. This requires a Global Positioning System (GPS) to carry out the tasks. A *mobilizer* may sometimes be needed, especially in Heterogeneneous Wireless Sensor Networks (HWSNs) to move the sensor nodes, when circumstances demand. All of these subunits may need to fit into a matchbox-sized module [1]. The required size may be smaller than even a cubic-centimeter [4], which is light enough to be suspended in air.

The Heterogeneous Wireless Sensor Networks (HWSN), a class of WSNs are distributed networks consisting of large number of tiny, typically the size of 35 mm film canister [7,8], static, low power sensor nodes along with a few mobile, high power nodes. These sensor nodes just like their WSN counterparts, have sensing, processing, co-ordinating and communicating abilities. They are used to monitor changes in unattended regions and relay information to the respective control center where necessary action would be taken. In order to complete a

given task, all sensor nodes have to collaborate by exchanging and forwarding measurement data.

3. Applications

The Wireless Sensor Network technology has the potential to change the way we live, work and do business, with applications in entertainment, travel, retail industry, disaster and emergency management. It forms an increasingly attractive means of monitoring environmental conditions and to bridge the gap between the physical and the virtual world. Application areas for WSNs include geophysical monitoring (seismic activity), precision agriculture (soil management), habitat monitoring (tracking of animal herds), transportation (traffic monitoring), military systems, business process (supply chain management) [9,10] etc.

With continued advances in Micro-Electro-Mechanical Systems (MEMS), Wireless Sensor Networks (WSNs) have and will play a vital role in our daily lives. Humans have relied on wired sensors for years, for simple tasks such as temperature monitoring, to complex tasks such as monitoring life-signs in hospital patients. Wireless Sensor Networks provide unforseen applications in this new field of design [1]. From military applications such as battlefield mapping and target surveillance, to creating context-aware homes [11] where sensors can monitor safety and provide automated services tailored to the individual user; the number of applications are endless. Smart Dust is an example of one such application [12,13]. However this new technology poses many design goals, [1] that up until recently, have not been considered feasible for these applications.

1) The sensor networks are used in a variety of applications which require constant monitoring and detection of specific events. The military applications include battle field surveillance and monitoring, guidance systems of intelligent missiles and detection of attack by weapons of mass destruction, such as chemical, biological or nuclear [14].

2) The WSNs are employed in environmental applications [15] such as forest-fire and flood detection and habitat exploration of animals [16-19].

3) Sensors are extremely useful in patient diagnosis and monitoring. Bio-sensors are implanted in the human body to monitor the patient's physiological parameters such as heart beat or blood pressure. The data so collected is sent regularly to alert the concerned doctor on detection of an anamoly. Such an arrangement provides patients a greater freedom of movement instead of being constantly confined to the hospital bed. Rapid advancements in MEMS technology has made bio-sensors so sophisticated as to enable correct identification of allergies and associated diagnosis [1,20].

4. Issues and Challenges

The WSN is subjected to various resource constraints. The constraints are energy, bandwidth, memory and processing ability. Among them, energy is of prime concern, since it is severely constrained at sensor nodes and it is not feasible to either replace or recharge the batteries of sensor nodes that are often deployed in hostile environment. As a result, these constraints impose an important requirement on any QoS support mechanism in WSNs. Energy efficiency is a critical design issue in WSNs, where each sensor node relies on its limited battery power for data acquisition, processing, transmission and reception.

As the sensor nodes are typically very small and powered by irreplaceable battery, energy control becomes primary and also the most challenging problem in designing sensor networks [21]. In WSNs, each sensor node has different energy consumption rate due to inequality in event sensing and distance from Base Station. This leads to energy disparity among sensor nodes in the network which in turn shortens the lifetime of the network.

Another important issue in WSN is satisfying the QoS parameters. QoS parameters are used for evaluating the performance of networks. The various QoS parameters under considerations are latency, throughput and reliability. Security is a major concern in wireless communications. Sensor network is susceptible to a variety of attacks, including node capture, physical tampering and denial of service while prompting a range of fundamental research challenges. The QoS parameters and energy conservation are the prime factors affecting the lifespan of sensor network. Energy efficient routing mechanisms are inculcated to boost the performance of the sensor network. Wireless sensor networks pose certain design challenges due to the following reasons,

1) The sensor nodes are randomly deployed and hence do not fit into any regular topology. Once deployed, they usually do not require human intervention. This implies that setup and maintenance need to be autonomous.

2) Sensor networks are infrastructureless. Therefore, all routing and maintenance algorithms need to be distributed.

3) An important bottleneck in the operation of sensor nodes is the available energy. Sensors usually rely on their battery for power, which in many cases should be considered as a major constraint while designing protocols. The wireless sensor node, being a micro-electronic device, can only be equipped with a limited power source. In most application scenarios, replenishment of power resources might become impossible. The sensor node lifetime, therefore, shows a strong dependence on battery lifetime.

4) Hardware design for sensor nodes should also consider energy efficiency as a primary requirement. The

micro-controller, operating system, and application software should be designed to conserve power.

5) Sensor nodes should be able to synchronize with each other in a completely distributed manner, so that TDMA schedules can be imposed and temporal ordering of detected events can be performed without ambiguity.

6) A sensor network should also be capable of adapting to changing connectivity due to the failure of nodes, or new nodes powering up. The routing protocols should also be able to dynamically include or avoid sensor nodes in their paths.

7) Real-time communication over sensor networks must be supported through provision of guarantees on maximum delay, minimum bandwidth, or other QoS parameters.

5. Energy Consumption of Sensor Node

The sensor nodes operate in the three modes of sensing, computing and communications, and all of which consume energy. Of the three modes, maximum energy is expended for the communications process. The sensing unit is entrusted with the responsibility to detect the physical characteristics of the environment and has an energy consumption that varies with the hardware nature and applications. However, sensing energy represents a meagre percentage of the entire energy consumption within the entire WSN. In comparison, computations energy is much more. The communication unit consists of a short-range RF circuit which performs the transmission and reception tasks.

Communication energy contributes to data forwarding and it is determined with the transmission range that increases with the signal propagation in an exponential way. The energy consumption model includes the five states: *Acquisition, Transmission, Reception, Listen and Sleep* [22]. These states are described in **Table 1**.

Since the sensor nodes can be in any of three main operations of sensing, computations and communications, each of them could be in different states depending on the component nature. Accordingly different levels of energy are expended in each of them.

Table 1. States of the energy consumption model.

(i) Acquistion: The acquistion state includes sensing, A/D conversion, preprocessing and eventually storage of these data.
(ii) Transmission: The transmission state includes processing, packet forming, encoding, framing, queuing and base band adapting to RF circuits.
(iii) Reception: This state is responsible for low noise amplification, down converter oscillator, filtering, detection, decoding, error detection, address checking and random reception.
(iv) Listen: The listen state is similar to reception and involves the processes of low noise amplification, down convertor oscillator, filtering and terminates at detection.
Sleep: The sleep state expends least energy as compared to the other states.

6. Wireless Sensor Networks Communication Protocols Architecture

Figure 4 depicts the communication protocol stack architecture of the WSN. The energy consumed in one sensor node is influenced by protocol layers structure and the way each layer manages the sensing data.

The protocol layers stack used by the sink and nodes within the network includes the *application layer, transport layer, network layer, data link layer, physical layer, power management plane, mobility management plane and task management plane* and described in **Table 2**.

7. Energy Conservation Challenges in Communication Protocols and Design Issues in WSNs

Despite the innumerable applications of WSNs, these networks have several restrictions, e.g., limited energy supply, limited computing power, and limited bandwidth of the wireless links connecting sensor nodes. One of the main design goals of WSNs is to carry out data communication while trying to prolong the lifetime of the network and prevent connectivity degradation by employing aggressive energy management techniques.

The design of routing protocols in WSNs is influenced by many challenging factors. These factors must be overcome before efficient communication can be achieved. In

Table 2. Protocol layer stack.

(i) Application Layer: This supports different softwares for applications depending on the sensing tasks. There are three types of protocols defined for this layer: (a) SMP - Sensor Management Protocol (b) TADAP - Task Assignment and Data Advertisment Protocol SQDDP - Sensor Query and Data Dissemination Protocol
(ii) Transport Layer: This layer helps to maintain the data flow when the application layer is in need. The protocol development on this layer is a real challenge because sensors are influenced by many factors and constraints such as limited power and memory.
(iii) Network Layer: The network layer allows routing of data through the wireless communication channel. There are several methods and strategies to route data such as routing power cost with available energy based on the energy metric and data-centric routing based on interest dissemination and attribute based naming [1,23].
(iv) Data Link Layer: This layer is responsible for the multiplexing of data streams, data frame detection, medium access control (MAC) and error detection and correction. The design issues of the MAC layer protocol must take into account the different constraints such as power conservation, mobility management and recovery failure strategies.
(v) Physical Layer: This is the lower-most layer and is responsible for frequency selection, carrier frequency generation, signal detection, modulation and data encryption.

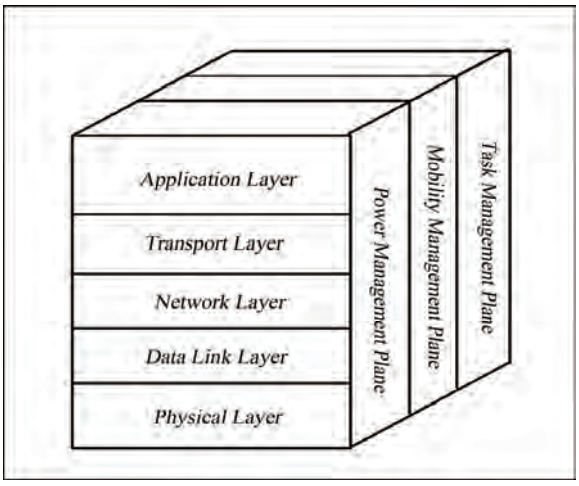


Figure 4. Wireless Sensor Network protocol stack.

the following subsections, some of the routing challenges and design issues that affect routing process in WSNs, are summarized [1,24,25,26].

1) *Node Deployment*

Node deployment in WSNs is application dependent and affects the performance of the routing protocol. The deployment can be either deterministic or randomized. In deterministic deployment, the sensors are manually placed and data is routed through pre-determined paths. However, in random node deployment, the sensor nodes are scattered randomly creating an infrastructure in an ad hoc manner.

If the resultant distribution of nodes is not uniform, optimal clustering becomes necessary to allow connectivity and enable energy efficient network operation. Inter-sensor communication is normally within short transmission ranges due to energy and bandwidth limitations. Therefore, it is most likely that a route will consist of multiple wireless hops.

2) *Energy Consumption without Losing Accuracy*

The sensor nodes can use up their limited supply of energy performing computations and transmitting information in a wireless environment. As such, energy-conserving forms of communication and computation are essential. Sensor node lifetime shows a strong dependence on the battery lifetime [1].

In a multihop WSN, each node plays a dual role as data sender and data router. The malfunctioning of some sensor nodes due to power failure can cause significant topological changes and might require re-routing of packets and reorganization of the network.

3) *Data Reporting Model*

Data sensing and reporting in WSNs is dependent on the application and the time criticality of the data reporting. Data reporting can be categorized as either time-driven (continuous), event-driven, query-driven, and hybrid [27]. The time-driven delivery model is suitable for applications that require periodic data monitoring. As such, sensor nodes will periodically switch on their sensors and transmitters, sense the environment and transmit the data of interest at constant periodic time intervals.

In event-driven and query-driven models, sensor nodes react immediately to sudden and drastic changes in the value of a sensed attribute due to the occurrence of a certain event or a query is generated by the BS. As such, these are well suited for time critical applications. A combination of the previous models is also possible. The routing protocol is highly influenced by the data reporting model with regard to energy consumption and route stability.

4) *Node/Link Heterogeneity*

In many studies, all sensor nodes are assumed to be homogeneous, *i.e.*, having equal capacity in terms of computation, communication, and power. However, depending on the application a sensor node can have different role or capability. The existence of heterogeneous set of sensors raises many technical issues related to data routing. For example, some applications might require a diverse mixture of sensors for monitoring temperature, pressure and humidity of the surrounding environment, detecting motion via acoustic signatures, and capturing the image or video tracking of moving objects.

These special sensors can be either deployed independently or the different functionalities can be included in the same sensor nodes. Even data reading and reporting can be generated from these sensors at different rates, subject to diverse quality of service constraints, and can follow multiple data reporting models. For example, hierarchical protocols designate a cluster-head node different from the normal sensors. These cluster-heads can be chosen from the deployed sensors or can be more powerful than other sensor nodes in terms of energy, bandwidth, and memory. Hence, the burden of transmission to the BS is handled by the set of cluster-heads [28].

5) *Fault-Tolerance*

Some sensor nodes may fail or be blocked due to lack of power, physical damage, or environmental interference. The failure of sensor nodes should not affect the overall task of the sensor network. If many nodes fail, MAC and routing protocols must accommodate formation of new links and routes to the data collection base stations. This may require actively adjusting transmit powers and signalling rates on the existing links to reduce energy consumption, or rerouting packets through regions of the network where more energy is available [29]. Therefore, multiple levels of redundancy may be needed in a fault-tolerant sensor network.

6) *Scalability*

The number of sensor nodes deployed in the sensing area may be in the order of hundreds or thousands, or more. Any routing scheme must be able to work with this huge number of sensor nodes. In addition, sensor network routing protocols should be scalable enough to re-

spond to events in the environment. Until an event occurs, most of the sensors can remain in the sleep state, with data from the few remaining sensors providing a coarse quality.

7) *Network Dynamics*

Most of the network architectures assume that sensor nodes are stationary. However, mobility of both BSs and sensor nodes is sometimes necessary in many applications [19]. Routing messages from or to moving nodes is more challenging since route stability becomes an important issue, in addition to energy, bandwidth etc.

Moreover, the sensed phenomenon can be either dynamic or static depending on the application, e.g., it is dynamic in a target detection/tracking application, while it is static in forest monitoring for early fire prevention. Monitoring static events allows the network to work in a reactive mode, simply generating traffic when reporting. Dynamic events in most applications require periodic reporting and consequently generate significant traffic to be routed to the BS.

8) *Transmission Media*

In a multi-hop sensor network, communicating nodes are linked by a wireless medium. The traditional problems associated with a wireless channel (e.g., fading, high error rate) may also affect the operation of the sensor network. In general, the required bandwidth of sensor data will be low, on the order of 1-100 kbps. Related to the transmission media is the design of medium access control (MAC). One approach of MAC design for sensor networks is to use TDMA based protocols that conserve more energy compared to contention based protocols like CSMA (e.g., IEEE 802.11).

9) *Connectivity*

High node density in sensor networks precludes them from being completely isolated from each other. Therefore, sensor nodes are expected to be highly connected. This, however, may not prevent the network topology from being variable and the network size from being shrinking due to sensor node failures. In addition, connectivity depends on the possibly random distribution of nodes.

10) *Coverage*

In WSNs, each sensor node obtains a certain view of the environment. A given sensors view of the environment is limited both in range and in accuracy; it can only cover a limited physical area of the environment. Hence, area coverage is also an important design parameter in WSNs.

11) *Data Aggregation/Fusion*

Since sensor nodes may generate significant redundant data, similar packets from multiple nodes can be aggregated so that the number of transmissions is reduced. Data aggregation is the combination of data from different sources according to a certain aggregation function, e.g., duplicate suppression, minima, maxima and average.

This technique has been used to achieve energy efficiency and data transfer optimization in a number of routing protocols. Signal processing methods can also be used for data aggregation. In this case, it is referred to as data fusion where a node is capable of producing a more accurate output signal by using some techniques such as beamforming to combine the incoming signals and reducing the noise in these signals.

12) *Quality of Service*

In some applications, data should be delivered within a certain period of time from the moment it is sensed; otherwise the data will be useless. Therefore bounded latency for data delivery is another condition for time-constrained applications. However, in many applications, conservation of energy, which is directly related to network lifetime, is considered relatively more important than the quality of data sent.

As the energy gets depleted, the network may be required to reduce the quality of the results in order to reduce the energy dissipation in the nodes and hence lengthen the total network lifetime. Hence, energy-aware routing protocols are required to capture this requirement.

8. Performance Evaluation Metrics

In order to study the challenges offered by the energy constrained wireless sensor nodes and to evaluate the performance and the QoS offered by the network, the performance metrics under consideration are discussed in **Table 3**.

The previous sections threw light on the WSNs, their characteristics, issues, challenges and applications. In order to understand their performance and behavior, the OMNET++ (Objective Modular Network Test-bed in C++) simulator may be employed. OMNET++ is a discrete-

Table 3. Performance metrics.

(i) Energy Consumption per successful data report This gives a good measure of the network lifetime. A routing algorithm which maximizes the lifetime of network, is desirable. This metric also shows how efficient the algorithm is, in energy consumption. This metric is an indication of the energy cost incurred to realize the achieved performance.
(ii) Network Lifetime Network Lifetime is defined as the time elapsed until the first node in the network is completed drained of its energy (dies).
(iii) Network Throughput This is defined as the total number of packets received at the sink divided by the simulation time.
(iv)Latency Latency is defined as the average time that a packet moves on the network.
(v) Delivery Ratio Delivery ratio of the network is specified in terms of the number of packets received at the sink divided by the number of packets generated at the source.

event simulator for WSNs [30]. It is a public-source, component-based, modular simulation frame work and used to simulate communication networks and other distributed systems.

Discrete-event simulation is a trusted platform for modelling and simulating a variety of systems. The design of WSNs requires the simultaneous consideration of the effects of several factors such as energy efficiency, fault-tolerance, Quality of Service (QoS) demands, synchronization, scheduling strategies, system topology, communications and coordination protocols.

9. Conclusions

A WSN is composed of tens to thousands of sensor nodes which communicate through a wireless channel for information sharing and processing. The sensors are deployed on a large scale for environmental monitoring and habitat study, for military surveillance, in emergent environments for search and rescue, in buildings for infrastructure health monitoring, in homes to realize a smart environment. WSNs have been made viable by the convergence of micro-electro-mechanical systems technology, wireless communications and digital electronics. The energy conservation challenges and related issues emphasize the need for energy saving and optimizing protocols to increase the lifetime of sensor networks.

10. References

[1] I. F. Akylidiz, W. L. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey on Sensor Networks," *IEEE Communications Magazine*, Vol. 40, No. 8, 2002, pp. 102-114.

[2] D. Estrin, R. Govindan, J. Heidemann and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking*, Seattle, Washington, USA, August 1999, pp. 263-270.

[3] P. B. Chun, N. R. Lo, E. Berg and K. S. J. Pister, "Optical Communication Using Micro-Corner Cube Reflectors," *Proceedings of the* 10*th IEEE International Micro Electro Mechanical Systems Conference* (*MEMS*'97), Vol. 40, No. 8, 1997, pp. 350-355.

[4] G. Pottie and W. Kaiser, "Wireless Integrated Network Sensors," *Communications of the ACM*, Vol. 43, No. 5, 2000, pp. 51-58.

[5] H. Hashemi, "The Indoor Radio Propagation Channel," *Proceedings of IEEE*, Vol. 81, No. 7, 1993, pp. 943-968.

[6] H. Baltes *et al.*, "Micromachined Thermally Based CMOS Microsensors," *Proceedings of IEEE*, Vol. 86, No. 8, 1998, pp. 1660-1678.

[7] D. Culler, D. Estrin and M. Srivastava, "Overview of Sensor Networks," *IEEE Computer*, Vol. 37, No. 8, 2004, pp. 41-49.

[8] S. Tarannum, D. Prakash, S. George, B. V. Tara, S. Ushe, L. Nalini, K. R. Venugopal and L. M. Patnaik, "Consolidate and Advance: An Efficient QoS Management in Heterogeneous Wireless Sensor Networks," *IEEE ICSCN* 2008, Chennai, January 2008, pp. 93-98.

[9] B. Akan, Y. Sankarasubramaniam and I. F. Akyildiz, "ESRT: Event-to-Sink Reliable Transport in Wireless Sensor Networks," *Proceedings of the* 4*th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Annapolis, Maryland, USA, 2003, pp. 177-188.

[10] M. Kuorilehto, M. Hännikäinen and T. D. Hämäläinen, "A Survey of Application Distribution in Wireless Sensor Networks," *EURASIP Journal on Wireless Communications and Networking*, Vol. 2005, No. 5, 2005, pp. 774-788.

[11] S. Meyer and A. Rakotonirainy, "A Survey of Research on Context-Aware Homes," *Workshop on Wearable, Invisible, Context-Aware, Ambient, Pervasive and Ubiquitous Computing*, Adelaide, 2003, pp. 159-168.

[12] B. Warneke, M. Last, B. Liebowitz and K. S. J. Pister, "Smart Dust: Communicating with a Cubic-Millimeter Computer," *Computer Magazine*, Vol. 34, No. 1, 2002, pp. 44-51.

[13] V. Hsu, M. Kahn and K. Pister, "Wireless Communication for Smart Dust," Electronic Research Laboratory Technical Memorandum, February 1998.

[14] A. G. Ruzzellli, R. Tynan, M. J. O'Grady and G. M. P. O'Hare, "Advances in Wireless Sensor Networks," *Encyclopaedia of Mobile Computing and Commerce* (*EMCC*), Vol. 1, 2006, pp. 1-12.

[15] D. C. Steere, A. Baptista, D. McNamee, C. Pu and J. Walpole, "Research Challenges in Environmental Observation and Forecasting Systems," *Proceedings of the* 6*th Annual International Conference on Mobile Computing and Networking*, Boston, Massachusetts, United States, 2000, pp. 292-299.

[16] E. Biagioni and K. Bridges, "The Application of Remote Sensor Technology to Assist the Recovery of Rare and Endangered Species," *In Special Issue on Distributed Sensor Networks for the International Journal of High Performance Computing Applications*, Vol. 16, No. 3, 2002, pp. 315-324.

[17] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton and J. Zhao, "Habitat Monitoring: Application Driver for Wireless Communications Technology," *Proceedings of the* 2001 *ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*, San Jose, Costa Rica, April 2001, pp. 20-41.

[18] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler and J. Anderson, "Wireless Sensor Networks for Habitat Monitoring," *In ACM International Workshop on Wireless Sensor Networks and Applications* (*WSNA*'02), Atlanta, September 2002, pp. 88-97.

[19] H. Wang, J. Elson, L. Girod, D. Estrin and K. Yao, "Target Classification and Localization in Habitat Monitoring," *Proceedings of the IEEE ICASSP*'03, Hong Kong, April 2003, pp. 597-600.

[20] L. Schwiebert, S. K. S. Gupta and J. Weinmann, "Research Challenges in Wireless Networks of Biomedical Sensors," *Mobile Computing and Networking*, Rome, Italy, 2001, pp. 151-165.

[21] M. Eltoweissy, M. Younis, K. Akkaya and A. Wadaa, "On Handling QoS Traffic in Wireless Sensor Networks," *Proceedings of the* 37*th Annual Hawaii International Conference on System Science*, Hawaii, 2004, pp. 5-8.

[22] M. Ilyas and I. Mahgoub, "Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems," CRC Press, Boca Raton, 2005.

[23] S. Tarannum, S. Aravinda, L. Nalini, K. R. Venugopal and L. M. Patnaik, "Routing Protocol for Lifetime Maximization of Wireless Sensor Networks," *International Journal on Information Processing*, Vol. 1, No. 2, 2007, pp. 58-67.

[24] T. Nieberg, S. Dulman, P. Havinga, L. Hoesel and J. Wu, "Collaborative Algorithms for Communication in Wireless Sensor Networks," University of Twente, Nederlands, 2003.

[25] J. N. A. Karaki and A. E. Kamal, "Routing Techniques in Wireless Sensor Networks: A Survey," *IEEE on Wireless Communications*, Vol. 11, No. 6, 2004, pp. 6-28.

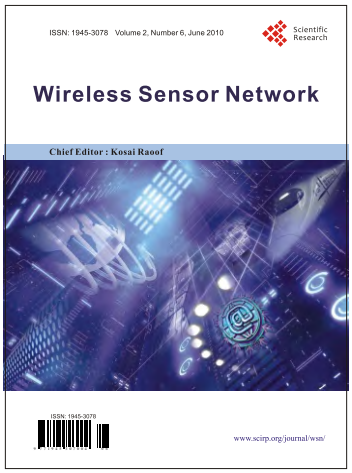
[26] M. Yebari, T. Addali, A. Z. Sadouq and M. Essaaidi, "Energy Conservation Challenges in Wireless Sensor Networks: A State-of-The-Art Study," *International Journal on Information and Communication Technologies*, Vol. 1, No. 2, 2008, pp. 29-35.

[27] Y. Yao and J. Gehrke, "The Cougar Approach to In-Network Query Processing in Sensor Networks," SIGMOD Record, September 2002.

[28] S. Tarannum, S. Srividya, D. S. Asha and K. R. Venugopal, "Dynamic Hierarchical Communication Paradigm for Wireless Sensor Networks: A Centralized, Energy Efficient Approach," *Wireless Sensor Networks*, Vol. 1, No. 4, 2009, pp. 340-349.

[29] S. Tarannum, V. Anitha, A. Priya, K. R. Venugopal and L. M. Patnaik, "Self-Healing AntChain for Increasing Lifespan in Wireless Sensor Networks," *International Engineering and Technology* (*IETECH*) *Journal of Communication Techniques*, Vol. 2, No. 4, 2008, pp. 239-246.

[30] A. Vargas, "OMNET ++ Discrete Event Simulator System," Version 2.3 Edition, 2003.



Wireless Sensor Network (WSN)

Call for Papers

http://www.scirp.org/journal/wsn

ISSN 1945-3078 (Print) ISSN 1945-3086 (Online)

WSN is an international refereed journal dedicated to the latest advancement of wireless sensor network and applications. The goal of this journal is to keep a record of the state-of-the-art research and promote the research work in these areas.

Editor-in-Chief

Dr. Kosai Raoof , GIPSA LAB, University of Joseph Fourier, Grenoble, France

Subject Coverage

This journal invites original research and review papers that address the following issues in wireless sensor networks. Topics of interest are (but not limited to):

- Network Architecture and Protocols
- Self-Organization and Synchronization
- Quality of Service
- Data Processing, Storage and Management
- Network Planning, Provisioning and Deployment
- Integration with Other Systems
- Software Platforms and Development Tools
- Routing and Data Dissemination
- Energy Conservation and Management
- Security and Privacy
- Developments and Applications
- Network Simulation and Platforms

We are also interested in short papers (letters) that clearly address a specific problem, and short survey or position papers that sketch the results or problems on a specific topic. Authors of selected short papers would be invited to write a regular paper on the same topic for future issues of the WSN.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. Authors are responsible for having their papers checked for style and grammar prior to submission to WSN. Papers may be rejected if the language is not satisfactory. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/wsn

Email: wsn@scirp.org

Wireless Sensor Network, 2010, 2, 411-491
Published Online June 2010 in SciRes (http://www.SciRP.org/journal/wsn/).

TABLE OF CONTENTS

Volume 2 Number 6 **June 2010**

