Medium access control and network planning in wireless networks

by

Zakhia (Zak) Abichar

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Computer Engineering

Program of Study Committee:
J. Morris Chang, Co-major Professor
Ahmed E. Kamal, Co-major Professor
Daji Qiao
Yong Guan
Lu Ruan

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKOWLEDGEMENT

I would like to thank my co-advisors Dr. Morris Chang and Dr. Ahmed Kamal for the opportunity of studying for the Ph.D. degree with them. I enjoyed learning about the field of wireless networks with them and about how to do research. I hope that I have been a good student of them and provided them with the respect that they deserve.

I also appreciate the help of my committee members Dr. Daji Qiao, Dr. Lu Ruan, and Dr. Yong Guan. I would like to thank Dr. Arun Somani who was earlier on my committee during my prelim exam before taking a sabbatical leave for this year.

During the years of my Ph.D., I received a lot of help from the graduate secretary in the Department of Electrical and Computer Engineering, Pam Myers. Also, the people at the Graduate College were very helpful in providing me with explanations about the rules and policies for the Ph.D. program.

I feel very fortunate to have met several people in my research groups who were good friends and showed me new ways to think about things. Several other people that I met at Iowa State University were also good companions.

Finally, I would like to thank my family for providing me with the support to embark on the graduate program and to the continuation of this dissertation.

ABSTRACT

Wireless Local Area Networks (WLANs) and Wireless Metropolitan Area Networks (WMANs) are two of the main technologies in wireless data networks. WLANs have a short range and aim at providing connectivity to end users. On the other hand, WMANs have a long range and aim at serving as a backbone network and also at serving end users. In this dissertation, we consider the problem of Medium Access Control (MAC) in WLANs and the placement of Relay Stations (RSs) in WMANs. We propose a MAC scheme for WLANs in which stations contend by using jams on the channel. We present analytic and simulation results to find the optimal parameters of the scheme and measure its performance. Our scheme has a low collision rate and delay and a high throughput and fairness performance. Secondly, we present a MAC scheme for the latest generation of WLANs which have very high data rates. In this scheme, we divide the stations into groups and only one station from each group contends to the channel. We also use frame aggregation to reduce the overhead. We present analytic and simulation results which show that our scheme provides a small collision rate and, hence, achieves a high throughput. The results also show that our scheme provides a delay performance that is suitable for real-time applications and also has a high level of fairness. Finally, we consider the problem of placing Relay Stations (RSs) in WMANs. We consider the Worldwide Interoperability for Microwave Access (WIMAX) technology. The RSs are used to increase the capacity of the network and to extend its range. We present an optimization formulation that places RSs in the WiMAX network to serve a number of customers with a pre-defined bit rate. Our solution also provides fault-tolerance by allowing one RS to fail at a given time so that the performance to the users remains at a predictable level. The goal of our solution is to meet the demands of the users, provide fault-tolerance and minimize the number of RSs used.

CHAPTER 1. Introduction

Wireless data networks have become extensively used and prevalent in the last decade. Two of the main technologies in wireless data networks are Wireless Local Area Networks (WLAN) and Wireless Metropolitan Area Networks (WMAN). While both type of these technologies provide wireless connectivity, they are different in their scopes and goals. WLANs have a small range and are used as a last hop in wireless to the end users. They are currently prevalent and are used in public and private venues. On the other hand, WMANs have a long range and aim at serving as a backbone network and also at serving end users.

The WLANs are specified by the IEEE 802.11 series of standards. The interfaces specified by IEEE 802.11b and IEEE 802.11g standards were the most used standards worldwide. Recently, a new standard called IEEE 802.11n has been ratified. This standard provides very high data rates in the WLAN by using the latest advances in the physical layer.

The IEEE 802.16 series of standards provides specifications for WMANs. In the last few years, there has been renewed interest in WMANs through a technology called Worldwide Interoperability for Microwave WiMAX (WiMAX). WiMAX has a strong industry support. This technology aims at using WMANs to serve fixed and mobile users. A fixed user can be a single user or a station that aggregates the traffic of multiple users. WiMAX also aims at bringing service to mobile users by supporting handset devices and mobile computers.

In this dissertation, we consider two types of problems. The first is the Medium Access Control (MAC) in WLANs. The MAC scheme arbitrates the channel access mechanism to multiple users. The goal of the MAC scheme is to allow the stations to share a channel in an efficient manner. The MAC scheme should provide a low collision rate, a high throughput, a low delay average and variance and a high fairness between the users.

The second problem that we consider is the placement of Relay Stations (RSs) in WiMAX networks. RSs are used to increase the capacity of the network and to extend its range. The goal of using RSs is to support the Base Station (BS) which is the main element in the infrastructure of the WiMAX networks. The BS provides access to the users. The BS is also connected to the main network, which is the Internet. The RSs' function is to be the middle man between the BS and the end users, called the Subscriber Stations (SSs) or Mobile Stations (MSs). Our solution to RS placement aims at finding the number of RSs that should be used and their locations. A placement solution should satisfy the users' demands while avoiding unnecessary cost for the deployment of the network.

In the rest of this chapter, we summarize our contribution in the MAC scheme for WLANs, the MAC scheme for the latest generation of WLANs that have high data rates and in the placement of RSs in WiMAX networks.

1.1 Medium Access Control in WLANs

In today's wireless networks, stations using the IEEE 802.11 Standard contend for the channel using the Distributed Coordination Function (DCF). Research has shown that DCF's performance degrades especially with the large number of stations. This becomes more concerning due to the increasing proliferation of wireless devices. In Chapter 2, we present a Medium Access Control (MAC) scheme for wireless LANs and compare its performance to DCF and to other efficient schemes. Our scheme, which attempts to resolve the contention in a constant number of slots (or constant time), is called CONTI. The contention resolution happens over a pre-defined number of slots. In a slot, the stations probabilistically send a jam signal on the channel. The stations listening retire if they hear a jam signal. The others continue to the next slot. Over several slots, we aim to have one station remaining in the contention, which will then transmit its data. We find the optimal parameters of CONTI and present an analysis on its performance. More comprehensive evaluation is presented in the simulation results where we compare CONTI, DCF and other efficient schemes from the literature. We consider the number of slots used, the collision rate, the throughput, the delay

and fairness. The highest throughput was achieved by CONTI. Moreover, our results provide measurements from each of the schemes that we consider and provide the insight on each scheme's operation.

1.2 Medium Access Control in WLANs with High Data Rates

The latest generation of Wireless Local Area Networks (WLANs) was initiated by the IEEE 802.11n Standard. The standard uses the latest advances in the physical layer that provide high data rates. It also aims at improving the efficiency of the Medium Access Control (MAC) scheme to provide a throughput that is higher than 100 Mbps. In Chapter 3, we present a Group-based MAC (GMAC) scheme that aims at reducing the contention and the overhead. The stations are divided into groups and only one station from each group does the contention, thus reducing the probability of a collision. A station that gains access to the channel transmits a polling frame to allow the other stations in its group to transmit without requiring RTS/CTS frames, thus reducing the overhead. GMAC is interoperable with legacy 802.11 devices. To evaluate GMAC, we present analytic and simulation results. We also consider other schemes from the literature and compare them to GMAC. The results show that GMAC achieves a high throughput, high fairness, low delay and high scalability with respect to the data rates.

1.3 Planning of Relay Station (RS) Locations in WiMAX Networks

The IEEE 802.16j standard specifies the use of Relay Stations (RS) in WiMAX networks. The standard does not specify how to plan the locations of the RSs within the network. There have been several papers that aim at planning the locations of RSs in the WiMAX architecture. However, placement of RSs in WiMAX networks such that an RS failure won't interrupt the service, hence making the network fault-tolerant, is an important design and planning problem that has not been addressed in the literature. In Chapter 4, we address this problem, and present an Integer Linear Program (ILP) formulation that provides the planning of RS locations with fault-tolerance. We allow one RS to fail while keeping the provided service at a designated level (defined in throughput to users). We present numerical results that show

how our model can be used to plan the positions of RSs. We also incorporate the existence of obstacles in the planning field, such as large structures or natural formations, that might happen in real life. To the best of our knowledge, this is the first work that addresses planning the RS locations in WiMAX in a fault-tolerant manner.

CHAPTER 2. A Medium Access Control Scheme for Wireless LANs with Constant-Time Contention

An article that will appear in *IEEE Transactions on Mobile Computing.*[1]

Zakhia Abichar[2] and J. Morris Chang[3]

2.1 Introduction

The use of wireless networks in everyday computing has been a success story and new wireless technologies continue to emerge. Nowadays, wireless networks are a necessary part of the computing world. This was made possible by the IEEE 802.11 Standard (1) which provides technical specifications for the wireless interfaces. In addition, the Wi-Fi Alliance was formed to certify interoperability of wireless products from various vendors. The Medium Access Control (MAC) scheme in the standard that is most widely used is the Distributed Coordination Function (DCF). Its function is arbitrate the use of the medium to multiple stations that are connected to one Access Point (AP) in the infrastructure mode. In addition, DCF can be used in the infrastructure-less, or ad hoc, mode in which there is no AP. This paper presents a study on the MAC schemes. We propose a new scheme and we make a comparison between our scheme, DCF and several other efficient schemes.

The contention with DCF works as the following. The stations use Contention Windows (CW) to randomize their access and try to avoid collisions. Initially, a station waits for DIFS (DCF Inter Frame Space) and transmits if the channel is idle. However, if the channel is busy,

[2]Primary researcher and author

[3]Faculty advisor

the CW is used. The CW is initially assigned to a preset value, CW_{min}, which depends on the physical layer. Then, a station sets a backoff (BO) counter to a random value chosen from a uniform distribution from $[0, CW]$. The station decreases the BO counter by one for every time slot the channel is idle. If a busy channel is detected, the BO counter is freezed and the countdown resumes from the freeze value after the channel is idle for a duration of DIFS. The station transmits when its BO counter reaches zero. If two or more stations reach zero at the same time, there will be a collision and the transmitted frames won't be received correctly. The colliding stations will not receive an ACK frame and they will double their CW (until it reaches the maximum value equal to CW_{max}). On the other hand, when a station transmits a data frame successfully, its CW is reset to the initial value CW_{min}.

There are numerous studies on DCF in the literature. Some studies presented performance models of DCF to characterize its performance (2),(3),(4). One observation from the results indicates that DCF's performance degrades significantly with an increase in the number of stations. While this wasn't an issue at the inception of DCF, now more and more people use wireless connections and this becomes a limitation practically. The decrease of performance in this case is attributed to the large number of collisions with the increase in number of stations. Other evaluations (5) of DCF show that its delay might be very large with busy traffic conditions. Finally, the fairness of DCF has been considered (6) and it was shown that DCF doesn't have a high fairness in the short-term, although its fairness increases as the stations contend for longer periods.

In this paper, we present a MAC scheme that provides access by resolving the contention between stations. The main feature of our scheme is that it attempts to resolve the contention in the same number of slots every time. Our scheme, which attempts to resolve the contention in a CONstant TIme, is called CONTI. The contention resolution has several slots. At the first slot, all the stations with frames to transmit contend. The stations, with a probability that we define, choose an event of sending a jam on the channel for the slot duration. The jam is simply a burst of energy (similar to the HIPERLAN scheme (7) and Blackburst scheme in (8)) and doesn't need to contain any specific information. With the complementary probability,

the stations choose an event of listening to the medium. During a slot, stations retire from the contention if they were listening and hear a jam, which we call preemption. The remaining stations move on to the next slot and repeat the contention. We aim to have one remaining station at the end of contention to provide it with access to the medium.

We present two methods for finding the parameters. The parameters are the number of slots in a contention and the probability corresponding to every slot. The first method aims at bounding the collision rate. The smallest number of slots and the corresponding probability vector are found so that the collision rate is bounded by a given value. It uses an algorithmic approach and finds the parameters with a low computation time. The second method uses an optimization approach to find the parameters that maximize the throughput. While this method requires a larger computation time, we only need to find the parameters offline and make them known to the stations. Thus, during the data transmission, this method doesn't need to operate; instead the parameters that it has produced are readily used.

Additionally, we present an analysis that characterizes the throughput of CONTI and DCF and compare them to each other. We also show analytically the effect of the slot duration on the performance achieved by CONTI.

In the simulation results, we compare CONTI, DCF and we program the simulation code of other proposed MAC schemes in the literature. We implement all of the schemes in the same simulation environment to ensure the fairness of comparison. We present results on the number of slots used in the contention resolution, the collision rate and the throughput. We show that CONTI achieves the highest throughput among the schemes and we provide the insight on the behavior of the various schemes. We also show the results of delay and the fairness of the schemes and explain the trends that were observed.

The rest of the paper is organized as follows. Section 2.2 presents the related work that we compare against and other schemes in the MAC area. Section 2.3 presents the specifications of our scheme and Section 2.4 gives an analysis to obtain the optimal parameters of CONTI by bounding the collision rate. Section 2.5 presents another type of analysis on the parameters in order to maximize the throughput. Section 2.6 characterizes the performance gain of CONTI

and analyzes the effect of the slot duration on the performance. Section 2.7 presents the simulation results comparison of multiple MAC schemes and, finally, Section 2.8 presents the conclusion of the paper.

2.2 Related Work

There have been numerous MAC schemes proposed in the literature. We reference in this section some of the well-known schemes and we compare against them in the simulation results. We highlight here two types of schemes: (1) the contention window (CW) based schemes and (2) the jamming based schemes.

2.2.1 PREMA

The scheme *Prioritized Repeated Eliminations Multiple Access* (PREMA) was proposed in (9). PREMA is a jamming-based scheme. It works as the following. Contending stations transmit a jam, whose length in slots is drawn from a geometric distribution with parameter q. After the last jam slot, the stations do one slot of carrier sense. If they hear another ongoing slot, they're out of this contention. If not, it means they passed this elimination. The stations with the longest burst will survive the elimination. Following, they do another elimination by choosing another random number from the same distribution and jamming and then one slot of carrier sense. The number of eliminations is a parameter called h. The authors of PREMA use the parameters $h = 4$ and $q = 0.5$. We use these parameters when we compare our work with PREMA.

2.2.2 k-EC

The scheme *k-Round Elimination Contention* (k-EC) was proposed in (10). k-EC, like PREMA, is a jamming-based scheme. It also has several rounds of eliminations in a contention. There are k rounds of elimination, where k is a parameter. A round of k-EC consists of at most m slots. The contending stations choose a random number uniformly from $[0, m-1]$ and transmit only one jam in the slot number. If a station chooses 0, then it's the first slot, etc. If

the station is not jamming, then it should be listening by carrier sense. When a station hears a jam while it's listening, it drops out of the contention and the round is finished for it. The other stations survive and move to the next round. Since the jam can happen in any slot, a round of k-EC is at most m slots long. The authors of k-EC use $k = 7$ and $m = 3$ as the best parameters. As a result, a contention of k-EC with these parameters takes anywhere from 7 slots to 21 slots.

In the proposal of k-EC in (10), the authors compare k-EC to our initial work CONTI that was published in (11). However, we have the following comments on their comparison.

- In k-EC (10), the authors indicate that they take the results of CONTI from our paper (11), rather than programming CONTI in the simulator. This comparison will reduce the accuracy of the result, since the computation environments of the two simulations are different and the physical layer characteristics would not be the same. However, in this paper, we implement both CONTI and k-EC in the same environment and use the same PHY characteristics.

- Secondly, both k-EC and CONTI use the jams in the same way. They need the same requirements in transmit length, since they serve a similar function in both schemes. In k-EC, they use the jam slot time of 10 μs. However, in CONTI, we used the slot time of 20 μs as in the standard. So, to have a fair comparison of the MAC schemes, we need to have the same parameters on the slot time, which we do in our simulations.

2.2.3 Idle Sense

The scheme *Idle Sense* was originally proposed in (12) and then it was revised in (13). In our work, we consider the revised Idle Sense scheme as in (13). Unlike PREMA and k-EC, Idle Sense is based on the contention window (CW) mechanism, like the standard's DCF scheme. The main idea of Idle Sense is observing that there is an optimal number of slots between two consecutive transmissions. This number if deemed to be $n_i^{target} = 3.91$ for the 802.11b physical layer. It is suitable for a large number of scenarios with varying number of contending stations. Hence, in Idle Sense, all the stations observe the number of slots and adjust the CW up or

down to match the number of observed idle slots to the target value. The CW is adjusted up by $CW \leftarrow CW + \epsilon$, and down by $CW \leftarrow \alpha.CW$. The optimal parameters presented are $1/\alpha = 1.0666$ and $\epsilon = 6.0$.

The above procedure of adjusting the CW is performed periodically after the elapse of a number of transmissions given by $maxtrans$. Initially, $maxtrans = 5$, but when the above procedure is done, $maxtrans$ might be changed if the number of idle slots is far off from n_i^{target} so as to speed up the convergence. Specifically, $if(|n_i^{target} - n_i| \geq \beta)$, then $maxtrans \leftarrow 5$; this value is considered small. Otherwise, $maxtrans \leftarrow CW/\gamma$; this value will keep $maxtrans$ proportional to the number of active stations and will be roughly equal to $3n$ (n is the number of active stations) in IEEE 802.11b as explained in (13). The optimal parameters are given as $\beta = 0.75$ and $\gamma = 4$. We use the same parameters when we compare our work to Idle Sense in the simulation.

2.2.4 Other Approaches

Other proposed approaches based on jams are in (14), (15), (16), (17), (18) and (7). Other proposed approaches to optimize the CW schemes are in (19), (20), (21), (22), (23) and (24). In this paper, we compare our scheme to the standard's DCF, PREMA, k-EC and Idle Sense.

2.3 Medium Access Control (MAC) with Constant-Time Contention

This section presents the proposed MAC scheme, CONTI, with attempts to resolve the contention using a constant number of slots. We start by defining the terms that are used in our work.

2.3.1 Notations

- The number of stations in the WLAN cell is given by n.
- The contention is resolved over a number of slots given by k, and the contention slots are labeled $\{s_1, \ldots, s_k\}$.

- During a contention slot, a station either transmits a pulse, called *signal 1* or listens to the channel, called *signal 0*.
- The probability vector used by the stations to decide whether to transmit a pulse or listen is given by $p : \{p_1, \ldots, p_k\}$. A station will choose signal 1 during slot s_i with probability p_i. Otherwise, a signal 0 is chosen with a probability $1 - p_i$.
- The number of remaining stations in the contention at the end of the slots is designated by the vector $r : \{r_0, \ldots, r_k\}$. So, $r_0 = n$ stations start the contention, and r_i stations remain in the contention at the end of slot s_i.
- An instance of CONTI is characterized by its parameters, the number of slots k and the probability vector p. Thus, an instance of the scheme, S, is designated by $S(k, p)$.

2.3.2 Contention

The contention of n stations is resolved using CONTI over k contention slots. Each of the stations uses the same probability vector p. All of the stations go through the following procedure. Before a contention slot s_i, a station chooses signal 1 with probability p_i or signal 0 with probability $1 - p_i$.

During a contention slot, the station will transmit a pulse on the channel if it has signal 1. Otherwise, the station will listen to the channel. The pulse that is transmitted doesn't need to contain information. Rather, its presence on the channel indicates to other stations that some stations have chosen a signal 1. A station that is listening and hears the presence of a signal on the channel is said to be preempted and this station doesn't contend anymore in this contention. But if a station with signal 0 doesn't hear a signal, it stays in the contention. If the station has signal 1, it transmits the pulse and moves to the next contention slot. At the end of the last slot, a station transmits its data frame if it has not been preempted.

During a contention slot, it is better to eliminate the largest number of stations possible. This means the contention resolution is occurring quickly and the amount of time spent on contention resolution is minimized. Before slot s_i, there are r_{i-1} stations. At the end of slot

s_i, there are r_i stations that are remaining in the contention. Thus, slot s_i has eliminated $(r_{i-1} - r_i)$ stations from the contention, which we seek to maximize.

With CONTI, it is possible that no stations are eliminated during a contention slot. This happens if all the stations choose signal 1. Then, no station is preempted. It also happens if all the stations choose a signal 0. If this event happens, then the following slots will continue the contention. But if it happens in the last slot and there are more than one station remaining, there will be a collision.

For an efficient contention resolution, the probability choices should be optimized to minimize the collision rate. The number of slots should be also minimized so that the time spent in the contention is reduced. For that purpose, the values of k and the vector p are optimized in the next section.

Finally, we add a stipulation that ensure compatibility with the Inter-Frame Spacing used in wireless networks, such as DIFS in the standard. In CONTI, there might be a few consecutive slots where all of the stations choose signal 0. Thus, a station that had already retired from contention should not count this silent time in its IFS timer. Thus, we require a station that has retired to stop its IFS timer until the contention is finished. Since the station knows the number of slots, k, a priori, it can do that.

2.3.3 Example

An example on the contention resolution using CONTI is presented in Figure 2.1. There are 6 stations. In the first slot, stations 2, 4 and 5 choose signal 1 and preempt stations 1, 3 and 6. Thus, stations 1, 3 and 6 don't contend anymore in this round. The graph on the left side of the figure shows the signals, while the graph on the right side depicts the jams. In the second slot, stations 2, 4 and 5 choose signal 0 and no station is preempted. All the stations move to the third slot. In the third slot, stations 2 and 4 preempt station 5. Finally, in the last slot, station 2 preempts station 4. Then, station 2 is able to transmit a data frame. In this example, $n = 6$, $k = 4$ and the vector r is $\{6, 3, 3, 2, 1\}$.

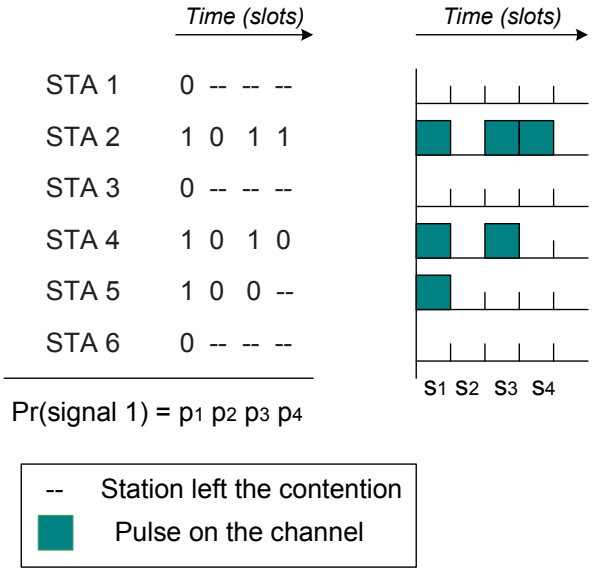

Figure 2.1 Contention resolution using CONTI.

2.3.4 Pseudocode

The contention resolution using CONTI is specified in the pseudocode in Algorithm 1. The pseudocode describes the operation of a CONTI instance $S(k, p)$. In Algorithm 1, the state variable $retire$ indicates if the station has been preempted, when $retire = 1$, or if the station is still in the contention, when $retire = 0$.

The function $defer(t)$ makes the station remain idle for a duration of t. The function $pulse(t)$ makes the station transmit a pulse for a duration of t. Finally, the function $pulseDetect(t)$ makes the stations detect if there has been a pulse on the channel during a time slot.

2.3.5 Why do we Need to Optimize the Parameters?

In this part, we present an example that shows how the parameters affect the contention resolution in CONTI. In other approaches in the literature that use similar probabilistic jams (14), the probabilities are chosen randomly. This example shows that we can have performance gain by optimizing the parameters. In this example, we compare two cases. First, we consider intuitively chosen values of $p_i = 0.5$ for all the slots. Then, we show another case with more

Algorithm 1 Contention Resolution with CONTI

```
retire := 0
i := 1
while (i ≤ k) do
    if retire = 1                /* Station has been preempted */
        defer(t_slot)
    else if retire = 0           /* Station in contention */
        proba := Uniform(0,1) /* Choose signal 1 or 0 */
        if proba < p_i
            signal := 1
        else signal := 0
        if signal = 1            /* Station with signal 1 */
            pulse(t_slot)
        else if signal = 0       /* Station with signal 0 */
            listen(t_slot)
            if pulseDetected(t_slot) = true
                retire := 1
    i := i + 1
```

Table 2.1 Percentage Collision Rate ($p_i = 0.5, (1 \leq i \leq k)$)

k	2	3	4	5	6	7	8
$n = 10$	80.73	50.95	28.33	14.89	7.62	3.86	1.94
$n = 25$	99.37	87.00	59.21	34.21	18.31	9.46	4.80

appropriate parameter choices.

In the case of $p_i = 0.5$ for $1 \leq i \leq k$, we show how the collision rate is affected by the number of slots used. The result is shown in Table 2.1 for the number of stations 10 and 25. With 5 contention slots, the collision rate[4] of 10 stations is 14.89% and with 25 stations it is 34.21%. Also, as one might think intuitively, the result shows that an increase in the number of slots reduces the collision rate.

Now we consider the use of CONTI with 5 slots and with the following probability vector[5] $p : \{0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$. For this vector, the collision rate for 10 stations is 7.59% compared to 14.89% with the intuitive choice of $p_i = 0.5$. This is a reduction by almost a factor of 2. For 25 stations, the collision rate is 13.65% compared to 34.21% using the intuitive

[4]The table shows analytic result of the collision rate. The expression of the collision rate in presented in the next section

[5]We show how to derive the numbers in this vector in the next section.

choice. This shows that optimized parameters have a significant effect on the performance of contention resolution with CONTI.

2.4 Parameters: Bounded Collision Rate

This section presents an analysis to find the optimal parameters of CONTI based on a bounded collision rate. The analysis also shows that for a wide range of n, we can use the same parameters, while remaining close to the optimal performance. Hence, the stations don't need to know or approximate n.

This section has the following contents. First, the definition of optimal solution is presented and then we show the algorithmic concepts that we use to get the parameters. Then, the collision rate is found analytically, which will be used to find the parameters. Then, k and p are found for a given number of stations, n. Finally, we show that there are certain k and p that can be used efficiently, independently of n for a range of realistic cases.

2.4.1 Defining The Optimal Solution

The optimal solution with CONTI can be defined in more than one way. If we seek to minimize the collision rate, say to make it equal to zero, we can do that but we risk making the number of slots too large. On the other hand, if we use a small number of slots to avoid wasting time in the contention, we might end up with a large collision rate that adversely affects the throughput. Thus, we use a definition that is a trade-off between these two factors. Our definition of optimal solution is based on the idea that we can tolerate a small collision rate. Then, we need to find the minimum number of slots and the corresponding probabilities such that the collision rate is bounded.

Definition 1 Optimal solution – *The minimum value of k and a probability vector p should be found such that the collision rate p_{coll} is bounded by a value given in p^*_{coll}.*

2.4.2 Solution Approach

In a problem satisfying the *optimal substructure property* (25) and the *greedy-choice property,* a greedy solution yields an optimal solution. In the greedy solution, there are several choices of parameters (different p-values at different contention slots) and the best choices are found at each step. For CONTI, that means the value of p that minimizes the number of stations remaining in contention are found in each contention slot. It remains to show that CONTI satisfies the two properties above.

A problem is said to have the *optimal substructure property* if an optimal solution to the problem contains within it optimal solutions to subproblems. This is proved for the case of CONTI by contradiction. Consider an optimal solution $S_\alpha(k_\alpha, p_\alpha)$ that satisfies the upper-bound condition on the collision rate p^*_{coll}. If at any slot s_i, the choice of p_i in S_α doesn't minimize the number of remaining stations (r_i), then p_i can be replaced with another optimal value p_i^{opt} that minimizes r_i. Then, k_α might be reduced or a smaller value for p^α_{coll} is obtained, which implies that S_α is not an optimal solution.

By definition, a problem has the *greedy-choice property* if a greedy choice that is already made combined with an optimal solution to the remaining subproblem yields an optimal solution. To show this property for CONTI, consider a case of contention resolution where a greedy choice is made at the first slot s_1. The greedy choice minimizes the number of remaining stations (r_1) at the end of the slot. An optimal solution for the subproblem with r_1 stations and $k-1$ slots, combined with the greedy solution at the first slot yields an optimal solution because it contains the minimum k value.

2.4.3 The Collision Rate

Let the term $\sigma(n, k, p)$ be the probability that the instance of CONTI, $S(k, p)$, resolves the contention successfully for n stations.

Next, the probability of preempting stations over one slot is defined. Consider a contention slot s_i where $r_{i-1} = u$ stations start the contention and $r_i = v$ stations remain at the end of the slot. Let the probability of this event be designated by $\tau_{u,v}(p_i)$. Its expression is the

following.

$$\tau_{u,v}(p_i) = \begin{cases} \binom{u}{v}.(p_i)^v.(1-p_i)^{u-v} & 1 \leq v \leq u-1 \\ (p_i)^u + (1-p_i)^u & v = u \end{cases} \tag{2.1}$$

In the first case, v out of u stations remain at the end of the slot. In the second case, all of the u stations remain at the end of the slot. This happens if all the stations choose the same signal, whether it is signal 1 or signal 0.

Before we can give the analytic expression of the collision rate, we give the following definition of the subvector of p.

Definition 2 *For a vector $p : \{p_1, p_2, \cdots, p_k\}$ with k elements, the term π_i $(1 \leq i \leq k)$ defines the subvector of p with $k - i + 1$ elements given by $\pi_i : \{p_i, p_{i+1}, \cdots, p_k\}$, so it is a suffix subvector.*

Finally, the collision rate of CONTI is given by the following theorem.

Theorem 1 *The probability that scheme $S(k, p)$ resolves the contention successfully for n stations is given by:*

$$\sigma(n, k, p) = \Sigma_{i=0}^{n-1} \Big[\tau_{n,n-i}(p_1).\sigma(n-i, k-1, \pi_2) \Big] \tag{2.2}$$

The trivial cases for the expression of σ are the following.

$$\sigma(1, i, \pi_{(k-i+1)}) = 1, \qquad 1 \leq i \leq k \tag{2.3}$$

$$\sigma(v, 1, \pi_k) = \tau_{v,1}(p_k), \qquad v \geq 1 \tag{2.4}$$

Proof 1 *First, notice that π_1 designates the same vector as p. Then, $\sigma(n, k, p)$ can be rewritten as $\sigma(n, k, \pi_1)$. After the elapse of one slot, the number of stations is reduced from $r_0 = n$ to r_1, where $r_0 \geq r_1$. The remaining problem is the contention resolution of r_1 stations in $k - 1$ slots using the vector π_2. This subproblem is solved successfully with a probability given by $\sigma(r_1, k-1, \pi_2)$.*

During the first slot, the number of stations that are preempted is between 0 *and* $n-1$*. Each of these events occur with a probability of* $\tau_{n,n}(p_1), \ldots, \tau_{n,1}(p_1)$*, respectively. Thus,* $\sigma(n,k,p)$ *is equal to the following expression:*

$$\sigma(n,k,\pi_1) = \tau_{n,n}(p_1).\sigma(n,k-1,\pi_2) + \tau_{n,n-1}(p_1).\sigma(n-1,k-1,\pi_2)+ \cdots + \tau_{n,1}(p_1).\sigma(1,k-1,\pi_2) \quad (2.5)$$

The trivial cases for the expression of σ *are given in Equations 2.3 and 2.4. Equation 2.3 represents the case where there is one station left and there are one or more slots. In this case, the contention is resolved successfully with probability 1. The case in Equation 2.4 represents the scenario when there is one slot left and there are one or more stations contending. In this case, the contention is resolved correctly if all the stations are preempted except one. This happens with a probability given by* $\tau_{v,1}(p_k)$*.*

2.4.4 Finding the Probability Choices

We need to find the values in the vector p that resolve the contention the quickest. Over a slot s_i, the number of stations should be reduced from r_{i-1} to r_i, with having r_i minimized.

One observation about the probability values p_i is the following. In the earlier slots, the value of p_i should be small. This will allow few stations to choose signal 1 and hopefully preempt most of the other stations. Then, the contention resolution proceeds fast. At the later slots, there should be few remaining stations and then p_i should have a larger value so that at least one station will choose a signal 1 and preempt the other stations. The values of p_i drawn from the analysis agree with this observation in that the values in p are non-decreasing. This trend of increasing probability values as the slots progress is also observed in the analysis of the Sift MAC scheme (24). However, Sift differs from CONTI in that it is based on Contention Windows, not jamming slots, and it has the goal of supporting event-driven sensor networks, not wireless LANs.

Consider a slot s_i with u contending stations. The probability that v stations remain at the end of the slot is given by $\tau_{u,v}(p_i)$ in Equation (2.1). The expected value of v is given by

Table 2.2 Contention over One Slot (CDF on v)

u	p_i^*	$E[v]$	v^*	$\Pr(v \leq v^*) = CDF(v^*)$
2	0.5	1	1	0.5
3	0.4314	1	2	0.735
4	0.4245	2	3	0.857
5	0.36715	2	4	0.891
10	0.2563	3	5	0.925
20	0.14955	3	9	0.960
30	0.11945	4	10	0.977
40	0.09425	4	10	0.980
50	0.0783	4	10	0.981
60	0.0673	4	10	0.982
70	0.064	5	10	0.985
80	0.05705	5	10	0.985
90	0.0516	5	10	0.984
100	0.04715	5	10	0.984

the following equation:

$$E[v] = \left[\sum_{v=1}^{u} v.\binom{u}{v}.(p_i)^v.(1-p_i)^{u-v}\right] + u.(1-p_i)^u \tag{2.6}$$

For a given value of u, Equation (2.6) is solved numerically to determine p_i that minimizes v. This value of p_i, denoted by p_i^* reduces the number of stations the quickest over one slot. The corresponding values of u, p_i and $E[v]$ are presented in the three leftmost columns of Table 2.2. An entry in the table is read as follows: with $u = 100$ stations at the beginning of a slot, using $p_i = 0.04715$ provides an expected number of remaining stations equal to $v = 5$.

2.4.5 CDF on the Number of Remaining Stations

The information in the three leftmost columns of Table 2.2 shows the value of p_i that minimizes the expected value of v. However, this information doesn't show the probability that this event ($E[v]$ is minimized) will happen. For the entry with $u = 100$, the event $v = 5$ will occur with a probability of $\tau_{100,5}(0.04715) = 0.178$. Thus, we will have to use the Cumulative Distribution Function (CDF) to have a higher confidence on how the number of stations is reduced over one contention slot.

Let the CDF on v be defined by the function $\theta_{u,v}(p_i)$. This function designates the probability that, starting with u stations, the number of stations at the end of the slot will be less or equal to v. The value of $\theta_{u,v}(p_i)$ is the following:

$$CDF(v) = \theta_{u,v}(p_i) \ = \ \sum_{j=1}^{v} \tau_{u,j}(p_i) \tag{2.7}$$

Using the CDF, a bound (in a probabilistic sense[6]) on the number of stations that survive a contention slot is given in Table 2.2. At the start of the slot, there are u stations. The column labeled v^* designates the upperbound, in probabilistic terms, on the number of stations that will survive the contention. The last column, labeled $CDF(v^*)$, indicates the probability of this event happening, that is, $\Pr(v \leq v^*)$. As an example, the last entry is interpreted as: starting with $u = 100$ stations, there will remain $v \leq 10$ stations at the end of the slot with a probability of 0.984.

The information in Table 2.2 is used to derive the number of slots k and the probability vector p that are needed to resolve the contention of n stations. The collision rate will be measured by the analytic expression given in Equation (2.2).

2.4.6 Finding the Parameters Given the Number of Stations

Now the number of slots and the probability vector can be found for a given number of stations using Table 2.2. In the definition of the optimal solution, it was required to minimize the number of slots k and find a corresponding vector p so that the collision rate is smaller than an upperbound, $p_{coll} < p^*_{coll}$. In this paper, we use $p^*_{coll} = 6\%$, which is considered the upperbound on the collision rate that is tolerable. The value of $p^*_{coll} = 6\%$ was chosen based on simulation results.

The parameters for $n = 100$ station are found as follows. In the first slot, $p_1 = 0.04715$ is used, which yields a number of remaining stations $r_1 \leq 10$, with a high probability (equal to 0.984 from Table 2.2). In the second slot, we assume that there are 10 stations and then $p_2 = 0.2563$. This leads to $r_2 \leq 5$ with a high probability. Similarly, the number of remaining

[6]In the rest of this section, the meaning of bound *in a probabilistic sense* apply to v^*.

Table 2.3 The Parameters and the Collision Rate

n	k	p	p_{coll}
2	5	$\{0.5, 0.5, 0.5, 0.5, 0.5\}$	3.12%
3	5	$\{0.4314, 0.4314, 0.5, 0.5, 0.5\}$	4.32%
4	5	$\{0.4245, 0.4245, 0.4314, 0.5, 0.5\}$	5.33%
5	5	$\{0.36715, 0.4245, 0.4314, 0.5, 0.5\}$	5.88%
10	6	$\{0.2563, 0.36715, 0.4245, 0.4314, 0.5, 0.5\}$	3.85%
20	6	$\{0.14955, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.52%
30	6	$\{0.11945, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.53%
40	6	$\{0.09425, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.40%
50	6	$\{0.0783, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.28%
60	6	$\{0.0673, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.21%
70	6	$\{0.064, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.20%
80	6	$\{0.05705, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.26%
90	6	$\{0.0516, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.35%
100	6	$\{0.04715, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$	5.48%

stations progresses as $r_3 = 4$, $r_4 = 3$, $r_5 = 2$ and $r_6 = 1$. The remaining probability choices are $p_3 = 0.36715$, $p_4 = 0.4245$, $p_5 = 0.4314$ and $p_6 = 0.5$. In total, it takes 6 slots to resolve the contention when starting with 100 stations. The collision rate measured by the analytic expression is $p_{coll} = 5.48\%$.

The parameters for other values of n are presented in Table 2.3. There are cases when the CDF values from Table 2.2 don't correspond to a high certainty. For example, starting with 2 stations, the best possible event is to have 1 stations remaining. This happens with a probability of 0.5. To satisfy the upperbound on the collision rate of 6%, the same value of $p_i = 0.5$ is used over several slots. For the case of $n = 2$, there are 5 slots each using $p_i = 0.5$. The collision rate is $p_{coll} = 3.12\%$.

2.4.7 Constant-Time Contention Resolution

In Table 2.3, the parameters for CONTI were optimized for a given number of stations. This was done independently for each value of n. However, there are several cases where $20 < n < 100$ that obtained the number of slots and the same probability values. For the cases where $n < 20$, the number of slots is 5, which is close to the other cases.

Next, the same parameters are used for all the cases of n. The parameters chosen are

Table 2.4 Collision Rate Comparison

n	k	p_{coll}	k°	p°_{coll}
2	5	3.12%	6	3.92%
3	5	4.32%	6	4.00%
4	5	5.33%	6	4.19%
5	5	5.88%	6	4.37%
10	6	3.85%	6	5.02%
20	6	5.52%	6	5.52%
30	6	5.53%	6	5.53%
40	6	5.40%	6	5.40%
50	6	5.28%	6	5.28%
60	6	5.21%	6	5.21%
70	6	5.20%	6	5.20%
80	6	5.26%	6	5.26%
90	6	5.35%	6	5.35%
100	6	5.48%	6	5.48%

the ones used for $n = 100$. Accordingly, the scheme $S^\circ(k^\circ, p^\circ)$ is defined where $k^\circ = 6$ and $p^\circ = \{0.04715, 0.2563, 0.36715, 0.4245, 0.4314, 0.5\}$. The collision rate of S° is measured and compared to the case with the previous parameters that are found in Table 2.3. The comparison results are presented in Table 2.4. The collision rate varies slightly and it remains below 6% for all cases. For some cases, the collision rate is lower for S° but this is because 6 slots are used, instead of 5. Using this observation, it is possible to use the same parameters, as in S°, to resolve the contention independently of the number of stations. This makes it easier to do contention using CONTI as the number of the stations doesn't need to be known and all the stations use the same parameters all the time.

2.5 Parameters: Maximized Throughput

In this section, we find the optimal parameters for CONTI, k and p, that maximize the throughput.

2.5.1 Time Utilization

The medium access with CONTI is shown in Figure 2.2. In the first attempt, a frame is transmitted successfully and an ACK is received in reply. However, the second attempt is a collision.

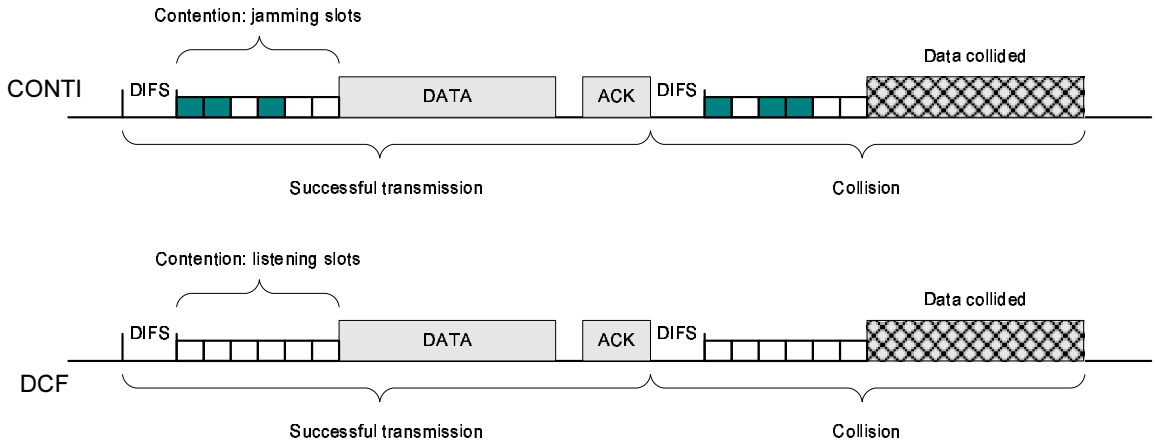

Figure 2.2 Transmission cycle of CONTI and DCF: a successful transmission and a collision.

Hence, the time utilization of CONTI, designated by ρ, is found as the following:

$$\rho = \frac{p_s.t_{data}}{p_s.T_{successful} + p_c.T_{collision}} \tag{2.8}$$

where t_{data} is the time to transmit the data frame. The probabilities of success and collision are given by p_s and p_c, respectively. $T_{successful}$ is the time for a successful transmission cycle, and $T_{collision}$ is the time consumed by a collision cycle, which are given as follows:

$$T_{successful} = t_{difs} + k.t_{slot} + t_{data} + t_{sifs} + t_{ack} \tag{2.9}$$

and,

$$T_{collision} = t_{difs} + k.t_{slot} + t_{data} \tag{2.10}$$

To be able to find the time utilization, we need to know the expression of the successful transmission event, p_s, which is given in the previous section in Equation (2.2).

2.5.2 Finding the Parameters

According to the above, if we know the number of stations, n, and the data transmission time, t_{data}, we can find the optimal parameters which are the number of slots, k, and the probability vector, p, that contains k elements. The optimal parameters maximize the time utilization of the access scheme.

We are interested in a number of scenarios in the Wireless LAN environment. Each scenario has a number of contending stations and a given frame size. Thus, we optimize the parameters having in mind all of the realistic scenarios in a WLAN environment. We consider the number of stations to be: 2, 5, 7, 10, 15 and 25. The frame sizes that we consider, in bytes, are 250, 600, 950, 1300, 1650, 2000 and the maximum frame size, 2346. The transmission rate is 11 Mbps.

Let j be the number of scenarios that vary the number of stations (here $j = 6$). And let k be the number of scenarios that vary the frame size (here $k = 7$). Hence, we have 42 scenarios, one for each value of n corresponding to a frame size. For a certain scenario with n stations and a given frame size, the time utilization is designated by $\rho^n_{FrameSize}$. Hence, we need to maximize the average throughput over all the scenarios given by: $\frac{1}{j.k}\sum_j \sum_k \rho^n_{FrameSize}$.

The parameters k and p are found by looking in the search space:

$$\begin{cases} k \geq 1 \\ 0 \leq p_1 \leq 1 \\ 0 \leq p_2 \leq 1 \\ \dots \\ 0 \leq p_k \leq 1 \end{cases} \tag{2.11}$$

and the objective function is to maximize the average time utilization:

$$\max(\frac{1}{j.k}\sum_j \sum_k \rho^n_{FrameSize}) \tag{2.12}$$

The optimal parameters obtained are $k = 7$ slots and the probabilities are $p = \{0.18, 0.31, 0.40, 0.48, 0.48, 0.49, 0.49\}$. For these parameters, the average time utilization over all the sce-

narios is 65.27%. These are the parameters that we use in the simulation results as elaborated in the discussion below.

2.5.3 Discussion

We have presented two methods to find the parameters of CONTI. The method in Section 2.4 was based on bounding the collision rate and using an algorithmic approach to finding the parameters k and p. On the other hand, the method in this section is based on maximizing the throughput and searching in the probability space to find the parameters. We compare these two methods as follows:

- The method in Section 2.4, which uses an algorithmic method to find the parameters, requires less computation time since it optimizes one slot and then moves to the next one. However, the optimization method presented in this section requires more computation time since it considers all of the combinations of p_i values over the slots.
- For the WLAN scenarios that we consider, the computation time is tractable. Hence, we use the results that were produced in this section.
- A comparison of the two methods in the simulation shows almost identical results. Hence, in the simulation we only use the results of the method in this section.

2.6 Characterization of the Time Utilization

This section presents an analysis on the throughput of CONTI and DCF. It also shows the effect of the slot length on the relative performance between CONTI and DCF. A more comprehensive evaluation is shown between CONTI, DCF, PREMA, k-EC and Idle Sense in the next section by simulation.

The access schemes of CONTI and DCF are shown in the illustration in Figure 2.2. While both of the schemes wait for the initial DIFS interframe space, the contention of CONTI uses the pulses, while the contention of DCF uses the backoff countdown. What follows the contention is similar for the two cases which is the data transmission, the wait of SIFS interframe

space, and the transmission of the ACK frame upon successful transmission. In the case of a collision, an ACK frame will not be received.

The characterization of the time utilization between CONTI and DCF can be written as the following.

$$gain_{conti} = \frac{\rho_{conti}}{\rho_{dcf}} \tag{2.13}$$

By observation from Figure 2.2, and using the derivation similar to Equation 2.8, we have:

$$gain_{conti} = \frac{p_s^{conti}}{p_s^{dcf}} \times \frac{p_s^{dcf}.T_{successful}^{dcf} + p_c^{dcf}.T_{collision}^{dcf}}{p_s^{conti}.T_{successful}^{conti} + p_c^{conti}.T_{collision}^{conti}} \tag{2.14}$$

The expressions for $T_{successful}^{conti}$ and $T_{collision}^{conti}$ are given in Equations (2.9) and (2.10), respectively.

The expressions for $T_{successful}^{dcf}$ and $T_{collision}^{dcf}$ are the following,

$$T_{successful}^{dcf} = t_{difs} + t_{contention} + t_{data} + t_{sifs} + t_{ack} \tag{2.15}$$

and,

$$T_{collision}^{dcf} = t_{difs} + t_{contention} + t_{data} \tag{2.16}$$

where $t_{contention}$ designates the average number of slots spent in a DCF contention.

2.6.1 Probability of a Successful Transmission

Since CONTI and DCF employ two different mechanisms for the contention resolution, it is obvious that they have different expressions for the probability of a successful transmission.

The characterization of the collision event in DCF was presented in (2). Accordingly, the probability that a station transmits in a slot, p_{tr}, and the probability that a stations has a successful transmission given a transmission attempt, p_{cond}, are given as follows:

Table 2.5 Analytic Results of Successful Transmission

No. of Stations	10	20	30	50	75
DCF	0.838	0.775	0.737	0.688	0.657
CONTI	0.982	0.976	0.970	0.959	0.944

$$p_{tr} = 1 - (1 - \tau)^n \tag{2.17}$$

$$p_{cond} = \frac{n\tau(1 - \tau)^{n-1}}{p_{tr}} \tag{2.18}$$

where n is the number of contending stations and τ and p are given as follows:

$$\tau = \frac{2(1 - 2p)}{(1 - 2p)(CW_{min} + 1) + pCW_{min}(1 - (2p)^m)} \tag{2.19}$$

$$p = 1 - (1 - \tau)^{n-1} \tag{2.20}$$

The analytic results of the successful probability of transmission are presented in Table 2.5 below. The results of DCF are based on the equations above from (2) and the results of CONTI are obtained using the equation we derived in Equation (2.2). For DCF and CONTI, as shown in Table 2.5, the probability of a successful transmission decreases as the number of stations increases.

2.6.2 Effect of the Length of the Contention Slot

To understand how the duration of the contention slot affects the system performance, we look at the number of slots that is consumed in each access to the channel. In CONTI, the parametrization has been studied for using 7 slots. However, it is a different story in DCF. By observing the number of idle slots spent between the end of DIFS and the initial moment of the transmission, this number is different and becomes smaller when the number of the stations increases.

For example, if there are 5 stations contending with the values of backoff counter equal to 7, 10, 13, 16, 19, it means in the first access, there will be 7 idle slots, and thereafter, there will be 3 contention slots preceding each data transmission. For the evaluation of the gain in time utilization, 3 slots are assumed for DCF and 7 slots are assumed for CONTI. Of course,

CONTI is using more slots but making up in providing a lower collision rate. In this study the data rate is 11 Mbps, the control rate is 1 Mbps and the frame size is 1500 bytes.

The results of the gain in time utilization are shown in Figure 2.3. CONTI has a gain in time utilization in comparison to DCF. Expectedly, with a larger time slot, the gain that CONTI has starts to decrease since CONTI is consuming more slots in one access. The other observation is, with the larger number of stations, the gain of CONTI increases due to its better collision rate. The time slots that are shown range in duration from 20 μs, which is commonly used in practice in IEEE 802.11b, to larger values up to 60 μs. The value of 60 μs represents the threshold where the gain of CONTI starts to disappear, for the number of 10 stations. This is seen in the figure, as the lowermost curve approaches 1 on the left-hand side.

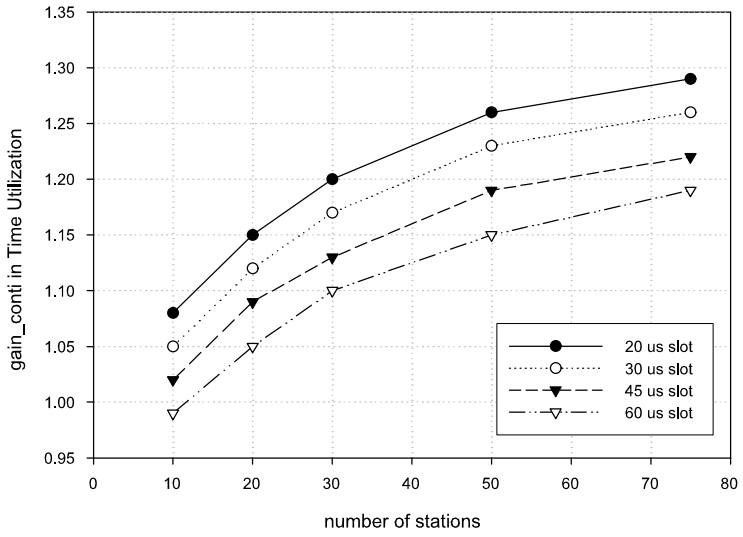

Figure 2.3 Effect of the Time Slot Duration.

More comprehensive evaluation between CONTI, DCF and the other schemes that we consider follow in the simulation results in the next section.

2.7 Simulation Results

This section presents the simulation results. We compare the performance of CONTI, DCF, PREMA, k-EC and Idle Sense. We consider the important measurements for the MAC such

Table 2.6 Physical Layer Characteristics (802.11b)

Characteristics	**Value**	**Comments**
tSlotTime	$20\mu s$	Slot time
tSIFSTime	$10\mu s$	SIFS time
tDIFSTime	$50\mu s$	DIFS = SIFS + 2 × Slot
aCWmin	15	min contention window size
aCWmax	1023	max contention window size
tPLCPOverhead	$192\mu s$	PLCP overhead

Table 2.7 Parameters of CONTI, PREMA, k-EC and Idle Sense

CONTI
Number of slots: $k = 7$
Probabilities: $p = \{0.18, 0.31, 0.40, 0.48, 0.48, 0.49, 0.49\}$
PREMA
Number of eliminations: $h = 4$
Geometric distribution parameter: $q = 0.5$
k-EC
Number of rounds: $k = 7$
Maximum length of a round: $m = 3$
Idle Sense
Target number of idle slots: $n_i^{target} = 3.91$
CW increase parameter: $\epsilon = 6.0$
CW decrease parameter: $1/\alpha = 1.0666$
Initial window size: $maxtrans = 5$
Window size adjustment parameters: $\beta = 0.75$, $\gamma = 4$

as the number of slots used, the collision rate, the throughput, the delay and the fairness to the users.

2.7.1 Parameters

We developed a discrete-event simulator for the MAC of wireless networks. The physical layer we consider is the 802.11b (26). Its parameters, with the parameters of DCF are summarized in Table 2.6. We program the simulation code and evaluate all the schemes in the same environment and in the same manner to ensure the fairness of evaluation.

The parameters of CONTI, PREMA, k-EC and Idle Sense are presented in Table 2.7. The parameters of all the other schemes are the optimal parameters from the papers in which they were proposed.

Table 2.8 Average Number of Slots in a Contention

No. of Stations	**CONTI**	**DCF**	**PREMA**	**k-EC**	**Idle Sense**
10	7	3.00	11.06	12.01	4.08
20	7	2.34	12.02	11.40	3.81
35	7	2.00	12.82	10.90	3.70
50	7	1.84	13.33	10.58	3.66
75	7	1.69	13.91	10.22	3.64
100	7	1.60	14.32	9.96	3.64

2.7.2 Number of Contention Slots

Every one of the schemes that we compare requires a certain number of contention slots. While the number of slots spent in contention isn't the only performance indicator, having a small number of slots is generally considered as preferable. On one side, CONTI takes a constant number of 7 slots and Idle Sense aims to achieve a number of slots equal to 3.91. On the other side, DCF, PREMA and k-EC spend a varying number of slots for each contention.

Table 2.8 shows the average number of slots spent in a contention. In this simulation experiment, the duration time is 1200 seconds and the frame size is 1500 bytes. The data rate is 11 Mbps and the control rate is 1 Mbps. The number of stations changes for different simulation runs as shown in the table. First, CONTI takes 7 slots for every contention, which is already known from the parameters. With DCF the number of slots is reduced with more stations even though the CW size is becoming larger. This happens since the number of slots that are spent is the minimum among all the backoff counters of stations. However, there are more collisions which will be shown next. PREMA has a different trend with more slots spent when there are more stations. This happens since the longest burst prevails in PREMA. With more stations, there's more chance to have a long burst. k-EC starts at 12 slots and takes less with more stations since the earliest jam finishes an elimination round. Finally, with Idle Sense, the number of slots fluctuates around the target number of 3.91. It is slightly higher with less stations slightly lower with more stations.

From this experiment, we conclude that the schemes with multi-round contentions, PREMA and k-EC use the largest number of slots, average more than 10 slots in the above scenarios. On the other hand, the schemes based on the contention window use the least number of slots,

average less than 4 slots. CONTI comes in the middle by using 7 slots.

2.7.3 Collision Rate

This part shows the collision rates of the schemes. The simulation parameters are similar to above with 1200 seconds of simulation time, 1500 bytes frames, 11 Mbps of data rate and 1 Mbps of control rate. The collision rate is shown in Figure 2.4.

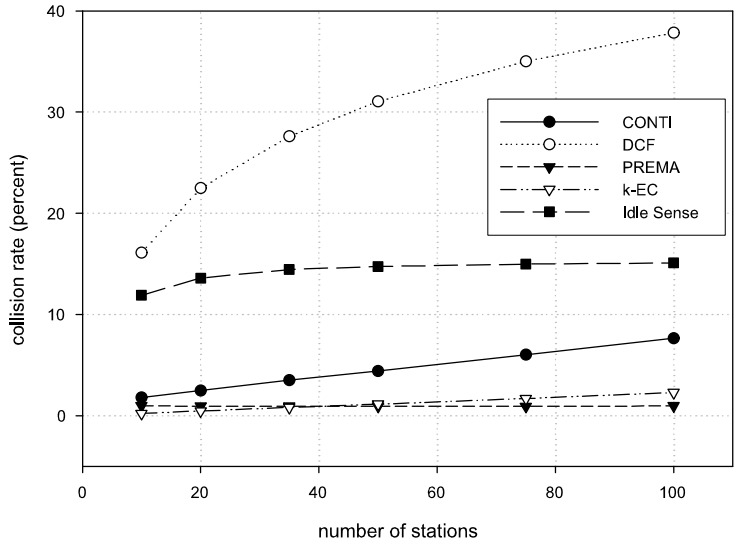

Figure 2.4 Collision rate.

The number of stations varies in the same numbers as in the previous experiment. First, we notice that PREMA and k-EC have the lowest collision rates. Across all scenarios, their rate is around 1 percent. However, their number of slots used per contention was the highest. CONTI has a collision rate that starts at 1 percent for 10 stations and climbs to 7 percent for 100 stations. Then, Idle Sense has a collision rate that varied between 11 and 14 percent. Finally, DCF had the highest rate of collisions that varied from 16 to 37 percent. From the previous experiment, DCF used the least number of slots per contention, but its collision rate turned out to be much higher than the other schemes. Typically, the trend is that the higher number of slots allows reducing the collision rate.

Finally, we mention an observation between the collision rate of PREMA and k-EC. At

a lower number of stations, k-EC has a smaller collision rate. However, their collision rate intersect at about $n = 35$, and for higher n, PREMA has a smaller collision rate. This trend happens since PREMA is based on a longest-burst-prevail policy. Therefore, for a large number of stations, there is more chance that the geometric distribution used will produce a large burst and therefore, there won't be a collision.

2.7.4 Throughput

The two previous experiments show the insightful measures of number of slots used and the collision rate. However, we need to understand their effect on the throughput that is achieved with the schemes. This part shows the throughput comparison of the schemes. The experiment environment is similar to above with 1200 seconds of simulation time, 11 Mbps for data transmission and 1 Mbps of control rate.

We would like to note that since the throughput of several of the schemes are close to each other, we used a quite large simulation time of 20 minutes to obtain stable results. We noticed that the throughput values at this time are stable for different simulation runs. For each scenario, we made 10 simulation runs and we show the results in Figures 2.5, 2.6 and 2.7. On top of every column in the chart, there is an error bar at a distance d from the column top. The error is defined as the maximum variation from the average value obtained. That is, the column top is at value x, then, all the values obtained are in the interval $[x - d; x + d]$. In the figures, the error bars are almost coincident with the column top since we encountered typical errors of 0.01 to a maximum 0.11; due to the large simulation time that we used.

The throughput in Figures 2.5, 2.6 and 2.7 corresponds to 5, 20 and 50 stations with 802.11b, respectively. Each figure has seven groups of bars, each corresponding to a frame size, ranging from 250 bytes to the maximum size of 2346 bytes. We note the following observations.

- The first observation is that, for a number of stations, the throughput increases with the larger frame size. This happens since with larger frames, there is less time spend in contention and more time spent in transmission in the simulation time. The same simulation time of 1200 seconds applies for all the cases.

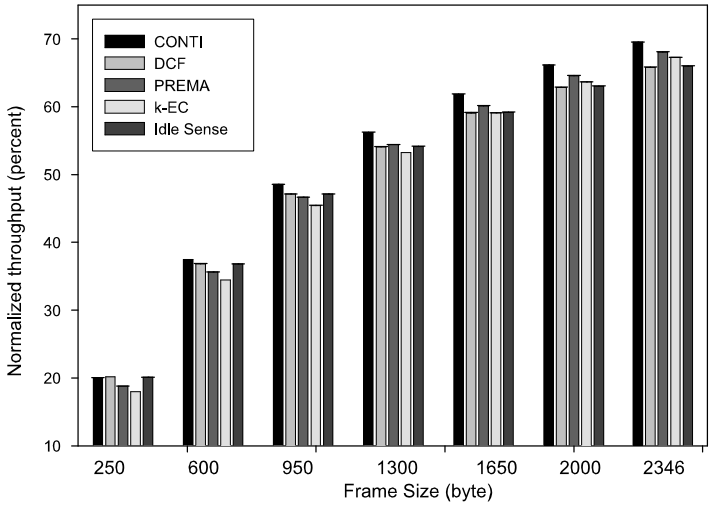

Figure 2.5 Throughput of 5 Stations with IEEE 802.11b.

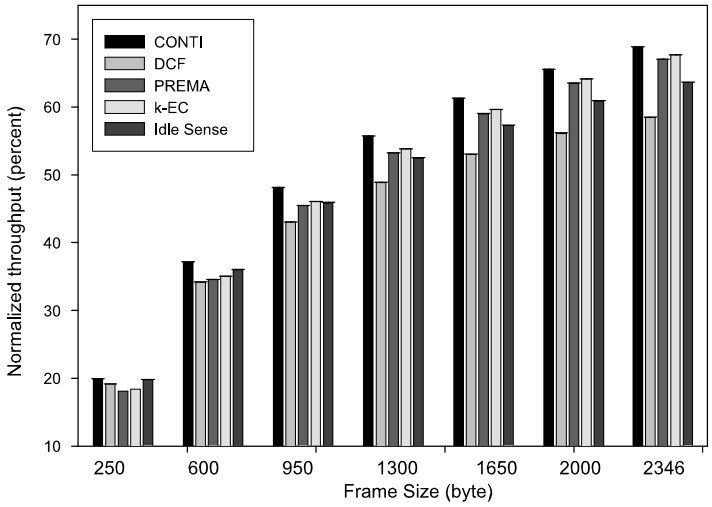

Figure 2.6 Throughput of 20 Stations with IEEE 802.11b.

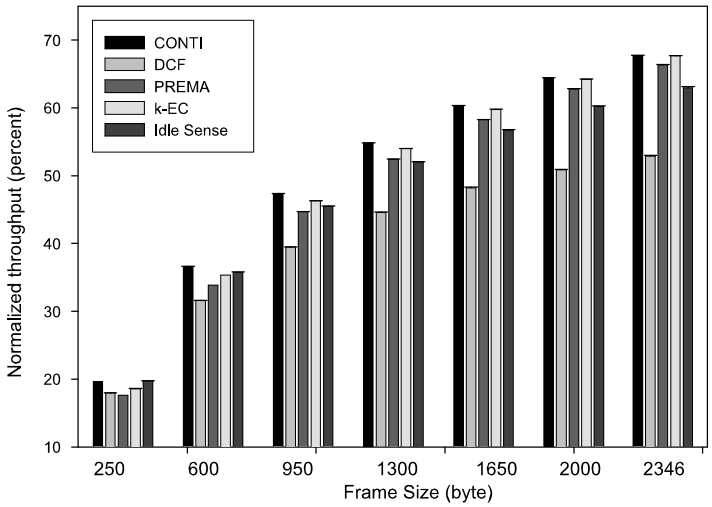

Figure 2.7 Throughput of 50 Stations with IEEE 802.11b.

- For a low number of stations, as in the first figure, the schemes have almost identical performance, although CONTI maintains a small advantage. The error bars along with the long simulation time validate the higher performance of CONTI, although by a small margin. More importantly, this figure shows that DCF has a similar performance to the other schemes for the low number of stations. This, however, is not the case with larger number of stations where DCF's performance starts to slip.

- We also notice that for the small frame size, PREMA and k-EC are behind DCF and Idle Sense in Figures 2.5 and 2.6. In other words, the jamming based schemes are behind the CW-based schemes (except for CONTI). This is because the advantage of PREMA and k-EC is their small collision rate. However, this advantage is not effective when the frame size is small. When the frame size is large, PREMA and k-EC surpass DCF and CONTI since these latter two's collision rate will waste a significant portion of time.

- Finally, we make the following observation on the throughput of PREMA and k-EC. In Figure 2.5, PREMA has a higher throughput. However, in Figures 2.6 and 2.7, k-EC's throughput surpasses PREMA's. This trend happens since k-EC requires less slot with more number of stations due to its earliest-jam-prevail policy.

2.7.5 Throughput with IEEE 802.11g

We also show the throughput evaluation with the IEEE 802.11g physical layer in Figure 2.8. We use the maximum data rate supported by 802.11g, which is 54 Mbps. The control rate we use is 2 Mbps, since this rate is mandatory to be supported in 802.11g. The simulation is also run for 1200 seconds. In 802.11g, even more frames will be transmitted in this time since the transmission time is shorter, and this will ensure stable results as well.

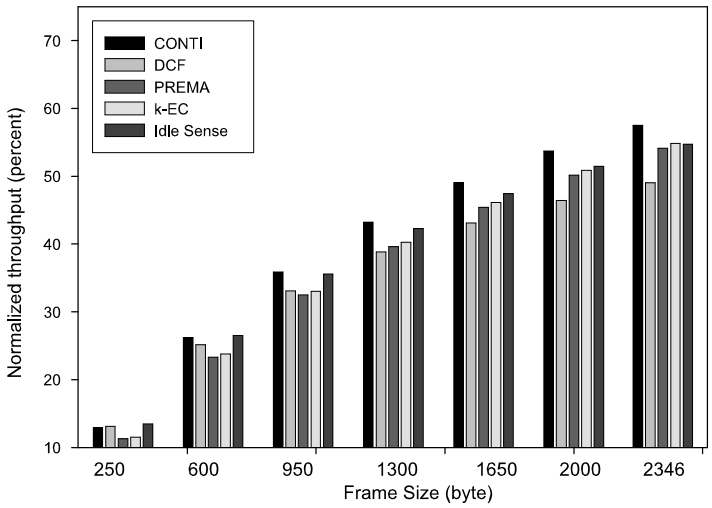

Figure 2.8 Throughput of 20 Stations with IEEE 802.11g.

In this part we use the following physical layer parameters, and not the ones in Table 2.6. The slot time is 9 μs, SIFS time is 10 μs, DIFS time is 28 μs, CWmin is 15, CWmax is 1023 and the PLCP overhead is 41.6 μs. However, we continue using 802.11b after this part for the delay and fairness measurements.

First off, since Figures 2.5, 2.6, 2.7 and 2.8 are on the same scale, we notice that the normalized throughput of 802.11g is lower. This is the normalized throughput percent, however, and 802.11g still has higher throughput in Mbps due to its higher rate. This trend happens because in this simulation there is a large gap between the control rate of 2 Mbps and the data rate of 54 Mbps. As a matter of fact, 802.11g is more sensitive to overhead since the opportunity to transmit a lot of data would be lost due to higher rates (27). From Figure 2.8, however, we

observe that CONTI maintains a slightly higher throughput than the other schemes for the frame sizes of 600 bytes and larger.

2.7.6 Delay

In this part we evaluate the delay of the schemes. The delay is defined as the time spent from when the frame arrives at the station's head of the queue to the time where it's transmitted successfully. Similar to above, the experiment to measure the delay has 1200 seconds of simulation time, 11 Mbps of data rate and 1 Mbps of control rate. The number of stations and the frame size are varied and the results are shown in Figures 2.9, 2.10 and 2.11. The figures also show the error bars, although it's not well visible in most of the cases since it's small. The error was typically between 0.01 and 0.45. Every result was obtained based on 10 runs of the simulations with 1200 seconds for each run.

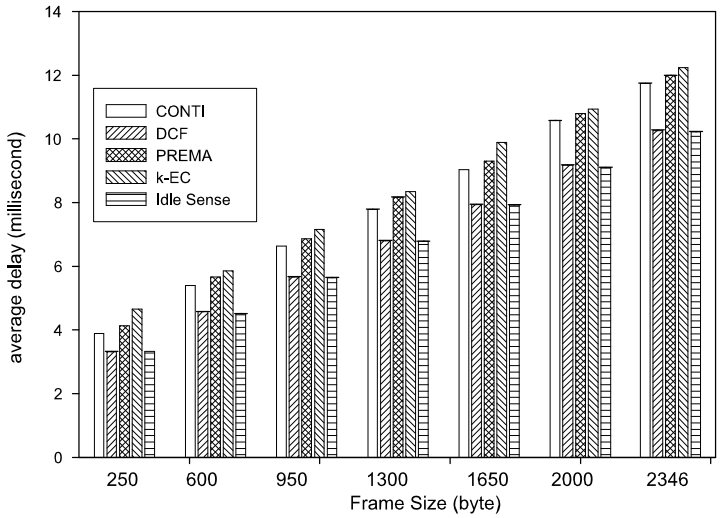

Figure 2.9 Delay of 5 Stations

With 5 stations in Figure 2.9, DCF and Idle Sense have the smallest delay. This is because they have the combination of a low collision rate with 5 stations, and also using a small number of slots. However, when there are more stations, in Figures 2.10 and 2.11, DCF's delay rises significantly and Idle Sense doesn't have an advantage anymore over PREMA and k-EC. For

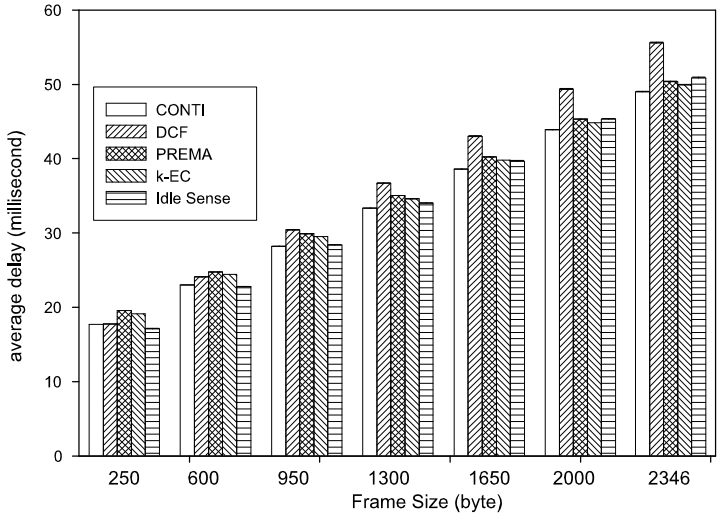

Figure 2.10 Delay of 20 Stations

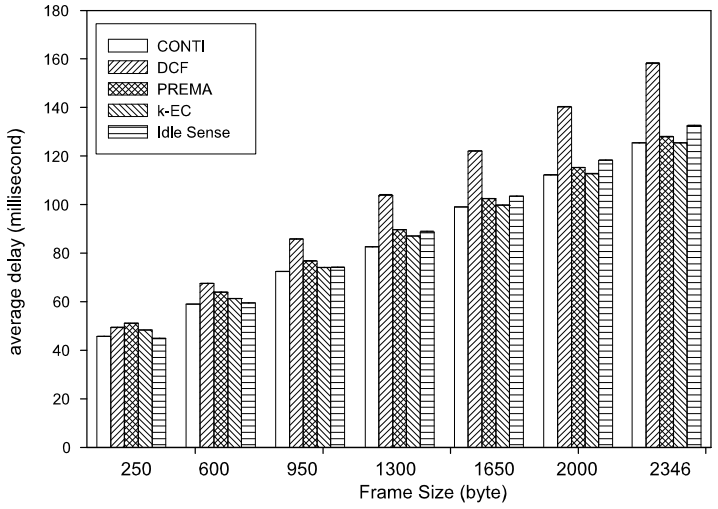

Figure 2.11 Delay of 50 Stations

20 and 50 stations, CONTI has a small advantage in the delay values over the other schemes.

2.7.7 Fairness

In this part, we evaluate the fairness of the MAC schemes. We use Jain's fairness index (28) for this measurement. To evaluate this index, we consider the number of active stations to be n, and the proportions of the transmitted frames with respect to the total number of successfully transmitted frames by each station is given by $x_1, x_2, ..., x_n$, respectively. The Jain's index is then given by the following:

$$f(x_1, x_2, ..., x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n. \sum_{i=1}^{n} x_i^2} \tag{2.21}$$

The value of this index ranges from 0 to 1. A value that is close to zero means a low fairness between the users, while a value approaching 1 designates a high fairness. We investigate several factors that affect the fairness: the time-scale (to understand short-term fairness and long-term fairness), the frame size and the number of the stations.

Short-term vs. Long-term Fairness: To evaluate the short-term fairness, we implement the sliding window mechanism that is presented in (29). This mechanism requires a transmission trace, which is a set of the stations' IDs, in the order in which they transmitted. First, we consider a window starting at the first element of the trace and of length w. We find Jain's index on this window. Then we slide the window (still of length w) by one element to the right and we find Jain's index again. We keep sliding the window until the right side of the window reaches the last element in the trace. We average all the index values on all the windows. This would be the average value associated with the window size w. Then, we plot the average index against the values of w.

In this scenario, we have 20 stations that are transmitting. The trace size is 5000. We vary the window size from 20 up to 1000 as we measure the averaged Jain's index. The results are shown in Figure 2.12. Apparently from the figure, DCF has the lowest fairness among all the window sizes. CONTI, PREMA and k-EC have a comparable fairness, while Idle Sense has a slightly higher fairness. As the time evolves into long-term, all of the schemes' fairness

approaches 1. However, the difference is in the speed of convergence which is the short-term fairness.

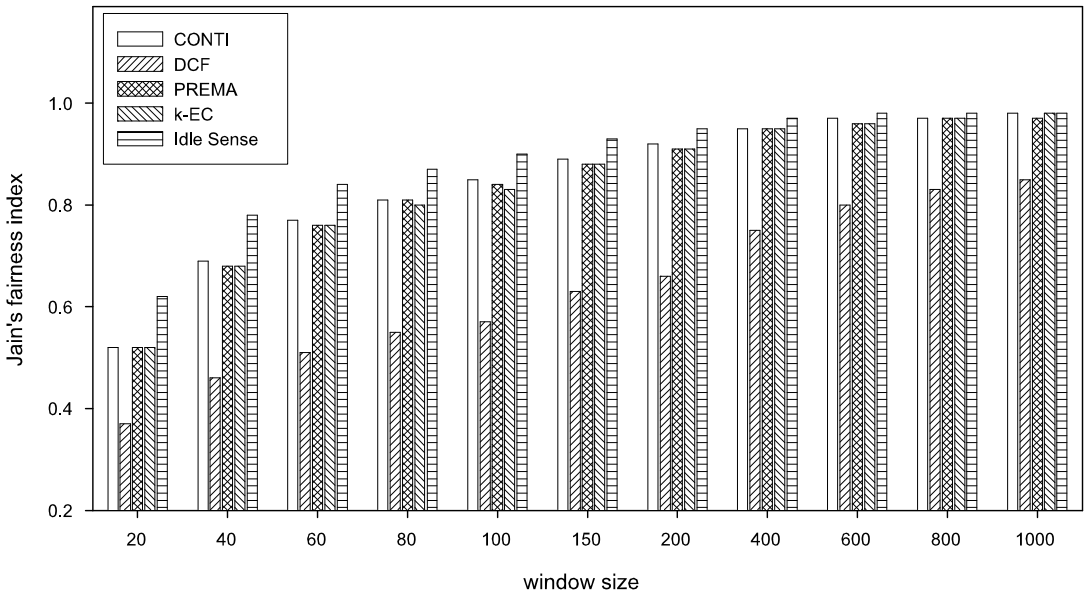

Figure 2.12 Averaged Jain's Index over Sliding Window.

One reason why Idle Sense has a better fairness than CONTI, PREMA and k-EC is that Idle Sense require each station to select one random number per access. On the other, CONTI, PREMA and k-EC, even though they give the same parameters to all the stations, they require the selection of multiple random numbers in one access. That might introduce more randomness in the distribution of access than with Idle Sense.

Fairness vs. Frame Size: We measure the fairness when the frame size changes. The results are shown in Figure 2.13. We observe that there is a trend that happens on a specific short term basis. For CONTI, PREMA, k-EC and Idle Sense, this short term is around 0.5 seconds. At this term basis, smaller frame sizes give a higher fairness. This is because there would be more frames transmitted with the smaller frame size; hence, fairness will increase in accordance with the result of Figure 2.12. At the simulation smaller than 0.5 seconds, the error bars were too large to draw any conclusion. A the simulation time that's larger than 0.5 seconds, this curve becomes too flat so that the trend disappears. It seems every scheme has

a short term interval where this trend would be observed.

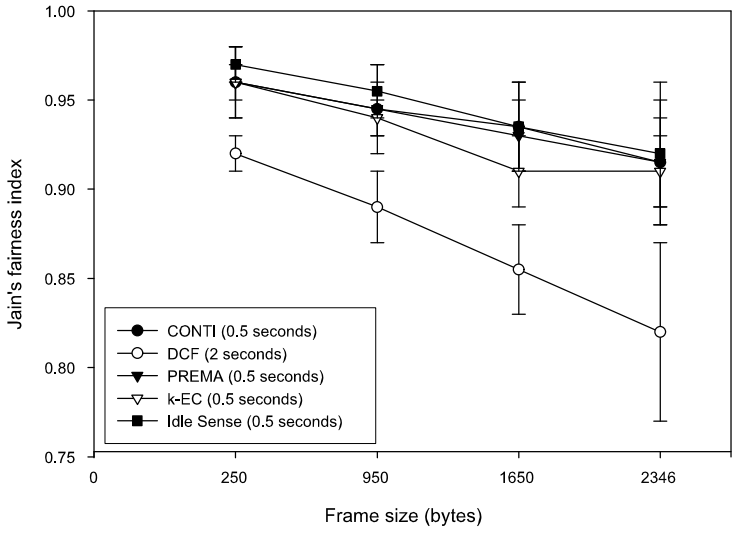

Figure 2.13 Fairness vs. Frame Size.

For CONTI, PREMA, k-EC and Idle Sense, this short term was 0.5 seconds. For DCF, this short term was 2 seconds, hence its result in Figure 2.13 are for 2 seconds of simulation time. All of the results are based on error from 10 simulation runs.

Number of Stations: We measure the fairness when the number of stations changes. Similarly to the previous experiment, this effect occurs on a time scale that is around 0.5 seconds for CONTI, PREMA, k-EC and Idle Sense, and at 2 seconds for DCF.

In this experiment, the frame size is 1500 bytes. The result is shown in Figure 2.14. Every value shows the error bar based on 10 simulation runs. The observation from this experiment is the following.

When there are more stations, it's more likely that unfairness will happen since there are more possibilities of the distribution of successful frames among the stations. This is the trend that's observed in the short term. As the simulation time increases, this trends diminishes.

Finally, we conclude the following from the fairness measurement.

1. In longer simulation time, more fairness is observed, unlike the short term where there's more difference between the schemes.

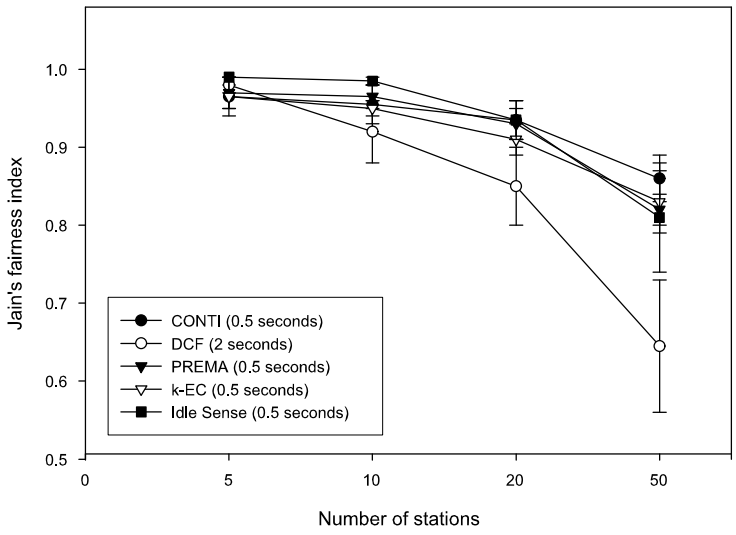

Figure 2.14 Fairness vs. Number of Stations.

2. The frame size affects the fairness in the short term only, where the small frame size gives more fairness. This disappears in the long term.

3. The number of stations affects the fairness in the short term only, where less stations give higher fairness. This observation diminishes in the long term.

2.8 Conclusion

This chapter presented a comparison of MAC schemes for wireless LANs. Our scheme, which attempts to resolve the contention in a constant-time (CONTI), was compared to other schemes, namely, DCF, PREMA, k-EC and Idle Sense. First, we reviewed the related work and described the operation of a few schemes that we compare against. Then, we presented the details of CONTI and obtained its optimal parameters in two ways: one algorithmic approach and one optimization approach. Following, we presented an analysis that shows the effect of the contention slot on the throughput of CONTI. Finally, in the simulation results, we compared the performance of CONTI to other schemes. From the experiments, we started by showing the number of slots in a contention and the collision rate. Then, we showed the throughput where

CONTI provided a small advantage over the schemes in the majority of the cases. We also showed the delay where DCF and Idle Sense provided the lowest delay for the small number of stations, equal to 5. However, CONTI provided the lowest delay for the medium and large size network, with 20 and 50 stations, respectively. Finally, we made a fairness comparison that showed that Idle Sense has the highest fairness, followed equally by CONTI, PREMA and k-EC, then finally by DCF. In the last part we showed that a smaller frame size and a smaller number of stations increase the fairness; but this trend happens temporarily on a short-term basis.

Acknowledgement

The authors would like to thank the anonymous reviewers for their comments. Especially, we believe that the addition of Section 2.5, which was made in response to the reviews, improves the quality of this chapter.

CHAPTER 3. Group-Based Medium Access Control for IEEE 802.11n Wireless LANs

An article that is submitted to *IEEE Transactions on Mobile Computing.*[1]

Zakhia Abichar[2] and J. Morris Chang[3]

3.1 Introduction

The latest generation of Wireless Local Area Networks (WLANs) is based on the IEEE 802.11n Standard (30). The standard was ratified in 2009 and has been in development for the prior few years. The standard uses the latest advances at the physical layer which achieve high data rates. The Multi-Input Multi-Output (MIMO) technology is used. MIMO systems use multiple antennae at the stations to increase the range and the link capacity. They also have other features such as allowing multiple users to transmit simultaneously without causing a collision.

One of the goals of the 802.11n standard is to achieve a throughput that is higher than 100 Mbps. To do that, the Medium Access Control (MAC) scheme of the WLANs was revised to improve its efficiency. The main mechanisms are summarized in the next section (Related Work). General information on the 802.11n standard can be found in (31) and (32).

The 802.11n standard maintains interoperability with 802.11 devices of previous generations (802.11b and 802.11g). As 802.11n devices start to become prevalent, a significant portion of the wireless devices would be based on previous generations. The interoperability allows all the devices to coexist.

[2]Primary researcher and author

[3]Faculty advisor

In this paper, we introduce a Group-based Medium Access Control (GMAC) protocol for wireless LANs that have high data rates and a large number of stations. With GMAC, stations are divided into groups based on their locations in the WLAN so that groups are free of hidden nodes. Each group has a leader and only group leaders participate in the contention for the medium. When a group leader wins the contention, it reserves time for all the stations in its group via RTS/CTS and it transmits a polling frame that contains the group's schedule. Stations in the same group are within the communication range of each other, thus they transmit in a contention-free and hidden-node-free environment without requiring RTS/CTS exchanges. GMAC achieves gain over DCF by allowing fewer stations to contend and by reducing the number of control frames. Moreover, GMAC is interoperable with legacy 802.11 devices; the legacy devices will content with the group leaders. When a group leader wins the contention and reserves time via RTS/CTS for all of its group, the legacy devices will observe a silent period as indicated in the RTS/CTS frames and will not interfere with the group operation of GMAC. Simulation results show that GMAC maintains a high throughput as the data rates and the number of stations increase.

To form groups that are free of hidden nodes, the stations should be able to estimate the distance to another stations. This will allow a station to know if it is able to join a group while keeping this group free of hidden nodes. This will be explained in the proposed scheme's details. There are schemes in the literature that propose a way to achieve localization and positioning, such as (33), (34) and (35). Alternatively, if the stations use a GPS receiver, they can find their locations and estimate the distance to other stations.

In a previous article (27), we presented the scheme of GMAC. In this article, we include analytic results and provide more extensive simulation results. We compare GMAC to the IEEE 802.11n standard which was not yet ratified at the time of the earlier publication. We also present comparison to state-of-the-art schemes that are found in the literature, TMAC (36) and E-GDCF (37), which we overview in the next section.

The simulation results show that GMAC provides a significantly higher throughput than the other schemes. GMAC also has one of the lowest collision rate. The results also compare

the delay characteristics of the MAC scheme and their fairness performance.

The rest of this paper is organized as follows. Section 3.2 presents the related work and Section 3.3 presents the system model. The proposed scheme is presented in Section 3.4 and the analysis is presented in Section 3.5. Finally, the simulation results are presented in Section 3.6 and the conclusion of the article is in Section 3.7.

3.2 Related Work

This section presents the related work. We review the IEEE 802.11n Standard, two schemes that we compare to our scheme in the simulations (TMAC and E-GDCF) and other schemes in the literature on the MAC of WLANs with high data rates.

3.2.1 The IEEE 802.11n Standard

The IEEE 802.11n Standard (30) presents several improvements to deal with the high data rates. First, a new Inter-Frame Space (IFS), called Reduced Inter-Frame Space (RIFS), is defined to reduce the amount of time between frames. RIFS may be used instead of Short Inter-Frame Space (SIFS) to separate multiple transmissions of a single station. These frames must be destined to the same recipient. Also, the station using RIFS must support the high data rates (i.e., it is not a legacy station). In 2.4 GHz band, RIFS is 2 μs, whereas SIFS is 10 μs.

The standard also supports Aggregate MSDUs (A-MSDU) in order to increase the efficiency. The MSDUs in an A-MSDU should be transmitted to the same receiver. The aggregated MSDUs should also have the same priority parameters. The lifetime timer of the A-MSDU expires when the timers of all MSDUs in it expire. The standard also supports the transmission of Aggregate MPDUs (A-MPDU). When the transmitter of the A-MPDU is the AP, the MPDUs can be addressed to multiple recipients.

Another mechanism in 802.11n is called Dual CTS Protection. It is used when the stations use the technique of Space Time Block Coding (STBC). STBC increases the range of the BS, however, this type of transmission is not understood by legacy devices. Thus, when STBC is

used, a station would transmit an RTS frame to the AP, the AP would reply with a CTS frame in STBC, followed by another CTS frame in non-STBC.

The standard introduces a Block ACK mechanism that aggregates ACK frames that are destined to a recipient. The ACK frames in a Block ACK should be in response to frames that have the same priority parameters. There are two variants of Block ACK. The Immediate Block ACK technique specifies that the sender transmits a number of data frames. They are followed by a Block ACK Request frame. The receiver then immediately transmits the Block ACK frame after a SIFS duration. The Delayed Block ACK mechanism specifies that the sender, after transmitting a number of data frames, transmits a Block ACK Request frame. The receiver replies with an ACK frame to acknowledge receiving the Block ACK Request. The Block ACK may be transmitted in subsequent TXOPs. In the mean while, the sender can transmit further data frames to the same recipient.

Another mechanism in the standard is called the Reverse Direction Protocol. This mechanism allows the transmission of data in both ways during a TXOP (originally, only the owner of the TXOP used to transmit data). To initiate this mechanism, the TXOP holder includes in a PPDU a Reverse Direction Grant. This allows the receiver station to transmits data frames in the TXOP.

The standard also introduces a 20/40 MHz BSS operation mode. In this mode, the AP and the associated stations of the BSS transmit either in a 20 MHz channel (the primary channel) or in a 40 MHz channel (the primary and secondary channels). The 20/40 MHz mode also defines measures to avoid interference with other BSSs. For example, if an AP is operating in the 20/40 MHz mode and detects an overlapping BSS whose primary channel is the AP's secondary channel, this AP will switch to 20 MHz mode. The AP might later switch to another pair of channels. The switching between the 20 MHz mode and 20/40 MHz mode and the switching to another pair of channels should happen after notifying all the associated stations, including the ones using the power-save mode.

3.2.2 Token-Coordinated Random Access MAC (TMAC)

The scheme Token-Coordinated Random Access MAC (TMAC) was presented in (36). TMAC divides the stations into groups. The AP rotates a token among the groups in a round-robin way. When a group has the token, its stations transmit. The stations in a group access the channel by contending among each other.

In TMAC, the number of groups is g and the groups are called $V_1, V_2, \ldots, V_g$. The number of stations in group V_i is N_{V_i} and is smaller than the maximum value, $\overline{N}_V$. When group V_i has the token, it has the right to transmit for a duration called the Token Service Period, denoted by TSP_i. This variable is defined by $TSP_i = (N_{V_i}/\overline{N}_V).\overline{TSP}$, where $\overline{TSP}$ is a parameter.

In the group contention, a station transmits at most once for each time its group gets the token. The contention in group V_i is done using the contention window value, CW_i, which is given to the group by the AP in the token frame. If a station collides, it doubles it contention window, up to m times.

When station s_i accesses the channel, it has the right to transmit for a duration that depends on its current rate, r_i. The transmission duration also depends on two parameters that are given by the AP in the token frame, which are: the reference rate R_f and the reference duration T_f. Accordingly, if $r_i \geq R_f$, then the station has the right to transmit for a duration of $(r_i/R_f).T_f$. Otherwise, the station has the right to transmit for a duration of T_f.

TMAC uses the Block ACK mechanism. The maximum number of ACK frames in a Block ACK is designated by A_f, which is a parameter.

In the duration given to group V_i, which is TSP_i, the stations of this group might finish their transmission before the end of TSP_i. This case is detected by finding the maximum number of idle slots in the contention, called TIFS. We have $TIFS = DIFS + (m.CW_t.\sigma)$, where σ is the slot duration. Thus, after a duration of TIFS where the channel is idle, the token is seized from the current group and the AP passes it to the next group.

In the simulation results in (36), the values of the parameters are: $\overline{N}_V = 15$, $\overline{TSP} = 35\ ms$, $m = 2$, $T_f = 2\ ms$ and $A_f = 2$. The value of R_f was varied in the simulation results.

3.2.3 Enhanced Grouping-based Distributed Coordination Function (E-GDCF)

The scheme Enhanced Grouping-based Distributed Coordination Function (E-GDCF) was proposed in (37). E-GDCF enhances the earlier scheme Grouping-Based DCF (GB-DCF) by the same authors (38).

In E-GDCF, when the number of active stations, M is larger than eight, a grouping mechanism is used to reduce the number of contending stations. The AP broadcasts the parameters N, N_h and k. The parameter N designates the number of groups. The target of E-GDCF is to have 2 stations in each group. Thus, $N = \lfloor M/2 \rfloor \geq 1$. The stations are divided based on their MAC addresses. The parameter k designates the starting bit position of the MAC address that should be used to divide the stations into groups. For example, if $N = 4$ and $k = 0$, the stations that have 00 in the LSBs of their MAC addresses are in group 0. Two bits are considered since $log_2 4 = 2$ and the LSB is considered since $k = 0$. Similarly, the stations that have 01 as the LSBs of their MAC addresses are in group 1, and so on. For another example, if $N = 4$ and $k = 1$ the stations with MAC address ending in 110 belong to group 11 (we start considering from bit position 1, not bit position 0). The parameter N_h is the group head, which is the group that should start transmission. The other groups follow in sequential order.

When a group starts transmission, the stations in it do contention using $CW_{min} = CW_{max}$ that is either 8 or 4. The results in (37) show better results for the value 8; thus, we use this value when we simulate E-GDCF. E-GDCF aims at obtaining groups of similar sizes, thus, N and k are chosen such that the groups have similar sizes. A scoring system gives a score for each (N, k) combination. The combination that gives the smallest standard deviation in the group sizes is selected.

One group cycle in E-GDCF starts with a DIFS, followed by the contention of the first group and data transmission. There is no ACK in reply to the data. Immediately following the data, the second group contends and transmits one data frame. Once the last group contends, a SIFS elapses and the AP transmits a Block ACK for all the data. The Block ACK has ACK frames for all the stations. Each station has transmitted only one data frame.

3.2.4 Other Works in the Literature

There have been several proposals that address various areas in the WLANs with high data rates. We summarize below a representative part of them.

A MAC scheme is proposed in (39) to allow the stations with high data rates to transmit in order to increase the throughput. The AP broadcasts a *reference data rate* and the stations contend using the CSMA/CA algorithm. When a station wins the contention, it transmits only if it is able to transmit at a rate that is higher than the reference data rate. Otherwise, the station gains credit. Once the station has enough credit, it transmits upon winning the contention at any data rate it is capable of.

In (40), the proposed MAC scheme introduces a delay to the frames. This allows more data to arrive from the higher layers and to use frame aggregation. The delaying procedure was shown to improve the throughput and keep the delay acceptably low. However, in some cases with TCP flows when the window size is small, the TCP layer will send only a few frames to the MAC layer. The MAC layer will not send this data as it is waiting for more data from the TCP layer to aggregate them. At the same time, the TCP layer is waiting for the ACK before sending more data. Thus, this might lead to starvation. Significant throughput dropped were reported in this case. To solve this issue, the TCP traffic was excluded from the delaying procedure.

In (41), a MAC scheme was proposed that provides access to the stations through a polling mechanism. Frame aggregation is also used to improve the efficiency. This scheme, however, does not provide a distributed mode. In (42), several enhancements in the 802.11n are evaluated. In (43) and (44), the reverse direction protocol of the 802.11n is evaluated. In (45), (46), (47), (48) and (49), frame aggregation mechanisms are introduced and evaluated. In (50) and (51), performance evaluation of the 802.11n is presented.

Several works have been proposed in the literature that use MIMO features in the MAC scheme. The MAC scheme in (52) allows multiple users to transmit simultaneously by using a MIMO-based physical layer. With the MAC scheme in (53), a station that accesses the medium might use a subset of the available channels. This would leave the remaining channels

available to other users. This scheme improves the utilization of the channels and reduces the number of collisions since multiple stations can transmit at the same time. Other schemes that use MIMO techniques to allow concurrent transmissions are in (54), (55) and (56).

Schemes have also been proposed to operate in the 20/40 MHz mode that was defined in the 802.11n standard, such as (57), (58) and (59). In (60), a comparison is provided between channel bonding (when contiguous channels are merged into a broadband channel) and multi-channel techniques (when a station transmits at the same time on multiple channels that are not contiguous).

3.3 System Model

This section presents the network configuration that we consider for the WLAN system.

3.3.1 Network

We consider a WLAN in which the AP connects to wireless devices which are the end users. The number of connected stations is potentially large due to the large bandwidth that is available at the AP. The transmissions are either form the station to the AP or from the AP to the station.

3.3.2 Distance Estimation

For the purpose of forming groups that are free of hidden nodes, the stations should be able to estimate their distance to the group leaders.

Distance estimation and localization technologies are becoming more available and more accurate. Localization technologies based on Received Signal Strength (RSS) have been proposed earlier. The articles (33), (34) and (35) present localization schemes based on RSS that can be implemented with off-the-shelf 802.11 devices. The stations might also find their position using a GPS receiver. Variants of GPS, such as A-GPS (Assisted GPS) and E-GPS (Enhanced GPS), relieve the line-of-sight (LOS) requirement to allow operation indoors and improve in the computation time.

3.3.3 Use of Time-Based Fairness

In the WLAN transmissions, the MAC scheme specifies for how long a station will transmit once it accesses the medium. If a station transmits one data frame upon each access, each of the station will transmit the same number of frames and obtain the same throughput. This is called throughput-based fairness. If the station transmits for a given duration of time upon each access, each of the station will transmit for the same duration. This is called time-based fairness.

Previous research results such as (61) and (62) have shown that in WLANs where the stations have different data rates (multi-rate networks), time-based fairness allows significant gain in the total throughput of all the stations.

If each station transmits only one frame upon access in the multi-rate WLAN, the stations with low rates will take a long time to transmit their data frame. Then, the stations with high rates will transmit their data frame quickly and will spend most of their time waiting for the slow stations. Then, the high rates that are available at some stations are not leveraged.

This is illustrated in the following analysis. There are two stations in the WLAN, one that has a high rate, r_{big} and another one that has a low rate, r_{small}. The rates are in Mbps. We find the throughput over a duration t that is achieved with either the throughput-based fairness policy or with time-based fairness policy.

3.3.3.1 Throughput-Based Fairness

With the throughput-based fairness, the stations take turn in transmitting one frame each. Over the duration t, the stations have transmitted the same number of frames. The first station takes a duration of L/r_{big} to transmit one frame, where L is the size of one frame in Megabits. The second station takes a duration of L/r_{small} to transmit one frame. Then, to transmit one frame from each station it takes this amount of time:

$$\frac{L}{r_{big}} + \frac{L}{r_{small}} = \frac{L(r_{big} + r_{small})}{r_{big} \times r_{small}} \tag{3.1}$$

Table 3.1 Throughput-Based Fairness vs. Time-Based Fairness

r_{small}	r_{big}	$\rho_{through}$	ρ_{time}
11	11	11	11
11	54	18.27	32.5
11	130	20.28	70.5
11	216.7	20.93	113.85

Then, the total throughput achieved with the throughput-based policy, called $\rho_{through}$ is:

$$\rho_{through} = \frac{2 \times r_{big} \times r_{small}}{(r_{big} + r_{small})} \tag{3.2}$$

3.3.3.2 Time-Based Fairness

When a time-based fairness policy is used, each of the stations will use the same amount of time. Each station will use a duration of $t/2$ to transmit data frames at its own rate. The first station will transmit $r_{big}.t/2$ Megabits. The second station will transmit $r_{small}.t/2$ Megabits. The total throughput that is achieved with time-based fairness, called ρ_{time}, is:

$$\rho_{time} = \frac{r_{big} + r_{small}}{2} \tag{3.3}$$

The numerical results are shown in Table 3.1. In the first line, the two rates are equal to 11 Mbps and the throughput under the two policies is the same, equal to 11 Mbps. In the second line, $r_{big} = 54$ Mbps, and now there is difference. The time-based fairness gives a much higher total throughput. In the third line, r_{big} becomes 130 Mbps and the time-based policy gives a throughput that is threefold that of the throughput-based policy. The time-based policy continues to give a much higher throughput in the fourth line. We also observe between the third and fourth lines, when r_{big} changed from 130 Mbps to 216.7 Mbps, the total throughput under throughput-based policy merely increased from 20.28 Mbps to 20.93 Mbps. But with the time-based fairness the increase was considerable from 70.5 to 113.85 Mbps. This shows that the throughput-based policy doesn't really leverage the high data rates in the presence of station with low rates. Accordingly, we use the time-based fairness policy in our proposed scheme.

3.4 Proposed Scheme

In this section, we present the Group-based MAC (GMAC) scheme in details. First, we describe how stations can be divided into groups that are free of hidden nodes. Then, we describe the contention of group leaders and the polling-based transmission of non-leaders. Finally, we discuss how groups are maintained when stations leave the network or move their location.

3.4.1 Overview

The main ideas of GMAC are:

- The stations are divided into groups based on their locations in the WLAN so that groups are free of hidden nodes. Each group has one station designated as leader.
- Only the group leaders contend for the medium using the DCF scheme. When a leader wins the contention, it reserves transmission time for all of its group using RTS/CTS. The leader then transmits a polling frame that contains the group's schedule.
- Stations in the same group transmit according to the schedule in the polling frame. There is no need for the RTS/CTS exchange as the leader has reserved time for the entire group and there are no hidden nodes in the group.
- Every station is allocated a time period that is sufficient to transmit one data frame at the lowest rate (time-based fairness). If a station can transmit at a higher rate, it aggregates multiple MSDUs for which it received a Block ACK frame.

3.4.2 Formation of Groups

The stations should be divided into groups that are free of hidden nodes. To achieve this, a new station joins a group if its distance to the leader is smaller or equal to $R/2$, where R is the communication range. In this way, the maximum distance between any two stations in a group is R. When there are several leaders within $R/2$ of the new station, the group with the

closest leader is selected. If the distance estimation error in bounded by δ, then a new station will join a group if its estimated distance to the leader is smaller or equal to $R/2 - \delta$.

The new station listens initially for a period of T_{listen} and estimates its distance to the group leaders that it can hear. If an RSS-based distance estimation scheme used, the station will estimate its distance to the leader based on the signal strength. If a positioning scheme is used, i.e. a scheme that provides each station with its position, the new station will know the position of the leaders from the polling frame. The positioning scheme can be based either on the signal strength or on a GPS technology. The polling frame format is shown in Figure 3.1. There is an optional field that is the "Location of Leader". If a distance estimation scheme is used, there is no need to know the location of the leader. Only the distance matters, and thus this field will not be used. However, if the position scheme provides the location of the stations, then the leader will broadcast its location in this field. The new station will find its location and compute the distance to the leader.

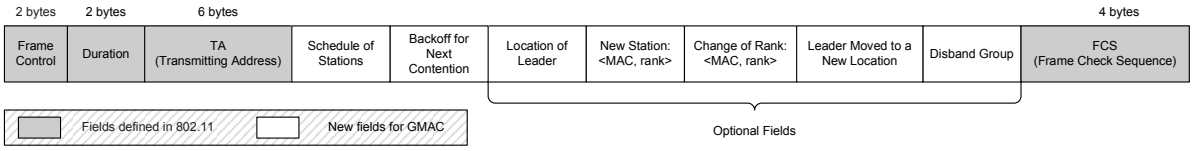

Figure 3.1 Polling Frame of GMAC.

In the Association Request frame that the new station sends to the AP, the new station indicates which leader it would like to join. This information is repeated in the *Association Response* frame. The group leader will decode the association frames and will take note of the new station joining its groups. Next time the group leader transmits a polling frame, it will use the "New Station" field and indicates the MAC address and the assigned rank of the new station.

If the new station cannot find a group leader within distance of $R/2$ from it, it will use group ID -1 in the Association Request frame to say that there is no existing group that it can join. This means that the new station will start a new group and become its group leader. From now on, this station will transmit a polling frame when it accesses the channel. The

procedure for a new station to join a group is summarized in Algorithm 2.

Algorithm 2 Procedure for a New Station to Join a Group

-Step 1: Listen to the network for a duration of T_{listen} and estimate the distance to the existing leaders.
-Step 2: Choose the group that has the closest group leader.
-Step 3: Do Association Request with the AP and indicate the group leader to join.
-Step 4: The selected group leader will take note and add the new station to the polling frame's schedule.

-Step 5: If no group is found within $\frac{R}{2}$, start own group. Indicate this in Association Request. Broadcast the polling frame.

3.4.3 Contention of the Group Leaders

The group leaders contend using a modified version of the DCF scheme (1). The modification we use is that the contending stations, which are the leaders, include in the value of the Back-Off (BO) timer that they will use in the next contention. This mechanism was proposed in (63) and (64) to allow contending stations to know each other's BO timers and avoid collisions. In GMAC, the BO timer of the leader is transmitted in the field "Backoff for Next Contention" in the polling frame. This field is used by the stations in the leader's group to detect if the leader has left the network without a formal Disassociation Request. This mechanism will be explained later in this section in the maintenance discussion.

When a leader wins the contention, it reserves time for itself and for all the stations in its group. This is done with an RTS/CTS exchange with the AP, as shown in Figure 3.2. For each station in its group, the leader reserves time that is sufficient to transmit the largest data frame (2346 bytes) at the lowest data rate in the physical layer. The reserved time for a station is also sufficient for the inter-frame spaces and the reception of the ACK or Block ACK frame. If the station is able to transmit at a rate that is higher than the lowest rate, it will transmit an Aggregate-MSDU (A-MSDU) frame which contains data frames that are destined to the same recipient. In response, the station will receive a Block ACK frame.

If the leader doesn't have data to transmit, it will contend and transmit the polling frame.

The group leader also has the responsibility to monitor the transmission of its group and

to end the reserved time in case some of the stations don't have data to transmit. If that's the case, the group leader transmits an End-NAV frame to end the reserved time and open medium for contention. The End-NAV is also repeated by the AP so that all the station in the WLAN receive it. This is shown in Figure 3.3. The details on the contention-free transmission within a group are presented next.

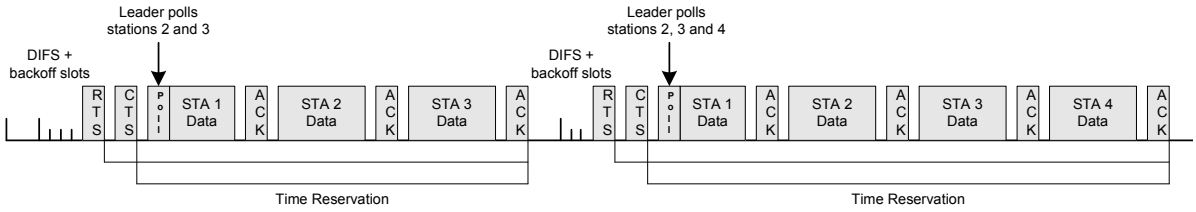

Figure 3.2 Transmission of GMAC. Scenario where no stations skip transmission.

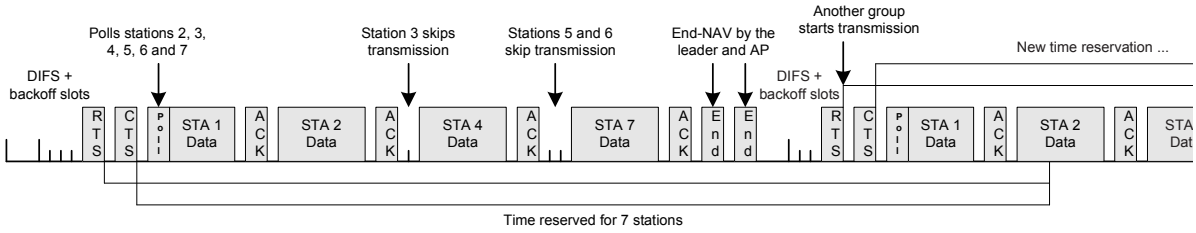

Figure 3.3 Transmission of GMAC. Scenario where some stations don't have data frames and skip transmission.

3.4.4 Transmission of Stations in a Group

When a group leader wins the contention, it transmits an RTS/CTS exchange with the AP to reserve time for all the stations in the group. Then, the leader transmits a polling frame. If the leader has data to transmit, the polling frame is aggregated in A-MPDU to the data frame.

The polling frame contains the schedule of the stations that are in the group. The stations, hence, will transmit in a contention-free mode. Also, no RTS/CTS frames are needed since the group doesn't have hidden nodes.

The stations transmit in the same order as the schedule of the polling frame. A station transmits one data frame, or an A-MSDU if it has a high rate, waits for a SIFS, to receive a potential ACK frame or a Block ACK frame.

After a SIFS duration elapses, the next station on the schedule transmits. Then, the next station transmits and so on. Some stations might not have data to transmit. These stations remain silent. Then, after the elapse of an additional SIFS duration, the next station on the schedule starts transmission. Since the groups don't have hidden nodes, the next station can rightly interpret the idle medium as the event of no data at the previous station.

3.4.5 Group Maintenance

There are several events that require the maintenance of the information in GMAC. A station, whether a leader or not, might leave the network. The station might leave the network by observing the protocol, i.e., issuing a Disassociation Request. In another event, the station might run out of battery or the system on it might crash; thus the station might leave without disassociating from the network. In addition, a station, whether a leader or not, might move within the WLAN area. Hence, the groups need to be maintained so that they stay free of hidden nodes. This part considers these scenarios and presents the corresponding operations in these cases.

3.4.5.1 A Station Issues a Disassociation Request

In this event, the station issues a Disassociation Request frame and leaves the network. This is the normal way of leaving the network. If this station is a non-leader, then the leader of its group will know that this station has left because the leader decodes the association and disassociation frames. In this case, there might be a gap in the ranks of the station. If the ranks of the group are 1, 2, 3, 4 and 5, and station 3 leaves the network, then the ranks will be 1, 2, 4 and 5. To avoid having a SIFS duration wasted in every transmission of the group, the leader will take the station with the highest rank (station 5) and change its rank to that of the station that has left (station 3). This is done with the "Change of Rank" field of the

polling frame; the MAC address and the new rank of the station are indicated. If the station with the highest rank (station 5) has left, then no action is needed.

If the station leaving the network is a leader, then its group is disbanded by using the "Disband Group" field of the polling frame. The stations of this group will join existing groups or form new groups according to the initial procedure in Algorithm 2.

3.4.5.2 A Station Leaves without Issuing a Disassociation Request

A station might leave the network without issuing a Disassociation Request. This might happen if the operating system on the station crashes suddenly or if its battery runs out. For a non-leader station, this event is detected as the following. When the station is idle for a duration that is greater than the parameter, T_{idle}^{max}, the leader assumes that the station has left the network without a Disassociation Request. The group leader keeps a table that has the Last Time of Transmission (LToT) for every station in its group. When the station with rank i in group j transmits data at time t, then the leader of group j records in the table $LToT_i^j = t$. Every while, the leader subtracts the $LToT$ of the stations from the current time $(currentTime - LTOT_i^j)$ for all the entries in the table. If the obtained result is greater than T_{idle}^{max}, it means the corresponding station has been idle for a long time and it would be removed from the group. If a station is idle for a while and doesn't wish to be removed from the group, then it can transmit an empty frame to itself periodically every T_{idle}^{max} to avoid being removed. The value of T_{idle}^{max} should be large, for example 10 seconds. Even if the user hasn't given an input on the computer for 10 seconds, usually, the applications running will transmit some traffic within this duration.

If a leader leaves the network without issuing a Disassociation Request, the stations in its group cannot transmit. This event is detected by having the stations observe the backoff timer that the leader will use in the next contention. This value is transmitted in the field "Backoff for Next Contention" in the polling frame. Using this field, the stations of the group can predict when the leader will transmit. If they don't hear the leader, then they will wait for a period of T_{leader} before assuming that the leader has left the network. This wait period

Table 3.2 Parameters & State Information of GMAC

$T_{initial}$	Duration for a new station to listen to leaders estimate its distance to them
T_{idle}^{max}	Maximum idle duration for a non-leader station before it is removed from its group
$LToT_{rank}^{groupID}$	Last Time of Transmission for station with ID $rank$ in group with ID $groupID$
T_{leader}	Time before declaring that the leader has left the network

will be also useful if the leader has collided and has not, in fact, left the network. The value of T_{leader} should not be very large since it affects the user experience. During a wait of this period, the user might not receive service. It could be set to 1 or 2 seconds; such a duration is tolerable by the end user but is considered a large duration in the MAC operation.

3.4.5.3 A Station Moves in the Network Area

The stations might move within the WLAN area while remaining associated to the same AP. The mobility of the stations might introduce hidden nodes in the groups. When a non-leader station moves, it will find its new distance to the leader. If this distance remains smaller or equal to $R/2$, then the station doesn't need to change anything. However, if the distance to its leader becomes larger than $R/2$, then the station will perform the initial procedure of joining a group as described in Algorithm 2.

If a leader stations moves, it will find out if it can join an existing group. If this is the case, the leader will disband its group by using the "Disband Group" field of the polling frame and it will join an existing group. However, if the leader moved its location and couldn't find an existing group in the new location, then it will remain a group leader. In this case, the leader will notify the stations in its group that it has moved by using the field "Leader Moved to a New Location". The stations in the corresponding group that are still within a distance of $R/2$ to the leader will remain in the group. The stations that are not within a distance of $R/2$ to the leader will perform the initial procedure described in Algorithm 2.

The parameters and the state information used in GMAC are summarized in Table 3.2.

3.4.6 Interoperability with Legacy 802.11

GMAC is interoperable with the legacy 802.11 devices such as 802.11g and 802.11b. The legacy devices will contend with the group leaders to access the medium. The legacy devices will not get involved in the contention-free group transmission. They will set their NAV timer according to the RTS/CTS frames that is done when a group starts to transmit.

3.5 Analysis

In this section, we present an analysis that compares the throughput performance of three schemes: (1) GMAC, (2) DCF with frame aggregation, such as the one specified in IEEE 802.11n, which we call DCF-Agg and (3) the regular DCF scheme without frame aggregation.

3.5.1 Performance Gain

We relate the performance of GMAC and DCF-Agg to the performance of the regular DCF scheme. A transmission pattern of GMAC, DCF-Agg and DCF is shown in Figure 3.4.

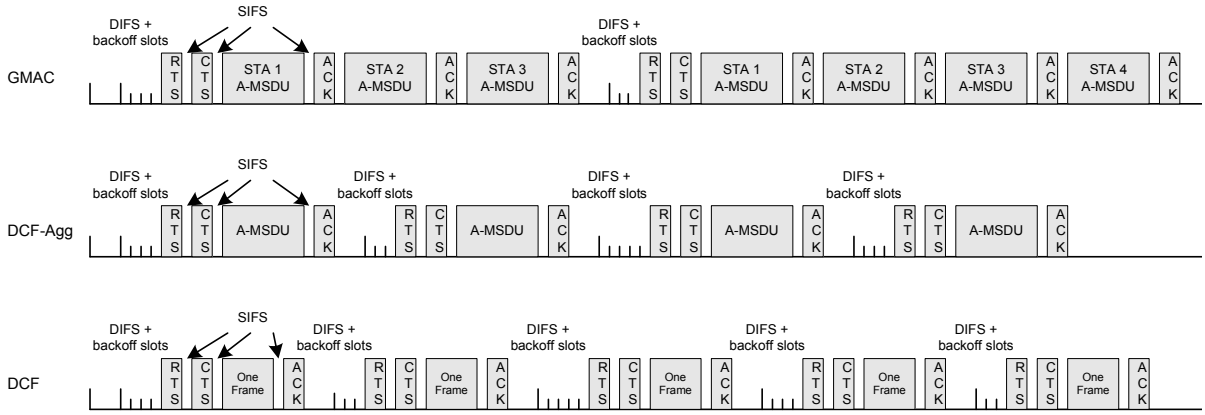

Figure 3.4 The Transmission Pattern of GMAC, DCF-Agg and DCF.

Let $\eta(n, b, T)$ be the number of frames transmitted successfully when n stations contend using DCF over a time duration T and each station transmits an A-MSDU that contains b data frames.

In the regular DCF scheme, the stations transmit one frame upon access. However, in the

DCF-Agg scheme, each station transmits an A-MSDU that contains b frames. The performance gain of DCF-Agg over DCF is given by the following term:

$$\gamma_{\text{DCF}}^{\text{DCF-Agg}} = \frac{\eta(n, b, T)}{\eta(n, 1, T)} \tag{3.4}$$

In this analysis, we consider that there are n stations that are using GMAC. These stations are divided into g groups equally. Every station transmits an A-MSDU with b frames upon access. Thus, the number of contending stations is g and the number of data frames transmitted after each contention is $(n/g).b$. The performance gain of GMAC over DCF is given by the following term:

$$\gamma_{\text{DCF}}^{\text{GMAC}} = \frac{\eta(g, \frac{n}{g}.b, T)}{\eta(n, 1, T)} \tag{3.5}$$

Equations (3.4) and (3.5) indicate that the throughput gain in DCF-Agg comes from reducing the contention overhead per frame while, on the other hand, the performance gain in GMAC comes from reducing the contention overhead per frame and also from reducing the number of collisions.

3.5.2 Time Utilization

The time utilization of the regular DCF scheme was analyzed in (2). It is given as the following. Let the minimum Contention Window be CW_{min} and the maximum backoff stage be m, then the time utilization of the DCF scheme is given by:

$$\mu = \frac{P_s.P_{tr}.T_{payload}}{(1 - P_{tr})\sigma + P_{tr}.P_s.T_s + P_{tr}.(1 - P_s).T_c} \tag{3.6}$$

In this equation, P_{tr} is the probability that a station transmits in a slot and P_s is the conditional probability of a successful transmission in a slot given that at least one station tries to transmit. The term T_s designates the time consumed by a successful transmission. The term T_c designates the time consumed by a collision event and σ designates the duration of a backoff slot.

The expressions for P_{tr} and P_s are given in the following equations.

$$P_{tr} = 1 - (1 - \tau)^n \tag{3.7}$$

$$P_s = \frac{n.\tau.(1 - \tau)^{n-1}}{P_{tr}} \tag{3.8}$$

The expression of τ is given by the following equations.

$$\tau = \frac{2(1 - 2p)}{(1 - 2p)(CW_{min} + 1) + p.CW_{min}(1 - (2p)^m)} \tag{3.9}$$

$$p = 1 - (1 - \tau)^{n-1} \tag{3.10}$$

3.5.3 Applying to GMAC, DCF-Agg and DCF

To use the result above, we need to find T_s and T_c for each of the three schemes. For DCF, we have the following equations where $t_{data(1)}$ is the time to transmit one data frame.

$$T_s^{\text{DCF}} = t_{difs} + t_{slots} + t_{rts} + t_{cts} + t_{data(1)} + t_{ack} + 3.t_{sifs} \tag{3.11}$$

$$T_c^{\text{DCF}} = t_{difs} + t_{slots} + t_{RTS} \tag{3.12}$$

For DCF-Agg, the difference from the two equations above is that we have an A-MSDU of b frames that is transmitted on every access. We have $T_c^{\text{DCF-Agg}}$ is the same as T_c^{DCF} and $T_s^{\text{DCF-Agg}}$ is the following equation where $t_{data(b)}$ is the time to transmit an A-MSDU that contains b data frames.

$$T_s^{\text{DCF-Agg}} = t_{difs} + t_{slots} + t_{rts} + t_{cts} + t_{data(b)} + t_{ack} + 3.t_{sifs} \tag{3.13}$$

For GMAC, after every contention the stations in a group transmit. We assume in the analysis that there are $\frac{n}{g}$ stations in a group. Each one transmits an A-MSDU that contains b frames. There are $\frac{n}{g}$ Block ACK frames transmitted, assuming that no erroneous transmissions occur, and $(\frac{n}{g} + 1)$ intermediate SIFS durations. We have T_c^{GMAC} is the same as T_c^{DCF} and T_s^{GMAC} is the following:

Table 3.3 Performance Comparison of GMAC, DCF-Agg and DCF

n	b	g	$\gamma_{\text{DCF}}^{\text{DCF-Agg}}$	$\gamma_{\text{DCF}}^{\text{GMAC}}$	$\gamma_{\text{DCF-Agg}}^{\text{GMAC}}$
10	5	2	3.00	4.55	1.38
20	5	4	3.03	4.24	1.39
40	5	5	3.07	4.53	1.47
60	5	5	3.10	4.75	1.53
100	5	5	3.15	5.03	1.59

$$T_s^{\text{GMAC}} = t_{difs} + t_{slots} + t_{rts} + t_{cts} + \frac{n}{g}.t_{data(b)} + \frac{n}{g}.t_{ack} + (\frac{n}{g} + 1).t_{sifs} \tag{3.14}$$

The number of frames transmitted in duration T by n contending stations where b frames are aggregated in each access is:

$$\eta(n, b, T) = \frac{\mu.T.r_{avg}}{L} \tag{3.15}$$

In the equation above, r_{avg} is the average data rate and L is the average frame length in bytes.

3.5.4 Analysis Results

Using the results of the analysis, we find the throughput gain of DCF-Agg and GMAC in comparison to DCF, given by the terms $\gamma_{\text{DCF}}^{\text{DCF-Agg}}$ and $\gamma_{\text{DCF}}^{\text{GMAC}}$, respectively. We also show the throughput gain of GMAC over DCF-Agg, indicated by the term $\gamma_{\text{DCF-Agg}}^{\text{GMAC}}$. The results are shown in Table 3.3. For these results, the data rate is 116 Mbps, the control rate is 6.5 Mbps and the frame size is 1000 bytes.

DCF-Agg achieves a throughput gain over DCF by a factor of about 3. This gain is achieved by the transmission of an A-MSDU that contains $b = 5$ data frames upon every access. GMAC achieves a performance gain over DCF by a factor of about 4.5 to 5. This is the result of less contention and the transmission of A-MSDUs with $b = 5$ data frames upon an access. Finally, GMAC achieves a throughput gain of DCF-Agg by a factor of about 1.3 to 1.5. While both GMAC and DCF-Agg benefit from A-MSDU transmissions with $b = 5$ data frames in each, only GMAC benefits from reduced contention.

Table 3.4 Physical Layer Characteristics

Characteristics	Value	Description
Slot Time	9 μs	Contention slot time
RIFS	2 μs	Reduced Inter-Frame Space
SIFS	10 μs	Short Inter-Frame Space
DIFS	28 μs	DCF Inter-Frame Space
CW_{min}	15	Minimum contention window size
CW_{max}	1023	Maximum contention windows size

DCF-Agg is similar to the IEEE 802.11n DCF scheme that is simulated in the next section. The comparison results between DCF-Agg and GMAC are similar to the results that are presented in the next section.

3.6 Simulation Results

In this section, we present the simulation results which compare the 802.11n standard, the proposed scheme GMAC and the schemes from the literature TMAC and E-GDCF.

The results were obtained from our own simulation code which simulates the MAC layer of the WLAN. All of the schemes were tested in the same environment and using the same physical layer. Hence, this ensures that the comparison is fair and focuses on the MAC scheme performance. For TMAC and E-GDCF, we used the parameters that were presented in their original publications as in Section 3.2.

The physical layer characteristics are presented in Table 3.4. These characteristics are used for the 802.11n DCF scheme and for GMAC. For TMAC and E-GDCF, we use the values from this table when applicable, otherwise, we use the values presented in the respective original articles. For example, TMAC and E-GDCF use DIFS which we take from this table. However, TMAC and E-GDCF define their own CW_{min} and CW_{max} values, which we set as defined by the scheme's authors.

The data rates that we use in the simulation are shown in Table 3.5. These rates are defined in the 802.11n standard (30). In the standard, rates are defined for several configurations. The rates in this table correspond for three spatial streams between the sender and the receiver. This means a MIMO-based physical layer is used with three antennae at each end. The

Table 3.5 Data Rates (in Mbps)

Control Frames	Data Frames
6.5	21.7 – 43.3 – 65.0 – 86.7 – 130.0 – 173.3 – 195.0 – 216.7

bandwidth used to support these rates is 20 MHz.

In the simulation results, a station uses one of the data rates for all of its data transmission. All of the control frames are transmitted at the control rate. Each data rate is used by an equal number of stations. For example, when there are 8 stations in the network, each station selects a rate. Similarly, when there are 40 or 120 stations in the network, each rate is used by 5 or 15 stations, respectively. We use a number of stations in the network that is a multiple of eight. As a result, the average rate in the network is the average of all the data rates, which is 116.46 Mbps.

3.6.1 Collision Rate

The results of the collision rate evaluation are shown in Figure 3.5. The simulation time is 1200 seconds and the frame size is 1300 bytes. First, we notice that the collision rate of DCF grows to be the highest. DCF focuses mostly on reducing the overhead (such as RIFS instead of SIFS) and aggregating the frames. However, DCF doesn't aim at reducing the collision rate. Thus, its rate is acceptable with a low number of stations but it can become large when there is a large number of stations.

The collision rate of E-GDCF remains constant when the number of stations changes. It is almost equal to 34.6 percent. The contention in E-GDCF happens within the groups. Each group has a size of two stations and use $CW_{min} = CW_{max} = 8$. As a result, an increase in the number of stations will have no effect on the number of stations that are contending with each other at the same time.

TMAC has a collision rate that is increasing, however, it remains smaller than that of DCF. TMAC uses contention like DCF. But in TMAC, the contention is limited to stations in a group. There are at most $\overline{N}_V = 15$ stations in a group. Thus, at most 15 stations are contending together. The collision rate in TMAC depends on the number of stations in the

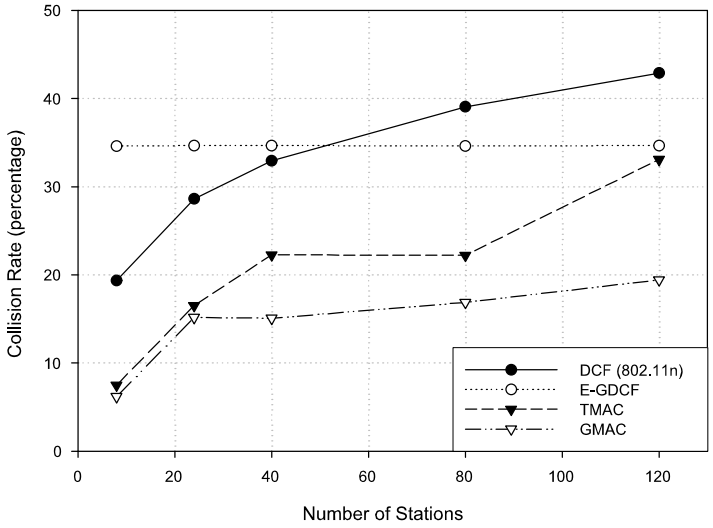

Figure 3.5 Collision Rate.

group. Figure 3.5 shows five instances of the collision rate of TMAC with the number of stations equal to 8, 24, 40, 80 and 120. The stations are distributed to the groups equally. With 8 stations, there is one group with 8 stations contending. With 24 stations, there are two groups with 12 stations each. Thus, the collision rate increases since now there are 12 stations contending together instead of 8. In the third point on the graph, there are 40 stations. Thus, there are three groups; two groups have 13 stations each and one group has 14 stations. In the fourth point on the graph, there are 80 stations that are divided over six groups; four groups have 13 stations and two groups have 14 stations. As a result, the third and fourth points on the graph have 13 or 14 stations in each group, and thus, they have similar collision rates. Finally, with 120 stations, there are six groups with 15 stations each. In this case, the collision rate is the highest since 15 stations contend together.

GMAC achieves the smallest collision rate among the schemes that we compare. This is because in GMAC, only the group leaders contend. The number of group leaders increases when there are more stations, but it doesn't increase a lot. In GMAC, a new station tries to join an existing group. If it cannot, then it starts a new group. This policy encourages the number of groups to remain small. Hence, the number of contending stations will stay small.

In Figure 3.5, when the number of stations is 8, 24, 40, 80 and 120, the corresponding number of groups in GMAC is 2, 3, 3, 5 and 6. Thus, at most six stations are contending at the same time and a small collision rate is achieved. We observed that when the number of stations increases up to 300, the number of groups obtained is about 10 to 12. When there are 10 to 12 group leaders in the WLAN area, a new station will most likely be within a distance of $R/2$ of one of these leaders and it will join their group; so it won't start a new group.

The collision rate results present a comparison between GMAC and TMAC. Both of these schemes use the grouping in similar ways. The contention in TMAC happens within the group. The size of a group is limited to 15. On the other hand, GMAC allows only one station in each group to contend. Since there was less than 15 groups for the scenarios with up to 120 stations, GMAC had a smaller number of stations contending and it achieved a smaller collision rate.

3.6.2 Throughput

The throughput results are shown in Figures 3.6, 3.7 and 3.8. The number of stations in the network for the three figures is 8, 40 and 120, respectively. The simulation time is 1200 seconds. The frame size varies from 300 bytes to the maximum allowed size of 2346 bytes. The stations always have data frames to transmit. DCF and GMAC use a time-based fairness. When a station has access to the channel, the amount of time it is permitted to use is sufficient to transmit one frame at the lowest rate. If a station is able to transmit at a higher data rate, it aggregates multiple MSDUs in an Aggregate-MSDU (A-MSDU) and it receives a Block ACK frame. For TMAC, the time for each station is defined in the scheme. If the station's rate r_i is higher than the reference rate R_f, the station transmits for a duration of $r_i/R_f.T_f$, where $T_f = 35\ ms$. Otherwise, the station transmits for a duration of T_f. We set the reference rate, R_f, for TMAC as 173.3 Mbps since this is considered as a high rate among the rates that we use. In the TMAC article, R_f was set to a high rate among the considered rates. For E-GDCF, each station transmits one frame upon access as specified in E-GDCF's article. Then, a Block ACK is sent by the AP which contains the ACKs to multiple users.

In Figure 3.6, the throughput results for the schemes are shown when the number of stations

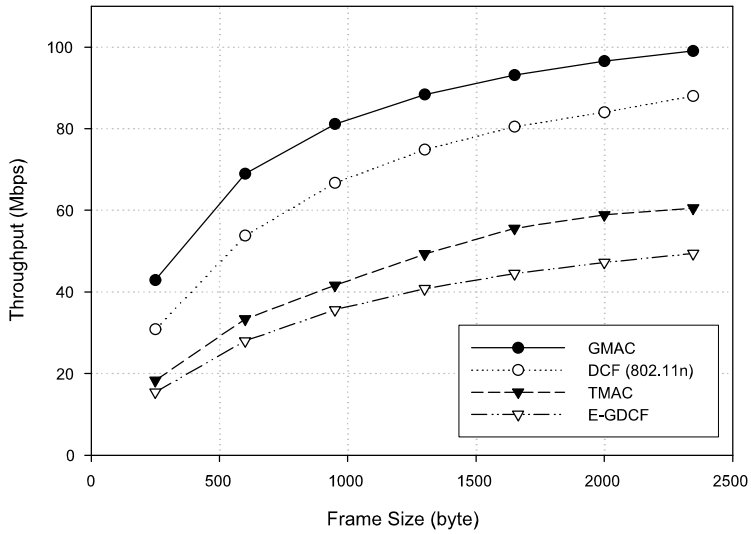

Figure 3.6 Throughput comparison of IEEE 802.11n DCF, GMAC, TMAC and E-GDCF. Number of stations is 8.

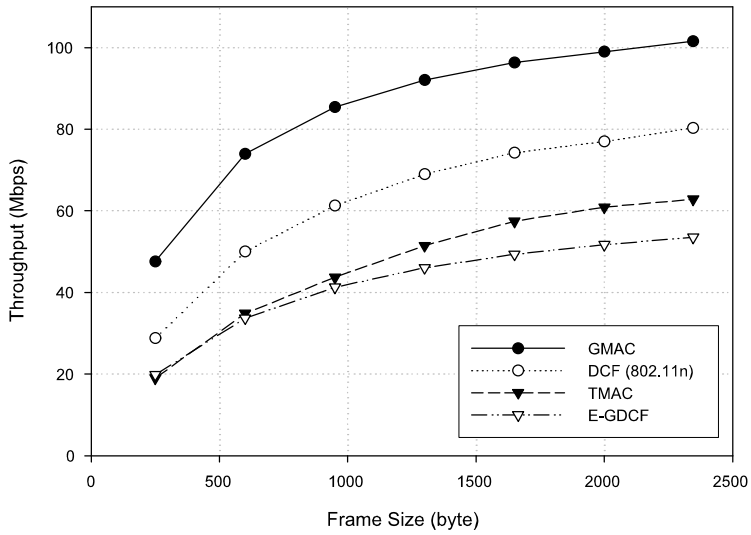

Figure 3.7 Throughput comparison of IEEE 802.11n DCF, GMAC, TMAC and E-GDCF. Number of stations is 40.

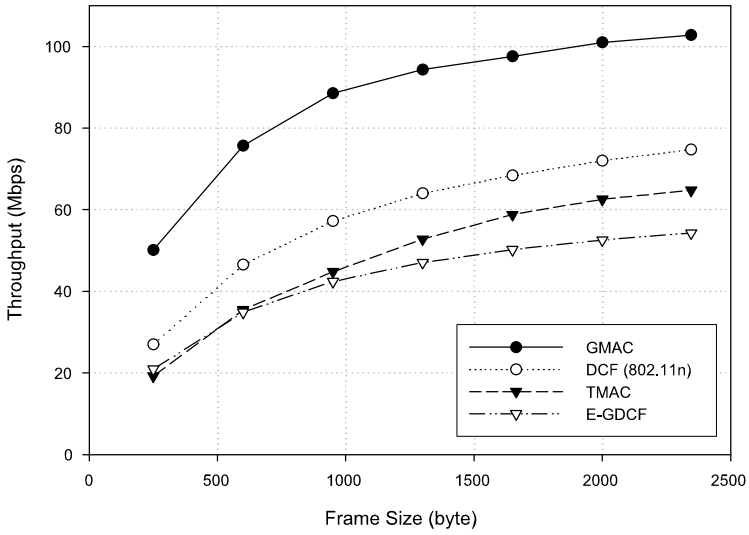

Figure 3.8 Throughput comparison of IEEE 802.11n DCF, GMAC, TMAC and E-GDCF. Number of stations is 120.

is eight. The average data rate in the network is 116.46 Mbps. The throughput that is achieved peaks at about 100 Mbps with GMAC. In the figure, the throughput of GMAC varies from 42 Mbps to 99 Mbps and remains about 10 to 15 Mbps higher than that of DCF. GMAC has a higher throughput since it has fewer collisions than DCF. GMAC also benefits from the mechanisms used by DCF, which are the frame aggregation and using the RIFS instead of SIFS.

The throughput of TMAC ranges from 18 to 60 Mbps and that of E-GDCF ranges from 15 to 49 Mbps. One of the reason that TMAC has a lower throughput than GMAC's and DCF's is that TMAC uses an RTS/CTS exchange prior to every data transmission. When the gap between the control rate (6.5 Mbps) and the average data rate (116.46 Mbps) is very large, the RTS/CTS exchange might not increase the throughput since the RTS and CTS frames take a very long time transmit compared to the data frame. It would be better in such cases to allow the data frame to have a possible collision. E-GDCF has a low throughput for two reasons. First, E-GDCF has a high collision rate, compared to the other schemes, although this collision rate remains stable with more stations in the network. Secondly, the stations

in E-GDCF transmit only one frame upon access to the channel. As a result, there is more channel access overhead for each frame transmitted.

Figures 3.7 and 3.8 show the throughput with 40 and 120 stations, respectively. The same trends are observed as in Figure 3.6. However, there is a significant observation. The throughput of GMAC and E-GDCF increase slightly when the number of stations increase. This is because the grouping-based schemes are able to benefit from a large number of stations. In GMAC a larger group means there are more data frames to transmit for each access overhead. In E-GDCF, a larger number of groups means there are less DIFS spaces used (since one DIFS is used for every grouping cycle). However, in DCF, a larger number of stations means there are more collision. This drops the throughput.

3.6.3 Delay

In this part, we show the results of the delay performance. The simulation time is 1200 seconds and the frame size is 1300 bytes. There are 40 stations in the network.

Table 3.6 shows the average delay and the standard deviation on the delay. The delay measures the time when the frame reaches the head of the queue to the time it is transmitted successfully. The smallest delay average was obtained by E-GDCF. The reason for this small delay is that the stations in E-GDCF transmit only one data frame upon access. With the other schemes, an A-MSDU with multiple data frames is transmitted for stations with high data rates. Transmitting A-MSDUs allowed the other schemes to obtain a higher throughput. E-GDCF, which obtained a lower throughput, did better in the delay measurement. The average delay of GMAC, DCF and TMAC is also considered to be small and allows supporting real-time applications.

The standard deviation on the delay of E-GDCF is also the smallest for the same reason that E-GDCF doesn't use frame aggregation. TMAC has a smaller standard deviation on the delay than GMAC since the rotation between groups in TMAC is done with the token. All the groups get access predictably in a round robin way. TMAC has more collisions, thus its average delay is larger than that of GMAC. However, the consistent rotation between the

Table 3.6 Delay Average and Standard Deviation (in Milliseconds)

	GMAC	DCF	TMAC	E-GDCF
Average	23.69	31.34	56.12	8.53
Standard Deviation	70.75	153.05	40.58	6.31

groups makes the delay variation small. While GMAC achieves a smaller average delay, the groups in GMAC get access through contention. Thus, if one group leader doesn't get access for a while, it means all of the stations in the group don't get access. Hence, GMAC has a higher delay variation. However, its delay fall within the acceptable range of supporting real-time traffic. Finally, the average delay of DCF is in the middle between GMAC and TMAC. But the delay variation of DCF is one of the largest experienced among the schemes. This is mostly due to the contention window mechanisms. In DCF, all the stations contend and the contention window could reach the maximum value of 1023. This event would produce a large delay.

3.6.4 Fairness

In this part, we show the results for the fairness measurements. We use Jain's Index (28) to find the fairness between the stations. The value of Jain's index ranges from 0 to 1. The best fairness scenario obtains a value of 1 while the worst case obtains a value of 0. First, we present fairness results based on the number of frames that is transmitted by each station. This is the measurement of the throughput-based fairness. Secondly, we present the fairness results based on the number of Transmission Opportunity (TXOP) obtained by each station. For GMAC and DCF, each station uses the same amount of time upon access. Then, the fairness on the number of TXOPs is the time-based fairness measurement.

In the fairness evaluation, we do a simulation time of 60, 300 and 1200 seconds to show how the fairness evolves with time. The number of stations in the network is 40 and the frame size is 1300 bytes. The results are shown in Figures 3.9 and 3.10.

Figure 3.9 shows the fairness on the number of frames that is transmitted by each station. This is the throughput fairness between the stations. The fairness of GMAC and DCF is

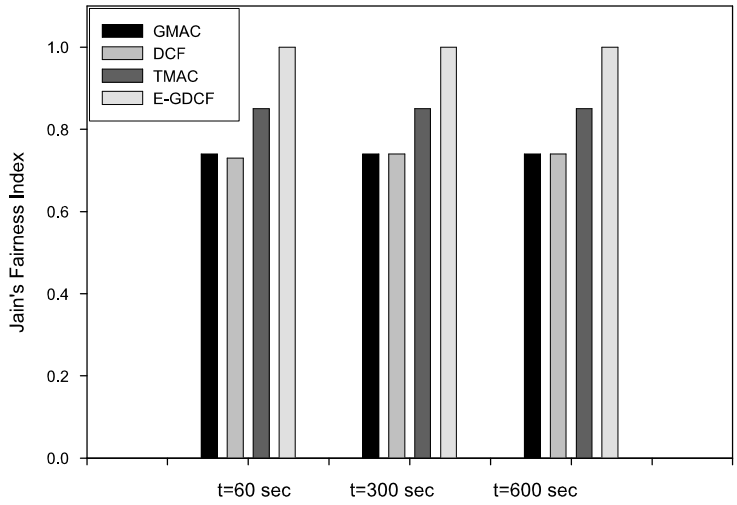

Figure 3.9 Jain's Fairness Index on the Number of Frames Transmitted by each Station.

around 0.74 for all the values of the simulation time. GMAC and DCF obtained a lower fairness than the other schemes in Figure 3.9 since we used time-based fairness (not throughput-based fairness) for these schemes. The time-based fairness increases the total throughput significantly when there are some stations with low rates, as we presented in Section 3.3 and as shown in (61) and (62). Accordingly, the significant measurement of fairness for DCF and GMAC is by measuring the number of times each station has accessed the medium. This will show if each of the stations has transmitted for the same amount of time. This will be presented later in this section.

From Figure 3.9, TMAC achieves a fairness index of 0.85 throughout all the values of the simulation time. In TMAC, the token is rotated among the groups in a round robin way. Thus, all the groups will access the medium for the same number of times. TMAC recognizes that having stations with low rates will decrease the total throughput of the system. Thus, TMAC allows stations with high rates to transmit more. In TMAC, a station i that can transmit at rate r_i that is higher than the reference rate, R_f, will transmit for a duration of $r_i/R_f.T_f$, where T_f is a pre-defined reference duration. A station i with a rate r_i that is smaller than R_f will transmit for T_f.

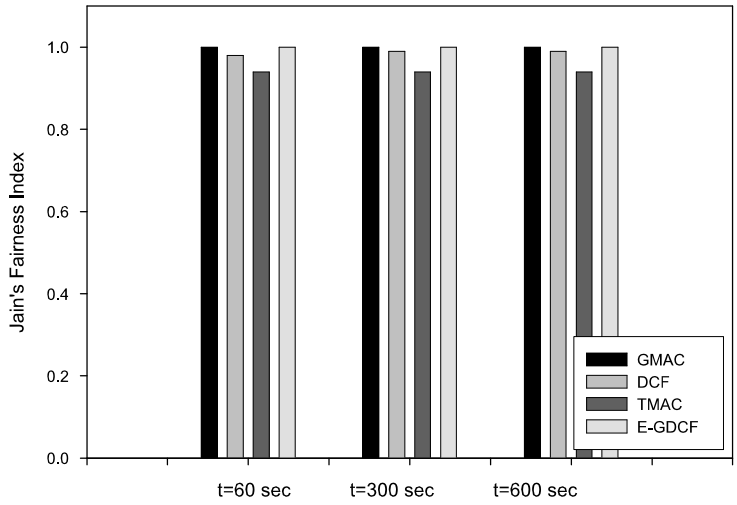

Figure 3.10 Jain's Fairness Index on the Number of TXOPs Obtained by each Station.

As shown in Figure 3.9, E-GDCF achieves a fairness index of 1 for all the values of the simulation time. E-GDCF has a high throughput-based fairness since each station transmits one frame upon accessing the medium. Each of the stations also has accessed the medium for the same number of times. The group has taken turns, one-by-one, in the transmission. Thus, there is fairness between the groups. Within a group, there are at most two stations that contend to the channel using the DCF scheme. The DCF scheme is fair in the long-term. E-GDCF provided a high level of throughput-based fairness. However, it provided the lowest throughput among the schemes that we compared. This agrees with the results that we showed in Section 3.3. The throughput-based fairness provides a low total throughput when there are stations with low rates in the network.

In Figure 3.10, the fairness on the number of TXOPs for each station is measured. GMAC provides a fairness index of 1 for all the values of the simulation time. Accordingly, all the stations in GMAC has accessed the medium for the same number of times. Also, each of the stations has transmitted for the same duration of time. Then, GMAC has achieved a high level of time-based fairness.

DCF has a fairness index of 0.98 when the simulation time is 60 seconds and a fairness

index of 0.99 when the simulation time is 600 or 1200 seconds. This is a high fairness and it means that each of the stations has accessed the medium for almost the same number of times. It also means that each of the stations has transmitted for almost the same duration since the time-based fairness is used with DCF.

In Figure 3.10, the fairness of TMAC is 0.94 for all the values of the simulation time. Thus, TMAC provides more time-based fairness than throughput-based fairness.

Finally, E-GDCF provides a fairness index of 1 on the number of times each station has transmitted. This means that each station in GMAC has transmitted for the same number of times. For GMAC, this doesn't mean that there is time-based fairness since each station transmitted one frame at its own rates. Since the stations has varying rates, the stations with low rates took more time to transmit and has used more time than the stations with high rates.

3.7 Conclusion

This article presented the Group-based Medium Access Control (GMAC) scheme for WLANs with high data rates. GMAC is based on the idea of grouping the contending stations to reduce the number of collisions. GMAC also uses MAC mechanisms to reduce the overhead, such as using smaller Inter-Frame Spaces and using frame aggregation. We presented the details of the proposed scheme and provided an analysis that characterizes the throughput gain that is obtained with GMAC and with the standard's DCF with frame aggregation. In the simulation results, we considered other MAC scheme from the literature that were designed for WLANs with high data rates. The results showed that GMAC provides a significantly higher throughput while maintaining a low collision rate. The simulation results also presented the delay and fairness characteristics of the schemes.

CHAPTER 4. Fault-Tolerant Planning of Relay Station Locations in IEEE 802.16j (WiMAX) Networks

An article that is submitted to *IEEE Transactions on Mobile Computing.*[1]

Zakhia Abichar[2], Ahmed E. Kamal[3] and J. Morris Chang[4]

4.1 Introduction

After the success and wide deployment of Wireless Local Area Networks (WLAN) (65), the area of wireless networks has witnessed the standardization process for broadband wireless access networks. The IEEE 802.16 standards provide last-mile connectivity and they have been touted to fill several needs: last-mile end user access, initial deployment of infrastructure in unwired areas and providing access to mobile users.

The networks running on the IEEE 802.16 standards are known as WiMAX (Worldwide Interoperability for Microwave Access). One of the 802.16 standards, developed by Task Group j, and recently standardized is the IEEE 802.16j Standard (66). The goal of 802.16j is to support the use of Relay Stations (RS). The main stations in a WiMAX network are the Base Station (BS) and the Subscriber Stations (SS), alternatively called Mobile Stations (MS) when the device is mobile.

The goal of an RS is to support the connectivity between the BS, on one side, and the SSs and MSs, on the other side. The RS can extend the range of a BS. For example, there could be users that are out of reach of the BS and cannot connect to the network. With the

[2]Primary researcher and author

[3]Faculty advisor

[4]Faculty advisor

placement of an RS between the user and the BS, the user would be able to connect; hence, the range of the BS is extended. The RS can be also used to enhance the capacity of the BS. For example, even if all the users are in range with the BS, placing one or more RSs in the cell allows higher data rates and enhancing the cell's capacity as a result. An introduction to the 802.16j standard is provided in (67), (68).

The 802.16j standard, however, doesn't specify how the RSs should be placed in the network. This issue remains open to research. It is the goal of this paper to devise a technique for planning the RS locations in a WiMAX network.

4.1.1 Motivation

In this paper, we consider the problem of placing several RSs to support a BS in order to provide a resilient operation of the WiMAX network. In real life, the users expect a reliable service. Especially, there are some users, like businesses, who rely on the Internet connection in order for them to function. If the network is planned with no fault tolerance, an RS failure might shut down totally the connection to some users.

We consider that only one RS might fail at a certain time. It could be any RS among the used RSs. With an adequate level of service, the RS should be fixed before another RS fails. This assumption (allowing *only one* RS to fail) also allows keeping the cost of the system reasonable.

In order to provide fault tolerance, we define for each user a full rate and a backup rate. When there is no failure among the RSs, the users get the full rate. However, in the case of a failure, we consider that a reduced rate to the user is better than no service at all. Thus, in the case of an RS failure, the users get the backup rate. This definition also allows users who primarily depend on the Internet for business to have a backup rate that is equal to the full rate. Thus, these users will function without service degradation even in the case of an RS failure.

The input to our problem is the location of the BS and the location of the users (SSs and MSs) and their respective demands represented by the bit rate. Customers are represented by

Test Points (TP). For example, if there are several houses located nearby of each other with demands of 2, 3 and 5 Mbps, they could be represented by a TP (located in a centric point to the houses) with a demand equal to the SS demands of 10 Mbps.

The planning solution we present aims at placing the RSs in the network to achieve several goals. These goals are:

1. All the service area should be covered with connection to the network. The area is defined through the TPs; thus by providing connectivity between all the TPs and the BS, the service area will be covered.

2. The throughput demand of all the TPs should be satisfied. There should be a connection between the TP and the BS with a flow equal to the pre-defined demand of the corresponding TP.

3. The number of RSs placed by our solution should be minimized in order to reduce the equipment, installation and operation cost.

4. We also provision for the case when an RS in the network might fail. In such a case, we define an adequate level of service, which we call the backup service rate, that should be provided and plan the RSs in a way to guarantee this service. Thus, our planning method is fault-tolerant.

4.1.2 Contribution

In a previous contribution (69), we provided a planning solution that satisfies only the first three goals listed above. There was no fault-tolerance provision in our initial approach. Thus, if an RS fails, there was no guarantee that the level of service provided would be adequate to the subscribers. Hence, in this paper, we extend our approach to incorporate fault-tolerance and ensure that users are served adequately in the case of an RS failure.

There are several papers in the literature that address the problem of placing RSs in the WiMAX network. These are reviewed in the next section. However, to the best of our knowledge, there is no work on planning the locations of RSs in WiMAX networks with fault-

tolerance. There have been some approaches on placing relays in a fault-tolerant manner in other types of networks such as wireless sensor networks.

We formulate the RS planning problem using a Mixed Integer Linear Program (MILP). We present numerical results by solving our model with CPLEX v12.2. We believe that solving the model directly to obtain results is a valid approach since a limited number of RSs will be installed in real-life to serve one BS and also planning is not a real-time operation. By using CPLEX v12.2, which runs on a modern multi-core machine, we obtained solution times for realistic scenarios that are under 10 seconds. Thus, the computation time is reasonable for practical cases.

We present numerical results that show how our model finds the number and locations of RSs. Our model also specifies all the links that are used and gives the rate on each link. Also, for every RS that is used in the main topology (used when no RS is in a failure condition), the model gives a corresponding backup topology in case this RS fails.

Our model can also consider topological conditions such as the existence of obstacles in the coverage area that might occur in real-life. For example, there might be a lake in the coverage area that would prohibit the installation of an RS in the lake. However, the wireless link is able to go over the lake. Alternatively, there might be a small hill or a large building in the coverage area. A hill with a rough terrain would prohibit installing an RS on it and it would also block a wireless link from traversing it. Our planning solution can accommodate both types of these obstacles.

The rest of this paper is organized as follows. Section 4.2 presents the related work and Section 4.3 presents the network model. The optimization model with fault-tolerance is provided in Section 4.4. Finally, the numerical results are given in Section 4.5 and the conclusions are given in Section 4.6.

4.2 Related Work

This section presents earlier work in the literature that is related to the problem of planning the RS locations with fault-tolerance.

4.2.1 Planning RS Location in WiMAX Networks

Our previous work in (69) presents a model for planning the RS locations in a WiMAX networks. However, in the previous work, there is no guarantee of service if an RS fails since fault-tolerance was not considered. Therefore, in this paper, we extend our model to provide resiliency to the network.

A planning model is presented in (70) to find the locations of BSs and RSs in the network. The model is formulated as an optimization problem. In this model, there is at most one RS between the SS/MS and the BS. A maximum of two hops is allowed. Since the standard doesn't have a limit on the number of hops going through the RSs, this assumption may impose unneeded restrictions.

The same authors of the work above present an extension of their work in (71). In this paper, they consider a large coverage area that makes the computation time of the model very large. They divide the area into clusters and apply the approach above to every cluster. Then, the cases on the boundaries of the clusters are solved to find the overall solution. This paper similarly limits the number of hops to two.

In (72), an RS placement model is presented. This work is based on cooperative transmission between the source node and the relay node to provide a better signal to the destination node. They consider the decode-and-forward scheme and the compress-and-forward scheme for cooperative transmission. This model is different from our work since it considers the placement of a single RS to serve multiple SSs. The authors present another RS placement solution in (73), where the cooperative transmission paradigm is used in multi-hop relaying for the purpose of range extension. Also, the same authors present in (74) and (75) an RS placement solution that uses the cooperative transmission technique for the purpose of capacity enhancement.

In (76), a relay placement algorithm is presented based on the concept of dual-relaying. In this concept, every SS or MS connects to the BS always by using two RSs. The SS transmits to the two RSs at the same time; then, the RSs forward the data to the BS in a cooperative way. The focus in this approach is on making benefit of the diversity gain by using the links

that go through the dual relay.

In (77), a model is presented to find the locations of RSs that extend the range of a BS in a WiMAX network. This work defines preset topologies and finds the RS placement for these topologies; in comparison, our model in this paper and in (69) can work with any topology. This work also considers RS location planning for sector-based topology. Each sector uses a frequency that is different from adjacent sectors in order to reduce the interference.

In (78), the problem of joint BS and RS deployment is considered and an optimization model is presented. Due to the large size of the problem, the model consumes a large amount of time to solve. Thus, the authors also present an approximation algorithm.

In (79), the problem of RS placement in the WiMAX network is considered. The location of the RSs and the bandwidth allocation to users are found. This work assumes that users' demands could change due to fluctuation in traffic demand and due to mobility. Thus, the optimization of the RS locations is found on a long-term basis and the bandwidth allocation to users is found on a short-term basis.

The authors of (80) consider using relays for the purpose of capacity enhancements as follows. There is a BS, an area that can be totally covered by the BS and a given number of relays. This work decides where to place the relays to maximize the system capacity.

In (81), the following paradigm is considered for the placement of RSs in WiMAX networks. The number and locations of BSs are given. The goal of the problem is to place RSs that use the transparent mode. In this mode, the RS don't transmit control information; the control information are only transmitted by the BS. The RSs are thus in range of the BS and the goal of the RS placement is capacity enhancement. Another approach by the same authors based on the transparent RS mode is presented in (82). The goal here is also capacity enhancement. However, this work aims at finding the locations of BSs and RSs.

4.2.2 Planning Relay Locations with Fault-Tolerance

There are approaches in the literature that provide relay location planning with fault-tolerance. But these works are not for WiMAX networks.

In (83), a Integer Linear Program (ILP) model is presented for placing relays in sensor networks to provide fault-tolerance in case some nodes fail. The main issue was connectivity, regardless of bandwidth requirements, which implies that all relay nodes may be operative all the time. The same authors present an extension of their work in (84), which takes into consideration the routing strategy in order to reduce the battery consumption.

Other approaches on fault-tolerant placement of relay nodes in sensor networks are in (85), (86), (87), (88) and (89).

The approaches presented in this section consider a multitude of issues and configurations for the RS use in WiMAX networks. There are multiple approaches on planning RS placement in WiMAX networks. There are also approaches that plan with fault-tolerance the locations of relay nodes, although not in WiMAX networks. To the best of our knowledge, there is no approach that provides the planning of RS locations in WiMAX network with fault-tolerance. Hence, our paper is the first to propose such an approach.

4.3 Network Model

This section presents the network model that we consider in this paper.

4.3.1 Relay Station Non-Transparent Mode

The IEEE 802.16j Standard presents a transparent mode and a non-transparent mode for the RS operation.

In transparent mode, the users (SSs and MSs) don't know of the presence of an RS. The RS doesn't transmit control information (such as downlink-map and uplink-map). These are transmitted by the BS. Thus, all the SSs and MSs are in range with the BS. However, the RSs are used in the transparent mode for the purpose of capacity enhancement.

In the non-transparent mode, the RS transmits control information to the SSs it serves. Multihop routes are allowed in the non-transparent mode. The goal for using non-transparent RSs is to extend the range of the network and also to enhance the capacity. In this paper, we use the non-transparent mode in order to achieve these two goals.

4.3.2 Duplexing Mode

When multi-hop RSs are used, the frames of two stations that are in range should be duplexed either in the time domain or in the frequency domain in order to avoid interference.

The IEEE 802.16j Standard (66) allows using different frequencies for RSs serving the same BS. Thus, we make the assumption that the RSs duplex their transmission in Frequency Division Duplex (FDD) mode. For example, on a two-hop route BS - RS - SS, we can have a transmission of rate r on the BS-RS hop and another transmission of the same rate on the RS-SS hop. This happens if the two hops are using different frequencies. With Time Division Duplex (TDD), the two hops will alternate in transmission. However, each hop will have a larger bandwidth since the bandwidth is not divided anymore. We use FDD for simplicity, but our model can be modified easily to work with TDD.

4.3.3 Link Capacity

Our model allocates a rate on each link that is used in the produced topology. The allocated rate on a link is bounded by the maximum capacity of the link.

The maximum capacity of a wireless link can be modeled with the Shannon-Hartley equation as given in (90). It is given by the equation: $C = B.\log_2(1 + \frac{S}{N})$, where C is the capacity in bit/sec, B is the channel bandwidth in Hz, S is the received signal power and N is the noise (S/N is the signal-to-noise ratio or SNR). The capacity changes with the distance since the SNR degrades when the distance increases.

Other factors also affect the link capacity such as the coding and modulation schemes. When a high SNR is measured on the link, coding and modulation schemes with high rates are used. However, when the SNR is low, robust coding and modulation schemes are preferred to limit the Bit-Error-Rate (BER), although they provide low data rates. Table 4.1 shows the achievable bit rates for the OFDMA physical layer as given in the standard (91). QPSK is more robust but achieves a small rate. On the other hand, 64-QAM is less robust but achieves a high rate.

The factors that affect a link's capacity can be combined in an equation. For a link from

Table 4.1 OFDMA Rates (in Mbps) for Various Modulation Schemes using 7 MHz Bandwidth

QPSK 1/2	QPSK 3/4	16-QAM 1/2	16-QAM 3/4	64-QAM 2/3	64-QAM 3/4
5.82	8.73	11.64	17.45	23.27	26.18

node i to node j, the maximum rate is: $m_{i,j} = \Gamma(SNR_{i,j},\ \beta,\ Cod,\ Mod)$, where β is the upperbound on the BER, Cod is the coding scheme and Mod is the modulation scheme. Γ is the function that maps all the three parameters to the maximum rate.

Any definition of the function Γ can work with our model. However, for simplicity, we assume that the maximum rate changes with distance. In real-life scenarios, there is usually a field survey which precedes the network deployment (92), (93) and (94). The link rates are selected based on the links characteristics such as, the SINR (Signal-to-Interference-plus-Noise-Ratio), fading, the specifics of the terrain and interference with other wireless systems.

4.3.4 Definition of Fault-Tolerance

The planning model we present in this paper allows the failure of an RS without interrupting service to the users, albeit at a reduced bit rate, hence tolerating equipment failure.

We assume that only one RS will fail at a given time. This is a reasonable assumption since usually in the time it takes the RS to be repaired, there is a very small probability that another RS will fail. This is true since the number of RSs supporting a BS cannot be very large. This assumption will keep the cost of RSs small, since tolerating the failure of two or more RSs at the same time, we would require installing many extra RSs. This might not be a cost effective approach.

For every set of customers, represented by a Test Point (TP), a tuple $\{r_i, rb_i\}$ defines the requested service rates. When all the RSs are operational, the full rate or a TP i, given by r_i is provided. However, when there is an RS failure, a reduced rate which is the backup rate, rb_i is provided. We have $rb_i \leq r_i$. Users who request the same service rate even in the case of an RS failure will have $rb_i = r_i$.

4.4 Optimization Model

This section presents the optimization model for the RS planning problem with fault-tolerance. The model takes as input (1) the possible sites where an RS can be installed, (2) the location of the Test Points (TP) that represent the users' traffic, (3) the rates (full and reduced) in Mbps of each TP; the full rate is provided when all the RSs are operational and the reduced rate is guaranteed when there is an RS failure. (4) Finally, the model takes as an input the maximum rate on any link: BS-to-RS, BS-to-TP, RS-to-RS and RS-to-TP, which depends on the link characteristics such as distance, SNR and bandwidth.

The output of our model is the full-rate (main) topology and the reduced-rate (backup) topologies. The full-rate topology is defined by the number of RSs used, their positions, the links used, the rate on each link and, finally, the connection node for each TP (either the BS or an RS). Each of the backup topologies corresponds to a failure in one of the RSs used in the main topology. For example, if the main topology uses RS 1, RS 3 and RS 8, then there will be three backup topologies that are used in the case any of these RSs fail.

For any TP, TP_i, its full rate is designated by r_i and its reduced rate is designated by rb_i (rb stands for backup rate), which is the minimum acceptable rate in the case of failure.

Let $R = \{RS_0, \ldots, RS_{n-1}\}$ be the set of candidate sites for RS with cardinality $|R| = n$. Similarly, let $T = \{TP_0, \ldots, TP_{m-1}\}$ be the set of Test Points (TP) that represent the user traffic with cardinality $|T| = m$.

4.4.1 Decision Variables

The following decision variables define the full-rate topology.

$$dR_i = \begin{cases} 1;\ \text{an RS is deployed in site RS i},\ (i \in R) \\ 0;\ \text{otherwise} \end{cases}$$

$$dBR_i = \begin{cases} 1;\ \text{a link is used between the BS and RS } i,\ (i \in R) \\ 0;\ \text{otherwise} \end{cases}$$

$$dBT_i = \begin{cases} 1;\ \text{TP } i \text{ is assigned to the BS},\ (i \in T) \\ 0;\ \text{otherwise} \end{cases}$$

$$dRR_{ij} = \begin{cases} 1;\ \text{a link is used between RS } i \text{ and RS } j,\ (i, j \in R) \\ 0;\ \text{otherwise} \end{cases}$$

$$dRT_{ij} = \begin{cases} 1;\ \text{TP } j \text{ is assigned to RS } i,\ (i \in R,\ j \in T) \\ 0;\ \text{otherwise} \end{cases}$$

We also define variables that are similar to the above in order to specify the backup topologies. These variables are: dR_i^k, dBR_i^k, dBT_i^k, dRR_{ij}^k and dRT_{ij}^k. The term k indicates the RS that has failed. For example, when $k = 3$, these variables define the backup topology that is used when RS 3 fails.

We also define decision variables that designate the assigned flow (in Mbps) on each link. While the previous variables were binary, the flow variables take integer values. We use integer values since the rate used on a link is usually one of a subset of the supported rates by the physical (PHY) layer. Thus, the rate is selected from a given set of rates.

In the full-rate topology, the variables fBR_i, and fRR_{ij} designate the flow on the links from BS-to-RS, and RS-to-RS, respectively, where i and j are indices of RSs ($i, j \in R$).

Similarly, in the backup topology, the variables fBR_i^k and fRR_{ij}^k designate the flow on the links from BS-to-RS and RS-to-RS, respectively, where i, j and k are indices of RSs ($i, j, k \in R$).

4.4.2 Topology Constraints

The following constraints define the topology of the RS domain. They ensure that when a link is used in the solution, the two end nodes of the link exist (i.e., the RSs are selected). They also ensure that a TP is connected either directly to the BS or to only one RS; we use this condition for simplicity of problem formulation and for simplicity of network operation.

First, when there is a link between the BS and RS i, there should be an RS deployed at site RS i. This is ensured by the following constraints in the full-rate and the backup topologies.

$$dBR_i \leq dR_i \ ; \qquad \forall i \in R \tag{4.1}$$

$$dBR_i^k \leq dR_i^k \ ; \qquad k \neq i, \ \forall i, k \in R \tag{4.2}$$

When there is a link between RS i and RS j, two RSs should be installed at sites RS i and RS j. This is ensured by the following constraints.

$$dRR_{ij} \leq \frac{dR_i + dR_j}{2} \ ; \qquad \forall i, j \in R \tag{4.3}$$

$$dRR_{ij}^k \leq \frac{dR_i^k + dR_j^k}{2} \ ; \qquad i \neq k, \ j \neq k, \ \forall i, j, k \in R \tag{4.4}$$

When there is a link between RS i and TP j, an RS should be deployed at site RS i. This is ensured by the following constraints.

$$dRT_{ij} \leq dR_i \ ; \qquad \forall i \in R, \ \forall j \in T \tag{4.5}$$

$$dRT_{ij}^k \leq dR_i^k \ ; \qquad i \neq k, \ \forall i, k \in R, \ \forall j \in T \tag{4.6}$$

The following constraints send all the traffic of a TP either through a direct link with the BS or through a single TP.

$$dBT_i + \sum_{j \in R} dRT_{ji} = 1 \ ; \qquad \forall i \in T \tag{4.7}$$

$$dBT_i^k + \sum_{j \in R, \ j \neq k} dRT_{ji}^k = 1 \ ; \qquad \forall i \in T \tag{4.8}$$

4.4.3 Flow Constraints

The flow constraints ensure that the amount of data that is transported is balanced and sufficient for the demands of all the TPs.

4.4.3.1 Flow Balance at the BS

In the main topology, the total traffic going out of the BS should be equal to the sum of the full rates, r_i, of all the TPs. This condition is ensured by the following equation.

$$\sum_{i \in R} fBR_i.dBR_i + \sum_{j \in T,\ mBT_j \geq r_j} r_j.dBT_j = \sum_{j \in T} r_j \tag{4.9}$$

At a backup topology, the rate provided to TP i is greater or equal than rb_i. Then, this condition is used.

$$\sum_{i \in R,\ i \neq k} fBR_i^k.dBR_i^k + \sum_{j \in T,\ mBT_j \geq rb_j} rb_j.dBT_j^k = \sum_{j \in T} rb_j\ ; \qquad \forall k \in R \tag{4.10}$$

We are interested in keeping the system linear. Thus, we use the following transform and substitute in Equation (4.9).

$$X_i = fBR_i.dBR_i$$

Equation (4.9) becomes:

$$\sum_{i \in R} X_i + \sum_{j \in T,\ mBT_j \geq r_j} r_j.dBT_j = \sum_{j \in T} r_j \tag{4.11}$$

X_i can be evaluated using the following set of linear constraints, where Q is a large number such that $Q\ > max(fBR_i),\ i \in R$.

$$X_i \geq Q.dBR_i - Q + fBR_i \qquad i \in R \tag{4.12}$$

$$X_i \leq fBR_i \qquad i \in R \tag{4.13}$$

$$X_i \geq 0 \qquad i \in R \tag{4.14}$$

$$X_i \leq Q.dBR_i \qquad i \in R \tag{4.15}$$

Similarly, we use the following transform for Equation (4.10).

$$X_i^k = fBR_i^k.dBR_i^k \ ; \qquad i \neq k, \ \forall i,k \in R \tag{4.16}$$

Then, Equation (4.10) becomes the following equation. X_i^k is evaluated like X_i was evaluated in Equation (4.12) to Equation (4.15).

$$\sum_{i \in R, \ i \neq k} X_i^k + \sum_{j \in T, \ mBT_j \geq rb_j} rb_j.dBT_j^k = \sum_{j \in T} rb_j \ ; \qquad \forall k \in R \tag{4.17}$$

4.4.3.2 Flow Balance at an RS

At an RS, the amount of traffic that is coming from the BS and from upstream RSs is equal to the amount of traffic that is going to downstream RSs and to TPs that are directly connected to the RS. This is ensured by the following constraint.

$$fBR_i.dBR_i + \sum_{j \in R} fRR_{ji}.dRR_{ji} = \sum_{j \in R} fRR_{ij}.dRR_{ij} + \sum_{y \in T, \ mRT_{iy} \geq r_y} r_y.dRT_{iy} \ ; \quad \forall i \in R \tag{4.18}$$

The equation above is made linear by using the transform $Z_{ij} = fRR_{ij}.dRR_{ij}$ and becomes the following. Z_{ij} is evaluated like X_i was evaluated in Equation (4.12) to Equation (4.15).

$$X_i + \sum_{j \in R} Z_{ji} = \sum_{j \in R} Z_{ij} + \sum_{y \in T, \ mRT_{iy} \geq r_y} r_y.dRT_{iy} \ ; \qquad \forall i \in R \tag{4.19}$$

For the backup topologies, the flow conservation at the RS is ensured by the following equation.

$$fBR_i^k.dBR_i^k + \sum_{j\in R,\ j\neq k} fRR_{ji}^k.dRR_{ji}^k = \sum_{j\in R,\ j\neq k} fRR_{ij}^k.dRR_{ij}^k + \sum_{y\in T,\ mRT_{iy}\geq rb_y} rb_y.dRT_{iy}^k$$
$$i \neq k,\ \forall i,k \in R \quad (4.20)$$

The equation above is made linear by using the transform $Z_{ij}^k = fRR_{ij}^k.dRR_{ij}^k$ and it becomes the following equation. Z_{ij}^k is evaluated similar to how X_i was evaluated in Equation (4.12) to Equation (4.15).

$$X_i^k + \sum_{j\in R,\ j\neq k} Z_{ji}^k = \sum_{j\in R,\ j\neq k} Z_{ij}^k + \sum_{y\in T,\ mRT_{iy}\geq rb_y} rb_y.dRT_{iy}^k\ ; \qquad i \neq k,\ \forall i,k \in R \qquad (4.21)$$

4.4.3.3 Flow Balance at a TP

In the main topology, the amount of traffic going from a TP to the BS or to an RS should be equal to the full-rate, r_i of the TP. This is ensured by the following constraint.

$$\sum_{i=i,\ mBT_i\geq r_i}^{i} r_i.dBT_i + \sum_{j\in R,\ mRT_{ji}\geq r_i} r_i.dRT_{ji} = r_i\ ; \qquad \forall i \in T \qquad (4.22)$$

For the backup topologies, the amount of traffic at the TP should be equal to rb_i. This is ensured by the following constraint.

$$\sum_{i=i,\ mBT_i\geq rb_i}^{i} rb_i.dBT_i^k + \sum_{j\in R,\ j\neq k,\ mRT_{ji}\geq rb_i} rb_i.dRT_{ji}^k = rb_i\ ; \qquad \forall i \in T,\ \forall k \in R \qquad (4.23)$$

4.4.4 Constraint on the Link Capacity

The maximum flow that can be transmitted on a link is limited by the power transmitted, the link distance and the coding and modulation schemes. The maximum rate that can be assigned on the link from the BS to RS i, fBR_i, is limited by $fBR_i \leq mBR_i$, where mBR_i

is the maximum rate on this link. A similar notation is used for all the other links. The constraints that ensure the upperbound are the following.

$$fBR_i \leq mBR_i \qquad i \in R$$

$$fRR_{ij} \leq mRR_{ij} \qquad i,j \in R$$

$$fBR_i^k \leq mBR_i \qquad i \neq k,\ i,k \in R$$

$$fRR_{ij}^k \leq mRR_{ij} \qquad i \neq k,\ j \neq k,\ i,j,k \in R$$

The above upper bounds (mBR_i and mRR_{ij}) are input parameters to the problem. Also, when an RS fails, the rate on all the links incident on it is zero. Thus, $fBR_i^k = 0$ when $i = k$, $fRR_{ij}^k = 0$ when $i = k$ or $j = k$.

4.4.5 Objective Function

The primary objective of our solution is to minimize the total number of RSs used. This will minimize the cost of RS installation. We define the term S_i which indicate if an RS is installed at site RS i either in the main topology or in any other topology.

$$S_i \geq dR_i\ ; \qquad i \in R \tag{4.24}$$

$$S_i \geq dR_i^k\ ; \qquad i \neq k,\ i,k \in R \tag{4.25}$$

To minimize the total number of RSs that are installed, S_i should be minimized.

We also aim at reducing the number of RSs used in backup topologies. Minimizing the number of RSs used in every topology allows us to remove lengthy paths. For example, if a TP can connect to the BS by going through one RS only, it is better not to use two RSs for this TP. Thus, minimizing the number of RSs in the backup topologies will make a TP use the necessary number of RSs it needs to connect to the BS; and not more than that.

The term T_k designates the number of RSs used in the backup topology when RS k fails. We have the following constraint.

$$T_k \geq \sum_{i \in R,\ i \neq k} dR_i^k \ ; \qquad \forall k \in R \tag{4.26}$$

For a similar reason as above, we aim to minimize the number of RSs that is used in the main topology, designated by the term V. We have the constraint:

$$V \geq \sum_{i \in R} dR_i \tag{4.27}$$

The term Obj combines the terms above. The main term from the above is $\sum_{i \in R} S_i$ since it gives the number of RSs that should be installed. It should be given a higher weight than the other terms. The maximum value of $\sum_{k \in R} T_k$ is n^2 and the maximum value of V is n. Thus, we give the weight $n^2 + n$ to the term with S_i.

$$Obj \geq (n^2 + n).\sum_{i \in R} S_i + \sum_{k \in R} T_k + V \tag{4.28}$$

Then, the objective function is:

$$\min(Obj) \tag{4.29}$$

4.5 Numerical Results

This section shows the numerical results. Planning problems are presented and solved with our model. Firstly, we show examples of planning without fault-tolerance. In these examples, when an RS fails, there is no guarantee of service to the TPs. Secondly, we show examples with fault-tolerance, where service will be guaranteed even in case of failure of an RS.

4.5.1 Planning without Fault-Tolerance

This section presents the initial results of planning the RS locations without fault-tolerance. In this section, where no fault-tolerance is considered, the variables and constraints in the model that are used for fault-tolerance are omitted. These are all the variables that have an index k, which are: dR_i^k, dBR_i^k, dBT_i^k, dRR_{ij}^k, dRT_{ij}^k, fBR_i^k and fRR_{ij}^k.

Also, the objective function will change. In the case where we do not have fault-tolerance, the objective is simply to minimize the total number of RSs used. Then, the objective function is:

$$\min \ (\sum_{i \in R} dR_i) \tag{4.30}$$

This section also shows how our model considers the existence of obstacles in the planning field. We consider two types of obstacles. A lake-type obstacle does not allow installing RSs on it. However, it allows a wireless link to cross the lake. The second one is the mountain-type obstacle, which could be a steep mountain or a tall structure. This type of obstacle also does not allow installing an RS on it, and it does not allow a wireless link to go through it.

4.5.2 Planning RS Locations

The following is an RS planning example that uses our solution. The problem is shown in Figure 4.1. The planning area is made discrete by the use of a square grid. The BS location is on the top line of the grid as shown in the figure. We select this setting since we consider that the BS is at the edge of the connected area. The area below the BS does not have the connection and we plan to connect this area through the BS. Without loss of generality, we can use any topology with our model.

The potential sites for an RS are the corners of a grid square. In the 4-by-4 grid, the RS sites are numbered 0 to 15, as shown in the figure. The possible site of a TP is in the center of a square. The TPs are numbered 0 to 9. In the figure, the TP numbers are 2, 4, 5, 6 and 8. The number shown in the figure next to each TP is its traffic demand in Mbps.

The maximum rate on the links is shown in Table 4.2. The distance unit is the side length of a square in the grid. The table shows the feasible rate for the corresponding distance interval (per the model in Section 4.3.3).

The solution to this planning example is shown in Figure 4.2. The shaded RS sites are the ones that have been selected. Three RSs are needed for this problem, which are RS1, RS5 and RS7. The solid line links are the BS-RS and RS-RS links. The dashed lines are the BS-TP and

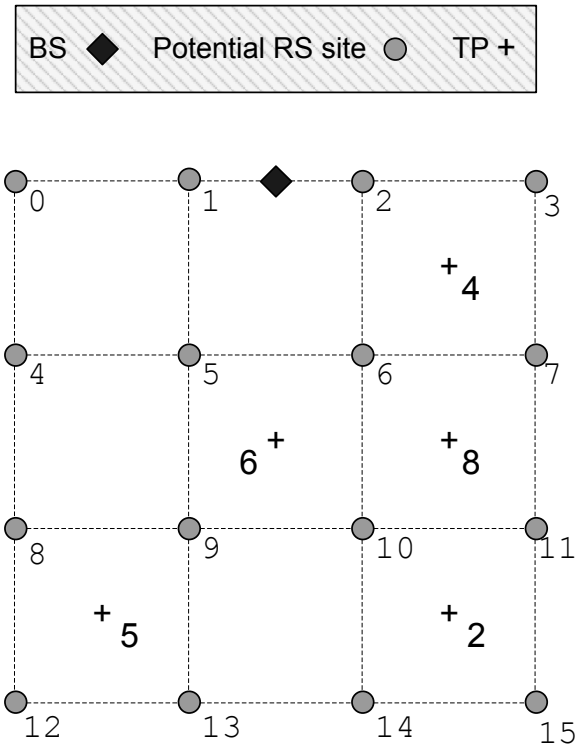

Figure 4.1 Planning RS Locations without Fault-Tolerance. Problem (TP Demands in Mbps)

RS-TP links. The underlined numbers are the link rates allocated by the solution. The rates of the dotted links are equal to the corresponding TPs' rates. The arrows on the links show the flow of traffic in the downlink to facilitate interpreting the results. However, the traffic may go in the uplink or downlink direction.

We note the following observations from this example:

- The distance from the TP to the BS doesn't necessarily indicate a direct or relayed connection. For example, the TP with demand of 2 Mbps is the farthest from the BS.

Table 4.2 Link Rates Used in Our Results

Distance (*unit*)	**Link Rate** (*Mbps*)
$if\ distance <= 1$	$rate = 10$
$else\ if\ distance <= 2$	$rate = 5$
$else\ if\ distance <= 3$	$rate = 2$
$else\ if\ distance <= 4$	$rate = 1$
$else$	$rate = 0$

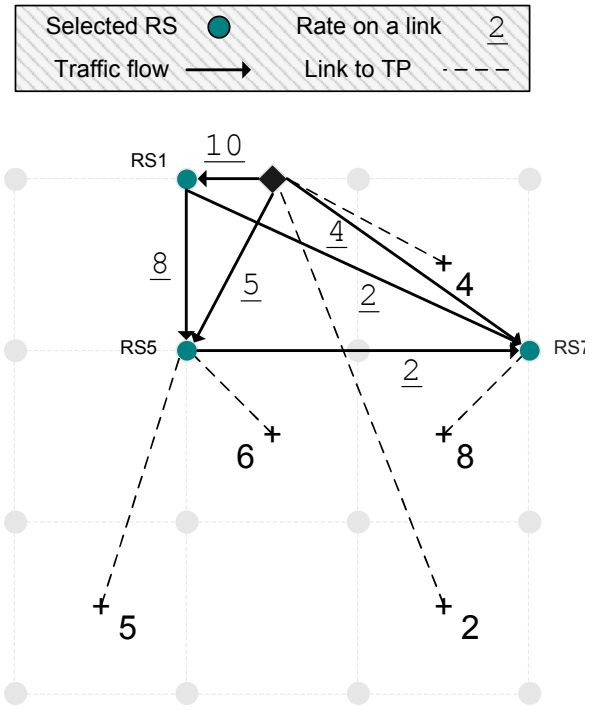

Figure 4.2 Planning RS Locations without Fault-Tolerance. Solution.

However, its demand is relatively low which can be satisfied by a single link. On the other hand, TPs which are closer to the BS have higher demands, and require the use of relay stations.

- Secondly, the RS-to-RS links help in reducing the number of relays. In our example, there is more traffic to the right of the BS ($4 + 8 + 2 = 14$) than the left (5) and the middle (6). Thus, in the solution, the diagonal and horizontal links, both with rate of 2 Mbps, between RSs (1,7) and (5,7), respectively, relay the traffic from the right side to the less congested left side. If this was not the case, more relays would be needed on the right side.

4.5.3 Obstacle Model

In this part, we consider the existence of obstacles in the planning area. In real life, obstacles could be natural such as a hill, a forest or a lake, or man-made structures such as buildings, water towers or others.

In this subsection, we consider two cases of obstacles as shown in Figure 4.3. First, there might be lakes or other similar obstacles in the area. This type of obstacles does not allow deploying an RS in it. However, two adjacent RSs can transmit on a link that goes over this type of obstacle, as shown in the figure. In the model, the RS location over this obstacle is canceled by setting the maximum rate on such RS to null. In the model, these are the parameters for when RS_i is canceled: $m_i^{BR} = 0$ as the rate to the BS, $m_{ij}^{RR} = 0, \forall j \in R$ as the rate to other RSs, and $m_{ik}^{RT} = 0, \forall k \in T$ as the rate to all TPs.

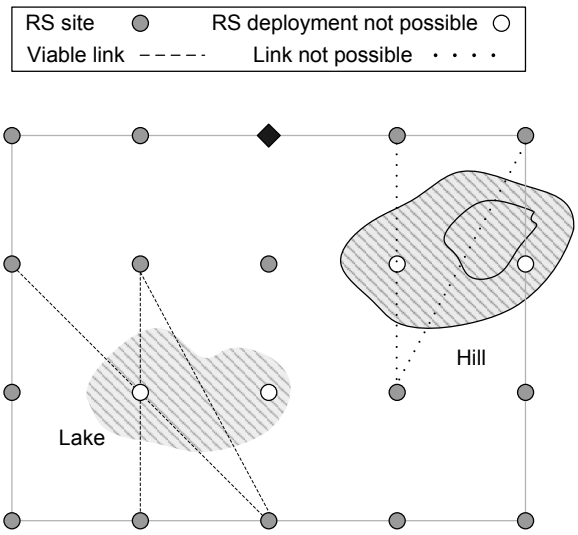

Figure 4.3 Obstacle Model.

The other type of obstacles that we consider doesn't allow a wireless link to penetrate it. For example, it could be a small hill in the area. We also consider that it might be rough terrain and it wouldn't allow deploying an RS on it. This type is also shown in Figure 4.3. The same conditions that were for the first type apply here to cancel the use of RSs that coincide with the obstacle. In addition, we need to cancel the links that traverse the obstacle too. So for every link that is blocked by the obstacle from the BS to RS_i, from the BS to TP_j or from RS_i to RS_k and from RS_i to TP_j, we have $m_i^{BR} = 0,\ m_j^{BT} = 0,\ m_{ik}^{RR} = 0,\ m_{ij}^{RT} = 0$.

4.5.4 Planning with Lake-Type Obstacles

Now we show how the planning is solved by our model with obstacles in the planning area. We take the planning case in Figure 4.1 and we insert obstacles in it, once a lake and once a hill. Then we compare the planning outcome due to the effect of the obstacles.

First, we start by inserting a lake in the planning area as shown in Figure 4.4. The lake covers the RS locations of RS5 and RS6. Hence, it is no longer possible to deploy RSs at these sites. We note the following observations in comparison to Figure 4.2.

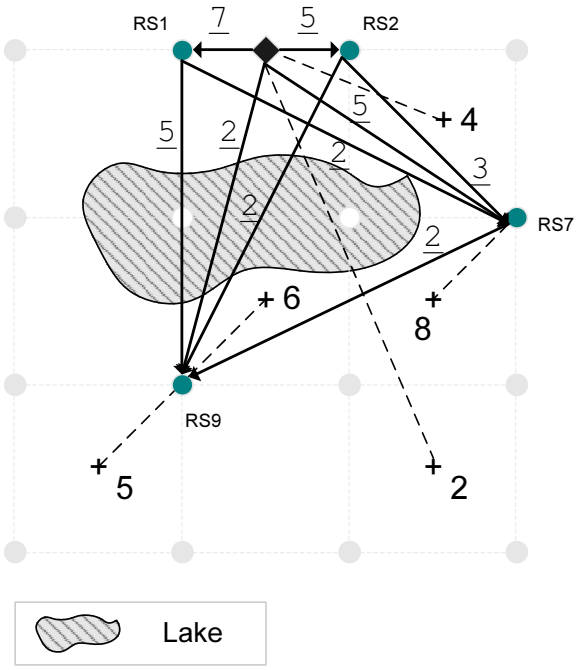

Figure 4.4 Planning RS Locations without Fault-Tolerance. Solution with a Lake-Type Obstacle.

- The number of RSs has changed from 3 to 4. This shows that an obstacle might increase the number of RSs. Previously, RS1, RS5 and RS7 were used. The obstacle makes RS5 no longer available. In the new planning, the RSs used are RSs 1, 2, 7 and 9.

- As we designated in the model, the links can traverse a lake-type obstacle. Hence, in the planning result, many links traversed the lake. Notably, from RS9, there are 3 links that go over the lake to RS1, the BS and RS2.

- Finally, the links that were direct from TPs to BS remain unchanged since the lake-type obstacle doesn't interfere with these links.

4.5.5 Planning with Mountain-Type Obstacles

In the following case, we insert a mountain-type obstacle in the planning area. This type of obstacle, as the lake-type, cancels the overlapping RS position. In addition, it cancels the links traversing it. Thus, the mountain-type obstacle is more restrictive than the lake-type. The result of the planning is shown in Figure 4.5. The result shows that here, 5 RSs are now needed to support the TPs. This is intuitive since this obstacle has more restrictions than in the previous case, which required 4 RS, and compared to the initial case with no obstacles, which required 3 RSs.

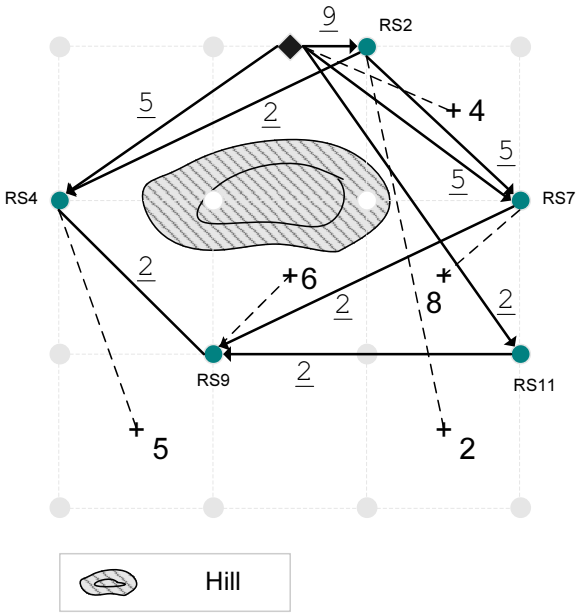

Figure 4.5 Planning RS Locations without Fault-Tolerance. Solution with a Mountain-Type Obstacle.

We note the following observations:

- The network used links that are dispersed around the mountain, since it is not possible to go through it.

- Notice for TP4 (with rate of 2 Mbps), it was able to communicate directly with the BS in the case with no obstacles and in the case with lake obstacle. Now, it cannot do this anymore, because of the mountain, and it goes through RS2.
- Finally, we saw that more restrictions in the planning area on the RS locations and the viable links will likely require more RSs to satisfy the TPs.

4.5.6 Planning with Fault-Tolerance

In this part, we present planning results with fault-tolerance. The problem input is shown in Figure 4.6 and it has the same TP locations and rates as the example in Figure 4.1. In this case, there are also backup rates for each TP, which are smaller than or equal to the the main rate.

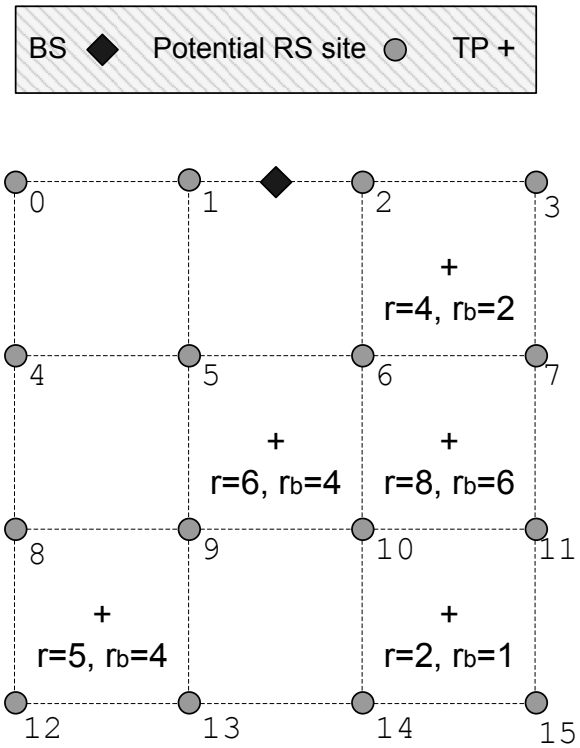

Figure 4.6 Planning RS Locations with Fault-Tolerance. Problem (main rates and backup rates in Mbps)

Figure 4.7 shows the main topology which supports the main rates of the TPs. This

topology, similar to that in Figure 4.2, supports the same normal operation rates, and also uses three RSs. However, unlike the topology in Figure 4.2, it uses RS 2, RS 5 and RS 7. Moreover, it also requires the installation of an additional relay at site RS 10. Even though RS 10 is not used in the main topology, it is required in case one of the three used RSs fails.

In Figure 4.7, the TPs with rates of 4 and 2 Mbps connect directly to the BS since their direct link can support the required rate. This is similar to Figure 4.2. Each of the other TPs connects to the RS that is closest to it.

Figure 4.8 shows the backup topology that is used when RS 2 fails. In this topology, the backup rates are supported, which are smaller than the main rates in this example. Due to the lower rates, now three TPs are able to have a direct connection to the BS (compared to two in the main topology). The other two TPs connect through RSs. In this topology, RS 10 is also not used since RS 5 and RS 7 are able to support the TPs' demands, which makes the recovery from RS 2 failure faster, since RS 10 need not be used.

The topology in Figure 4.9 is the case when RS 5 fails. In this case, three RSs are needed to support the TPs. Notice that in the previous topology, RS 5 was strategically located between the BS and the TP in the low-left corner. Since RS 5 has failed, there is no RS that can satisfy this TP. Thus, two RSs are used to connect this TP.

In Figure 4.10, the topology that is used when RS 7 fails is shown. Now the TP in the low-left corner is able to connect via RS 5. However, the TP with demand 6 Mbps, which was previously relying on RS 7 cannot connect with only one RS. Then, RS 2 and RS 10 convey the traffic of this TP in this case. However, the TP connects only to RS 10, and RS 10 connects to both the BS and RS 2 in order to receive the data.

Finally, when RS 10 fails, we can continue to use the main topology as in Figure 4.7 since this topology doesn't use RS 10.

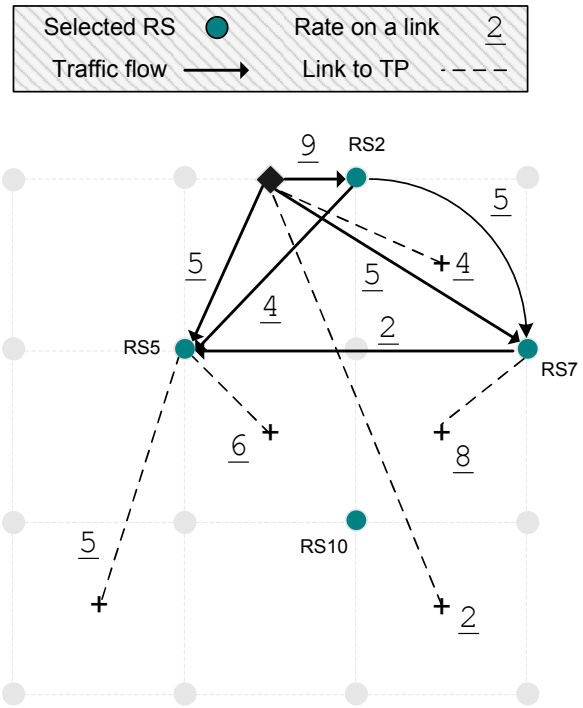

Figure 4.7 Planning RS Locations with Fault-Tolerance. Solution: The Main Topology.

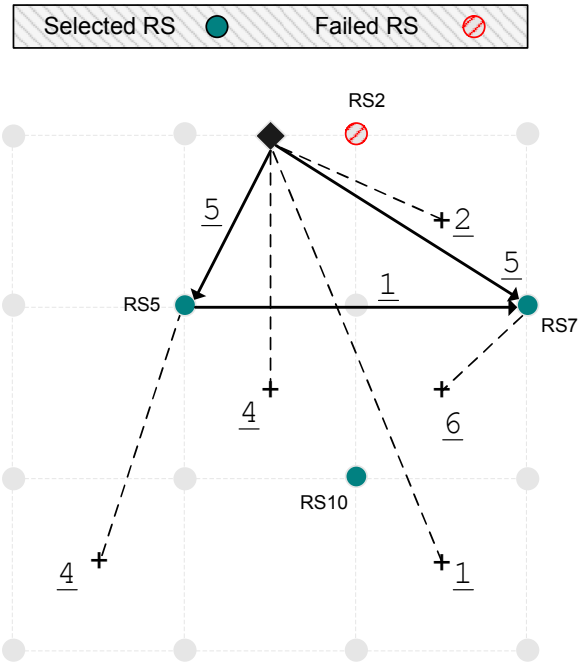

Figure 4.8 Planning RS Locations with Fault-Tolerance. Solution: Backup Topology when RS2 Fails.

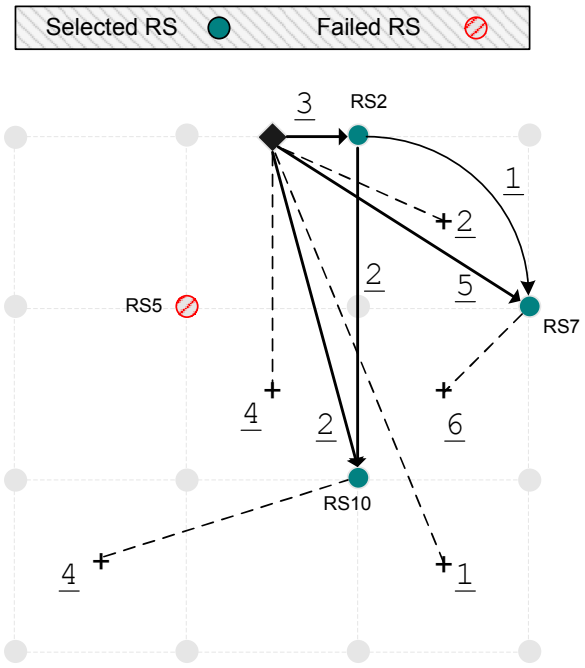

Figure 4.9 Planning RS Locations with Fault-Tolerance. Solution: Backup Topology when RS5 Fails.

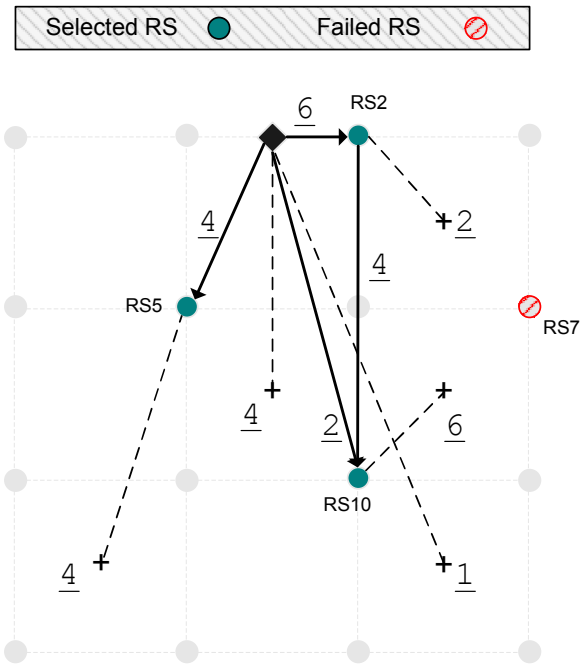

Figure 4.10 Planning RS Locations with Fault-Tolerance. Solution: Backup Topology when RS7 Fails.

4.5.7 A Realistic Planing Case

In this part, we consider a case where we need to plan the locations of RSs. This scenario is similar to ones that could happen in real life where our solution can be used. This planning case is shown in Figure 4.11.

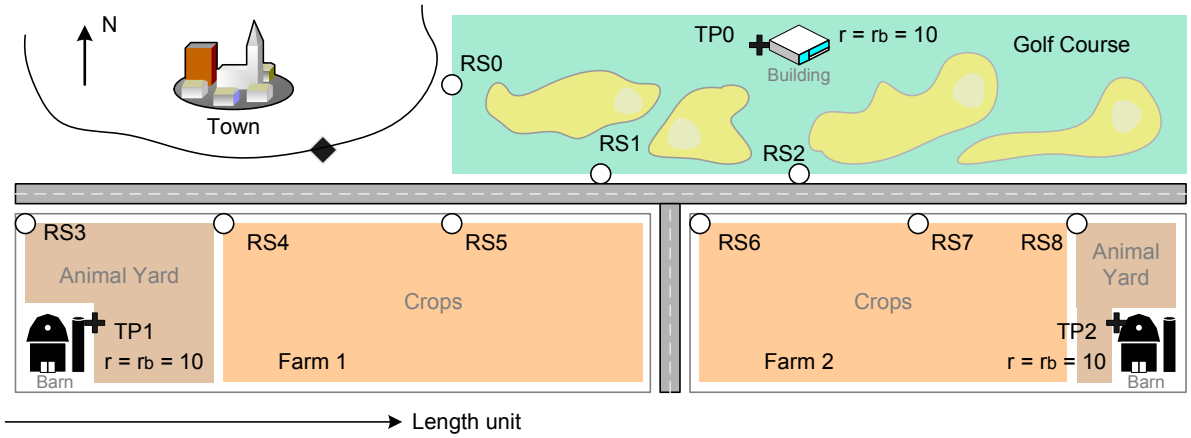

Figure 4.11 Planning of RS Locations in a Realistic Scenario. Fault-Tolerance is Provided. The Planning Problem.

On the top-left corner of the figure, there is a town that has Internet connectivity. However, there is no connectivity at the areas outside of the city limits. There is a BS at the edge of the town, marked by the black diamond.

The goal of this planning case is to connect the three properties adjacent to the town through the BS. They are a golf course east of the town, a farm south of the town and another farm southeast of the town. Each of the three entities has a single building: the main building at the golf course and the barns at the farms. In each property, the TP is at the corresponding building. Each of the TP has a regular demand rate of $r_i = 10$ Mbps. In case of an RS failure, it is required to receive the same level of service, i.e., no drop in rate should be experienced. Therefore, $rb_i = 10$ Mbps.

Each property has provided possible locations for RSs to be deployed at their site. There are three RS sites at each property marked RS 0 through RS 8 in Figure 4.11. In the layout of the properties, there are large areas that cannot have an RS, such as in the middle of the golf course or in the middle of the crop field. Therefore, the RS sites are located at the edges

of fields. Our model can work with any layout of the potential RS sites. Once the X- and Y-coordinates of the RS sites, BS and TPs are known, our model generates the optimal number and locations of the RSs.

In this example, the data rates on a link are determined according to Table 4.2. The length unit is shown in Figure 4.11.

Figure 4.12 shows the main topology. It turns out that 4 RSs are needed for this planning case. They are RS 1, RS 6, RS 7 and RS 8. In the main topology, RS 7 is not used.

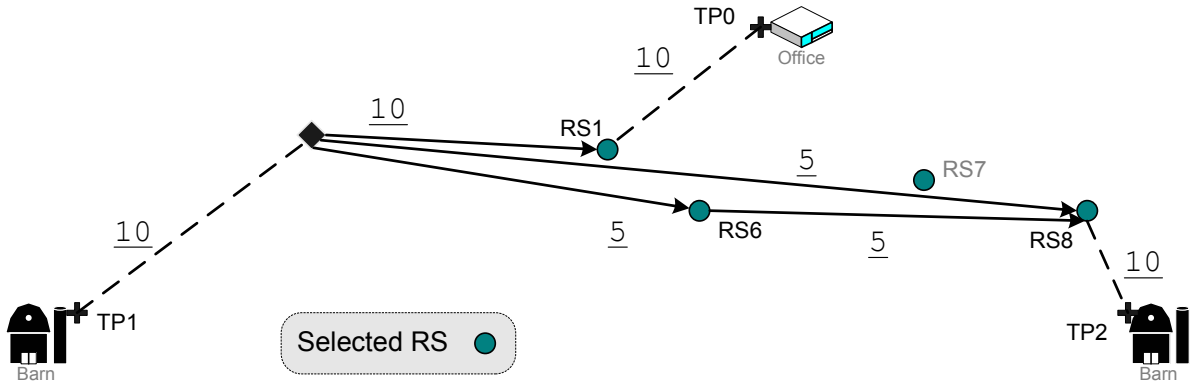

Figure 4.12 Main Topology

In the main topology, TP 1 connects directly to the BS since it is at proximity and is able to obtain 10 Mbps through a direct link. TP 0, which is a bit farther, requires two hops, which go through RS 1. TP 2 is even further than TP 0. Therefore, it connects to the TP through RS 6 and RS 8.

In Figure 4.13 the topology when RS 1 fails is shown. TP 1 doesn't have any change and still has a direct connection to the BS. But TP 2 connects to RS 7 even tough RS 8 is closer to it. This happens since we restrict the TP to connect to only one RS (or to the BS). RS 7 is closer to the BS and therefore is able to obtain a higher data rate than RS 8 can do. As a result, TP 2 is linked to RS 7.

In Figure 4.14, a similar observation happens for TP 2 which connects through RS 7 even though RS 8 is closer to it than RS 7. In this topology, TP 0 uses a single relay, RS 1, to reach the BS. And, finally, TP 1 also uses a direct connection to the BS.

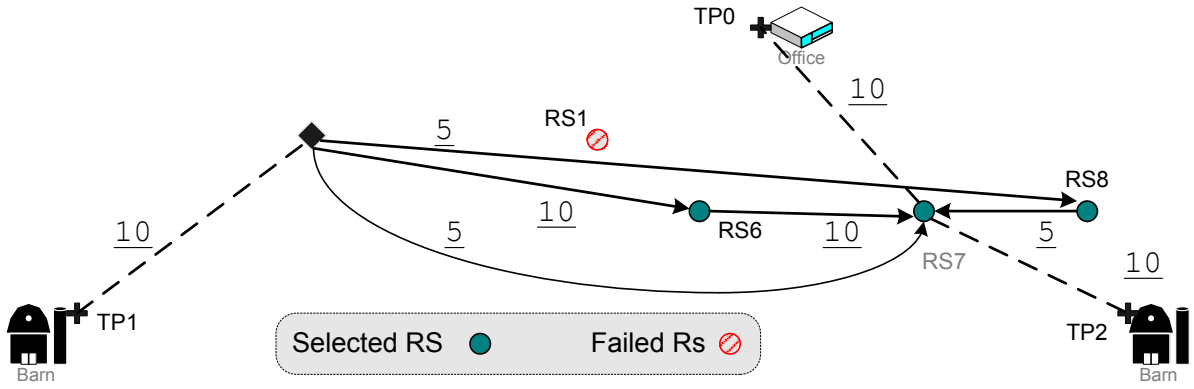

Figure 4.13 Backup Topology when RS 1 fails.

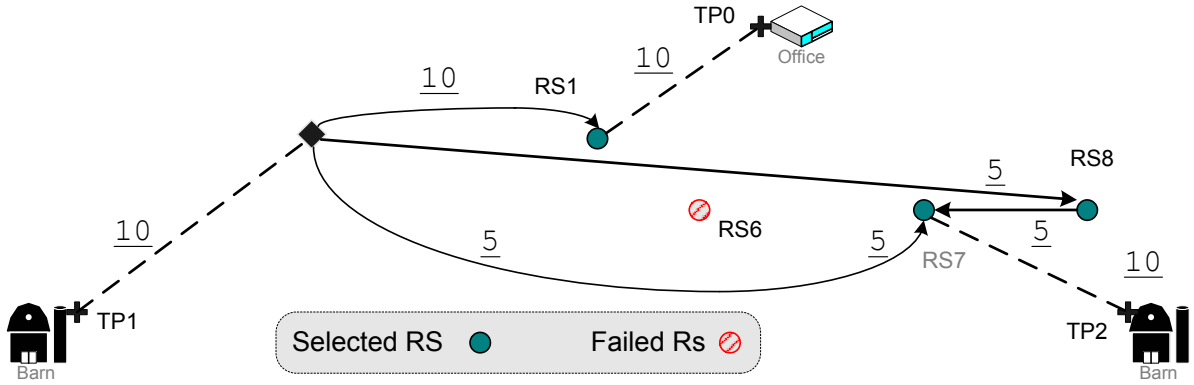

Figure 4.14 Backup Topology when RS 6 fails.

Figure 4.15 shows the topology when RS 8 fails. Here, TP 0 uses a two-hop connection through RS 6. TP 2 takes its traffic from RS 7.

4.6 Conclusions

In this paper, we considered the problem of planning the RS locations in the WiMAX network in a fault-tolerant manner. To the best of our knowledge, this is the first work that provides fault-tolerance in planning RS locations in WiMAX. We provided a Mixed-Integer Linear Program (MILP) that formulates the planning problem. We solved the problem with CPLEX and obtained numerical results that show how our model produces the main topology and the backup topologies of a network. Finally, we considered the existence of obstacles in the planning field, such as a large structure or a natural obstacle. We showed how our model

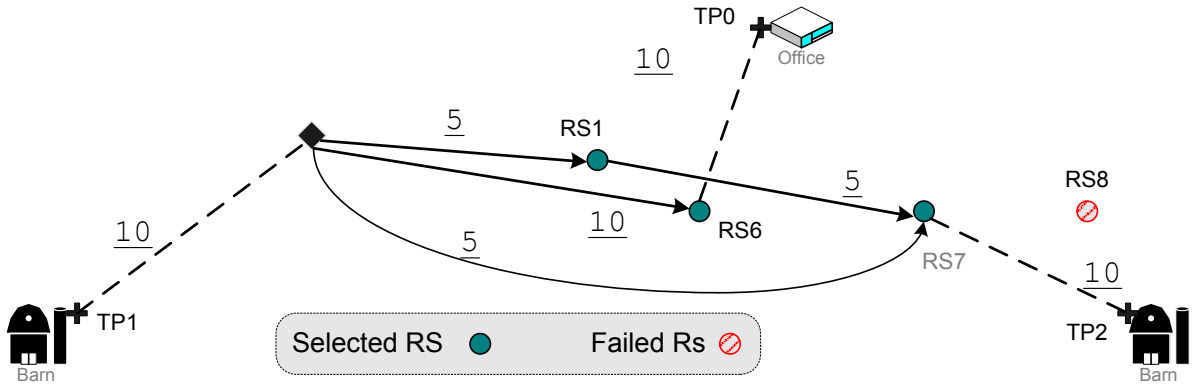

Figure 4.15 Backup Topology when RS 8 fails.

can deal with these obstacles and plan the network around them effectively.

CHAPTER 5. Conclusion

In this dissertation, we considered the MAC scheme of WLANs and the RS placement in WiMAX networks. First, we proposed the CONTI scheme for access in WLANs. CONTI has the feature of resolving the contention in a constant number of slots. Also, by optimizing its parameters, CONTI achieves a very low collision rate. Hence, this allows it to obtain a high throughput. The results we provided also evaluated the delay and fairness performance of CONTI. Secondly, we presented the GMAC scheme which provides access to the stations in a WLAN with high data rates. The grouping mechanism provided an efficient contention resolution. The frame aggregation and eliminating the use of control frames within the groups also increased the efficiency. The results show that GMAC provides a significantly higher throughput than the standard's scheme and other schemes found in the literature. Our results also evaluated the delay and fairness performance of GMAC. Finally, we presented an optimization formulation to plan the RS locations in WiMAX networks. We considered fault-tolerance so that a predictable level of performance is obtained in the case an RS fails. Our solution found the minimum number of RSs that should be used, their respective locations, the links that should be used and the rate on each link. We also considered the existence of obstacles in the planning field and showed how our model can accommodate them.

BIBLIOGRAPHY

[1] IEEE Std 802.11, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications." 1999.

[2] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 18, no. 3, Mar. 2000.

[3] E. Ziouva and T. Antonakopoulos, "CSMA/CA Performance under High Traffic Conditions: Throughput and Delay Analysis," *Elsevier Computer Communications*, vol. 25, no. 3, 2002.

[4] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2005.

[5] M. Carvalho and J. Garcia-Luna-Aceves, "Delay Analysis of IEEE 802.11 in Single-Hop Networks," in *IEEE International Conference on Network Protocols (ICNP)*, 2003.

[6] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed Fair Scheduling in a Wireless LAN," in *Sixth Annual International Conference on Mobile Computing and Networking (MobiCom)*, August 2000.

[7] P. Jacquet, P. Minet, P. Muhlethaler, and N. Rivierre, "Priority and Collision Detection with Active Signaling - The Channel Access Mechanism of HIPERLAN," *Wireless Personal Communications*, vol. 4, no. 1, pp. 11–26, January 1997.

[8] J. L. Sobrinho and A. S. Krishnakumar, "Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 17, no. 8, pp. 1353–1368, Aug. 1999.

[9] G. Wikstrand, T. Nilsson, and M. Dougherty, "Prioritized Repeated Eliminations Multiple Access: A Novel Protocol for Wireless Networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, April 2008, pp. 1561–1569.

[10] B. Zhou, A. Marshall, and T.-H. Lee, "A k-Round Elimination Contention Scheme for WLANs," *Mobile Computing, IEEE Transactions on*, vol. 6, no. 11, pp. 1230–1244, Nov. 2007.

[11] Z. Abichar and J. M. Chang, "CONTI: Constant-Time Contention Resolution for WLAN Access," in *IFIP Networking*, 2005.

[12] M. Heusse, F. Rousseau, R. Guillier, A. Duda, "Idle Sense: An Optimal Access Method for High Throughput and Fairness in Rate Diverse Wireless LANs," in *SIGCOMM*, 2005.

[13] Y. Grunenberger, M. Heusse, F. Rousseau, and A. Duda, "Experience with an Implementation of the Idle Sense Wireless Access Method," in *Proceedings of the 2007 ACM CoNEXT conference (CoNEXT'07)*. ACM, Dec. 10–13 2007, pp. 1–12.

[14] H. Wu, A. Utgikar, and N. Tzeng, "SYN-MAC: A Distributed Medium Access Control Protocol for Synchronized Wireless Networks," *Mobile Networks and Applcations (MONET)*, vol. 10, no. 5, Oct. 2005.

[15] T. You, C. Yeh, and H. Hassanein, "A New Class of Collision Prevention MAC Protocols for Wireless Ad Hoc Networks," in *IEEE International Conference on Communications (ICC)*, 2003.

[16] C. Yeh and T. You, "A QoS MAC Protocol for Differentiated Service in Mobile Ad Hoc Networks," in *IEEE International Conference on Parallel Processing (ICPP)*, 2003.

[17] J. Stine, G. DeVeciana, K. Grace, and R. Durst, "Orchestrating Spatial Reuse in Wireless Ad Hoc Networks Using Synchronous Collision Resolution (SCR)," *Journal of Interconnection Networks*, vol. 3, no. 3-4, Sep.-Dec. 2002.

[18] J. Galtier, "Analysis and optimization of MAC with constant size congestion window for WLAN," in *Second International Conference on Systems and Networks Communications, 2007. ICSNC 2007.*

[19] L. Bononi, M. Conti and E. Gregori, "Runtime Optimization of IEEE 802.11 Wireless LANs Performance," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 1, Jan. 2004.

[20] K. C. Tay, K. Jamieson and H. Balakrishnan, "Collision-Minimizing CSMA and Its Applications to Wireless Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, Aug. 2004.

[21] Q. Ni, I. Aad, C. Barakat, and T. Turletti, "Modeling and Analysis of Slow CW Decrease for IEEE 802.11 WLAN," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2003.

[22] H. Wu, S. Cheng, Y. Peng, K. Long, and J. Ma, "IEEE 802.11 Distributed Coordination Function (DCF): Analysis and Enhancement," in *IEEE International Conference on Communications (ICC)*, 2002.

[23] F. Cali, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Transactions on Networking (ToN)*, vol. 6, no. 8, pp. 785–799, Dec. 2000.

[24] K. Jamieson, H. Balakrishnan, and Y. C. Tay, "Sift: A MAC protocol for Event-Driven Wireless Sensor Networks," in *Proc. of the 3rd European Workshop on WSNs (EWSN)*, 2006.

[25] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: The MIT Press, 2001.

[26] IEEE Std 802.11b, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band." 1999.

[27] Z. Abichar, J. M. Chang, and D. Qiao, "Group-Based Medium Access for Next-GenerationWireless LANs," in *WOWMOM '06: Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*. IEEE Computer Society, 2006, pp. 35–41.

[28] R. Jain, *The Art of Computer Systems Performance Analysis*. John Wiley and Sons, 1991.

[29] C. E. Koksal, H. Kassab, and H. Balakrishnan, "An Analysis of Short-Term Fairness in Wireless Media Access Protocols," in *ACM Sigmetrics 2000*, June 2000.

[30] "IEEE Std 802.11n-2009, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications – Amendment 5: Enhancements for Higher Throughput," pp. c1–502, 2009.

[31] E. Perahia, "IEEE 802.11n Development: History, Process, and Technology," *Communications Magazine, IEEE*, vol. 46, no. 7, pp. 48 –55, jul. 2008.

[32] Y. Xiao, "IEEE 802.11n: enhancements for higher throughput in wireless LANs," *Wireless Communications, IEEE*, vol. 12, no. 6, pp. 82 – 91, dec. 2005.

[33] P. Bahl and V. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2000.

[34] D. Madigan, E. Alnahrawy, R. Martin, W. Ju, P. Krishnan, and A. Krishnakumar, "Bayesian Indoor Positioning Systems," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2005.

[35] M. Robinson and I. Psaromiligkos, "Received Signal Strength Based Location Estimation of a Wireless LAN Client," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2005.

[36] Y. Yuan, W. A. Arbaugh, and S. Lu, "Towards Scalable MAC Design for High-Speed Wireless LANs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007.

[37] K.-C. Ting, H.-H. Lee, and F. Lai, "Design and Analysis of Enhanced Grouping DCF Scheme for the MAC Layer Enhancement of 802.11n with Ultra-high Data Rate," in *Wireless Communication Systems, 2007. ISWCS 2007. 4th International Symposium on*, Oct. 2007.

[38] K.-C. Ting, M.-y. Jan, S.-h. Hsieh, H.-H. Lee, and F. Lai, "Design and analysis of grouping-based DCF (GB-DCF) scheme for the MAC layer enhancement of 802.11 and 802.11n," in *MSWiM '06: Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems.* New York, NY, USA: ACM, 2006, pp. 255–264.

[39] Y. Yuan, D. Gu, W. Arbaugh, and J. Zhang, "High-Performance MAC for High-Capacity Wireless LANs," in *IEEE Int. Conference on Computer Communications and Networks (ICCCN)*, 2004.

[40] D. Skordoulis, Q. Ni, and C. Zarakovitis, "A Selective Delayed Channel Access (SDCA) for the High-Throughput IEEE 802.11n," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, apr. 2009, pp. 1 –6.

[41] S. Kim, Y. Kim, S. Choi, K. Jang, and J. Chang, "A High-Throughput MAC Strategy for Next-Generation WLANs," in *IEEE Int. Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2005.

[42] C.-Y. Wang and H.-Y. Wei, "IEEE 802.11n MAC Enhancement and Performance Evaluation," *Mobile Networks and Applications*, vol. 14, no. 6, pp. 760–771, 2009.

[43] M. Ozdemir, D. Gu, A. McDonald, and J. Zhang, "Enhancing MAC Performance with a Reverse Direction Protocol for High-Capacity Wireless LANs," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, sep. 2006, pp. 1 –5.

[44] D. Akhmetov, "802.11N: Performance Results of Reverse Direction Data Flow," in *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, sep. 2006, pp. 1 –3.

[45] T. Selvam and S. Srikanth, "A frame aggregation scheduler for IEEE 802.11n," in *Communications (NCC), 2010 National Conference on*, jan. 2010, pp. 1 –5.

[46] H. Jin and Y. Zhu, "Research on an Improved MRA Mechanism in 802.11n Network," in *Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09. 5th International Conference on*, sep. 2009, pp. 1 –4.

[47] T. Li, Q. Ni, D. Malone, D. Leith, Y. Xiao, and T. Turletti, "Aggregation With Fragment Retransmission for Very High-Speed WLANs," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 2, pp. 591 –604, apr. 2009.

[48] D. Skordoulis, Q. Ni, H.-H. Chen, A. Stephens, C. Liu, and A. Jamalipour, "IEEE 802.11n MAC frame aggregation mechanisms for next-generation high-throughput WLANs," *Wireless Communications, IEEE*, vol. 15, no. 1, pp. 40 –47, feb. 2008.

[49] B. S. Kim, H. Y. Hwang, and D. K. Sung, "Effect of Frame Aggregation on the Throughput Performance of IEEE 802.11n," mar. 2008, pp. 1740 –1744.

[50] L. Cai, X. Ling, X. Shen, J. Mark, and H. Long, "Capacity analysis of enhanced mac in ieee 802.11n," in *Communications and Networking in China, 2006. ChinaCom '06. First International Conference on*, oct. 2006, pp. 1 –5.

[51] Y. P. Fallah and H. M. Alnuweiri, "Modeling and performance evaluation of frame bursting in wireless lans," in *IWCMC '06: Proceedings of the 2006 international conference on Wireless communications and mobile computing.* New York, NY, USA: ACM, 2006, pp. 869–874.

[52] A. Ettefagh, M. Kuhn, C. Esli, and A. Wittneben, "Performance of a cluster-based MAC protocol in multiuser MIMO wireless LANs," in *Smart Antennas (WSA), 2010 International ITG Workshop on*, feb. 2010, pp. 262 –269.

[53] A. Ashtaiwi and H. Hassanein, "MIMO-Based Collision Avoidance in IEEE 802.11e Networks," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 3, pp. 1076 –1086, mar. 2010.

[54] E. Kartsakli, N. Zorba, L. Alonso, and C. Verikoukis, "Multiuser MAC Protocols for 802.11n Wireless Networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, jun. 2009, pp. 1 –5.

[55] M. Siam and M. Krunz, "An overview of MIMO-oriented channel access in wireless networks," *Wireless Communications, IEEE*, vol. 15, no. 1, pp. 63 –69, feb. 2008.

[56] Z. Hu and C.-K. Tham, "CCMAC: coordinated cooperative MAC for wireless LANs," in *MSWiM '08: Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems.* New York, NY, USA: ACM, 2008, pp. 60–69.

[57] S. Pollin and A. Bahai, "Performance Analysis of Double-Channel 802.11n Contending with Single-Channel 802.11," in *Communications, 2009. ICC '09. IEEE International Conference on*, jun. 2009, pp. 1 –6.

[58] Y. Utsunomiya, T. Tandai, T. Adachi, and M. Takagi, "A MAC Protocol for Coexistence between 20/40 MHz STAs for High Throughput WLAN," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 3, may. 2006, pp. 1136 –1140.

[59] A. E. Xhafa, A. Batra, and A. Zaks, "On the 20/40 MHz Coexistence of Overlapping BSSs in WLANs," *Journal of Networks*, vol. 3, no. 7, pp. 56–63, July 2008.

[60] L. Xu, K. Yamamoto, and S. Yoshida, "Performance Comparison Between Channel-Bonding and Multi-Channel CSMA," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, mar. 2007, pp. 406 –410.

[61] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance Anomaly of 802.11b," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2003.

[62] G. Tan and J. Guttag, "Time-based Fairness Improves Performance in Multi-rate Wireless LANs," in *The USENIX Annual Technical Conference*, Boston, MA, June 2004.

[63] J. Choi, J. Yoo, S. Choi, and C. Kim, "EBA: An Enhancement of IEEE 802.11 DCF via Distributed Reservation," *IEEE Trans. on Mobile Computing*, vol. 4, no. 4, Jul./Aug. 2005.

[64] Y. Xiao, H. Li, K. Wu, K. Leung, and Q. Ni, "Reservation and Grouping Stations for the IEEE 802.11 DCF," in *IFIP Networking*, 2005.

[65] Z. Abichar and J. M. Chang, "A Medium Access Control Scheme for Wireless LANs with Constant-Time Contention," *Mobile Computing, IEEE Transactions on*, Volume 10, Number 2, February 2011.

[66] "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification," *IEEE Std 802.16j-2009 (Amendment to IEEE Std 802.16-2009)*, pp. c1 –290, june 2009.

[67] S. Peters and R. Heath, "The future of WiMAX: Multihop relaying with IEEE 802.16j," *Communications Magazine, IEEE*, vol. 47, no. 1, pp. 104 –111, january 2009.

[68] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *Communications Magazine, IEEE*, vol. 47, no. 10, pp. 100 –105, october 2009.

[69] Z. Abichar, A. E. Kamal, and J. M. Chang, "Planning of Relay Station Locations in IEEE 802.16 (WiMAX) Networks," in *Proc. of IEEE Wireless Communications and Networking Conference, WCNC 2010*.

[70] Y. Yu, S. Murphy, and L. Murphy, "Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks," *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, pp. 922–926, Jan. 2008.

[71] ——, "A Clustering Approach to Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks," *Communications, 2008. ICC '08. IEEE International Conference on*, pp. 2586–2591, May 2008.

[72] B. Lin, P.-H. Ho, L.-L. Xie, and X. Shen, "Optimal relay station placement in IEEE 802.16j networks," in *IWCMC '07: Proceedings of the 2007 international conference on Wireless communications and mobile computing*. New York, NY, USA: ACM, 2007, pp. 25–30.

[73] B. Lin and P.-H. Ho, "Dimensioning and location planning of broadband wireless networks under multi-level cooperative relaying," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 11, pp. 5682 –5691, nov. 2009.

[74] B. Lin, M. Mehrjoo, P.-H. Ho, L.-L. Xie, and X. Shen, "Capacity Enhancement with Relay Station Placement in Wireless Cooperative Networks," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, apr. 2009, pp. 1 –6.

[75] B. Lin, P.-H. Ho, L.-L. Xie, X. Shen, and J. Tapolcai, "Optimal Relay Station Placement in Broadband Wireless Access Networks," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 2, pp. 259 –269, feb. 2010.

[76] B. Lin, P.-H. Ho, L.-L. Xie, and X. Shen, "Relay Station Placement in IEEE 802.16j Dual-Relay MMR Networks," in *Communications, 2008. ICC '08. IEEE International Conference on*, may. 2008, pp. 3437 –3441.

[77] S.-J. Kim, S.-Y. Kim, B.-B. Lee, S.-W. Ryu, H.-W. Lee, and C.-H. Cho, "Multi-Hop Relay Based Coverage Extension in the IEEE802.16j Based Mobile WiMAX Systems," in *Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on*, vol. 1, sep. 2008, pp. 516 –522.

[78] H.-C. Lu and W. Liao, "Joint Base Station and Relay Station Placement for IEEE 802.16j Networks," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, nov. 2009, pp. 1 –5.

[79] D. Niyato, E. Hossain, D. Kim, and Z. Han, "Relay-centric radio resource management and network planning in IEEE 802.16j mobile multihop relay networks," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 6115 –6125, dec. 2009.

[80] C.-Y. Chang, C.-T. Chang, M.-H. Li, and C.-H. Chang, "A Novel Relay Placement Mechanism for Capacity Enhancement in IEEE 802.16j WiMAX Networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, jun. 2009, pp. 1 –5.

[81] Y. Yu, S. Murphy, and L. Murphy, "Interference aware relay station location planning for IEEE 802.16j mobile multi-hop relay network," in *PM2HW2N '09: Proceedings of the 4th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks*. New York, NY, USA: ACM, 2009, pp. 201–208.

[82] ——, "Planning Base Station and Relay Station Locations for IEEE 802.16j Network with Capacity Constraints," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, jan. 2010, pp. 1 –5.

[83] A. Bari, A. Jaekel, and S. Bandyopadhyay, "Optimal Placement of Relay Nodes in Two-Tiered, Fault Tolerant Sensor Networks," in *Computers and Communications, 2007. ISCC 2007. 12th IEEE Symposium on*, 1-4 2007, pp. 159 –164.

[84] A. Bari, Y. Xu, and A. Jaekel, "Integrated Placement and Routing of Relay Nodes for Fault-Tolerant Hierarchical Sensor Networks," in *Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on*, 3-7 2008, pp. 1 –6.

[85] S. Misra, S. D. Hong, G. Xue, and J. Tang, "Constrained Relay Node Placement in Wireless Sensor Networks: Formulation and Approximations," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 2, pp. 434 –447, apr. 2010.

[86] J. Bredin, E. Demaine, M. Hajiaghayi, and D. Rus, "Deploying Sensor Networks With Guaranteed Fault Tolerance," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 1, pp. 216 –228, feb. 2010.

[87] X. Han, X. Cao, E. Lloyd, and C.-C. Shen, "Fault-Tolerant Relay Node Placement in Heterogeneous Wireless Sensor Networks," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 5, pp. 643 –656, may. 2010.

[88] W. Zhang, G. Xue, and S. Misra, "Fault-Tolerant Relay Node Placement in Wireless Sensor Networks: Problems and Algorithms," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, may. 2007, pp. 1649 –1657.

[89] B. Hao, H. Tang, and G. Xue, "Fault-tolerant relay node placement in wireless sensor networks: formulation and approximation," in *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, 2004, pp. 246 – 250.

[90] W. Stallings, *Wireless Communications and Networks*. Prentice Hall, 2001.

[91] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001), pp. 1–857, 2004.

[92] T. Theodoros and V. Kostantinos, "Wimax network planning and systems performance evaluation," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, March 2007, pp. 1948–1953.

[93] M. Molina-Garcia and J. I. Alonso, "Planning and sizing tool for wimax networks," *Radio and Wireless Symposium, 2007 IEEE*, pp. 403–406, Jan. 2007.

[94] J. Garcia-Fragoso and G. Galvan-Tejada, "Cell planning based on the wimax standard for home access: a practical case," in *Electrical and Electronics Engineering, 2005 2nd International Conference on*, Sept. 2005, pp. 89–92.

Biography

I have finished the requirements of the Ph.D. degree in Computer Engineering at Iowa State University in December 2010. I earned the Bachelor of Science degree in Computer Engineering from Iowa State University in August 2003. My research interest is in wireless networks and mobile computing.

I have worked as a Teaching Assistant and Research Assistant in the Department of Electrical and Computer Engineering for several semesters while I was enrolled in the Ph.D. program. I was also a Graduate Technical Intern at Intel Corporation in Parsippany, New Jersey during May to August of 2006. At Intel, I worked on the base band module that implements the MAC and physical layers of the WiMAX network.

I received the Professional Advancement Grant (PAG) from Iowa State University in 2005 and 2006. PAG provides a stipend to graduate students to travel and attend a conference. I have also designed the digital catalog and album art for the proceedings of the 2008 IEEE International Conference on Electro/Information Technology (EIT'2008), which was held at Iowa State University.

I can be contacted at my permanent email address at: zakhiaa@gmail.com .