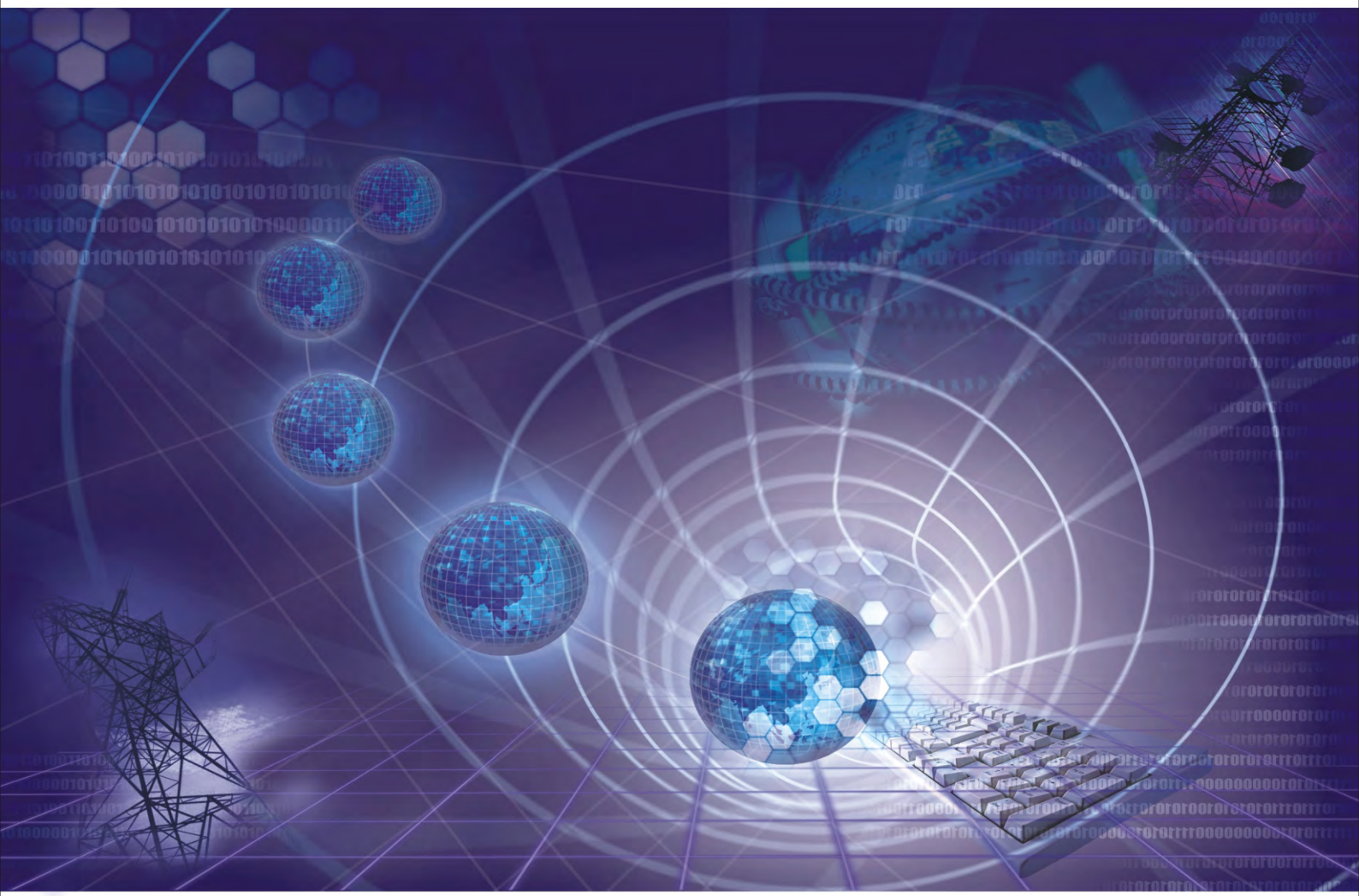
International Journal of

Communications, Network and System Sciences

ISSN: 1913-3715

Volume 2, Number 8, November 2009



www.scirp.org/journal/ijcns/

Scientific Research

JOURNAL EDITORIAL BOARD

ISSN 1913-3715 (Print) ISSN 1913-3723 (Online)
http://www.scirp.org/journal/ijcns/

Int. J. Communications, Network and System Sciences, **2009, 2, 695-803**
Published Online November 2009 in SciRes (http://www.SciRP.org/journal/ijcns/).

TABLE OF CONTENTS

Volume 2 Number 8 November 2009

A Comparative Study of Medium Access Control Protocols for Wireless Sensor Networks
M. GUNN, S. G. M. KOO……695

Service Adaptable 3G Turbo Decoder for Indoor/Low Range Outdoor Environment
C. CHAIKALIS, N. S. SAMARAS……704

Application of a Dynamic Identity Authentication Model Based on an Improved Keystroke Rhythm Algorithm
W. C. YANG, F. FANG……714

Evaluation of Network Stack Optimization Techniques for Wireless Sensor Networks
J. JEONG……720

A Cooperative Location Management Scheme for Mobile Ad Hoc Networks
D. LI, J. C. WANG, L. P. ZHANG, H. LI, J. ZHOU……732

A Perceptual Approach to Reduce Musical Noise Using Critical Bands Tonality Coefficients and Masking Thresholds
C. V. R. RAO, M. B. R. MURTHY, K. S. RAO……742

A Real-Time Measurement Algorithm for Available Bandwidth
Y. YIN, W. D. WU…… 746

Modified Ceiling Bounce Model for Computing Path Loss and Delay Spread in Indoor Optical Wireless Systems
K. SMITHA, A. SIVABALAN, J. JOHN……754

A MAC Scheme with QoS Guarantee for MANETs
Y. B. YANG, Y. L. WEI……759

A Reputation-Based Multi-Agent Model for Network Resource Selection
J. F. TIAN, J. LI, L. D. YANG……764

Subcarrier Availability in Downlink OFDM Systems with Imperfect Carrier Synchronization in Deep Fading Noisy Doppler Channels
L. NOOR, A. ANPALAGAN, S. KANDEEPAN……775

Performance Analysis of MAC Protocol for LEO Satellite Networks
M. X. GUAN, R. C. WANG……786

Ant Colony Optimization Based on Adaptive Volatility Rate of Pheromone Trail
Z. Q. CAI, H. HUANG, Y. QIN, X. H. MA……792

A Review of Wireless Body Area Networks for Medical Applications
S. ULLAH, P. KHAN, N. ULLAH, S. SALEEM, H. HIGGINS, K. S. KWAK……797

International Journal of Communications, Network and System Sciences (IJCNS)

Journal Information

SUBSCRIPTIONS

The *International Journal of Communications, Network and System Sciences* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc.,USA.

E-mail: service@scirp.org

Subscription rates: Volume 2 2009

Print: $50 per copy.

Electronic: free, available on www.SciRP.org.

To subscribe, please contact Journals Subscriptions Department, E-mail: service@scirp.org

Sample copies: If you are interested in subscribing, you may obtain a free sample copy by contacting Scientific Research Publishing, Inc at the above address.

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: service@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:

E-mail: ijcns@scirp.org

Int. J. Communications, Network and System Sciences, **2009, 2, 695-703**
doi:10.4236/ijcns.2009.28080 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Comparative Study of Medium Access Control Protocols for Wireless Sensor Networks

Meghan GUNN, Simon G. M. KOO
Department of Mathematics and Computer Science, University of San Diego, San Diego, USA
Email: {*meghangunn-09, koo*}*@sandiego.edu*
Received July 7, 2009; *revised August* 12, 2009; *accepted September* 21, 2009

Abstract

One of the major constraints in wireless sensor networks (WSNs) is power consumption. In recent years, a lot of efforts have been put into the design of medium access control (MAC) protocols for WSN, in order to reduce energy consumption and enhance the network's lifetime. In this paper, we surveyed some MAC protocols for WSN and compared their design tradeoffs. The goal is to provide a foundation for future MAC design, and to identify important design issues that allow us to improve the overall performances.

Keywords: Wireless Networks, Sensor Networks, Performance Evaluation

1. Introduction: Wireless Sensor Networks

A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices that use sensors to cooperatively monitor physical or environmental conditions such as temperature, sound, vibration, pressure, motion, or pollutants at different locations. The purpose of a WSN is to collect and process data from a target domain and transmit information back to specific sites. WSN technology is an emerging technology that can be utilized in a wide range of potential applications including but not limited to, biomedical treatment, military applications, traffic surveillance, fire detection, structural and earthquake monitoring, industrial control, and rescue operations.

Such a network usually consists of a number of wireless sensor nodes that arrange themselves into a multihop network. Each node consists of one or more sensors, a low power radio transceiver or other wireless communication device, an embedded processor, and an energy source, usually a battery. The size of a wireless node can vary from the size of a shoebox, down to size of a grain of dust, and cost varies depending on size. These size and cost constraints result in corresponding constraints on nodes resources, including energy, memory, computational speed and bandwidth.

2. Factors for Design of a MAC Layer Protocol

Considering that sensor nodes are likely to be battery powered, and because they are often implemented in environments where it proves to be difficult to change or recharge batteries, prolonging the lifetime of nodes is a critical issue for a successful wireless sensor network. Not only does the transmission of data cost energy, but receiving, and scanning for data also use a significant amount of energy. In addition to being energy efficient, WSN should be scalable and adaptable to change. Change can come in the form of network size, node density, or topology. Additionally, nodes may die over time, new nodes may join, or nodes may move to a different location. A good MAC protocol should gracefully accommodate such network changes. Lastly fairness, latency, throughput, bandwidth utilization are also concerns for WSN. However, these goals may be primary concerns in traditional wireless networks, but they prove to be secondary for WSN. This is due to the fact that in a traditional wireless network, usually a number of different applications may be competing for use of the communication channel; however in a WSN, the nodes are typically working for the same application.

3. Sources of Energy Waste

In a sensor, the Radio Frequency (RF) module, which consumes most of the energy, becomes the crucial entity to be optimized. Therefore, designing an energy-efficient Medium Access Control (MAC) protocol is significant factor in reducing energy consumption based on its direct control over RF module [2]. There are four distinctive sources of energy waste for wireless sensor nodes, colli-

sions, overhearing, control packet overhead, and idle listening. Collisions are caused by contention, when two nearby sensor nodes both attempt to access the communication channel at the same time. Overhearing is a result of a node picking up packets that are destined to other nodes. Common control packets used in WSN include Ready-to-Send (RTS), Clear-to-Send (CTS), and Acknowledge (ACK). The transmission of these packets contributes to energy consumption, therefore a minimal number of control packets should be used to make a data transmission. Idle listening has proved to be one of the major sources of energy waste. Given that a node does not know when it will be the receiver of a message from one of its neighbors, it must keep its radio in receive mode at all times, resulting in idle listening. Studies have shown that idle listening can consume 50-100% of the power required for receiving [1].

4. Proposed MAC Protocols

There are a considerable number of MAC protocols that have been designed and implemented for WSN. This section will discuss a few of these protocols and their essential behaviors.

4.1. Sensor MAC (S-MAC)

The key idea behind S-MAC [1] is the utilization of managed synchronized duty cycles. A duty cycle utilizes a periodic awake and sleep schedule, allowing nodes in sleep mode turn off their radio [1]. A duty cycle is represented as a ratio of wake time to total cycle time, S-MAC limits it duty cycles to about 10%, reducing energy waste by attempting to minimize idle listening. Sleep and listen periods are predefined and constant in S-MAC.

Additionally, nodes in S-MAC create virtual clusters by periodically exchanging sleep schedules with their neighboring nodes [1]. This exchange is implemented by sending a SYNC packet, which is very short, and includes the address of the sender and the time of its next sleep. Nodes that receive the SYNC packet will adjust their timers immediately after they receive the SYNC packet and will go to sleep when the timer fires. Thus the schedules are updates and the nodes are synchronized.

Nodes that reside in two virtual clusters wake up for the listen phases for both clusters. This however is one of the drawbacks of the S-MAC algorithm, the possibility of a node following two different schedules resulting in more energy consumption via idle listening and overhearing.

Lastly, S-MACs design includes the utilization of adaptive listen, overhearing avoidance techniques, and message passing. With adaptive listen, neighboring nodes wake up for a short period of time at the end of each transmission to listen for possible data transmissions. To avoid overhearing, all immediate neighbors of sender and receiver are put to sleep upon receiving RTS/CTS. Resultantly neighbors do not overhear data packets and following ACKS. The nodes use the duration field in the packet, which indicates how long to sleep. Message passing is a technique in which long messages are divided into frames and sent in a burst. With this method, nodes may achieve energy savings by minimizing communication overhead and latency at the expense of unfairness in medium access.

4.2. Timeout MAC (T-MAC)

T-MAC is similar to S-MAC in that it utilizes an active/sleep duty cycle. However, TMAC improves upon the design of S-MAC by introducing an adaptive duty cycle in which the active part is dynamically ended, increasing the efficiency of the algorithm for variable traffic loads. The idea behind the design of T-MAC is as follows. While latency requirements and buffers space are generally fixed, the message rate will usually vary. Therefore, the nodes must be implemented with an active time that can handle the highest expected load. Whenever the load is lower than that which is expected, the active time is not optimally used and energy will be wasted on idle listening. To solve this inefficiency, the T-MAC protocol implementation reduces idle listening by transmitting all messages in bursts of variable length, and sleeping between bursts. To maintain an optimal active time under variable load, the length of the active time is dynamically determined, ending in an intuitive way by timing out when the node hears nothing [4].

Every node periodically wakes up to communicate with its neighbors during active time periods. The nodes communicate using a modified RTS-CTS-DATA-ACK four-step exchange to deliver messages, which provides both collision avoidance and reliable transmission [4]. A node will keep listening and potentially transmitting, as long as it is in an active period. An active period ends when no activation event has occurred for a time TA. An activation event includes but is not limited to, the reception of any data on the radio, the sensing of communication of the radio, the end-of-transmission of a node's own data packet or an ACK data packet. If no activation event is sensed, the node then goes to sleep again until the next frame. During the sleep mode, new messages are queued.

An important aspect of T-MAC is determining TA, the time that a node must wait before it times out, and goes to sleep. The idea is that a node should not go to sleep while its neighbors are still communicating, since it may be the receiver of a subsequent message [4]. Receiving the start of the RTS or CTS packet from a neighbor is

enough to trigger a renewed interval TA. Additionally, a node may be out of range, and therefore may not hear the RTS that starts a communication with its neighbor, so the interval TA must be long enough to receive at least the start of the CTS packet.

4.3. S-MACL, a Global Sleeping Schedule

As mentioned the S-MAC protocol creates virtual clusters in which the clustered nodes follow a common sleeping schedule. In order to connect these virtual clusters, nodes residing between clusters have to adopt multiple schedules. These nodes, known as border nodes, constitute nearly 50 percent of the nodes in some networks and may have to adopt up to 4 different schedules. These border nodes have to stay in active mode longer than other node, which means that they waste more energy than non-border nodes. Resultantly, these nodes will die sooner, and the network coverage rate is reduced. A more serious problem happens in multi-hop sensor networks, in which border nodes have to act as intermediate outers to relay packets. The death of these border nodes may increase the routing difficulty, even segment a network. Some nodes will not be able to communicate to the rest of the network [6].

To resolve the problem of multiple sleeping schedules, S-MACL attempts to merge all the virtual synchronization clusters into one cluster to ensure that only one sleeping schedule will be used in a fully connected network. To do this, S-MACL utilizes the node id, a unique identifier that is mounted on each sensor node. More specifically it uses the id of the synchronizing sender node and applies it as a schedule id. The scheduling process in S-MACL is presented as follows. When a node does not receive any SYNC frame after its first listening period, it will arbitrarily choose one schedule and announce this schedule and assign its own id as the schedule id. We call such a node a synchronizer, since it chooses its schedule independently and other nodes will synchronize with it. Otherwise, the node will receives a schedule from a neighbor SYNC frame before having a chance to choose its own schedule, and will follow that schedule by setting its schedule as the same, and announcing this schedule to its neighbors. We call such a node a follower. When a node receives a different schedule from its neighbors' SYNC frame, it will compare the current schedule id and the new schedule id. Then it will start following the schedule with the higher id. If the new schedule in the incoming SYNC frame has a lower id, this node will announce its own schedule during the listening time of the new schedule. This operation ensures that nodes will always use the schedule with highest id. The authors show through various scenarios, with different numbers of nodes and different topologies that S-MACL performs better than S-MAC in most cases [6].

4.4. Patten MAC (P-MAC)

P-MAC [14] is unique in that instead of having fixed sleep and awake schedules as with S-MAC, the sleep-wakeup schedules of the sensor nodes are adaptively determined, based on a node's own traffic and that of its neighbors. This improves throughput under heavy traffic and reduces unwanted energy consumption while the networks is performing under light loads when compared to the performance S-MAC.

Similar to S-MAC, P-MAC is a time-slotted protocol, however unlike S-MAC in which a node sleeps for a duration of the time slot and is awake for the remainder, in P-MAC, the node must either be awake or asleep for the entire duration of the time slot. With P-MAC, a sensor node gets information about the activity in its neighborhood before sending communication packets through patterns. Based on these patterns, a sensor node can put itself into a long sleep for several time frames when there is no traffic in the network. If there is any activity in the neighborhood, a node will know this through the patterns and will wake up when required. Thus P-MAC saves more power than S-MAC as well as T-MAC, without compromising on the throughput.

A sleep-wakeup pattern is a stream of bits indicating the tentative sleep-wakeup plan for a sensor node over several slot times [14]. A 1 in the stream indicates that the node intends to stay awake during a slot time, while a 0 indicates that the node intends to sleep. Since the pattern is only a tentative plan, it is subject to change. This pattern stream of 1s and 0s is generated for each individual node. These patterns are used to convey activity from one node to its neighbors. Thus, the schedule for a node is derived from its own pattern and, as well as the patterns of its neighboring nodes, resulting in a schedule for the network.

Pattern generation based on the binary strings that are associated with a node over some number of time slots, this is referred to as a period [14]. The nodes' pattern is updated during each period using local traffic information available at the node and exchanged between the neighboring nodes at the end of each period. When the network is activated, the pattern at every node has just one bit for the first period, which is 1. If there is no data for a node to send at the first time slot of bit 1, then it indicates that the traffic load is light, and the node can afford to go to sleep. Consequently, the node updates its pattern to 01, and so on. If during the next time slot, the node still has no data to send, the node is encouraged to sleep longer by doubling the number of 0 bits, ie. 001. By exponentially increasing the sleep time during light traffic the node is able to save a considerable amount of energy. On the other hand, if a node has any data to transmit at any time slot, regardless of the pattern bit at that time slot, the next bit in the pattern becomes a 1.

These patterns are not the decisive sleep schedule for the nodes; they are only a tentative sleep-wakeup plan [14]. As mentioned P-MAC obtains it schedule based on the node's pattern, and the pattern of its neighbors. The nodes broadcast newly generated patterns at the end of the current period. As a result, the time is divided into time frames, referred to as super time frames (STF). Each STF has two sub-frames. In the first, the Pattern Repeat Time Frame (PRTF), each node repeats its current pattern. The second time frame, the Pattern Exchange Time Frame (PETF), is used for the exchange of new patterns between neighbors. To obtain the actual sleep-wakeup itinerary, the strings of bits are compared between neighbors at each time slot as well as looking for data packets in the buffer of the neighboring nodes at each time slot. Based on a series of rules, the bits are compared and a new pattern is created and followed. In addition to the use of 1s and 0s, 1- bit is introduced. A 1-bit will be used when the nodes pattern bit is 1 and either the receivers bit is 0, or the node has no packets to be sent. Therefore, 1- implies that the node should wakeup at the beginning of the time slot and listen for a short amount of time. If it hears no communication from its neighbors, then it goes back to sleep. The reason for this is that since the pattern bit for the node is 1, the node is a candidate to be a receiver of communication and its neighbors may try to send data to it. Thus, if the node goes to sleep, the packet destined to it will be lost, and energy is wasted.

4.5. Traffic-Adaptive MAC (TRAMA)

As a traffic load increase, the probability of collisions of control or data packets occurring in any contention-based scheme increases. This degrades channel utilization and further reduces battery life [7]. TRAMA implementation attempts to provide energy-efficient conflict free channel access in wireless sensor networks by creating transmission schedules that are adaptive to changes, prolongs the battery life of each node, and is robust to wireless loses [7]. The protocol consists of three components: the Neighbor Protocol (NP), the Schedule Exchange Protocol (SEP) and the Adaptive Election Algorithm (AEA). Additionally, TRAMA uses single, time-slotted channel access that is divided up into random and scheduled access periods.

The main function of the Neighbor Protocol is to gather two-hop neighborhood information by using signaling packets. This protocol operates periodically during random access periods. Schedule Exchange Protocol utilizes a schedule consisting of intended receivers for future transmission slots. Schedules are established based on the current traffic information at the node, and are periodically propagated to the neighboring node. SEP maintains consistent schedules for the one-hop neighbors of each node. The Adaptive Election Algorithm uses the schedule information from SEP and the neighborhood information to elect a transmitter, receiver and stand-by nodes for the current time slot. Nodes that are not selected to transmit or receive data for a particular time slot are removed from the election process, allowing them to switch to sleep mode and improving the channel utilization. As a result, the sleep schedule of a node is a direct function of the traffic going through the node and its neighbors, and is synchronized automatically when nodes exchange information about their identifiers and their traffic [7].

TRAMA organizes access to the communication channel into time slots allowing random and scheduled access. Random Access periods are used for signaling, synchronization, and updating two-hop neighbor information. The scheduled access periods are used for contention free data exchange between nodes.

4.6. B-MAC, a Versatile Low Power MAC

B-MAC is a carrier sense media access (CSMA) protocol that utilizes low power listening and an extended preamble to achieve low power communication [10]. Furthermore, B-MAC is designed for duty cycled WSN, so nodes have an awake and a sleep period, and each node can have an independent schedule.

Periodic channel sampling or low-power listening (LPL) is the primary technique that B-MAC employs. LPL is carried out as follows. A node wakes up every check-interval; it turns on the radio and samples the channel. If activity (a preamble) is detected, the node remains awake for the time required to receive the incoming data packet. After reception, the node returns to sleep. However, if no packet is received, a timeout forces the node back to sleep.

If a node wishes to transmit, it precedes the data packet with a preamble that is slightly longer than the sleep period of the receiver. The preamble is predefined data automatically appended at the beginning of transmitted data. By using an extended preamble, that is at least as long as the sleep period, a sender is assured that at some point during the transmission of the preamble, the receiver will wake up and detect the preamble, and remain awake to receive the data packet.

A key challenge of B-MAC is implementing check intervals that are very short which then ensure a reasonable length for the preamble. Carrier sense duration also has to be very short so that receiver does not have to spend too much energy listening to the communication channel. A carrier sense must be accurate to reduce latency of transmission and energy consumption at sender.

B-MAC additionally utilizes software automatic gain control as a method of Clear Channel Assessment (CCA), which accurately determines if the channel is clear, thus

effectively avoiding collisions. This is a necessity so that the node can determine what is a noise and what is a signal, due to the fact that ambient noise is prone to environmental changes. This is achieved by taking signal strength samples when the channel is assumed to be free, such as immediately after transmitting a packet. These samples are stored in a FIFO queue and the median of the queue is added to an exponentially weighted moving average with decay. This value gives a fairly accurate estimate of the noise floor of the channel. Effectively, a node, before transmission, takes a sample of the channel; if the noise is below the noise floor, the channel is clear and it can send immediately [10].

4.7. X-MAC, a Short Preamble MAC

While being simple and improving energy efficiency, the low power listening approach used by B-MAC which employs a long preamble is suboptimal in terms of energy consumption, is subject to overhearing, as well as introducing excess latency at each hop [11]. This issue is threefold. First, the receiver typically has to wait the full period until the preamble is finished before the DATA/ACK exchange can begin, even if the receiver has woken up at the start of the preamble. Second, LPL suffers from the overhearing problem, where receivers who are not the target of the sender also wake up during the long preamble and have to stay awake until the end of the preamble to find out if the packet is destined for them. This wastes energy at all non-target receivers within transmission range of the sender. Third, because the target receiver has to wait for the full preamble before receiving the data packet, the per-hop latency is lower bounded by the preamble length. Over a multi-hop path, latency can accumulate to become substantial [11].

X-MAC is a low power MAC protocol that strives to overcome these shortcomings by employing a shortened preamble approach. The ideas behind this approach is to embed address information of the target node in the preamble so that non-target receivers can realize that they are not the receiver and quickly go back to sleep. This solution addresses the overhearing problem. Furthermore, X-MAC introduces the strobed preamble. This approach allows the target receiver to interrupt the long preamble as soon as it wakes up and determines that it is the target receiver. This is accomplished by dividing the one long preamble into a series of short preamble packets, each containing the id of the target node. Accordingly, instead of sending a constant stream of preamble packets, the protocol inserts small pauses between the series of short preamble packets, during which time the transmitting node pauses to listen to the medium. These gaps enable the receiver to send an early ACK packet back to the sender by transmitting the ACK during the short pause between preamble packets. When a sender receives an ACK from the intended receiver, it stops sending preambles and sends the data packet. This allows the receiver to cut short the excessive preamble, which reduces per-hop latency and energy spent unnecessarily waiting and transmitting [11].

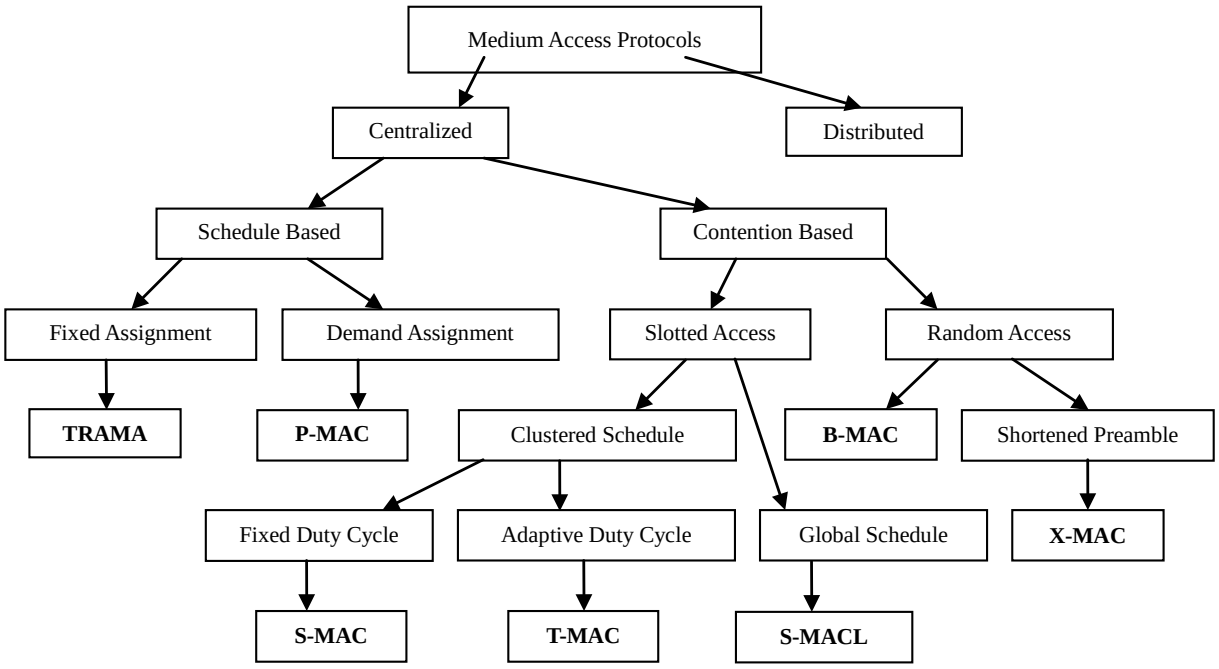


Figure 1. MAC design options.

Table 1. Tradeoff analysis.

	Energy	Fairness	Latency	Throughput
S-MAC	(+) Periodic Sleep (+) Message Passing (-) Idle Listening (-) Overhearing	(-) Message Passing	(-) Periodic Sleep (+) Adaptive listen (+) Message Passing	(-) Periodic Sleep
T-MAC	(+) Adaptive Duty Cycle		(-) Adaptive Duty Cycle	(-) Adaptive Duty Cycle
S-MACL	(+) Global Sleep Schedule			
P-MAC	(+) Adaptive Sleep Schedules			(+) Adaptive Sleep Schedules
TRAMA	(+) Transmission Schedules (-) Overhearing	(+) Transmitter Electron Algorithm		(+) Transmitter Electron Algorithm
B-MAC	(+) LPL (-) Long Preamble (-) Overhearing			
X-MAC	(+) Shortened Preamble		(+) Strobed Preamble	

5. A Comparison

5.1. Comparison of Design

Centralized MAC protocol design can be divided into two sub sections, schedule- based, and contention-based. A schedule based design schedules nodes into different sub-channels. Schedules protocols are successful in that they avoid collisions thus promoting energy efficiency. However, they tend to have poor scalability and adaptability. On the other hand in a contention-based schedule, nodes compete in a probabilistic coordination for access to the communication channel. Contention-based protocols have proved to be more scalable and flexible to topology change. However, when compared with schedule-based designs, they are not as energy efficient. Figure 1 illustrates which of the previous protocols discussed use each of the schemes and partitions the design of the protocols into more detailed subsets.

5.2. Analysis of Tradeoffs

Table 1 represents comparison of the tradeoffs in protocol design based on the statistics available. Due to resource constraints, the table is not complete. A (+) indicates a positive outcome of the subsequent design method. A (-) indicates a tradeoff of a network performance metric as a result of the design technique.

5.3. Protocol Comparison

This section offers a more detailed discussion of the advantages and disadvantages of the protocol design. It also offers a direct comparison between some of the protocols. Again, for some of the protocols, the information is minimal due to a lack of available information.

5.3.1. S-MAC

S-MAC reduces the amount of energy wasted by idle listening, which is accomplished by introducing sleep schedules. Its implementation is simple, and time synchronization overhead is prevented with sleep schedule announcements. Lastly, adaptive listening is used to reduce multi-hop latency due to periodic sleep modes and nodes waiting until the subsequent listen period of the intended receiver. Adaptive listen saves more energy for heavy loads by reducing latency by at least half.

On the other hand, the S-MAC protocol essentially trades energy efficiency for reduced throughput and increased latency. Throughput is reduced because only the active part of the frame is used for communication. Latency increases because a message-generating event may occur during sleep time. Additionally, adaptive listening incurs overhearing or idle listening if the packet is not destined to the listening node. Lastly, sleep and listen periods are predefined and constant, which decreases the efficiency of the algorithm under variable traffic load.

For light traffics loads S-MAC offers significant energy efficiency over always listening MAC protocols. Simulation experiments have shown that the S-MAC protocol reduces the energy used by the radio with up to 30%, after optimal tuning. The energy savings and increased throughput of S-MAC as compared with traditional protocols without sleep cycles such as CSMA and IEEE 802.11 without duty cycle control is shown in Figures 2 and 3.

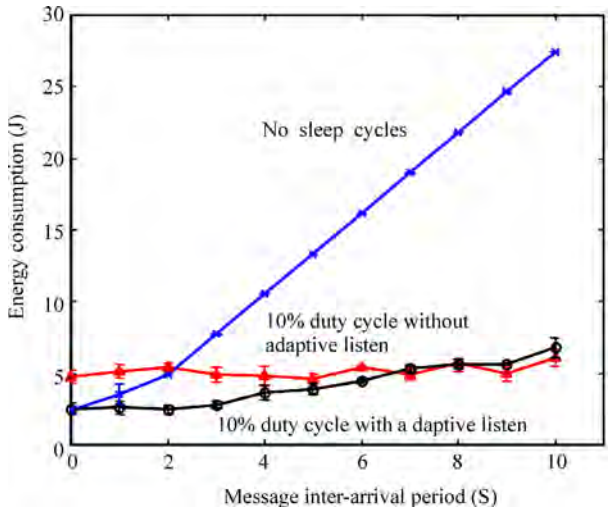


Figure 2. Energy consumption at different traffic loads.

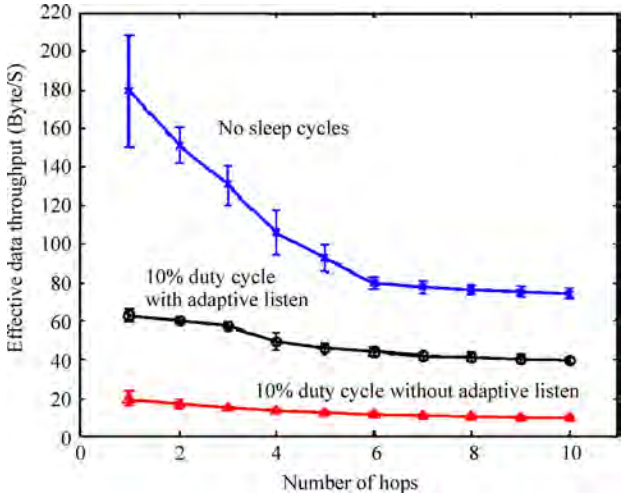


Figure 3. Effective through put under highest traffic load.

Figure 2 shows that at light traffic load, periodic sleeping has significant energy savings over fully active mode and adaptive listen saves more at heavy load by reducing latency. In Figure 3 one can see that adaptive listen significantly increases throughput.

5.3.2. T-MAC

Simulation experiments have shown that the T-MAC protocol reduces the energy used by the radio with as much as 80% in a typical scenario when compared to classical protocols like CSMA. The S-MAC protocol saves only 30% in this scenario, after optimal tuning. Implementation of the T-MAC protocol on real wireless sensor hardware has shown that, in an idle situation, the radio can be turned off for as much as 97.5% of the time, reducing the total energy used more than 96%. In a situation with high message rates, the T-MAC protocol does

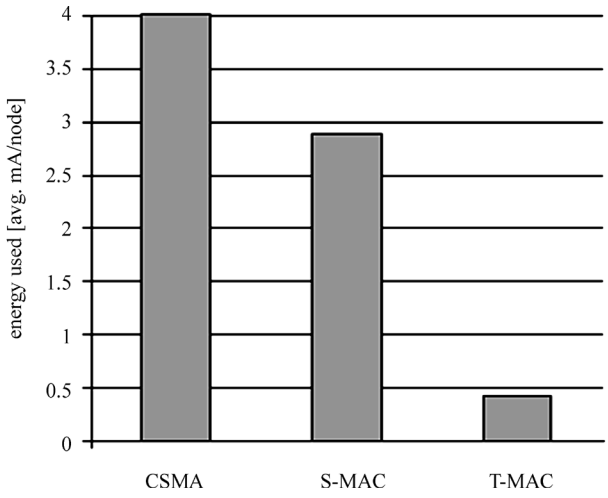


Figure 4. Energy consumption based on event triggered reporting.

not increase the latency, since nodes do not sleep in that case. Furthermore, the authors show that, for variable workloads, T-MAC uses one fifth of the energy used by S-MAC. While this adaptive duty cycling reduces energy usage for variable workloads, these gains come at the cost of reduced throughput and increased latency. Results of simulations are illustrated in Figure 4, which compares the amount of energy used for CSMA, S-MAC, and T-MAC in a typical scenario.

5.3.3. S-MACL

With S-MACL, all nodes consume less energy, especially the border nodes that act as intermediate routers, greatly increasing the lifetime of these nodes. Additionally, as a result of the global synchronization schedule, the number of collisions is reduced, which also reduces the amount of energy wasted. The contrastive simulation of S-MAC with S-MACL results showed that S-MACL achieves a great level of energy efficiency compared with S-MAC.

5.3.4. P-MAC

Based on simulations done by the authors, in comparison to S-MAC under light traffic loads, P-MAC consumes less energy, though throughput remains the same. However, under heavy traffic loads, P-MAC consumes less energy and achieves a higher throughput. This is due to the fact that with S-MAC, sensor nodes must periodically go to sleep, even if the traffic load is high. On the other hand, the implementation of P-MAC allows the nodes to stay awake due to the varying schedule patterns. Because PMAC is able to adaptively schedule sleep and awake periods, it offers more energy savings under light loads, and higher throughput under heavy loads as compared to S-MAC.

5.3.5. TRAMA

TRAMA is able to improve energy efficiency by utilizing transmission schedules that avoid collisions of data packets at the receiving nodes. Additionally nodes switch to low power radio mode when there are no data packets intended for those nodes. Furthermore, TRAMA achieves conflict-free transmission by scheduling access among two-hop neighboring nodes during a particular time slot and by allowing nodes to switch to sleep mode if they are not selected to transmit or are not the intended receivers of traffic for a particular time slot. Lastly, adequate throughput and fairness is achieved based on the transmitter-election algorithm that is inherently fair and promotes channel reuse as a function of the competing traffic around any given source or receiver. On the other hand, TRAMAs efficiency is limited by its complex election algorithm and data structure. Moreover, it incurs overhead due to explicit schedule propagation as well as higher queuing delays [7].

TRAMA implementation results in a higher percentage of sleep time and less collision probability when compared to CSMA based protocols, which greatly improves energy savings. TRAMA has a higher delay but higher maximum throughput than contention-based S-MAC. Through extensive simulations, TRAMAs performance is compared against a number of contention and a scheduled based MACs. It is evident from the simulation results that significant energy savings can be achieved by TRAMA depending on the offered load. TRAMA also achieves higher throughput (around 40% over S-MAC and CSMA and around 20% over 802.11) when compared to contention-based protocols because it avoids collisions due to hidden terminals [7].

5.3.6. B-MAC

The authors have show that testing the communication channel for activity is about 10x less expensive than listening for a full contention period. Idle listening is reduced in the B-MAC protocols by shifting the burden of synchronization to the sender: when a sender has data, the sender transmits a preamble that is at least as long as the sleep period of the receiver; thus, the sender and receiver can be completely decoupled in their duty cycles [10]. This removes the need for, and the overhead introduced by, synchronized wake/sleep schedules.

The authors show that B-MAC surpasses existing protocols in terms of throughput, latency, and for most cases energy consumption. It is simple in both design and implementation. While B-MAC performs quite well, it suffers from the overhearing problem, and the long preamble dominates the energy usage. Additionally, while, unscheduled sleep reduces control overhead, consequentially, the sender incurs greater overhead to wake up the unsynchronized receiver from sleep.

The performance benchmark has shown that B-MAC outperforms S-MAC with greater energy savings and network performance [7].

5.3.7. X-MAC

B-MAC requires more time to transfer packets from the source to the destination. This is because the entire preamble has to be always sent, even though the receiver was already awake. X-MAC saves this time, thus conserving energy.

6. Conclusions

When developing a MAC protocol, prolonging lifetime for nodes is a critical issue to consider in order to promote for a successful wireless sensor network. Many of the developed protocols are developed with specific assumptions in mind and for specific applications. In this article, we surveyed wireless MAC protocols for wireless sensor networks, and we can conclude that no protocol is the "best" implementation. However, each of these protocols addresses different issues that arise from energy waste in sensor nodes.

7. References

[1] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor network," In Proceedings of the INFOCOM'02. IEEE Computer Society, San Francisco, 2002.

[2] J. Ai, J. F. Kong, and D. Turgut, "An adaptive coordinated medium access control for wireless sensor networks," In Computers and Communications Proceedings, ISCC'04, 2004.

[3] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," In Computer Networks, Vol. 38, No. 4, pp. 393–422, 2002.

[4] T. Zheng, S. Radhakrishnan, and V. Sarangan, "PMAC: An adaptive energy-efficient MAC protocol for Wireless Sensor Networks," In Parallel and Distributed Processing Symposium Proceedings, 19th IEEE International**,** 2005.

[5] LAN MAN Standards Committee of the IEEE Computer Society, IEEE Std 802.11-1999, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE, 1999.

[6] L. Zhang, G. Somnath, V. Prakash, and S. Samar, "An energy efficient wireless sensor MAC protocol with global sleeping schedule," In Computer Science and its Applications, CSA'08, International Symposium, pp. 303– 308, October 2008.

[7] V. Rajendran, K. Obraczka, and J. J. Garcia-Luna-Aceves "Energy-efficient, collision-free medium access control for wireless sensor networks," In Wireless Networks, Vol. 12, pp. 63, 2006.

[8] J. M. So and N. Vaidya, "Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using

a single transceiver," Talk at Workshop with Intl. School on WSN, Dagstuhl, Germany, August 30, 2005.

[9] C. L. Fullmer and J. J. Garcia-Luna-Aceves, "Solutions to hidden terminal problems in wireless networks," In Proceedings ACM SIGCOMM'97, Cannes, France, September 14-18, 1997.

[10] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," In Second ACM Conference on Embedded Networked Sensor Systems, 2004.

[11] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: A short preamble MAC protocol for duty-cycled wireless sensor networks," In Proceedings of 2nd ACM Conference on Embedded Networked Sensor Systems (SenSys'06), pp. 307–320, 2006.

[12] S. Singh and C. Raghavendra, "PAMAS: Power aware multi-access protocol with signalling for ad hoc networks," ACM SIGCOMM Computer Communication Review, Vol. 28, No. 3, pp. 5–26, July 1998.

[13] P. Karn, "MACA - a new channel access method for packet radio," In ARRL/CRRL Amateur Radio 9th Computer Networking Conference, pp. 134–140, 1990.

[14] LAN MAN Standards Committee of the IEEE Computer Society, IEEE Std 802.11-1999, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE, 2000.

Int. J. Communications, Network and System Sciences, **2009, 2, 704-713**
doi:10.4236/ijcns.2009.28081 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Service Adaptable 3G Turbo Decoder for Indoor/Low Range Outdoor Environment

Costas CHAIKALIS, Nicholas S. SAMARAS
Department of Informatics & Telecommunications, TEI of Larissa, Larissa, Greece
Email: *kchaikalis@teilar.gr*
Received July 28, 2009; *revised September* 6, 2009; *accepted October* 7, 2009

Abstract

For the well-known 3G mobile communications standard UMTS, four different service classes have been specified. Considering two turbo decoding algorithms, like SOVA and log-MAP, it would be desirable to use an efficient turbo decoder. In this paper this decoder is shown to adapt dynamically to different service scenarios, considering parameters like performance and complexity for indoor/low range outdoor operating environment. The scenarios show that for streaming service class real-time class applications the proposed decoding algorithm depends on data rate; for the majority of scenarios SOVA is proposed, whereas log-MAP is optimal for increased data rates and medium-sized frames. On the other hand, conversational service class real-time applications cannot be established. For the majority of non real-time applications (interactive and background service classes) either algorithm can be used, while log-MAP is proposed for medium data rates and frame lengths.

Keywords: Reconfigurable Systems, Turbo Decoder, UMTS, Flat Reyleigh Fading, Indoor/Low Range Outdoor Operating Environment

1. Introduction and UMTS Data Flow

Channel coding is a critical signal processing element in modern mobile communications systems. Turbo codes [1] represent a powerful channel coding technique. Universal Mobile Telecommunications System (UMTS) belongs to the third generation (3G) of mobile communication systems. Turbo codes have been incorporated as a channel coding scheme in UMTS for data rates higher or equal to 28.8 kbps [2]. They also provide high coding gains in flat fading channels with the use of outer block interleaving [3,4]. Soft-input/soft-output (SISO) decoder is part of a turbo decoder and two candidate algorithms to be used in a SISO decoder are soft output Viterbi algorithm (SOVA) and log maximum a-posteriori (log-MAP) algorithm [2,5–7].

A reconfigurable turbo decoder can be derived according to the common operations of the two algorithms, optimal in terms of performance and latency [8,9,10]. We consider just SOVA and log-MAP and not other turbo decoding algorithms like max-log-MAP or MAP, because SOVA is better in terms of delay, while log-MAP is better in terms of performance [3,5].

SOVA and log-MAP algorithms share common operations which have been addressed in [8–10]. These common operations form a turbo decoder which can be reconfigured and choose the suitable turbo decoding algorithm for different applications (reconfigurable SOVA/log-MAP turbo decoder). In [8] and [10] is also shown that in a reconfigurable SOVA/log-MAP turbo decoder scaling of the extrinsic information is possible with a common scaling factor, which is constant and independent of signal-to-noise ratio for additive white Gaussian noise (AWGN) channels. In [9] it is shown that in the case of a flat Rayleigh fading channel for a reconfigurable SOVA/log-MAP decoder a common scaling factor with value 0.7 is the optimal choice.

Nowadays, UMTS represents the dominant 3G system in the mobile communications market. According to UMTS specifications, a transport channel transfers data over radio interface from Medium Access Control sublayer of layer 2 to physical layer and is characterized by its transport format set, which consists of different transport formats. They must have the same type of channel coding and time transmission interval (TTI), while the transport block set or data frame size can vary. The transport block set determines the number of input bits to the channel encoder and can be transmitted every

TTI, with possible values for TTI of 10, 20, 40 and 80 msec [2,11]. After channel coding, outer block interleaving is performed, and since the frame duration in UMTS is 10 msec, the number of columns of the outer block interleaver can be 1, 2, 4 or 8, depending on TTI value. Therefore, the TTI values and the number of columns of the outer block interleaver are interrelated. Furthermore, every transport channel is assigned a radio access bearer with a particular data rate, which provides the transfer of the service through the radio network. A mobile terminal may use several parallel transport channels simultaneously, each having its own characteristics (transport format set).

UMTS radio interface transfers multiple applications. Parameters like bit error rate (BER) performance and delay are assigned to these applications. Four different service traffic classes are defined: conversational, streaming, interactive and background. For real-time conversational and streaming classes BER has to be less than 10^{-3}, while for non-real time interactive and background classes BER has to be less than 10^{-5}. The maximum acceptable delay for conversational class is 80 msec, for streaming it is 250 msec, for interactive it is 1 sec, while for background it is higher than 10 sec [2,11].

2. Simulation Parameters

The discrete representation of flat Rayleigh fading channel is given by the following equation:

$$y_k = \alpha_k \cdot x_k + n_k \quad (1)$$

where k is an integer symbol index, x_k is a binary phase shift keying (BPSK) symbol amplitude (± 1), n_k is a Gaussian random variable and y_k is a noisy received symbol. The fading amplitude a_k is a sample from a correlated Gaussian random process with zero mean and is generated using the Sum of Sines or Jakes-model, which is described in [12]. This model is based on summing 9 sinusoids whose frequencies are chosen as samples of the Doppler spectrum. The properties of Jakes model are further analysed in [13].

For the simulation model a carrier frequency $f_c = 2$ GHz is considered. It is also assumed that 1000000 bits are transmitted and grouped into frames whose length k_f must be ≥ 40 and ≤ 5114, according to UMTS specifications [2,14]. For a particular transport channel, every TTI the data with the characteristics specified in a transport format of the transport channel (k_f bits), is turbo encoded (constraint length $K = 4$ and rate $r_c = 1/3$) at the transmitter. Furthermore, each time instant it is assumed that the two recursive systematic convolutional encoders of the turbo encoder start encoding from all-zero state. After turbo coding and block interleaving using the UMTS parameters, the bits are BPSK modulated and transmitted through the mobile channel. At the receiver, outer block deinterleaving and turbo decoding is performed. The received values are not quantized which means that floating point arithmetic is used. The receiver is also assumed to have exact estimates of the fading amplitudes (perfect channel estimation without side information), while eight iterations are used in the turbo decoder.

Table 1 illustrates eight different UMTS dedicated transport channels with different transport format sets, which represent different implementation scenarios of the reconfigurable turbo decoder. The transport format set for each transport channel consists of different example transport formats and also of dynamic and semi-static parts. The semi-static part (turbo encoder parameters, TTI) is the same for all transport formats of the transport format set, while the dynamic part (frame size) differs [2, 11,15]. Moreover, as published simulation results have shown in [3,4] for flat Rayleigh fading channels, data rate, outer block interleaving (thus TTI) and signal-to-noise ratio (SNR) greatly affect BER performance: for each scenario of Table 1 these three parameters differ considering also the examples presented in [15].

Table 1. Implementation scenarios.

Transport channel type	Transport format set				Data rate R_b (kbps)	SNR (dB)	Scenario
	Dynamic part	Semi-static part					
	Transport block set or frame sizes (bits)	Turbo encoder parameters		TTI (msec)			
		K	Code rate				
Dedicated channel	576, 1152	4	1/3	40	28.8	32	1
	576, 1152, 1728, 2304	4	1/3	40	57.6	30	2
	336, 672, 1008, 1344	4	1/3	20	64	30	3
	336, 672, 1344, 2688	4	1/3	20	128	30	4
	336, 672, 1344, 2688, 3024	4	1/3	20	144	28	5
	168, 336, 672, 1344, 2016, 2688, 3360, 4032	4	1/3	20	384	28	6
	2560	4	1/3	40	64	30	7
	336, 1344, 2688, 4032, 4704	4	1/3	40	2000	40	8

Table 2. Quality of service and proposed decoding algorithm for scenarios 1, 2 and 3 of Table 1.

		Frame size (bits)	t_d using SOVA (msec)	t_d using log-MAP (msec)	Max latency (msec)	Log-MAP BER	SOVA BER	BER range	Proposed decoding algorithm
Scenario 1	**Conv. class**	576	240	528	80	0.000472	0.000523	$<10^{-3}$	Cannot be applied
		1152	400	976	80	0	0	$<10^{-3}$	
	Streaming class	576	240	528	250	0.000472	0.000523	$<10^{-3}$	SOVA
		1152	400	976	250	0	0	$<10^{-3}$	Cannot be applied
	Non-real time classes	576	240	528	Up to 1 sec interactive, >10 sec background	0.000472	0.000523	$<10^{-5}$	Cannot be applied
		1152	400	976		0	0	$<10^{-5}$	Log-MAP or SOVA
Scenario 2	**Conv. class**	576	160	304	80	0.001836	0.002096	$<10^{-3}$	Cannot be applied
		1152	240	528		0.000988	0.001036	$<10^{-3}$	
		1728	320	752		0.000582	0.000634	$<10^{-3}$	
		2304	400	976		0	0	$<10^{-3}$	
	Streaming class	576	160	304	250	0.001836	0.002096	$<10^{-3}$	Cannot be applied
		1152	240	528		0.000988	0.001036	$<10^{-3}$	
		1728	320	752		0.000582	0.000634	$<10^{-3}$	
		2304	400	976		0	0	$<10^{-3}$	
	Non-real time classes	576	160	304	Up to 1 sec interactive, >10 sec background	0.001836	0.002096	$<10^{-5}$	Cannot be applied
		1152	240	528		0.000988	0.001036	$<10^{-5}$	
		1728	320	752		0.000582	0.000634	$<10^{-5}$	
		2304	400	976		0	0	$<10^{-5}$	Log-MAP or SOVA
Scenario 3	**Conv. class**	336	82	157.6	80	0.003485	0.003888	$<10^{-3}$	Cannot be applied
		672	124	275.2		0.00146	0.00183	$<10^{-3}$	
		1008	166	392.8		0.000779	0.000984	$<10^{-3}$	
		1344	208	510.4		0.000519	0.000538	$<10^{-3}$	
	Streaming class	336	82	157.6	250	0.003485	0.003888	$<10^{-3}$	Cannot be applied
		672	124	275.2		0.00146	0.00183	$<10^{-3}$	
		1008	166	392.8		0.000779	0.000984	$<10^{-3}$	SOVA
		1344	208	510.4		0.000519	0.000538	$<10^{-3}$	SOVA
	Non-real time classes	336	82	157.6	Up to 1 sec interactive, >10 sec background	0.003485	0.003888	$<10^{-5}$	Cannot be applied
		672	124	275.2		0.00146	0.00183	$<10^{-5}$	
		1008	166	392.8		0.000779	0.000984	$<10^{-5}$	
		1344	208	510.4		0.000519	0.000538	$<10^{-5}$	

Table 3. Quality of service and proposed decoding algorithm for scenarios 4 and 5 of Table 1.

		Frame size (bits)	t_d using SOVA (msec)	t_d using log-MAP (msec)	Max latency (msec)	Log-MAP BER	SOVA BER	BER range	Proposed decoding algorithm
Scenario 4	**Conv. class**	336	61	98.8	80	0.003465	0.004047	$<10^{-3}$	Cannot be applied
		672	82	157.6		0.001584	0.001713	$<10^{-3}$	
		1344	124	275.2		0.000796	0.000934	$<10^{-3}$	
		2688	208	510.4		0	0	$<10^{-3}$	
	Streaming class	336	61	98.8	250	0.003465	0.004047	$<10^{-3}$	Cannot be applied
		672	82	157.6		0.001584	0.001713	$<10^{-3}$	
		1344	124	275.2		0.000796	0.000934	$<10^{-3}$	SOVA
		2688	208	510.4		0	0	$<10^{-3}$	SOVA
	Non-real time classes	336	61	98.8	Up to 1 sec interactive, >10 sec background	0.003465	0.004047	$<10^{-5}$	Cannot be applied
		672	82	157.6		0.001584	0.001713	$<10^{-5}$	
		1344	124	275.2		0.000796	0.000934	$<10^{-5}$	

		2688	208	510.4		0	0	$<10^{-5}$	Log-MAP or SOVA
Scenario 5	**Conv. class**	336	58.6	92.26	80	0.005771	0.006268	$<10^{-3}$	Cannot be applied
		672	77.3	144.5	80	0.003005	0.003287	$<10^{-3}$	
		1344	114.6	249.06	80	0.000704	0.0010007	$<10^{-3}$	
		2688	189.3	458.13	80	0	3.091e-05	$<10^{-3}$	
		3024	208	510.4	80	0	0	$<10^{-3}$	
	Streaming class	336	58.6	92.26	250	0.005771	0.006268	$<10^{-3}$	Cannot be applied
		672	77.3	144.5	250	0.003005	0.003287	$<10^{-3}$	
		1344	114.6	249.06	250	0.000704	0.0010007	$<10^{-3}$	Log-MAP
		2688	189.3	458.13	250	0	3.091e-05	$<10^{-3}$	SOVA
		3024	208	510.4	250	0	0	$<10^{-3}$	SOVA
	Non-real time classes	336	58.6	92.26	Up to 1 sec interactive, >10 sec background	0.005771	0.006268	$<10^{-5}$	Cannot be applied
		672	77.3	144.5		0.003005	0.003287	$<10^{-5}$	
		1344	114.6	249.06		0.000704	0.0010007	$<10^{-5}$	
		2688	189.3	458.13		0	3.091e-05	$<10^{-5}$	Log-MAP
		3024	208	510.4		0	0	$<10^{-5}$	Log-MAP or SOVA

Table 4. Quality of service and proposed decoding algorithm for scenarios 6 and 7 of Table 1.

		Frame size (bits)	**t_d using SOVA (msec)**	**t_d using log-MAP (msec)**	**Max latency (msec)**	**Log-MAP BER**	**SOVA BER**	**BER range**	**Proposed decoding algorithm**
Scenario 6	**Conv. class**	168	43.5	49.8	80	0.0073	0.008	$<10^{-3}$	Cannot be applied
		336	47	59.6	80	0.0065	0.007	$<10^{-3}$	
		672	54	79.2	80	0.0051	0.0062	$<10^{-3}$	
		1344	68	118.4	80	0.0022	0.0028	$<10^{-3}$	
		2016	82	157.6	80	0.000654	0.001307	$<10^{-3}$	
		2688	96	196.8	80	0	0	$<10^{-3}$	
		3360	110	236	80	0	0	$<10^{-3}$	
		4032	124	275.2	80	0	0	$<10^{-3}$	
	Streaming class	168	43.5	49.8	250	0.0073	0.008	$<10^{-3}$	Cannot be applied
		336	47	59.6	250	0.0065	0.007	$<10^{-3}$	
		672	54	79.2	250	0.0051	0.0062	$<10^{-3}$	
		1344	68	118.4	250	0.0022	0.0028	$<10^{-3}$	
		2016	82	157.6	250	0.000654	0.001307	$<10^{-3}$	Log-MAP
		2688	96	196.8	250	0	0	$<10^{-3}$	Log-MAP or SOVA
		3360	110	236	250	0	0	$<10^{-3}$	Log-MAP or SOVA
		4032	124	275.2	250	0	0	$<10^{-3}$	SOVA
	Non-real time classes	168	43.5	49.8	Up to 1 sec interactive, >10 sec back-ground	0.0073	0.008	$<10^{-5}$	Cannot be applied
		336	47	59.6		0.0065	0.007	$<10^{-5}$	
		672	54	79.2		0.0051	0.0062	$<10^{-5}$	
		1344	68	118.4		0.0022	0.0028	$<10^{-5}$	
		2016	82	157.6		0.000654	0.001307	$<10^{-5}$	
		2688	96	196.8		0	0	$<10^{-5}$	Log-MAP or SOVA
		3360	110	236		0	0	$<10^{-5}$	Log-MAP or SOVA
		4032	124	275.2		0	0	$<10^{-5}$	Log-MAP or SOVA
Scenario 7	**Conv. class**	2560	400	976	80	0	0	$<10^{-3}$	Cannot be applied
	Streaming class	2560	400	976	250	0	0	$<10^{-3}$	Cannot be applied
	Non-real time classes	2560	400	976	Up to 1 sec interactive, >10 sec back-ground	0	0	$<10^{-5}$	Log-MAP or SOVA

Table 5. Quality of service and proposed decoding algorithm for scenario 8 of Table 1.

		Frame size (bits)	t_d using SOVA (msec)	t_d using log-MAP (msec)	Max latency (msec)	Log-MAP BER	SOVA BER	BER range	Proposed decoding algorithm
Scenario 8	**Conv. class**	336	81.34	83.76	80	0.001815	0.0019574	$<10^{-3}$	Cannot be applied
		1344	85.37	95.05	80	0.001414	0.0016548	$<10^{-3}$	
		2688	90.75	110.1	80	9.97e-07	9.97e-07	$<10^{-3}$	
		4032	96.12	125.1	80	0	0	$<10^{-3}$	
		4704	98.81	132.68	80	0	0	$<10^{-3}$	
	Streaming class	336	81.34	83.76	250	0.001815	0.0019574	$<10^{-3}$	Cannot be applied
		1344	85.37	95.05	250	0.001414	0.0016548	$<10^{-3}$	
		2688	90.75	110.1	250	9.97e-07	9.97e-07	$<10^{-3}$	Log-MAP or SOVA
		4032	96.12	125.1	250	0	0	$<10^{-3}$	Log-MAP or SOVA
		4704	98.81	132.68	250	0	0	$<10^{-3}$	Log-MAP or SOVA
	Non-real time classes	336	81.34	83.76	Up to 1 sec interactive, >10 sec background	0.001815	0.0019574	$<10^{-5}$	Cannot be applied
		1344	85.37	95.05		0.001414	0.0016548	$<10^{-5}$	
		2688	90.75	110.1		9.97e-07	9.97e-07	$<10^{-5}$	Log-MAP or SOVA
		4032	96.12	125.1		0	0	$<10^{-5}$	Log-MAP or SOVA
		4704	98.81	132.68		0	0	$<10^{-5}$	Log-MAP or SOVA

According to [2] and [11], three different operating environments have been specified for UMTS:

- Rural outdoor operating environment with maximum supported mobile terminal speed 500 km/h and maximum data rate of 144 kbps. Here, it has to be mentioned that a speed of 500 km/h corresponds to high speed vehicles (e.g. trains). More typical value for this environment is 300 km/h.
- Urban or suburban outdoor operating environment with maximum supported mobile speed 120 km/h and maximum data rate of 384 kbps.
- Indoor or low range outdoor operating environment with maximum supported mobile speed 10 km/h and maximum data rate of 2 Mbps.

In [9] the approach is similar, but we considered the first two operating environments: a terminal speed of 300 km/h for a rural outdoor environment and a terminal speed of 100 km/h for an urban/suburban outdoor environment. In this paper we focus on the last operating environment and we choose a low terminal speed of 4 km/h. This means that the maximum data rate of 2 Mbps can be considered. A terminal speed of 4 km/h is a typical common value and it is important to be explored: represents walking human speed. In other words, each implementation scenario of the reconfigurable decoder of Table 1 is applied to indoor or low range outdoor operating environment. Moreover, similarly to [8,9,10], for the calculation of total maximum delay per frame for SOVA and log-MAP we use the following equations assuming a pipeline turbo decoder architecture and a processor that runs at the same rate for both SOVA and log-MAP:

Total max delay using SOVA:

$$t_d = 2 \times TTI + \left(\frac{k_f}{R_b} \times N \right) \quad (2)$$

Total max delay using log-MAP:

$$t_d = 2 \times TTI + \left(\frac{k_f}{R_b} \times N \times 2.8 \right) \quad (3)$$

where t_d is the total delay, k_f is the frame size, R_b is the data rate of the radio bearer assigned to the transport channel and N is the number of turbo decoder iterations. In these equations the higher complexity of log-MAP compared to SOVA (2.8 times) is also considered.

3. Simulation Results

The suitable decoding algorithm for each scenario is chosen according to performance and delay. Therefore, for each scenario of Table 1 all four service classes are applied to determine the quality of service profile parameters for different applications. Delay is calculated for each algorithm using Equations (2) and (3), while the simulated BER for each scenario is given in the following subsections together with a brief analysis of the results. Particularly, Table 2 shows quality of service for the different frame lengths of scenarios 1, 2, 3, while Tables 3 and 4 present quality of service for scenarios 4, 5 and 6, 7, respectively. Finally, Table 5 presents quality of service for scenario 8.

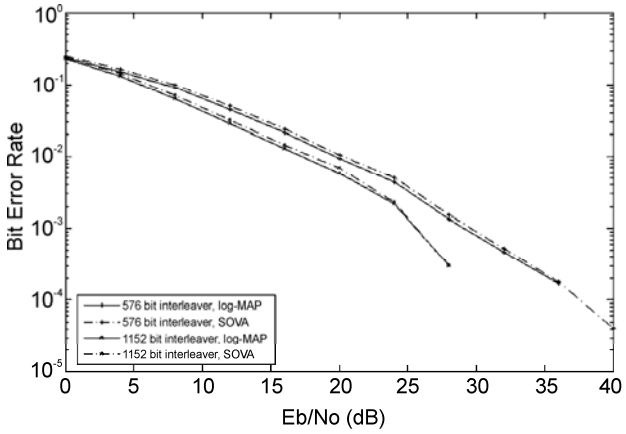


Figure 1. BER vs Eb/No for scenario 1.

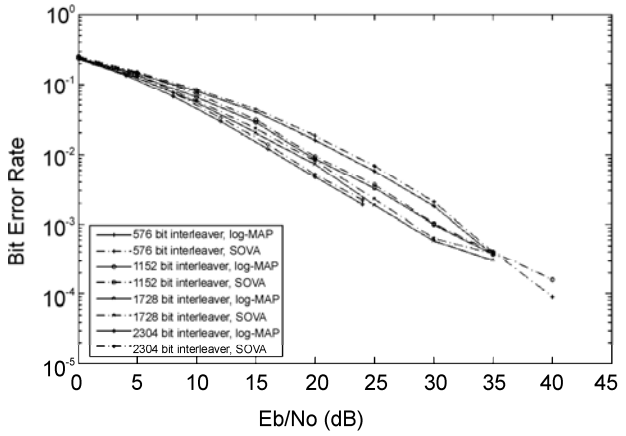


Figure 2. BER vs Eb/No for scenario 2.

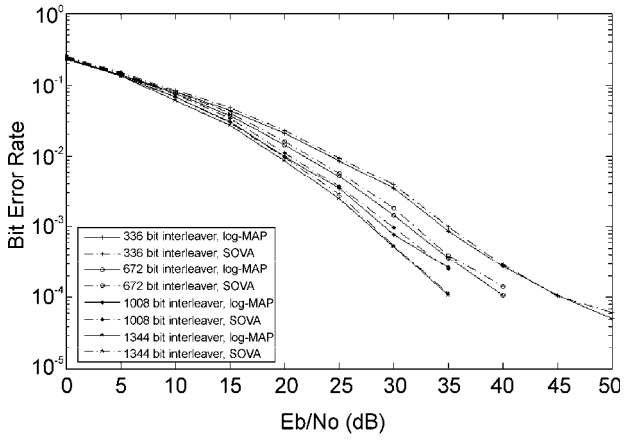


Figure 3. BER vs Eb/No for scenario 3.

3.1. Scenario 1

The simulated BER for this scenario is shown in Figure 1 assuming a symbol rate R_s of 86.4 Kbaud, normalised fade rate $f_d T_s = 0.000085$ with Doppler frequency $f_d = 7.407$ Hz. Two frame lengths are considered in this scenario: 576 and 1152 bits, as Table 2 illustrates.

3.1.1. Conversational Service Class

At a SNR of 32 dB, the conversational class cannot be considered for this scenario because even though the BER criterion is satisfied, latency is too high for all frame lengths for either SOVA or log-MAP.

3.1.2. Streaming Service Class

For this class only a frame length of 576 bits can be applied. In this case SOVA satisfies both requirements, while log-MAP exceeds the maximum acceptable delay limit. For a frame of 1152 bits delay for SOVA and log-MAP is too high to achieve the limit for this class.

3.1.3. Interactive/Background Service Classes

For a frame length of 576 bits neither algorithm can be used because of the low BER criterion, while both requirements are achieved from both algorithms for a frame length of 1152 bits. Thus, a 576 bit frame service can not be applied, whereas in an 1152 bit frame service either SOVA or log-MAP can be used.

3.2. Scenario 2

The simulated BER results for this scenario are shown in Figure 2 assuming a symbol rate R_s of 172.8 Kbaud, normalised fade rate $f_d T_s = 0.000042$ and a SNR of 30 dB.

3.2.1. Conversational Service Class

According to Table 2, for this class the four different frame lengths cannot be applied because of the tight delay limit (80 msec).

3.2.2. Streaming Service Class

Similarly, as illustrated in Table 2, the four frame lengths are not applicable. Particularly, for frame lengths of 576 and 1152 bits SOVA satisfies the delay criterion, but does not satisfy BER criterion. On the other hand, the use of log-MAP gives unacceptable delay. For frame lengths of 1728 and 2304 bits although BER is satisfied from both algorithms, maximum acceptable delay is exceeded.

3.2.3. Interactive/Background Service Classes

For these service classes it is well-known that BER must be low and latency limits are not very strict. Thus, the first three frame lengths cannot be applied due to not acceptable BER. For a frame length of 2304 bits the two criteria are achieved by both decoding algorithms: either SOVA or log-MAP can be used.

3.3. Scenario 3

Figure 3 presents the simulation results for this scenario using the following parameters: $R_s = 192$ Kbaud, $f_d T_s = 0.000038$ and a SNR of 30 dB.

3.3.1. Conversational Service Class

According to the analysis of Table 2, the four frame lengths give too high delay. Thus, their application is not possible for SOVA or log-MAP.

3.3.2. Streaming Service Class

The analysis of Table 2 clearly shows that for all frame lengths SOVA satisfies the delay limit of 250 msec at 30 dB. On the other hand the BER limit is not achieved for the small frames of 336 and 672 bits. Thus, SOVA can be used for frames of 1008 and 1344 bits. For log-MAP and frames of 672, 1008, 1344 bits the delay limit cannot be achieved. For a small frame of 336 bits the delay limit is achieved, but the BER limit is not achieved.

3.3.3. Interactive/Background Service Classes

For these non-real time service classes and for all four frames the achieved BER is lower than the acceptable limit. Therefore, although the delay limit is achieved the four frames can not be applied.

3.4. Scenario 4

Figure 4 presents the simulated BER for this scenario using the following parameters: $R_s = 384$ Kbaud, $f_d T_s = 0.000019$ with $f_d = 185.1$ Hz and a SNR of 30 dB.

3.4.1. Conversational Service Class

Again, for this class the four frames cannot be applied because of high delay. For a frame of 336 bits although delay is acceptable for SOVA, BER criterion is not satisfied. According to Table 3 it is obvious that this service scenario is not possible to be implemented.

3.4.2. Streaming Service Class

The analysis of Table 3 clearly shows that for all frame lengths SOVA satisfies the delay limit of 250 msec at 30 dB. On the other hand the BER limit is not achieved for the small frames of 336 and 672 bits. Thus, SOVA is the proposed turbo decoding algorithm for frames of 1344 and 2688 bits. For log-MAP and frames of 1344, 2688 bits the delay limit cannot be achieved. For small frames of 336 and 672 bits the delay limit is achieved, but the BER limit is not achieved.

3.4.3. Interactive/Background Service Classes

According to Table 3, for these classes and for the first three frames the achieved BER is lower than the acceptable limit. Therefore, although the delay limit is achieved these frames can not be applied. On the other hand, for a frame of 2688 bits the two parameters (BER, delay) are satisfied by both algorithms.

3.5. Scenario 5

For Figure 5 the following parameters are assumed: $R_s = 432$ Kbaud, $f_d T_s = 0.000017$ and a SNR of 28 dB. Figure 5 shows BER performance for the five different frame lengths specified in Table 1 for this scenario.

3.5.1. Conversational Service Class

For this class (Table 3) for all five frames the delay criterion is too low to be achieved from both algorithms. There is an exception for the small frames of 336 and 672 bits, where the delay criterion is achieved for SOVA but BER criterion is not. It is obvious that the constraints

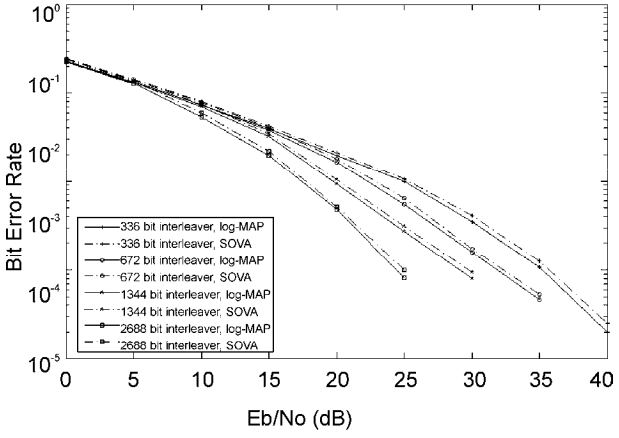


Figure 4. BER vs Eb/No for scenario 4.

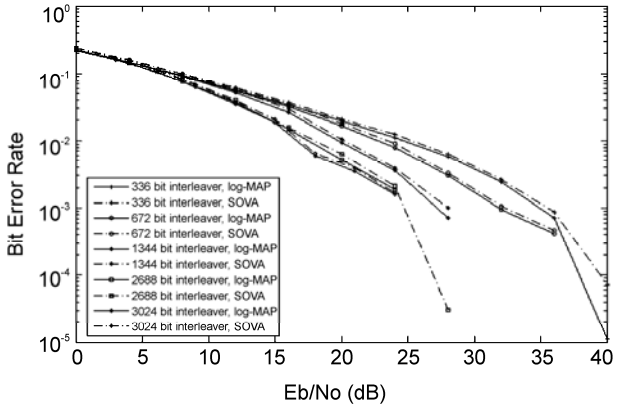


Figure 5. BER vs Eb/No for scenario 5.

of the two parameters cannot be achieved by both algorithms.

3.5.2. Streaming Service Class

The analysis of Table 3 identifies three cases:

- Small frames of 336 and 672 bits. Here, the delay limit is achieved, but the BER limit is not for SOVA and log-MAP. This means that these frames cannot be implemented.
- Medium frame of 1152 bits. Here, the delay limit is achieved by both algorithms. Log-MAP is the proposed choice because it can achieve the BER limit as well. SOVA cannot achieve the BER limit. Thus, log-MAP represents the proposed algorithm.
- Large frames of 2688 and 3024 bits. Here, SOVA is the algorithm that can be implemented. The reason is the following: BER limit is achievable by both algorithms, whereas delay limit is achieved only by SOVA.

3.5.3. Interactive/Background Service Classes

According to Table 3, the delay limit is achieved by both SOVA and log-MAP for all frames. Furthermore, for the first three frames the BER limit is not achieved, but for 2688 bits frame it is achieved only by log-MAP. In this case log-MAP is proposed. For a frame of 3024 bits the limits of the two parameters are achieved by both algorithms.

3.6. Scenario 6

Figure 6 illustrates the simulated BER of the different frame lengths for this scenario using the following parameters: $R_s = 1152$ Kbaud, $f_d T_s = 0.0000064$ and a SNR of 28 dB.

3.6.1. Conversational Service Class

For this class (Table 4) for the first three frames although delay criterion is satisfied, BER criterion is not satisfied. For the next four frames either BER, or delay limits are not achieved for SOVA and log-MAP. Thus, this service class is not possible to be implemented for all frames.

3.6.2. Streaming Service Class

The analysis of Table 4 identifies four cases:

- Frames of 168, 336, 672 and 1344 bits. Here, the delay limit is achieved, but the BER limit is not achieved for SOVA and log-MAP. This means that these frames cannot be implemented.
- Frame of 2016 bits. Here, the delay limit is achieved by both algorithms. Log-MAP is the proposed choice because it can achieve the BER limit as well, while SOVA cannot achieve the BER limit. Thus, log-MAP represents the proposed algorithm.
- Frames of 2688 and 3360 bits. Here, SOVA and log-MAP achieve both limits. Therefore, both algorithms can be used.
- Frame of 4032 bits. Here, SOVA is the algorithm that can be implemented. The reason is the following: BER limit is achievable by both algorithms, whereas log-MAP gives unacceptable delay.

3.6.3. Interactive/Background Service Classes

According to Table 4, the delay limit is achieved by both SOVA and log-MAP for all frames. Furthermore, for the first five frames the BER limit is not achieved. Thus, they cannot be implemented. For frames of 2688, 3360 and 4032 bits the limits of the two parameters are achieved by both algorithms.

3.7. Scenario 7

In Figure 7 BER performance for the different frame lengths for this scenario can be seen using the following parameters: $R_s = 192$ Kbaud, $f_d T_s = 0.000038$ and a SNR of 30 dB.

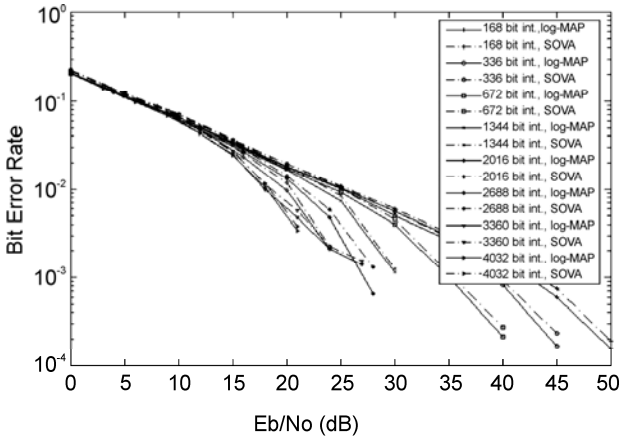


Figure 6. BER vs Eb/No for scenario 6.

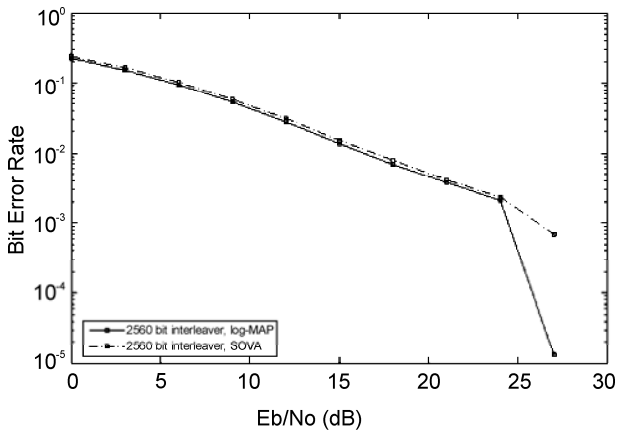


Figure 7. BER vs Eb/No for scenario 7.

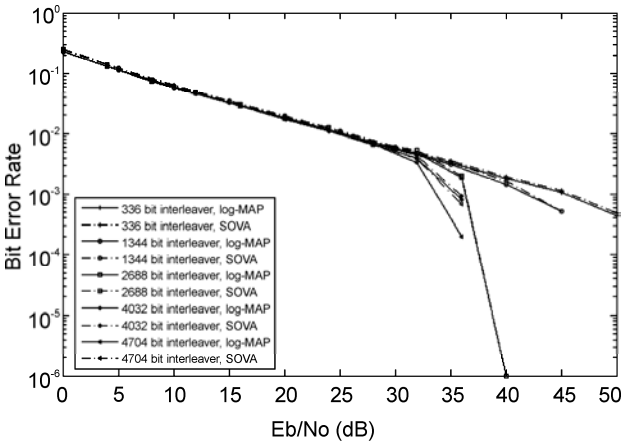


Figure 8. BER vs Eb/No for scenario 8.

3.7.1. Conversational/Streaming Service Classes

The analysis of Table 4 clearly shows that the frame of 2560 bits gives unacceptable delay for both real time classes and both decoding algorithms. Therefore, they cannot be implemented.

3.7.2. Interactive/Background Service Classes

For non-real time classes both limits are achieved by both algorithms, which mean that they are both suitable for this application.

3.8. Scenario 8

Figure 8 presents BER performance of the different frame lengths for this scenario using the following parameters: $R_s = 6000$ Kbaud, $f_d T_s = 0.0000012$ and a SNR of 40 dB.

3.8.1. Conversational Service Class

For all five frames the calculated delay, according to Table 5, is too high. Thus, this scenario cannot be implemented for this service class.

3.8.2. Streaming Service Class

Here, delay criterion is achieved by both algorithms and for all frames. Furthermore, for frames of 336 and 1344 bits the BER limit is not achievable by the two algorithms. This means that these two frames cannot be implemented. On the other hand, for frames of 2688, 4032 and 4704 bits the two criteria are satisfied by both algorithms: they are equally suitable.

3.8.3. Interactive/Background Service Classes

From Table 5 it can be seen that the analysis is similar to the previous section: the first two frames cannot be established, whereas for the last three frames either SOVA or log-MAP can be used.

4. Conclusions

In this paper we have presented possible reconfiguration scenarios applied to an important receiver technique, namely, channel decoding. It has been shown that reconfigurability is a desirable feature towards the implementation of energy efficient receivers without performance sacrifices.

For a UMTS turbo decoder SOVA and log-MAP correspond to the main decoding algorithms. Considering performance and complexity or delay, SOVA is the best choice in terms of complexity, while log-MAP is the best choice in terms of performance. The similarities in the data-flow of the two algorithms support the idea of a reconfigurable SOVA/log-MAP turbo decoder [8,9,10]. Moreover, according to [3] at low terminal speeds BER is worse than at higher terminal speeds. For UMTS some applications require the lowest possible delay, while for others the lowest possible performance is sufficient. Having in mind the results of [9] it is observed that at rural and urban/suburban outdoor operating environments more frames can be established compared to indoor/low range outdoor environment. Thus, for indoor/low range outdoor environment there are many applications which cannot be established.

Our results for indoor/low range outdoor environment show that for all implementation scenarios real time conversational class cannot be established. The reason is the low terminal speed which gives high BER. Comparing with urban/suburban environment in [9], this class can be applied to medium sized frames and high data rates, whereas in rural outdoor operating environment this class can be applied to small frames and low or medium data rates.

For real time streaming class the proposed algorithm choice depends on data rate. For low data rates all frames cannot be applied, except for small frames where SOVA is optimal. For medium data rates (64 kbps, 128 kbps) small frames cannot be applied, while for medium-sized frames SOVA is proposed. For 144 kbps again SOVA is proposed for larger frames, while for medium-sized frames log-MAP is optimal. For high data rates (384 kbps) small frames cannot be considered, for medium frames log-MAP is proposed, while for large frames SOVA is proposed. For the other frame lengths either algorithm is proposed. For very high data rates (2 Mbps) small frames cannot be established: for the other frames either SOVA or log-MAP can be used. On the other hand, in [9] for streaming class applications urban/suburban and rural outdoor operating environments SOVA is optimal for the scenarios that can be established. It is remarkable that, similarly to [9], as data rate increases more and larger frames can be applied.

For non-real time applications performance is the priority and delay requirements are looser. We observe that

for all scenarios small frames cannot be applied due to tight BER. For larger frames both algorithms are equally suitable. Furthermore, for medium data rates and medium frames log-MAP is the proposed algorithm choice. For urban/suburban outdoor environment the conclusions are similar in [9], whereas for rural outdoor environment log-MAP is optimum for the small frames and the two algorithms are equally suitable for larger frames.

5. References

[1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo codes," IEEE Trans. on Communications, Vol. 44, No. 10, pp. 1261–1271, 1996.

[2] H. Holma and A. Toskala, "WCDMA for UMTS: Radio access for third generation mobile communications," J. Wiley, 2000.

[3] J. Woodard and L. Hanzo, "Comparative study of turbo decoding techniques: An overview," IEEE Transactions on Vehicular Technology, Vol. 49, No. 6, pp. 2208–2233, 2000.

[4] E. Hall and S. Wilson, "Design and analysis of turbo codes on Rayleigh fading channels," IEEE Journal on Selected Areas in Communications, Vol. 16, No. 2, pp. 160–174, 1998.

[5] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," IEEE ICC'95, Seattle, USA, pp. 1009–1013, 1995.

[6] J. Hagenauer and P. Hoher, "A Viterbi algorithm with soft outputs and its applications," IEEE GLOBECOM'89, Dallas, USA, pp. 1680–1686, 1989.

[7] S. Pietrobon, "Implementation and performance of a turbo/MAP decoder," International Journal of Satellite Communications, Vol. 16, No. 1, pp. 23–46, 1998.

[8] C. Chaikalis, "Reconfigurable structures for turbo codes in 3G mobile radio transceivers," PhD thesis, School of Engineering Design and Technology, University of Bradford, UK, 2003.

[9] C. Chaikalis, "Implementation of a reconfigurable turbo decoder in 3GPP for flat Rayleigh fading," Elsevier Digital Signal Processing Journal, 2008.

[10] C. Chaikalis and J. M. Noras, "Reconfigurable turbo decoding for 3G applications," Elsevier Signal Processing Journal, Vol. 84, No. 10, pp. 1957–1972, 2004.

[11] 3GPP TS 25.201 V3.3.0. "Physical layer-General description," Release 1999, 2002.

[12] W. C. Jakes, "Microwave mobile communications," J. Wiley & Sons, New York, 1974.

[13] M. Patzold, U. Killat, F. Laue, and Y. Li, "On the statistical properties of deterministic simulation models for mobile fading channels," IEEE Transactions on Vehicular Technology, Vol. 47, No. 1, pp. 254–269, 1998.

[14] 3GPP TS 25.212 V3.9.0. "Multiplexing and channel coding (FDD)," Release 1999, 2002.

[15] 3GPP TR 25.944 V3.5.0. "Channel coding and multiplexing examples," Release 1999, 2001.

Int. J. Communications, Network and System Sciences, **2009, 2, 714-719**
doi:10.4236/ijcns.2009.28082 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Application of a Dynamic Identity Authentication Model Based on an Improved Keystroke Rhythm Algorithm

Wenchuan YANG, Fang FANG
School of information and communication Engineering, Beijing University of Posts and Telecommunication, Beijing, China
*Email: yangwenchuan@bupt.edu.cn, fang.*16898*@gmail.com*
Received July 15, 2009; *revised August* 29, 2009; *accepted October* 2, 2009

Abstract

Keystroke rhythm identification, which extracts biometric characteristics through keyboards without additional expensive devices, is a kind of biometric identification technology. The paper proposes a dynamic identity authentication model based on the improved keystroke rhythm algorithm in Rick Joyce model and implement this model in a mobile phone system. The experimental results show that comparing with the original model, the false alarm rate (FAR) of the improved model decreases a lot in the mobile phone system, and its growth of imposter pass rate (IPR) is slower than the Rick Joyce model's. The improved model is more suitable for small memory systems, and it has better performance in security and dynamic adaptation. This improved model has good application value.

Keywords: Biometric Identification Technology, Keystroke Rhythm, Identity Authentication, Keystroke Latency Time, IPR, FAR

1. Introduction

Along with the technical development of computers and networks, unauthorized users or hackers do more and more invasions on the information system through the network. Therefore, protection of computer security has become a matter of urgency. User's identity authentication is an important means to carry out system security. The traditional login method only depends on a single password content, which has the drawback of password leak and security problems, so biometric identification technology is proposed to enhance system security by taking use of human biological characteristics.

Biological characteristics, the only difference from other people, can automatically identify and verify the physical characteristics or behavior patterns. The biometric identification technologies currently contain hand recognition, fingerprint identification, facial recognition, voice recognition, iris recognition and signature recognition, etc [1]. Although these methods can accomplish the identification, the processes of feature extractions need expensive hardware devices which limit the application range to a large extent.

Keystroke rhythm identification is a kind of biometric identification, and it has the advantages of biometric identification, in addition, the greatest advantage in terms of keystroke rhythm is low cost. It costs almost no hardware and can be applied in majority of systems which only require keyboards to accomplish the identification. Moreover, once it is applied, it can play a significant role on improving the security of keyboard input, so keystroke rhythm has a wide range of application.

20th century 80's, biometric identification using personal keystroke features was first proposed. This method can effectively prevent the illegal invasions. Keystroke rhythm identification extracts characteristics through keyboards without additional devices. So many scholars have paid much attention on this issue and have made some methods of pattern recognition applied to the identity authentication based on keystroke characteristics [2]. Rick Joyce and Gopal Gupta have also done some specific research on keystroke characteristics. In this paper, we will further improve the Rick Joyce algorithm to design a dynamic identification model based on keystroke rhythm which is suitable for small memory systems, then we realize it in a mobile phone system and do analysis of experimental data.

2. Keystroke Rhythm Algorithm of Rick Joyce Model and an Improved Model

Some studies have shown that users' keystroke rhythm characteristics like fingerprints can reflect persons' biological characteristics by the keystroke duration time and keystroke latency time [3,4]. For example, the definition of the n-th keystroke duration time means the time between the n-th button is pressed and released, the n-th keystroke latency time stands for the time between the n-th button is pressed and the (n+1)-th button is pressed. The following are two models of identity authentication based on keystroke latency time.

2.1. Rick Joyce Identity Authentication Model [5] and its Algorithm

Using keystroke rhythm as a means to study identity authentication at first is Rick Joyce and GopaIGupta. They design an identity authentication model based on keystroke rhythm – the keystroke latency time. In this system, a user is asked to type eight times of his username, password, firstname, lastname. Then the system processes the information extracted to establish a four-dimensional vector:

$$M = \left(M_{username}, M_{password}, M_{firstname}, M_{lastname}\right)$$

Each component of this four-dimensional vector is the average on eight vectors, and they use M as this user's behavioral characteristic profile. In the process of identity authentication, it is required to check the information typed by tested user (assuming he/she has been aware of the content typed), thus the input information is processed to form a test vector :

$$T = \left(T_{username}, T_{password}, T_{firstname}, T_{lastname}\right)$$

and the system just verify user's identity by comparing M with T.

The user's keystroke rhythm can be expressed as a n-dimensional vector, in which n is the total number of input intervals, and the last number of the vector means the latency time between the last button is pressed and the enter button is pressed. Then M and T can be recorded as $M = (M_1, M_2 \cdots M_n)$, $T = \left(T_1, T_2 \cdots T_n\right)$. And the difference between M and T is expressed as $\left\|M - T\right\|_1$, namely $\sum_{i=1}^{i=n} \left|m_i - t_i\right|$, which is a norm for comparison.

Then let the first user enter eight times of information again to form eight training sequences $S_1, S_2 \cdots S_8$ as the reference vectors, and they can be used to calculate the value of $\left\|M - S_i\right\|_1, i = 1, 2 \cdots 8$, then the mean and standard deviation can be gained. Consider the false alarm rate (FAR) and imposter pass rate (IPR), the verification threshold for this system is the mean plus one-and-one-half standard deviation, namely:

$$\Phi = \frac{1}{8}\sum_{i-1}^{i-8}\left\|M - S_i\right\| + 1.5\sigma$$

If the tested user meets the inequality $\left\|M - T\right\|_1 > \Phi$, he/she is considered as an imposter. If the tested user meets the inequality $\left\|M - T\right\|_1 < \Phi$, he/she is verified as an authentic user stored in system.

2.2. A Dynamic Model Based on the Improved Algorithm

The shortcomings of the algorithm above lie in: first of all, once the characteristic profile is formed, it is difficult to modify. Secondly, the Rick Joyce algorithm ignores the consideration of signature curve shape [5], and it just compares M with T by $\left\|M - T\right\|_1 = \sum_{i-1}^{i-n}\left|m_i - t_i\right|$, in which they do not refer to the detailed information of signature curve shape. Therefore, in terms of the shortcomings above, we make a corresponding improvement.

Because the calculation amount of the algorithm above is large, we propose a dynamic model based on the improved algorithm with a small amount of calculation.

The user is required to type his password in our system, and then the system processes the information typed to establish a one-dimensional vector:

$$M = (M_{password})$$

$M_{password}$ is the mean vector of eight times of keystroke latency time typed. Because of the possibility of long waiting time, the time between the last button is pressed and enter button is pressed can be ignored in this algorithm.

In general, both the keystroke rhythm of user and the tested user can be expressed as n-dimensional vectors, in which n is the total number of input intervals, namely:

$$M = \left(M_1, M_2 \cdots M_n\right), \quad T = \left(T_1, T_2 \cdots T_n\right)$$

The difference between M and T is expressed as $\left\|M - T\right\|_2$, namely:

$$\vec{\Delta} = \left\|M - T\right\|_2 = \left(\left|m_1 - t_1\right|, \left|m_2 - t_2\right| \cdots \left|m_n - t_n\right|\right)$$

Again, the first user enter eight times to form training sequences $S_1, S_2 \cdots S_8$ as the reference vectors, then we get eight difference vectors:

$$\overrightarrow{\Delta_{s_i}} = \left\| M - S_i \right\|_2, i = 1,2 \cdots 8$$

And use these eight difference vectors to calculate the mean difference vector $\overrightarrow{\Delta_s}$ and the standard deviation vector $\vec{\sigma}$.

$$\overrightarrow{\Delta_s} = \frac{1}{8}(\sum_{i=1}^{i=8} \overrightarrow{\Delta_{s_i}}), i = 1,2 \cdots 8 \,, \vec{\sigma} = \frac{1}{8}\sum_{i=1}^{i=8} (\overrightarrow{\Delta_{s_i}} - \overrightarrow{\Delta_s})^2, i = 1,2 \cdots 8$$

Then a threshold vector is pre-established in our model:

$$\vec{\Phi} = \overrightarrow{\Delta_s} + 3\vec{\sigma}$$

Here, we compare the difference vector $\left\| M - T \right\|_2$ with the threshold vector $\vec{\Phi}$. If $\vec{\Delta} < \vec{\Phi}$, namely:

$$(\Delta_1, \Delta_2 \cdots \Delta_n) < (\Phi_1, \Phi_2 \cdots \Phi_n)$$

It means that each component of difference vector Δ_i is less than the corresponding component of threshold vector Φ_i. At this time, the identity of the tested user is authentic. If $\vec{\Delta} > \vec{\Phi}$, the tested user is verified as an imposter. Consider the non-regularity of user's operations, the matching percentage of corresponding components comparison can be adjusted a little. However, in this paper we adopt the matching percentage of comparison is 100% when the tested user will be considered as authentic user.

Besides, the system will merge the test vector into the user keystroke rhythm vector M to modify the user characteristic profile when the tested user is considered as authentic user [6]. For example, when the (k+1)-th result is legal, we can integrate this test vector with former user characteristic profile in system:

$$M_{k+1} = \frac{k \cdot M_k + T}{k+1}$$

M_k is the k-th characteristic profile of user. In this way the matching model for keystroke rhythm is variable along with the input of legal user. Accordingly, it is able to modify the keystroke rhythm and benefit the performance in authentication .

Suppose that an imposter enters a password and the certain parts of his/her keystroke latency time are longer (or shorter) but the others are normal, because Rick Joyce model considers the latency time as a whole, it may be verified as authentic user. It is required to enhance the security for some privacy systems [7]. The improved model in this paper take shape of the signature curve into account, and it contains each interval's threshold value, so the security performance is stronger.

3. System Implement

The improved dynamic identity authentication model is suitable for systems with small calculation amount, so in our study, a mobile phone system will be applied.

As the limited space and computing capability in the mobile phone system, it needs an algorithm of small calculation amount and the less storage space. For one thing, the improved algorithm does not require too much calculation, the calculation amount about comparison and modification of rhythm characteristics are small. For another thing, mobile phones are used by individuals, and it should not be required to identify more individuals and store more keystroke rhythm profiles for them, so the storage space requirements will be reduced. In addition, the amount of buttons on the mobile phone keyboard is small and the users' keystroke actions are relatively concentrated, so that it will help to form steady keystroke characteristics to enhance the security performance.

We have established a model of identity authentication using the characteristic profile of keystroke rhythm in a mobile phone system [8]. First of all, the tested users type their passwords, and then the system extracts information to form their keystrokes rhythm characteristics which are compared with the existed user characteristic profile in the mobile phone system. Only when users type the same content and their keystrokes rhythm characteristics are within the threshold value of model, they can successfully login. The system process is as follows (Figure 1):

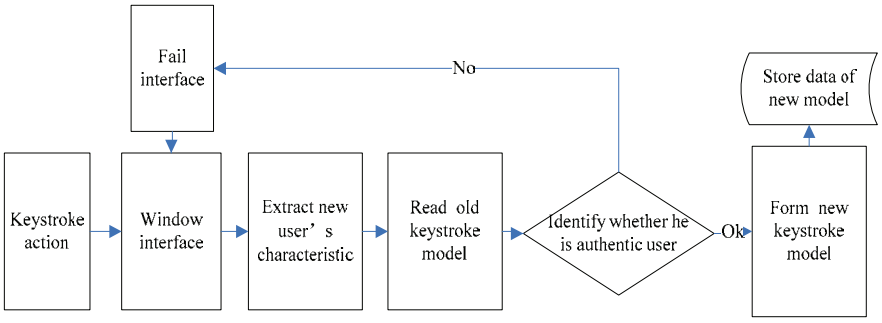


Figure 1. System process of identity authentication based on keystroke rhythm.

Table 1. Experimental data of latency time.

unit: Millseconds	latency time (rhythm features stored in system)				Latency time (authentic user)				latency time (imposters)			
1	19	33	19	32	16	38	19	37	23	32	27	30
2	17	37	20	35	17	37	19	32	21	32	26	29
3	18	37	17	36	22	40	18	35	23	28	24	27
4	16	42	22	37	17	41	19	31	25	37	29	29
5	18	41	19	32	17	39	17	37	22	32	26	28
6	18	35	19	39	18	37	19	36	23	36	26	31
7	17	37	18	34	17	34	18	34	24	29	26	32
8	16	36	20	36	18	37	17	36	21	32	27	32
9	15	38	19	32	17	34	19	35	22	29	26	30
10	19	39	14	37	18	35	20	34	19	30	25	25
11	17	40	18	30	17	35	15	31	25	30	21	26
12	17	38	14	32	17	39	17	32	20	34	21	27

4. Data Analysis and Results

The keystroke rhythms typed by users are our experimental data resources which are transformed from the mobile phone system to the computer through serial ports. The identification system requires users to login a password which contains five numbers [9]. Therefore, the vectors processed in system are four-dimensional vectors. We just choose three representive group of data in the Table 1 below. In these groups, the users need to type the password for twelve times. Then we do analysis based on these data.

Two groups of keystroke rhythm features can be gained by processing the data above, then we compare them with the rhythm features already stored in our system. The latency time curves of signature typed are shown in Figures 2-5.

Figures 2-5 show that only when the keystroke rhythm vector is nearly the same with the user characteristic profile, the tested user is considered as authentic user. However, the keystroke rhythm vector of imposters is far different from the user characteristic profile which means the difference is beyond the threshold value pre-established in system. Furthermore, the result demonstrates the improved model is able to accomplish the identity authentication based on keystroke rhythm in the mobile phone system.

In Table 2, M is the characteristic profile vector of user in system, and we choose four representive data to illustrate the performance of the improved model. Through the table we can clearly see that by the increase of threshold value, it has a good security performance in this system. Though we do not show the whole data, those listed above certainly have representive meanings in our study to a large extent. And the Rick Joyce model is also implemented in our system to make comparison with the improved model. The performance comparisons are showed in Figure 6.

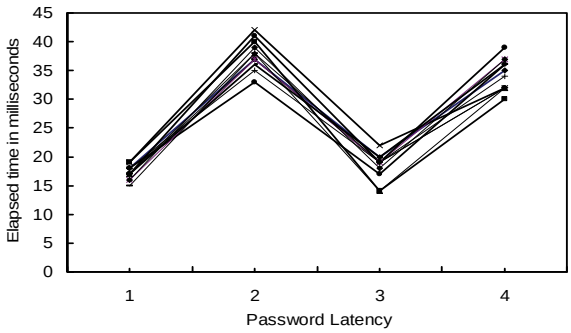


Figure 2. Keystroke rhythm features stored in system.

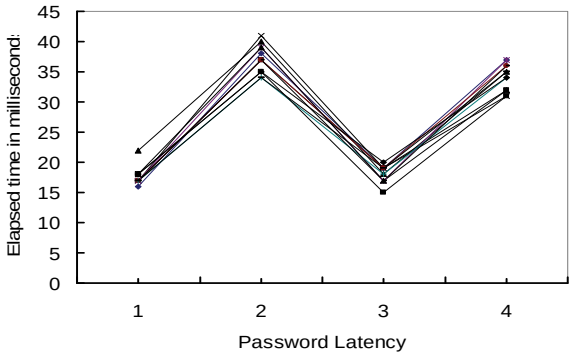


Figure 3. Twelve login attempts of authentic user.

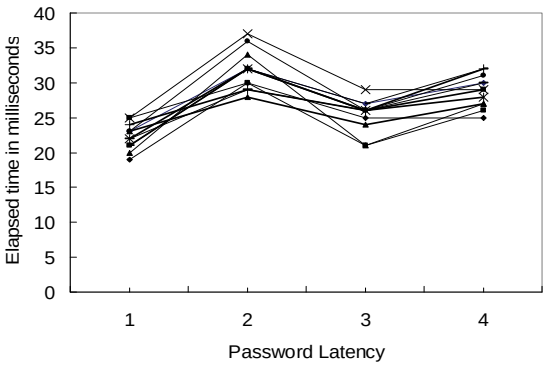


Figure 4. Twelve login attempts of imposters.

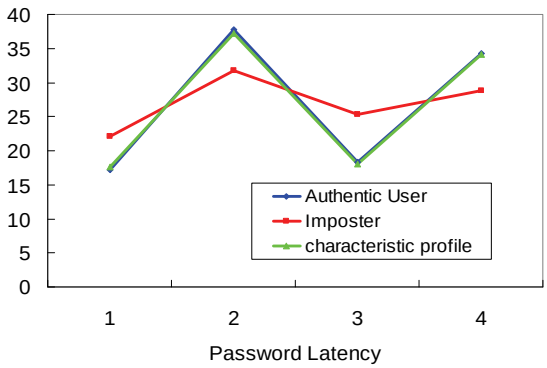


Figure 5. Comparison with the rhythm characteristic profile.

Table 2. Results comparison under different thresholds.

	latency	latency	latency	latency	Identity result
M	17.25	37.75	18.33	34.3	
$T_{authentic1}$	16	38	19	37	
σ:1	Ok	Ok	Ok	Ok	pass
σ:3	Ok	Ok	Ok	Ok	pass
$T_{authentic2}$	22	40	18	35	
σ:1	No	Ok	Ok	Ok	imposter
σ:3	Ok	Ok	Ok	Ok	pass
$T_{imposter}$	22.08	31.75	25.33	28.83	
σ:1	No	No	No	No	imposter
σ:3	Ok	No	No	No	imposter
σ:4	Ok	Ok	No	No	imposter

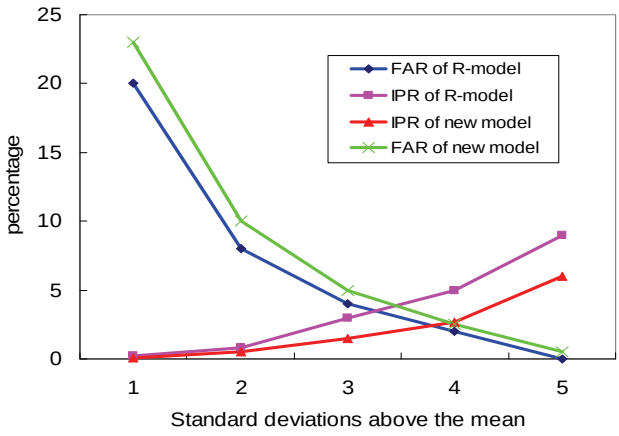


Figure 6. IPR and FAR versus threshold.

Figure 6 shows that though the IPR is not high in the Rick Joyce model (R-model) when the threshold value using one-and-one-half standard deviation [5], the FAR is not as ideal as IPR, nearly 14%, so it is not suitable for daily operations in the small memory system due to the troubles which will take to the phone users. In the improved new model, the growth rate of IPR is less than the R-model's, and it is only 1.5% in the improved model when the threshold value using triple standard deviation, which means the threshold work well. In terms of FAR, though the value is higher than the R-model's under the same threshold, the FAR is only about 5% in new model, and it is much lower than the R-model's FAR. With the increase of threshold, the FAR decreases a lot and the IPR increases slowly in the new model. Consider roundly the application in a small memory system, the new model is better to meet the actual operation needs, and this algorithm is dynamic, so along with the increase of authentication times, the identity authentication system can be improved automatically [10]. In the actual system design, through changing the multiple of the standard

deviation so as to influence the threshold value, we can easily adjust the performance in security and dynamic adaptation.

5. Conclusions

In this paper, a dynamic model based on the improved keystroke rhythm algorithm has been proposed and a mobile phone authentication system has been implemented. In this mobile phone system, the users could login the system only by typing a password which contains five known numbers. The system verifies the identity by using the dynamic model based on the improved algorithm. The experimental results show that it is able to accomplish the identification. Then, we compare the improved model with the original model in this system. It has illustrated that the FAR is 14% in the original model and 5% in the improved one, in which the FAR decreases by 9%, and the IPR increases slowly in the new model comparing with the original model. Therefore, the improved model has a high capability of authentication in the mobile phone system. Moreover, the performance in security and dynamic adaptation can be easily adjusted by changing the threshold value in the actual system design. And the improved model can be applied to meet different requirements and protect individual privacy. This model has good application value.

Although the dynamic identity authentication model based on the improved algorithm has been demonstrated in the small memory system, more experiments will be required to investigate how the keystroke rhythm model could be applied to other aspects. For the further study, other parameters, such as the length of password vector, the number of reference vectors, should be considered to enhance our ability of devising new and better models against unauthorized users or hackers.

6. References

[1] M. Zhu, J. Zhou, and J. K. Wang, "A new approach for user authentication based on biometrics," Computer Engineering, Vol. 28, No. 10, October 2002.

[2] S. Bleha, C. Slivinsky, and B.Hussien, "Computer-access security systems using keystroke dynamics[J]," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 12, December 1990.

[3] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication. future generation computing systems (FGCS)," Journal: Security on The Web (Special issue), March 2000.

[4] F. Monrose, M. K. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamic," Proceedings of the 6 ACM Conference on Computer and Communication Security, 1999.

[5] R. Joyce and G. Gupta, "Identity authentication based on keystroke latencies," Communications of the ACM, February 1990.

[6] F. Bergadano, D. Gunetti, and C. Picardi, University of Torino, "User authentication through keystroke dynamics [J]," ACM Transactions on Information and System Security, Vol. 5, No. 4, November 2002.

[7] W. G. de Ru and J. Eloff, "Enhanced password authentication through fuzzy logic[J]," IEEE Expert, Vol. 12, No. 6, November/December 1997.

[8] Z. H. Deng, S. Q. Zhuo, etc., "The development of applications in mobile phone systems," Science Press, March 2004.

[9] R. Gaines, W. Lisowski, and S. Press, "Authentication by keystroke timing: Some preliminary results[R]," Rand Corporation:Rand Report R-2560-NSF, 1980.

[10] J. Sleggett, G. Williams, and J. Usnick, "Dynamic identity verification via keystroke characteristics," International Journal of Man-Machine Studies, 1991.

Int. J. Communications, Network and System Sciences, **2009, 2, 720-731**
doi:10.4236/ijcns.2009.28083 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Evaluation of Network Stack Optimization Techniques for Wireless Sensor Networks

Jaein JEONG
Computer Science Division, UC Berkeley, Berkeley, California, USA
Email: *jaein@eecs.berkeley.edu*
Received August 19, 2009; *revised September* 20, 2009; *accepted October* 10, 2009

Abstract

We present a network stack implementation for a wireless sensor platform based on a byte-level radio. The network stack provides error-correction code, multi-channel capability and reliable communication for a high packet reception rate as well as a basic packet-level communication interface. In outdoor tests, the packet reception rate is close to 100% within 800 ft and is reasonably good up to 1100 ft. This is made possible by using error correction code and a reliable transport layer. Our implementation also allows us to choose a frequency among multiple channels. By using multiple frequencies as well as a reliable transport layer, we can achieve a high packet reception rate by paying additional retransmission time when collisions increase with additional sensor nodes.

Keywords: Wireless Sensors, Network Stack, Error Correction Code, Reliable Transport, Multi Channel

1. Introduction

In wireless sensor networks, it is desirable to have sensor nodes communicate without packet loss allowing information from a sensor node to be transferred reliably. In reality, however, sensor nodes suffer different levels of packet loss due to signal attenuation and multi-path effect [1–3]. In this paper, we take a practical approach to address this problem. We implement a network stack as an experiment vehicle based on a byte-level radio, the CC1000 transceiver [4]. Besides the basic packet-level communication, the network stack provides additional functionality for performance improvement such as error-correction code, retransmission and channel diversity. Based on the network stack we have implemented, we have evaluated how this extended functionality improves communication performance. Our network stack has reasonable performance in an outdoor environment with the packet reception rate close to 100% within an 800 ft range. We have found that error-correction code, retransmission and channel diversity improves this performance even further. For error-correction code, we have used SECDED (Single-Error Correction and Double-Error Detection) code [5,6]. Using SECDED code improved the packet reception rate, but it was not effective when packets were completely lost due to the multi-path effect. Retransmission was effective in reducing packet losses from the multi-path effect and contention from traffic. We also saw that using multiple radio channels was very effective in reducing collisions when multiple senders burst packets.

The rest of this paper is organized in the following way: Section 2 overviews background and related work; Section 3 describes the design principles of our network stack; Section 4 empirically evaluates our network stack and Section 5 concludes this paper.

2. Background and Related Work

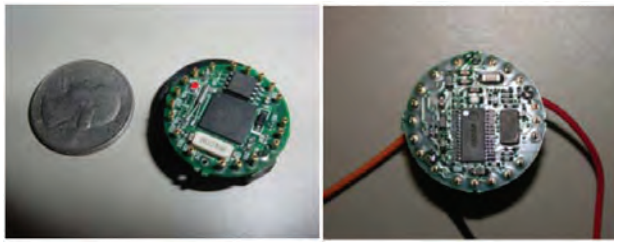
TinyOS provides communication capability using multiple layers like other networked systems: application, transport, network, and link layer. An *application* sees the radio as a service interface through which it can send and receive data in fixed sized packets. The *transport layer* provides best effort delivery and process-to-process communication. The *network layer* provides a routing tree. The *link layer* converts a packet to and from the byte data and interacts with the underlying radio chip. Since the link layer is so closely coupled with the radio chip, writing a network stack for a new platform usually requires rewriting the link layer.

Our project is based on some earlier works. Hill wrote a preliminary version of a network stack for the CC1000

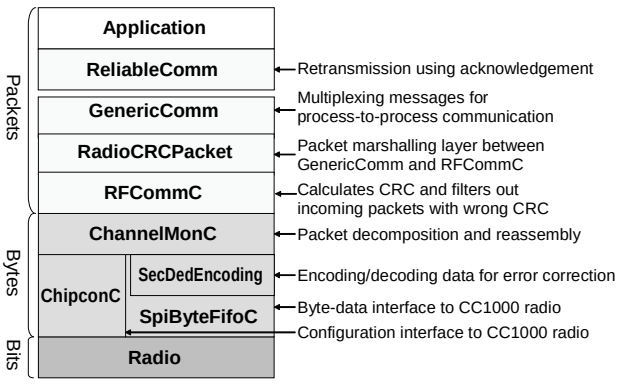
radio supporting Active Message number multiplexing and packet framing [7]. The CC1000 can operate in one of three frequency bands (433MHz, 866MHz and 900 MHz) and each band requires different values for external components and initialization parameters. The choice of frequency band is determined by the coverage and the number of legally usable channels: 900MHz is preferred for its relatively large selection of channels and 433MHz for its longer range. Hill initially wrote a network stack for the CC1000 radio in the 900MHz range and Crossbow technology modified it to support the 433MHz range. From the earlier work done by Hill and Crossbow technology, we found some room for improvement: 1) supporting error correction code to protect data from transient errors, 2) utilizing multiple channels of the CC1000 radio chip, 3) supporting reliable communication.

3. Design

In this section, we describe mechanism and design philosophy of our network stack for the Mica2dot platform [8] that is based on a byte-level radio transceiver CC1000 radio chip and the ATMega103L microcontroller [9], illustrated in Figure 1. First, we describe how we provide byte-level data and configuration interfaces (SpiByteFifoC and ChipconC modules in Figure 1) to abstract the low-level interface of the CC1000 radio with generic byte-level interfaces. Second, we describe how we provide a packet-level interface for the CC1000 radio (ChannelMonC module). Third, we describe how we provide error-correction capability to protect data from transient errors (SecDedEncoding module). Fourth, we describe how we provide an interface to utilize the diversity of multiple radio channels (ChipconC module). Fifth, we describe how we support reliable communication using retransmission (ReliableComm module).



(a) Mica2dot node – front and back



(b) Mica2dot network stack and description of its layers

Figure 1. Mica2dot and its network stack.

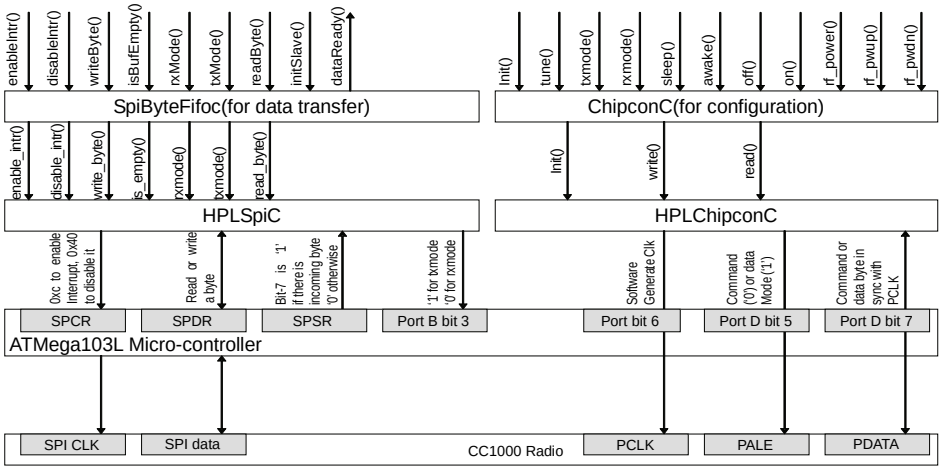
3.1. Abstracting Radio Hardware

1) *Byte-level data interface*: While at the physical level, the ATMega103L microcontroller communicates with the CC1000 radio in bits over two serial pins, it provides a byte-level interface through the SPI (Serial Peripheral Interface) registers: SPSR, SPDR and SPCR.

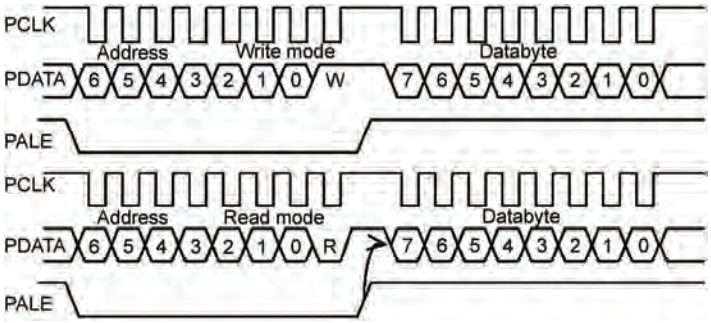
- SPSR (SPI Status Register) can be used to check whether there is an incoming byte or not. The most significant bit (bit 7) of the SPSR becomes high when there is an incoming byte in the buffer, low otherwise. The CC1000 radio can be switched between send and receive mode by changing the data direction of the data pin (bit-3 of port B of ATMega103L microcontroller).
- SPDR (SPI Data Register) is a byte buffer, which assembles either outgoing or incoming bits into a byte before reading or writing data to and from the CC1000 radio. When incoming bits are assembled into a byte, the ATMega103L microcontroller triggers SIG_SPI interrupt as well as setting bit 7 of the SPSR. This allows incoming data for the CC1000 to be efficiently processed using an interrupt instead of polling. One note is that only a send or a receive can be done at any time because the CC1000 radio has only a single buffer.
- SPCR (SPI Control Register) can be used to enable or disable interrupts. The SPI clock interrupt is triggered when SPCR is set to 0xc0, and is disabled when set to 0x40. Finally, data should be read or written at the same rate with that of the external device. The CC1000 radio is synchronized to the ATMega103L microcontroller through the SPI clock. The clock triggers an output pin of the microcontroller (bit-1 of port B) to the data clock pin of the radio chip (SPI CLK).

HPLSpiC and SpiByteFifoC are TinyOS modules that abstract the data interface from the ATMega103L microcontroller to the CC1000 radio. The methods for these modules are summarized in Figure 2(a) and Table 1.

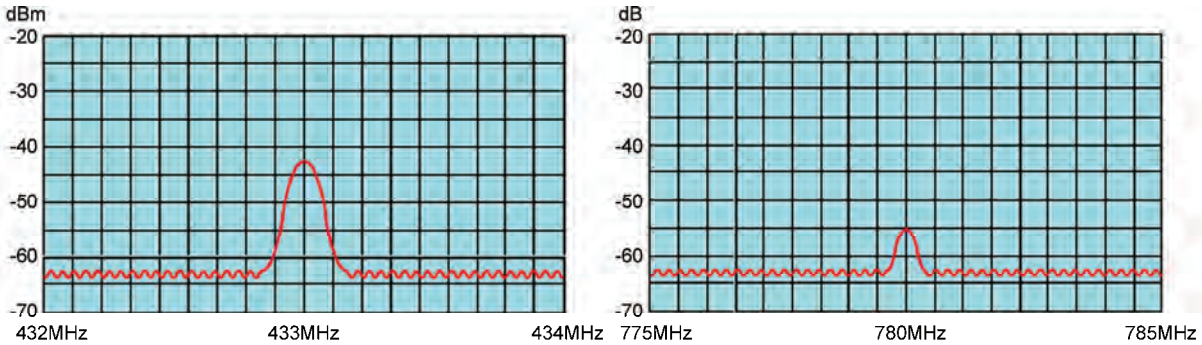
2) *Serial configuration interface*: While the CC1000 radio can transmit or receive data through the data interface, it needs to be configured for initialization or property changes such as radio frequency (explained in the next subsection) or transmission power level. The CC1000 radio exposes three pins PALE, PCLK and



(a) Hardware abstraction for CC1000 radio – SpiByteFifoC and ChipconC modules provide higher-level abstraction over CC1000 radio hardware



(b) Bit sequence for accessing CC1000 control registers – i) write a byte, ii) read a byte



(c) Waveform of CC1000 in spectrum analyzer – i) correctly configured, ii) misconfigured. Notice that the peak is over 10dB higher when the radio is correctly configured compared to when it is misconfigured.

Figure 2. Abstracting radio hardware.

PDATA for this purpose as shown in Figure 2(a). These pins are mapped to data port pins (port D bit 5, 6 and 7). By setting or clearing these pins, the microcontroller can send a sequence of bits (control register address, a byte of data and a bit representing whether the operation is write or read as it is shown in Figure 2(b). This is implemented as init(), write() and read() in the HPLChipconC module (Table 2). The ChipconM module implements high level functions such as setting the radio frequency or transmission power using these primitives.

3) *Using multiple channels*: The CC1000 radio allows us to select a channel at run time among a number of frequencies. The purpose of selecting channels is to reduce any interference from neighboring sensor nodes or other wireless devices. For the CC1000 radio chip to operate at a specific frequency, it needs to be configured with the correct frequency words and clock divisor byte. CC1000 transmits and receives at different frequencies and these frequencies are represented by two 24-bit frequency words. These frequencies are generated by dividing the

Table 1. Byte-level data interface to the CC1000 radio.

HPLSpiC

Method	Description
HPLSpi. enable_intr()	Enables SPI clock interrupt (sets SPCR to 0xC0)
HPLSpi. disable_intr()	Disables SPI clock interrupt (sets SPCR to 0x40)
HPLSpi. write_byte()	Writes a byte (writes a byte to SPDR)
HPLSpi. read_byte()	Reads a byte (reads a byte from SPDR)
HPLSpi. is_empty()	Bit-7 is '1' if there is an incoming byte, '0' otherwise
HPLSpi. txmode()	Switches to transmit mode (sets bit-3 of PortB as '1')
HPLSpi. rxmote()	Switches to transmit mode (sets bit-3 of PortB as '0')

SpiByteFifoC

Method	Description
SpiByteFifo. initSlave()	Initializes the SPI registers
SpiByteFifo. dataReady()	Called when there is an incoming byte
SpiByteFifo. enableIntr()	Calls HPLSpi.enable_intr( )
SpiByteFifo. diableIntr()	Calls HPLSpi.diable_intr( )
SpiByteFifo. writeByte()	Calls HPLSpi.write_byte( )
SpiByteFifo. readByte()	Calls HPLSpi.read_byte( )
SpiByteFifo. isBufEmpty()	Calls HPLSpi.is_empty( )
SpiByteFifo. txMode()	Calls HPLSpi.txmode( )
SpiByteFifo. rxMode()	Calls HPLSpi.rxmode( )

Table 2. Serial configuration interface to the CC1000 radio.

HPLChipconC

Method	Description
HPLChipcon.init()	Initializes CC1000 configuration registers
HPLChipcon.write()	Writes a byte to the CC1000 register with the given 7-bit address
HPLChipcon.read()	Reads a byte from the CC1000 register with the given 7-bit address

ChipconM

Method	Description
Chipcon. init()	Initializes CC1000 configuration registers and tunes the radio frequency to the given value
Chipcon. tune()	Tunes the radio frequency to the given value
Chipcon. txmode()	Sets the CC1000 in transmit mode
Chipcon. rxmode()	Sets the CC1000 in receive mode
Chipcon. sleep()	Puts the CC1000 in the sleep mode
Chipcon. awake()	Awakes the CC1000 radio from the sleep mode
Chipcon. off()	Turns off the CC1000 radio
Chipcon. on()	Turns on the CC1000 radio
Chipcon. rf_power()	Sets the transmit power for the CC1000 radio
Chipcon. rf_pwup()	Increments the transmit power for the CC1000 radio
Chipcon. rf_pwdn()	Decrements the transmit power for the CC1000 radio

Table 3. Channels available for Mica2dot in 433MHz band.

	CH1	CH2	CH3	CH4
Parameter for Chipcon.init( ) or Chipcon.tune( )	0	1	2	3
Tx freq (MHz)	433.02	433.64	434.20	434.71
Reg4-6	57f785	581785	583785	585785
Rx freq (MHz)	433.09	433.71	434.27	434.78
Reg4-6	580000	582000	584000	586000
Reg12 divisor (PLL)	60	60	60	60
Output power (dBm)	-45	-45	-47	-47

frequency synthesizing clock (we are using 14.7456 MHz) with the clock divisor byte. These values are set up in the ChipconM module.

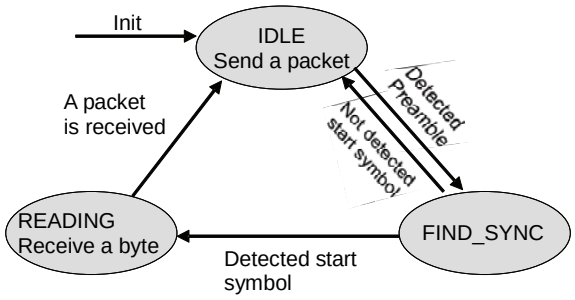
The CC1000 radio can operate on one of the three frequency range 433 MHz, 868 MHz or 900 MHz, depending on the selection capacitor and inductor values for the resonator and the filter in the assembled hardware. While 900 MHz is preferable for the wider selection of frequencies, we chose the 433 MHz band in favor of longer range. Recommended values are listed in [4], but none of them worked for the 433MHz band. We found four working channels by measuring signal strength for different values between 433 MHz and 435 MHz, as shown in Table 3.

Figure 2(c) shows the waveforms from a spectrum analyzer when the configuration is correct and wrong. In Figure 2(c)-(i), all the external components such as the resonator and filter are set to 433MHz band and the control register of CC1000 is correctly configured. The waveform has a peak at around 433MHz and its peak output power had around -45dBm using an inducting antenna in the spectrum analyzer input. In Figure 2(c)-(ii), control registers are set to 433MHz, but the inductor in the resonator is set to the value used in the 900MHz band. Since this resonator value doesn't match the other external components and the configuration value, its results at its peak is somewhere in the middle between 433MHz and 900MHz and its output power is much weaker than it should be. Actually, the initial build of the 433MHz Mica2dot nodes had this bug, and had to be addressed in the early stages of our project.

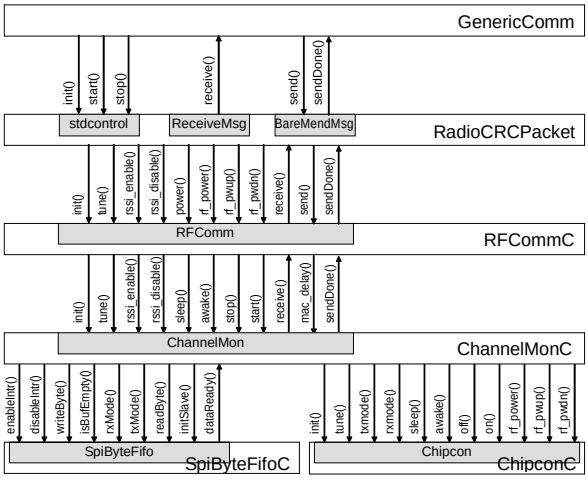
3.2. Providing Packet-Level Interface

Packet decomposition and reassembly is done in ChannelMonC with the help of the SpiByteFifoC module. SPI is synchronized to the microcontroller by the clock and generates an interrupt at regular intervals. At each interrupt invocation, the interrupt handler SIG_SPI in the SpiByteFifoC module is called. Then, we can determine if we can send or receive a byte by looking at the control register (SPSR) as it is shown in Figure 3(a).

Initially, it is in the IDLE state. If no incoming bytes are available, ChannelMonC sends a packet. Since the radio chip transfers data in bytes, we need to signal the beginning and the end of a packet. This is done by having a special sequence of bytes (preamble and start symbol) at the beginning and a fixed number of bytes after that. After sending the preamble and start symbol, ChannelMonC sends the data bytes. Data bytes can be



(a) State transition diagram for packet decomposition and reassembly



(b) Packet level interface in the network stack

Figure 3. Providing packet-level interface.

sent as they are or can be sent after being encoded with error correction code for integrity. We used the SecDedEncoding module which implements a single-error-correction-and-double-error-detection (SECDEC) code. The version of ChannelMonC with error correction code is ChannelMonEccC. When there is an incoming byte, ChannelMonC reads the byte and sees whether the sequence of bytes received matches the preamble. If it does match, it goes to the FIND_SYNC state. If the next incoming bytes match the start symbol, it goes to the READING state. After reading the fixed length of data (36 bytes is the default), ChannelMonC notifies the arrival of a packet to the RFCommM module. The packet interface in the network is summarized in Figure 3(b).

3.3. Providing Error-Correction Code

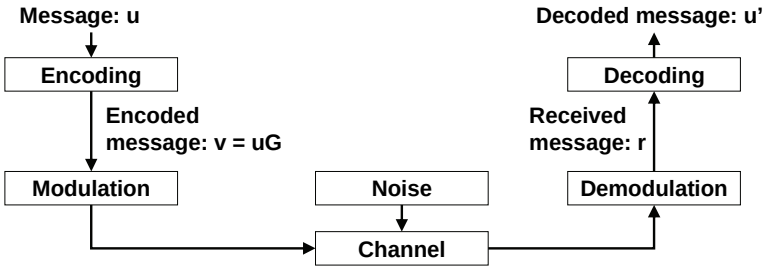
In this section, first, we describe a coding theory based on linear block codes over field GF(2), which can be implemented in a simple and efficient way on resource constrained wireless sensor nodes. Then, we describe our implementation of error correction code based on linear block codes. The general process of encoding, transmission and decoding of the message is shown in Figure 4(a).

1) *Theory*: The encoding of message **u** into codeword **v** can be achieved by multiplying the message **u** with the generation matrix **G**. For data of width *k*-bits, **G** is of the form $[\mathbf{I}_k : \mathbf{C}]$, where $\mathbf{I}_k$ is the *k*-by-*k* identity matrix and **C** is the *k*-by-*r* binary matrix.

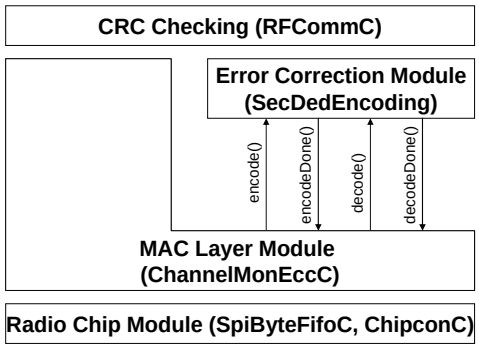
On the receiver end, a syndrome is calculated for detecting and possibly correcting errors. The syndrome **s** is calculated from the received signal **r** and the parity matrix **H**. The parity matrix **H** is constructed from the generator matrix **G** and is of the form $\mathbf{H} = [\mathbf{C}^T : \mathbf{I}_r]$, where **r** is the number of parity bits. Denoting the error vector by **e**, we have

$$\mathbf{s} = \mathbf{rH}^{\mathbf{T}} = (\mathbf{v} + \mathbf{e})\mathbf{H}^{\mathbf{T}} = \mathbf{uGH}^{\mathbf{T}} + \mathbf{eH}^{\mathbf{T}} = \mathbf{eH}^{\mathbf{T}}$$

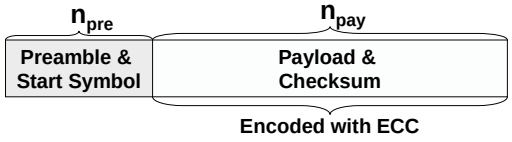
Here, $\mathbf{GH}^T = [\mathbf{I}_k : \mathbf{C}][\mathbf{C}^T : \mathbf{I}_r]^T = \mathbf{C} + \mathbf{C} = \mathbf{0}$ (addition is in GF(2)). The non-zero **s** implies that an error occurred. Depending on the capability of the error correction code, the non-zero syndrome is compared with a row or a sum of rows of $\mathbf{H}^T$. If there are such rows, the correct codeword can be determined by flipping corresponding bits in the received signal.



(a) Encoding, transmission and decoding of message



(b) Implementing error correction code with network stack



(c) Packet length with error correction code

Figure 4. Providing error-correction code.

The receiver can decode the corrected codeword by solving the equation **v** = **uG**. Especially, for a systematic code where the first *k*-columns of generator matrix **G** form an identity matrix, u is just first *k*-bits of **v**.

2) *Odd-weight-column code*: Odd-weight column code can correct single bit errors and detect double bit errors (SECDED) [5]. As the name implies, each column of the parity matrix **H** has an odd number of 1s. To construct an odd-weight column code for an input of *k* data bits, the parity matrix **H** should include a sufficient number of parity bits *r* so that the number of columns of **H** is at least *k* + *r*. For example, *r* = 5 parity bits are needed to recover *k* = 8 bits of data, and *r* = 6 parity bits are needed to recover *k* = 24 bits of data.

The columns of **H** are constructed as follows:

- The last *r* columns of **H** form the identity matrix $\mathbf{I}_r$.
- The first *k* columns of **H** are chosen from any other odd weight column vectors than the ones used in $\mathbf{I}_r$.

For example, **G**, **H** for *k* = 8, *r* = 5 are:

$$\mathbf{G} = [\mathbf{I}_8 : \mathbf{C}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{H} = [\mathbf{C}^T : \mathbf{I}_5] = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We now give an example of how data is encoded, and how the received message can be corrected. Let the message being sent be $\mathbf{u} = [0100\ 0010]$. The encoded message **v** is therefore

$$\mathbf{v} = \mathbf{uG} = [0100\ \ 0010\ \ 10111]$$

Assume that the second bit of the encoded message is inverted due to noise in the wireless channel. Then, received message **v'** is $[0100\ 0010\ 10111]$. Multiplying the received message by the transpose of the parity matrix **H**, we get the syndrome **s** as follows:

$$\mathbf{s} = \mathbf{v'}\mathbf{H}^T = [0\ \ 1\ \ 0\ \ 1\ \ 1]$$

We note that the syndrome obtained is the second row of $\mathbf{H}^T$, which implies that second bit of the codeword is inverted. Thus, we can get correct codeword $\mathbf{v} = [0100\ \ 0010\ \ 10111]$. Since the generator matrix **G** is a systematic code, the data bits can be decoded by taking the first-*k* bits of the codeword. Thus, $\mathbf{u} = [0100\ \ 0010]$.

3) *Implementation*: A wireless sensor node communicates with other sensor nodes using the network stack shown in Figure 1. Although error correction code (ECC) can be located in any other layer, we decided to have the error-correction- code module in the MAC layer, which interfaces application / network layer with physical layer by packetizing the received data bytes and fragmenting a packet to be sent into data bytes. Since this approach does not change the interface to the application/network layer, it has an advantage that any applications written for non-ECC version of MAC can run without any code modification. We implemented an error correction code module for odd-weight-column code SecDedEncoding, which takes 8-bit data and generates a 13-bit codeword.

Figure 4(b) shows how our error-correction-code implementation interacts with the MAC layer through the interface RadioEncoding. When a packet is to be sent, the MAC layer module ChannelMonEccC calls the method encode for each byte of data in the packet. After a sufficient number of input data bytes have been received, the input data bytes are encoded by the internal encoding function radio_encode_thread and codeword is passed to the ChannelMonEccC through encodeDone event. When data bytes are passed by the physical layer, the MAC layer module ChannelMonEccC calls the method decode for each byte of its received packet. After a sufficient number of input data bytes have been received, the received data bytes are decoded by the internal decoding function radio_decode_thread and the original data bytes are sent to ChannelMonEccC through

decodeDone event. The internal encoding function radio_encode_thread calculates parity bits by comparing each bit of input data bytes. This corresponds to calculating **mG**, where **m** is the message and **G** is the generator matrix.

The internal decoding function radio_decode_thread calculates the syndrome **s** by looking at each bit of received data bits. This is equivalent to $\mathbf{rH}^T$ where **r** is received data and $\mathbf{H}^T$ is the transpose of the parity matrix. A non-zero syndrome implies an error and the position of bit errors can be found by comparing the syndrome with column vectors of **H** matrix. We made this lookup fast by using an array that maps any possible syndrome value to the error bit position.

In general, the MAC layer of WSN consists of the following fields: preamble, start symbol, payload and checksum (Figure 4(c)). When error-correction code is used, the data bytes in the payload and checksum are encoded into codeword while the data bytes for the preamble and start symbol are not encoded. Then, we can estimate the transmission overhead of an error-correction code for the following parameters:

- n_{pre} : length of preamble and start symbol (bytes)
- n_{pay} : length of payload and checksum (bytes)
- r_{ecc} : length of codeword for one-byte data
- n_{ecc} : data bytes transmitted with ECC (bytes)
- n_{no_ecc} : data bytes transmitted without ECC (bytes)

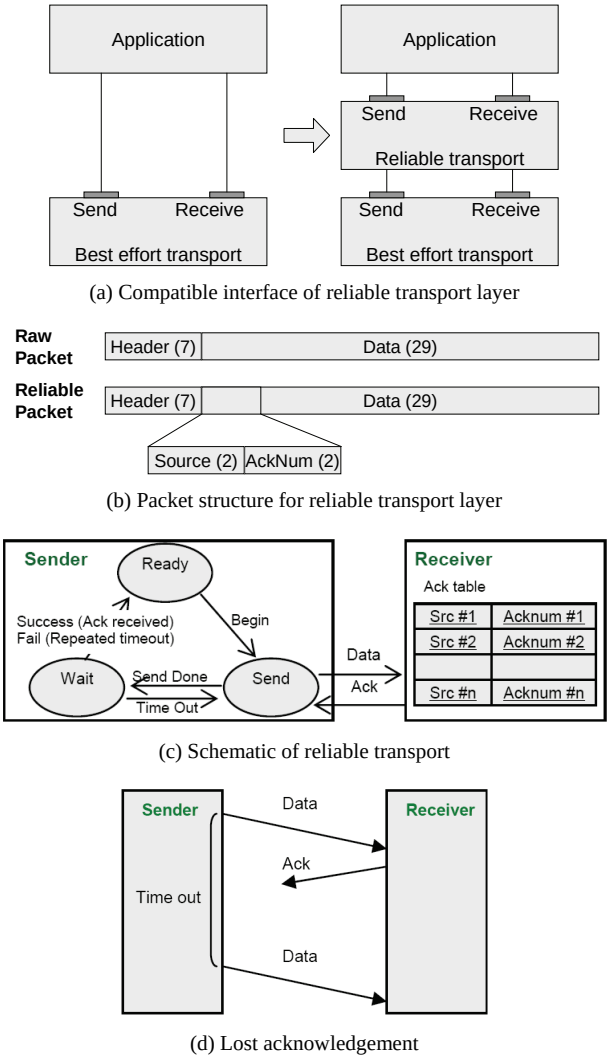
$$\begin{aligned} Overhead &= \frac{n_{ecc} - n_{no\_ecc}}{n_{no\_ecc}} \\ &= \frac{(n_{pre} + n_{pay} \cdot r_{ecc}) - (n_{pre} + n_{pay})}{n_{pre} + n_{pay}} \\ &= \frac{n_{pay}}{n_{pre} + n_{pay}} \cdot (r_{ecc} - 1) \end{aligned}$$

Since the TinyOS distribution has n_{pre} = 20 and n_{pay} = 36 for CC1000 radio, the overhead for our ECC implementation can be calculated as follows:

$$\begin{aligned} Overhead &= \frac{n_{pay}}{n_{pre} + n_{pay}} \cdot (r_{ecc} - 1) \\ &= \frac{36}{20 + 36} \cdot (\frac{2B}{1B} - 1) = 64.3\% \end{aligned}$$

3.4. Reliable Transport Layer

We wanted reliable communications. But we also wanted compatibility and ease of use. So we designed it to give it the same interface as that of existing best-effort transport layer. As shown in Figure 5(a), we designed the reliable transport layer so that it can be inserted between a best-effort transport layer and application layer without any significant modification to the application.



(a) Compatible interface of reliable transport layer

(b) Packet structure for reliable transport layer

(c) Schematic of reliable transport

(d) Lost acknowledgement

Figure 5. Reliable transport layer.

We also wanted a lightweight layer. A compatible and lightweight approach steered us to implement connection-less communication. Interfaces of existing best-effort transport layer support only connection-less communication. In the reliable transport layer, Connection information is managed globally.

To guarantee reliable communication, we mainly used acknowledgment and retransmission. Packet structure came to incorporate more information. Sender and receiver use this information and react properly according to it.

1) *Reliable message*: For reliable communication, additional meta-data (source address, acknowledgment number) needs to be included in each packet. The packet size is 36 bytes. 7 bytes are already used by lower layers for meta-data. And 29 bytes are used as data field. 4 more bytes (2 for source address, 2 for acknowledgment) are taken from the data field for meta-data in the reliable

transport layer. Now the length of the data field has decreased from 29 bytes to 25 bytes (13.8 percent loss).

Sender gets a packet from an upper application layer. And it realigns data, adds source address and acknowledgment number, and sends it to lower best-effort transport layer. Receiver also does the same conversion. The packet structure is shown in Figure 5(b). The use of additional meta-data is transparent to applications except the decreased size of the data field.

2) *Sender*: Sender is a finite state machine. When sender gets a packet from an upper application layer, it adds a source address and acknowledgment number as shown in the Figure 5(b), realigns data, and passes the packet to the lower best-effort transport layer. If the sender receives an acknowledgment from the receiver, it reports a success to the upper application layer. If it does not receive an acknowledgment but times out, it retransmits the unacknowledged packet. The amount of waiting time is a random number between T and 2T. After N successive time-outs, the sender reports a failure to the upper application layer. For simplicity, the sender uses block-and-wait strategy. The sender only needs to remember the current receiver's information. Figure 5(c) shows the main part of a state diagram of the sender. Since there is no queue in the lower layer, if sender tries to send a packet while the receiver of the same node is also replying by sending an acknowledgment, the packet from the sender can be lost. So a buffer of size 1 is used by the sender. When an acknowledgment is being processed, sender saves the packet in the buffer and transmits after the receiver of that sensor completes the acknowledgment.

3) *Receiver*: Receiver is also a finite state machine. When the receiver gets a packet from the lower best-effort transport layer, it looks at the source address and acknowledgment number. If it is a new packet, it sends an acknowledgment and passes the packet to the upper application layer. If it is a packet already received, it only sends an acknowledgment to the sender. To decide whether the received packet is a new packet or an already received one, it maintains connection information in the Ack table. The table has a pair of source addresses and acknowledgment number as an entry. In case of a lost acknowledgment, as in Figure 5(d), duplicate data packets can arrive. Then we should not report the second packet, and we need a table for this purpose.

Since the size of table is limited, it cannot handle an arbitrary number of connections all the time. So a FIFO algorithm is used to replace entries in the table. To reduce table lookup time, a reverse chronological search is used. It looks up the most recent connection first, and then next recent connection, and so on.

4. Evaluation

In order to see the effectiveness of our network stack implementation, we set up two kinds of experiments: range test and contention test. The range test, which was set up in the middle of the UC Berkeley campus (Figure 6), is to measure the performance of the network stack in a non-contentious environment. In the range test, the sender sends a number of packets and the receiver counts how many packets it received from the sender as we moved the sender farther from the receiver. The contention test, which was set up in an indoor environment, is to measure the performance of the network stack in the presence of traffic as we vary the number of senders or the number of radio channels being used. For each test, we used the packet reception rate as an indicator of effectiveness for the transmission method.

4.1. Range Test

In order to see whether our network stack performs reasonably in a non-contentious environment, we measured the packet reception rate as we vary the distance between the sender and the receiver. As a preliminary test, we measured the performance only with the basic configuration of the network stack without using any additional functionality such as error-correction code, retransmission or multiple radio channels. Figure 7(a) shows that the network stack performs well up to 800 ft with the packet reception rate higher than 90%. But, the packet reception rate drops severely after this point.

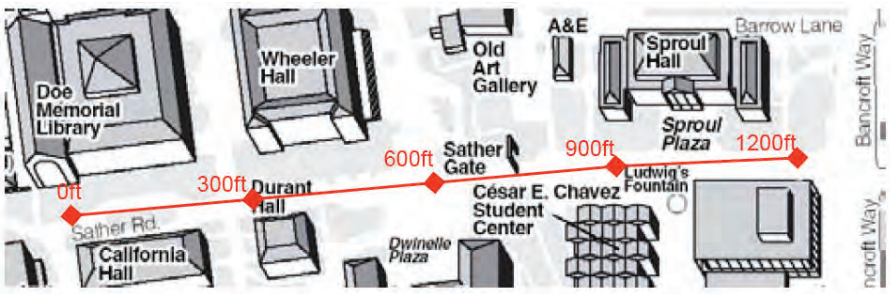
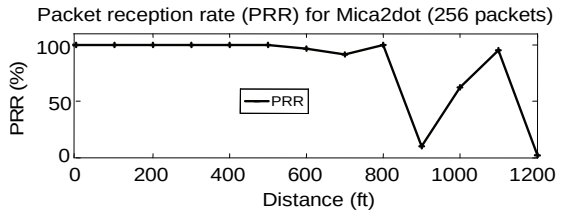
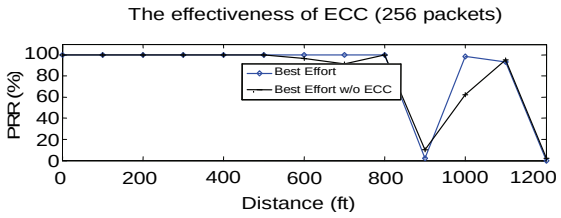


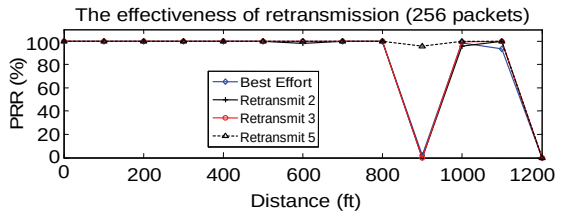
Figure 6. Locations of outdoor experiment.



(a) Base-case



(b) Effectiveness of ECC



(c) Effectiveness of retransmission

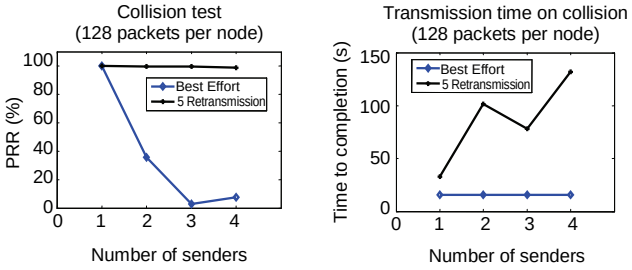
Figure 7. Range test.

In order to see whether additional functionality can improve the performance, we first tested the case with error-correction code. We measured the performance of the network stack with SECDED error-correction code and compared it with the performance with the base case. Figure 7(b) shows that the ECC implementation was more resilient to errors and had a better packet reception rate. However, the SECDED code is not effective when the packets were completely lost (900 ft).

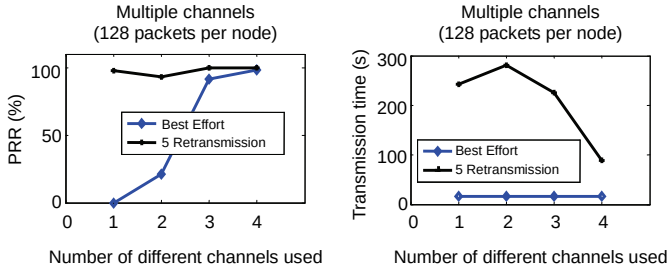
We have seen that using SECDED error correction improves the performance of the base case, but it still has a problem of packet losses. In the next test, we used both error-correction code and retransmission to see whether we can further improve the performance of the network stack. We measured the packet reception rate of the network stack for the four different implementations: the implementation with no retransmission and the ones with 2,3 or 5 retransmissions (Figure 7(c)). All implementations used SECDED for integrity. Retransmission was slightly better than the best effort transmission. The difference between the three retransmission schemes were not that noticeable except that 5 retransmission could receive the message while all the other methods failed at 900 ft. We found this was possible because radio waves from the sender took different paths when the sender retransmitted the packets.

4.2. Contention Test

In order to see how effective retransmission handles contentions, we set up an experiment. We placed multiple senders (1, 2 or 4) and a receiver close by and measured the packet reception rate and the transmission time for two extreme cases: no retransmission and 5 retransmissions. Figure 8(a) shows that the effect of retransmission in a closely populated area is very noticeable.



(a) Effect of retransmission



(b) Effect of using multiple channels

Figure 8. Contention test.

Table 4. Time to send / receive 512 packets.

Best Effort	Retransmission (5 retries)	Retransmission (0 retries)
31 sec	64 sec	32 sec

It reduced most of the packet drops due to collision with increased transmission time. This shows that packets are very likely to be dropped when multiple nodes are sending packets in bursts and that packet drops can be avoided with retransmission. We can also see that retransmission takes more time as the rate of collision increases.

In the next experiment, we wanted to see the effect of multiple radio channels. We prepared 8 nodes, dividing them into groups, each of which is composed of one sender and one receiver. We measured the packet reception rates and the transmission time for the non-retransmission implementation and the 5 retransmission implementation of the four senders. We varied the number channels used (1, 2 and 4) to see the effect of multiple channels on the two implementations. Figure 8(b) shows that using multiple channels was more effective with the non-retransmission implementation than retransmission implementation. This was expected from the results of the range test in that retransmission received most of the packets whereas non-retransmission received only 10% of the packets. Using multiple channels also helped the retransmission time. It used smaller amounts of transmission time when more channels are available. When there are fewer channels available, it spent more for transmission but it still achieved a high packet reception rate. This implies that retransmission and the use of multiple channels can be beneficial for reliable packet delivery. We can also infer that there is some interference among channels. Otherwise, for the case of 4 channels, the ratio of received packets should be very close to 100% for the best effort transport.

4.3. Overhead of Reliable Transport Layer

To measure the overhead of the sender, we eliminated the wait for acknowledgment on the sender side. And for 512 packets, we measured completion time. The result is shown in Table 4. The overhead is negligible for the sender. To measure the overhead of the receiver, we made the receiver send data to another sensor node. However, the two sensor nodes interfered with each other. Unfortunately we were not able to get correct results. We surely expect some overhead for the receiver side, because it should send a packet for each incoming packet while this is not needed in the best effort transport. In retransmission, every packet involves two transmissions. This explains why retransmission takes about twice as long as the best effort does.

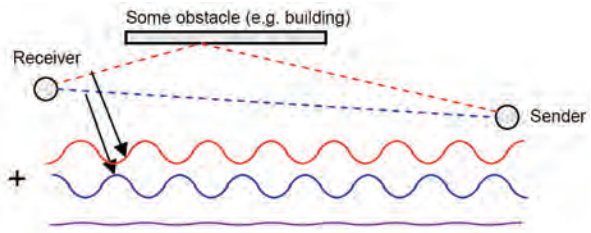


Figure 9. Multi-path effect.

4.4. Multi-Path Effect and Theoretical Limit of Range

If we look at the range test results in the previous Figures, the graphs consistently had dips at 900 ft. Once the sender moves beyond that distance, the receiver received the packets from the sender again. This happened because the radio signal is propagated through waves. Radio waves from the sender take paths while they travel and their phase can change when they reflect on some obstacles. Waves of opposite phase cancel each other out and the resulting signal becomes weaker than the sensitivity of the receiving node, thus packets cannot be heard. This phenomenon is called multi-path effect (Figure 9). More complicated devices like CDMA cellular phones use multiple antennas of different phase to avoid the antenna of different phase to avoid problem, but we cannot depend on this method because CC1000 has only single antenna. However, we can around this by having intermediate nodes between the two nodes and by having the intermediate nodes relay the packets.

5. Conclusions

We presented an implementation of the network stack for the CC1000 radio transceiver. It has reasonable performance in an outdoor environment with the packet reception rate close to 100% up to 800ft. To improve its performance in a more contentious environment, we extended the network stack with error-correction code, reliable transmission and multiple radio channels. For error-correction code, we used a SECDED code, and it was effective in improving the packet reception rate except when the packets were completely lost due to the multi-path effect. The reliable transmission scheme was effective in reducing packet loss from the multi-path effect and contention from traffic, but its overhead was a bit high when there was a great deal of collision. This was caused by some of the implementation decisions. In our reliable transmission scheme, senders try to retransmit unacknowledged packets after a random amount of time within the timing window of fixed size. We found

that this does not help under high collision even though the waiting time varied within the timing window. We expect that increasing the timing window size similar to exponential back-off will reduce the transmission rate so that the overall system can make progress. In the reliable transport layer, the sender's window size is one and this causes the sender to block and wait. Increasing the window size will reduce the waiting time and improve the transfer rate. This requires that the sender keeps an 'Ack table' to buffer unacknowledged packets. Using multiple radio channels was very effective in reducing collisions when multiple senders burst packets. Currently, the channel is statically tuned at compile time, but performance can suffer when the channel is mis-configured or there is contending traffic at the configured channel. A dynamic frequency allocation mechanism is needed to address this problem.

6. Acknowledgements

This work is supported by the Defense Advanced Research Projects Agency under a contract F33615-01-C1895 ("NEST"), the National Science Foundation under grants #0435454 ("NeTS-NR") and #0454432 ("CNS-CRI"), a grant from the Keck Foundation, and generous gifts from HP and Intel.

7. References

[1] K. Sohrabi, B. Manriquez, and G. J. Pottie, "Near ground wideband channel measurement in 800 - 1000mhz," IEEE Vehicular Technology Conference, July 1999.

[2] A. Woo and D. E. Culler, "A transmission control scheme for media access in sensor networks," In The Annual International Conference on Mobile Computing and Networking (MobiCom'01), July 2001.

[3] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," In The ACM Conference on Embedded Networked Sensor Systems (SenSys'03), November 2003.

[4] Cc1000 data sheet. http://www.chipcon.com/files/CC1000_Data_Sheet_2_1.pdf.

[5] M. Y. Hsiao, "A class of optimal minimum odd-weight-columnsec-dedcodes," IBM Journal of Research and Development, Vol. 14, No. 4, July 1970.

[6] J. Jeong and C.-T. Ee, "Forward error correction in sensor networks," In The First International Workshop on Wireless Sensor Networks (WWSN'07), June 2007.

[7] J. Hill. Cc1000 network stack. http://local.cs.berkeley.edu/grad/jaein/xbow.tgz.

[8] Mica2dot. http://www.xbow.com/products/Product_pdf_files/ Wireless_pdf/MICA2DOT_Datasheet.pdf.

[9] Atmega 103l microprocessor data sheet. http://www.atmel.com/dyn/ resources/prod_documents/doc0945.pdf.

Int. J. Communications, Network and System Sciences, **2009, 2, 732-741**
doi:10.4236/ijcns.2009.28084 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Cooperative Location Management Scheme for Mobile Ad Hoc Networks

Demin LI[1], Jiacun WANG[2], Liping ZHANG[3], Hao LI[1], Jie ZHOU[4]
[1]*College of Information Science and Technology, Donghua University, Songjiang District, Shanghai, China*
[2]*Department of Software Engineering, Monmouth University, West Long Branch, NJ, USA*
[3]*College of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Songjiang District, Shanghai, China*
[4]*School of Science, Donghua University, Songjiang District, Shanghai, China*
Email: *deminli@dhu.edu.cn, jwang@monmouth.edu, zhangliping@sues.edu.cn, haoli@mail.dhu.edu.cn*
Received July 8, 2009; *revised August* 17, 2009; *accepted September* 22, 2009

Abstract

A mobile ad hoc network (MANET) is a kind of wireless ad hoc network. It is a self-configuring network of mobile routers connected by wireless links. Since MANETs do not have a fixed infrastructure, it is a challenge to design a location management scheme that is both scalable and cost-efficient. In this paper, we propose a cooperative location management scheme, called CooLMS, for MANETs. CooLMS combines the strength of grid based location management and pointer forwarding strategy to achieve high scalability and low signaling cost. An in-depth formal analysis of the location management cost of CooLMS is presented. In particular, the total location management cost of mobile nodes moving at variable velocity is estimated using the Gauss_Markov mobility model for the correlation of mobility velocities. Simulation results show CooLMS performs better than other schemes under certain circumstances.

Keywords: Ad hoc Network, Grid, Home Region, Location Management, Forwarding Pointer, Cost Estimation

1. Introduction

Mobile ad hoc networks consist of wireless hosts that communicate with each other in the absence of a fixed infrastructure. Some examples of the possible uses of ad hoc networks include soldiers on the battlefield, emergency disaster relief personnel, and networks of laptops. Routing a packet from a source to a destination in a mobile ad hoc network is challenging because nodes in the network may move and cause frequent and unpredictable topological changes. Thus, when two nodes travel apart, they may no longer have a direct link between them. Likewise, if a node moves behind a hill, its links to its neighbors may be severed because of fading. Other reasons for changes in topology include jamming and the entry of new nodes to the network or departure of the existing nodes from the network.

Location management enables an ad hoc network to track the locations of mobile terminals between call arrivals. Since mobile users are free to move within the coverage area, the network can only maintain the approximate location of each user. When a connection needs to be established for a particular user, the network has to determine the user's exact location within the defined granularity. Location management for an ad hoc network contains three components: *location update*, *maintaining home regions*, and *locating a node*. The operation of informing the home region about the current location of the mobile user is known as location update or location registration, and the messaging cost, measured in packets per second per node, of performing these activities is the *cost of location update*. When a node moves from region *A* into region *B*, it needs to inform nodes in region *A* of its departure and meanwhile, it needs to inform nodes in region *B* of its arrival. It also needs to collect location information about all the nodes registered in region *B*. The operation of these activities is known as maintaining home regions, and the messaging cost of performing these activities is the *cost of maintaining home regions*. When a node receives a data packet for some destination, it needs to find the current location of the destination before sending packets to it. The operation of determining the location of the mobile user is called terminal locating or paging, and the messaging cost of performing

these activities is the *cost of locating a node*.

Several approaches have been proposed to address ad hoc network routing and costs [1–6]. In [4] a scalable routing protocol is presented. This protocol relies on a location update mechanism that maintains approximate location information for all nodes in a distributed pattern. As nodes move, this approximate location information is constantly updated. To maintain the location information in a decentralized way, this paper maps a node ID to a geographic sub-region of the network. Any node present in this sub-region is then responsible for storing the current location of all the nodes mapped to this sub-region. In order to send packets to a node, the sender first queries the destination's sub-region for the approximate location of the destination, and then uses a simple geographic routing protocol to forward the packets to the destination's approximate location. It is therefore easy to see that the location update cost in this protocol is dependent on the speed of node movement.

SLALoM, a grid based location management scheme, scales well in large, mobile ad-hoc networks [5]. The scheme divides a square into some unit regions which is called order-1 squares. It then combines K^2 of the order-1 squares to form order-2 squares. A node's home region will consist of an order-1 square. With some exceptions, every node has a home region in each order-2 square. If an original square area is A, every node has A/K^2 home regions. The scheme partitions those home regions into near home region. By constructing the tree of home regions of a node based on some rules, the location update information of a node is easily disseminated by the span tree. When a node moves from one order-1 to another, it not only sends a message to its nearby home regions, but also sends messages to the eight other home regions. Hence the location management cost is increased.

A novel multi-level hierarchical grid location management protocol with only one home region for a node that is called HGRID, for large scale ad hoc networks, is introduced in [6]. The paper shows that the average per node signaling cost in HGRID grows only logarithmically in the total number of nodes in a uniformly randomly distributed network—a substantial improvement over the signaling cost incurred by current location management schemes. If S (source) and D (destination) are co-located in the same grid, the location of D, as indicated by the location reply, is accurate and S can forward the data directly to D's location. Otherwise, the location is approximate, and S forwards the data packet to the location server specified in the location reply. When the data packet reaches the specified grid, the server v that receives the packet checks its neighbor table to see if D is co-located in the same grid. If so, the packet is successfully forwarded to the destination. Otherwise, the server searches for the D in its location database. By construction, v must have an entry for D in its location database. If v has accurate information about D, it further forwards the packet to D, otherwise, the packet is forwarded to the next location server which is in a level lower than v in hierarchy, but has more accurate information about D's position. This process continues until the packet reaches D, or it reaches a lower grid, and the node that receives the packet drops it since it has no information about D. This can happen because D would have left this grid, and D's location update to its new hierarchical leader failed to reach the leader before the data packet was forwarded by the leader to D's previously visited grid. Some times for location updating, the mobile node may inform two or more leaders, and those actions also increased the location management cost.

In order to decrease the location management cost, we propose a cooperative location management scheme, CooLMS, which is a nice integration of the grid model and pointer forwarding strategy. Pointer forwarding strategy was developed for location tracking in PCS networks. It helps reduce the cost of location update because when a node changes its location, it doesn't need to update its location information in its home region. It also helps reduce the cost of finding nodes because, with a forwarding pointer in place, a node can find out the location information of a destination node without the communication to the home region of the destination node. Pointer forwarding strategy has been recently studied quite intensively. Some variations have been proposed, such as a one-step painter forwarding strategy [13], location tracking pointer forwarding [14], two-level pointer forwarding strategy [15], K-step pointer forwarding strategy [16], and a combination of local anchor and forwarding pointer [17]. But to the best of our knowledge, nobody has ever introduced the pointers forwarding strategy into to the mobility management of mobile ad hoc networks. With the adoption of the pointer forwarding strategy, CooLMS reduces location update costs significantly.

In addition to the new effective location management scheme CooLMS, the paper also presents a more realistic location management cost model. Considering mobile users' movement is generally confined to a limited geographical area, the motion velocity of a mobile node is *variable*. Furthermore, the change of a mobile's velocity within a short time is limited due to physical restrictions. Therefore, a mobile user's future velocity is variable and likely to be correlated with its past and current velocity. The Gauss–Markov model [9] represents a wide range of user mobility patterns, including, as its two extreme cases, the random walk [10,11] and the constant velocity fluid-flow models. Since it captures the essence of the correlation of a mobile's velocities in time, we use it to specify the characteristics of mobile node movement.

Considering the measurements of mobile motion velocity are not necessarily consecutive, this paper presents an approach to the total cost discrete estimation of variable velocity mobile location management for ad hoc mobile networks.

The rest of the paper is organized as follows. In Section 2, we describe the cooperative location management scheme, which is based on grids and pointer forwarding strategy. In Section 3, we discuss the total cost of the cooperative location management scheme under constant velocity. Considering mobile users often travel at variable velocity, total location management cost estimation is presented based on a Gauss–Markov mobility model in Section 4. Simulation results which compare the performance of our CooLMS scheme with that of SLALoM and HGRID schemes are presented in Section 5. Section 6 concludes the paper.

2. Cooperative Location Management Scheme

In CooLMS, each node is assigned one and only one home region. All nodes in a home region act as the location server for any node in that region. This is different from schemes that select a single node in a region as the location server, in which the selected node can easily run out of battery.

In this section, we introduce CooLMS and explain its advantages over existing location management schemes.

2.1. Selection of Service Regions

We assume that mobile nodes are capable of knowing their current location, for example, using the Global Positioning System (GPS), and are equipped with radios with certain transmission range. We also assume the nodes move about in a square region of area A so as to simplify the discussion. Our scheme, just like the one reported in [5], divides the square into unit regions which are called *order*-1 squares. It then merges every $K \times K$ order-1 square to form order-2 square. Figure 1 illustrates this idea, where solid line bordered squares are order-2 squares and each order-2 square contains 3×3 order-1 squares.

Using a predefined function f, each mobile node is assigned a single home region based on its ID, which is an order-1 region. The home region is also called service region. Notice that our grid model is different from the one proposed in SLALoM, in which every node has a home region in each order-2 square, hence every node has $O(A/K^2)$ home regions.

2.2. Location Management Process

The overhead cost of a location management scheme can be divided into three parts: location update cost, location maintenance cost and location finding cost. The location update cost covers all the signaling messages that nodes send to their home servers (in home region) whenever they move to a new location. The location maintenance cost covers all the signaling messages that nodes a) send to their previous order-1 squares to inform them of their departure, b) send to their current order-1 squares to inform them of their arrival, and c) collect as they are now location servers for the nodes currently registered in their order-1 squares. The location finding cost covers all the signaling messages sent for locating a mobile node.

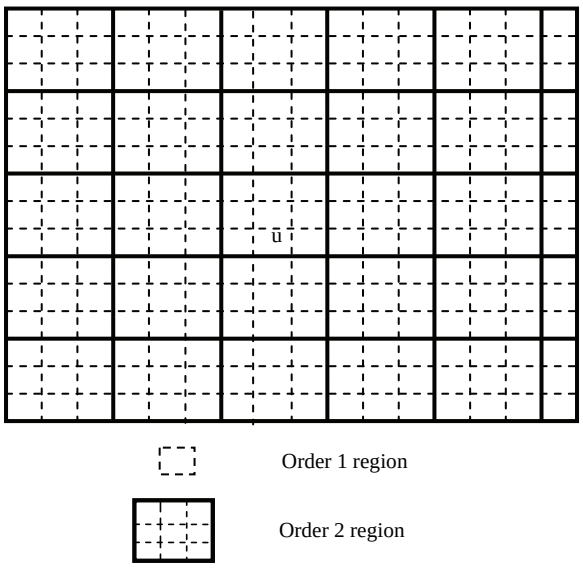


Figure 1. The location server region of a node.

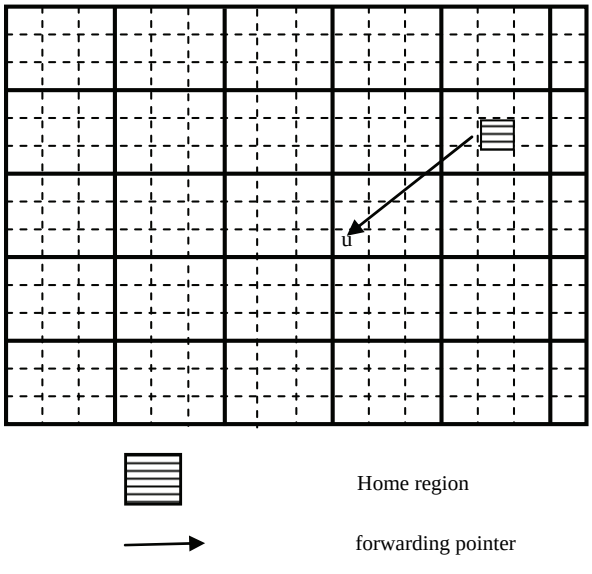


Figure 2. Home region and the forwarding pointer.

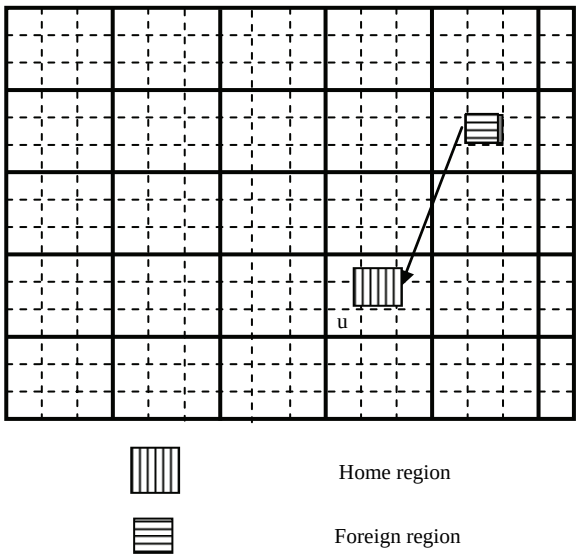


Figure 3. Home region and foreign region.

2.2.1. Assigning Home Regions

A predefined one-to-one mapping *f* is used by the scheme to map the ID of a node to a region-1 square as its home region. This mapping is known to every node, so that each node can determine the home region of any other node in constant time.

2.2.2. Maintaining Location with Forwarding Pointers

When a node *u* is in the network and located in the order-2 square of its home region, the home region knows the exact location information of node *u*. But when the node *u* moves to one of the eight sibling order-2 squares of the order-2 square of its home region, the home region may set a forwarding pointer to point to the order-1 square that node *u* arrives, as illustrated in Figure 2.

If node *u* is not in the eight order-2 squares nearby its home region, the home region authorizes the order-2 square in which node *u* resides to select an order-1 square as its foreign region, as shown in Figure 3. The home region of node *u* knows roughly (not exactly) the location information of node *u* in this situation. Just like the home region does, when the node *u* further moves to one of the eight neighboring order-2 squares of the order-2 square of its foreign region, the foreign region may set a forwarding pointer to point to the order-1 square that node *u* arrives.

2.2.3. Locating a Node

When a source node S wants to send data packets to a destination node D, S finds the home region of D using the D's ID and mapping *f*. It then sends a query packet to this home region to inquire of D's location. Four cases arise:

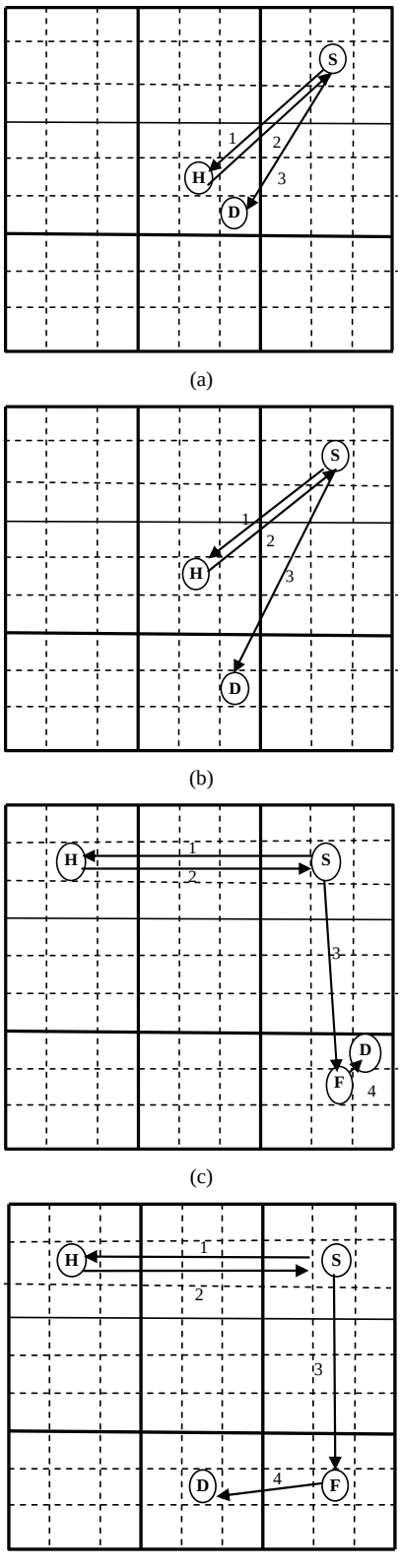


Figure 4. The location discovery process. (S: source node; D: destination node; H: home region; F: foreign region; 1, 2, 3, and 4: the order of the location discovery process)

1) If D is in the order-2 square of its home region, S requests the location information of destination node D, the home region sends directly the location information of node D to source node S, and S can send directly the data packets to destination node D using this information, as shown in Figure 4(a).

2) If D is in one of the eight neighboring order-2 squares to its home region order-2 square, S sends request to the home region of D. S can send directly the data packets to destination node D using this information, as shown in Figure 4(b)

3) If D is in the order-2 square of its foreign region, the home region sends the location information of the foreign region to source node S, and S requests D's foreign region for D's location information. After this process, the data packets from S can be retransmitted by the home region and foreign region to D, as shown in Figure 4(c).

4) If D is in one of the eight sibling order-2 squares to its foreign region order-2 square, S sends data packets to the foreign region, and the foreign region retransmit the packets by forwarding pointer to destination node D, as shown in Figure 4(d).

2.3. Forwarding Pointers

When mobile node *u* leaves the order-2 square of its home region, the home region sets a forwarding pointer pointing to the order-1 square where *u* arrives. The forwarding pointer is called *first class forwarding pointer*, and the order-1 square that node *u* arrive is called a foreign region. When *u* further moves to another sibling order-1 square as shown in Figure 5, the previous order-1 square can generate a forwarding pointer pointing to the new order-1 square where u is currently located; it doesn't need to inform its home region. This reduces the updating and paging packets cost. We call the forwarding pointer generated by order-1 squares the *second class forwarding pointer*.

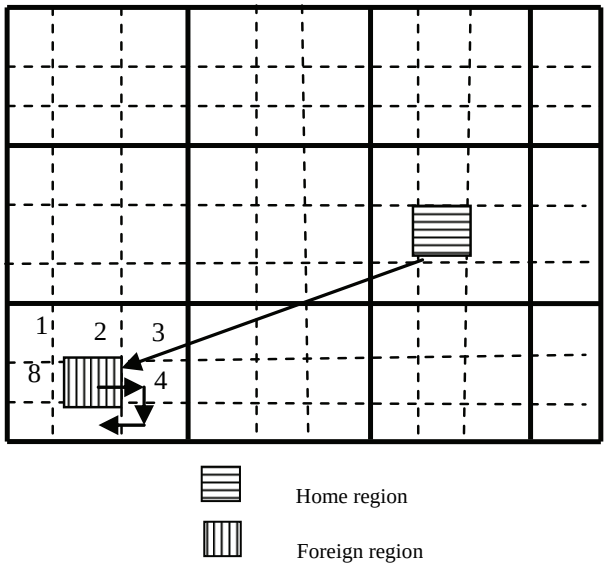


Figure 5. The process of setting up forwarding pointers.

We illustrate the process of setting up forwarding pointers in Figure 5. The square filled with horizontal strips is the home region of node *u*, while the square filled with vertical strips is its foreign region. The arrow from the home region to the foreign region is the first class pointer, while the arrows around the foreign region are the second class forwarding pointers. If *u* moves from the foreign region to the square numbered 4, a second class forwarding pointer is set to indicate the current location of *u*. Similarly, if it moves to the square numbered 5 from the one numbered 4, another second class forwarding pointer is set pointing to the square numbered 5, and so on.

We see from the process that increasing the number of forwarding pointers will reduce the location updating cost, but may also increase the paging cost. It is a trade-off phenomenon in location management. If a node moves across many region-2 square, we can further setup forwarding pointers from one foreign region to another around its home region, just like the process of setting up forwarding pointers from one order-1 region to another neighboring foreign region as shown in Figure 5. We will discuss the optimal number of forwarding pointers in Section 5.

3. The Cost of Cooperative Location Management

In this section, we discuss how to evaluate the total location management cost, which is measured in packets per second per node to transfer. We assume that nodes distribute uniformly, move randomly and independently of each other. Each node selects a direction to move, chosen uniformly between $[0, 2\pi]$. Each node selects its speed, chosen uniformly between $[v\text{-}c, v\text{+}c]$ for some time t, where the t is exponentially distributed with a mean of τ. After a mobile has traveled for time t, it selects another direction, speed and time to travel. The mobility model and the geographic routing algorithm, MFR (most forward with fixed radius routing, which forwards packets to the neighbor closest to the destination node) [12] is the same as in [5] and [6], which allows us to compare our scheme's performance with [5] and [6].

We also assume that the ad hoc network under study is in a rectangular region; all nodes in the network are equipped with Global Positioning System (GPS) that provides them with their current location; they are aware of the identities of their neighbors; and each node has a unique ID (such as an IP address). We use the following notations in this section:

A area of the total networks
N number of nodes
M the threshold of forwarding pointer number
K network area A is divided into K^2 order-1 squares
a area of a order-1 region
b broadcast cost in a region
v speed (meters per second)
u cost of sending location update message to home region
δ cost of collecting location information
γ average density of nodes in ad hoc networks
r radius of a node
z average distance for one hop of a node
λ Poisson distribution parameter of the data packets arrive at

As mentioned in Section 2, the network area is partitioned into unit regions or squares. Based on the distribution and mobility assumption made in the beginning of the section, the size of the unit region is chosen so that its average node density γ is approximately a constant. Thus, the total area of the networks $A = N/\gamma$, Woo and Singh [4] noted the following:

1) The cost of broadcasting in an order-1 square by a node, b, is proportional to the number of transmissions needed to cover the said square. The latter is in turn proportional to the area of the order-1 square divided by the area covered by a single transmission, and also is the location updating cost for just only time. Thus, $b = O(a/r^2)$.

2) The distance a node u has to cover to cross an order-1 square is proportional to the side of an order-1 square. Thus, the number of order-1 squares that a node crosses per second, is proportional to $n_1 = O(v/\sqrt{a})$.

3) Similarly, the number of order-2 squares that a node u crosses per second, can also be estimated by $n_2 = O(n_1/K) = O(v/K\sqrt{a})$ order-1 squares per second.

4) The cost of updating all severing region include foreign regions and order-1 squares estimated by $n_3 = O(n_1/M) = O(v/M\sqrt{a})$.

3.1. Cost of Location Update

When a node u moves out of the current region and enters a new region, three cases of location update may occur:

1) When the new region locates in the order-2 square of its home region, the home region knows the exact location information of node u. In this case, we only broadcast the n_1 order-1 nodes. The location update cost is $O(n_1(b))$.

2) When the node u moves to one of the other neighboring eight order-2 squares of its home region order-2 square, the home region may set a forwarding pointer to point the order-1 square that node u arrives. The location update cost is $O(n_2(b+K/z))$.

3) When the node u is not in the nine order-2 squares nearby its home region discussed above, the home region authorizes the order-2 square where the node u locates to select an order-1 square as a foreign region of the node u. The home region of node u knows roughly the location information of node u in this situation. Just like the home region does, when the node u moves to one of the other eight order-2 squares (not include the order-2 square the foreign region locates) that neighbor to the order-2 square of its foreign region, the foreign region may set a forwarding pointer to point to the order-1 square that node u arrives. The location update cost is $O(n_3(Ab/K^2+A/K))$.

So, the total location updating cost c_u is given by

$$\begin{aligned} c_u &= O(n_1(b) + n_2(b + K/z) + n_3(Ab/K^2 + A/K)) \\ &= O(\frac{v}{\sqrt{a}}\frac{a}{r^2} + \frac{v}{K\sqrt{a}}(\frac{a}{r^2} + \frac{K}{z}) + \frac{v}{M\sqrt{a}}(\frac{N}{\gamma}\frac{a}{r^2}\frac{1}{K^2} + \frac{N}{\gamma}\frac{1}{K})) \\ &= O(v + \frac{v}{K} + \frac{vN}{MK} + \frac{vN}{MK^2}) \end{aligned}$$

3.2. Cost of Maintaining Location Information

There is no such a cost in cell phone network because of the existence of base stations. However, in mobile ad hoc networks, when a node moves from region A into region B, it needs to inform nodes in region A that it has left and meanwhile, it needs to inform nodes in region B of its arrival. The nodes need to take three steps to maintain location information, namely

1) Inform the original order-1 region of its departure;
2) Inform the new order-1 region of its arrival;
3) As a location server in the new order-1 region, cache the registration information of nodes in the region.

So the cost for maintaining nodes location information c_m is

$$\begin{aligned} c_m &= O(n_1(b + b + \delta)) = O(n_1(2b + \delta)) \\ &= O(\frac{v}{\sqrt{a}}(\frac{2a}{r^2} + \delta)) = O(v) \end{aligned}$$

3.3. Cost of Finding Nodes

The cost of finding the location might be zero or a constant if either the node itself or one of its neighbors has the location information available in a cache, which happens if there are several packets destined for the same destination. In our analysis, however, we make the pessimistic assumption that the location information is not cached; therefore, the source node needs to contact the destination node's home region to find the destination's location.

The location finding process includes two situations (combined from the four cases described in Subsection 2.2.3)

1) When the source node A wants to send data packets to its destination B, node A queries the home region of node B. If node B is in its home region, then node A directly sets up a linkage with node B. From Figure 4 (a) we can deduce the cost is asymptotically to3*d*(S, H)/*z*, where *d*(S, H) is the distance between the source node S and the home region, and it is proportional to *K*. So we know that the location finding cost is proportional to *K*. Similarly, from Figure 4 (b) we can also conclude that the location finding cost is proportional to *K*;

2) If node B is not in its home region, then node A sets up a linkage through B's home region, the first and second class forwarding pointers, and then sends packets using this linkage. From Figure 4 (c) and (d), the location finding cost is asymptotically to [2*d*(S, H) + *d*(S, F)]/*z*, where the *d*(S, F)] is the distance between the source node S and the foreign region, and it is proportional to *K*. For forwarding pointer cost, we can easily get that it is proportional to *M* and λ.

So the average total cost of locating a node is

$$c_d = O((K/z) + O(\lambda M/z)$$

3.4. The Total Cost of Location Management

The total cost of location management

$$\begin{aligned} c_{total} &= N(c_u + c_m + c_d) \\ &= O(2vN + \lambda NM + KN + vN/K + vN^2/(KM) + vN^2/(K^2M)) \end{aligned} \quad (1)$$

4. Gauss-Markov Model Based Total Location Management Cost Estimation

4.1. Gauss – Markov Mobility Model

Mobile nodes often have to change speed during the course of motion. We assume that mobile nodes move at inconstant velocity and the velocity change follows a Gauss–Markov process. According to [9], the 1-D discrete version of the Gauss-Markov mobility model can be described as:

$$v_n = \alpha v_{n-1} + (1-\alpha)\mu + \sigma\sqrt{1-\alpha^2}\, w_{n-1} \quad (2)$$

where v_n is a node's mobile velocity during the *n*-th period, α is the memory level, which reflects the relationship between v_{n-1} and v_n, μ the mean of v_n, σ^2 the variance of v_n, and w_n an uncorrelated Gaussian process with zero mean, unit variance. w_n is independent of v_n.

Let $u_n = v_n - \mu$, $\beta = \sigma\sqrt{1-\alpha^2}$, the Equation (2) can be rewritten in the following simple and clear form

$$u_n = \alpha u_{n-1} + \beta w_n \quad (3)$$

4.2. Total Location Management Cost

After we have estimated each of the individual costs involved in location updating information, maintaining location information and finding the location information, we can estimate the total cost of routing packets. Assume that packets arrive at each node at a rate of λ packets per second according to a Poisson process. Then the average cost of routing N nodes can be calculated as (1). Recall the assumption that mobile motion process is Gauss–Markov. Let

$$v_n = u_n + \mu \quad (4)$$

We know from Section Ⅲ that the location updating cost $c_u = O(v) = K_1 v$, the maintaining location cost $c_m = O(v) = K_2 v$, and the finding node cost is constant K_3. So the total cost $c_{total} = (K_1 + K_2)v + K_3$. Substituting v with (4), we obtain

$$\begin{aligned} c_{total} &= (K_1 + K_2)(u_n + \mu) + K_3 \\ &= (K_1 + K_2)u_n + (K_1 + K_2)\mu + K_3 \end{aligned} \quad (5)$$

Let

$$\begin{aligned} \eta &= K_1 + K_2 \\ \kappa &= (K_1 + K_2)\mu + K_3 \end{aligned} \quad (6)$$

Then it follows from (5) and (6) that

$$c_n = \eta u_n + \kappa \quad (7)$$

From (3), we have

$$u_n = \alpha^n u_0 + (\sum_{k=0}^{n-1} \alpha^k w_{n-k})\beta$$

where the u_0 is the initial value of u_n. It results from (7) that

$$c_n = \eta[\alpha^n u_0 + (\sum_{k=0}^{n-1} \alpha^k w_{n-k})\beta] + \kappa \quad (8)$$

Consider w_n is an uncorrelated Gaussian process with zero mean, unit variance, and w_n is independent of v_n, then mean and variance of c_n are given as

$$\begin{aligned} Ec_n &= u_0\eta\alpha^n + \kappa \\ Dc_n &= \sum_{k=0}^{n-1} \alpha^{2k}\beta^2\sigma^2 = \frac{\beta^2\alpha^2(1-\alpha^{2n})}{1-\alpha^2} \end{aligned}$$

When the memory level $0 < \alpha < 1$, we have

$$\lim_{n \to +\infty} Ec_n = \kappa$$

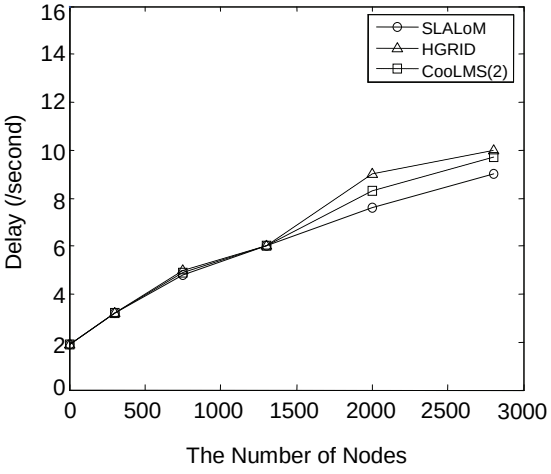
$$\lim_{n \to +\infty} Dc_n = \frac{\beta^2 \alpha^2}{1 - \alpha^2}$$

We give approximate estimation of the total cost of location management in this section, and we will give some simulation results in the next section.

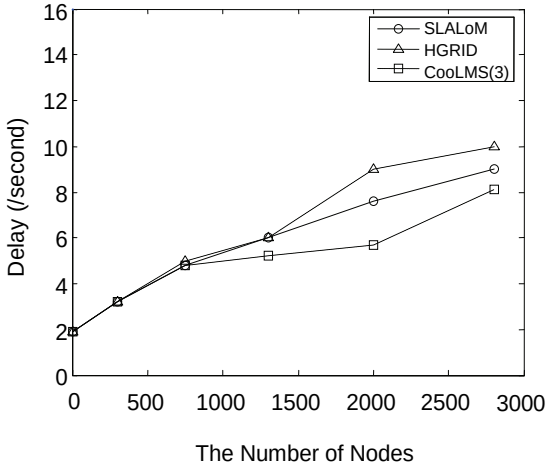
5. Simulation Results

In this section, we present the simulation results for CooLMS, SLALoM [5] and HGRID [6] in network time delay and total location management cost, using OPNET technology [18].

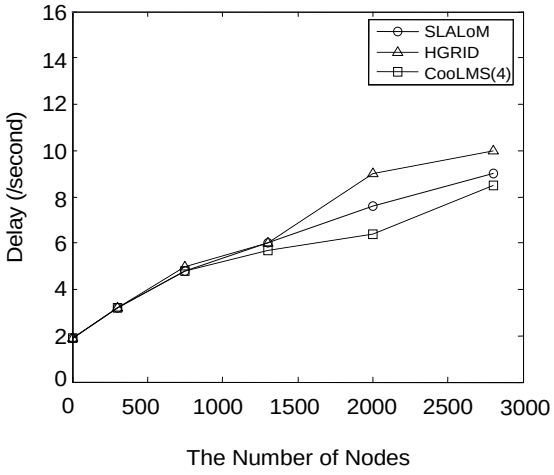
To compare the performance of these three schemes, a location management layer was built in to the TCP/IP protocol stack that is operated in conjunction with IP as the network layer protocol. Main data structures [6] in the location management layer consist of a *location table* and a *neighbor table*. When a location server node receives a location update packet from a node, the current location of that node is updated in the location table.



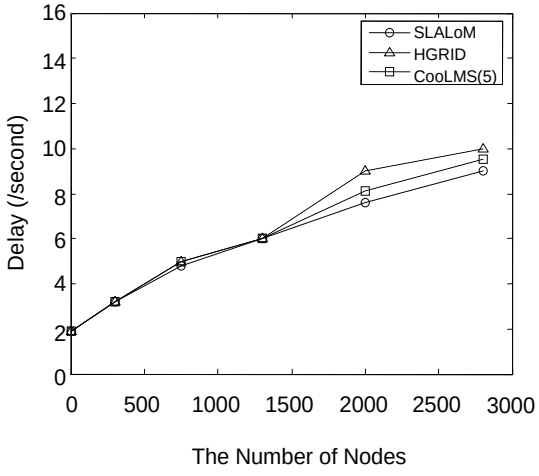
(a) M=2 for time delay



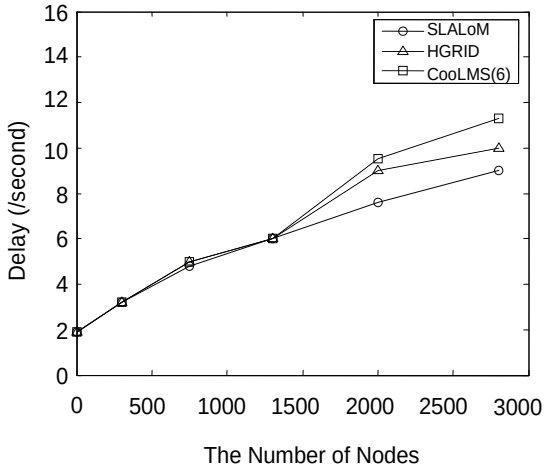
(b) M=3 for time delay



(c) M=4 for time delay



(d) M=5 for time delay

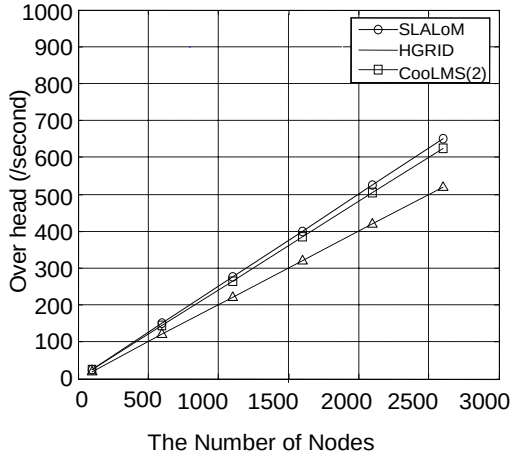


(e) M=6 for time delay

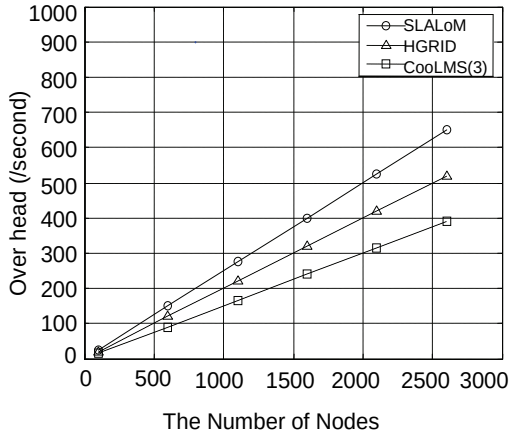
Figure 6. Network time delay.

MFR [12] without backward progression, in which packets are dropped if no forward progress can be made, was implemented as the geographic routing algorithm.

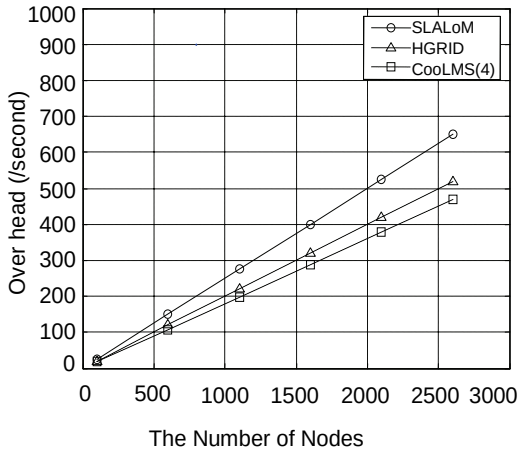
We assume that nodes distribute uniformly, move randomly and independently of each other. Each node selects a direction to move, chosen uniformly between [0, 2π].



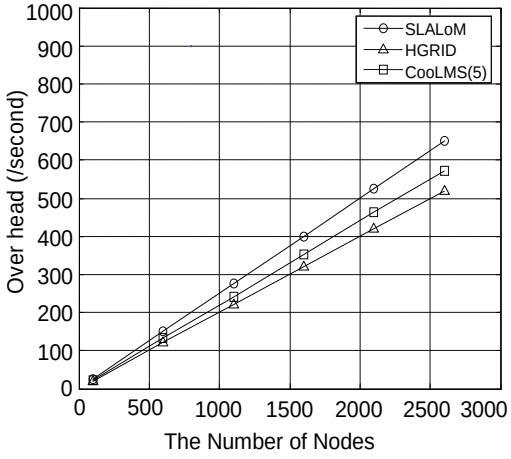
(a) M=2 for overhead



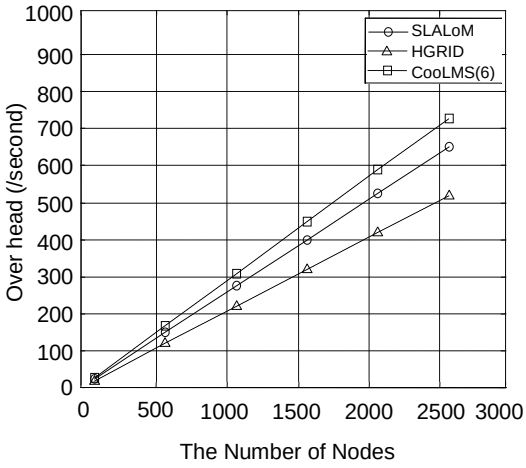
(b) M=3 for overhead



(c) M=4 for overhead



(d) M=5 for overhead



(e) M=6 for overhead

Figure 7. Location management overhead.

Each node selects its speed, chosen uniformly between [0, 10*m*/*s*] for some time *t*, where *t* is exponentially distributed with mean τ. After a mobile has traveled for certain time *t*, it selects another direction, speed and time to travel. The topology consists of from 50 to 3000 nodes. When the number of nodes increase, the areanetwork time delay and total location management cost. As shown in Figure 6, there is a tradeoff between paging, updating cost and network time delay. We can balance the tradeoff by selecting the threshold of forwarding pointer number. When threshold of forwarding pointer number M = 3 and 4, the network time delay with CooLMS is lower than the networks with SLALoM and HGRID; when threshold of forwarding pointer number M = 2 and 5, the network time delay with CooLMS is between the networks with SLALoM and HGRID; When threshold of forwarding pointer number M = 6, the network time delay with CooLManet is higher than the networks with SLALoM and HGRID.

Figure 7 shows that when the threshold of forwarding pointer number M =3 and 4, the total location manage-

ment cost with CooLManet is lower than the networks with SLALoM and HGRID; When the threshold of forwarding pointer number M = 2 and 5, the total location management cost with CooLMS is between the networks with SLALoM and HGRID; When the threshold of forwarding pointer number M = 6, the total location management cost with CooLManet is higher than the networks with SLALoM and HGRID.

We conclude from the simulation that using suitable number of forwarding pointers including the first and second pointers, the network time delay and location management cost can be reduced significantly.

6. Conclusions

We proposed a cooperative location management scheme, CooLMS, for mobile ad hoc networks. CooLMS combines the strength of grid based location management and cooperative location management to achieve low signaling cost as well as high scalability. We also discussed the total cost estimation of mobile location management for ad hoc mobile networks with missing measurements. Simulation results indicate that the network time delay and location management cost can be reduced significantly by using suitable number of forwarding pointers. Practically, the CoolMS scheme is good for networks used in metropolis areas where mobile users typically travel across several location areas from home to work on a daily basis.

Our future work includes location management for special routing protocols and location management for QoS of mobile decision support in ad hoc mobile network.

7. Acknowledgment

This work is partially supported by NSFC granted number 60874113 and 70271001; China Postdoctoral Fund granted number 2002032191; Shanghai Fund of Science and Technology granted number 00JG05047; Shanghai Key Scientific Research Project under grant number 05dz05036; and 2008 Fund of Engineering Research Centre of Digitized Textile & Fashion Technology, Ministry of Education.

8. References

[1] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," Proceedings of 4th Annual Conference on Mobile Computing and Networking, Dallas, TX, October 25–30, 1998.

[2] T.-W. Chen and M. Gerla, "Global state routing: A new routing scheme for ad-hoc wireless networks," Proceedings of IEEE International Communications Conference, 1998.

[3] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," Proceedings of 4th Annual Conference on Mobile Computing and Networking, Dallas, TX, October 25–30, 1998.

[4] S. M. Woo and S. Singh, "Scalable routing protocol for ad hoc networks," Wireless Networks, Vol. 7, pp. 513–529, 2001.

[5] C. Cheng, H. Lemberg, S. Philip, E. van den Berg, and Tao Zhang, "SLALoM: A scalable location management scheme for large mobile ad-hoc networks," Proceedings of IEEE Wireless Communications and Networking Conference, Vol. 2, pp. 574–578, March 2002.

[6] S. J. Philip, J. Ghosh, and C. M. Qiao, "Performance evaluation of a multilevel hierarchical location management protocol for ad hoc networks," Computer Communications, Vol. 28, pp. 1110–1122, 2005.

[7] S. M. S. Masajedian and H. Khoshbin, "Cooperative location management method in next generation cellular networks," Proceedings of the Ninth International Symposium on Computers and Communications, pp. 525–530 2004.

[8] J. Zhou, B. Tang, and D. Li, "Partition digraph for location area management in mobile computing environment," International Journal of Nonlinear Sciences and Numerical Simulation, Vol. 5, No. 4, pp. 393–396, 2004.

[9] B. Liang, and Z. J. Haas, "Predictive distance-based mobility management for multidimensional PCS networks," IEEE/ACM Transactions on Networking, Vol. 11, No. 5, pp. 718–732, 2003

[10] J. S. M. Ho and I. F. Akyildiz, "Mobile user location update and paging under delay constraints," ACM-Baltzer J. Wireless Networks, Vol. 1, pp. 413–425, December 1995.

[11] Y.-B. Lin, "Reducing location update cost in a PCS network," IEEE/ACM Trans. Networking, Vol. 5, pp. 25–33, February 1997.

[12] T.-C. Hou and V. O. K. Li, "Transmission range control in multihop packet radio networks," IEEE Transactions on Communications, Vol. COM-34, No. 1, pp. 38–44 1986.

[13] K. Sue and C. Tseng, "One-step pointer forwarding strategy for location tracking in distributed HLR environment," IEEE Journal on Selected Areas in Communications, Vol. 15, No. 8, pp. 1455–1466, 1997

[14] Y. Lin and W. Tsai, "Location tracking with distributed HLR's and pointer forwarding," IEEE Transaction on Vehicular Technology, Vol. 47, No. 1, pp. 58–64, 1998

[15] W. Ma and Y. Fang, "Two-level pointer forwarding strategy for location management in PCS networks," IEEE Transaction on Mobile Computing, Vol. 1, No. 1, pp. 32–45, 2002

[16] C.-M. Weng and C.-H. Chu, "K-step pointer forwarding strategy for location tracking in distributed HLR environment," IEE Proceedings of Communications, Vol. 150, No. 3, pp. 207–213, June 2003.

[17] W. Ma and Y. Fang, "An efficient mobility management scheme based on location anchoring and pointer forwarding," IEEE 58th Vehicular Technology Conference, VTC'03-Fal1, Vol. 4, pp. 2764–2768,6-9 October 2003.

[18] http://www.opnet.com/solutions/network_rd/modeler.html.

Int. J. Communications, Network and System Sciences, **2009, 2, 742-745**
doi:10.4236/ijcns.2009.28085 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Perceptual Approach to Reduce Musical Noise Using Critical Bands Tonality Coefficients and Masking Thresholds

Ch. V. Rama Rao[1], M. B. Rama Murthy[2], K. Srinivasa Rao[3]
[1]*Department of ECE, Gudlavalleru Engineering College, Gudlavalleru, India*
[2]*Jayaprakash Narayan College of Engineering, Dharmapur, Mahabubnagar, India*
[3]*TRR College of Engineering, Pathancheru, India*
Email: *chvramaraogec@gmail.com*
Received August 18, 2009; *revised September* 27, 2009; *accepted October* 19, 2009

Abstract

Traditional noise reduction techniques have the drawback of generating an annoying musical noise. A new scheme for speech enhancement in high noise environment is developed by considering human auditory system masking characteristics. The new scheme considers the masking threshold of both noisy speech and the denoised one, to detect musical noise components. To make them inaudible, they are set under the noise masking threshold. The improved signal is subjected to extensive subjective and objective tests. It is observed that the musical noise is appreciably reduced even at very low signal to noise ratios.

Keywords: Noise Reduction, Musical Noise, Masking Threshold

1. Introduction

In many speech communication systems, enhancing the corrupted speech is a challenging task especially at high noise level. A large number of noise reduction techniques have been proposed in the past. They are based on spectral subtraction [1] and Wiener filtering [2] techniques. The main drawback of these methods is the appearance of an annoying residual noise, often referred to as musical noise. Later techniques developed rely on psychoacoustical considerations. Mainly they exploit the masking properties of the human auditory system. For example according to the enhancement scheme proposed in [3], only audible noise components are estimated and suppressed. Other approaches introduce a perceptual modification on traditional denoising systems [4,5].

In the present paper a new speech enhancement technique is developed for reducing the musical noise. In this work, the auditory masking threshold is estimated for musical noise detection and reduction. Musical noise is detected based on fact that musical noise components present in the enhanced signal lie above the noise masking threshold. On the other hand, the frequency components of noisy speech lie below the noise masking threshold. Hence, by using some comparison rules musical noise is detected. The detected musical noise components are set under the noise masking threshold and their closet neighbours are smoothed resulting in musical noise reduction.

2. Basic Speech Enhancement System

Let the corrupted speech signal $y(n)$ be represented as

$$y(n) = s(n) + d(n) \quad (1)$$

where $s(n)$ is the clean speech signal and $d(n)$ is the noise signal. The processing is done on a frame-by-frame basis. The Short Time Fourier transform (STFT) is used and the previous model is re-written as

$$Y(m, f) = S(m, f) + D(m, f) \quad (2)$$

where m indicates the frame index and f is the frequency index. The denoised speech short time magnitude $|S(m, f)|$ is obtained using a spectral denoising approach. In this paper, modified Wiener filter [6] is used to denoise the speech signal. The denoised speech is obtained as follows

$$|S(m,f)| = W(m,f).|Y(m,f)| \quad (3)$$

where $W(m,f)$ is the modified Wiener filter gain [6], obtained by including the cross correlation between clean speech signal and noise signal. $W(m,f)$ is given by

$$W(m,f) = \frac{\xi(m,f) + \frac{\delta.Y(m,f)D(m,f)}{E\{D^2(m,f)\}}}{\xi(m,f) + 1 + 2\frac{\delta.Y(m,f)D(m,f)}{E\{D^2(m,f)\}}} \quad (4)$$

where $\xi(m,f) = E\{S^2(m,f)\}/E\{D^2(m,f)\}$ is apriori signal to noise ratio (SNR). $\xi(m,f)$ is calculated according to the decision directed approach reported in [7]. δ is the cross correlation coefficient for estimating the correlation between the noisy speech and noise signal in a frame [6]. The modified wiener filter gain function is not only controlled by $\xi(m,f)$ as for conventional Wiener filter but also by δ. When δ is zero noise and clean speech signals are uncorrelated and $W(m,f)$ is reduced to conventional Wiener filter gain function. The proposed approach consists on reducing musical noise existing in denoised speech signal spectrum denoted by $|\hat{s}(m,f)|^2$. The temporal domain enhanced speech is obtained with the following relationship

$$\hat{s}(n) = IFFT\left[|\hat{S}(m,f)|.e^{j\arg(Y(m,f))}\right] \quad (5)$$

3. Proposed Enhancement Technique

The proposed enhancement technique consists of different steps described below.

- Modified Wiener filter gain function is applied to get denoised speech.
- The noise masking threshold NMT is calculated for both noisy speech and denoised one.
- A musical noise detector is used. For each frequency, it gives a Boolean flag M which indicates the presence or absence of musical noise.
- The musical noise is reduced when present.

3.1. Musical Noise Detection

In order to detect musical noise in denoised speech, perceptual properties of human auditory system are used. There are two steps in detecting musical noise: calculation of noise masking threshold, detection of tonal components in both noisy speech and denoised speech.

3.1.1. Noise Masking Threshold Calculation

The NMT is obtained through modelling the frequency selectivity of the human ear and its masking property. By using masking threshold we distinguish "tone masking noise" and "noise masking tone". In our context of musical noise detection, we consider only the situation of "noise masking tone". In fact, the musical noise is a tone signal which is audible during noise components. The NMT is calculated according to principle explained in [8].

3.1.2. Tonal Components Detection

Tonal and non tonal components are identified because their masking models are different. The power spectrum and noise masking threshold of both noisy speech and denoised speech are calculated. Components above noise masking threshold in noisy speech are treated as tonal and belong to speech components. Components above noise masking threshold in denoised speech are marked as tonal and belong to either speech components or musical noise components. Hence, musical noise components can be detected and they are the marked tonal components appearing in denoised speech and not appearing in noisy speech. Figure 1 shows locations of

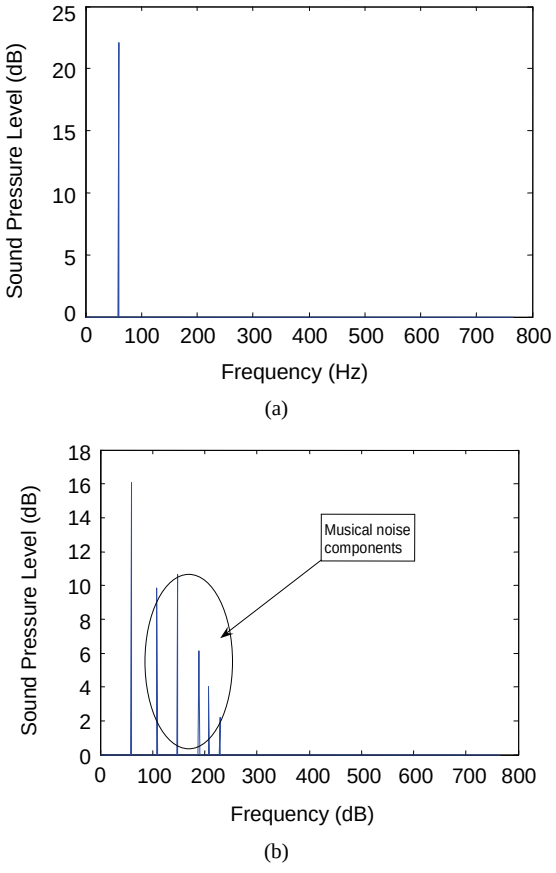


Figure 1. Location of tones in (a) noisy speech (b) denoised speech.

tones in noisy speech and denoised speech. In this work to identify the musical noise tonality coefficient is used. The tonality coefficient α_d is computed for each critical band of denoised speech and α_n for the noisy speech. Musical noise appears in any i^{th} critical band if α_d^i is greater than α_n^i. It becomes audible if the difference $\Delta\alpha_i = \alpha_d^i - \alpha_n^i$ is greater than a certain predetermined threshold τ_i. The threshold τ_i of the i^{th} band depends on critical band order and masking properties of human ear. We are interested in the audibility of tones in the presence of narrow-band noise. A narrow-band noise having 1 bark bandwidth can mask a tone within the same critical band if intensity is below the noise masking threshold NMT_i where the NMT is calculated as follows [5,8]

$$NMT_i = E_i - 5.5 \tag{6}$$

3.1.3. The Experimental Determination of τ_i

The experimental procedure to determine τ_i is as follows:

A white Gaussian noise is considered and power spectrum of each frame is subdivided in critical bands. For each critical band, its energy E_i and its tonality coefficient α_i are computed. For the i^{th} critical band, the power P_i of an additive audible tone which is equal to the noise masking threshold $P_i = NMT_i$ is computed. A sinusoid of the power P_i is injected in the center of the i^{th} critical band and tonality coefficient α_i' is computed. The difference $\alpha_i' - \alpha_i$ represents the threshold τ_i over which an additive tone becomes audible in the presence of narrow-band noise. Experimentally it is observed that τ_i is quite constant for all critical bands and is about τ_i=0.06. Hence in present work τ_i equal to 0.06 is used. Finally, a Boolean flag M, indicating musical noise presence in any critical band is computed using

$$M_i = \begin{cases} 1 & if\ \alpha_d - \alpha_n \geq 0.06 \\ 0 & otherwise \end{cases} \tag{7}$$

3.2. Musical Noise Reduction

Musical noise reduction is to remove only the parts responsible of the musical noise character by shifting down the power spectrum of detected musical components under the denoised speech noise masking threshold. In this work correction term, $C(f)$ is used to shift down sufficiently the power spectrum. The estimated power spectrum of corrected speech is written as

$$\left|\hat{S}(m,f)\right|^2 = \begin{cases} NMT_{\hat{s}}(m,f) - C(f) & if\ M(m,f) = 1 \\ \left|\hat{S}(m,f)\right|^2 & otherwise \end{cases} \tag{8}$$

where $C(f)$ the correction term is chosen according to subjective listening tests. Values of $C(f)$ for speech and pause frames as given by *Sofia Ben Jebara* [9] indicated in Table 1 are used in the present work. It is observed that the attenuation is small for low frequency and is considerable for high frequency components. During pause, it is constant since distortion and musical component appear in the same way in all frequency bands.

Table1. Correction constants for musical noise reduction.

	Frequency band(KHz)	[0,1]	[1,2]	[2,3]	[3,4]
Speech	C(f)	0.5	2	5	10
Pause	C(f)	10	10	10	10

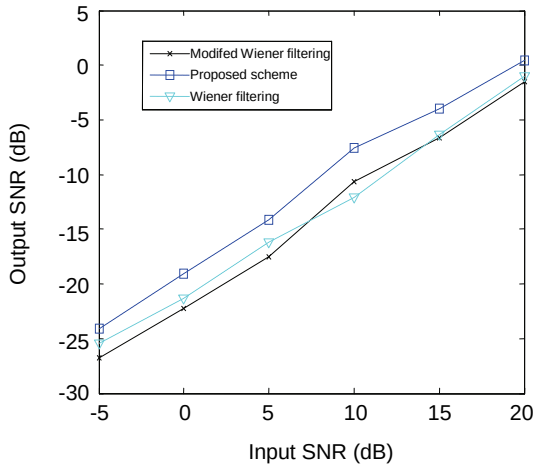


Figure 2. Output segmental SNR values.

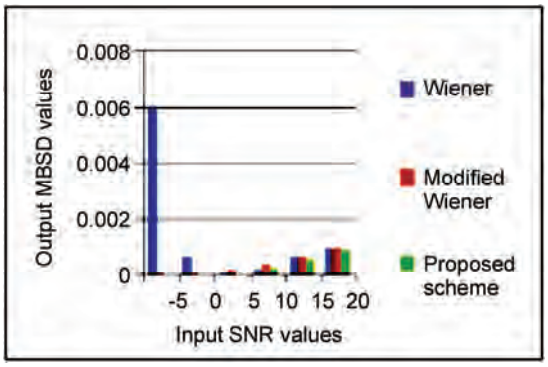


Figure 3. Output MBSD values

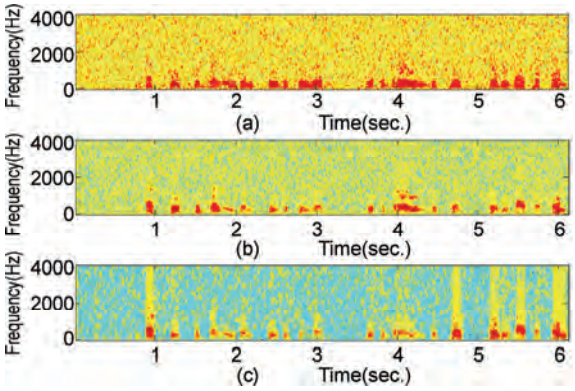


Figure 4. Spectrograms of (a). Noisy speech (b). Denoised speech (c). Enhanced speech by proposed scheme.

4. Results and Discussions

The proposed technique is evaluated using temporal, spectral and perceptual criteria. Segmental signal to noise ratio (SNR_{SEG}) is used as quantitative temporal criteria. For spectral criteria, spectrograms are used and the Modified Bark spectral Distance (MBSD) is used as perceptual criteria [10].

In our simulations, recorded speech samples are used and corrupted with white Gaussian noise and simulations are performed on MATLAB platform. Figure 2 and Figure 3 shows the comparison of performance results of the classical Wiener filtering, modified Wiener filtering and the proposed scheme for different values of signal to noise ratio in terms of SNR_{SEG} and MBSD values respectively. Figure 4 shows spectrogram plots.

Interpretations from the Figures 2, 3 and 4 are as follows:

- The proposed scheme leads to better performance in terms of quality and intelligibility speech signal for all criteria and also it is well noticeable for spectral and perceptual criteria which have good correlation with listening tests.
- Spectrograms are considered in Figure 4. The noisy speech signal is a speech corruptedby a white Gaussian noise whose SNR=10 dB. The denoised speech signal by a modified Wiener filtering is affected by a musical noise (isolated points randomly distributed in time and frequency). The amount of such noise is reduced by the proposed scheme.

5. Conclusions

In this work, a new enhancement scheme for reducing musical noise imposed by a modified Wiener filtering is proposed. The masking characteristics of the human ear are used to detect and to reduce musical noise. Simulation results show that this scheme provides better results in terms of temporal, spectral and perceptual criteria.

6. References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transaction Acoustics, Speech and Signal Processing, Vol. ASSP-27, No. 2, pp. 113–120, April 1979.

[2] Y. Ephraim and D. Mallah, "Spectral enhancement using optimal non-linear spectral amplitude estimation," on Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 1118–1121, 1983.

[3] A. Akbari-Azrani, R. Le bouquin Jannes, and G. Faucon, "Optimizing speech enhancement by exploiting masking properties of the human ear," on Proceedings International Conferences on Acoustic, Speech and Signal Processing ICASSP, IEEE, pp. 800–803, 1995.

[4] N. Virag, "Single channel speech enhancement based on masking properties of human auditory system," IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 2, pp. 126–137, February 1999.

[5] K. A. Sheela, CH. V. R. Rao, K. S. Prasad, and A. V. N. Tilak, "A new noise reduction pre-processor for mobile voice communication using perceptually weighted spectral subtraction method," 3rd International Conferences on Mobile Ubiquitous and Pervasive Computing, VIT University, 16-19 December 2006.

[6] CH. V. R. Rao, M. B. R. Murthy, and K. S. Rao, "Speech enhancement using modified Wiener filter," National Conference on Futuristic Advancements in Computing & Electronics, Deccan College of Engineering & Technology, 19-21 March, 2009.

[7] Y. Ephraim and D. Mallah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," IEEE Transaction on Speech Audio Processing, Vol. ASSP-32, pp. 1109–1121, 1984.

[8] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," IEEE, Journal on Selected Areas of Communication, Vol. 6, pp. 314–323, 1988.

[9] S. B. Jebara, "A perceptual approach to reduce musical noise phenomenon with wiener denoising technique," proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2006.

[10] W. Yan, M. Dixon, and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold," on Proceedings of the Speech Coding Workshop IEEE, pp. 55–56, 1997.

Int. J. Communications, Network and System Sciences, **2009, 2, 746-753**
doi:10.4236/ijcns.2009.28086 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Real-Time Measurement Algorithm for Available Bandwidth

Yi YIN, Weidong WU
School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
Email: yinyi1023@sina.com, wwdtylwt@163.com
Received March 26, 2009; *revised July* 10, 2009; *accepted August* 21, 2009

Abstract

Available bandwidth estimation is useful for route selection in overlay networks, QoS, and traffic engineering. Many measurement algorithms, such as Pathload, Pathchar, and Packet Transmission Rate (PTR) method, etc. have been proposed. PTR method sends a sequence of packet trains to characterize the interaction between probing packets and the competing traffic, and uses the average rate of the packet train as an estimate of the available bandwidth. However, this PTR algorithm does not fully consider the situation that the detection packets lost themselves. This paper improves the original PTR algorithm which considers the specialty of the burst of the network background flow. The improved PTR algorithm uses the method to match the initial gap value and gap step value to solve the problem about the burst of background flow, and the improved PTR algorithm record and control the number of packets with source and destination to solve the lost of some packets. Finally, theory and experiments, verified by the improved algorithm of PTR, can reflect the changes of the network stably and timely under the circumstance of the network fluctuates frequently. It improves the accuracy of a network measurement and makes the measurement results, which can reflect the changes of the network more clearly.

Keywords: Available Network, Background Flow, Detection Packet Pair

1. Introduction

Network measurement is the network application which analyzes the available bandwidth with some kind of technique for Internet. It acquires the availability of bandwidth through the analysis of the data and then obtains the state of the network [1]. Accuracy, precision and timeliness with network measurement will directly impact on network routing, quality of service, network load, volatility, and other issues. Therefore, choosing a good network measurement algorithm can not only measure the result more accurately, but also reflect more changes of the available network bandwidth.

With network measurement, the probe rate model (PRM) is based on the concept of the self-induced congestion [2]. The principle of PRM is to send a series of detection flow which has differently sent rate R_0 from the source to the destination host. The destination detects the rate of flow R_m. The PRM uses the data between R_m and R_0 with available bandwidth to measure the network. This measurement is based on the concept of available bandwidth by lead congestion. Tentative available bandwidth as A, the measurement method of the PRM principle is:

1) if $R_0 < A$, $R_m = R_0$

2) if $R_0 > A$, $R_m < R_0$

where the value of R_0 grows from small to large.

When the rate of destination size changes between R_m and R_0 by detection flow, it means the available bandwidth has been depleted, and now R_0 is the available bandwidth.

The PTR (packet transmission rate) algorithm is based on the PRM principle [3,4].

This paper attempts to consider the burst of background flow and the instability of the measurement packets to improve the PTR algorithm effectively, and uses the network measurement tool to validate that the improved PTR algorithm has smaller volatility and higher measurement accuracy than the original algorithm, and the measurement results are smooth and close to the

theoretical value.

2. PTR Algorithm

PTR algorithm is to measure available bandwidth of network based on PRM principle. Its function is that detection packets send it out from the source to the destination with time interval from small to large. When the packets are received at the destination, the packet rate is calculated and PTR algorithm is used to analyze the background flow and measure the bottleneck link, and then regard it as the available bandwidth to measure the condition of network [4]. Here, the available bandwidth means the fact that the end-to-end path of the n links' minimum value in some time respective to pass by the maximum amount of the useful data [5,6]; bottleneck link means the fact that the end-to-end path of the n links' minimum value in some time respective passing by the non-maximum amount of the useful data [7]; and the background flow means the fact that the end-to-end path of the n links' value in some time respective to pass by the non-useful data.

Discussed PTR algorithm in the bottleneck link, assumption the background flow is transmitted in adjacent links. In order to detect the bandwidth available, the source sends more detection packets continuously and the destination will measure the time interval of these packets after the packets thread the router. Considering that the source sends n detection packets continuously, two back-to-back packets p_i and p_{i+1} $(1 \le i \le n-1)$ through the router from the end of the queue arrives at the other end. As shown in Figure 1.

In Figure 1, there are two detection packets and the background flow through the router from the end of the queue arriving at the other end and then recombining. After these time interval of detection packets thread the router and recombine, the background flow and router include the time interval is different before the packet thread the router [8].

The PTR algorithm needs theoretical analysis from some variable, and has been received by the detection exploration.

Assumption g_I (the initial gap) is the detection packet pair [9] which has the initial time interval at the source; g_B (the bottleneck gap) is the detection packet pair time length on the output link; g_O (the output gap) is the detection packet pair which has the time interval at the destination; g_i is the detection packet pair i which has the time interval at the destination in the background flow; $\sum_{i=1}^{M} g_i^+$ is the time interval of a group had increased with packets pair M in packets n; $\sum_{i=1}^{K} g_i^=$ is the time interval of a group had equaled with packets pair K in packets n; $\sum_{i=1}^{N} g_i^-$ is the time interval of a group had decreased with packets pair N in packets n; B_O is the total bandwidth of the link; B_C is the background flow of the link; s is the detection packet size.

By describing PTR algorithm, the situation will make the following general regulation: detection packets pair has queued and through router, and detection packets has not existed deadlock in transmission network.

It is assumed that the background flow is stable at the network. Background flow occupies a little network bandwidth and has a little fluctuation; and the detection packets have not been lost in the process of transmission. Now as premise to analysis the algorithm [10].

When the background flow occupies the bandwidth constantly, according to the necessary measurement data and known data at PRM principles can be obtained by the packet rate which is equivalent to the ratio that the length of packet and the time of packet arrived destination in the background flow at one time. That is:

$$R_m = \frac{s}{g_O - g_B + g_I} \tag{1}$$

Here g_O means the time interval with the head of packets arrive at the destination and then all of the packets have been passed.

And now,

$$g_O = g_B + \frac{B_C}{B_O} \cdot g_I \tag{2}$$

Here $\frac{B_C}{B_O} \cdot g_I$ means the latency that the bandwidth has been occupied by the background flow that make the detection packets have not arrived at the destination on time.

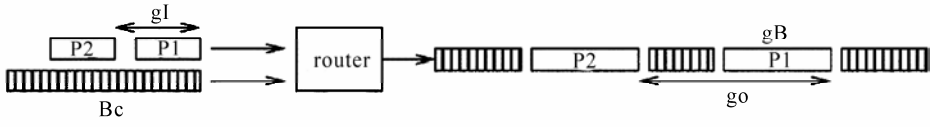


Figure 1. Detection packets and background flow competition through route.

However, the background flow based on a constant of bandwidth is only a basic situation that Equation (1) cannot be used for the actual network. In general, background flow of bandwidth is always changing. The variable bandwidth occupied by the background flow and the rate of the detection packets arrival terminal is:

$$R_m = \frac{(M+K+N)\cdot s}{\sum_{i=1}^{M} g_i^{+} + \sum_{i=1}^{K} g_i^{=} + \sum_{i=1}^{N} g_i^{-}} \quad (3)$$

Equation (3) as PTR equation, and PTR algorithm is used by this equation.

The significance of this equation is the ratio that sending a total length of several detection packets and receiving all time of the detection packet at the destination of the link. According to the PRM principle, the source rate will gain the network available bandwidth information timely. The background flow, which has occupied the bandwidth variably in the network, is the application of measurement in the real situation.

Evidently, PTR algorithm is built on the PRM principle, which can be used by the changes of the background flow and detected the packet rate.

3. Improved Algorithm

In the actual network measurement, either volatility or loss detection packet has not been avoided. But the PTR algorithm in measurement has not considered these two issues. Therefore, it must remove these assumptions and discuss verification in the actual network.

The original PTR algorithm, based on the PRM model, has higher accuracy and faster speed of convergence with the background flow relatively constant and the utilization of the link is not high and the influence is small for networks. As mentioned earlier that this is the two assumptions of a description. When the flow is small or stable with the network background flow traffic, detection packet by the volatility of the time interval can be more objective response network conditions. However, when the network environment is poor, like the background flow changes on the network path, it is larger or the utilization of the bottleneck link is higher, as background flow volatility is more obviously, the existing PTR algorithm sends detection packets with the time interval from small to large, which have some false measurement value, and then the measurement results significant ups and downs, which means the measurement is not stable. As the detection packet is instable, it cannot be obtained with accurate value when the network condition has large fluctuation.

3.1. Theoretical Exploration with the Improved Algorithm

Through Equation (3), the detection packet rate is the ratio by two summations. Since the objective over PTR algorithm reflects the changes in the bottleneck rather than accurately to measure the timely rate with each detection packet, it can be arranged the PTR algorithm further.

Assumption n detection packets will be divided into l groups once a time, each group has k detection values. Firstly, calculate average send gap ($avg\_sed\_gap(i)$) and average receive gap ($avg\_rec\_gap(i)$) with the packet sequence i, ($i \in (0,k)$). Secondly, calculate the summation of average send gap and average receive gap with the group of packet sequence l. The significance by Equation (3) could be re-described

$$R_m = \frac{s \times l}{\sum_{i=1}^{l} avg\_rec\_gap(i)} \quad (4)$$

Apart from PTR algorithm over two data with g_i and s, that needs to introduce two new values.

From the Equation (4), other than the interval time g_i and the size s from the known detection packet with the background flow at the source in Equation (3), it introduces two new values, that are the average send gap($avg\_sed\_gap(i)$) and the average receive gap($avg\_rec\_gap(i)$).These two values mean the weighted average with the time interval which is the source and destination from the detection packets. As the PTR algorithm does not measure the timely rate accurately by detection packets, it only needs to measure the changes of the network truly. It makes every time interval values not accurate reflected, thus it just sum those average value which could direct to reflect some changes.

This methods estimate detection rates used to sample mean can be reduced the consideration of the detection packets that is not necessary to consider a small quantity of the detection packets lost.

3.2. Improvement of the Algorithm in the Circumstance of Frequent Fluctuations in the Background Flow

Reference Equation (4) can improve the PTR algorithm appropriately. Assumptions the detection packets have been sent to the initial time interval *init_gap* and transmitted to the time interval *gap_step*. *init_gap* and *gap_step* will be set the fixed value but not as described in the PTR algorithm from small to large to send the detection packet with time interval [11].

Write algorithm for these conditions.

Algorithm PTR:

{


```
probe_num=PROBENUM;
   packet_size=PACKETSIZE;
   gB=GET_GB();
   init_gap=gB/2;
   gap_step=gB/8;
   dst_gap_sum=0;

for(packet_size=PACKETSIZE;packet_size!=0;packet
_size--)
    {
SEND_PROBING_PACKETS(probe_num;packet_size);
        inc_gap_sum=GET_INCREASED_GAPS();
        dst_gap_sum=GET_DST_GAPS();
     }
   c_bw=gB*inc_gap_sum/dst_gap_sum;
   a_bw=b_bw-c_bw;
}
```

There are two fixed value in algorithm, that is, init_gap=gB/2, gap_step=gB/8. These two values are constant. When the gap values are constant, no matter how changes in the network background flow, the bandwidth is always constant over the detection packet. Therefore, it is easy to be consistent with measurements of the fluctuations and change of the background flow.

3.3. Improvement of the Algorithm in the Circumstance of Desert with the Detection Packets

If the original PTR algorithm encounters the high-intensity in network background flow, as the total bandwidth is always limited, even if the gap value is constant and exploration data are nice match for the background flow, detection packets may not arrival at the destination but lost, thereby affect the analysis of the data. Analysis the PTR equation, by the Equation (4) can be seen, if detection packet is lost, it will affect the convergence conditions of the summation, thus affect the accuracy of the detection data.

Now it describes the flow chart with improved PTR algorithm, and increases the data with recording and controlling in source and destination. At the same time, it takes advantage of an improved approach to matching the transmitter and receiver data, does accurate records, and finally analyzes the network (Figure 2).

Source:

1) Take some group by detection packets from the host, each group have k packets, recorded as P_r, where $r \in [0,k]$, deliver to send cache

2) Saving a constant time t_s in send cache,

3) Deliver detection packet and t_s from send cache, then go 1)

Destination:

1) Waiting, record wait time t_w at the same time

2) Receive detection packet P_r from the source and send it to receive cache, and intercalate a variable R to record the detection packets number from the source, the initial value of R is 0

3) Waiting time t_w and detecting packet will be handed over the host, and endue a new variable t_n from t_w

4) Go (1), then $t_w \leftarrow 0$, $R++$

5) Compare t_n with t_s, and define three value $M=0$, $K=0$, $N=0$. If $t_n > t_s$, $M++$; if $t_n = t_s$, $K++$; if $t_n < t_s$, $N++$

6) When the compare is complete, $t_n \leftarrow 0$

It can be seen from the algorithm description to provide a very important parameter t_w. With the value of t_w, it can be distinguished in the detection packets which have been sent out to the destination, and the source is arrived or lost due to the net reasons. If time is over, the source without retransmission and the destination can also automatically be received and recorded in the number of packets that insure match for the packets at each side.

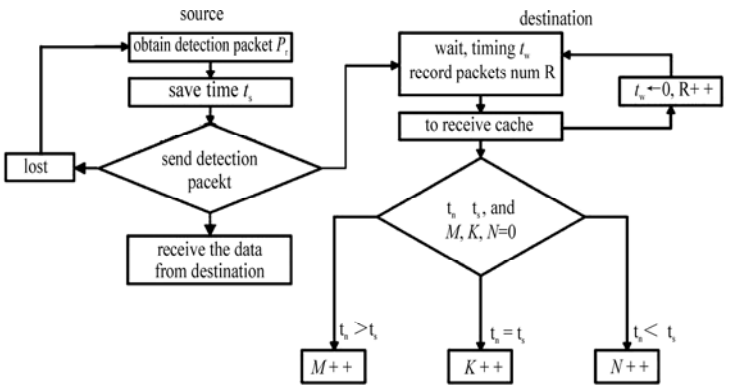


Figure 2. Flow chart with improved PTR algorithm.

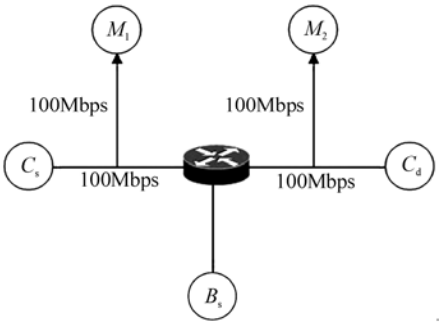


Figure 3. Measurement topology.

Improving these two points by the original PTR algorithm is that PTR algorithm is a method to change the detection packets through its own to measure network. This algorithm requires its own change to directly reflect the network, but if the change of background flow is also obviously, it is impossible to know the variation packets are in congestion state. So the measurement result is deviation. At the same time, once the background flow is too big, may lead the detection packets lost, which is another reason to deviation of the measurement. These deviations are PTR algorithm need to improve.

Improved algorithm, on the one hand, grasps the background flow of the changes is more accurately; on the other hand, the transmission of detection packet hold maximize control. Both of these improvements, the accuracy of measurement have increased greatly, and the changes of the load may too faster.

4. Experimental Results

First, using the improved PTR algorithm to measure network, and calculate the theoretical data and circulate algorithm to compare whether the results match. Second, through network to analysis receive data to prove its precision and timeliness, and compare the original algorithm to know the improved algorithm has higher accuracy.

Measurement topology as is shown in Figure 3. C_s is source and C_d is destination above detection packet. M_1 is monitor the variation in time interval with detection packet by source and M_2 is monitor by destination. Monitor and compare data by two ends of the router when detection packets have been sent. The topology of experiments use the emulator [12], On this basis, compare the original algorithm and the improved algorithm with the theoretical value, and accord statistical properties of the Internet flow [13,14].

According to the Equation (4), $\sum_{i=1}^{l} avg\_rec\_gap(i)$ can be used in the algorithm to express the value of dst_gap_sum, so

$$R_m = \frac{s \times 8 \times l}{dst\_gap\_sum} = \frac{s \times 8 \times n}{\sum g_O},$$

That acquire the output time interval g_O. Compare the measurement curves with the value of $\frac{g_O}{g_I}$ as is shown in Figure 4, it also can reflect the basic values of the network.

For $g_B = 0.08ms$, $g_I = 0.31ms$, $B_C = 7.2Mb/s$, $B_O = 20Mb/s$, the length of the detection packet is 700 Byte, and there are 60 packets in a group. Using the emulator tools to make the background flow intensity increased gradually from $0Mb/s$ to $20Mb/s$. The reason to take these values is whether it is time or rate value, the size of these values are relatively modest which have not been lost with changes in the quality of network easily and have not been obliterated with the

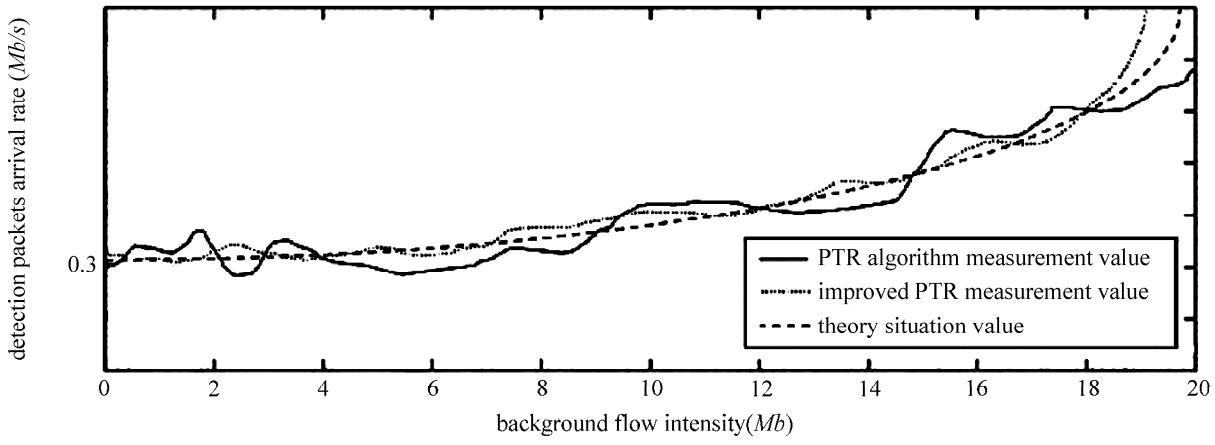


Figure 4. The theoretical value compared with the measured value.

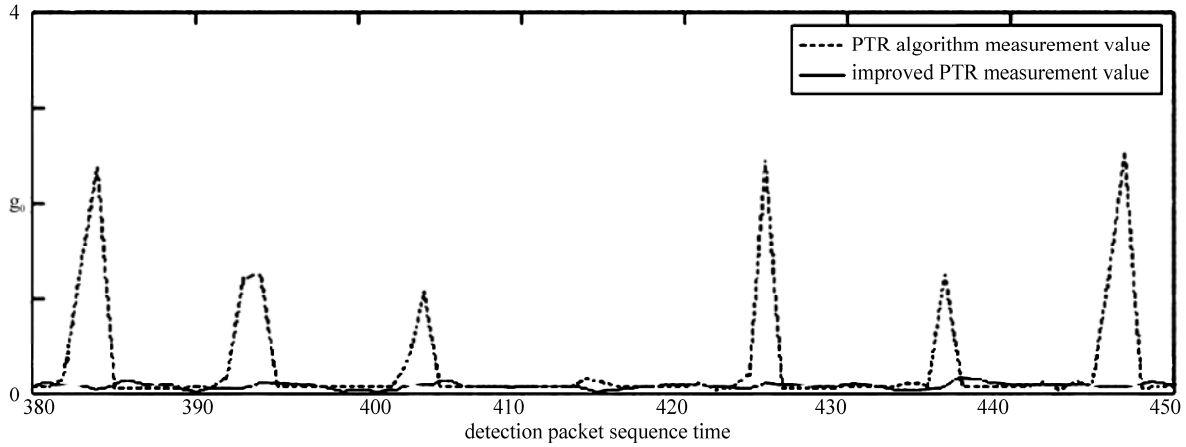


Figure 5. The comparison of the value of receiving which are based on the same source rate between the original algorithm and the improved algorithm.

excessive bandwidth facilely [4,15]. The theory is that before the background flow intensity reached $12.8Mb/s$ ($20Mb/s - 7.2Mb/s$), the rate of the detection packets arrived at the destination will not obviously change which have been sent. So, before the background flow intensity reaches $12.8Mb/s$, the detection packets line in the Figure 4 shows as a curve that the slope grows slowly. When the background flow intensity from $12.8Mb/s$ to $20Mb/s$, detection packet rate will slow down gradually until the background flow can occupied available bandwidth fully and causing bottlenecks that lead to detection packets are inaccessible to destination. At the destination, the detection rate of the available bandwidth is $0Mb/s$. At this point in Figure 4, the rate of the detection packets have been showed the curve that the slope as $+\infty$. The measurements and theoretical values inosculate basically. At the same time, the improved algorithm compared to the original algorithm can be clearly seen, the original algorithm with a strong volatility, and the improved algorithm performance the value have been smoothed. It is just consistent with the situation of the algorithm performance which after setting the two fixed values in Caption 3.2 of the above described.

According to the Equation (2), when the background flow in the path and compete bandwidth with the detection packet flow, the output data stream for the time interval is $g_O = g_B + \frac{B_C \cdot g_I}{B_O}$. This equation is going to be validated with theoretical and experimental, and compare the improved algorithm to the original algorithm, as is shown in Figure 5.

Here, get $B_O = 100Mb/s$. The remaining data with the same on the past experimental data that the theoretical value is:

$$g_O = 0.08 + \frac{7.2 \times 0.31}{100} = 0.10ms$$

In the experimental model, the source send detection packets in the C_s can be measured in the value of g_I in the M_1 as shown in Figure 5 of the initial interval value. When detection packets compete with background flow which occupy the bandwidth $B_C = 7.2Mb/s$ to though the router, the measurement value of the improved PTR algorithm g_O is as shown in Figure 5 with the value "improved algorithm" in M_2, and the measurement the same value of the original PTR algorithm as is shown in Figure 5 with the value "original algorithm". The experimental results indicate that the improved algorithm of the measurement results are nearly match with the calculation results in Equation (2) , and the data from the improved algorithm are more stable than original algorithm.

The experiment result is proved the usefulness of the detecting packets records as Caption 3.3.

The measurement result, which is gained by the experimental environment, and using the tool of MRTG (Multi Router Traffic Grapher), compares with the improved algorithm and the original algorithm. MRTG is a software tool of monitoring network link flux load. It use the snmp agrement to gain the flow information of the equipment, and displays flux load to users by HTML documents of graphics which included PNG format, displaying the flux load in a very intuitive form. In the interception 24-hour process of measurement, it has entered 40Mb/s background flow for one hour initiatively, but the intensity of other times background flow is only 20Mb/s. The measurement results are shown in Figure 6. From the measurement results, we can see the measurement results of improved algorithm and the stability of MRTG are almost always the same and the improved algorithm data are more stable than the original algorithm data at the same time. Because the loss of detection group is almost inevitable in the network, it has not

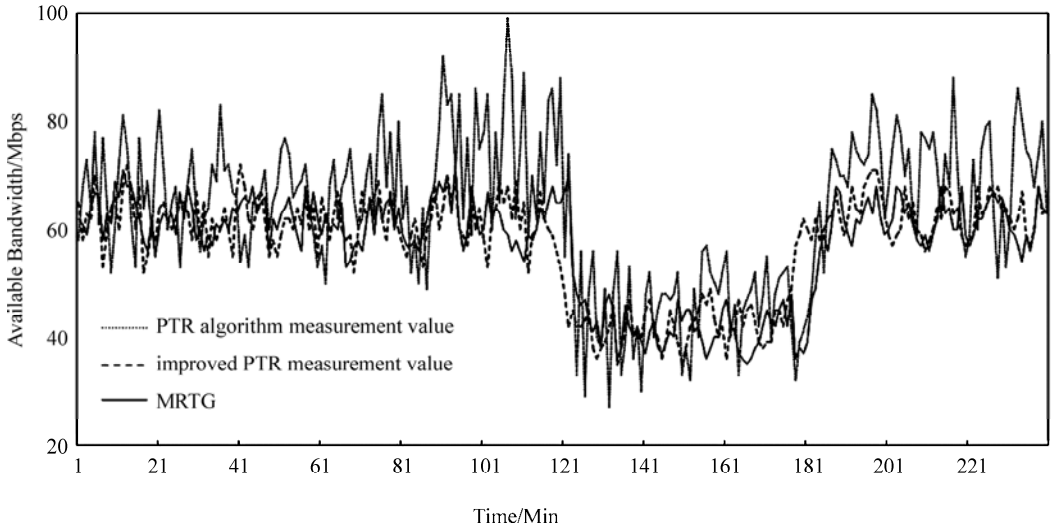


Figure 6. The original algorithm compete the same sending rate with the improved algorithm.

considered the factor of loss of detecting packets in original algorithm. We can obtain the velocity of detecting packets by data overall, the original algorithm would produce some peak value as Figure 6 after detecting group loss. The improved PTR algorithm, by dealing with the lost detection packets, the results of detect is more stable and accord with the measurement value which was gained from MRTG.

When the background flow compete the whole bandwidth with the data flow, the improved PTR algorithm measure the data always accurate.

5. Conclusions

This paper discusses the theory and application of PTR algorithm of network measurement. After the analysis of the limitations and shortcomings of the algorithm, it proposed the improved algorithm and processes, and by comparing the algorithm's theoretical value and the actual measured value to induce conclusions for performance improved algorithm can match with the theoretical value better. At the same time, it also put up that improved algorithm can measure the ins and outs of the network accurately, increasing the veracity of the network measuring and boosting up the precision of the network measuring. Especially when the network speed fluctuates frequently, it has a greater improvement in reflecting of network conditions timely and accurately.

6. References

[1] S. Banerjee and A. Agrawala, "Estimating available capacity of a network connection [A]," IEEE International Conference on Networks [C], Singapore, pp. 131–138, September 2000. "NetDyn: Network Measutrements Tool," http://www.CS.umd.edu/suman/netdyn/.

[2] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools [A]," Proceedings of ACM Intemet Measurement Conference (IMC) [C], Miami Beach, Florida, October 2003.

[3] N. N. Hu and P. Steenkiste, "Evaluation and characterization of available bandwidth probing techniques [J]," IEEE Journal on Selected Areas in Communications, Vol. 21, No. 6, pp. 879–894, August 2003.

[4] C. Dovrolis, P. Ramanathan, and D. Moore, "What do packet dispersion techniques measure?" in Proceedings of Conference Computer Communication, pp. 905–914, April 2001.

[5] M. Jain and C. Dovrofis, "End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput [A]," Proceedings of ACM SIGCOMM Symposium on Communication Architectures Protocols [C], Pittsburgh, PA, USA, pp. 295–308, August 2002.

[6] K. Lai and M. Baker, "Nettimer: A tool for measuring bottleneck link bandwidth," in Proceeding of USENIX Symposium on Internet Technologies and Systems1, pp. 123–134, March 200.

[7] R. Prosad, C. Davrolis, M. Murray, et al., "Bandwidth estimation: Metrics measurement techniques and tools [J]," IEEE Network, Vol. 17, No. 6, pp. 27–35, 2003.

[8] V. Paxson, "Measurements and analysis of end-to-end internet dynamics," Ph. D. dissertation, Computer Science Division, U. C. Berkeley, Berkeley, CA, May 1996.

[9] N. N. Hu and P. Steenkiste, "Estimating available bandwidth using packet pair probing [J]," Carnegie Mellon University (CMU), 9 September 2002.

[10] Pasztor A, Veitch D. The Packet Size Dependence of Packet Pair Like Methods[C]. Proc. of IWQoS' 02, Mi-

ami Beach, Florida, USA, 2002.

[11] Ns2 [Online]. Available: http://www.isi.edu/nsnam/ns.

[12] K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: Recent traffic measurements from an internet backbone," presented at the ISOC INET Conf., July 1998.

[13] Y. Zhang, N. Duffield, V. Paxson, and S. Shenker, "On the constancy of Internet path properties," in Proc. ACM SIGCOMM Internet Measurement Workshop, San Francisco, CA, Nov. 2001, pp. 197–211.

[14] K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: Recent traffic measurements from an internet backbone," presented at the ISOC INET Conf., July 1998.

Int. J. Communications, Network and System Sciences, **2009, 2, 754-758**
doi:10.4236/ijcns.2009.28087 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Modified Ceiling Bounce Model for Computing Path Loss and Delay Spread in Indoor Optical Wireless Systems

K. SMITHA [1], A. SIVABALAN [2], J. JOHN [2]
[1]*Delphi Technical Center India, Bangalore, India*
[2]*Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India*
Email: Smitha.Chandrasekhar@delphi.com, {sbalan, jjohn}@iitk.ac.in
Received July 18, 2009; *revised August* 18, 2009; *accepted September* 23, 2009

Abstract

This paper proposes modifications to the tradional Ceiling Bounce Model and uses it to characterize diffuse indoor optical wireless channel by analyzing the effect of transceiver position on signal propagation properties. The modified approach uses a combination of the tradional ceiling bounce method and a statistical approach. The effects of different transmitter-receiver separations and height of the ceiling on path loss and delay spread are studied in detail.

Keywords: Indoor Optical Channel, Modified Ceiling Bounce Model, Path Loss, Delay Spread

1. Introduction

The increasing demand for high data rates along with high mobility of data terminals has resulted in the expanding popularity of optical wireless local area networks (LANs) [1–3]. The optical spectral region has plenty of unused unregulated bandwidth making it possible to establish high bit rate data links. Since optical signals are blocked by the walls of the rooms, optical wireless communication systems are secure from eavesdropping and interference. The square law photo detector used at the receiver end is always thousands of times larger than the wavelength of the light and hence, multipath propagation does not produce fading in a direct detection system.

Among different IR system configurations, the diffuse topology is the most robust one for local area networks as it does not require either LOS path between the transmitter and receiver or strict alignment between them. The problems associated with such a configuration are high path loss and intersymbol interference (ISI) due to multipath dispersion. Multipath propagation results in ISI because of the spreading out of pulses in time due to the availability of different paths of varying path lengths for propagation. This limits the maximum bit rate achievable.

Detailed characterization of multipath medium is essential for the successful design of indoor wireless systems. Modeling and simulation of indoor infrared channel has been addressed in the literature with the pioneering work of Gfeller *et al.* [1,2], who introduced the idea of using infrared for indoor wireless communications. They presented a method for determining the power distribution throughout a room given the geometry of the channel. Barry *et al.* [4–7] proposed the recursive method for evaluating the impulse response of an indoor free-space optical channel with Lambertian reflectors through which accurate analysis of the effects of multipath dispersion can be carried out for any multiple reflections of any order. Perez Jimenez *et al.* [8,9] suggested a closed-form expression for the RMS delay spread which can be used to find the impulse response of an optical wireless channel, based on several experiments. Carruthers *et al.* [10] proposed the Ceiling Bounce model which adopts a simple modeling approach assuming an infinitely large room, i.e. considering only a single reflection from the transmitter to receiver via ceiling in the room.

In this paper, we present a detailed characterization of the indoor optical wireless channel by combining the statistical approach [8,9] and the traditional ceiling bounce model [10]. The effect of transmitter and receiver location on different system features viz., RMS delay spread, path loss and system bandwidth are analysed in detail. In Section 2, we define our impulse response calculation method. Effect of transceiver position on RMS Delay Spread is discussed in Section 3. Section 4 discusses the effect of transmitter and receiver position

on path Loss. The last section describes the conclusions of the work.

2. Impulse Response Calculation Using Modified Ceiling Bounce Approach

For the calculation and analysis the room is assumed to be empty. It is assumed that the transmitter is pointed straight upwards and emits a Lambertian pattern which corresponds to a transmitter semi angle (at half power) of 60°. The reflecting elements are also assumed to be perfect Lambertian reflectors with reflectivity between 0.6 and 0.8 [7]. The receiver is assumed to be pointed straight upward. The transmitter and receiver are located, respectively, at coordinates (x_1, y_1) and (x_2, y_2) in the horizontal (x,y) plane and h_1 and h_2 represents the transmitter-ceiling and receiver-ceiling vertical separations. The room is assumed to be empty. Figure 1 represents the configuration explained above.

2.1. Traditional Ceiling Bounce Approach

We base our study on the Ceiling Bounce method developed by Carruthers *et al.* [10] where the impulse response of a channel is given by,

$$h(t) = G_o 6(a)^6 (t+a)^7 u(t) \qquad (1)$$

where, $a = 2H/c$ is called the ceiling bounce parameter. $G_o = \rho A 3\pi H^2$ represents the DC (optical) gain, $u(t)$ is unit function, ρ is the plane reflectivity, A the receiver photodiode area, H is the separation distance from the ceiling (transmitter and receiver assumed to be co-located), and c the velocity of light.

The above model represents the impulse response due to diffuse reflection from a single infinite plane reflector, which is a good approximation to a large ceiling. As an approximation, this model considers only the first bounce off the ceiling and ignores the higher order reflections from the walls and ceiling. When the transmitter and receiver are near the center of a large room, the impulse response of a diffuse configuration is dominated by the single bounce off the ceiling. When the receiver is further separated from the transmitter, the ceiling no longer appears to the transmitter to be well approximated by an infinite plane, and the contribution of the walls to the impulse response will increase relative to that of the ceiling. This method cannot consider different separations of transmitter and receiver from the ceiling. So a new approach is followed in this work to find the impulse response and the latter channel analysis using this response.

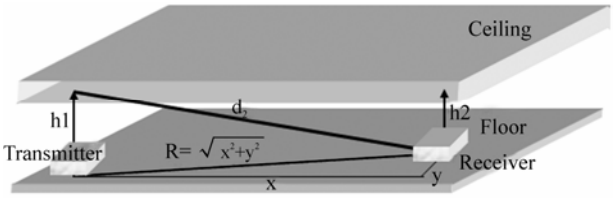


Figure 1. Configuration chosen for channel parameter analysis.

2.2. Modified Ceiling Bounce Approach

2.2.1. Calculation of RMS Delay Spread

For fixed transmitter and receiver locations, multipath dispersion is completely characterized by the channel impulse response $h(t)$. The RMS delay spread τ_{rms}, is commonly used to quantify the time dispersive properties of multipath channels .

In this work, the RMS delay spread is calculated initially using the statistical approach proposed by Perez Jimenez *et al.* [8,9]. According to this approach, the value of rms delay spread depends upon the distance between transmitter-reflector-receiver d, the transmission angle between transmitter and receiver θ, and mode number of the source radiation pattern.

The general expression for rms delay is given by [8]

$$\tau_{rms}(ns) = a + b\cos(c\theta + d) \qquad (2)$$

For the configuration considered in this paper, after substituting the statistical parameters, the above equation becomes [8],

$$\tau_{rms} = -2.37 + 0.007\, n + (0.8 - 0.002\, n)\, d \qquad (3)$$

The estimated values of rms delay spread obtained using this closed form expression is very accurate. This value is used as the parameter in the ceiling bounce model to find the actual impulse response.

2.2.2. Calculation of Path Loss

The path loss of an unshadowed diffuse configuration can be estimated using the expression [7]:

$$G_o = \rho A_R h_1^2 h_2^2 \pi^2 \iint_{Ceiling} 1(h_1^2 + (x-x_1)^2 + (y-y_1)^2)^2 \, dx\, dy (h_2^2 + (x-x_2)^2 + (y-y_2)^2)^2 \qquad (4)$$

The expression assumes a detector field-of-view half angle of 90° and a Lambertian source. The results almost follow closely the experimental results, but it shows variations at large horizontal separations, where the effect of neglected higher order reflections is relatively important.

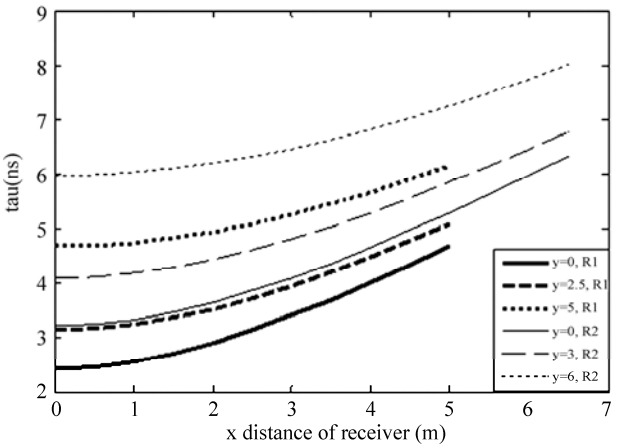
2.2.3. Impulse Response Calculation

In order to find the exact response of, first the ceiling bounce parameter is estimated accurately using the relation, $a = 12\sqrt{\frac{11}{13}}\tau_{rms}$, where τ_{rms} is obtained using (3). This value of ceiling bounce parameter and the value of path loss obtained above substituted in (1) to get the impulse response of a diffuse infrared channel. The model so developed is much better than the traditional model due to following reasons. The values of G_o and a are determined by the locations and orientations of the transmitter and receiver within the room. This model can take into account different separations of transmitter and receiver from the ceiling rather than assuming the transmitter and receiver to be co-located. Thus it can analyse the effects of multipath dispersion effectively and determine the power distribution profile. This approach can also estimate the variations in rms delay spread, system bandwidth and path loss due to change in position of transmitter and receiver.

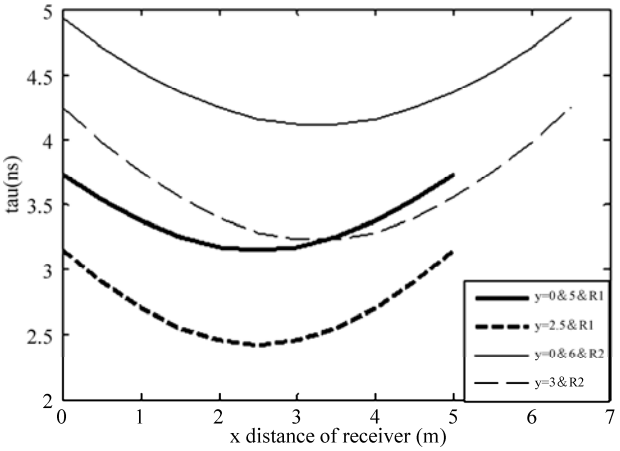
3. Effect of Transceiver Position on RMS Delay Spread

In all our computations, one corner of the room is assumed to be the origin (0,0). The length of the room is assumed to be the x co-ordinate, breadth to be the y co-ordinate and height to be the z co-ordinate. The rms delay spread variations with different transmitter and receiver position in different rooms were calculated. Two cases are considered in each room. In the first case, transmitter is kept at one corner of the room and receiver is moved all over the room. The second case considers transmitter to be placed at the center of the room and the receiver moved all around the room. Figure 2 show the variation in the delay spread with transmitter and receiver location for two rooms.

From the Figure 2 (a) and (b) it is clear that the value of rms delay spread depends on the distance between the transmitter and the receiver, as well as the separation from the ceiling. It increases with increase of both the quantities. If we observe the two figures, it can be noted that the value of the rms delay spread in Figure 2(a) is larger than the value in Figure 2(b) for the same receiver position. This is because of the change in the transmitter location between the two. We can also see that the maximum value of rms delay spread increases with room size. This is because of the increase in the number of paths and the path lengths which causes more time to reach the destination after multiple reflections. Thus the value of rms delay spread depends on the position of transmitter, receiver and the room size chosen. Even in the same room, by properly locating transmitters, we can reduce the rms delay spreads.



(a) Transmitter placed at one corner



(b) Transmitter placed at centre

Figure 2. Variation of rms delay spread with receiver positions for Room1-5x5x3m, Room2-6.5x6x3.5m.

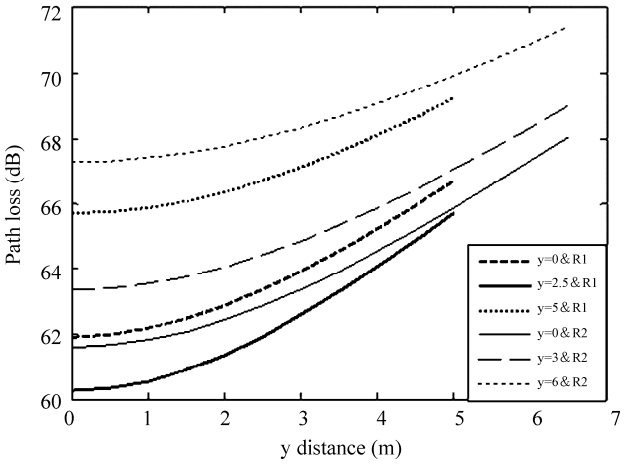
4. Effect of Transmitter and Receiver Position on Path Loss

DC gain and path loss are calculated using Equation (4) through numerical integration. Figure 3(a) and (b) shows the variation of path loss with change in receiver position for two different room sizes. These figures clearly show that, when the separation between the transmitter and receiver increases, path loss also increases.

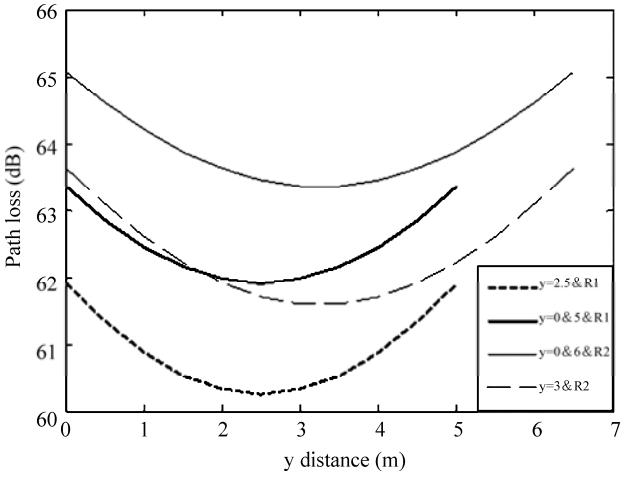
4.1. Impulse and Frequency Responses

The modified model has been used to find the impulse response of infrared channel for different transmitter and receiver positions. The corresponding frequency response is obtained by taking the Fourier transform of the impulse response.

Figure 4 represents the impulse response and frequency response plots obtained in a room of size 5mx5m

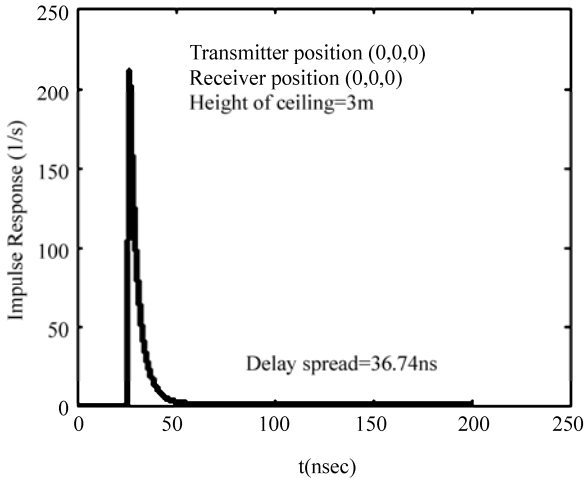


(a) Transmitter placed at one corner

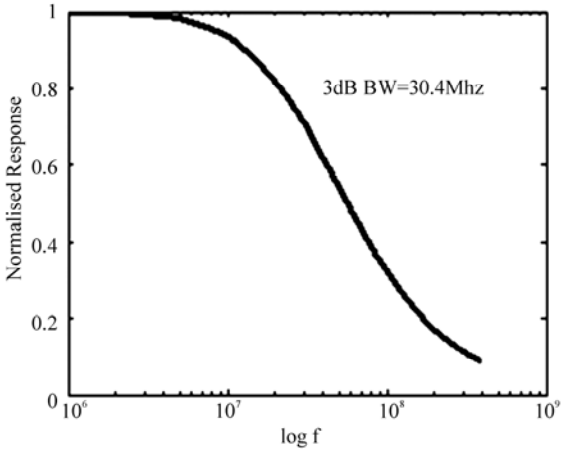


(b) Transmitter placed at centre

Figure 3. Variation of path loss with receiver positions for Room1-5x5x3m, Room2-6.5x6x3.5m.



(a) Impulse response



(b) Frequency response

Figure 4. The impulse response and the frequency response for room size of 5x5x3m.

Table 1. Variation of channel parameters with receiver position.

Room 5x5x3m ; Tx(0,0)				
Rx	Path loss(dB)	Tau(ns)	Delay spread (ns)	BW(MHz)
(0,0)	60.25	2.43	36.74	30.4
(2.5,2.5)	63.35	3.73	50.69	19.7
(5,5)	69.20	6.16	73.20	11.9
Room 6.5x6x3.5m; Tx(0,0)				
Rx	Path loss(dB)	Tau(ns)	Delay spread(ns)	BW(MHz)
(0,0)	61.59	3.22	45.46	22.8
(3,3)	64.84	4.82	61.2	15.3
(6,6)	70.88	7.75	86.4	9.4

x3m. Table 1 shows all the important channel parameters obtained using the modified ceiling bounce approach in two rooms of size 5x5x3m and 6.5x6x3.5m. As the separation between the transmitter and receiver increases, the system bandwidth also decreases as is evident from the Table 1. This has effect on the maximum bit rate achievable. With distance the multipath effects are more pronounced, which causes a decrease in the bandwidth, thus resulting in reduction of the maximum bit rate achievable.

5. Conclusions

Modified ceiling bounce method to find the propagation properties of the channel is propossed. This method allows computing impulse response of the diffuse channel with less computational complexity than the simple Ceiling bounce model. The influence of transceiver position on the indoor diffuse channel parameters is analyzed. Results clearly show that path loss is a function of separation between the transmitter and receiver.

6. References

[1] F. R. Gfeller, and U. Bapst, "Wireless in-house data communication via diffuse infrared radiation," Proceedings of IEEE., Vol. 67, pp. 1474–1485, 1979.

[2] F. R. Gfeller, H. R. Muller, and P. Vettiger, "Infrared communication for in-house applications," Proceedings of IEEE Conference on Computer Communication, pp. 132–138, 1978.

[3] Z. Ghassemlooy and A. C. Boucouvalas, "Guest editorial: Indoor optical wireless communications system and networks," International Journal Communication System, Vol. 18, pp. 191–193, 2005.

[4] J. R. Barry, J. M. Kahn, W. J. Krause, E. A. Lee, and D. G. Messerschmitt, "Simulation of multipath impulse response for wireless optical channels," IEEE Journal of Selected Areas on Communication, Vol. 11, pp. 367–379, April 1993.

[5] J. R. Barry, "Wireless infrared communication," Boston: Kluwer, 1994.

[6] J. M. Kahn and J. R. Barry, "Wireless infrared communications," Proceedings of IEEE, Vol. 85, pp. 265–298, February 1997

[7] J. M. Kahn, W. J. Krause, and J. B. Carruthers, "Experimental characterization of nondirected indoor infrared channels," IEEE Transaction Communication, Vol. 43, pp. 1613–1623, 1995.

[8] R. P. Jimenez, J. Berges, and M. J. Betancor, "Statistical model for the impulse response of infrared indoor diffuse channels," IEE Electronic Letters, Vol. 33, pp. 1298–1300, 1997.

[9] R. P. Jimenez, V. M. Melian, and M. J. Betancor, "Analysis of multipath impulse response of diffuse and quasi-diffuse optical links for IR-WLAN," Proceedings of 14th Conference on IEEE Computer Communication Society, Vol. 1, pp. 924–930, 1995.

[10] J. B. Carruthers and J. M. Kahn, "Modelling of nondirected wireless infrared channels," IEEE Transa- ction on Communication, Vol. 45, pp. 1260–1268, 1997.

Int. J. Communications, Network and System Sciences, **2009, 2, 759-763**
doi:10.4236/ijcns.2009.28088 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A MAC Scheme with QoS Guarantee for MANETs

Yanbin YANG[1,2], Yulin WEI[2]
[1]*School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China*
[2]*Department of Communications, Guangdong Vocational Technology College of Posts & Telecom, Guangzhou, China*
Email: *sinhei*@163.*com, TX*06*wls*@126.*com*
Received July 30, 2009; *revised August* 31, 2009; *accepted September* 25, 2009

Abstract

IEEE 802.11 distributed coordination function (DCF) can alleviate the collision and hidden station problem, but it doesn't differentiate traffic categories (TC). Therefore, it can't provide sufficient QoS support for different traffic categories. Recently, a new contention-based enhanced distributed channel access (EDCA) scheme was proposed which provides a probabilistic QoS support. In this paper, an adaptive EDCA scheme with QoS guarantee for mobile ad hoc networks (MANETs) is proposed. In this scheme, the EDCA scheme and the token bucket algorithm (TBA) are combined to adjust the contention window (CW). Our scheme provides the traffic differentiation.

Keywords: Mobile Ad Hoc Network, Medium Access Control, Distributed Coordination Function, Enhanced Distributed Channel Access, Contention Window

1. Introduction

Mobile ad hoc network (MANET) becomes popular because of its low cost and easy deployment. The medium access control (MAC) protocol for this wireless communication system employs a mandatory contention-based channel access function called Distributed Coordination Function (DCF) [1], which is based on the Carrier Sense Multiple Access (CSMA) mechanism. Mobile stations deliver MAC Service Data Units (MSDUs) after detecting that there is no other transmission on the same wireless medium. However, if two stations detect that the channel is free at the same time, a collision occurs. The IEEE 802.11 standard defines a Collision Avoidance (CA) mechanism to reduce the probability of such collisions. As a part of CA, before starting a transmission a station performs a backoff operation. It has to keep sensing the channel for an additional random time after detecting the channel as being idle for a minimum duration called DCF Interframe Space (DIFS). Only if the channel remains idle for this additional random time period, the station is allowed to initiate the transmission. The duration of this random time is determined as a multiple of a slot. Each station maintains a so-called Contention Window (CW), which is used to determine the number of slots a station has to wait before transmission.

To avoid the hidden station problem inherent in CSMA, IEEE 802.11 defines a Request to Send/Clear to Send (RTS/CTS) scheme, which can be used optionally. Before transmitting data frames, a station has the option to transmit a short RTS frame, followed by the CTS transmission by the receiving station. The RTS and CTS frames include the information of how long it does take to transmit the next data frame, i.e., the first fragment, and the corresponding ACK response. Thus, other stations close to the transmitting station and hidden stations close to the receiving station will not start any transmissions; their timer called Network Allocation Vector (NAV) is set. RTS/CTS helps to protect long data frames against hidden stations. With fragmentation, multiple ACKs are transmitted, whereas with RTS/CTS the MSDU can be efficiently transmitted in a single data frame. Between two consecutive frames in the sequence of RTS, CTS, data, and ACK frames, a Short Interframe Space (SIFS) offers transceivers time to turn around. Because RTS/CTS handshake may introduce some delay especially in case of wireless networks, [2] proposed a method to improve overall throughput of the network where selected RTS packets are dropped based on a node sequence number.

802.11 DCF reduces collision and hidden terminal problem, but it doesn't differentiate TC. Therefore it can't provide sufficient QoS support for different traffic categories. Recently, a new contention-based EDCA scheme was proposed which provides a probabilistic QoS support. [3] presented a new protocol, called dis-

tributed end-to-end allocation of time slots for real-time traffic (DARE). It allocates and uses periodic time slots for QoS-demanding applications. DARE reserves these time slots in a fully distributed way, schedules the real-time data packets, repairs broken reservations, and disseminates the reservation information to potential interferers using a piggyback technique. It works well for high traffic load network but poorly for low traffic load network. [4] presented a distributed MAC scheme called opportunistic synchronous array method (SAM) for a large network of wireless routers. [5] presented a novel tone-based contention resolution mechanism that exploits space-time uncertainty and high latency to detect collisions and count contenders, achieving good throughput across all offered loads for underwater acoustic sensor networks.

In this paper, an adaptive EDCA scheme with QoS guarantee for MANETs is proposed. In this scheme, the EDCA algorithm and the token bucket algorithm (TBA) are combined to adjust the CW. Our scheme provides the traffic differentiation.

2. Token Bucket Algorithm

The number of bytes in the bucket (bucket length) and the occupancy of the transmission buffer (queue length) as input parameters are employed in the token bucket algorithm [6] (see Figure 1).

The bucket size (bsize) determines the accepted burstiness of source. The bucket length (blen) represents the resources that the user can use to transmit packets. The bucket limit (blim) represents the maximum packet length allowed. The queue size (qsize) represents the source capacity. The queue length (qlen) denotes the willingness of a station to transmit packets.

This algorithm computes a value which is used to scale the CW values defined in IEEE 802.11. It is described as follows:

If (overload) then p=(1+Δ4)p
Else if (qlen=0) then p=(1+Δ1)p
Else if (blen<blim) then p=(1+Δ2)p
Else p=(1-Δ3)p
P=min{p,1}

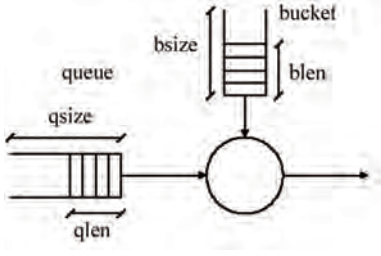


Figure 1. Brief description of token bucket algorithm.

CW=p*CW802.11

where Δ1=0.025, Δ4=0.25, $\Delta_2 = \frac{b\lim - blen}{b\lim}\Delta_1$, $\Delta_3 = \frac{blen - b\lim}{bsize - b\lim}\Delta_1$. A solution to determine the overload is described as follows:

If (av_nr_coll>c) then overload=true
Av_nr_coll=(1-t)*num_coll+av_nr_coll

where av_nr_coll is the average collision number; c is a constant with value in [0,8] and is always set to 4; t=0.25; num_coll is the collision number after transmission.

3. Adaptive EDCA

EDCA [7] is designed to provide prioritized QoS by enhancing the contention-based DCF. It provides differentiated, distributed access to the wireless medium for QoS stations (QSTAs) using 8 different user priorities (UPs). Before entering the MAC layer, each data packet received from the higher layer is assigned a specific user priority value. How to tag a priority value for each packet is an implementation issue. The EDCA scheme defines four different first-in first-out (FIFO) queues called access categories (ACs) that provide QoS support for the delivery of traffic with UPs at the QSTAs. Each data packet from the higher layer along with a specific user priority value is mapped into a corresponding AC according to Table 1. Different kinds of applications (e.g., background traffic, best effort traffic, video traffic, and voice traffic) can be casted into different ACs. For each AC, an enhanced version of the DCF, called enhanced distributed channel access function (EDCAF), contends for TXOPs using a set of EDCA parameters from the EDCA Parameter Set Element or from the default values of the parameters when no EDCA Parameter Set Element is received from the QAP of the QBSS with which the QSTA is associated. [8] proposed an effective and simple model for two-level collision to demonstrate that EDCA

Table 1. User priority to access category mappings.

User priority (UP)	Access category (AC)	Designation
1	0	Background
2	0	Background
0	1	Best effort
3	1	Best effort
4	2	Video
5	2	Video
6	3	Voice
7	3	Voice

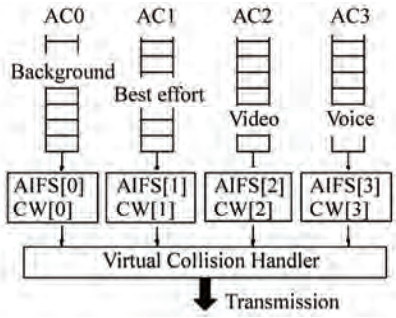


Figure 2. Implementation model.

mechanism, which adopts priority classification and two-level collision, can not only preferentially guarantee the throughput of voice and video, but also greatly decrease the average delay of the system.

Figure 2 shows the implementation model with four transmission queues, where each AC behaves like a virtual station: It contends for access to the medium and independently starts its backoff procedure after sensing the medium idle for at least AIFS period. In EDCA a new type of IFS is introduced, the arbitrary IFS (AIFS), in place of DIFS in DCF. Each AIFS is an IFS interval with arbitrary length as follows:

$$AIFS[AC] = SIFS + AIFSN[AC] \times \text{slot time}$$

where AIFSN[AC] is called the arbitration IFS number and is determined by the AC and the physical settings. The timing relationship of EDCA is shown in Figure 3. The AC with the smallest AIFS has the highest priority. The values of AIFS [AC], CWmin [AC], and CWmax [AC], which are referred to as the EDCA parameters, are announced by the AP via beacon frames. The purpose of using different contention parameters for different queues is to give a low-priority class a longer waiting time than a high-priority class, so the high-priority class is likely to access the medium earlier than the low-priority class. [9] presented the effect of different values of AIFS on MAC access delay. An internal collision occurs when more than one AC finishes the backoff procedure at the same time. In such a case, a virtual collision handler in every QSTA allows only the highest-priority AC to transmit frames, and the others perform a backoff operation with increased CW values. [10] proposed an analytical model to calculate the EDCA parameter set that ideally achieves a predetermined utilization ratio between uplink and downlink flows.

Transmission opportunity (TXOP) is defined in IEEE 802.11e as the interval of time when a particular QSTA has the right to initiate transmissions. There are two modes of EDCA TXOP defined, the initiation of the EDCA TXOP and the multiple frame transmission within an EDCA TXOP. An initiation of the TXOP occurs when the EDCA rules permit access to the medium. A multiple frame transmission within the TXOP occurs when an EDCAF retains the right to access the medium following the completion of a frame exchange sequence, such as on receipt of an ACK frame. The TXOP limit duration values are advertised by the QAP in the EDCA Parameter Set Information Element in Beacon frames. During an EDCA TXOP, a STA is allowed to transmit multiple MAC protocol data units (MPDUs) from the same AC with a SIFS time gap between an ACK and the subsequent frame transmission. A TXOP limit value of 0 indicates that a single MPDU may be transmitted for each TXOP. This is also referred to as contention free burst (CFB).

We propose an adaptive EDCA scheme for the contention period (CP). In our scheme, we combine the EDCA scheme and the token bucket algorithm (TBA) to adjust the CW. In the adjustment of the CW, there are additional aspects that have to be taken into account: 1) We do not want the CW to increase above the values used by the Best Effort terminals, since this would lead to a worse performance than Best Effort. For backward compatibility, the CW for Best Effort should be the one defined by the 802.11 standard; 2) If the low transmission rate of the application is the reason for transmitting below the desired rate, then the CW should obviously not be decreased; 3) When estimating the transmission rate, it would be desirable to control the allowed burstiness of the source; 4) CWs should not be allowed to decrease in such a way that they negatively influence the overall performance of the network. After considering all the above issues, we describe the adaptive EDCA scheme as follows.

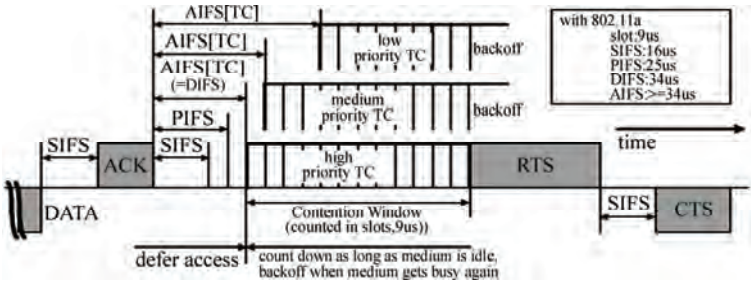


Figure 3. The timing relationship of EDCA.

3.1. Successful Transmission

Step 1. Compute the collision rate

$$f_{curr}^{j} = \frac{E(collisions_j[p])}{E(data\_sent_j[p])}.$$

Step 2. Introduce exponentially weighted moving average (EWMA) to compute the average collision rate

$$f_{avg}^{j} = (1-\alpha) * f_{curr}^{j} + \alpha * f_{avg}^{j-1},$$

where α is between 0.2 and 0.25.

Step 3. Compute p by using the TBA

If (overload) then p=(1+Δ4)p

Else if (qlen=0) then p=(1+Δ1)p

Else if (blen<blim) then p=(1+Δ2)p

Else p=(1-Δ3)p

P=min{p,1},

where Δ1=0.025, Δ4=0.25, $\Delta_2 = \frac{b\lim - blen}{b\lim}\Delta_1$,

$\Delta_3 = \frac{blen - b\lim}{bsize - b\lim}\Delta_1$.

Step 4. Use the multiply factor (MF)

MF[i]=min(p,0.8).

Step 5. After successful transmission, CW[i] is set to CWnew[i]=max(CWmin[i],CWold[i]*MF[i]).

3.2. After Collision

Step 1. Compute the collision rate

$$f_{curr}^{j} = \frac{E(collisions_j[p])}{E(data\_sent_j[p])}.$$

Step 2. Introduce exponentially weighted moving average (EWMA) to compute the average collision rate

$$f_{avg}^{j} = (1-\alpha) * f_{curr}^{j} + \alpha * f_{avg}^{j-1},$$

where α is between 0.2 and 0.25.

Step 3. Compute p by using TBA

If (overload) then p=(1+Δ4)p

Else if (qlen=0) then p=(1+Δ1)p

Else if (blen<blim) then p=(1+Δ2)p

Else p=(1-Δ3)p

P=min{p,1}

If (overload) then p=(1+Δ4)p

If ($f_{avg}^{j} > c$) then overload=true,

where c=0.5, Δ1=0.025, Δ4=0.25, $\Delta_2 = \frac{b\lim - blen}{b\lim}\Delta_1$,

$\Delta_3 = \frac{blen - b\lim}{bsize - b\lim}\Delta_1$.

Step 4. Use the multiply factor (MF)

MF[i]=min(p,0.8).

Step 5. After transmission failure, CW[i] is set according to each TC's PF[i] to guarantee that high priority TC has low PF[i] value

CWtemp[i]=min(CWmax[i],CWold[i]*MF[i])

CWnew[i]=max(CWold[i],CWtemp[i]*MF[i]).

4. Simulation Results

In this section we evaluate the performance of EDCF (IEEE 802.11e) [11], AEDCF and our scheme (Adaptive EDCA). AEDCF [12] is another improved scheme of EDCA. It also improves the throughput through changing CW.

The simulation is with respect to two scenarios, namely Scenario 1 and Scenario 2. Scenario 1 is a light traffic load environment. Scenario 2 is a high traffic load environment. The parameters CWmin, CWmax, AIFS are set to 5, 100, 5 in the scenario 1 and 30, 500, 15 in the scenario 2. In our scheme p, α, bsize, and c are set to 0.5, 0.25, 10, 0.5, respectively. Through Figure 4 we find that there is a slump of EDCF and AEDCF in the 10 stations because of collision among these stations. But our scheme (Adaptive EDCA) keep good throughput by using TBA.

Through Figure 5 we find that the throughput of AEDCF is 4.9% higher than EDCF in the high traffic load environment. But the throughput of our scheme (Adaptive EDCA) is 11.4% higher than EDCA.

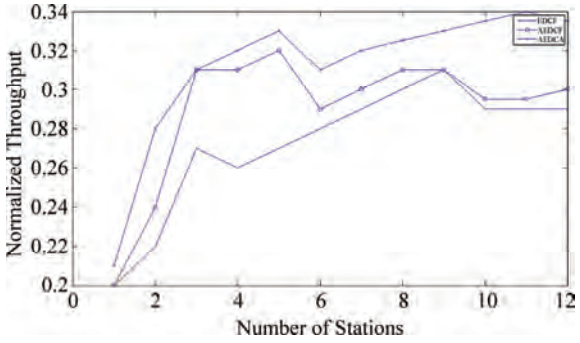


Figure 4. Scenario 1.

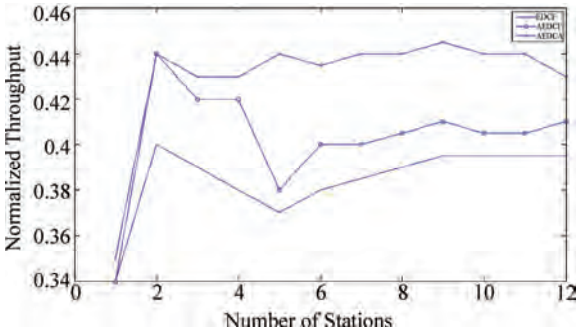


Figure 5. Scenario 2.

5. Conclusions

The real-time and multimedia services are more and more popular. We find that providing service differentiation and high throughput is a new challenge in MANETs. The MAC protocol of the IEEE 802.11e is changed in order to improve its effectiveness. The EDCA scheme and the TBA are combined to adjust the CW. Our scheme provides the traffic differentiation. The results show the good performance of our scheme under both light and high traffic loads.

6. References

[1] IEEE Std, "802.11-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," Reference number ISO/IEC 8802-11:1999(E), IEEE Std. 802.11, 1999 Edition.

[2] P. V. Krishna and Iyengar, "Sequencing technique: An enhancement to 802.11 medium access control to improve the performance of wireless networks," N.Ch.S.N., International Journal Communication Networks and Distributed Systems, Vol. 1, No. 1, pp. 52–70, 2008.

[3] E. Carlson, C. Prehofer, C. Bettstetter, *et al*, "A distributed end-to-end reservation protocol for IEEE 802.11-based wireless mesh networks," IEEE Journal on Selected Areas in Communications, Vol. 24, No. 11, November 2006, pp. 2018–2027.

[4] B. Zhao and Y. B. Hua, "A distributed medium access control scheme for a large network of wireless routers," IEEE Transactions on Wireless Communications, Vol. 7, No. 5, pp. 1614–1622, May 2008.

[5] A. A. Syed, W. Ye, and J. Heidemann, "T-Lohi: A new class of MAC protocols for underwater acoustic sensor networks," IEEE INFOCOM'08, The 27th Conference on Computer Communications, pp. 231–235, 13-18 April 2008.

[6] A. Banchs and X. Perez, "Providing throughput guarantees in IEEE 802.11 wireless LAN," Wireless Communications and Networking Conference, IEEE WCNC'02, Vol. 1, pp. 130–138, 17-21 March 2002.

[7] S. Sehrawat, R. P. Bora, and D. Harihar, "Performance analysis of QoS supported by enhanced distributed channel access (EDCA) mechanism in IEEE 802.11e," IAENG International Journal of Computer Science, IJCS_33_1_6, Vol. 33, No. 1, February 2007.

[8] J. R. Yan, S. Y. Zhang, H. Long, and Y. F. Sun, "An analytical model for EDCA mechanism," Journal of Electronics & Information Technology, Vol. 30, No. 4, April 2008.

[9] X. Y. Zhou, H. F. W, L. M. Sun, *et al*., "MAC access delay analysis of EDCA mechanism of wireless LANs," Journal of Software, Vol. 19, No. 8, pp. 2127−2139, 2008. http://www.jos.org.cn/1000-9825/19/2127.htm.

[10] F. Keceli, I. Inan, and E. Ayanoglu, "Weighted fair uplink/downlink access provisioning in IEEE 802.11e WLANs," ICC'08, IEEE International Conference on Communications, pp. 2473−2479, 19-23 May 2008.

[11] S. H. Choi, J. del Prado, N. S. Shankar, *et al*., "IEEE 802.11 e contention-based channel access (EDCF) performance evaluation," ICC'03, IEEE International Conference on Communications, pp. 1151−1156, 11-14 May 2003.

[12] L. Romdhani, Q. Ni, T. Turletti, "Adaptive EDCF: Enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks," IEEE Wireless Communications and Networking, Conference, Vol. 2, pp. 16−20, March 2003.

Int. J. Communications, Network and System Sciences, **2009, 2, 764-774**
doi:10.4236/ijcns.2009.28089 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Reputation-Based Multi-Agent Model for Network Resource Selection

Junfeng TIAN, Juan LI, Lidan YANG
Network Technology Institute, Hebei University, Baoding, China
Email: *tjf@hbu.edu.cn, {antylj, mousekidcn*1984}@163.*com*
Received July 7, 2009; *revised August* 12, 2009; *accepted September* 27, 2009

Abstract

Because of the anonymity and openness of on-line transactions and the richness of network resources, the problems of the credibility of the on-line trading and the exact selection of network resources have become acute. For this reason, a reputation-based multi-agent model for network resource selection (RMNRS) is presented. The model divides the network into numbers of trust domains. Each domain has one domain-agent and several entity-agents. The model prevents the inconsistency of information that is maintained by different agents through the periodically communication between the agents. The model enables the consumers to receive responses from agents significantly quicker than that of traditional models, because the global reputation values of service providers and consumers are evaluated and updated dynamically after each transaction. And the model allocates two global reputation values to each entity and takes the recognition value that how much the service provider knows the service into account. In order to make users choose the best matching services and give users with trusted services, the model also takes the similarity between services into account and uses the similarity degree to amend the integration reputation value with harmonic-mean. Finally, the effectiveness and feasibility of this model is illustrated by the experiment.

Keywords: Trust, Reputation, Trust-Domain, Multi-Agent, Similarity

1. Introduction

The World Wide Web has evolved at an extreme rate due to its capacity to provide an endless amount of resources to the public users. Hence, the user finds him lost in a pool of information, without knowing how to select resources and which resources are credible [1]. A wide range of mechanisms such as contracts and commercial laws as well as face to face meetings help reduce the likelihood of risks to the consumers in the traditional businesses. However, When doing online trading, users often have little or no prior knowledge of their potential business partner(s) and the absence of these mechanisms and face-to-face encounter make them can not check goods before paying and can not distinguish cheat from honest effectively. As a result, there has always produced a trust deficit in e-commerce [2].

Thus, the trust has become an integral part of traditional transaction and e-commerce transaction. Recommender system which is based on trust have proven to be an important method to effectively find those resources that users are interested in from endless resources in the network, by providing users with more proactive and personalized information services. And the recommender system based on trust collaborative filtering is to recommend trusted and satisfying information services to users [3,4].

The exiting recommender systems based on trust filtered recommendation information just through authenticating users' identity or removing users with low trust value [5]. They didn't consider the following four problems: 1) the role of transaction behavior for users in e-commerce has two types: buyer and seller, so we can't only use one trust value or reputation value to measure the users' trust level with different transaction behavior. Because the malicious users may use honest buy behavior to cover up the dishonest sell behavior, or vice versa [6]; 2) if recommendations are from different recommenders, we should treat them differently not only because of the recommenders' various trust levels but also because the recommenders have different knowledge to their recommendation [7]; 3) in order to make users quickly find resources, the time from making a request to receiving the response should short as much as possible

[8]; 4) after each transaction, both the participators not only should update their trust values or reputation values but also should share their trust information of transactions which can increase the spread of trust information and raise the performance of network.

2. Related Works

In Peer-to-Peer (P2P) e-commerce transaction, users exchange information or transact with others through direct communication, but those users don't know each other before and they also don't trust each other. And the openness of P2P system makes the users can't avoid others' malicious behaviors. Once users transact with one malicious node, they may incur substantial losses [9]. In order to solve the problem of security for the network service, M. Blaze proposed the concept of Trust Management firstly in 1996 [10], and its basic thought was to admit the imperfection of security information in open system. It proposed that making safety decision for system needed additional security information. Nowadays the researches on trust are mainly classified into two types, identity trust and behavior trust. The identity trust which is based on code, authentication protocol or digital signature technology checks entity's authenticity and makes the decision whether authorize the entity to access. But the behavior trust pays more attention to the trusted problem in broader meaning. According to the past behavior experiences, it updates the trust relationship between users dynamically and timely. International research indicates [11] that the network security is developing toward the direction of credible network. The future network security is the credible network with increasing credible behavior, which is a new consensus that is agreed by the network security research areas in recent years. The research on whether the users' behavior is credible not only increases the security of network through decreasing or avoiding transaction with malicious users but also improves the success rate of transactions and decreases the extra spending caused by monitor or precaution which are caused by distrust. Thereby the overall performance of the network is improved.

There have been lots of researches about behavior trust at home and abroad. Based on the transitivity of trust, the model named EigenREP was based on global reputation in P2P environment [12]. But its drawback is its astringency, high communication costs and relative global reputation, thus it can not evaluate whether the node is credible just through the value of global reputation. The model proposed in paper [8] presents that each grid domain is associated with multiple brokers and each broker with multiple entities. It eases the network traffic at the broker sites and makes the service providers' (SPs) response to the consumers' request significantly quicker, but it might lead to the information inconsistency maintained by different brokers and solving the problem has its costs. A trust model [13] based on behaviors was proposed to achieve the resource sharing and cooperation among different domains in grid environment, which dynamically reflects the entity's subject characteristic. But its limitations are that it doesn't update the trust value, so it cannot reflect the dynamics of trust computing. Traditional resource selection method [14–16] always selects service providers with the highest trust value. They don't consider whether the selected services are the services that consumer expected, that is to say whether the selected services and the expected services are accordant. Paper [17] proposed a similarity measurement about ontology-based semantic web services and paper [18] proposed a method of similarity search for web services. Both of them measure the similarity between the services with ontology and find the expected services to consumers. But the weakness is the high complexity of algorithm. Paper [3–5] proposed the idea of trust filtering in recommender systems, which considered that the recommenders should have similar tastes and preferences, should be trustworthy in the sense that they had a history of making reliable recommendations and should have different trust degree in entity trust and content trust. The weakness is that they don't solve the sparseness of similarity well when they find similar users. Paper [7] proposed a role-based recommendation and trust evaluation model which firstly takes the role of recommender into account. But it didn't present how to organize and storage a rational role hierarchy. A novel distributed trust model is propose in [6], which iteratively calculates for each node a global seller reputation value and a global buyer reputation value based on transaction history, and whether a node is credible or not can be identified from them. But it doesn't provide the computation of some coefficients.

To solve the problems of existing distributed trust model, the paper proposes a novel reputation-based multi-agent model for network resource selection (RMNRS). With nodes' identity and their recognition to services, the model computes for each node a global buyer reputation and a global seller reputation. And the model estimates whether one node is credible or not through the final Trust-Value which is the harmonic-mean of Trust-Value and similarity degree between request services and provided services. The model's characters are listed bellow.

1) Our model is based on trust domain which adopts multi-level trust management mechanism to manage agents belonging to different levels. The periodically communication between agents prevents the information's inconsistency between the agents.

2) Our model computes the similarity between request service and provided service by ontology. We want to find the most similar services to the requestor. Combined with the computing of Trust-Value, the services that we

supplied will be not only trusty but also matching to the request.

3) Our model takes the recognition value that how much the service provider knows the service into account, which makes the provided services more similar to the request.

4) Our model doesn't use one trust value to determine whether a node is credible or not. It keeps global buyer reputation value and global seller reputation value for each node, thus it can reflect node's different trust level with different transaction behavior.

5) After each transaction, the participators can share mutual trust information under certain condition, namely trust propagation.

6) Our model provides the computation of coefficient which is the weight when integrating two global reputation values.

This paper is organized as follows. Section 3 presents the related definitions, algorithms and fundamental principle of the model. Section 4 presents the simulation to validate our model's effectiveness and feasibility. And finally Section 5 concludes our work.

3. The Related Definitions, Algorithms and Fundamental Principle of the RMNRS Model

3.1. Related Definitions

Definition 1 (Trust): Trust is the subjective probability expectation of trustor to trustee's specific behavior which is relied on experiences and continuous to modify its value as the change of trustee's behavior. Paper [7] proposed that trust is a complex subject relating to an entity's belief in honesty, trustfulness, competence and reliability of another entity. Paper [19] proposed that trust is to believe others, which is established on their own knowledge and experiences and is a subjective behavior between entities. Trust is different from the belief that person believes in object things, which is a subjective judgment. The trust itself is not a fact or proof, it is acknowledge of observed fact. According to the different achieving trust way when entities interact with each other, Beth [20] divided the trust into direct trust and recommendation trust. To trust an entity directly means to believe in its capabilities with respect to the given trust class. Recommendation trust expresses the belief in the capability of an entity to decide whether another entity is reliable in the given trust class and in its honesty when recommending third entities.

Definition 2 (Trust-Domain): According to the Web-Based activities and related application, we divide the virtue network into numbers of self-government domains and define the self-government domain as Trust-Domain.

Definition 3 (Domain-Entity): Domain-Entity is the node or object who has some resources in network. It can be a user, service or resource. The interaction between entities has two types: the interaction of intra-domain and inter-domain.

Definition 4 (Transaction): One transaction is one interactive behavior which happens between two nodes when they need mutual services in the P2P network, such as one business dealing in e-commerce, one file download and so on. The buyer is the one who requests the services and the seller is the one who provides the requested services.

Definition 5 (Reputation): Reputation is the expectation of one entity's future behavior based on the observation of the entity's past behavior or evaluation information in transactions [9]. There are two types, local reputation and global reputation. The local reputation is defined as the expectation of a node's future behavior based on the past evaluation information which is provided by one of its buyers. The global reputation is defined as the expectation of a node's future behavior based on the past evaluation information which is provided by those nodes who had transacted with node j ago.

In this paper, each node has four types of reputation value: local buyer reputation, local seller reputation, global buyer reputation and global seller reputation.

Definition 6 (Trust-Value): The Trust-Value of Trust-Domain is defined as the mean of the Trust-Value of all entity-agents. The Trust-Value of entity-agent is the mean of the global reputation value of all the entities managed by the entity-agent. The Trust-Value of one entity is the weighted mean of global buyer reputation value and global seller reputation value of the entity.

Definition 7 (Identity): The Identity is not the role of entities in transaction. It refers to the identity symbol of one entity's recognition degree to one service in certain service domain, such as the social positions, social titles or certificates.

3.2. Trust Domain and Agent

In social network, everyone or group has his or her own interesting and joins in trusted communication circle, they have high credibility on the people or group in the same circle [21]. While in virtue network, because of the disparate resources, the sharing, cooperation and high performance of resources has become difficult. The wide connectivity of network requires to establish public and effective security mechanism between different nodes and peoples and to implement consistent security strategy. It also requires doing specific security management according to the application of multi-network. In this paper, we import the agent mechanism to abstract the

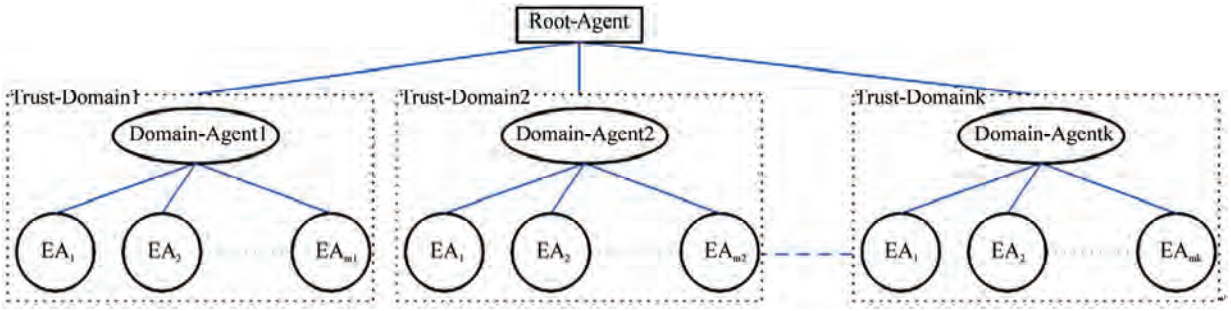


Figure 1. Trust-Domain and agent management architecture.

network, as shown in Figure 1 which implies the architecture of the model. We use this mechanism to manage the entities' trust computing and trust relationship between entities. Our model is not only manageable but also permits the managed object to cooperate independently, which is in accordance with the network computing mode and development tend [23].

We assume that there is an absolute trusted agent: Root-Agent (RA) who takes charge of every agent in each domain. Every domain has a Domain-Agent (DA) and numbers of Entity-Agents (EA)

1) The Root-Agent is the manager of global trust relationship of the system. It is an absolute trusted root. It manages and collects all the Domain-Agents' information and maintains the global trust relationship. The information that the RA maintains is the ID of DA, the Trust-Value of domain and semantic base.

2) The Domain-Agent is the manager of part trust relationship of the system and the trust relationship of domain. If there are entities request to join in the domain, the DA has the right that permits them to join in or judges which EA they belong to and it passes the information of entities to RA, so that RA can update the semantic base timely. When one EA's Trust-Value is less than threshold, the DA has the authority to retake the authority and awards the authority to other entities. According to the Trust-Value of the entity who wants to be the EA, DA makes the decision that whether it can be or not. The entity has the status to be an EA only when the Trust-Value of one entity is over the threshold. All the entities that an EA maintains provide the same or similar services. The DA collects EAs' information periodically (the period is decided by the size of system or the request of interaction) that it maintains and broadcasts the information as guide Trust-Value to each EA. The information that the DA maintains is the ID of EA and the direct transaction table (the domain-ID of entity, the entity-agent-ID of the entity, global buyer reputation value, global seller reputation value, service type, time, cost and the satisfaction degree of transaction).

3) The entities that maintained by Entity-Agent is organized as Binary Sorting Tree which has such characters: if the left sub-tree is not null then all the Trust-Value of entities in left sub-tree will be smaller than that of root; if the right sub-tree is not null then all the Trust-Value of entities in right sub-tree will be bigger than that of root; both the sub-tree are Binary Sorting Trees. We can add, delete, update and lookup needed information from the Binary Sorting Tree. The information that maintained by EA is illustrated as follows.

a) The storage structure of entities in Binary Sorting Tree.

```
Typedef struct BiTree {
DataType degree; // the Trust-Value of entity;
struct BiTree   * lchild,* rchild;   // the pointer of the left sub-tree and right sub-tree;
} BiTnode,*Bitree;
```

The binary tree is sorted by the trust degree of entities. The EA checks the Trust-Value of entities periodically and updates its position in the tree. If the Trust-Value of one entity is smaller than threshold, the EA can remove it from the tree. When one entity wants to join in, the EA puts it in the feat position.

b) The storage structure of entities in Binary Sorting Tree of entity's identity.

```
Typedef struct RoTree {
DataType Re-degree;        // the recognition degree of entity's identity;
struct BiTree   * lchild,* rchild;
}RoTnode,* Rotree;
```

The Binary Sorting Tree of entity's identity is established based on the recognition level of one entity to the services that it provides. The tree nodes are ordered by the recognition degree. The operation that we can do to the tree is similar to that of a). The reason that why the EA maintains the role tree is that it helps to select suitable entity.

c) Direct transaction table: the domain-ID of transaction entity, the entity-agent-ID of the entity, global buyer reputation value, global seller reputation value, service type, time, cost and the satisfaction degree of transaction.

The communication between entities is through Trusted Communication Agent Interface (TCAI) [22], which makes the communication between entities more

reliable. The agents pass information on to each other periodically. Once one entity's information was updated, it can push [23] its updated information to the related agents, which can avoid the information disaccording to others.

Our model is based on domain and has three levels, so we divide the trust relationship between entities into two types, the trust relationship of intra-domain and inter-domain.

1) The trust relationship of inter-domain: The evaluation objects of trust are based on domain. The trusted level of a domain is evaluated through the behavior that the entities showed in the transaction. Thus the trust level of the domain reflects all the entities' trust level in the domain.

2) The trust relationship of intra-domain: The trust of intra-domain is mainly the interaction between entities; the evaluation objects of trust are the entities in the domain. After the initialization, the trust level will be updated according to the behavior of entities in the following transaction.

3.3. The Fundamental Principles of RMNRS

This section mainly shows the basic principles of this model and the specific algorithms are listed in Subsection 3.4-3.9.

Step 1: The consumer sends the request which contains the service type and the mini-trust threshold to the agents.

Step 2: The agents select those nodes with higher global reputation value and calculate their integrated Trust-Value.

Step 3: The agents firstly calculate the similarity degree between buyer and Candidates, and then the agents calculate the harmonic-mean of similarity degree and integrated Trust-Value.

Step 4: If there are entities whose final Trust-Value (harmonic-mean) is bigger than the mini-trust threshold, go to Step 5, conversely go to Step 11.

Step 5: If the number of candidates is over one, go to Step 6, on the other hand, go to Step 7.

Step 6: Under the same credible condition, we select the nodes with the highest global buyer reputation.

Step 7: The buyer transacts with the selected entity and both parties give each other the satisfaction degree to this transaction.

Step 8: Both their entity-agents and domain-agents will update their reputation value and transaction table after transaction.

Step 9: According to the satisfaction degree, if both parties want to share their trust information, go to Step 10, on the other hand, go to Step 13.

Step 10: Both parties share their trust information of transaction.

Step 11: The agent ask the buyer whether it wants to modify the mini-trust threshold, so that it can find the suitable entity to transact with. If the user wants to decrease the threshold, go to Step 1, and on the other hand go to Step 12.

Step 12: The transaction is failure.

Step 13: The transaction is success.

3.4. The Initialization of Reputation

When one entity didn't have any interactions with other entities ago, should we trust it? In this paper we show several methods. According to specific environment you can select suitable method.

1) You can set the initial global buyer reputation value and global seller reputation value to be $\omega_1 \in [0,1]$ and $\omega_2 \in [0,1]$ separately.

2) According to the security information (such as the information of identity [7]) that the user provides, you can convert the information into the initial reputation value of entity through the function, where R is the set of role information and D is the domain that R belongs to.

After the initialization, the global reputation value will be updated according to the behavior of entities in the following transaction. To simplify the experiment, we set both the initial global buyer and seller reputation value of entity to be 0.5.

3.5. The Computation of Local Reputation Value

According to the definition of the local reputation, the value of local reputation is calculated in the light of the historical transaction evaluation that the buyer sent to the seller. After each transaction, both the participators feed back the degree of satisfaction to each other. And the user also should give out a threshold, so if the degree of satisfaction is higher than the threshold, we think that this transaction is satisfactory, on the other hand, we think it is unsatisfactory. We use the percentage of the number of satisfactory transactions in all transactions to represent the local reputation. And we let Lb_{ij} represent the local buyer reputation, which mean that the local reputation given by node i as the buyer to the node j as the seller and let Ls_{ij} represent the local seller reputation, which mean that the local reputation given by node i as the seller to node j as the buyer. The formula is illustrated as follows.

$$Lb_{ij} = \frac{N_{bg}(i,j)}{N_{bg}(i,j) + N_{bb}(i,j) \times N_{punish}} \tag{1}$$

$$Ls_{ij} = \frac{N_{sg}(i,j)}{N_{sg}(i,j) + N_{sb}(i,j) \times N_{punish}}. \tag{2}$$

where $N_{bg}(i, j)$, $N_{bb}(i, j)$, represent the number of satisfactory transaction and the number of unsatisfactory transaction between buyer i and seller j. $N_{sg}(i,j)$, $N_{sb}(i,j)$ represent the number of satisfactory transaction and the number of unsatisfactory transaction between seller i and buyer j. N_{punish} is the coefficient which is the punishment to malicious transaction. Especially when the cost of the transaction is very high, the punishment coefficient should be bigger. So the formula is

$$N_{punish} = ((\alpha+1)\times \sum_{k=1}^{N_{bb}(i,j)} C_{mk}(i,j) - \alpha) \Big/ \sum_{k=1}^{N_{bb}(i,j)} C_{mk}(i,j),$$

where $C_{mk}(i,j)$ is the cost of k^{th} unsatisfactory transaction between node i and node j. We assume all the cost of transaction is bigger than 1. On the contrary, we set it to be 1; The formula not only reflects the higher the cost is, the bigger the weight is, but also satisfies the monotonic increasing with the increasing transaction cost and its result is in the cope of [1,1+α], where α is the regulatory factor. So that the user can adjust the scope of punishment according to his needs.

On the basis of the analysis above, the properties of the local reputation formula are shown as follows.

1) The model evaluates whether the node is credible or not directly. The more the number of satisfactory transaction is, the closer to 1.0 the reputation value is. The more the number of unsatisfactory transaction is, the closer to 0.0 the reputation value is.

2) The introduction of the punishment coefficient $N_{punish} \geq 1$ makes the reputation value decrease quicker than rise. It embodies the punishment to the malicious nodes and especially to those nodes who make use of high cost transactions to cheat.

3.6. The Computation of Global Reputation Value

The global reputation value is the integrated evaluation of one node, which is obtained form those nodes who had transacted with the node. We let Gb_i denote the global buyer reputation, where node i is the buyer and Gs_i to be the global seller reputation, where node i is the seller. But there are some factors we should consider when we calculate the value of global reputation.

1) The global buyer (seller) reputation of node i should be the integrated evaluation of all those nodes who had transacted with node i.

2) The evaluation of different nodes with different Trust-Value should be kept separate. That is because the evaluation of the nodes with higher Trust-Value is more important than that of nodes with lower Trust-Value.

3) The more the number of transaction between both parties is, the more credible the evaluation is.

4) The global reputation is an accumulative process, so only through persistent credible transaction, the value will be higher.

The formulas are listed as follows.

$$Gb_i(k+1) = \frac{\sum_{j\in Vsi} Gs_j(k)\times(1-e^{-\frac{Nsg(j,i)-Nsb(j,i)}{5}})\times Ls_{ji}}{\sum_{j\in Vsi} Gs_j(k)\times(1-e^{-\frac{Nsg(j,i)-Nsb(j,i)}{5}})} \quad (3)$$

$$Gs_i(k+1) = \frac{\sum_{j\in Vbi} Gb_j(k)\times(1-e^{-\frac{Nbg(j,i)-Nbb(j,i)}{5}})\times Lb_{ji}}{\sum_{j\in Vbi} Gb_j(k)\times(1-e^{-\frac{Nbg(j,i)-Nbb(j,i)}{5}})} \quad (4)$$

where the meanings of $N_{sg}(j,i)$, $N_{sb}(j,i)$, $N_{bg}(j,i)$ and $N_{bb}(j,i)$ are the same as that of Chapter 3.5. Vb_i represents the set of nodes who had transacted with node i and they are buyer. Vs_i represents the set of nodes who had transacted with node i and they are seller.

The weighted average method not only can embody the views of all the nodes, but also keep the meaning of global reputation unchanged. The character of 1-exp(-($N_g(j,i)$-$N_b(j,i)$)/5) negative exponent increase as $N_g(j,i)$-$N_b(j,i)$ in accordance with the feature that credible transaction can improve the reputation and incredible transaction can decrease the reputation.

3.7. The Computation of Trust-Value

The Trust-Value of node j represented as T_{ij} is the critical factor for node i to determine whether to transact with node j. Because it is the integrated value of Ls_{ji} and Gs_j, its computation must consider the trust level of node i to j's local reputation and global reputation. If node i has transacted with node j, the Trust-Value of node j is the weighted sum of Ls_{ji} and Gs_j. If node i didn't have any transactions with node j before, node i can ask his entity-agent and domain-agent to recommend providers. Both the agents search those entities that had transacted with node j and ask them to recommend node j. Those recommenders give views about node j according to the history performance of node j. Then the agents feed back these information to the requestor i. To ensure the validity of recommendation, the rules of recommendation are shown as follows.

1) The recommendations have time limitation, the scope of time is [σ_1, σ_2] which can be defined as the transaction needs. That is because the previous transaction information can not exactly reflect the *SPs*' credibility of the current situation.

2) Because the recommendation is finite, the depth that the agent searches recommenders in the binary sorting tree must be smaller than h.

3) In order to reduce the likelihood of collusion, we set the number of recommenders must more than certain

threshold ϑ where $\vartheta > 0$. If the number of recommenders is less than ϑ, we can add some virtue nodes as recommender whose reputation is the same as the initial Trust-Value and the number of nodes that had transacted with each recommender is $+\infty$.

Because of the asymmetry of trust, the recommendation involves the consumer's trust to the recommender's recommendation. The higher trust level the consumer is to the recommender, the more trust the consumer is to the recommendation. If recommender has high recognition about his services, the trust level of his recommendation also can be increased. We let T_{ij} denote the Trust-Value of node *j* as opposite to node *i*. The formula is shown as follows.

$$T_{ij} = \frac{\sum_{j \in Dir} (\alpha Ls_{ji} + \beta Gs_{ji})}{\|Dir\| + \|Undir\|} + \frac{\sum_{r \in Undir} T_{ir} \times rec_r \times (\alpha Ls_{jr} + \beta Gs_{jr})}{\|Dir\| + \|Undir\|} \quad (5)$$

where *Dir* is the set of nodes who had transacted with node *i* directly and *Undir* is the set of nodes who are recommended by recommender *r* to transact with node *i*. The $\|Dir\|$ and $\|Undir\|$ represent the length of the corresponding set. $rec_r \in [0,1]$ is the recognition degree of recommender *r* to service. α and β is the weighting of local reputation and global reputation separately and $\beta=1-\alpha$. The value of α is computed as follows.

The local reputation is the evaluation of one node according to his history transaction behavior with another node. So if the number of transactions is more, the transaction cost and the evaluation of the transaction are higher, the buyer will more trust the local reputation which is achieved through his direct experiences.

$$\alpha = \sqrt{(\sum_{j=1}^{m} C_{mbig} \Big/ \sum_{i=1}^{n} C_m) \times (\sum_{j=1}^{m} Eva_{good} \Big/ \sum_{i=1}^{m} Eva)}$$

where *n* is the total number of transaction, *m* is the number of transactions with high cost and good evaluation. The user can defined a threshold to determine the value of m. C_m is the cost of transaction and C_{mbig} is the high cost of transaction. Eva is the evaluation of transaction, Eva_{good} is the good evaluation of transaction and *Eva*, $Eva_{good} \in [0,1]$.

3.8. The Harmonic-Mean of Trust-Value and Similarity Degree

The traditional recommender systems based on trust believe that the higher the global reputation of recommender is, the more credible the recommendation is. But in fact the value of global reputation is not in conformity with the importance of recommendation. For example, some malicious nodes may achieve high global reputation value through fake or some nodes collude in order to remote their Trust-Value or to destroy other competitor. Thus the recommender not only has the high trust degree of recommendation but also should make sure that the content of recommendation is credible. In this paper, in order to ensure the content of recommendation is credible, the model imports the semantic-based service matching method which calculates the similarity degree between the request service and the provided service. The service is reliable only when the provided service satisfies the user's needs. The computation of similarity adopts one simple and efficient method provided by paper [24], denoted as *WS*.

This paper uses the similarity degree to harmonize the Trust-Value with weighting. The higher the similarity degree is, the bigger the harmonic-mean is. The harmonic-mean is the reciprocal of the arithmetic mean of the reciprocals, which is mainly used to the situation that the initial digitals is not the direct initial digitals but its frequency had been computed. The formula is shown as follows.

$$ST_{ij} = \frac{2 \times WS_{ij} \times T_{ij}}{\delta \times T_{ij} + (1-\delta) \times WS_{ij}} \quad (6)$$

where δ is the adjustment factor; WS_{ij} is the similarity degree of the request service and provided service; T_{ij} is the trust level that how much node *i* trust in node j; ST_{ij} is the harmonic-mean of similarity and Trust-Value which directly determines whether node *i* transacts with node *j* or not.

3.9. The Updates of Trust-Value and the Sharing of Trust Information

After each transaction, both parties will feed back the evaluation of this transaction to each other, namely the satisfaction degree. Both the entity-agents of two parties need to respectively update the global buyer reputation (Gb_i) of buyer and the global seller reputation (Gs_j) of seller. The updates in time ensure that the users don't need to calculate the reputation value when they send the request, so that the model can respond the user's request quickly. The formulas are listed as follows.

$$Gb_i^{k+1} = \varphi \times Gb_i^k + (1-\varphi) \times S_k(i,j) \times C_k, \quad (7)$$

$$Gs_j^{k+1} = \phi \times Gs_j^k + (1-\phi) \times S_k(j,i) \times C_k \times rec_k \quad (8)$$

where $S_k(i,j) \in [0,1]$ represents the satisfaction degree of buyer *i* to seller *j* after the k^{th} transaction. $S_k(j,i) \in [0,1]$ represents the satisfaction degree of seller *j* to buyer *i* after the k^{th} transaction. The satisfaction degree indicates the level of satisfaction achieved by the consumer on the

service provided by the *SP*. A normalized value between 0 and 1 is used, with 1.0 indicating 100% satisfaction and 0.0 indicating the lowest satisfaction. $C_k = (C_{mk}-1)/C_{mk}$ represents the weight of transaction cost in the k^{th} transaction. $rec_k \in [0,1]$ is the recognition degree of entity to the service that he provided in k^{th} transaction which is maintained in the binary sorting tree of entity's identity by his entity-agent and is registered by entity when he joined the domain. $\varphi, \phi \in [0,1]$ are the weights which are assigned according to last transaction time.

In addition to this, we also need to update the trust tables which are related to the entity and maintained by EA and DA. The update method is pushing the information to the needed agent.

Updating the reputation value in time can effectively reduce the response time that the buyer waits from his agents and it is done among free time without influencing the transaction. Thus the user can use the reputation value directly only with time decay such as linear decline, exponential decline and so on.

Both the transaction parties can set a threshold to determine whether to share trust information or not. That is to say, they share their trust information only when their Trust-Value is more than the threshold. The sharing of trust information makes the trust propagate quickly and improves the performance of network.

4. Simulations and Analysis

In this paper, the simulation is based on the PeerSim which is written in the Java language and is based on components [25]. It can support the extensibility and dynamic of the P2P network better. And it adopts the modular design and uses the configuration file to custom the modules and parameters. Thus it has high expandability. It also provides the interfaces and statistical methods to generate the network and makes the simulation and the evaluation of one algorithm more easily.

The principle of organization of the RMNRS model makes us not use the existing protocols of PeerSim directly. We are obliged to inherit one existing protocol and write our own algorithm to simulate. The protocol that we inherit is: Idle Protocol-Average Function. The control is cycle-based. In this paper we just inherit the interconnect relation between nodes and we write the algorithms about the model of EigenRep and RMNRS autonomously.

According to the character of the nodes in the network, we divide the nodes into four types: absolute trust nodes, trust nodes, critical trust nodes and distrust nodes.

Definition 1: absolute trust is the trust that is established on the basis of both parties in the partnership experience long-term transaction and major event test;

Definition 2: trust is the trust that is established on the basis of both parties in the partnership experience long-term transaction and general event test;

Definition 3: critical trust is the trust that both parties in the partnership don't have sufficient reason to trust or distrust each other;

Definition 4: distrust is the trust that after the long-term transaction, both parties in the partnership don't trust one at least.

We define the absolute trust node's value to be T where $T \geq T_{max}$, trust node's value to be T where $T_{mid} < T < T_{max}$, critical trust node's value to be T where $T_{min} \leq T \leq T_{mid}$, distrust node's value to be T where $T < T_{min}$, and $0 \leq T_{min} \leq T_{mid} \leq T_{max} \leq 1$.

4.1. Experiment 1

According to the paper, we define the network size is 1000, the simulation cycle is 50, the degree of a node is 6 and the init global reputation value is 0.5. We assume that T_{min}=0.3, T_{mid}=0.6and T_{max}=0.9. We define the percent that the four type nodes respectively are 10%, 30%, 30% and 30%. All of the nodes belong to different trust domains. Each domain has four types of nodes. Entities' Trust-Value will be changed timely along with the transaction in the domain or between domains. Along with the increase of the number of transactions the variation of the four types of entities' average Trust-Value is showed in Figure 2.

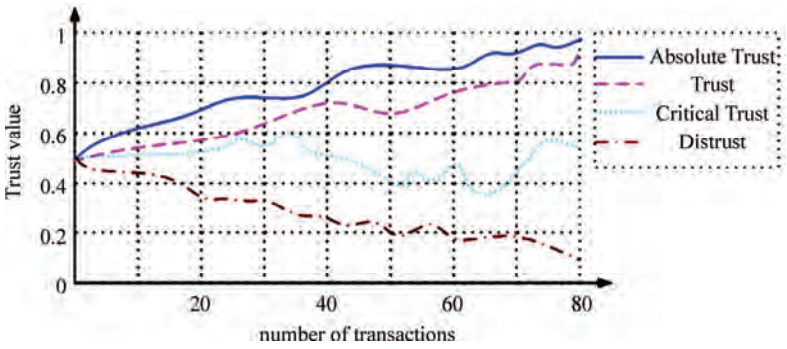


Figure 2. The variation of the average Trust-Value of four types of nodes.

As shown in Figure 2, the initial reputation value of all nodes is 0.5 and along with the increasing number of trades, the model updates the reputation value according to the Equation (5). Finally we calculate the Trust-Value of each node. According to the definitions of the four types of nodes, we can find that the result of RMNRS model is in accordance with the expected analysis. And along with the increase of the number of trades, it well reflects that the variation of the nodes' Trust-Value.

4.2. Experiment 2

The assumption of nodes is the same as the Experiment 1 in Chapter 4.1. In this simulation, we assume the degree of each node is 3, that is to say, the direct transaction table of each node maintains three nodes' trust information in the primary stages and the three nodes are selected randomly by the PeerSim protocols. When we write the program, we let the nodes' initial reputation value be random which is between 0 and 1. The experiment is designed to record the selection number from the node sending out the request to it receives the request. The network consumption is based on the number of selection step when the consumer selects the transaction partner.

In this model, we allocate two global reputations for each node. Thus the user can select the nodes with higher global seller reputation in the first stage which expands the range of selection. Then the model computes the harmonic mean of the similarity degree and the integration Trust-Value. The final Trust-Value sincerely reflects whether the provided service satisfies the users' needs, which can reduce the unnecessary selection. However the EigenRep model just follows the traditional method to select the provider and the comparison number is more than the RMNRS model. Figure 3 realistically reflects the analysis of this paper and verifies the efficiency of the RMNRS model.

4.3. Experiment 3

Whether the transaction is success or not is decided by the satisfaction degree which is fed back by user. If the satisfaction degree is larger than 0.6 then this transaction is success and vice versa. We define the success ratio of transaction to be the proportion of the success number.

In this simulation, we assume that the absolute trust nodes provide the credible services with the probability of 100%. And we also assume that the RMNRS and the EigenRep select the absolute nodes with the probability of 80%. As seen in Figure 4, when there are not malicious nodes in the environment, the success ratio of transaction is 95%. Along with the increase of malicious nodes, the transaction success ratio of EigenRep declines obviously. When the ratio of malicious nodes is 50%, the transaction success ratio of EigenRep is only about 60%. This is because there is lack of punishment to the malicious nodes in the EigenRep. So its success ratio has bigger drop. The RMNRS punishes the malicious nodes and matches the services between the request and the provided services. And it uses the similarity degree to amend the Trust-Value with the harmonic mean which ensures that the provided service is the needed service. The RMNRS avoids the transaction with the malicious nodes and improves the success ration of transaction. Under the condition of existing malicious nodes with the property of 50%, the transaction success ration of the RMNRS is about 80%. The experiment verifies the feasibility and efficiency of the RMNRS.

5. Conclusions

In this paper we present a reputation-based multi-agent model for network resource selection (RMNRS) which prevents the inconsistency of information maintained by different agents through the periodically communication between the agents. The model enables the consumer to receive the response from the agent significantly more

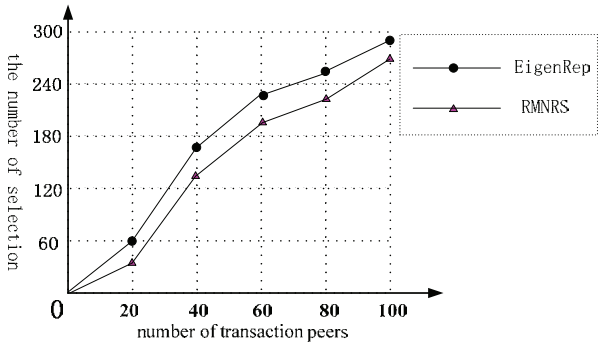


Figure 3. The comparison between RMNRS and EigenRep in the network consumption.

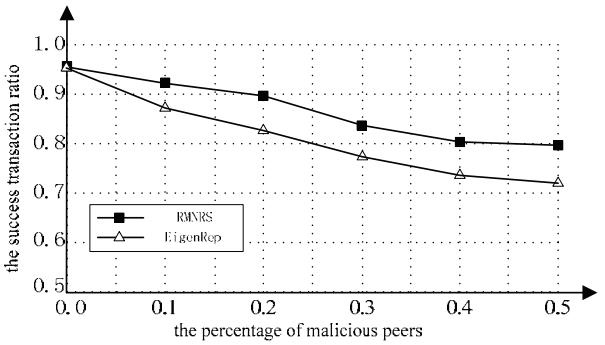


Figure 4. The comparison between RMNRS and EigenRep for success ratio of the transaction.

quickly than that of traditional models because the global reputation values of both parties are evaluated and updated dynamically after the completion of each transaction. And the model allocates two global reputation values to each entity and takes the recognition value that how much the service providers know the service into account. In order to make users choose the best matching services to their request and give users with trusted services, the model also takes the similarity between services into account and uses the similarity degree to amend the integration reputation value with the harmonic-mean. The following work is to research how to avoid the sharing of incredible information with malicious nodes and how to punish those malicious nodes.

6. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 60873203), the Natural Science Foundation of Hebei Province (Grant No. F2008000646) and the Guidance Program of the Department of Science and Technology in Hebei Province (Grant No. 072135192).

7. References

[1] H. Ibrahim, P. K. Atrey, and E. S. Abdulmotaleb, "Semantic similarity based trust computation in websites," International Multimedia Conference, New York, ACM, pp. 65–72, 2007.

[2] S. K. Chong and J. H. Abawajy, "Feedback credibility issues in trust management systems," 2007 International Conference on Multimedia and Ubiquitous Engineering: proceedings: MUE'07, Los Alamitos, Calif., IEEE Computer Society, pp. 387–391, 2007.

[3] Donovan J O and Smyth B, "Trust in recommender systems," in proceedings of the 10th international conference on Intelligent user interfaces, New York, ACM, pp. 167–174, 2005.

[4] P. Massa and P. Avesani, "Trust-aware recommender systems," in proceedings of the 2007 ACM conference on Recommender systems. New York, ACM, pp. 17–24, 2007.

[5] Y. Gil and D. Artz, "Towards content trust of web resources," in proceedings of the 15th international conference on World Wide Web, New York, ACM, pp. 565–574, 2006.

[6] D. S. Peng, C. Lin, and W. D. Liu, "A distributed trust mechanism directly evaluating reputation of nodes," Journal of Software, Vol. 19, No. 4, pp. 946–955, April 2008.

[7] Y. Wang and V. Varadharajan, "Role-based recommendation and trust evaluation," in the 9th IEEE International Conference on E-Commerce. Technology and the 4th IEEE International Conference on Enterprise Computing, E–Commerce and E–Services, Tokyo, IEEE, pp. 278–288, 2007.

[8] P. Varalakshmi, S. Thamarai Selvi and M. Pradeep, "A multi-broker trust management framework for resource selection in grid," in Communication Systems Software and Middleware, COMSWARE'07, 2nd International Conference on Bangalore, IEEE, pp. 7–12 January 2007.

[9] S. X. Jiang and J. Z. Li, "A reputation–based trust mechanism for p2p e-commerce systems," Journal of Software, Vol. 18, No. 10, pp. 2551–2563, 2007.

[10] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized trust management," in proceedings of the 17th Symposium on Security and Privacy, CA, IEEE Computer Society Press, pp. 164–173, 1996.

[11] C. Lin, L. Q. Tian, and Y. Z. Wang, "Research on user behavior trust in trustworthy network," Journal of Computer Research and Development, Vol. 45, No. 12, pp. 2033–2043, 2008.

[12] S. D. Kamvar and M. T. Schlosser, "EigenRep: Reputation management in P2P networks," in Lawrence S, ed. Proceedings of the 12th International World Wide Web Conference Budapest, ACM Press, pp. 123−134, 2003.

[13] C. F. Wang and F. C. Sun, "Hierarchical entity self-determined trust model based on behaviors in grid environment," Computer Engineering and Applications, Vol. 43, No. 16, pp. 135–138, 2007.

[14] F. Maheswaran and M. Maheswaran, "Evolving and managing trust in grid computing systems," in proceedings of the 2002 IEEE Canadian Conference on Electrical Computer Engineering, pp. 1424–1429, 2002.

[15] F. Azzedin and M. Maheswaran, "A trust brokering system and its application to resource management in public resource grid," Parallel and distributed Computing Symposium. pp. 22, 2004.

[16] X. Li and L. Liu, "Peertrust: supporting reputation-based trust for peer-to-peer electronic communities," IEEE Transactions on Knowledge and Data Engineering, Special Issue on Peer to Peer Based Data Management, Vol. 16, No. 7, pp. 843–857, 2004.

[17] X. Dong, A. Halevy, J. Madhavan, *et al*., "Similarity search for web services," in Proceedings of the Thirtieth International Conference on Very Large Data Bases, VLDB Endowment, pp. 372–383, 2004.

[18] X. Wang, Y. H. Ding, and Y. Zhao, "Similarity measurement about ontology-based semantic web services," in Conjunction with 4th European Conference on Web Services (ECOWS'06), pp. 4–6, 2006.

[19] X. Y. Li and X. L. Gui, "Research on dynamic trust model for large scale distributed environment," Journal of Software, Vol. 18, No. 6, pp. 1510−1521, 2007.

[20] T. Beth, M. Borcherding and B. Klein, "Valuation of trust in open networks," in proceedings of the European Symposium on Research in Computer Security, Brighton UK, Springer-Verlag, pp. 3–18, 1994.

[21] X. Z. Zhang, "The research of virtual enterprise trust mechanism—the innovation of trust management in the network environment," Hunan, Hunan People's Publishing House, July 2005.

[22] Steve Hanna, Co-Chair, TNC Work Group, TCG. TNC: Open Standards for Network Access Control. https://www.trustedcomputinggroup.org.

[23] J. F. Tian, B. Xiao, X. X. Ma, *et al*, "The trust model and its analysis in TDDSS," Journal of Computer Research and Development, Vol. 44, No. 4, pp. 598–605, April 2007.

[24] X. Qing, "Semantic-based web service discovering algorithm," Master's degree, Jinan, Shandong University, pp. 17–26, 2006.

[25] PeerSim: A simulation environment for P2P protocols in java. Version 1.0.4. 2008

Int. J. Communications, Network and System Sciences, **2009, 2, 775-785**
doi:10.4236/ijcns.2009.28090 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Subcarrier Availability in Downlink OFDM Systems with Imperfect Carrier Synchronization in Deep Fading Noisy Doppler Channels

Litifa NOOR[1], Alagan ANPALAGAN[1], Sithamparanathan KANDEEPAN[2]
[1]*WINCORE Research Lab, Ryerson University, Toronto, Canada*
[2]*Wireless Signal Processing Group, National ICT, Australia*
Email: *alagan@ee.ryerson.ca, lnoor@ee.ryerson.ca*
Received July 11, 2009; *revised August* 30, 2009; *accepted September* 19, 2009

Abstract

Multicarrier systems such as orthogonal frequency division (OFDM) are considered as a promising candidate for wireless networks that support high data rate communication. In this article, we investigate the performance of a multiuser OFDM system under imperfect synchronization. Analytical results indicate that the SNR degrades as the average power of the channel impairments such as AWGN, carrier frequency offset due to Doppler frequency and fading gain is increased. The SNR degradation leads to imperfect synchronization and hence decreases the total number of subcarriers available for allocation. Monte Carlo analysis shows up to 22% loss in the number of allocatable subcarriers can be expected under a specific imperfect synchronization condition as compared to perfect synchronization. We utilize empirical modelling to characterize the available number of subcarriers as a Poisson random variable. In addition, we determine the percentage decrease in the total number of allocatable subcarriers under varying channel parameters. The results indicate 19% decrease in the number of available subcarriers as average AWGN power is increased by 10dB; 44% decrease as the Doppler frequency is varied from 10Hz to 100Hz; and 56% decrease as the fading gain is varied from 0dB to -30dB.

Keywords: OFDM, Subcarrier Availability, Synchronization

1. Introduction

The growing demand for high-speed wireless communications has led to the investigation of spectrally effcient systems for downlink transmission in multicarrier systems [1]. Multicarrier systems such as Orthogonal Frequency Division Multiplexing (OFDM) enable the network to provide high data rate communication by using adaptive subcarrier allocation. Although high data rate communication is subject to inter-symbol interference (ISI) due to dispersive nature of wireless channels, OFDM enables the radio network to support high data rates while reducing the effect of ISI [2,3].

However, OFDM based systems are sensitive to carrier frequency offset (CFO), which leads to the loss of orthogonality between the subcarriers and thus introduction of inter-channel interference (ICI) [4]. CFO is caused by differences in transmitter and receiver oscillator frequencies, Doppler frequency due to relative mobility between transceivers and phase noise introduced by non-linear channels. Since accurate frequency synchronization is very important for reliable signal reception [5], OFDM systems utilize different synchronization schemes to facilitate acquisition and tracking of carrier frequency. However, the distortion inherent in wireless channels requires special design techniques and rather sophisticated adaptive coding and modulation algorithms to achieve accurate synchronization. Different methods have been suggested to reduce the effect of CFO [6–8], but obtaining perfect synchronization in wireless channels remains a challenging task. Hence, in this article we investigate the performance of multiuser OFDM under imperfect synchronization.

In the literature, various subcarrier allocation algorithms have been presented for multiuser OFDM systems. Many of these algorithms support high aggregate data

*This work was supported in part by a grant from National Science and Engineering Research Council of Canada..

rate and maximize system capacity [9–14]. However, the performance improvement is achieved when the system is based on perfect synchronization and the utilization of instantaneous channel information to adaptively allocate subcarriers. Assuming that the channel variation in a frequency selective fading environment is independent of each other, adaptive subcarrier allocation based on instantaneous channel information is an effective method to resource allocation in multiuser OFDM systems. However, obtaining the instantaneous channel condition under hostile wireless channels is not attainable in practical systems. Hence identifying the subcarriers with relatively low SNR and avoiding allocation of such subcarriers is a more practical approach to subcarrier allocation.

In this article, we determine the percentage loss in the available number of subcarriers under imperfect synchronization in comparison to perfect synchronization. The frequency synchronization scheme used is the open-loop maximum likelihood (ML) estimator. The performance of synchronization depends on noise, Doppler frequency and deep-fades in the channel that reduce the effective SNR associated with the subcarrier, which in turn degrades the allocatability of the subcarrier. The purpose of evaluating the system performance under imperfect synchronization is to restrict allocation of the subcarriers that are not suitable for transmission. Although the number of allocatable subcarriers as compared to perfect synchronization decreases which translates to a decrease in the aggregate data rate for a given number of users, the BER performance of the system could be improved by avoiding allocation on subcarrriers that are not suitable for transmission. Hence, any subcarrier allocation algorithm can be utilized while considering the variations in the total number of subcarriers. To the best of our knowledge, this is the first paper to characterize the subcarrier availability in deep fading noisy Doppler channels.

The organization of this article is as follows. In Section 2, the OFDM system model is discussed. In Section 3, we discuss the frequency synchronization utilized for the OFDM system. Following, a detailed analysis of SNR is provided in Section 4. In Section 5, we discuss the number of available subcarriers under imperfect synchronization and model the statistical characteristics of allocatable subcarriers. Following, in Section 6 we evaluate the number of available subcarriers under variable SNR. In Section 7, the paper is concluded.

2. System Model

Figure 1 shows the baseband equivalent model of the OFDM system that is being considered in this paper. This analysis is independent of mapping of the transmitted data as complex values $a_{0,i} \ldots a_{N-1,i}$, and is therefore applicable to all forms of modulation which is utilized in OFDM systems. As indicated in Figure 1, the IFFT is performed on the complex data symbols $a_{k,i}$ for $k = 0, 1, \ldots, N-1$, to produce the time-domain samples $b_{n,i}$ for $n = 0, 1, \ldots, N-1$ as follows:

$$b_{n,i} = \frac{1}{N}\sum_{k=0}^{N-1} a_{k,i} e^{j2\pi(\frac{n}{N})k} \tag{1}$$

where N is the number of data samples. The OFDM symbol $b_{n,i}$ is transmitted through the frequency selective Doppler channel. The frequency response of the channel is indicated by:

$$H_K = \alpha_K e^{j2\pi\frac{k}{N}(\Delta fT)} + \eta_k \tag{2}$$

where α_k is the (Rayleigh) fading gain for the kth subcarrier, $e^{j2\pi\Delta fT}$ is the CFO due to Doppler frequency with Δf indicating the Doppler frequency and T indicating the symbol period, and η_k is the Additive White Gaussian Noise (AWGN) component on the kth subcarrier.

The system is based on the characteristics of multiuser frequency-selective fading channels where different subcarriers are subject to different fading levels and the channel variations in a multiuser environment are independent of each other. To ensure frequency selectivity, the coherence bandwidth of the channel, which is the reciprocal of the multi-path spread, is assumed to be smaller in comparison to the bandwidth of the transmitted signal.

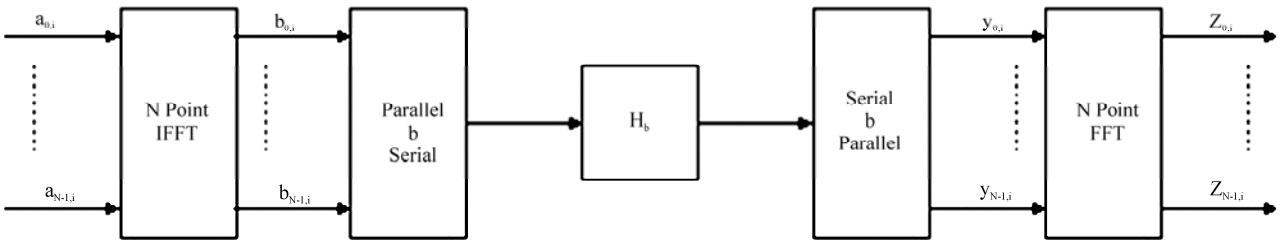


Figure 1. Baseband Equivalent OFDM System Model.

Under the assumption of AWGN channel, the received signal with the frequency offset and the fading gain is given as:

$$\begin{aligned} y_k &= \sum_{n=0}^{N-1} \alpha_k a_k e^{j2\pi(\frac{n}{N})k} e^{j2\pi(\frac{\Delta fT}{N})k} + \eta_k \\ &= \sum_{n=0}^{N-1} \alpha_k a_k e^{j2\pi(\frac{n}{N}+\frac{\Delta fT}{N})k} + \eta_k \end{aligned} \tag{3}$$

The received signal on the *kth* subcarrier and in the *ith* symbol period can be written as:

$$y_{k,i} = a_{k,i}\alpha_{k,i} e^{j2\pi N(\Delta fT)} + \eta_k \ , \tag{4}$$

where $a_{k,i}$ is the transmitted data on the *kth* subcarrier in the *ith* symbol period, and $\Delta f\ T$ is the normalized frequency offset.

The received signal after FFT is expressed as:

$$\begin{aligned} z_{m,i} &= \sum_{k=o}^{N-1} y_{k,i} e^{-j2\pi\frac{km}{N}} + \eta_m \\ &= \frac{1}{N}\sum_{l=o}^{N-1} \alpha_{l,i} a_{l,i} \sum_{k=0}^{N-1} e^{j2\pi\frac{k}{N}(l-m+\Delta fT)} + \eta_m \end{aligned} \tag{5}$$

Using the properties of geometric series, $z_{m,i}$ can be expressed as:

$$z_{m,i} = \frac{1}{N}\sum_{l=0}^{N-1} \alpha_{l,i} a_{l,i} \frac{\sin\pi(l-m+\Delta fT)}{\sin\frac{\pi}{N}(l-m+\Delta fT)} e^{j\frac{N-1}{N}(l-m+\Delta fT)} + \eta_m \tag{6}$$

The analysis of ICI can be simplified by defining N complex weighting coefficients, c_0, ..., c_{N-1}, which give the contribution of each of the N point values $a_{0,I}$, ...$a_{N-1,i}$ to the output value. Based on this, $z_{m,i}$ is written as:

$$z_{m,i} = c_o \alpha_{m,i} a_{m,i} + \sum_{l=0,l\neq m}^{N-1} c_{l-m} \alpha_{l,i} a_{l,i} + \eta_m \tag{7}$$

where the first term is the desired signal and second term is the ICI. c_o is the attenuation factor, $a_{m,i}$ is the transmitted data on the *mth* subcarrier in the *ith* symbol period, and $\alpha_{m,i}$ is the (Rayleigh) fading gain on the *mth* subcarrier during *ith* symbol period.

Using geometric series expansion, $z_{m,i}$ can be written as:

$$z_{m,i} = \frac{1}{N}\frac{\sin\pi(\Delta fT)}{\sin\frac{\pi}{N}(\Delta fT)} e^{j\pi(\frac{N-1}{N})(\Delta fT)} \alpha_{m,i} a_{m,i} + \sum_{l=0,l\neq m}^{N-I} \frac{1}{N}\frac{\sin\pi(l-m+\Delta fT)}{\sin\frac{\pi}{N}(l-m+\Delta fT)} \times \alpha_{l,i} a_{l,i} e^{j\pi(\frac{N-1}{N})(l-m+\Delta fT)} + \eta_m \tag{8}$$

For noiseless case, when $\Delta f\ T = 0$, $z_{m,i} = a_{m,i} + \eta_m$ which indicates the transmitted data plus noise. In the case where $\Delta f \neq 0$, the transmitted data is subject to attenuation and ICI.

In addition, the relationship between the attenuation component $c0$ and $\Delta f\ T$ indicates that attenuation in the desired signal component increases as the $\Delta f\ T$ is increased. Given that an increase in CFO is caused by an increase in the Doppler frequency, it can be seen that as the Doppler frequency is increased the attenuation in the desired signal increases leading to a decrease in the SNR.

The $\alpha_{m,i}$ is the (Rayleigh) fading gain for the *mth* subcarrier in the *ith* symbol period which attenuates the desired signal, which in turn reduces the SNR. Each subcarrier experiences a different level of fading and the maximum number of subcarriers that are in deep fade are dependent on the channel condition and vary every transmission time. The deep-fading in the channel degrades the corresponding subcarriers by reducing the effective SNR associated with the subcarrier.

3. Frequency Synchronization

To emulate imperfect synchronization, the system is considered under channel impairments such as Doppler frequency and frequency selective fading. The Doppler frequency leads to substantial CFO and the frequency selective fading subjects the subcarriers to independent fading gains. The CFO and the deep-fades in the channel degrade the corresponding subcarriers which reduces the effective SNR associated with the subcarrier. This makes the synchronization system at the receiver to perform poorly, and eventually lose lock with the corresponding subcarriers. Hence, the system experiences imperfect synchronization.

The frequency synchronization scheme utilized for the OFDM system is the well-known open-loop maximum likelihood (ML) estimator [15–17]. The frequency estimator gives rise to some jitter in the estimated frequency, increasing with decreasing SNR, due to the input noise. Correspondingly, the symbol error probability performance also degrades due to imperfect carrier recovery at low SNR. To avoid this, we set a threshold limit on the synchronizer, making sure that the frequency jitter does not exceed a certain limit. The frequency jitter has to be less or equal to the threshold frequency for the subcarrier to be declared as available for allocation. In this work, the threshold frequency is set as 10Hz. If the synchronizer is unable to lock to the carrier with this threshold limit, due to noise, Doppler frequency or deep fades, then we declare the subcarrier to be un-lockable during the given transmission time. Hence, the subcarrier is declared as not suitable for allocation.

Under the assumption of constant received signal power, as the noise power increases the number of sub-

carriers available for transmission decreases. The decrease in the total number of subcarriers available for reliable transmission imposes limitation on the total achievable data rate by the system. Hence, determining the decrease in the number of available subcarriers under imperfect synchronization which is the case in practical systems is important in assessing the accurate system performance.

4. SNR Analysis

In this section, we derive the expression for the SNR on the *mth* subcarrier as:

$$SNR_m = \frac{p_{D_m}}{P_{I_m} + P_{N_m}} \tag{9}$$

where P_{Dm} is the average power of the desired signal, P_{Im} is the average interference power and P_{Nm} is the average noise power on the *mth* subcarrier. The average power of the desired, interference and noise on the *mth* subcarrier is defined as

$P_{D_M} = E\left[\left|D_m\right|^2\right], P_{I_m} = E\left[\left|N_m\right|^2\right]$, and $P_{N_m} = E\left[\left|N_m\right|^2\right]$

respectively. Hence, the average SNR on the *mth* subcarrier is formulated as:

$$SNR_m = \frac{E\left[\left|D_m\right|^2\right]}{E\left[\left|I_m\right|^2\right] + E\left[\left|N_m\right|^2\right]} \tag{10}$$

where the average power of the desired signal is calculated as:

$$E\left[\left|D_m\right|^2\right] = E\left[\left|c_o \alpha_{m,i} a_{m,i}\right|^2\right] \tag{11}$$

Assuming that the transmitted data and the fading gain are independent, the average power of the desired signal can be written as:

$$E\left[\left|D_m\right|^2\right] = \left|c_o\right|^2 E\left[(\alpha_{m,i})^2\right] E\left[(\alpha_{m,i})^2\right] \tag{12}$$

The ICI power is expressed as:

$$\begin{aligned} E\left[\left|I_m\right|^2\right] &= E\left[\left|\sum_{l=0, l\neq m}^{N-1} c_{l-m}\alpha_{l,i} a_{l,i}\right|^2\right] \\ &= \sum_{l=0, l\neq m}^{N-1} \left|c_{l-m}\right|^2 E\left[\left(\alpha_{l,i}\right)^2\right] E\left[\left(a_{l,i}\right)^2\right] \end{aligned} \tag{13}$$

The above equation can be simplified as [4]:

$$E\left[\left|I_m\right|^2\right] = \left(1 - \left|c_o\right|^2\right) E\left[\left(\alpha_{l,i}\right)^2\right] E\left[\left(a_{m,i}\right)^2\right] \tag{14}$$

The power of the noise is expressed as :

$$E\left[\left|N_m\right|^2\right] = No \tag{15}$$

Hence, the SNR on the *mth* subcarrier is written as:

$$SNR_m = \frac{\left|c_o\right|^2 E\left[(\alpha_{m,i})^2\right] E\left[(a_{m,i})^2\right]}{(1 - \left|c_o\right|^2) E\left[(\alpha_{l,i})^2\right] E\left[(a_{m,i})^2\right] + N_o} \tag{16}$$

The above equation indicates the dependence of the SNR on the noise, frequency offset due to Doppler frequency and fading gain. Hence, as the average noise power increase the SNR decreases. In addition, to show the dependence of SNR on the CFO due to Doppler frequency, we express c_o as a function of $\Delta f\ T$. Hence, the SNR is written as:

$$SNR_m = \frac{\varnothing(\Delta fT) E\left[(\alpha_{m,i})^2\right] E\left[(a_{m,i})^2\right]}{(1 - \varnothing(\Delta fT)) E\left[(\alpha_{l,i})^2\right] E\left[(a_{m,i})^2\right] + N_o} \tag{17}$$

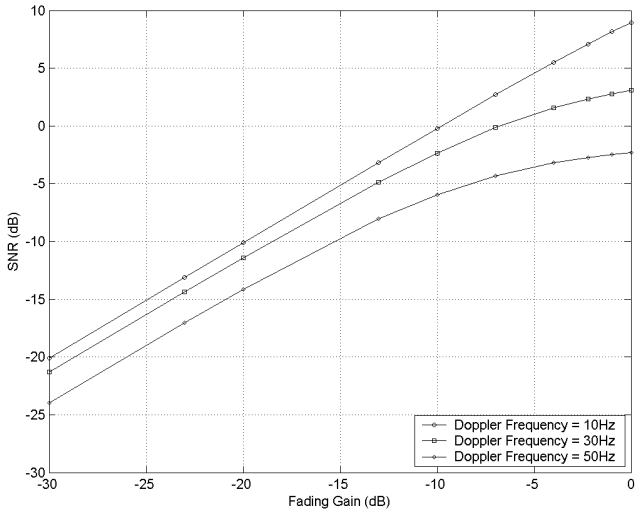


Figure 2. SNR versus Fading gain with different Doppler Frequency and constant average noise power of -10dB.

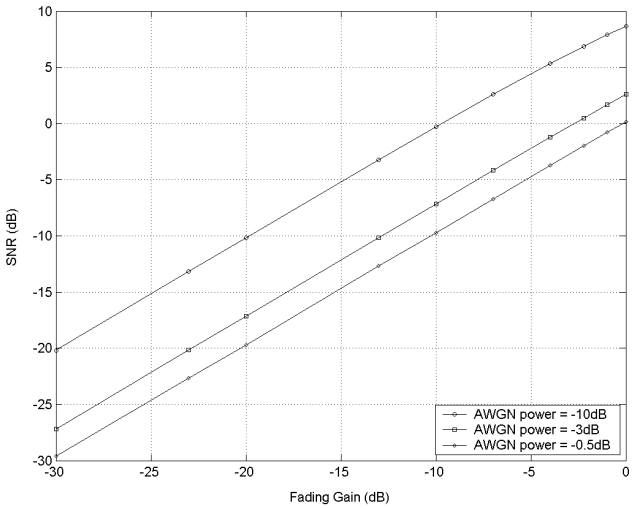


Figure 3. SNR versus Fading gain with different AWGN power and constant Doppler Frequency of 10Hz.

This expression clearly indicates that the desired signal and interfering power decrease due to CFO which in turn causes the SNR to decrease.

Analysis is performed to study the changes in the SNR as the average power of channel parameters such as average noise power, Doppler frequency and fading gain are varied.

Figure 2 indicates the changes in the SNR as the average fading gain is varied between -30dB to 0dB when average noise power is kept constant at -10dB. The SNR performance is evaluated under Doppler frequency of 10Hz, 30Hz, and 50Hz. As evident from (17), the average power of the fading gain degrades the signal power as well as the interference power. Hence, the increase in the average power of the fading gain degrades the SNR as indicated in Figure 2. Also, the increase in the Doppler frequency causes the interference power to increase. It is evident from this figure that as the Doppler frequency increases, the SNR decreases. The increase in the Doppler frequency from 10Hz to 50Hz results in 12dB decrease in the SNR for an average fading gain of 0dB.

Figure 3 indicates the changes in the SNR as the average fading gain is varied between -30dB to 0dB and the Doppler frequency is kept at 10Hz. The SNR performance is evaluated under average noise power of -10dB, -3dB and -0.5dB. As evident from (17), the increase in the average noise power increases the noise power and hence degrades the SNR. As illustrated in Fig. 3, when the average noise power is increased by 9.5dB, the SNR decreases by 9dB for an average fading gain of 0dB.

To maintain the adaptivity of the system further simplifying of (17) is avoided. Instead, Monte Carlo analysis is performed with values generated from Rayleigh distributions. Due to random nature of time variant multipath channels, it is reasonable to characterize such channels statistically. The random variable α_m is usually modelled using Rayleigh distribution and the subcarriers that are in deep-fade are not utilized during any transmission time interval in our case. To analyze the random effect, Monte Carlo analysis is performed to obtain the characteristics of the channel. Although the fading gains follow Rayleigh distribution, only the subcarriers that are identified as allocatable by the ML threshold limit are declared as allocatable.

5. Available Subcarriers with Imperfect Synchronization

In this section, an analysis is performed to quantify the variations in the number of available subcarriers under imperfect synchronization which results from noise, Doppler shift and frequency selective fades in the channel.

To emulate the effect of imperfect synchronization in a multiuser OFDM system, the average noise power, Doppler frequency and average fading level are selected as −3dB, 25Hz and −4dB respectively. The threshold frequency is set as 10Hz. The carrier frequency of the system is selected to be 4GHz and the forward link channel bandwidth is 20MHz. The total number of subcarriers, *N*, is 64 and the subcarrier bandwidth is 312.5kHz. The objective of the simulation is to obtain the average number of subcarriers for each user under imperfect synchronization and the variations in the total available subcarriers as the number of users is varied.

5.1. Number of Available Subcarriers under Constant SNR

The simulation results indicate that the number of subcarriers available for users varies in each transmission

time slot and under imperfect synchronization the total number of subcarriers available for reliable transmission is smaller compared to perfect synchronization. The availability of subcarriers for a certain user can be defined as follows:

$$N_a^k = N_T^K - N_{\bar{a}}^k \tag{18}$$

where N_a^k is the number of available subcarriers, N_T^K is the total number of subcarriers and $N_{\bar{a}}^k$ is the number of subcarriers that are not suitable for allocation for the *kth* user. The $N_{\bar{a}}^k$ varies for different users supporting the fact that the channel conditions in a multipath environment is random and changes for each user.

Under the specified channel conditions, in a multiuser environment the average number of subcarriers available for User-1 is 49 and for User-2 is 48. This translates to 77% and 75% available subcarriers for User-1 and User-2 respectively in comparison to perfect synchronization. To ensure reliable data trans mission, subcarrier allocation algorithms should consider the variations in the total available subcarriers for each user and avoid allocating subcarriers from $N_{\bar{a}}^k$.

In addition, we determine the maximum and minimum number of available subcarriers under a given SNR in a multiuser environment. As indicated in Figure 4, as the number of users increase, the minimum number of available subcarriers for transmission decreases while the maximum number of subcarriers available for transmission increases. This indicates that over a large number of trials the average number of allocatable subcarrier is 46.

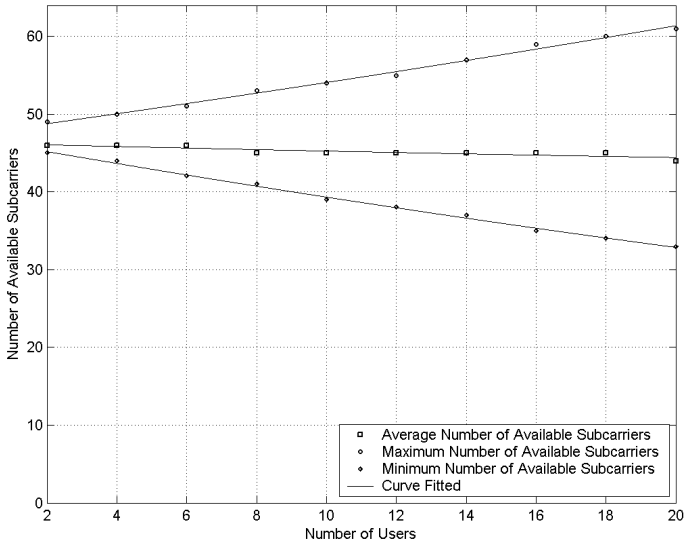
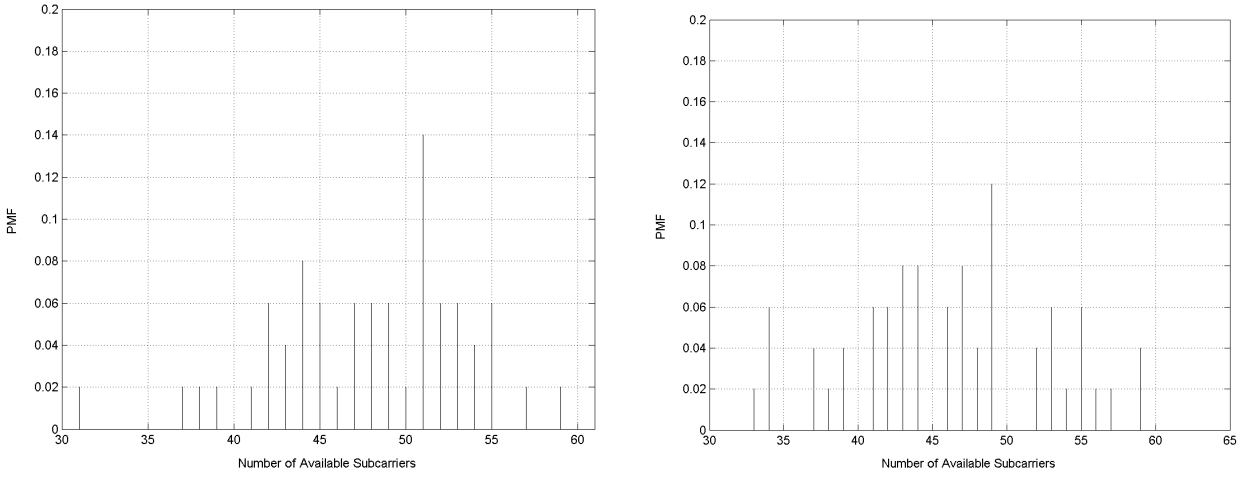
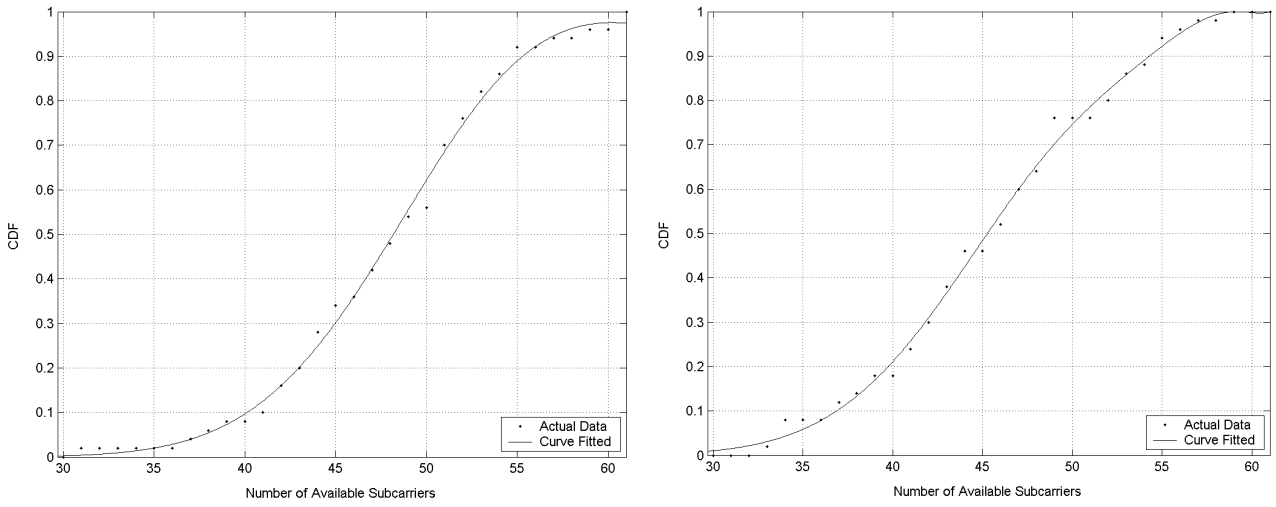


Figure 4. Number of available subcarriers versus number of users.



(a) User-1 (b) User-2

Figure 5. Probability mass function of the available subcarriers.



(a) User-1 (b) User-2

Figure 6. Cumulative distribution function of the available subcarriers.

In addition, the maximum and minimum available subcarriers indicate the upper and lower bound of the achievable data rate for the users since the data rate is directly proportional to available subcarriers. It can be stated that while the average available subcarriers does not vary significantly with the increase in the number of users, the minimum and maximum available subcarriers indicate noticeable variations. As the number of users increases, the difference between the maximum and minimum available subcarrier increases, which provides flexibility in the total number of allocatable subcarriers. This system characteristic supports multiplexing gain which in turn increases the aggregate data rate.

5.2. Statistical Model of the Subcarrier Availability

In this section, empirical analysis is used to determine the statistical characteristics of the number of available subcarriers. Given that the number of available subcarriers for transmission is a random variable that takes countable values, it is modelled as a discrete random variable (DRV) [18]. The obtained results are used to develop the probability mass function (PMF) of the available subcarriers for User-1 and User-2, which is indicated in Figure 5. The PMFs are used to obtain the cumulative distribution function (CDF) for both users which is given in Figure 6.

Based on the PMF of the number of available subcarriers, it is identified as a Poisson random variable (PRV) with the parameter λ that defines the average number of available subcarriers in a given time interval.

In our simulation, it is observed that λ for User-1 and User-2 are 51 and 49 respectively. In general, for all users λ is the same on average and hence the number of allocatable subcarriers is a PRV.

The DRV is identified as a PRV because it has the characteristics of a PRV which are explained bellow:

1) The number of available subcarriers is determined in transmission time slot with a fixed length, t_s.

2) The number of available subcarriers for each user varies in the time interval t_s but over a time period of Nt_s the number of available subcarriers for each user has a constant average.

3) The number of subcarriers that are available in disjoint time intervals are statistically independent because the fading gains are uncorrelated.

Although the number of available subcarriers is a PRV, the availability of each subcarrier for trans- mission can be modelled as a Bernoulli random variable (BRV) where the availability of the each is indicated by a 0 indicating not available for allocation and 1 available for allocation. Each subcarrier availability can be viewed as a Bernoulli trial because the number of available subcarrier in each trial is independent, and at most a certain number of subcarriers is determined in each trial. This further supports modelling of the total number of available subcarrier as a PRV since a sequence of Bernoulli trials occurring in time is modelled as a PRV.

Based on the above observations, the number of available subcarriers is identified as a PRV with mean, $\lambda = 0.78N$, where N is the total number of subcarriers. Hence, the average percentage loss of the subcarriers under imperfect synchronization is 22% under the specified channel condition.

To avoid performance degradation in terms of subcarrier availability for transmission, perfect synchronization is required. Since practical systems are subject to imperfect synchronization, the analysis indicates that determining the availability of subcarriers for transmission is essential in the optimization process of radio resource allocation for multiuser systems.

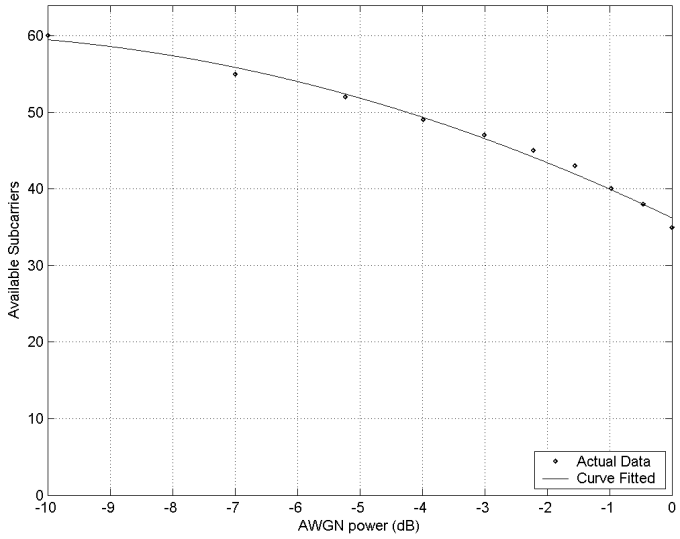


Figure 7. Number of available subcarriers versus AWGN power.

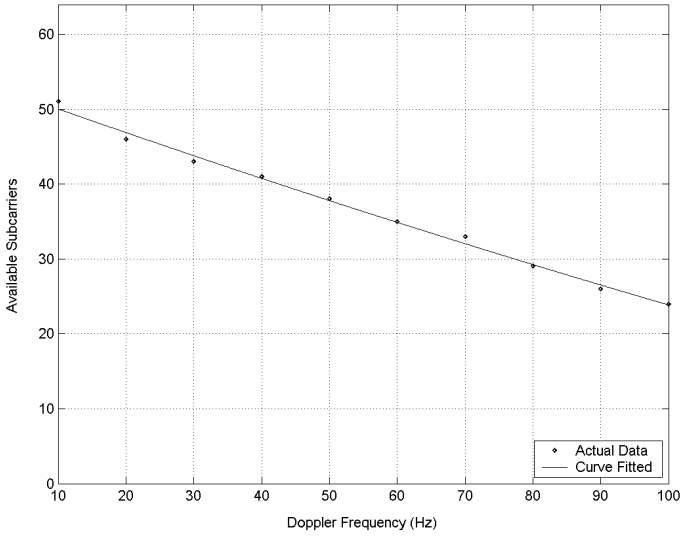


Figure 8. Number of available subcarriers versus Doppler Frequency.

6. Number of Available Subcarriers under Variable SNR

In this section, we investigate the variations in the total number of subcarriers as the factors such as, average noise power, which is modelled as AWGN, Doppler frequency, and fading gain are changed.

6.1. AWGN

To determine the changes in the number of available subcarriers, the AWGN power is varied while the Doppler frequency and the fading gain are kept constant at values of 25Hz and −4dB respectively. Figure 7 shows the number of available subcarriers under different AWGN power. Based on the results, as the AWGN power is varied between −10dB to 0dB, the number of available subcarriers decreases by 19%.

6.2. Doppler Frequency

To investigate the variations in the number of available subcarriers as the Doppler frequency is changed while the AWGN power and the fading gain are kept constant at values of −3dB and −4dB respectively. As illustrated in Figure 8, as the Doppler frequency is varied between 10Hz to 100Hz the number of available subcarriers decreases by 44%.

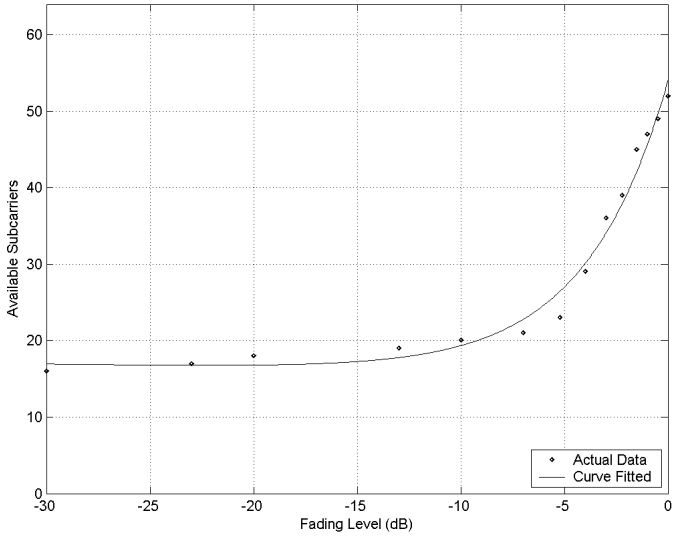


Figure 9. Number of available subcarriers versus fading level.

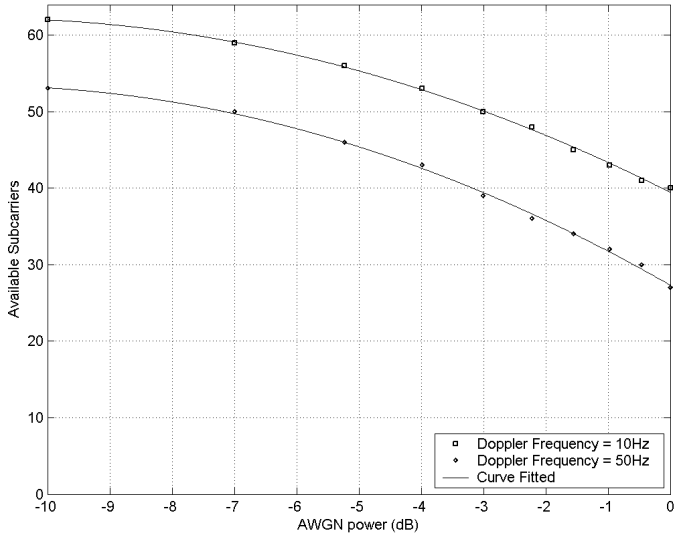


Figure 10. Number of available subcarriers versus AWGN power for different Doppler Frequency.

6.3. Frequency Selective Fading

To determine the changes in the number of available subcarriers as the fading level is varied, the AWGN power and the Doppler frequency are kept constant at −3dB and 25Hz respectively. Figure 9 depicts the corresponding number of subcarriers as the fading level is varied between −30 to 0dB. Hence, as the fading level is varied between 0dB to −30dB the number of available subcarries decreases by 56%.

To further analyze the effect of the hostile channel conditions such as AWGN power, deep-fading and Doppler frequency, the number of available subcarriers is determined with the variation in both the AWGN power and the Doppler frequency under a constant fading gain of −4dB. Figure 10 shows the number of available subcarriers with the variations in AWGN power and Doppler frequency. Based on the results, it can be stated that 40Hz increase in the Doppler frequency leads to 20% decrease in the number of available subcarriers for an AWGN power of 0dB.

In addition, the number of available subcarriers is determined with the variation in both the AWGN power and the fading level under a constant Doppler frequency of 25Hz. Figure 11 shows the number of available subcarriers with the variations in AWGN power and fading levels. As evident from this figure, 20dB increase in fading gain results in 12% decrease in the number of available subcarriers for an AWGN power of 0dB.

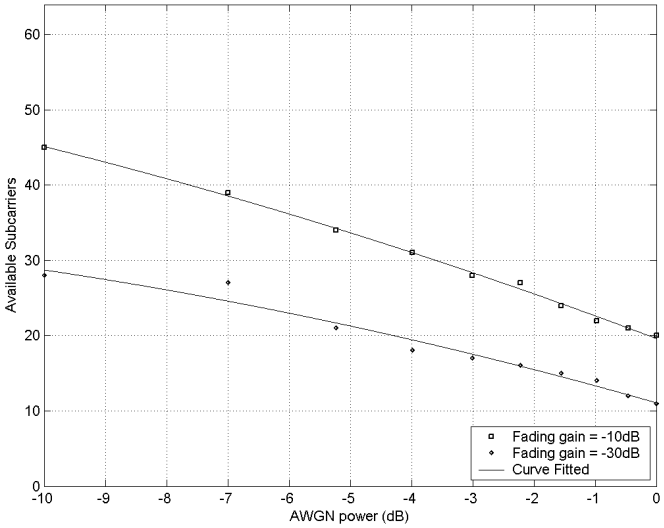


Figure 11. Number of available subcarriers versus AWGN power for different fading gain.

Hence, channel constraints such as AWGN, Doppler effect and frequency selective fading impose noticeable implication on system performance. The total number of available subcarriers changes with the variation in the AWGN power, Doppler shift and fading level. Thus, under imperfect synchronization, which is the case in most communication systems, not all the subcarriers are available for transmission.

7. Conclusions

To avoid performance degradation in terms of subcarrier availability for transmission, perfect synchronization is required. Since practical systems are subject to imperfect synchronization, the analysis indicates that determining the availability of subcarriers for transmission is essential in the optimization process of radio resource allocation. In this paper, we perform an analysis to determine the SNR degrades as the average power of the channel impairments such as AWGN, CFO due to Doppler frequency and fading gain are increased. The decrease in SNR causes imperfect synchronization and hence reduces the total number of available subcarriers for allocation. We use empirical modelling to characterize the number of available subcarriers as Poisson random variable and it is determined that under imperfect synchronization up to 22% of the subcarriers are not suitable for transmission as compared to perfect synchronization under certain channel conditions. We have determined the variations in the number of available subcarriers with the changes in the parameters such as AWGN, Doppler frequency and deep fades that introduce imperfect synchronization. It has been illustrated that a 10dB increase in the average AWGN power leads to 19% decrease in the total number of allocatable subcarriers; a variation of 10Hz to 100Hz in Doppler frequency causes 44% decrease in the number of allocatable subcarriers, and changes in the fading level between 0dB to −30dB result in 56% decrease in the number of allocatable subcarriers. Thus, under imperfect synchronization all the subcarriers are not available for transmission. Given that the data rate is directly proportional to the total number of available subcarriers for transmission, to provide a realistic measure of the system capacity subcarrier allocation algorithms should be based on the number of avail- able subcarriers under imperfect synchronization. Although subcarrier allocation under the constraint of imperfect synchronization does not support more users or higher data rates, it improves system reliability by eliminating allocation on unavailable subcarriers and hence improving the system BER performance.

8. References

[1] D. H. J. Sun and J. SauVola, “Features in future: 4g visions from a technical perspective,” IEEE GLOCOM, Vol. 6, pp. 3533–3537, November 2001.

[2] L. J. Cimini, “Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing,” IEEE Transaction on Communications, pp. 665–675, July 1995.

[3] J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” IEEE Communication Magazine, pp. 5–14, 1990.

[4] M. v. B. T. Pollet and M. Moeneclaey, “Ber sensitivity of ofdm systems to carrier frequency offset and wiener phase noise,” IEEE Transaction on Communications, Vol. 43, pp. 191–193, 1995.

[5] M. Luise and R. Reggiannini, “Carrier frequency acquisition and tracking for OFDM systems,” IEEE Transaction

on Communications, Vol. 44, 1996.

[6] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," IEEE Transaction on Communications, Vol. 42, pp. 2908–2913, 1994.

[7] P. B. J. van de Beek and M. Sandell, "Ml estimation of timing and frequency offset in ofdm systems," IEEE Transaction on Signal Procesings, Vol. 45, pp. 1800–1805, 1997.

[8] S. K. N. Lashkarian, "Globally optimum ml estimation of timing and frequency offset in ofdm systems," IEEE International Conference on Communications, Vol. 2, pp. 1044–1048, 2000.

[9] K. L. R. M. C. Y. Wong, and R. S. Cheng, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," IEEE Journal, Selected Areas in Communications, Vol. 17, 1999.

[10] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," IEEE Journal, Selected Areas in Communications, Vol. 21, 2003.

[11] B. E. Z. Shen and J. G. Andrews, "Optimal power allocation in multiuser OFDM systems," IEEE Global Telecommunications Conference, Vol. 1, 2003.

[12] J. C. W. Rhee, "Increased in capacity of multiuser ofdm system using dynamic subchannel allocation," IEEE 51st, Vehicular Technology Conference, Vol. 2, 2000.

[13] S. Y. C. Suh and Y. Cho, "Dynamic subchannel and bit allocation in multiuser OFDM with a priority user," IEEE Eighth International Symposium, Spread Spectrum Techniques and Applications, pp. 919–923, 2004.

[14] P. Song and L. Cai, "Multi-user subcarrier allocation with minimum rate request for downlink OFDM packet transmission," IEEE 59th Vehicular Technology Conference, Vol. 4, 2004.

[15] D. Rife and R. Boostyn, "Single-tone parameter estimation from discrete-time observations," IEEE Transaction on Information Theory, Vol. 5, 1974.

[16] S. Kandeepan, "Synchronisation techniques for digital modems," PhD thesis, University of Technology, Sydney, July 2003.

[17] S. Kandeepan and S. Reisenfeld, "Performance analysis of a correlator based maximum likelihood frequency estimator," SPCOM, pp. 169–173, 2004.

[18] S. Ross, "Introduction to probability models," Academic Press, 2003.

Int. J. Communications, Network and System Sciences, **2009, 2, 786-791**
doi:10.4236/ijcns.2009.28091 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Performance Analysis of MAC Protocol for LEO Satellite Networks

Mingxiang GUAN, Ruichun WANG
Department of Electronic Communication Technology, Shenzhen Institute of Information Technology, Shenzhen, China
Email: *gmx*2020@126.*com*
Received July 14, 2009; *revised August* 16, 2009; *accepted September* 22, 2009

Abstract

Considering that weak channel collision detection ability, long propagation delay and heavy load in LEO satellite communications, a valid adaptive APRMA MAC protocol was proposed. Different access probability functions for different services were obtained and appropriate access probabilities for voice and data users were updated slot by slot based on the estimation of the voice traffic and the channel status. In the proposed MAC protocol limited wireless resource is allocated reasonably by multiple users and high capacity was achieved. Three performance parameters: voice packet loss probability, average delay of data packets and throughput of data packets were considered in simulation. Finally simulation results demonstrated that the performance of system was improved by the APRMA compared with the conventional PRMA, with an acceptable trade-off between QoS of voice and delay of data.

Keywords: LEO Satellites, Adaptive Packet Reservation Multiple Access, MAC Protocol

1. Introduction

Due to various economic and technical constraints, terrestrial mobile networks can only provide communication services with a limited coverage. Recently, in response to increasing demand of real-time multimedia services and the truly global coverage required by personal communication services, there is a vast research on non-geostationary orbit (NGSO) satellites systems, especially on low earth orbit (LEO) satellite constellations with an altitude between 700 km and 1 500 km. LEO satellite constellations equipped with inter-satellite links, such as Iridium, Teledesic, Courier and so on, usually have onboard switching and onboard routing facilities and form an independent network in space. Direct connectivity between any pair of satellite mobile users can be achieved through the satellites and ISLs without any essential usage of the terrestrial core network. For the wide application prospect, they have already been the focus of the research on the satellite communication systems. This LEO system can provide real time voice and data traffics in the global range. It is the trend that various kinds of traffics will be provided by LEO satellites system. It is of great importance that an effective medium access control (MAC) protocol will be required to make full use of limited resource and to provide services with strict quality for users. MAC protocol is used to allow many mobile users to share simultaneously a finite amount of radio spectrum. The sharing of spectrum is required to achieve high capacity by simultaneously allocating the available bandwidth to multiple users. Thus the appropriate access control protocol will be a key problem for wireless mobile communications development.

LEO satellites will provide not only the real-time traffic such as video and voice but also data traffic with burst character. The efficiency of resource utilization will sharply decrease if fixed assignment multiple access is applied. Also, voice and video traffic will not be supported sufficiently if competitive multiple access (ALOHA, CSMA *et al.*) is used completely [1–3]. Since PRMA (Packet Reservation Multiple Access) as an access protocol for wireless local networks was introduced by D.J Goodman *et al*. in 1989 [4], its high efficiency for voice packet transmission captured much attention, since then new versions have been proposed to support multi-media traffic which is very important in the future mobile system. In literature [5] this protocol was researched profoundly and the author pointed out that PRMA is competitive protocol with the limit of traffic and connection number at one time. Three main problems will be encountered if this protocol is applied in

LEO satellites system. They are: 1) The channel collision detection ability is quite weak. 2) The propagation time delay is long comparatively to terrestrial communications system. 3) Heavy load will be supported because of many users in the coverage of the LEO satellites. The three characteristics will bring on increase of packet loss probability, severity of channel congestion and decrease of QoS (Quality of Service).

More improved PRMA protocols were provided based on [4–5]. In literature [6] PRMA-HS with re-transmission character was provided in order to overcome long time delay problem in satellite communication. But the access contention becomes serious and performance of the system degrades under the environment of large number of users or heavy load. Moreover this protocol has a changeable channel access time delay and a certain packet loss probability which are not suitable for services with strict QoS requirement. In literature [7] IPRMA (Integrated Packet Reservation Multiple Access) protocol was proposed for satellite communication. A user can reserve many slots to improve performance of this protocol. But it possibly exists that one user occupies the resource totally. In literature [8] MPRMA (Mini-Packet Reservation Multiple Access) was provided. In the protocol an available slot will be divided into many mini-slots in which competitive packets are transmitted. From [8] we can see that probability of collision in a mini-slot decreases. But this protocol can not support a mass of real-time traffic due to the decline of the efficiency of transmission. In literature [9] the author provided NC-PRMA (Non-Collision Integrated Packet Reservation Multiple Access) protocol which adopted queue model to avoid collision resulted from competition and performance was improved. But this protocol is not perfect in the long propagation delay because implementation of this protocol will be in the environment of short RTD (Round Trip propagation Delay) period.

From the analysis above, we can find that access probability for voice and data is obtained from the same access function, without considering the different traffic characteristics and requirements of voice and data users (voice users require real-time delivery but can accommodate higher bit error rates; data users do not need real-time transmission and can be queued but require low bit error rates or error-free transmission). Therefore it will be expected to be more efficient if priority is given to the transmission of voice, whilst minimizing the effects on the data packets. Considering that weak channel collision detection ability, long propagation delay and a mass of load in communication system in LEO satellites, an adaptive access control protocol improved form PRMA based on channel status, quality of service and estimation of traffic was proposed with priority of voice traffic referred to [10,11]. Thus in this method, voice packets access to the channel with a priority and a updated access probability and then, if the resource is available, data packets can be accepted with an updated access probability slot by slot. Our simulation results show that the efficiency is improved by the new adaptive PRMA protocol.

This paper is organized as follows. In Section 2, two kinds of access probability functions are derived for voice and data traffic respectively. Compared to conventional PRMA protocol, three performance parameters of voice packet loss probability, average delay of data packets and throughput of data packets are analyzed by simulation in Section3. Finally system performance and conclusions are obtained.

2. Protocol Model

2.1. Traffic Analysis

For a voice terminal, the voice source can be characterized by a two-state Markov chain model, as shown in Figure 1. Four parameters are required for the description of the model. They are: the mean duration of a talk burst t_1, the mean duration of the silence t_2, the transition probability from the talking state to the silent state γ and the inverse transition probability δ. The parameters γ and δ can expressed as follows:

$$\gamma = 1 - \exp(-\tau / t_1) \tag{1}$$

$$\delta = 1 - \exp(-\tau / t_2) \tag{2}$$

where τ is the width of one time slot. The empirical values for t_1 and t_2 are 1s and 1.35s. The voice terminal generates one packet per frame which is first composed of an information field with length of $R_s T_f$, where R_s is the bit-rate of voice and T_f is the duration of one frame, and a packet header with length H bits.

Next data users were concerned. A data terminal has a discontinuous stream. Denote the average bit rate of the data terminal by R_s and a data packet which is the same as voice packet is generated independently in each slot with a probability of δ_d. Hence, the mean bit rate of a data terminal is:

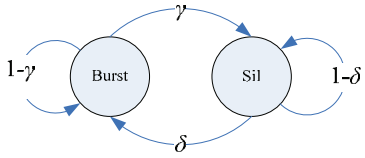


Figure 1. Two states Markov model of voice.

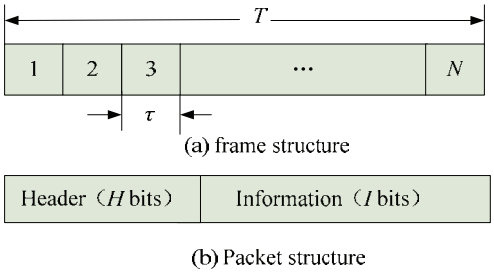


Figure 2. Structure of frame and packet.

$$R_d = \delta_d R_s A \tag{3}$$

where A is the number of slots per frame and can be calculated by:

$$A = \text{int}[R_p T_f / (R_s T_f + H)] \tag{4}$$

R_p is the channel rate before coding.

2.2. Frame Structure

These frames are further subdivided into *N* time slots as illustrated in Figure 2(a). Information packets transmitted from terminals to satellites consist of both a payload (actual information) and a header (control information) as illustrated in Figure 2(b). The time slot duration is τ and *T* is the duration of a single frame.

2.3. APRMA Protocol

In the beginning of this paper we have got the conclusion that conventional and improved PRMA protocols supported mixed voice and data traffic with a low efficiency in LEO satellites. In the APRMA protocol, P_v is the access probability of voice and P_d is the access probability of data respectively. Voice user is given priority compared with data user. Appropriate access probabilities P_v and P_d are broadcast from the LEO satellite, which is updated slot by slot based on the estimation of the voice traffic and the channel status.

The purpose of APRMA protocol is to guarantee real-time transmission of voice packets by priority transmission compared with data packets. When the channel load is light, transmitting data packets is allowed. On the other hand, when the channel load is heavy, transmission of data packet is postponed. We assume that the LEO satellites can recognize the total number of users in a cell and the number of users in reservation mode. Then, based on the statistical characteristics of the traffic models, the number of contending voice terminals is estimated and the access probability is calculated. From the voice model described in 2.1, voice terminals are M_v and voice terminals in reservation mode are M_{rsv}, The probability of n new terminals arrival talk burst and one terminal departing from talk burst can be considered as a binomial distribution:

$$\begin{cases} B(n,k,p) = C_n^k (p)^k (1-p)^{n-k} \\ P(n \mid M_{sil}) = B(M_{sil}, n, \delta) \\ M_{sil} = M_v - M_{rsv} \end{cases} \tag{5}$$

where M_{sil} is number of voice terminals in silent mode. When the population of users is large or the probability of p is small, the binomial model approaches the Possion model. Thus $\lambda = M_{sil}\delta$ is arrival rate of the voice users and $u = M_{rsv}\gamma$ is departure rate of the voice users respectively. Hence, in the current time slot the estimated value of M_{rsv} can be expressed by:

$$\begin{cases} M_{rsv} = M_{rsv1} + \lambda P_0 / \mu \\ Max(M_{rsv} - M_{rsv1}) = \Delta C \\ \Delta C = K_{OPT} - K_{RSV} \end{cases} \tag{6}$$

$$\begin{cases} r\gamma_f + c_v\gamma - s_v\delta = 0 \\ r(1-\gamma_f) + c_v P (1-P)^{c_v - 1}(1-r)(1-\gamma) - r = 0 \\ s_v - c_v + Br = M_v \end{cases} \tag{7}$$

where M_{rsv1} is the number of voice users in reservation mode in the previous time slot and P_0 is the access probability of voice users at current time slot. K_{OPT} is the available number of access channel for voice users. In the voice system, the balance equation is shown in the following: Here $\gamma_f = 1-(1-\gamma)^B \approx B\gamma$, P is the access probability. From the Equation (7), the following Equation (8) can be found.

$$(1+\frac{\gamma}{\delta})c_v + (B+\frac{\gamma_f}{\delta})r = M_v \tag{8}$$

For the data system, the probability of a data user to send a data packet successfully is w:

$$w = p_d u_d (1 - r(1 + b / M_d B)) \tag{9}$$

where $u_d = (1-p_v)^c (1-p_d)^{(1-\Gamma)b-1}$. Therefore, the data user and voice user access probability at the balance point can be shown by the Equation (10):

$$\frac{b\Delta}{c(1-\gamma_f)}[cp(1-\gamma)(1-r(1+\frac{b}{M_d B}))] = \delta_d M_d \tag{10}$$

where $\Delta = p_d(1-p) / p(1-p_d)$. In the voice system, the Equation (8) can be shown in the following.

$$h_1 c + h_2 r = M_v \tag{11}$$

where $h_1 = (1+\gamma/\delta)$, $h_2 = (1+\gamma_f/\delta)$. Combination Equations (10) and (11), the relation between b and c is in the following.

$$b = \min\{\frac{c}{\Delta}\delta_d M_d \frac{1-\gamma}{\gamma_f}(\frac{h_2}{M_v - h_1 c}), M_d\} \quad (12)$$

2.3.1. Voice Packet Loss Probability

Packet loss probability is defined as the ratio of the number of loss packets and the number of the generated packets at the terminals.

$$P_{vdrop} = \left\{1 - \frac{\gamma_f[1-(1-\gamma_f)v^{2B}]}{[1-(1-\gamma_f)v^B]^2}\right\}\frac{v^N\gamma_f(1-\gamma_f)}{1-v^B} + \frac{\gamma_f v^B}{[1-(1-\gamma_f)v^B]^2} \quad (13)$$

where N is the maximum time delay of voice packet and v is the successful access probability of voice packet.

$$v = 1-(1-r(1+b/M_d A))\cdot p(1-p)^{c-1}(1-p_d)^{(1-\Gamma)b} \quad (14)$$

2.3.2. Average Time Delay of Data Packet

Average time delay is defined as the lasting time from packet generated to packet received successfully at the LEO satellites. When a data packet arrived at a data user, j data packets were waited for transmission. Average time delay of this data packet was $(j+1)/w$. Therefore the average time delay of data packet is:

$$W_{ad} = \frac{1-\Gamma}{w(1-p_d)} \quad (15)$$

2.3.3. Throughput of Data Packet

Throughput is defined as the ratio of the number of packets received successfully and the number of packets generated at the terminals in a time unit. The Throughput of data packet is defined as the proportion of timeslots that successfully carry information packets.

$$T_{throughput} = r(1+\frac{b^2\Gamma\delta_d}{M_d A}) + \delta_d M_d(1-\Gamma) \quad (16)$$

3. Results Analysis

All the LEO mobile satellites system have taken CDMA technology except Iridium system (TDMA technology was taken). LEO satellites adopt multi-beam formation technology to make full use of the finite radio frequency resource. Therefore many cells are formed and users separated by space can re-use the radio channel. In [12] CDMA channel attenuation model in AGWN was proposed which adopted BPSK modulation technology and BCH coding (511, 229, 38). Figure 3 shows the access probability of voice users in APRMA protocol and Table 1 shows the simulation parameters.

Figure 4, Figure 5 and Figure 6 show the simulation results with equal loads of voice and data traffic. For 2%

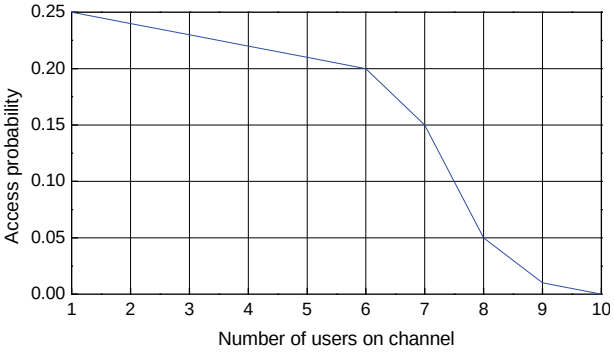


Figure 3. APRMA voice terminals access probability.

Table 1. Simulation parameters.

Parameter	Value
Channel rate	5.3Mbps
CDMA information rate	4599Kbps
Channel rate after coding	1022Kbps
Channel rate before coding	558Kbps
Voice rate	16Kbps
Average data rate	3.4Kbps
Frame duration	10ms
Information bit per packet	160bits
Frame header	69bits
Slots per frame	10
Maximum delay (Voice packet)	20ms
Mean duration of talk burst t_1	1s
Mean duration of silence t_2	1.35s

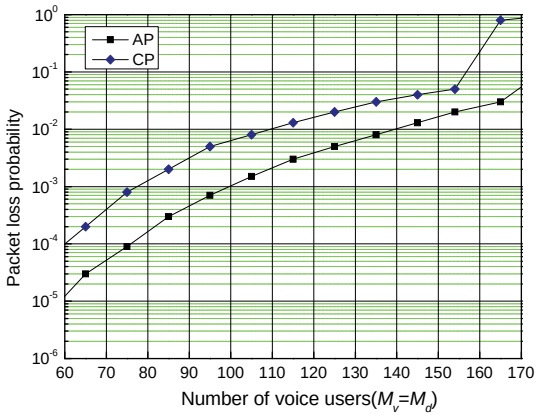


Figure 4. Voice packet loss probability.

packet loss probability as an acceptable level, then the system capacity is improved about 18% by APRMA compared to CP (Convention PRMA). Furthermore, Table 2 shows the comparison between AP and CP protocol. From the performance comparison in the Table 2, the system performance is improved by the APRMA protocol in all three cases.

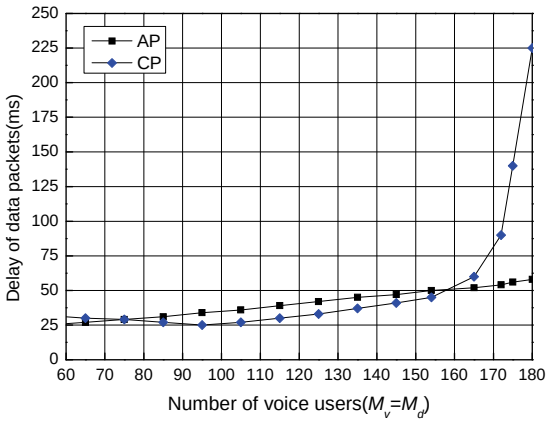


Figure 5. Average time delay of data packets.

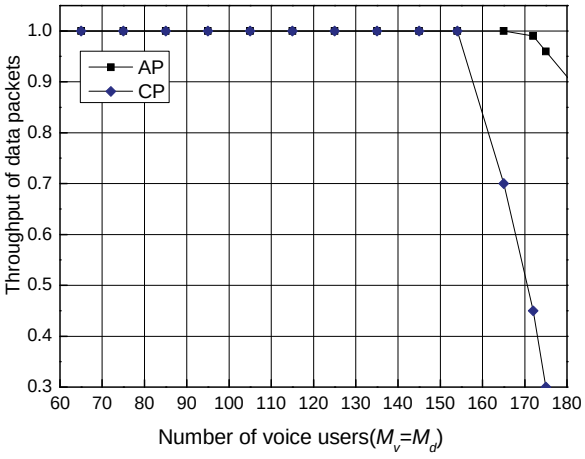


Figure 6. Throughput of data packets.

Table 2. Performance comparison.

Data user		100	150	200
AP	Voice user (PLP)	188	108	78
	Voice user (ATDDP)	282	170	138
	Voice user (TDP)	260	185	160
CP	Voice user (PLP)	142	85	58
	Voice user (ATDDP)	240	134	116
	Voice user (TDP)	220	150	135
Performance improvement (%)	Voice user (PLP)	24.6	21.2	25.6
	Voice user (ATDDP)	14.9	21.7	15.9
	Voice user (TDP)	15.4	18.9	15.6

PLP, ATDDP and TDP stand for packet loss probability, average time delay and throughput of data packet respectively.

4. Conclusions

With the development of LEO satellite communication, it is the base requirement that various kinds of services will be provided. Considering that weak channel collision detection ability, long propagation delay and heavy load in LEO satellite communication system, a valid adaptive access control protocol APRMA is proposed. Different access probability functions for different services are obtained and appropriate access probabilities for voice and data users are updated slot by slot based on the estimation of the voice traffic and the channel status. Simulation results demonstrate that the performance of system is improved by the APRMA compared with the conventional PRMA, with an acceptable trade-off between QoS of voice and delay of data. Also the APRMA protocol will be suitable for HAPS (high altitude platform station) with the character of weak channel collision detection ability, long propagation delay and heavy load.

5. Acknowledgement

This paper supported by the 3rd natural science foundation of institute (No. LG-08010) and 2nd doctoral innovation foundation of institute. The author would like to thank Dr. Zhong Weizhi and Li Lu for their revisions of the text, and the editor and the anonymous reviewers for their contributions that enriched the final paper.

6. References

[1] R. Fantacci, T. Pecorella, and Giombini, "Stability and performance analysis of a MAC protocol for a time-code air interface in LEO mobile satellite systems [J]," IEEE Transactions on Vehicular Technology, Vol. 52, No. 3, pp. 607–621, May 2003.

[2] F. Chiti, R. Fantacci, and F. Marangoni, "Advanced dynamic resource allocation schemes for satellite systems [C]," IEEE International Conference on Communications, Vol. 3, pp. 1469–1472, 2005,.

[3] M. Emmelmann and H. Bischl, "An adaptive MAC layer protocol for ATM-based LEO satellite networks [C]," IEEE Vehicular Technology Conference, Vol. 4, 2698–2702, 2003.

[4] D. J. Goodman, R. A. Valenzuela, and K. T. Gayliard, "Packet reservation multiple access for local wireless communications [J]," IEEE Transactions on Communications, Vol. 37, No. 8, pp. 885–890, 1989.

[5] S. Nanda, D. J. Goodman, and U. Timor, "Performance of PRMA: A packet voice protocol for cellular systems

[J]," IEEE Transactions on Vehicular Technology, Vol. 40, No. 3, pp. 584–598, 1991.

[6] E. Del Re and R. Fantacci, "Performance analysis of an improved PRMA protocol for low earth orbit-mobile satellite systems [J]," IEEE Transactions on Vehicular Technology, Vol. 48, No. 3, pp. 985–1001, 1999.

[7] R. Fantacci, T. Pecorella, and I. Habib, "Proposal and performance evaluation of an efficient multiple-access protocol for LEO satellite packet networks [J]," IEEE Journal on Selected Areas in Communications, Vol. 22, No. 3, pp. 538–545, 2004.

[8] G. Benelli, R. Fantacci, and G. Giambene, "Performance analysis of a PRMA protocol suitable for voice and data transmissions in low earth orbit mobile satellite systems [J]," IEEE Transactions on Wireless Communications, Vol. 1, No. 1, pp. 156–168, 2002.

[9] R. Fantacci, G. Giambene, and R. Angioloni, "A modified PRMA protocol for voice and data transmissions in low earth orbit mobile satellite systems [J]," IEEE Transactions on Vehicular Technology, Vol. 49, No. 5, pp. 1856–1876, 2000.

[10] J. W. So, "Adaptive traffic prediction based access control in wireless CDMA systems supporting integrated voice/data/video services [J]," IEEE Communications Letters, Vol. 8, No. 12, pp. 703–705, 2004.

[11] M. C. Vuran and I. F. Akyildiz, "A-MAC adaptive medium access control for next generation wireless terminals [J]," IEEE/ACM Transactions on Networking, Vol. 99, pp. 1–10, 2007.

[12] S. Z. Ozer and S. Papavassiliou, "Performance analysis of CDMA systems with integrated services [J]," IEEE Transactions on Vehicular Technology, Vol. 52, No. 4, pp. 823–836, 2003.

Int. J. Communications, Network and System Sciences, **2009, 2, 792-796**
doi:10.4236/ijcns.2009.28092 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Ant Colony Optimization Based on Adaptive Volatility Rate of Pheromone Trail

Zhaoquan CAI[1], Han HUANG[2,3], Yong QIN[4], Xianheng MA[2]
[1]*Educational Technology Center, Huizhou University, Huizhou, China*
[2]*School of Software Engineering, South China University of Technology, Guangzhou, China*
[3]*State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China*
[4]*Center of Information and Network, Maoming University, Maoming, China*
Email: {*gdzqcai, bssthh*}@163.*com*
Received May 25, 2009; *revised July* 15, 2009; *accepted August* 13, 2009

Abstract

Ant colony optimization (ACO) has been proved to be one of the best performing algorithms for NP-hard problems as TSP. The volatility rate of pheromone trail is one of the main parameters in ACO algorithms. It is usually set experimentally in the literatures for the application of ACO. The present paper first proposes an adaptive strategy for the volatility rate of pheromone trail according to the quality of the solutions found by artificial ants. Second, the strategy is combined with the setting of other parameters to form a new ACO method. Then, the proposed algorithm can be proved to converge to the global optimal solution. Finally, the experimental results of computing traveling salesman problems and film-copy deliverer problems also indicate that the proposed ACO approach is more effective than other ant methods and non-ant methods.

Keywords: Ant Colony Optimization (ACO), Adaptive Volatility Rate, Pheromone Trail

1. Introduction

ACO was first proposed by M. Dorigo and his colleagues as a multi-agent approach to deal with difficult combinatorial optimization problems such as TSP [1]. Since then, a number of applications to the NP-hard problems have shown the effectiveness of ACO [1]. Up till now, Ant Colony System (ACS) [2] and MAX-MIN Ant System (MMAS) [3] are so successful and classical that their strategies such as pheromone global-local update and Maximum-Minimum of pheromone are widely used in recent research [1].

The main parameters of ACO may conclude: k, ρ, α and β, where k is the number of artificial ants used for solution construction, ρ is the parameter for volatility of pheromone trail and α, β determines the relative importance of pheromone value and heuristic information [2,4,5]. All of the parameters are usually set with experimental methods in the application of ACO [5–7]. For the adaptive parameter setting, M. Dorigo and L.M. Gambardella presented a formula for the optimal number of ants k based on the value of ρ and q_0 in ant colony system. I. Watanabe and S. Matsui proposed an adaptive control mechanism of the parameter candidate sets based on the pheromone concentrations [8]. M. L., Pilat, and T. White put forward the ACSGA-TSP algorithm [9] with an adaptive evolutionary parameters β, ρ, q_0 and gave the experimental values of these parameters for some TSP problems. For the parameters α and β, which regulate the relative importance of pheromone trail and closeness [10], H. Huang proposed a dynamic strategy for a bi-directional searching ant colony system [11]. However, other parameters should be set experimentally.

This paper presents a trial work of setting the parameters of ACO adaptively. First, a tuning rule for ρ is designed based on the quality of the solution constructed by artificial ants. Then, we introduce the adaptive ρ to form a new ACO algorithm, which is tested to compute several benchmark instances of traveling sales-man problem and film-copy deliverer problem. Finally, the experimental result indicates that the new ACS with adaptive ρ performs better than GA [12], ACO [13] and ACS [2,14]. Furthermore, the convergence of the proposed ACO algorithm is proved.

2. Adaptive Volatility Rate of Pheromone Trail

The framework of ACO [1–2] is inspired by the ants' foraging behavior in selecting the shortest path between the nest and the food. Each ant builds a tour (i.e. a feasible solution to the TSP) by repeatedly applying a stochastic greedy rule (the state transition rule) as Equation (1) shows.

$$P_{gs}^{\ m}(t)=\begin{cases}\dfrac{[\tau_{gs}(t)]^{\alpha}[\eta_{gs}]^{\beta(g,t)}}{\sum_{r\in J_m(g)}[\tau_{gr}(t)]^{\alpha}[\eta_{gr}]^{\beta(g,t)}} & if \quad s\in J_k(g)\\ 0 & otherwise\end{cases} \tag{1}$$

where $P^{m}_{\ gs}$ is the probability with which the ant m chooses to move from city g to city s in iteration t, τ is the pheromone, $\eta=1/d$ is the reciprocal of distance d_{gs}, and $J_m(g)$ is the set of cities not having been visited yet when ant m is at city g.

After constructing its tour, an artificial ant also modifies the amount of pheromone on the visited edges by applying the pheromone updating rule. The rule is designed so that it tends to give more pheromone to the edges which should be visited by ants. The classical pheromone updating rule is:

$$\tau_{gs}(t+1)=(1-\rho)\tau_{gs}(t)+\rho\Delta\tau_{gs}(t) \tag{2}$$

where $\Delta\tau_{gs}(t)$ is the increment for the pheromone of edge (g,s) at the t-th iteration, and ρ is the volatility rate of the pheromone trail. The optimal ρ was set $\rho=0.1$ experimentally [1,2,4], which means that 90 per cent of the original pheromone trail remains and its 10 per cent is replaced by the increment.

In order to update the pheromone according to the quality of solutions found by ants, an adaptive rule for volatility of the pheromone trail is designed as follows:

$$\rho_m=L_m^{-1}/(L_m^{-1}+L_P^{-1}) \tag{3}$$

where L_m is the length of the solution S_m found by ant m, and L_P is the length of the solution S_P built based on the pheromone matrix, shown as Equation (4).

$$s=\arg\max_{u\in J_m(r)}\{[\tau(r,u)\} \tag{4}$$

where s is the city selected as the next one to city r for any $(r,s)\in S_P$.

The motivation of the proposed rule is: better solutions should contribute more pheromone, and the worse ones contribute less. We will use this rule to design a new ACO algorithm in the following section.

3. An ACO Algorithm with the Adaptive Parameter

In this section, a new ACO algorithm with the adaptive rule (shown as Equation 3) is introduced as follows:

Algorithm new ACO
input: An instance of TSP or FDP problems
Initialize solutions and pheromone value.
$S_{best}\leftarrow NULL$.
while termination conditions not met **do**
Construct S_P
for $i=1$ **to** k **do** {k is the number of artificial ants}
$S_i\leftarrow ConstructSolution(t)$.
ρ_i is calculated based on S_i.
if ($Length(S_i)<Length(S_{best})$) or ($S_{best}=NULL$) **then**
$S_{best}\leftarrow S_i$
Endif
Endfor
ρ_{best} is calculated based on S_{best}.
Carry out the pheromone updating rule with ρ_i ($i=1,...,k$) and ρ_{best}.
Endwhile
Output: S_{best}.
End_Algorithm

The framework of the proposed algorithm is similar to ant colony system (ACS) [2], so are the initialization, solution construction and setting of the parameters $q_0=0.9$, k=10, α=1 and $\beta=2$. There is only an updating rule in the algorithm shown as Equation 5 and 6.

$$\tau_{gs}(t+1)=(1-\rho_i)\tau_{gs}(t)+\rho_i L_i^{-1} \tag{5}$$

where $\forall(g,s)\in S_i$ and $\rho_i=L_i^{-1}/(L_i^{-1}+L_P^{-1})$ for the t-th iteration.

$$\tau_{gs}(t+1)=(1-\rho_{best})\tau_{gs}(t)+\rho_{best}L_{best}^{\ -1} \tag{6}$$

where $\forall(g,s)\in S_{best}$ and $\rho_{best}=L_{best}^{-1}/(L_{best}^{-1}+L_P^{-1})$ for the t-th iteration.

4. Convergence of the Proposed Algorithm

In this section, we give the convergence proof of the new ACO algorithm.

Given an arbitrary path (g,s),

$$\tau_{gs}(t)\le(1-\rho_1)\tau_{gr}(t)+\rho_1 U\le(1-\rho_1)^{t'}\tau_{gs}(0)+\frac{1-(1-\rho_1)^{t'}}{1-(1-\rho_1)}\rho_1 U \tag{7}$$

where $0 < t' \le t$, $\rho_1 = L_{\min}^{-1} / (L_{\min}^{-1} + L_{\max}^{-1})$, $U = \max\{\tau_{gs}(0), (L_{\min})^{-1}\}$, $L_{\max}$ is the length of the worst tour and $L_{\min}$ is the length of optimal tours.

$$\tau_{gs}(t) \ge (1-\rho_2)\tau_{gr}(t) + \rho_2 D \le (1-\rho_2)^{t'}\tau_{gs}(0) + \frac{1-(1-\rho_2)^{t'}}{1-(1-\rho_2)}\rho_2 D \tag{8}$$

where $0 < t' \le t$, $\rho_2 = L_{\max}^{-1} / (L_{\max}^{-1} + L_{\min}^{-1})$, $D = \min\{\tau_{gs}(0), (L_{\max})^{-1}\}$.

Because $0 < \rho_1, \rho_2 < 1$, $D \le \tau_{gs}(t) \le U$ when $t \to \infty$. Therefore, $\tau_{gs}(t)$ has an upper boundary and a lower boundary, we assume $0 < P_{low} \le \tau_{gs}(t) \le P_{high} < +\infty$ without a loss of generality.

When S^* is the optimal solution to a n-city TSP and $(a,b) \in S^*$ as an arbitrary path, the probability $P_{ab}(t_0)$, with which (a,b) is found by artificial ant in iteration $t_0 (t_0 > 0)$, can meet:

$$P_{ab}(t_0) \ge P\{q > q_0\} \frac{{\tau_{ab}}^{\alpha}(t_0) \cdot {\eta_{ab}}^{\beta}}{\sum_{j \in J(a)} {\tau_{aj}}^{\alpha}(t_0) \cdot {\eta_{aj}}^{\beta}} \tag{9}$$

$$P_{ab}(n_0) \ge p_0 \frac{\tau_{ab}(t_0) \cdot {\eta_{ab}}^{\beta}}{\sum_{j \in J(a)} \tau_{aj}(t_0) \cdot {\eta_{aj}}^{\beta}} \ge p_0 \frac{[{\tau_a}^{\min}(t_0)] \cdot [{\eta_a}^{\min}]^{\beta}}{\sum_{j \in J(a)} P_{high} \cdot [{\eta_a}^{\max}]^{\beta}} \ge p_0 \frac{P_{low} \cdot {\eta_{\min}}^{\beta}}{k P_{high} {\eta_{\max}}^{\beta}} \tag{10}$$

where $p_0 = P\{q > q_0\}$ [2], $\eta_{\min} = \min_{(i,j) \in S^*}\{\eta_{ij}\}$ and $\eta_{\max} = \max_{(i,j) \in S^*}\{\eta_{ij}\}$.

Given $a_0 = p_0 \frac{P_{low} \cdot {\eta_{\min}}^{\beta}}{k P_{high} {\eta_{\max}}^{\beta}} < 1$, the probability, by which S^* can be found by ants in iteration t_0, is $P_{S^*}(t_0) = \prod_{(a,b) \in S^*} P_{ab}(n_0) \ge {a_0}^{n-1}$, where n is the number of cities. The probability, by which S^* can never be found from iteration t_0, is:

$$\begin{aligned} \tilde{P}_{S^*}(t_0) &= \prod_{t=t_0}^{\infty}[1 - \prod_{(a,b) \in S^*} P_{ab}(t_0)]^k \\ &\le \prod_{t=t_0}^{\infty}[1 - {a_0}^{n-1}]^k = 0 \end{aligned} \tag{11}$$

where k is the number of artificial ants and t_0 can be arbitrary.

Hence, S^* can be found by probability one when the iteration $t \to \infty$, which theoretically confirms the capacity of global optimization of the proposed ACO algorithm.

5. Numerical Results

This section indicates the numerical results in the experiment that the proposed ACO algorithm is used to solve TSP problems [15] and FDP problems [14]. Other approaches for the problems ACS [2], ACO [13], GA-FDP [12] and ACS-FDP [14] are also tested in the same machines as the comparison with the proposed ACO.

Several TSP instances are computed by ACS [2], ACO [13] and the proposed ACO on a PC with an Intel Pentium 550MBHz Processor and 256MB SDR Memory, and the results are shown in Table 1. It should be noted that every instance is computed 20 times. The algorithms are both programmed in Visual C++6.0 for Windows System. They would not stop until a better solution could be found in 500 iterations, which is considered as a virtual convergence of the algorithms.

Table 1 shows that the proposed ACO algorithm (PACO) performs better than ACS [2] and ACO [13]. The shortest lengths and the average lengths obtained by PACO are shorter than those found by ACS and ACO in all of the TSP instances. Furthermore, it can be concluded that the standard deviations of the tour lengths obtained by PACO are smaller than those of another algorithms. Therefore, we can conclude that PACO is proved to be more effective and steady than ACS [2] and ACO [13]. Computation time cost of PACO is not less than ACS and ACO in all of the instances because it needs to compute the value of volatility rate $k+1$ times per iteration. Although all optimal tours of TSP problems cannot be found by the tested algorithms, all of the errors for PACO are much less than that for another two ACO approaches. The algorithms may make improvement in solving TSP when reinforcing heuristic strategies like local search like ACS-3opt [2] and MMAS+rs [3] are used.

FDP problem is an extended style of TSP problem. Two FDP instances in the literature [14] are computed by GA-FDP [12], ACS-FDP [14] and the proposed ACO-FDP on a PC with an Intel Pentium 400MBHz Processor and 128 MB EMS memory, and the results are shown in Table 2. It should be noted that every instance is computed 20 times. The algorithms are both programmed in Visual C++6.0 for Windows System. They would not stop until a better solution could be found in 500 iterations, which is considered as a virtual convergence of the algorithms.

Table 1. Comparison of the results obtained by ACS [2], ACO [3] and the proposed ACO (PACO) in TSP instances.

Problem	Algorithm	best	ave	time(s)	standard deviation
kroA100	ACS	21958	22088.8	65	1142.77
	ACO	21863	22082.5	94.6	1265.30
	PACO	21682	22076.2	117.2	549.85
ts225	ACS	130577	133195	430.6	7038.30
	ACO	130568	132984	439.3	7652.80
	PACO	130507	131560	419.4	1434.98
pr226	ACS	84534	86913.8	378.4	4065.25
	ACO	83659	87215.6	523.8	5206.70
	PACO	81967	83462.2	762.2	3103.41
lin105	ACS	14883	15125.4	88.8	475.37
	ACO	14795	15038.4	106.6	526.43
	PACO	14736	14888	112.2	211.34
kroB100	ACS	23014	23353.8	56.2	685.79
	ACO	22691	23468.1	102.9	702.46
	PACO	22289	22728	169.6	668.26
kroC100	ACS	21594	21942.6	54.8	509.77
	ACO	21236	21909.8	78.1	814.53
	PACO	20775	21598.4	114.8	414.62
lin318	ACS	48554	49224.4	849.2	1785.21
	ACO	48282	49196.7	902.7	2459.16
	PACO	47885	49172.8	866.8	1108.34

Table 2. Comparison of the results obtained by GA-FDP [12], ACS-FDP [14] and the proposed ACO-FDP in FDP instances [14].

Problem	Algorithm	best	ave	time(s)
Problem I	GA-FDP	4240.67	4261.4	153
	ACS-FDP	4122.33	4138.5	78
	ACO-FDP	4122.33	4126.2	80
Problem II	GA-FDP	4208	4250.6	184
	ACS-FDP	4163	4289.2	130
	ACO-FDP	4163	4165.8	135

The results in Table 2 indicate that PACO-FDP performs better than GA-FDP [12] and ACS-FDP [14] in the item of average length though it cannot find better solution than ACS-FDP [14]. PACO-FDP can be also considered as the improvement of ACS-FDP because the special strategies [14] are also used in PACO-FDP.

6. Discussions and Conclusions

This paper proposed an adaptive rule for volatility rate of pheromone trail, attempting to adjust the pheromone based on the solutions obtained by artificial ants. Thus, a new ACO algorithm is designed with this tuning rule. There is a special pheromone updating rule in the proposed algorithm whose framework is similar to Ant Colony System. Then, the convergence of the proposed ACO algorithm is proved to ensure its capacity of global capacity. Moreover, there are some experimental comparisons among the proposed ACO approach and other methods [2,12–14] in solving TSP and FDP problems. The results also show the effectiveness of the proposed algorithm.

Further study is suggested to explore the better management for the optimal setting of the parameters of ACO algorithms, which will be very helpful in the application.

7. Acknowledgements

This work has been supported by Natural Science Foundation of Guangdong Province (9151600301000001), Key Sci-tech Research Projects of Guangdong Province (2009B010800026), funded project of State Key Lab. for Novel Software Technology of Najing University and student research project (SRP) of South China University of Technology.

8. References

[1] M. Dorigo, G. D. Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," Massachusetts Institute of Technology, Artificial Life 5, pp. 137–172, 1999.

[2] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, pp. 53–66, 1997.

[3] T. Stützle and H. H. Hoos, "MAX-MIN ant system, future gener," Computer System, Vol. 16, No. 8, pp. 889–914, 2000.

[4] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," Theoretical Computer Science, Vol. 344, pp. 243–278, 2005.

[5] A. C. Zecchin, A. R. Simpson, H. R. Maier, and J. B. Nixon, "Parametric study for an ant algorithm applied to water distribution system optimization," IEEE Transactions on Evolutionary Computation, Vol. 9, No. 2, April 2005.

[6] M. Dorigo and L. M. Gambardella, "Ant colonies for the traveling salesman problem," Bio-systems, Vol. 43, pp. 73–81, 1997.

[7] K. M. Sim and W. H. Sun, "Ant colony optimization for routing and load-balancing: Survey and new directions," IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, Vol. 33, No. 5, September 2003.

[8] I. Watanabe and S. L. Matsui, "Improving the performance of ACO algorithms by adaptive control of candidate set, evolutionary computation," CEC'03, Vol. 2, pp. 1355–1362, 2003.

[9] M. L. Pilat and T. White, "Using genetic algorithms to optimize ACS-TSP," M. Dorigo *et al.* (Eds.):ANTS'02, LNCS 2463, pp. 282–287, 2002.

[10] L. M. Gambardella and M. Dorigo, "Ant-Q: A reinforcement learning approach to the traveling salesman problem," Appeared in: Proceedings of ML-95, Twelfth Intern. Conference on Machine Learning, Morgan Kaufmann, pp. 252–260, 1995.

[11] H. Huang and Z. F. Hao, "A bi-directional searching ant colony system," Proceedings of 2006 International Conference on Intelligent Systems and Knowledge Engineering (ISKE'06), April 6-7, 2006.

[12] R. W.Cheng and M. Gen, "Film-copy deliverer problem using genetic algorithms," Computers & Industrial Engineering, Vo1. 29, No. l-4, pp. 549–553, 1995

[13] J. Sun, S. W. Xiong, and F. M. Guo, "A new pheromone updating strategy in ant colony optimization," Proceedings of 2004 International Conference on Machine Learning and Cybernetics, Vol. 1, pp. 620–625, 2004.

[14] Z. F.Hao, H. Huang, X. W. Yang, Y. C. Liang, "Solve the film-copy deliverer problem using ant colony system," Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.

[15] G. Reinelt, "A traveling salesman problem library," ORSA Journal on Computing, TSPLIB, Vol. 3, No. 4, pp. 376–384, 1991.

Int. J. Communications, Network and System Sciences, **2009, 2, 797-803**
doi:10.4236/ ijcns.2009.28093 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

A Review of Wireless Body Area Networks for Medical Applications

Sana ULLAH[1], Pervez KHAN[1], Niamat ULLAH[1], Shahnaz SALEEM[2], Henry HIGGINS[3], Kyung Sup KWAK[1]
[1]*Graduate School of Telecommunication Engineering, Inha University Incheon, Nam-Gu, South Korea*
[2]*Graduate School of Computer Engineering, Inha University Incheon, Nam-Gu, South Korea*
[3]*Zarlink Semiconductor Company, Portskewett, Caldicot, United Kingdom*
Email: {*sanajcs, pervaizkanju, roshnee*13}*@hotmail.com, niamatnaz@gmail.com, kskwak@inha.ac.kr, henry.higgins@zarlink.com*
Received March 8, 2009; *revised May* 16, 2009; *accepted July* 27, 2009

Abstract

Recent advances in Micro-Electro-Mechanical Systems (MEMS) technology, integrated circuits, and wireless communication have allowed the realization of Wireless Body Area Networks (WBANs). WBANs promise unobtrusive ambulatory health monitoring for a long period of time, and provide real-time updates of the patient's status to the physician. They are widely used for ubiquitous healthcare, entertainment, and military applications. This paper reviews the key aspects of WBANs for numerous applications. We present a WBAN infrastructure that provides solutions to on-demand, emergency, and normal traffic. We further discuss in-body antenna design and low-power MAC protocol for a WBAN. In addition, we briefly outline some of the WBAN applications with examples. Our discussion realizes a need for new power-efficient solutions towards in-body and on-body sensor networks.

Keywords: Wireless Body Area Networks, Low Power MAC, Body Sensor Networks, BSN, WBAN

1. Introduction

Cardiovascular disease is the foremost cause of death in the United States (US) and Europe since 1900. More than ten million people are affected in Europe, one million in the US, and twenty two million people in the world [1–3]. The number is projected to be triple by 2020. The ratio is 17% in South Korea and 39% in UK [4–5]. The healthcare expenditure in the US is expected to increase from $2.9 trillion in 2009 to $4 trillion in 2015 [6]. The impending health crisis attracts researchers, industrialists, and economists towards optimal and quick health solutions. The non-intrusive and ambulatory health monitoring of patient's vital signs with real time updates of medical records via internet provides economical solutions to the health care systems.

A WBAN contains a number of portable, miniaturised, and autonomous sensor nodes that monitors the body function for sporting, health, entertainment, and emergency applications. It provides long term health monitoring of patients under natural physiological states without constraining their normal activities. In-body sensor networks allow communication between implanted devices and remote monitoring equipments. They are used to collect information from Implantable Cardioverter Defibrillators (ICDs) in order to detect and treat ventricular tachyarrhythmia[1] and to prevent Sudden Cardiac Death (SCD) [7].

A number of ongoing projects such as CodeBlue, MobiHealth, and iSIM have contributed to establish a proactive WBAN system [8–10]. A system architecture presented in [11] performs real-time analysis of sensor's data, provides real-time feedback to the user, and forwards the user's information to a telemedicine server. UbiMon aims to develop a smart and affordable health care system [12]. MIT Media Lab is developing MIThril that gives a complete insight of human-machine interface [13] HIT lab focuses on quality interfaces and innovative wearable computers [14]. NASA is developing a wearable physiological monitoring system for astronauts called LifeGuard system [15]. IEEE 802.15.6 aims to provide low-power in-body and

[1]Ventricular tachyarrhythmia are abnormal patterns of electrical activity originating within ventricular tissue.

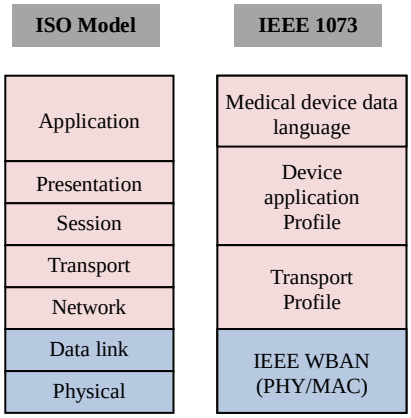


Figure 1. Model ISO and IEEE 1073.

on-body wireless communication standards for medical and non-medical applications [16]. IEEE 1073 is working towards a seven layers solution for wireless communication in a WBAN [17]. Figure 1 shows IEEE 1073 model.

The rest of the paper is organized into five sections. Section 2 presents a WBAN infrastructure for medical and non-medical applications. Section 3 and 4 discuss in-body antenna design and low-power MAC protocol for a WBAN. Section 5 outlines some of the WBAN applications. The final section concludes our work.

2. WBAN Infrastructure

A WBAN consists of in-body and on-body nodes that continuously monitor patient's vital information for diagnosis and prescription. Some on-body nodes are used for multimedia and gaming applications.

A WBAN uses Wireless Medical Telemetry Services (WMTS), unlicensed Industrial, Scientific, and Medical (ISM), Ultra-wideband (UWB), and Medical Implant Communications Service (MICS) bands for data transmission. WMTS is a licensed band used for medical telemetry system. Federal Communication Commission (FCC) urges the use of WMTS for medical applications due to fewer interfering sources. However, only authorized users such as physicians and trained technicians are eligible to use this band. Furthermore, the restricted WMTS (14 MHz) bandwidth cannot support video and voice transmissions. The alternative spectrum for medical applications is to use 2.4 GHz ISM band that includes guard bands to protect adjacent channel interference. A licensed MICS band (402-405 MHz) is dedicated to the implant communication.

Figure 2 shows the proposed WBAN infrastructure for medical and non-medical applications.

The WBAN traffic is categorized into On-demand, Emergency, and Normal traffic. **On-demand traffic** is initiated by the coordinator or doctor to acquire certain

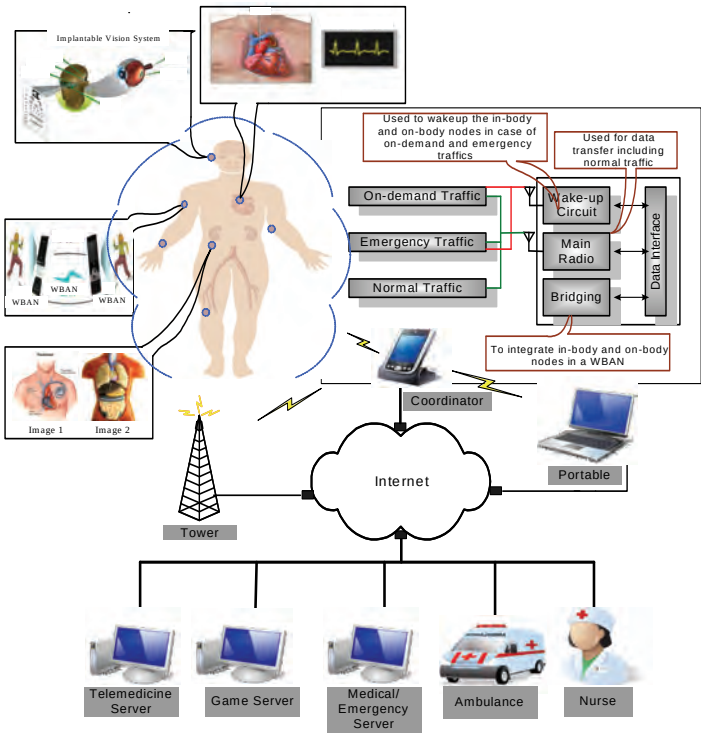


Figure 2. A WBAN infrastructure for medical and non-medical applications.

information, mostly for the purpose of diagnostic recommendations. This is further divided into continuous (in case of surgical events) and discontinuous (when occasional information is required). **Emergency traffic** is initiated by the nodes when they exceed a predefined threshold and should be accommodated in less than one second. This kind of traffic is not generated on regular intervals and is totally unpredictable. **Normal traffic** is the data traffic in a normal condition with no time critical and on-demand events. This includes unobtrusive and routine health monitoring of a patient and treatment of many diseases such as gastrointestinal tract, neurological disorders, cancer detection, handicap rehabilitation, and the most threatening heart disease. The normal data is collected and processed by the coordinator. The coordinator contains a wakeup circuit, a main radio, and a bridging function, all of them connected to a data interface. The wakeup circuit is used to accommodate on-demand and emergency traffic. The Bridging function is used to establish a logical connection between different nodes working on different frequency bands. The coordinator is further connected to telemedicine, game, and medical servers for relevant recommendations.

3. In-Body Antenna Design

The band designated for in-body communication is MICS and is around 403MHz. The wavelength of this frequency in space is 744mm so a half wave dipole will be 372mm. Clearly, it is not possible to include an antenna of such dimensions in a body [19]. These constraints make the available size much smaller than the optimum.

The electrical properties of a body affect the propagation in several ways. First, the high dielectric constant increases the "electrical length" of E-field antennas such as a dipole. Second, body tissue such as muscle is partly conductive and will absorb some of the signal but it can also act as a parasitic radiator. This is significant when the physical antenna is much smaller than the optimum. Typical dielectric constant (ε_r), conductivity (ρ) and characteristic impedance $Z_0(\Omega)$ properties of muscle and fat are shown in Table 1.

1) Dipole Antenna: For a dipole of length 10mm, at 403MHz, the radiation resistance is 45m.Ω in air. The electrical length of the dipole is increased when surrounded by material of a high dielectric constant such as the body.

2) Loop Antenna: For a loop of 10mm diameter the area is 78.5mm^2, this gives the radiation resistance of 626μΩ. However, the loop acts, as a "magnetic dipole" producing a more intense magnetic field than a dipole. The loop is of use within the body as the magnetic field is less affected by the body tissue compared to a dipole or a patch and it can be readily integrated into existing structures.

3) Patch Antenna: A patch antenna can be integrated into the surface of an implant. Without requiring much additional volume, the ideal patch will have dimensions as shown in Figure 3 and acts as a λ/2 parallel-plate transmission line with an impedance inversely proportional to the width.

The radiation occurs at the edges of the patch, as

Table 1. Body electrical properties [19].

Frequency	Muscle			Fat		
	(ε_r)	ρ (S.m^{-1})	$Z_0(\Omega)$	(ε_r)	ρ	$Z_0(\Omega)$
100	66.2	0.73	31.6	12.7	0.07	92.4
400	58	0.82	43.7	11.6	0.08	108
900	56	0.97	48.2	11.3	0.11	111

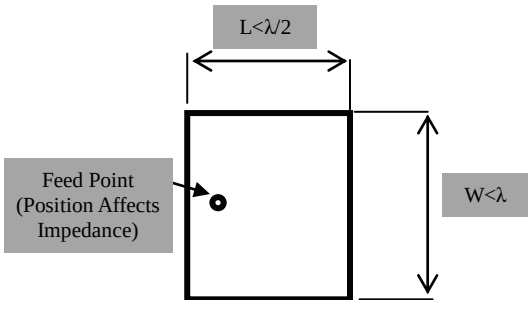


Figure 3. Patch antenna plan view, λ in the surrounding medium.

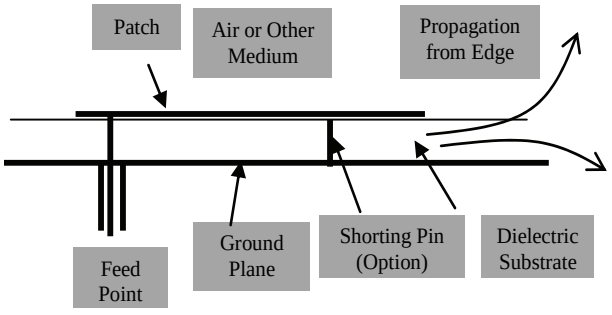


Figure 4. Patch antenna side view.

shown in Figure 4. For in-body use a full size patch is not an option. However, as it is immersed in a body tissue that has a dielectric constant in the order of 50, the electrical size of the patch becomes larger than would be in air. An electrically small patch will have low real impedance and therefore impaired performance compared to the ideal one. There are several other options for antenna such as Planar Inverted-F Antenna (PIFA), loaded PIFA, the bow tie, spiral and trailing wire. These antennas may have properties that may make them better suited for some applications.

4) Impedance Measurement: The impedance of the patch and dipole will be affected considerably by being surrounded by the body tissue. The doctor who fits it determines the position of an implant within a body. It may move within the body after fitting. Each body has a different shape with different proportions of fat and muscle that may change with time. This means that a definitive measurement of antenna impedance is of little value. Measuring it immersed in a body phantom can make an approximation of impedance liquid [20]. Using this impedance, the antenna-matching network can be designed with the provision of software controlled trimming as can be done with variable capacitors integrated into the transceiver. The trimming routine should be run on each power up or at regular intervals to maintain optimum performance.

4. MAC Protocol

The design and implementation of a low-power MAC protocol for a WBAN is currently a hot research topic. The most challenging task is to accommodate the in-body nodes in a power-efficient manner. Unlike on-body nodes, the in-body nodes are implanted under human skin where the electrical properties of the body affect the signal propagations. The human body is a medium that poses many wireless transmission challenges. The body is composed of several components that are unpredictable and subjected to change.

Li *et al.* proposed a novel TDMA protocol for an on-body sensor network that exploits the biosignal features to perform TDMA synchronization and improves the energy efficiency [21]. Other protocols like WASP, CICADA, and BSN-MAC are proposed in [22–24]. The performance of a non-beacon IEEE 802.15.4 is investigated in [25], where the authors considered low upload/download rates, mostly per hour. Furthermore, the data transmission is based on periodic intervals that limit the performance to certain applications. There is no reliable support for on-demand and emergency traffic.

The WBAN traffic requires sophisticated low-power techniques to ensure safe and reliable operations. Existing MAC protocols such as SMAC [26], TMAC [27], IEEE 802.15.4 [28], and WiseMAC [29] give limited answers to the heterogeneous traffic. The in-body nodes do not urge synchronized wakeup periods due to sporadic medical events. Medical data usually needs high priority and reliability than non-medical data. In case of emergency events, the nodes should access the channel in less than one second [30]. IEEE 802.15.4 Guaranteed Time Slots (GTS) can be utilized to handle time critical events but they expire in case of a low traffic. Furthermore, some in-body nodes have high data transmission frequency than others. Figure 5 shows the required MAC mapping of the WBAN traffic.

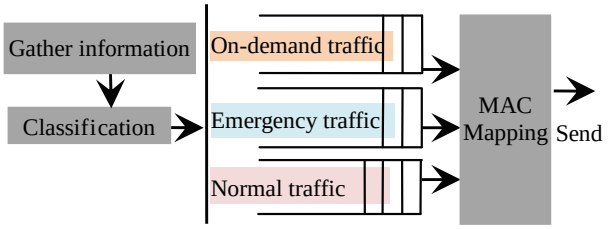


Figure 5. WBAN MAC mapping.

The IEEE 802.15.4 can be considered for certain on-body sensor network applications but this does not achieve the required power level of in-body nodes. For critical and non-critical medical traffic, the IEEE 802.15.4 has several power consumption and QoS issues [31–34]. Also, this standard operates in 2.4 GHz band, which allows the possibilities of interference from other devices such as IEEE 802.11 and microwave. Dave *et al.* studied the energy efficiency and QoS performance of IEEE 802.15.4 and IEEE 802.11e [35] MAC protocols under two generic applications: a wave-form real time stream and a real-time parameter measurement stream [36]. Table 2 shows the Packet Delivery Ratio and the Power (in mW) for both applications. The *AC_BE* and *AC_VO* represent the access categories voice and best-effort in the IEEE 802.11e.

IEEE 802.15.4 uses CSMA/CA mechanism that does not provide reliable solutions in the in-body sensor networks. The path loss inside human body results in improper Clear Channel Assessment (CCA). For a thresh-

Table 2. Packet delivery ratio and power (in mW).

Sensor Nodes		IEEE 802.15.4	IEEE 802.11e (AC_BE)	IEEE 802.11e (AC_VO)
Packet Delivery Ratio	Wave-form	100%	100%	100%
	Parameter	99.77%	100%	100%
Power (mW)	Wave-form	1.82	4.01	3.57
	Parameter	0.26	2.88	2.77

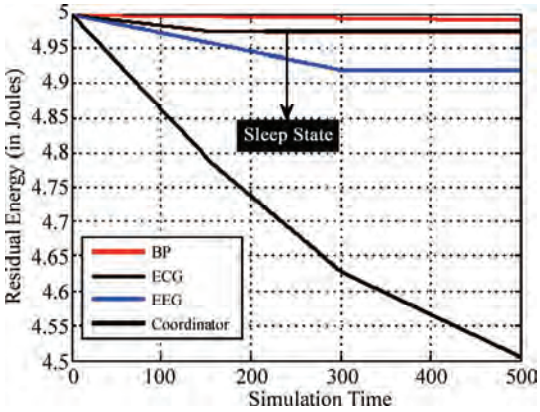


Figure 6. Residual emergency at on-body nodes.

old of -85dBm and -95dBm, the on-body nodes cannot see the activity of in-body nodes when they are away at 3 meters distance from the body surface [37]. An alternative solution is to use TDMA-based protocols for a WBAN. Therefore, we analyze the performance of a preamble-based TDMA [38] protocol for an on-body sensor network. We use ns-2 [39] for extensive simulations. Figure 6 shows the residual energy at the on-body nodes and the coordinator. After the nodes finish their transmissions, they go into sleep mode. The ECG node sleeps after 150 seconds. When the EEG node finishes its transmission at 300 seconds, the coordinator consumes less energy as indicated by the slight change in the curve.

5. WBAN Applications

WBANs have great potential for several applications including remote medical diagnosis, interactive gaming, and military applications.

Table 3 shows some of the in-body and on-body applications [40]. In-body applications include, monitoring and program changes for pacemakers and implantable cardiac defibrillators, control of bladder function, and restoration of limb movement [41]. On-body medical applications include monitoring ECG, blood pressure, temperature, and respiration. Furthermore, on-body non-medical applications include monitoring forgotten things, establishing a social network, and assessing soldier fatigue and battle readiness.

The following part discusses some of the WBAN applications:

1) Cardiovascular Diseases: Traditionally, holter monitors were used to collect cardio rhythm disturbances for offline processing without real-time feedback. However, transient abnormalities are sometimes hard to capture. For instance, many cardiac diseases are associated with episodic rather than continuous abnormalities, such as transient surges in blood pressure, paroxysmal arrhythmias or induced episodes of myocardial ischemia and their time cannot be accurately predicated [42]. A WBAN is a key technology to prevent the occurrence of myocardial infarction, monitor episodic events or any other abnormal condition and can be used for ambulatory health monitoring.

2) Cancer Detection: Cancer remains one of the biggest threats to the human life. According to National Center for Health Statistics, about 9 million people had cancer diagnosis in 1999 [43]. A set of miniaturised sensors capable of monitoring cancer cells can be seamlessly integrated in a WBAN. This allows physician to diagnose tumors without biopsy.

3) Asthma: A WBAN can help millions of patients suffering from asthma by monitoring allergic agents in the

Table 3. In-body and on-body sensor networks applications.

Application Type	Sensor Node	Date Rate	Duty Cycle (per device)% per time	Power Consumption	QoS (Sensitive to Latency)	Privacy
In-body Applications	Glucose Sensor	Few Kbps	<1%	Extremely Low	Yes	High
	Pacemaker	Few Kbps	<1%	Low	Yes	High
	Endoscope Capsule	>2Mbps	<50%	Low	Yes	Medium
On-body Medical Applications	ECG	3kbps	<10%	Low	Yes	High
	SpO2	32bps	<1%	Low	Yes	High
	Blood Pressure	<10bps	<1%	High	Yes	High
On-body Non-Medical Applications	Music for Headsets	1.4Mbps	High	Relatively High	Yes	Low
	Forgotten Things Monitor	256kbps	Medium	Low	No	Low
	Social Networking	<200kbps	<1%	Low	Low	High

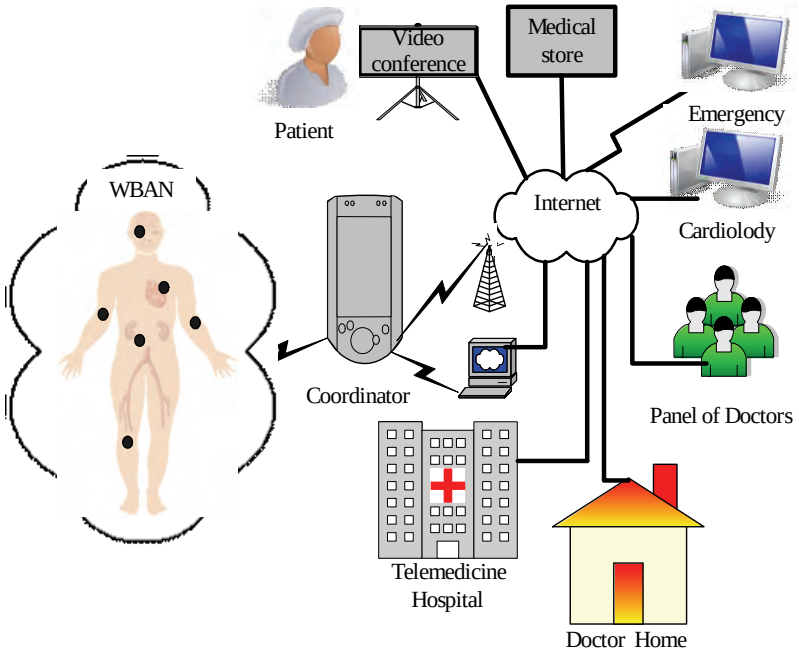


Figure 7. A real-time telemedicine infrastructure for patient rehabilitation.

air and providing real-time feedback to the physician. Chu *et al* proposed a GPS-based device that monitors environmental factors and triggers an alarm in case of detecting information allergic to the patient [44].

4) Telemedicine Systems: Existing telemedicine systems either use dedicated wireless channels to transfer information to the remote stations, or power demanding protocols such Bluetooth that are open to interference by other devices working in the same frequency band. These characteristics limit prolonged health monitoring. A WBAN can be integrated into a telemedicine system that supports unobtrusive ambulatory health monitoring for long period of time. Figure 7 shows a real-time telemedicine infrastructure for patient rehabilitation.

5) Artificial Retina: Retina prosthesis chips can be implanted in the human eye that assists patient with limited or no vision to see at an adequate level.

6) Battlefield: WBANs can be used to connect soldiers in a battlefield and report their activities to the commander, i.e., running, firing, and digging. The soldiers should have a secure communication channel in order to prevent ambushes.

6. Conclusions

In this paper, we proposed a WBAN infrastructure that supports on-demand, emergency, and normal traffic using wakeup and main radios. This infrastructure proves to be adequate for unobtrusive health monitoring. We further provided a technical discussion on the in–body antenna design and supported patch antenna for in-body communication. We also discussed low-power MAC protocol for a WBAN. Existing low-power MAC protocols have several limitations to accommodate the heterogeneous traffic in a reliable manner and hence require new power-efficient solutions. We finally outlined the potential of a WBAN for ubiquitous healthcare, entertainment, and military applications.

7. References

[1] J. G. Cleland, K. Swedberg, and F. Follath, “A survey of the quality of care among patients with heart failure in Europe. Part 1: Patient characteristics and diagnosis,” The Euro Heart Failure Survey Programme, Euro Heart Journal, pg 24, pp. 442–463, 2003.

[2] http://www.foxnews.com/story/0,2933,142436,00.html Date Visited, 15 December 2008.

[3] Heart Failure Facts and Figures: OU Medical Centre.

[4] http://www.who.int/whosis/mort/profiles/mortwprokor repofkorea.pdf.

[5] A. Barroso, J. Benson, *et al*., “The DSYS25 sensor platform,” In Proceedings of the ACM sensys, 2004.

[6] C. Borger, *et al*.: “Health spending projections through 2015: Changes on the horizon,” In Health Affairs Web Exclusive W61, February 22, 2006.

[7] Electromagnetic compatibility and Radio spectrum Matters (ERM), ETSI TR 102 655, pg 21, 2008.

[8] http://fiji.eecs.harvard.edu/CodeBlue, Date Visited, 21 November 2008.

[9] http://www.mobihealth.org, Date Visited, 20 January 2009.

[10] http://www.cs.uoregon.edu/research/wearables/index.htm

l, Date Visited, 21 July 2008.

[11] E. Jovanov, A. Milenkovic, C. Otto, and P. de Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," Journal of Neuro Engineering and Rehabilitation, Vol. 2, No. 6, March 2005.

[12] http://www.ubimon.net Date Visited, 16 March 2008.

[13] http://www.media.mit.edu/wearables/mithril, Date Visited, 09 February 2008.

[14] http://www.hitl.washington.edu, Date Visited, 25 April 2008.

[15] http://lifeguard.stanford.edu, Date Visited, 11 January 2009.

[16] http://www.ieee802.org/15/pub/TG6.html, Date Visited, 16 December 2008

[17] IEEE P1073.0.1.1/D01J, "Draft guide for health informatics-point-of-care medical device," Communication-technical report-Guidelines for the use of RF wireless technology, 2006.

[18] http://www.wireless-world-research.org/?id=92, Date Visited, 12 February 2009.

[19] G.-Z Yang, "Body sensor networks," Springer, pp. 117–143, 2006.

[20] J. Wojclk, *et al.*, "Tissue recipe calibration requirements, SSI/DRB-TP-D01-003," Spectrum Sciences Institute RF Dosemetry Research Board, 51 Spectrum Way, Nepean, Ontario, K2R 1E6, Canada.

[21] H. M. Li, H Jindong H. Tan, "Heartbeat driven medium access control for body sensor networks," Proceedings of the 1st ACM SIGMOBILE, International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments, pp. 25–30, 2007.

[22] Braem, B. Latre, I. Moerman, C. Blondia, and P. Demeester, "The Wireless Autonomous Spanning tree Protocol for multihop wireless body area networks," in Proceedings of the First International Workshop on Personalized Networks. San Jose, California, USA, 2006.

[23] B. Latre, B. Braem, I. Moerman, C. Blondia, E. Reusens, W. Joseph, and P. Demeester, "A low-delay protocol for multihop wireless body area networks," Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services, Mobi Quitous, pp. 1–8, 6–10, 2007.

[24] H. M. Li and J. D. Tan, "Medium access control for body sensor networks," Computer Communications and Networks, ICCCN'07, pp. 210–215, 13–16 August 2007.

[25] N. F. Timmons and W. G. Scanlon, "Analysis of the performance of IEEE 802.15.4 for medical sensor body area networking," IEEE SECON, 2004.

[26] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," IEEE Communications Letters, Vol. 12 No. 3, pp. 493–506, 2004.

[27] T. V. Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in SenSys'03, Los Angeles, pp. 171–180, 2003.

[28] IEEE Std.802.15.4, "Wireless medium access control (MAC) and physical layer (PHY) specifications for low data rate wireless personal area networks (WPAN)," 2006.

[29] A. El-Hoiydi and J.-D. Decotignie, "WiseMAC: An ultra low power MAC protocol for the downlink of infrastructure wireless sensor networks," in the Proceedings of the Ninth IEEE Symposium on Computers and Communication, ISCC'04, Alexandria, Egypt, pp. 244–251, 2004.

[30] Technical Requirement Document, IEEE 802.15.6, January 2009.

[31] A. Sikora and V. Groza, "Coexistence of IEEE 802.15.4 with other systems in the 2.4 GHz ISM-band," IEEE IMTC Proceeding, May 2005.

[32] N. Golmie, D. Cypher, and O. Rebala, "Performance analysis of low rate wireless technologies for medical applications," Computer Communications, Vol. 28, No. 10, pp. 1255–1275, June 2005.

[33] N. Chevrollier, N. Montavont, and N. Golmie, "Handovers and interference mitigation in healthcare environments," IEEE MILCOM Proceeding, October 2005.

[34] Howitt and J. Gutierrez, "IEEE 802.15.4 low rate-wireless personal area network coexistence issues," IEEE WNCN Proceeding, 2003.

[35] IEEE 802.11e Std, "Amendment to Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," Medium Access Control Quality of Services Enhancements, November 2005.

[36] D. Cavalcanti, R. Schmitt, and A. Soomro, "Performance analysis of 802.15.4 and 802.11e for body sensor network applications," 4th International Workshop on Wearable and Implantable Body Sensor Networks, BSN, 2007

[37] B. Zhen, H. B. Li, and R. Kohno, **"**IEEE body area networks and medical implant communications," Proceedings of the ICST 3rd International Conference on Body Area Networks, Tempe, Arizona, 2008.

[38] S. Ullah, R. Islam, *et al.*, "Performance analysis of a preamble based TDMA protocol for wireless body area network", Journal of Communications Software and Systems, Vol. 4, No. 3, pp. 222–226, 2008.

[39] http://www.isi.edu/nsnam/ns, Date Visited, 17 March 2009.

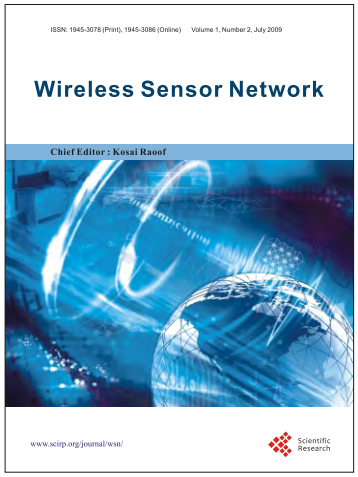
[40] IEEE 15-08-0644-07-0006-TG6, Technical Requirements Document, 2008

[41] S. Ullah, H. Higgins, Y. W. Cho, H. S. Lee, and K. S. Kwak, "Towards RF communication and multiple access protocols in a body sensor network," JDCTA: International Journal of Digital Content Technology and its Applications, Vol. 2, No. 3, pp. 9–16, 2008.

[42] B. Lo and G. Z. Yang, "Key technical challenges and current implementations of body sensor networks," IEEE Proceedings of the 2nd International Workshop on Body Sensor Networks (BSN'05), pp. 1–5, April 2005.

[43] National Center for Health Statistics, URL: http://www.cdc.gov/nchs/Default.htm, Date Visited, 12 March 2009

[44] H.-T. Chu, C.-C. Huang, Z.-H. Lian, and T. J. P. Tsai, "A ubiquitous warning system for asthma-inducement," in IEEE International Conference on Sensor networks, Ubiquitous and Thrustworthy Computing, Taichung, Taiwan, pp. 186–191, 2006.



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 System
- 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

Call for Papers

International Journal of

Communications, Network and System Sciences (IJCNS)

ISSN 1913-3715 (Print) ISSN 1913-3723 (Online)

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

IJCNS is an international refereed journal dedicated to the latest advancement of communications and network technologies. The goal of this journal is to keep a record of the state-of-the-art research and promote the research work in these fast moving areas.

Editors-in-Chief

Subject Coverage

This journal invites original research and review papers that address the following issues in wireless communications and networks. Topics of interest include, but are not limited to:

MIMO and OFDM technologies	Sensor networks
UWB technologies	Ad Hoc and mesh networks
Wave propagation and antenna design	Network protocol, QoS and congestion control
Signal processing and channel modeling	Efficient MAC and resource management protocols
Coding, detection and modulation	Simulation and optimization tools
3G and 4G technologies	Network security

We are also interested in:

- Short reports—Discussion corner of the journal:
  2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data.
- Book reviews—Comments and critiques.

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. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/ijcns ijcns@scirp.org

Int. J. Communications, Network and System Sciences, **2009, 2, 695-803**
Published Online November 2009 in SciRes (http://www.SciRP.org/journal/ijcns/).

TABLE OF CONTENTS

Volume 2 Number 8 **November 2009**

A Comparative Study of Medium Access Control Protocols for Wireless Sensor Networks
M. GUNN, S. G. M. KOO..... 695

Service Adaptable 3G Turbo Decoder for Indoor/Low Range Outdoor Environment
C. CHAIKALIS, N. S. SAMARAS..... 704

Application of a Dynamic Identity Authentication Model Based on an Improved Keystroke Rhythm Algorithm
W. C. YANG, F. FANG..... 714

Evaluation of Network Stack Optimization Techniques for Wireless Sensor Networks
J. JEONG..... 720

A Cooperative Location Management Scheme for Mobile Ad Hoc Networks
D. LI, J. C. WANG, L. P. ZHANG, H. LI, J. ZHOU..... 732

A Perceptual Approach to Reduce Musical Noise Using Critical Bands Tonality Coefficients and Masking Thresholds
C. V. R. RAO, M. B. R. MURTHY, K. S. RAO..... 742

A Real-Time Measurement Algorithm for Available Bandwidth
Y. YIN, W. D. WU..... 746

Modified Ceiling Bounce Model for Computing Path Loss and Delay Spread in Indoor Optical Wireless Systems
K. SMITHA, A. SIVABALAN, J. JOHN..... 754

A MAC Scheme with QoS Guarantee for MANETs
Y. B. YANG, Y. L. WEI..... 759

A Reputation-Based Multi-Agent Model for Network Resource Selection
J. F. TIAN, J. LI, L. D. YANG..... 764

Subcarrier Availability in Downlink OFDM Systems with Imperfect Carrier Synchronization in Deep Fading Noisy Doppler Channels
L. NOOR, A. ANPALAGAN, S. KANDEEPAN..... 775

Performance Analysis of MAC Protocol for LEO Satellite Networks
M. X. GUAN, R. C. WANG..... 786

Ant Colony Optimization Based on Adaptive Volatility Rate of Pheromone Trail
Z. Q. CAI, H. HUANG, Y. QIN, X. H. MA..... 792

A Review of Wireless Body Area Networks for Medical Applications
S. ULLAH, P. KHAN, N. ULLAH, S. SALEEM, H. HIGGINS, K. S. KWAK..... 797

ISSN: 1913-3715
9771913371005 12