On Markov Modeling of Random Access in Communication Systems

by

Yousry Salaheldin Abdel-Hamid
B.Sc. Ain Shams University, 1987
M.A.Sc. University of Victoria, 2003

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

On Markov Modeling of Random Access in Communication Systems

by

Yousry Salaheldin Abdel-Hamid
B.Sc. Ain Shams University, 1987
M.A.Sc. University of Victoria, 2003

Supervisory Committee

Dr. Fayez Gebali, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. T. Aaron Gulliver, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. Panajotis Agathoklis, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Colin Bradley, Outside Member
(Department of Mechanical Engineering)

Supervisory Committee

Dr. Fayez Gebali, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. T. Aaron Gulliver, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. Panajotis Agathoklis, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Colin Bradley, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

This dissertation considers the random access process in the Medium Access Control (MAC) of communications system. New MAC models are developed to improve the performance of random access based systems.

The first contribution is the introduction of a general multichannel random access model with a variable radix. This model is general and can be applied to many existing MAC protocols that utilize random access. It is shown that using the standard Binary Exponential Backoff (BEB) to resolve collisions is not always the best choice. By adjusting the radix, contention efficiency can be improved significantly. The analytical results obtained are confirmed by simulation.

The second contribution is the investigation of the variable radix backoff strategy with the contention-based bandwidth request (BW-REQ) mechanism in IEEE 802.16 systems. An analytical model of the BW-REQ procedure is presented which includes a variable radix in the backoff process. Analytical results are presented which show that the variable radix can easily be adjusted to the number of users and the available resources to enhance the efficiency of the Random Access Channel in the uplink subframe. Simulations results are presented to confirm the theory.

The third contribution is the development of a reliable Quality of Service (QoS) mechanism for random access systems. The available resources are quantitatively

categorized to provide differential services to two classes of users. The model is extended to employ a variable radix strategy. Results show that this strategy can be used in combination with differential services to provide an efficient QoS technique for random access.

The fourth contribution is an optimized packet-based finite state Markov chain (FSMC) model for the physical channel. This model employs an equal average fade range duration (AFRD) strategy to partition the signal-to-noise ratio (SNR). The Nakagami-m fading channel model is used as it can span a wide range of fading conditions. The accuracy of the analytical results is confirmed by simulation. A cross-layer Markov model encompassing the FSMC model and a general multichannel random access model is introduced.

Finally, a simulation toolbox using object oriented programming is presented. It was used to accurately simulate the models developed in this dissertation. This toolbox is general and can be used for a wide range of MAC models.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

"In the name of Allah, the most Gracious, the most Merciful"

"My Lord! Increase me in knowledge." (20/114).

I prostrate to Allah thanking him for granting me the strength and preserverence to complete my Ph.D.

First and foremost, I express my sincere thanks and appreciation to my supervisors, Prof. Aaron Gulliver and Prof. Fayez Gebali for their exceptionally enthusiastic supervision. Without their help and assistance, this work would not have been possible. I also express my gratitude to the other members of my supervisory committee Prof. Panajotis Agathoklis and Prof. Colin Bradley, and my external examiner Dr. Wail ElKilani at Ain Shams University, Cairo, Egypt. I would like to express my deepest gratitude to Prof. Wu-Sheng Lu for his exceptional knowledge and generous help. The success of this dissertation depends greatly on the encouragement and guidance of many others. I thank Dr. Hossam Fattah and Mr. Saamaan Pourtavakoli for their incomparable help and support in the completion of this dissertation. I would also like to thank my friend Dr. Mohamed Watheq El-Kharashi at Ain Shams University for his sincere guidance and management. I also thank Dr. Belaid Moa at Westgrid, Compute/Calcul Canada, for his participation in this work. Finally, the honor and love goes to my mother, brothers, my sister-in-law, my nephew, my niece and my friends for their endless love throughout the duration of my studies.

All thanks to Allah.

DEDICATION

To my beloved mother, whose supplications are never rejected

Chapter 1

Introduction

Multiple access is the capability of multiple users to access a resource. It is a critical aspect of many wireless communication systems. The key issue in designing the Medium Access Control (MAC) layer of a communication system is the multiple access. Research in delivering reliable multiple access mechanisms is gaining importance with the dramatic growth in the number of users and the significant increase in data rates due to multimedia and real-time applications. Multiple access protocols can be classified into fixed assignment and contention (random) based.

In fixed assignment techniques, a fixed portion of the resources is intentionally dedicated to each user either by a fixed channel slot or via a scheduled assignment. The most common of these techniques are Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA).

Although scheduled access may provide some level of guaranteed data delivery and scheduling, it has several disadvantages such large overhead, extensive delays caused by long idle periods when only few nodes are transmitting. In addition, it is not feasible to use scheduled access for network entry and connection setup purposes since user access to the network occurs randomly.

In random access, the resource (bandwidth) is available to all users all the time as a single channel. Users contend to capture the channel randomly, and therefore packet collisions are unavoidable. If collisions occur, collided packets have to be retransmitted. Retransmissions can cause significant delays that degrade network performance, therefore contention resolution is a very important component of a reliable random access protocol.

1.1 The History of Random Access

The ALOHA protocol developed by Abramson in the early 80s was the foundation of random access protocols. Initially, ALOHA was termed *pure* ALOHA. This is a very simple protocol in which users send packets and hope that they do not collide with the packets of other user. If the transmission is unsuccessful, the user independently schedules retransmission. The retransmission attempt can be scheduled randomly at some time in the future.

The *slotted* version of ALOHA is simply applying the above protocol on a slotted channel. Since data networks are based on discrete units of time (packets/frames), slotted ALOHA is considered an important variation of the simple (pure) protocol. Obviously, with slotted ALOHA, transmission is successful in a time slot if and only if one transmission occurs during that slot. Kleinrock and Lam [4], [5] were the first to develop a model for a slotted ALOHA random access system. They considered a multiple access system with a finite population consisting of N users.

Much research has focused on slotted ALOHA with numerous multiple access protocols, particularly for wireless environments. Historically, the main task during the 80's and early 90's was to investigate the performance of satellite networks where the ALOHA protocol is applicable [6].

1.2 Random Access in Wireless Networks

With the dramatic increase in demand by the mid 90s, much interest has been focused on designing wireless networks for local area communications. In a Wireless Local Area Network (WLAN), users must access the wireless channel to send data. This triggered intensive research to develop reliable contention resolution algorithms in order to minimize the number of collisions. Bianchi et al. in [7] developed the Distributed Coordination Function (DCF) as the basic mechanism for wireless users to access the network. The DCF is based on Carrier Sense Multiple Access with collision avoidance (CSMA/CA) with a truncated Binary Exponential Backoff (BEB) algorithm.

In DCF with CSMA/CA [8], a central *access point* provides wireless channel access to a group of users (nodes). A node that wishes to transmit its data via the access point has to first listen to the channel for a predetermined amount of time to make sure that no other node is transmitting on the channel within the wireless range. If

the channel is sensed free, the node transmits. If the channel is busy, the node defers transmission for a random period of time derived from the basic network timing unit (slot) and the number of attempts to retransmit, hence the name *backoff*. The term *binary* means that after i collisions, the random number of slot times chosen is between 0 and $2^i - 1$ inclusive. Therefore, as the number of retransmission attempts increases, the number of possibilities for the access delay increases exponentially. The term truncated means that after a certain number of retransmission attempts, the exponentiation ceases, i.e., the maximum backoff delay is $2^m - 1$ slot times where m is a specified number of attempts.

Random access systems with multiple orthogonal communication channels are called multichannel systems. Multichannel MAC protocols have emerged to satisfy the high bandwidth demands of next generation wireless communication systems. Multichannel slotted ALOHA with binary exponential backoff is an attractive random access technique. These systems are robust to failure of one or more channels, as the remaining channels can provide acceptable performance [9].

Generally, a wireless local area network (WLAN) has less than ten users so that collisions occur only occasionally. In this case backoff and retransmissions add marginal overhead which can be tolerated. If the number of users or access points increases to dozens or hundreds, many more users will collide, backoff, and retransmit data. As a result, network efficiency (throughput) is severely degraded, which results in reduced network capacity and noticeable delays for users. The IEEE 802.16 standard [10] (WiMAX) is designed to provide wireless access to a metropolitan area with thousands of users. Even with hundreds of users, using a CSMA-based protocol such as DCF to minimize collisions is not feasible. Thus, IEEE 802.16 employs a multichannel slotted ALOHA-based random access protocol with binary exponential backoff to provide resources in the form of network bandwidth. Bandwidth (BW) is granted by the base station (BS) to a subscriber station (SS) on a per connection basis in response to SS requests via a given request strategy. Thus WiMAX is termed a demand-assigned multiple access (DAMA) system.

Random access in IEEE 802.16 occurs in the request portion of the request-grant process for network users. BSs and SSs exchange data on frame basis. A portion of each frame is allocated to the contention-based initial access. This contention time interval known as the *random access channel* is divided into ranging (RNG) and BW request regions. RNG regions are used for initial network entry, ranging and power adjustments, while the BW request regions are used to send requests for bandwidth

for uplink data transmission [11]. In addition, best-effort (BE) data may be sent on the random access channel, but this is typically suitable for only small amounts of data. This data may also include additional requests for BW resources [12].

The random access process is slotted because of the WiMAX frame structure. Duplex operation is achieved either through frequency division duplexing (FDD) or time division duplexing (TDD). The latter technique is typically used. Communications between the BS and the SS is thus bidirectional, with an uplink (UL) channel (subframe) from SS to BS and a downlink (DL) subframe from BS to SS. Time in the UL channel is usually slotted (termed minislots) and shared using time-division multiple access, whereas on the DL channel the BS uses a continuous time-division multiplexing scheme. The duration of the downlink or uplink subframes in TDD mode is determined by the BS in a dynamic manner.

The Random Access Channel is located at the start of the uplink subframe, and it represents a multichannel system since it is divided into minislots. These minislots represent the transmission opportunities for BW requests and initial ranging (RNG) transmissions sent by the SSs. The transmission opportunity (TO) can take several forms such as time slots as in the case of fixed IEEE802.16(d). In the case of the mobile version IEEE802.16(e), the TO is subdivided into Orthogonal Frequency Division Multiplexing (OFDM) subcarriers. The Random Access Channel can also use subchannelization where selected subcarriers in a TO are grouped into a cluster (subchannels) to employ Frequency Division Multiple Access mechanisms [13].

In summary, each subframe consists of a number of time slots. The UL subframe is divided into a sequence of minislots which SSs can access in a synchronized manner under the control of the BS. Each SS that needs to send data in the uplink has to first request BW from the BS. The frame structure and BW request schemes in IEEE802.16 are illustrated and explained in more detail in Chapter 3. The most commonly used BW request technique is the *contention mode*. When collisions occur in contention mode, binary exponential backoff is specified in the standard as the collision resolution algorithm.

It is clear that the random access process and the binary exponential backoff contention resolution algorithm are key factors in the performance of most applications, particularly wireless networks. An efficient random access mechanism to improve connection setup, resource requests and best effort data transmission is an essential part of the design of a reliable communication system. Thus, this dissertation is dedicated to exploring random access and backoff mechanisms in multichannel multiple access

systems. It is important to note that much research has been done to model the random access channel, but little to model the random access process with respect to the user. For this reason the work presented in this dissertation focuses on modeling the random access and collision resolution from the user perspective. The results obtained can be used not only for initial access purposes, but also to provide reliable quality of service (QoS) as shown in Chapter 4.

1.3 Random Access and the Physical Layer

Traditionally, the underlying Physical (PHY) layer has been viewed as a black box that is completely separate from the MAC layer collision model. Including the physical channel properties such as the signal to noise ratio (SNR) and bit error rates (BER) in the MAC layer has been the focus of recent research in order to improve modeling of the multiple access process. Developing better models that combine communication and network theory still remains a challenge in a wide range of wireless communications research [14].

Combining the MAC and PHY layer properties is widely known as *cross-layer* design. As the word indicates, cross-layer design means a joint design of two or more layers to optimize system-wide performance via an exchange of parameters across the layers. For this reason, the foundation for development a cross-layer model for use in random access system design and performance evaluation is introduced in Chapter 5. Discretization of the physical channel at the packet level is the foundation for the combination of the MAC and PHY layers. This is mainly because at the MAC layer, data delivered from the underlying PHY layer is composed of blocks of data on which channel encoding and decoding, modulation and signal processing techniques have been applied to mitigate degradations due to noise, fading and inter-symbol interference. A first order finite state Markov channel (FSMC) model, also known as the Markov *block* fading model, is the most widely used technique for discrete modeling of the continuous wireless fading channel.

1.4 Dissertation Organization and Contributions

The work in this dissertation is organized as follows.

In Chapter 2, an extension to the simple model published in [15] is presented. This model accurately characterizes a multistage random access process from the

user perspective. The model is general and can efficiently be used to model a wide variety of MAC protocols that utilize a multichannel random access mechanism. An analytic Markov chain model for non-saturation conditions is developed to investigate the efficiency based on the average system throughput. A key contribution of this work is the use of a variable radix parameter in the backoff algorithm. Performance results show that an variable radix can significantly improve performance by reducing the amount of resources wasted during the backoff process. The model accuracy is confirmed by simulation results.

In Chapter 3, the variable radix backoff strategy introduced in Chapter 2 is further investigated in a practical application, namely the contention based BW request procedure in the IEEE 802.16 standard. A brief explanation of the IEEE 802.16 frame structure is presented. The BW request procedure is then analyzed in detail and extended to adopt the variable radix backoff strategy. Results show that the throughout can be significantly improved by varying the radix according to the number of contending nodes (SSs) and the available minislots in the random access channel. Simulation results are presented which confirm the accuracy of the analytic results.

In Chapter 4, a QoS technique for random access systems is presented. This novel approach to providing QoS divides the available resources into distinct service classes for contending users. Two classes of users are considered in this dissertation, however the model can easily be extended to any number of classes. The technique presented can provide a simple QoS mechanism for the contention-based bandwidth request (BW-REQ) process, and the best effort and non-real time data transmission classes in IEEE802.16. This mechanism is first implemented using a three-state Markov chain model [16] and [17]. It is then extended to exploit the variable radix strategy introduced in Chapter 2. Efficiency and average delay results show that this is an efficient QoS technique for multichannel random access systems. Simulation results are presented which confirm the accuracy of the results.

In Chapter 5, an optimized FSMC model for flat fading channels is developed. An equal average fade range duration (AFRD) partitioning methodology is used to discretize the signal to noise ratio (SNR) probability distribution function (PDF) to derive the Markov state thresholds. Since data transmission is packet-based, the SNR thresholds bounding each state are derived such that the average state duration (fade range) is a multiple of packet time units. This methodology provides a versatile technique to link the fade time to the packet duration. The model employs equal

duration partitioning and is applied to the Nakagami-m flat fading channel. The state bit error rate vector, steady state vector and transition probabilities are derived. The accuracy of the model is confirmed by simulation results for various fading intensities and Doppler frequencies.

A cross-layer Markov model which includes PHY layer errors based on the FSMC model is implemented. This model provides a foundation for studying the effects of practical (non-ideal) channel conditions on the random access process. This is achieved by adding an error state to the model introduced in Chapter 2 to incorporate channel errors which effect the system transitions.

In Chapter 6 some conclusions are presented as well as suggestions for future work.

The appendix provides pseudo code for the simulation tool that was developed to verify the analytic results in the dissertation.

Chapter 2

Backoff Strategies in Random Access Systems

2.1 Introduction

Random access is a critical part in any medium access protocol layer as it controls the initial user access to the system. It is the procedure that takes place during the initial network entry or connection setup of any multiuser system. Users randomly access the medium so that network resources can be acquired to transmit data. Depending on the available resources and the number of contending users, collisions are inevitable. The Binary Exponential Backoff (BEB) in which the backoff window is doubled at every contention attempt, is the most widely used contention resolution protocol to reduce the probability of collisions. It is usually combined with another contention resolution protocol such as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) used in the IEEE 802.11 Distributed Coordination Function (DCF) [8], [18], [7] and [19]. In IEEE 802.16 standard [10], BEB is the main contention resolution mechanism to resolve collisions during the contention interval in the uplink.

In this chapter, a general random access model is first presented. The model accurately characterizes the random access process for a wide range of multichannel systems. It is shown that the BEB, which exponentially increases the backoff window by a factor of 2, is not always a good solution to minimize collisions. Doubling the backoff window at every retransmission attempt may result in a significant amount unused resources, especially in a low to intermediate contention conditions. A variable radix parameter r is therefore introduced to the backoff procedure to provide a reli-

able adaptive strategy to minimize wasted resources. Performance results show that using a variable radix can significantly improve system performance by decreasing the unused resources corresponding to the retransmission attempts due to collisions.

2.2 Model Overview

The performance is best exemplified using a well known multichannel contention-based system, namely the bandwidth request process in IEEE 802.16 networks [10]. In these networks, the system is composed of a central controller or base station (BS) that controls a finite number of users or subscriber stations (SSs) N within its managed area. Data exchange between the BS and any SS is maintained via a request-grant protocol that is performed on a time division duplexing basis using Time Division Multiple Access (TDMA).

The system time is divided into frames, and each frame is subdivided into an uplink and a downlink subframes. If a particular SS wishes to acquire the medium to send data, it initially requests an adequate amount of bandwidth from the BS to serve its application [20]. This request strategy is performed by sending an access packet known as a *bandwidth request* (BW-REQ) message. The BW-REQ is a MAC-specific message sent during a dedicated contention interval at the start of an uplink subframe. Upon receiving a valid request in the uplink subframe, the BS grants the amount bandwidth sought by the SS in the subsequent downlink subframe. The BW grant in the DL subframe acts as the only acknowledgment by which an SS knows that its request has successfully reached the BS. In other words, if the SS does not receive a bandwidth grant during the subsequent downlink subframe, it assumes that the request has collided or no bandwidth is available.

The contention interval is divided into K resources. Each represents a transmission opportunity that can take several forms such as a time *slot* in an OFDM/TDMA system, or a subchannel in an Orthogonal Frequency Division Multiple Access (OFDMA) system [11]. In this chapter a transmission opportunity is referred to a time slot, simply termed as a *slot* henceforth.

Only one SS can acquire a slot at a time, i.e., a *success* occurs when only one SS chooses a free slot to access the medium at a given time step. A *collision* occurs when two or more SSs access the same slot simultaneously.

In Section 2.3, the SS analytical model represented by a finite state Markov chain is introduced. The proposed variable radix r is employed in deriving the steady

state solution as a function of the transition probabilities. Numerical results using are given in Section 4.3 to illustrate the system performance based on the average throughput[1]. To confirm the accuracy of the model, a simulation tool using object oriented programming is developed. An overview of the tool classes and associated methods is presented in Appendix A.

2.3 The Markov Chain Model

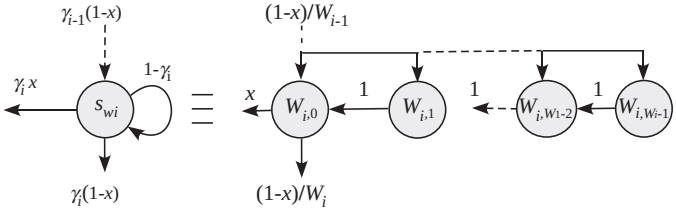


Figure 2.1: An i^{th} backoff window stage abstraction to a single wait state.

In [19], Bianchi presented a Markov model with i backoff stages, $0 \leq\ i \leq m$. Each stage has a backoff window W_i corresponding to the states at that stage. Transitions occur between states in every stage with probability 1. Before the first transmission attempt $i = 0$, an SS randomly chooses a uniformly distributed number (state) within the range $[0, W_0 - 1]$ and decrements its counter each slot and transmits when the counter reaches zero. If a collision occurs, the SS enters the next retrial stage using a binary exponential backoff (BEB) mechanism by doubling its backoff window so that $W_i = 2^i W_0$. If further collisions occur, this process continues until the last stage is reached with a maximum window size of W_m, where m is the maximum number of retrial stages. This backoff window scheme is employed in much of the literature in the area, but it results in a two dimensional chain with a large number of states. Since each stage i represents the time an SS waits until it begins the i^{th} transmission attempt, the backoff states at stage i can be represented by a single wait state s_{w_i} and a retransmission probability γ_i.

This abstraction as shown in Fig. 2.1, results in a one dimensional Markov chain providing a more tractable model especially when additional states are needed to model other system parameters such as channel errors or buffer occupancy conditions as will be shown in Chapter 5.

[1]Being the main performance measure in wide range of contention systems, the terms *average throughput* and *performance* are used interchangeably.

The relation between W_i and the retransmission probability γ_i can be derived as follows [21] Since on every stage, the location of each backoff state in the interval $[0, W_i - 1]$ is uniformly distributed, the average wait time at every stage is simply $\frac{W_i}{2}$. We have

$$\begin{aligned} \frac{W_i}{2} &= (1-\gamma_i)\gamma_i + 2(1-\gamma_i)^2\gamma_i + \cdots + (W_i-1)(1-\gamma_i)^{W_i-1}\gamma_i \\ &= \sum_{k=0}^{W_i-1} k(1-\gamma_i)^k\gamma_i \\ &= \frac{1-\gamma_i}{\gamma_i} - \frac{(W_i\gamma_i - \gamma_i + 1)(1-\gamma_i)^{W_i}}{\gamma_i} \end{aligned} \tag{2.1}$$

For a sufficiently large initial backoff window size (e.g., $W_0 \geq 32$) [18], [22], $(1-\gamma_i)^{W_i} << 1$ in (2.1), making the second term on the right-hand side ≈ 0. The retransmission probability at stage i can therefore be approximated by

$$\gamma_i \approx \frac{2}{W_i + 2} \tag{2.2}$$

Instead of doubling the backoff window (waiting interval) after every unsuccessful retransmission attempt [19], a variable radix r is introduced such that $W_i = r^{i-1}W_0$ which can take values other than 2. Therefore, we have

$$\begin{aligned} \gamma_{i+1} &= \frac{\gamma_i}{r} \\ \gamma_i &= \frac{\gamma_1}{r^{i-1}} \qquad 1 \leq i \leq m \end{aligned} \tag{2.3}$$

In a multiuser system, the current state of a request sent by an SS during the random access process only depends on the previous state, therefore the process can be efficiently modeled as a discrete time Markov chain (DTMC) system [23].

Since the focus in this chapter is improving the performance of the random access process in multichannel systems, for the sake of simplicity and clarity in the analysis, the following assumptions are adopted.

1) The system is frame synchronized, and thus the current system state is determined by the requests acknowledged by the BS at the end of the downlink subframe.

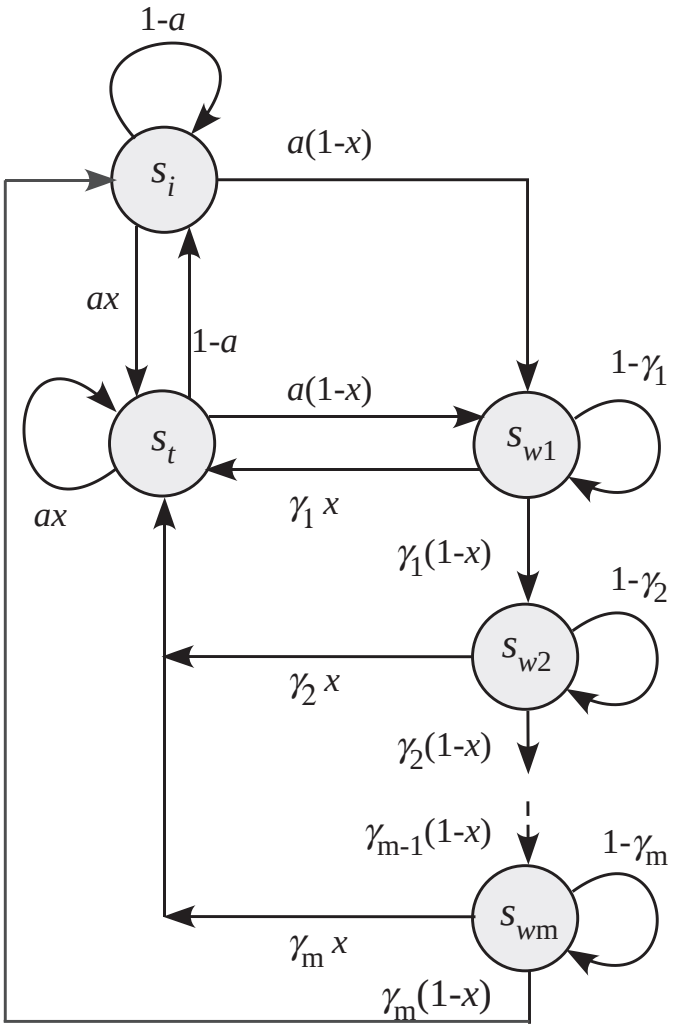


Figure 2.2: The Markov chain model of the contention interval.

2) There is abundant bandwidth, i.e., if a request successfully reaches the BS, a bandwidth grant is guaranteed.

3) The channel conditions are ideal, i.e., failure of a SS to receive a specific grant by the BS is only due to a collision of a corresponding request. Consequently, a success signifies that the request is granted.

4) If collision occurs in the last (m^{th}) retransmission attempt, the packet is discarded and the SS returns to the idle state s_i.

5) Starting from an idle state s_i, a SS does not wait before making the first transmission attempt as in [19]. This is reasonable and practical since here

a non-saturation condition is assumed so that users begin transmitting requests randomly. Thus $W_0 = 0$ and $W_i = r^{i-1}W_1$, $1 \leq i \leq m$.

Figure 2.2 presents the system Markov model. The system is assumed to have a one packet transmission buffer where requests arriving during a frame interval are Bernoulli distributed with an arrival probability a. The probability of success is x indicating that the request has successfully reached the BS. Based on the assumptions above, this also indicates that a grant has been received by the SS and thus the system migrates to the transmit state s_t. If a collision occurs, the system enters the first wait state s_{w_1} where the retransmission probability is γ_1. If another collision occurs, the SS enters the next state of a finite retrial phase consisting of m wait stages $s_{w_{1\ldots m}}$. If a collision occurs after m retransmission attempts (state s_{w_m}), the request packet is discarded and the SS exits the contention process by returning to the idle state s_i. At the i^{th} retransmission attempt, the system exits the contention process from state i with probability $\gamma_i x$ and migrates to the transmit state s_t.

At steady state, the system balance equations are given by

$$\begin{aligned}
s_{w_1} &= \frac{a(1-x)}{\gamma_1}\,(s_i + s_t) \\
s_{w_2} &= \gamma_1(1-x)s_{w_1} + (1-\gamma_2)s_{w_2} \\
&= \frac{\gamma_1(1-x)}{\gamma_2}s_{w_1} \\
&= \frac{a(1-x)^2 r}{\gamma_1}\,(s_i + s_t) \\
&\vdots \\
s_{w_i} &= \frac{a(1-x)^i}{\gamma_1}\,r^{i-1}\,(s_i + s_t) \qquad 1 \leq i \leq m
\end{aligned} \tag{2.4}$$

From (2.4) the wait states are given by

$$\begin{aligned}
\sum_{i=1}^{m} s_{w_i} &= \frac{a}{\gamma_1}\,(s_i + s_t) \sum_{i=1}^{m} (1-x)^i r^{i-1} \\
&= \frac{a}{\gamma_1 r}\,(s_i + s_t) \sum_{i=1}^{m} (r - rx)^i \\
&= \frac{a}{\gamma_1}\,(s_i + s_t) \left[\frac{1-x}{rx - r + 1} - \frac{(r-rx)^{m+1}}{r(rx - r + 1)} \right] \\
&= \frac{a}{\gamma_1}\,(s_i + s_t)\; Q
\end{aligned} \tag{2.5}$$

where $Q = \frac{r(1-x)-(r-rx)^{m+1}}{r(rx-r+1)}$

Using the terms s_{w_1}, s_{w_2}, ..., s_{w_m} in (2.4), the transmit state s_t is given by

$$\begin{aligned}
s_t &= axs_t + axs_i + \gamma_1\, x\, s_{w_1} + \gamma_2\, x\, s_{w_2} + \cdots + \gamma_m\, x\, s_{w_m} \\
&= ax\,(s_i + s_t) + ax(1-x)\,(s_i + s_t) \\
&+ ax(1-x)^2\,(s_i + s_t) + \cdots + ax(1-x)^m\,(s_i + s_t) \\
&= ax\,(s_i + s_t) \sum_{i=0}^{m} (1-x)^i \\
&= a\,(s_i + s_t)\left[1 - (1-x)^{m+1}\right] \\
&= \frac{a\,G\,s_i}{1 - a\,G}
\end{aligned} \tag{2.6}$$

where $G = 1 - (1-x)^{m+1}$

Using the normalization condition $\sum_j s_j = 1$ we have

$$\begin{aligned}
1 &= s_i + s_t + \sum_{i=1}^{m} s_{w_i} \\
&= s_i + \frac{a\,G\,s_i}{1 - a\,G} + \frac{a}{\gamma_1}\,Q\,s_i \left[1 + \frac{a\,G}{1 - a\,G}\right]
\end{aligned} \tag{2.7}$$

Substituting for s_i from (2.7) into (2.6) gives

$$s_t = \frac{a\,G}{1 - a\,G} \left[\frac{\gamma_1\,(1 - aG) + \gamma_1\,a\,G + a\,Q}{\gamma_1(1 - aG)} \right] \tag{2.8}$$

With a finite population of N equal priority SSs, a particular SS successfully

acquires a slot if all other contending (active) SSs do not choose this particular slot in the same frame period [23]. ¿From the MAC perspective, the system throughput is defined as the average number of successful transmissions per network unit time (frame). According to the Markov model in Fig. 2.2, the system throughput is the average number of SSs that are in the transmit state s_t during a given frame period $N\,s_t$. The average input traffic is given by the average request arrival per frame $N\,a$

The probability of activity p is defined as the probability an SS has a request to send at the start of a frame period. This is given by

$$\begin{aligned} p &= s_i a + s_t a + \gamma_1 s_{w_1} + \gamma_2 s_{w_2} + \cdots + \gamma_m s_{w_m} \\ &= a\,(s_i + s_t) \sum_{i=1}^{m} (1-x)^i \\ &= a\,(s_i + s_t) \left[1 + \frac{1 - x - (1-x)^{m+1}}{x} \right] \end{aligned} \tag{2.9}$$

The success probability x for a SS is defined as the probability that it acquires one of the K slots during a frame period. Equivalently, none of the remaining SSs accesses the same slot in that frame period. This probability is given by [21]

$$x = \left(1 - \frac{p}{K}\right)^{N-1} \tag{2.10}$$

Assuming an initial value of x and solving (2.10) with (2.6) and (2.8) numerically, the system throughput is obtained.

2.4 Numerical Results

In this section, the effect of the proposed variable radix startegy on the system performance is evaluated using the average system throughput. Different traffic levels are represented by the average request arrival from different numbers of SSs N. The number of slots in the contention interval is fixed at $K = 16$, and the maximum number of retransmission attempts is $m = 5$. The initial backoff window is $W_1 = 32$, giving a retransmission probability of $\gamma_1 \approx 0.06$. Simulation results which confirm the analytical results were obtained using the OOP Matlab tool in Appendix A. The results in the following figures are based on an average of $20,000$ trials.

In the following results, the benchmark for comparison is the BEB performance ($r = 2$), which is denoted by a solid line in the figures (except Fig. 2.3 (d)).

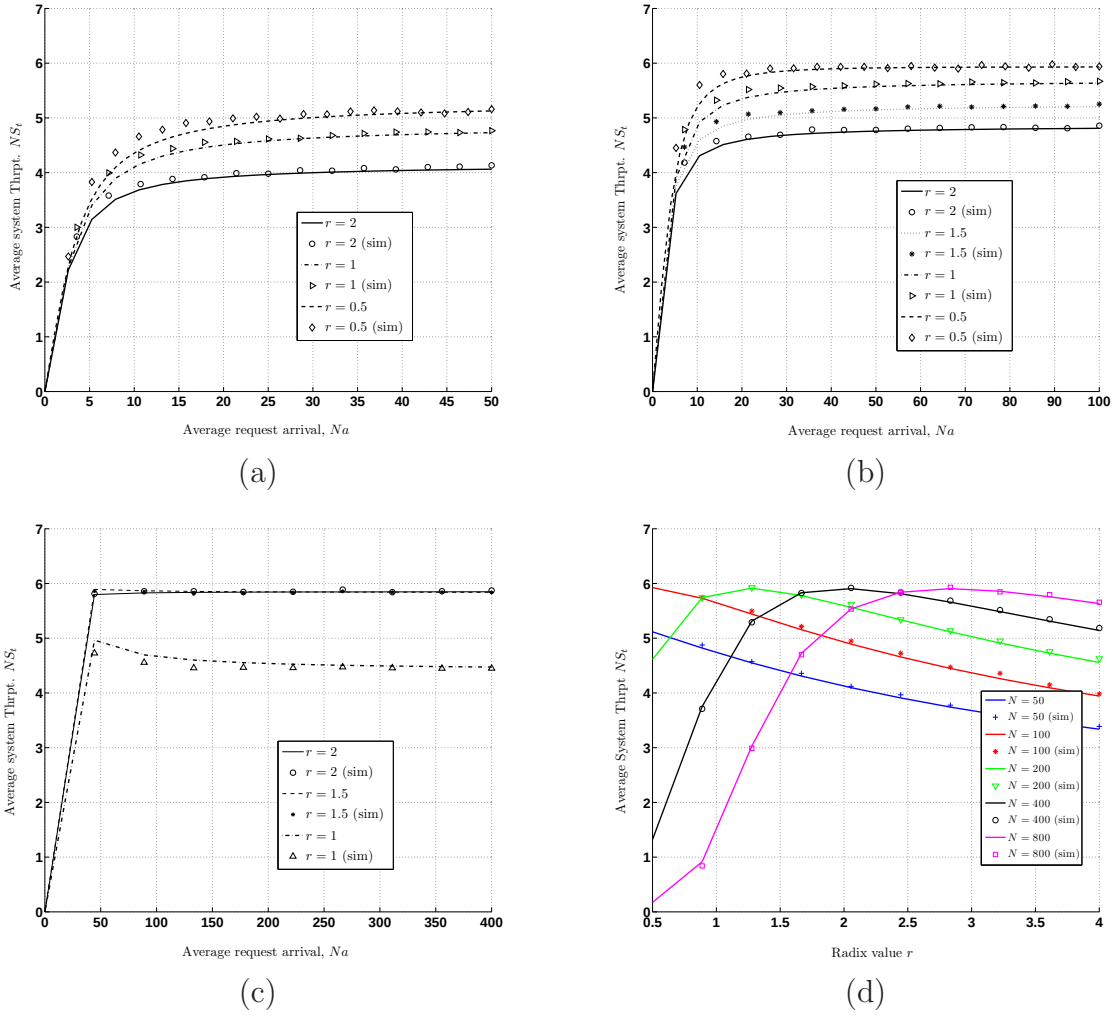


Figure 2.3: System throughput for different average request arrival and radix values.

Figure 2.3 (a) shows that over the range of traffic loads examined $0 \leq Na \leq 50$, the system throughput is monotonically increasing. This indicates that at this contention level, i.e., for the given values of N and K, the system efficiently accommodates retransmissions resulting from collided requests and thus increasing the average throughput. Three radix values were used: 2 (BEB), 1 and 0.5. We notice that with $r = 1$ (fixed window size), the system throughput is increased by 18% over BEB. This indicates that doubling the backoff window size W_i, (decreasing the probability of retransmission γ_i) , has a negative effect on the system performance due the increased number of unused slots per frame. In other words, better slot utilization can be achieved by fixing the window size rather than increasing it by a factor of 2 as

in the case of BEB. An interesting result is even observed at $r = 0.5$. In this case, the backoff window is reduced every stage as opposed to being increased with $r > 1$. Clearly, this reduction provides an increasing retransmission probability at every retransmission stage however, the system throughput is further increased providing a 20% improvement over BEB. This shows that for the given values of N, K and W_1, the system is in a low contention environment that can tolerate a more aggressive retransmissions. The improvement in performance indicates that it is using a smaller initial backoff window W_1 is recommended. However, decreasing W_1 on individual basis may not be practically feasible, as the window size would have to be dynamically assigned by the BS to each SS. Conversely, a global radix can easily be varied according to the traffic environment by each SS as an adaptive strategy. In addition, selected or polled users can individually choose their radix accordingly to provide a given QoS level with a given W_1 which is fixed for the entire network. This method will be investigated in more detail in Chapter 4.

In Figure 2.3 (b), a medium contention system is shown with $N = 100$, and three radix values 1.5, 1 and 0.5 are compared with BEB ($r = 2$). Note that the system performance increases as r decreases. It is shown that using $r = 1.5$ and $r = 1$ provides an improvement in the system throughput of approximately 15% and 20%, respectively, over BEB. At $r = 0.5$ a marginal increase in performance is achieved compared to $r = 1$ however, the performance does not increase for $Na \geq 50$. This indicates that retransmissions are to overwhelm (saturate) the system which indicates that a further decrease in r should not be used.

A heavily populated system is shown in Fig. 2.3 (c) for three radix values 2, 1.5 and 1. Note that using $r = 1.5$ provides the same performance as BEB, indicating that a further decrease in r will not improve the system performance. This is verified with a non increasing window strategy ($r = 1$), which results in a performance degradation of 25%. This indicates that the system is becoming *unstable* [23], and this contention level, r should be increased to reduce the probability of collisions.

A perspective view of the radix strategy is given in Figure. 2.3 (d). The system performance using different radix values for five system population values $N = 50$, $N = 100$, $N = 200$ and $N = 400$ is shown. At a medium load ($N = 50$ and 100), the system can easily accommodate the larger number of retransmissions at lower radix values $r < 1$. As r increases, the performance degrades by 20% and 35%, respectively at BEB due to the large number of unused slots at this contention level. At a higher load ($N = 200$), performance peaks at $r = 1.25$ and decreases by 8% with BEB. At

$N = 400$, BEB gives the best performance which indicates that BEB is a good choice at moderate to heavy contention levels. However, a higher radix than BEB is required in a very heavily loaded system ($N = 800$), where best performance is achieved with $r = 3$ where the loss in performance with BEB is approximately 8%.

The moving peak in Figure. 2.3 (d) representing the maximum performance indicates that r should be adjusted according to the system population. While BEB may be a good overall conservative choice, significant gains can be made by allowing a variable radix.

2.5 Summary

In this chapter, a new variable radix strategy to improve the system performance of multichannel random access systems was presented. An analytic Markov chain was introduced to model a multichannel contention system. The model presented provides an accurate study of the performance of a wide range the random access systems during the contention phase. Results using this model show that varying the radix r according to the number of contending stations can significantly improve performance compared to the binary exponential backoff algorithm. An efficient and general simulation toolbox was developed to verify the analysis using OOP. It was shown that a variable radix can seamlessly be applied to most multichannel random access systems such as the contention-based bandwidth request mechanism in IEEE802.16, as will be shown in Chapter 3.

Chapter 3

Variable Radix Backoff in IEEE 802.16

3.1 Introduction

In this chapter, a necessary study is performed to bring the variable radix technique implemented in Chapter 2 to a real and practical environment. Although the IEEE 802.16 services are intended to be minimal to contention-free, random access is still an essential process to perform network entry procedures. In addition, random access still plays a major role in non-real time and best effort data delivery. The Binary Exponential Backoff is utilized by the standard as the only contention resolution algorithm to resolve collisions. Performance results show that by varying the radix parameter, the bandwidth request performance can be improved significantly. Simulation results are presented to validate the accuracy of the presented model analysis.

3.1.1 IEEE 802.16 and Frame Sructure

The IEEE 802.16 (WiMAX) standard is a cost-effective solution to providing last-mile wireless services to end users. In this standard, a base station allocates wireless network resources (channel bandwidth) to mobile stations via a request mechanism. Dedicated time slots within the uplink subframe are available on a contention basis for mobile stations to transmit bandwidth request messages. Since mobile stations contend to send messages during this period, collisions are inevitable. According to the standard, subscriber stations follow the binary exponential backoff (BEB) protocol to resolve collisions.

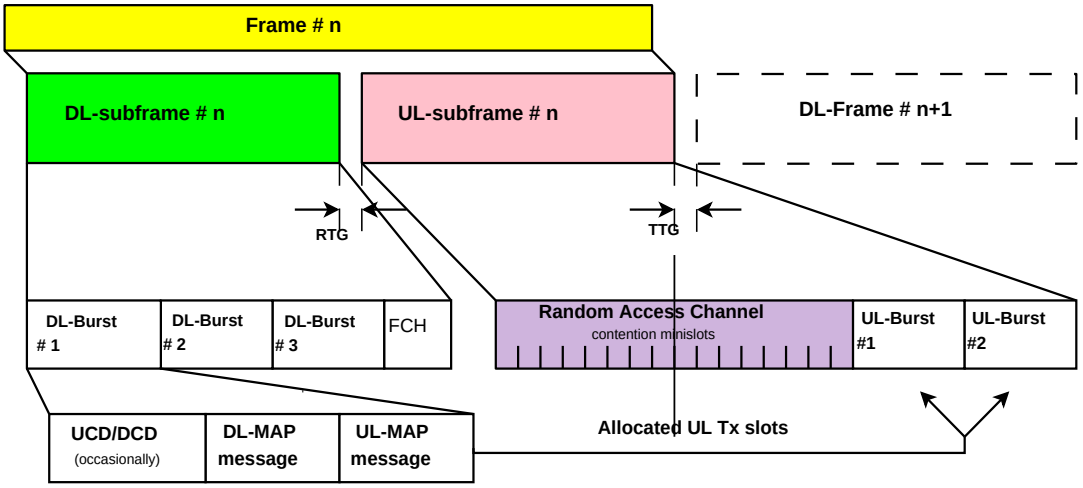


Figure 3.1: The IEEE 802.16 frame structure.

The IEEE 802.16 standard [10] defines two main operational modes, a mandatory point-to-multipoint (PMP) mode and a mesh mode. In PMP mode, a central base station (BS) controls a group of subscriber mobile stations (SSs) whereas in mesh mode, SSs manage a cooperative access and routing protocol in a distributed or self-organizing manner. In IEEE 802.16 PMP mode, channel resources in the form of bandwidth (BW) are allocated by the BS to SSs on a demand-grant basis. An SS reserves BW by sending a request to the BS. Upon receipt of a bandwidth request, the BS determines the amount of BW to be granted, schedules the grant confirmation, and thus grants a transmission opportunity to the SS.

The BW grant mechanism is implementation dependent, so the BS grants resources to the SS according to a given scheduling algorithm and the available channel resources at the time of the request. To explain this mechanism, a brief description of the WiMAX frame structure is given below.

In figure 3.1 a simplified IEEE 802.16 frame structure is shown. Communication between the BS and SSs is a bidirectional exchange of frames using time division multiple access (TDMA), and time division duplexing (TDD) via uplink (UL) and downlink (DL) subframes [11].

The transmission gaps TTG and RTG are inserted to allow the SSs to transition between transmit and receive modes.

The downlink subframe contains the downlink and uplink medium access (MAC) protocol messages, DL-MAP and UL-MAP, respectively. These broadcast MAC messages indicate the position of the time slots allocated to the SSs in the UL and DL

subframes. The uplink slots are used to send data to the BS, and data is received from the BS in the downlink slots.

The UL subframe contains a random access channel (RACH) which is the contention period in the IEEE 802.16 frame structure. The RACH is divided into a number of transmission opportunities or time minislots in which ranging request (RNG-REQ) and bandwidth request (BW-REQ) messages can be transmitted. RNG-REQ is the request for initial network entry to obtain system information such as distance from the BS, modulation, and error-correction coding. BW-REQ is used by the SSs to request bandwidth for data transmission[1].

3.1.2 Bandwidth Request in IEEE 802.16

The IEEE 802.16 standard defines two BW-REQ mechanisms.

1) A contention-free mode in which BW requests are sent in pre-assigned time slots to the BS in the UL subframe. It can also be in the form of piggybacking on UL data slots as SSs are allowed to send BW requests in these dedicated time slots.

2) A contention-based mode where SSs contend to send BW requests during the dedicated RACH in the UL subframe. Each minislot in the RACH can accommodate only one BW request. If more than one SS transmits in the same minislot, a collision occurs.

The focus in this chapter is on contention based BW requests.

Similar to the distributed coordination function (DCF) in the IEEE 802.11 standard, contention resolution is regulated using a variable transmission window based on a truncated binary exponential backoff (BEB) mechanism [7], [24]. The minimum and maximum backoff window sizes W_{min} and W_{max}, respectively, are defined in the uplink channel descriptor (UCD) message which is transmitted periodically (every 10 s) or aperiodically in the DL subframe.

Unlike IEEE 802.11 [18], 802.16 random access does not define an acknowledgement mechanism to indicate that the BS has received a BW request. This is because

[1]It is important to note that, due to the difference in the timing procedure and to be consistent with the notation in [10], using the CI notation for the the general model presented Chapter 2 might be inaccurate. Therefore in this model, the CI notation is referred to as the **Random access Channel** (RACH)

in IEEE 802.16, it is not possible to sense the channel using an approach such as carrier sense medium access (CSMA) due to the large number of SSs widely spread in an area that can reach several kilometers. Instead, the standard defines a time out period $T16$ with a minimum one frame duration of 10 ms. If an SS does not receive a response to a request after the $T16$ expires, the BW-REQ message is assumed to have been lost, collided, or no bandwidth is available to grant. In all three cases, the SS reenters the contention process according to the BEB mechanism.

The IEEE 802.16 BEB algorithm specifies that each SS chooses a random integer uniformly from the interval $[0, W_{i-1}]$, where i denotes the i^{th} transmission attempt. Thus, before the first transmission attempt, the SS chooses a random integer from the interval $[0, W_0]$ where $W_0 = W_{min}$. In the case of an unsuccessful request, the SS doubles the backoff window and retransmits the request. Thus, after the i^{th} unsuccessful request, the window size is $W_i = 2^i W_{min}$. This doubling continues until the maximum window size (defined in the UCD) is reached. $W_{max} = 2^m W_{min}$, so after the m^{th} transmission stage the window size does not increase and the SS remains in this stage. Once a request is successful, the SS exits the contention process and the backoff window is reset to the initial value W_{min}.

Some related work in the literature is presented here. In [12], a variable radix in a multichannel random access environment was introduced. It was shown that a variable radix can significantly improve the system performance for a range of active users in the system. In [25], [26], the authors studied the performance of the BW request mechanism based on varying the number of minislots for a fixed number of contending users (SSs) under saturation conditions. An extension to this work [27], considered the non-saturated conditions based on a Bernoulli arrival distribution. Fallah *et. al* [22] introduced the first analytic model of the BW request process in IEEE 802.16 using a discrete time Markov chain. A two-dimensional Markov model was developed with backoff and wait planes for various load conditions. Hossam *et. al* [28] extended the results in [22] by introducing the concept of subchannelization for BW requests. They showed that the throughput and capacity can be improved significantly by using subchannelization.

In this chapter, the results in [22] are extended and improved by introducing a variable radix parameter r where $r \neq 2$. It is shown that using a radix value other than the conventional binary value $r = 2$ can improve system performance. A variable r can be used to provide an adjustable backoff strategy to achieve better RACH performance according to the request traffic intensity.

In Section 3.2, the system is described and the BW request process is explained. This process is modelled using a discrete time Markov chain, and this chain is used to obtain the steady probabilities. Section 4.3 provides some analytic and simulation performance results, and finally a chapter summary is given in Section 3.4.

3.2 Contention-Based BW-REQ Markov Model

As shown in Fig. 3.1, a random access channel (RACH) is allocated by the BS in each uplink subframe for SSs to transmit BW requests. The RACH is divided into N minislots. When an SS wishes to send a BW request, it chooses a number randomly from within its assigned backoff window size. This represents the number of minislots the SS must wait before sending its request. The SS decrements a counter that starts at this number, and transmits in the frame at which the counter reaches zero.

After sending a request, an SS waits for a maximum of M frames (excluding the current frame), to receive an uplink grant from the BS. If N_r is the number of minislots remaining in the current frame, the SS must wait for $N_r + NM$ minislots before declaring a BW request failure (the BW-REQ message was collided, lost or no BW is available). At the end of the M^{th} frame, if no grant has been received, the SS re-enters the contention process by increasing its backoff window and retransmitting the request.

The BW request process Markov model is shown in Fig. 3.2. In this model, saturation is assumed so that every SS has a request to send. Although this assumption does not hold for all applications, it is useful when investigating the system performance under high traffic conditions where requests are generated on a continuous (persistent) basis. The contention interval is divided into two planes. The *backoff* plane is denoted by circular states in Fig. 3.2, and the *wait* plane is denoted by rectangular states.

At the beginning of the i^{th} transmission of a request, the SS chooses a random number from the interval $[0, W_{i-1}]$, which corresponds to a backoff state $b_{i,k}$ in the backoff plane. The probability of being in this state is $1/W_i = 1/r^i W_0$. The SS decrements its counter each minislot, so the transition probability between states in the backoff plane is 1. When the SS counter reaches zero, the request is transmitted at state $b_{i,0}$. After transmission, the SS enters the wait plane represented by the rectangular states. There are two states for each stage i of the wait plane. The upper states represent a non-collision situation where the BW request is assumed by the SS

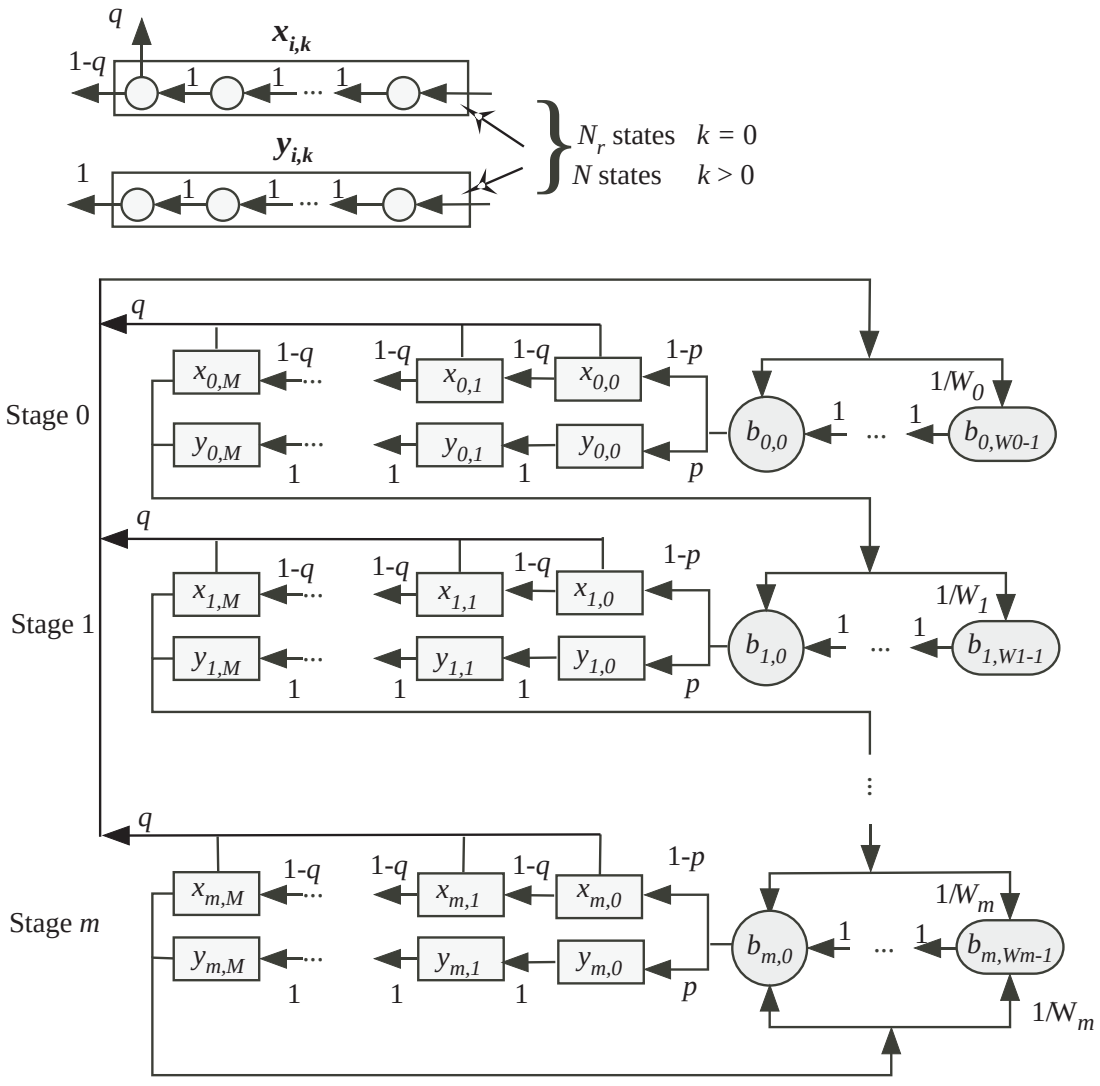


Figure 3.2: The IEEE 802.16 bandwidth request Markov chain model.

to have been received by the BS. The SS is then in wait state $x_{i,k}$, and the BS will grant the bandwidth with probability q. Thus a transition occurs to the next wait state with probability $1-q$, corresponding to no bandwidth being granted in that particular frame. If the BW is granted, the SS exits the contention process and starts a new transmission attempt at the first stage in the backoff plane.

The lower state of a wait stage represents a collision. In this case, the SS enters wait stage i according to the probability of collision, p. The SS is then in wait state $y_{i,k}$, where transitions occur between states with probability 1.

The wait states $x_{i,k}$ and $y_{i,k}$ represent a frame duration, so each wait state constitutes N circular (minislot) states, which is the number of minislots in the contention period. Thus, transitions occur between these states with probability 1. The wait plane contains M or $M+1$ states which is equal to the wait period in frames after which the SS considers the BW request to have failed.

If no grant is received after the last frame at stage i, the SS enters the next request stage $i+1$ (lower row in the Markov chain) at a random backoff state in the backoff plane with probability $1/r^{i+1}W_0$. With repeated request failures, the maximum backoff window size W_{max} is reached in the last (mth) stage, after which there is no further increase in the window size.

Because of the size of the 2-dimensional Markov chain in Fig. 3.2, the system has a very large number of unknowns. Therefore, the balance equation technique is employed to obtain a set of homogeneous equations [21]. This can be done by representing all states in stage i in terms of state $b_{i,0}$ where $0 \le i \le m$. This yields a set of $m+1$ linear equations with $m+1$ unknowns.

From the Markov chain in Fig. 3.2, we have

$$b_{i,k} = \frac{W_i - k}{W_i} b_{i,0} \qquad 0 \le k < W_i \tag{3.1}$$

$$x_{i,k} = (1-p)(1-q)^k b_{i,0} \qquad 0 \le i \le m,\ \ 0 \le k \le M \tag{3.2}$$

$$y_{i,k} = p b_{i,0} \qquad 0 \le i \le m,\ \ 0 \le k \le M \tag{3.3}$$

The system balance equations are derived as follows. For stage 0, we have

$$b_{0,0} = q \sum_{i=0}^{m} \sum_{k=0}^{M} x_{i,k} \tag{3.4}$$

¿From (3.2), state $b_{0,0}$ is given by

$$\begin{aligned} b_{0,0} &= \sum_{i=0}^{m}\sum_{k=0}^{M}(1-p)q(1-q)^k b_{i,0} \\ &= (1-p)\left[1-(1-q)^{M+1}\right]\sum_{i=0}^{m} b_{i,0} \end{aligned} \tag{3.5}$$

For all other stages, state $b_{i.0}$ is represented by

$$b_{i,0} = (1-q)x_{i-1,M} + y_{i-1,M} \qquad 1 \le i \le m \tag{3.6}$$

and from (3.3), $b_{i.0}$ is given by

$$b_{i,0} = \left[(1-p)(1-q)^{M+1} + p\right] b_{i-1,0} \qquad 1 \le i \le m \tag{3.7}$$

Equations (3.5) and (3.7) constitute a set of $m+1$ equations in $m+1$ unknowns. However, the nature of the Markov chain produces a rank-deficient system of equations which results in more than one possible solution. This can be solved by replacing one of the system equations (any row of the system matrix), with an independent equation based on the normalization condition to produce a unique solution.

From the Markov model, this equation is given by

$$\sum_{i=0}^{m}\sum_{k=0}^{W_i-1} b_{i,k} + \sum_{i=0}^{m}(x_{i,0}+y_{i,0}) + \sum_{i=0}^{m}\sum_{k=1}^{M}(x_{i,k}+y_{i,k}) = 1 \tag{3.8}$$

Substituting (3.1), (3.2) and (3.3) in (3.8) gives

$$\begin{aligned} \sum_{i=0}^{m}\left(\frac{W_i+1}{2}\right) b_{i,0} + \left\{N_r + MNp + N\sum_{k=1}^{M}\left[(1-p)(1-q)^k\right]\right\}.\sum_{i=0}^{m} b_{i,0} &= 1 \\ \sum_{i=0}^{m}\left\{\left(\frac{W_i+1}{2}\right) + N_r + MNp + N(1-p)\left[\frac{1-q-(1-q)^{M+1}}{q}\right]\right\} b_{i,0} &= 1 \end{aligned} \tag{3.9}$$

Let the vector $\mathbf{b}$ represent the $m+1$ unknowns that are obtained by solving the system of $m+1$ equations. Therefore we have

$$\mathbf{b} = \left[\begin{array}{cccc} b_{0,0} & b_{1,0} & \ldots & b_{m,0} \end{array}\right]^t \tag{3.10}$$

where t denotes transpose. The system of linear equations can be expressed as

$$\mathbf{A}\,\mathbf{b} = \mathbf{c} \tag{3.11}$$

where $\mathbf{A}$ is the system matrix and $\mathbf{c}$ is a column vector. This set of equations is homogeneous if $\mathbf{c} = \mathbf{0}$. The last row of $\mathbf{A}$ is replaced by the normalization equation.

From the above set of balance equations, the *non-zero* elements of $\mathbf{A}$ are given by

$$\begin{aligned}
a(0,0) &= 1-(1-p)\left[1-(1-q)^{M+1}\right] && (3.12)\\
a(0,i) &= -(1-p)\left[(1-(1-q)^{M+1}\right] \qquad 0<i\le m && (3.13)\\
a(i,i-1) &= -(1-p)(1-q)^{M+1}-p \qquad 1\le i<m && (3.14)\\
a(i,i) &= 1 \qquad 1\le i<m && (3.15)\\
a(m,i) &= \left(\frac{W_i+1}{2}\right)+N_r+MNp \\
&+ N(1-p)\left[\frac{1-q-(1-q)^{M+1}}{q}\right] \qquad 0\le i\le m && (3.16)
\end{aligned}$$

Therefore the vector $\mathbf{c}$ with $m+1$ elements has the form

$$\mathbf{c} = \left[\begin{array}{cccc} 0 & 0 & \ldots & 1 \end{array}\right]^t \tag{3.17}$$

On average, during the backoff process, a random number is selected from the interval $[0, \overline{W}-1]$, where $\overline{W}$ is the expected width of the contention window. This spans a number of contention periods and the last contention period may only be partial. If the value of the randomly selected backoff counter is in a contention period completely within $\overline{W}$, the average number of wait minislots is $N/2$. However, when this random number falls in the last partial contention period, the average number of wait minislots is

$$N_X = \frac{\overline{W}-\left\lfloor \overline{W}/N \right\rfloor N}{2} + \left(\left\lceil \overline{W}/N \right\rceil N - \overline{W}\right). \tag{3.18}$$

The probability of the former case occurring is given by

$$P_N = \left\lfloor \frac{\overline{W}}{N} \right\rfloor \frac{N}{\overline{W}} \tag{3.19}$$

while the probability of the latter case is

$$P_X = \frac{\overline{W} - \left\lfloor \overline{W}/N \right\rfloor N}{\overline{W}} \tag{3.20}$$

The probability averaged over both cases is equal to the average number of minislots remaining in the current frame N_r, so that

$$N_r = P_X N_X + P_N \frac{N}{2}. \tag{3.21}$$

Hence to calculate the remaining minislots in the current frame N_r, the expected value of the contention window $\overline{W}$ must be obtained. This is equal to the weighted sum (average) of all backoff states over all stages

$$\begin{aligned} \overline{W} &= \sum_{i=0}^{m} \sum_{k=0}^{W_{i-1}} k b_{i,k} \\ &= \sum_{i=0}^{m} \left(\frac{W_i^2 - 1}{6} \right) b_{i,0} \end{aligned} \tag{3.22}$$

where $W_i = r^i W_0$.

Having obtained the SS transmission probabilities at every stage $b_{i,0}$, the probability that an SS transmits in a given minislot is given by

$$\tau = \sum_{i=0}^{m} b_{i,0} \tag{3.23}$$

Considering a given SS, the conditional collision probability p is defined as the probability that one or more of the other SSs transmits in a given minislot [23]. This is given by

$$p = 1 - (1 - \tau)^{n-1} \tag{3.24}$$

where n is the total number of SSs controlled by the given BS.

From the MAC point of view, the throughput (performance) P_{TH} is defined as the probability that a BW request has been successfully received by the BS $P(A)$ and there is enough BW to grant the request in one of the M frames in the wait period

with a probability $P(B)$. These probabilities are given by

$$\begin{aligned} P(A) &= n\tau(1-\tau)^{n-1} && (3.25) \\ P(B) &= q + (1-q)q + (1-q)^2 q + \ \ldots (1-q)^M q \\ &= \sum_{k=0}^{M} q(1-q)^k && (3.26) \end{aligned}$$

Therefore, the RACH throughput is given by

$$\begin{aligned} P_{TH} &= P(A)P(B) \\ &= n\tau(1-\tau)^{n-1} \,.\, (1-(1-q)^{M+1}) && (3.27) \end{aligned}$$

3.3 Performance Results

To determine P_{TH}, (3.11) and (3.23) are solved numerically to obtain τ in (3.27). Then P_{TH} versus the number of contending SSs n is investigated while varying the radix r. the initial backoff window is $W_0 = 32$ with $m = 5$. The system is examined for different values of request grant probability q, $T16$ timeout period in frames M, and number of minislots in the RACH N. The benchmark for comparison is the performance with the binary exponential backoff ($r = 2$), which is denoted by a solid line in the figures. Simulations results to confirm the analysis were obtained based on an average of $20,000$ frames.

In Fig. 3.3, the system behavior with $N = 20$ is illustrated for a wide range of SS values n. Five values of r 3, 2.5, 2 (BEB), 1.5 and 1 are considered. The bandwidth grant probability is $q = 0.5$ and $M = 4$. this figures shows that in the low contention range $n < 50$, performance increases as r decreases. This is because a lower radix r provides a backoff window which increases more slowly. This decreases the number of unused slots at every stage during contention. As n increases, contention increases and a higher radix is needed to minimize collisions. In the range of approximately $150 < n < 200$, BEB gives the best results. When $n > 200$, performance degrades with BEB and is outperformed by $r = 2.5$. In a densely populated system $n > 400$, a higher radix than BEB is necessary to provide good performance. Note that $r = 3$ provides the best performance at $n > 450$, and is significantly better than BEB.

In Fig. 3.4 three values of r are used, 2 (BEB), 1.5 and 1, with $q = 0.5$ and $N = 20$ while varying the timeout period M. Figure 3.4 (a) shows that at $M = 4$,

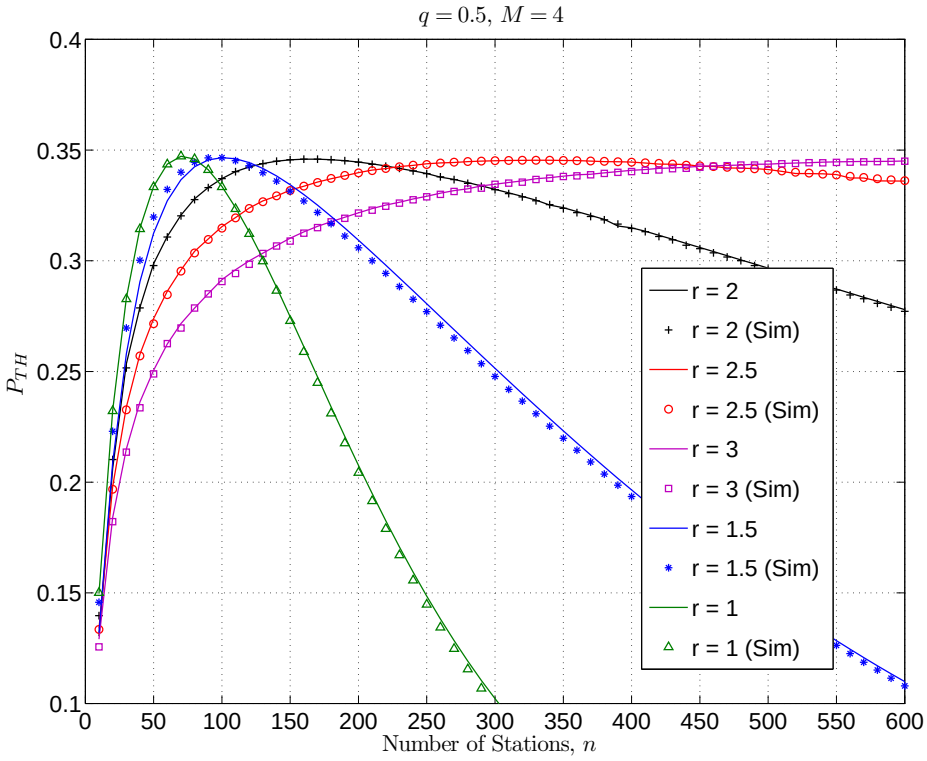
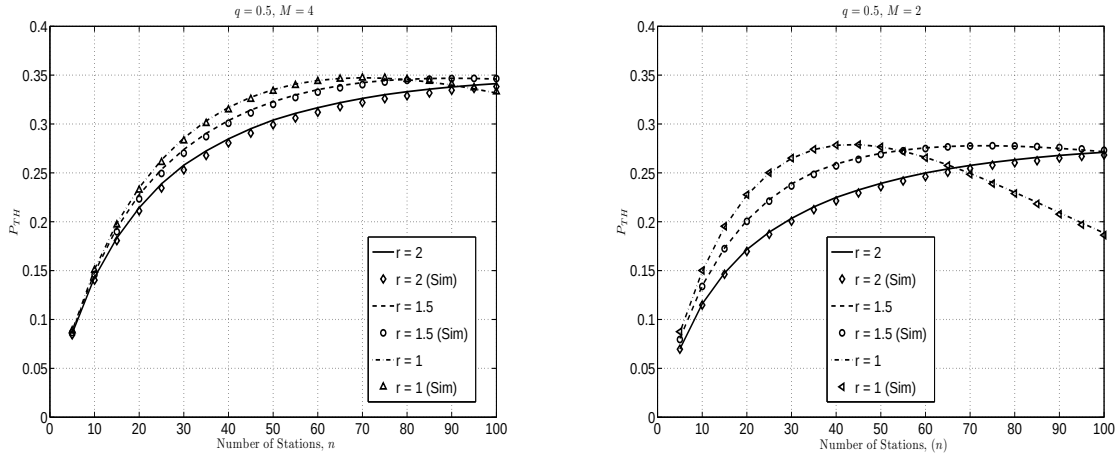


Figure 3.3: P_{TH} versus the number of SSs over a range of system populations with $N = 20$. $q = 0.5$ and $M = 4$

there is a performance improvement of approximately 5% with $r = 1.5$ and 10% with $r = 1$ over BEB ($r = 2$) in the range $30 \leq n \leq 60$. In Fig.3.4 (b), the corresponding improvement over BEB is 10% with $r = 1.5$ for $30 \leq n \leq 60$. A further increase of 25% is observed with $r = 1$ in the range $20 \leq n \leq 40$. These results show that as the number of SSs increases, contention increases thus degrading performance rapidly with $r = 1$. This indicates that a larger backoff window is needed for $n \geq 50$, however using a radix $r = 1.5$ still outperforms BEB for $n \leq 90$. These results show that decreasing the radix can improve the performance by reducing the number of unused minislots during the backoff phase. This improvement is more significant at lower values of M because a smaller M decreases the time between request retransmissions, creating a higher contention environment. In addition, using a lower values of M (decreasing the timeout period), increases the probability of an SS being in a backoff phase rather than a waiting phase. Thus decreasing the radix reduces wasted slots due to backoff which can provide a significant improvement in performance.

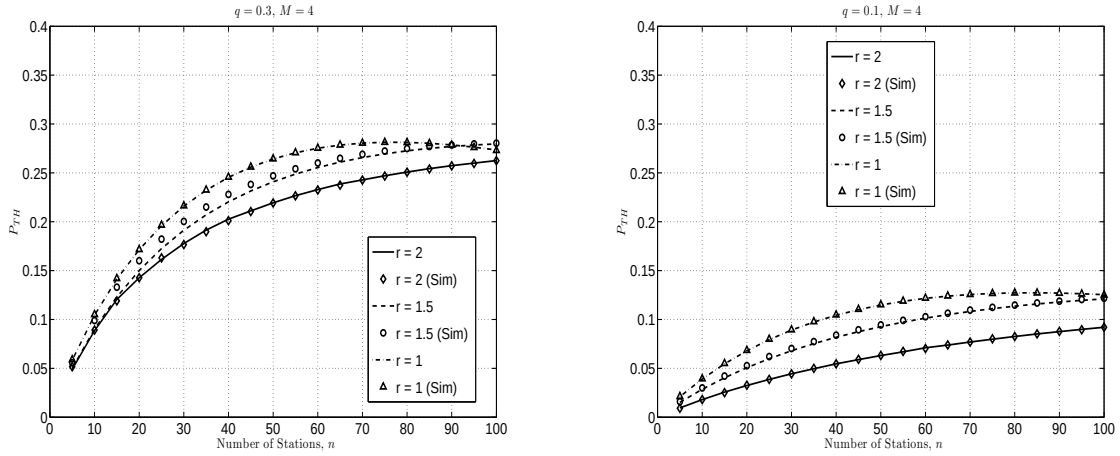


(a) (b)

Figure 3.4: P_{TH} versus the number of SSs n for $q = 0.5$ and $N = 20$ with timeout periods $M = 4$ and 2.

In Fig. 3.5 three values of r are used, 2 (BEB), 1.5 and 1 with $M = 4$ while varying the grant probability q. We see that when the probability of bandwidth available to be granted decreases to $q = 0.3$, the probability of migration to the backoff phase increases. Thus, varying the radix has a more significant effect on performance than at higher values of q used in Fig.3.4. In Fig. 3.5 (a), a radix value $r = 1.5$ provides a 5% performance improvement over BEB in the range $50 \leq n \leq 100$. At $r = 1$, an improvement of 10% is achieved in the range $40 \leq n \leq 60$. The effect of a variable radix with $q = 0.1$ is shown in Fig. 3.5 (b). The improvement in performance over BEB with $r = 1.5$ is approximately 35% for $n \geq 30$, and with $r = 1$ is approximately 50% for $20 \leq n \leq 70$.

Finally, in Fig. 3.6 the effect of a variable radix when contention is high due to a low number of minislots $N = 5$ is investigated. We observe that at $r = 1.5$ a performance improvement of 20% is achieved over BEB in the range $15 \leq n \leq 50$ and $35 - 40\%$ at $10 \leq n \leq 40$ when using $r = 1$. Even in this case, with $n \leq 50$, $r = 1$ outperforms BEB, and $r = 1.5$ outperforms BEB over a wide range up to $n \approx 100$. These results show that with a low value of N, the number of contention periods used in the backoff process has a significant effect on performance. By using a lower radix value, the number of minislots over several contention periods that are wasted during the backoff phase can be reduced. It is clear from the above results that the value of the radix r should be adjusted according to the BW request load to maintain an efficient system over a wide range of parameters.



(a) (b)

Figure 3.5: P_{TH} versus the number of SSs n for $M = 4$ and $N = 20$ with grant probabilities $q = 0.5$ and 0.1.

3.4 Summary

A new variable radix strategy was presented to improve the performance of contention based BW requests as defined in the IEEE 802.16 (WiMAX) wireless access standard. The proposed strategy shows that varying the radix according to the number of contending subscriber stations (SSs) can significantly improve system performance compared to that with binary exponential backoff (BEB). Performance results were presented for various network parameters and a wide range of traffic loads which confirm this claim. A variable radix strategy can be very beneficial in WiMAX applications where Quality of Service (QoS) is required. A BS can simply include the radix value in the UCD with the minimum and maximum window sizes. A variable radix can also be employed in a contention-based BW request mechanism for time-varying channel conditions.

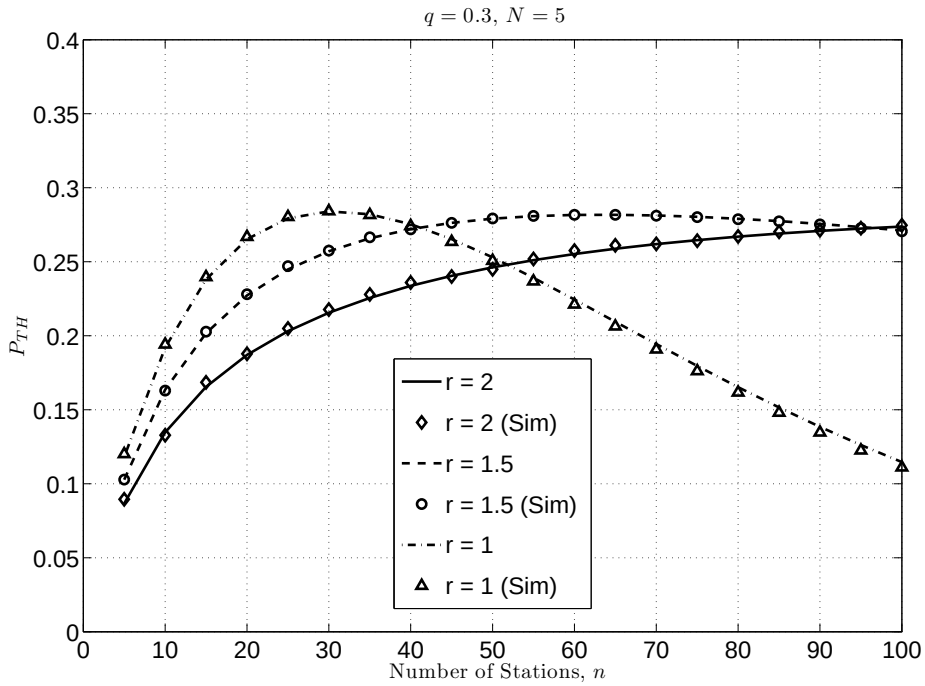


Figure 3.6: P_{TH} versus the number of SSs n for $q = 0.3$ with $M = 4$ and $N = 5$.

Chapter 4

A QoS Model for Random Access Systems

4.1 Introduction

In point-to-multipoint (PMP) communication systems, a base station (BS) serves several subscriber stations (SSs) by centrally coordinating the communications in the downlink (BS to SS) and uplink (SS to BS). In most systems, the BS controls access to the medium in the uplink direction by providing resources in the form of slots which are granted to the SSs on demand. All SSs wishing to transmit initially access the medium in an uncoordinated manner, i.e., randomly. For this purpose, random access is the main mechanism employed by the SS for connection initialization and network recognition to acquire resources and send channel information to the BS.

Resources requested by the SS can take several forms, e.g. time slots, spreading codes or frequency subchannels. A good example to illustrate the request procedure is the IEEE 802.16 *contention interval* (CI) in the uplink subframe as explained in Chapters 2 and 3

The CI consists of a finite number of available resources (time slots) for fixed interface (IEEE 802.16d), or spectrum spreading (CDMA) codes in the mobile interface (IEEE 802.16e) mainly used for for initial ranging (IR) and bandwidth (BW) requests. SSs contend to send bandwidth request messages (BW-REQ) during the CI period to acquire the sufficient amount of bandwidth needed to send their data. When a request successfully reaches the BS, the BS provides appropriate transmission opportunities in the next uplink subframe [10] [11].

In addition to network entry purposes, there are also contention based services that utilize random access mechanisms for data transmission, e.g. IEEE 802.16.
1) Non Real Time Polling Services (nrtPS) - In this service class, contention based polling is employed by the BS with groups of SSs. SSs belonging to a group can request resources during the contention based polling opportunities, which can result in collisions and additional attempts.
2) Best Effort Services (BE)- This class is designed for traffic which does not require any QoS or real-time guarantees, i.e., data is sent whenever resources are available. The SS uses contention based request for bandwidth. BE services may also allow transmission of small amounts of delay-tolerant data within the request interval, for example many applications such as interactive on-line games and Voice over IP (VoIP) which is currently the dominant Internet service [29],[30].

Although contention resolution techniques have been the focus of much research, very little has been done to exploit the random access mechanism, in particular to provide differential QoS amongst contending SSs. This is because the QoS techniques considered to date have employed a qualitative approach of granting specific resources based on the scheduling during the connection between a BS and the SSs. The novelty of the mechanism presented here lies in the use of the CI not only for resource requests and BE data, but also to provide a simple QoS technique prior to the connection by simply categorizing these resource quantitatively. The size of the CI has been considered in the literature, but not to provide a QoS mechanism.

For example, in [30], the authors introduced a short comparative study of the random access mechanisms for bandwidth requests for both OFDM and OFDMA systems in IEEE 802.16. In [31], an analysis of the size of the optimal contention period in the uplink subframe was presented with respect to the number of SSs to produce a cost function for throughput. In [32], an uplink scheduler architecture based on Data Over Cable Services Interface (DOCSIS) was proposed for nrtps and BE services. In [25], an analytical model of random access in WiMAX was introduced for a finite number of SS. The authors considered minimizing the request transmission delay as an optimization problem with respect to the size and number of contention slots. An adaptive technique of switching between random access and polling in IEEE 802.16 was proposed in [27] to improve channel performance. A study of QoS mechanisms supported in IEEE 802.16 was presented in [33]. The authors investigated the difference in average delay between the uplink and downlink traffic and proposed

a weighted round robin (WRR) scheduling algorithm for the downlink. In [16] and [17] the authors introduced a differential service technique for random access systems based on a two-class incoming traffic categories.

4.2 The QoS Model

In this chapter, a simple and realistic QoS model for random access systems is presented. The model provides a general and effective approach to a QoS classification by categorizing the available resources (slots)[1] of the CI into distinct service classes. The main strategy is based on the fact that lower class SSs are prone to higher collisions. For simplicity, a two-class model is considered here however, the technique can easily be extended to any number of classes. The model presented can provide a simple QoS mechanism for a wide range of contention based system, in particular the bandwidth request BW-REQ process, contention-based and non-real time data transmission classes in IEEE802.16.

Performance metrics based on average throughput and acceptance probabilities are investigated. This model provides a useful analytical tool for most contention systems that utilize random access as a mechanism for network entry, resource acquisition and contention-based or best effort data transmission. For example, the IEEE 802.16 bandwidth request and initial ranging mechanisms in the uplink contention channel. Following this example, the access packet in the presented model will henceforth be referred to as a *request*. For simplicity, only two classes are considered, however the technique can easily be extended to any number of classes. Simulation results using the OOP toolbox described in Chapter 2 were obtained to verify the accuracy of the analysis.

The proposed QoS mechanism is obtained by quantitatively dividing the entire CI into two overlapping contention regions K and K_2 as shown in Figure 4.1.

High (Class 1) and Low (Class 2) SSs are defined such that Class 1 SSs have access to all K slots in the contention region, while Class 2 SSs have access to only a subset of K_2 slots where $K > K_2$.

To verify the proposed QoS technique, a simple three-state (Tristate) Markov to model a praticular (tagged) SS is used.

Figure 4.2 shows the three-state Markov state diagram of a particular SS based on

[1] Time slots or simply *slots* is used to indicate resources in this model

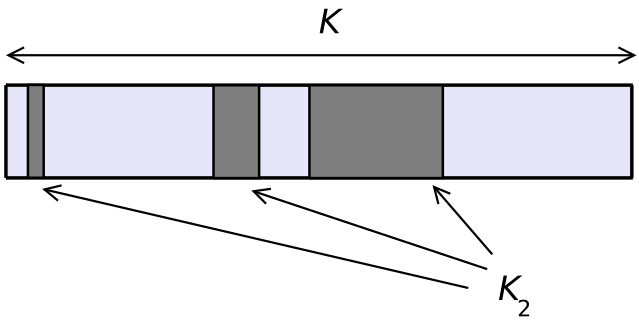


Figure 4.1: Slot categorization in CI.

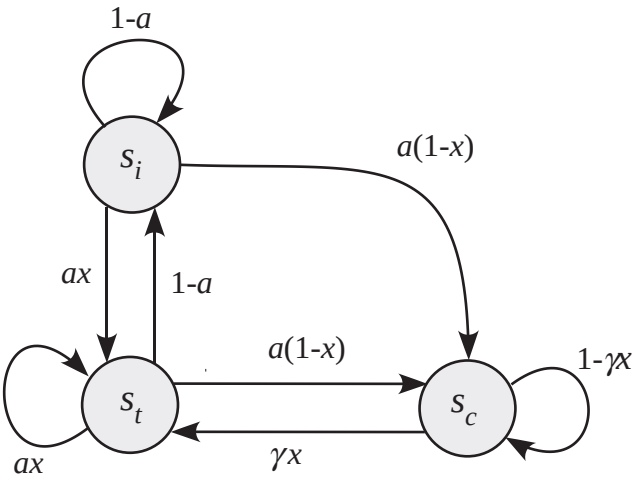


Figure 4.2: A simplified (tristate) SS Markov model.

the steady state and transition probabilities. Having a request to send, an idle SS at stage s_i enters transmission stage s_t or s_c if its request is successfully transmitted or collided, respectively. The probability of a request arrival is a and the probability of success x is defined as the probability of a particular SS accessing a slot successfully. If collisions occurs, the SS issues a retransmission with a probability γ till the request is successfully transmitted. It is noticed that, the model in Figure 4.2, represents a special case of the SS Markov model introduced in Chapter 2 by setting the number of retrial stages $m \to \infty$ and the radix $r \to 1$. Assumptions 1), 2) and 3) adopted in 2.3 hold for this model as well. For simplicity, the tristate model is considered however, backoff strategies used in Chapter 2 can be easily applied to the model as shown later in Section 4.4

From Figure 4.2, the transition matrix is given by

$$\mathbf{P} = \begin{bmatrix} 1-a & 1-a & 0 \\ ax & ax & \gamma x \\ a(1-x) & a(1-x) & 1-\gamma x \end{bmatrix} \tag{4.1}$$

At equilibrium, the equilibrium equation [21] is given by

$$\mathbf{Ps} = \mathbf{s} \tag{4.2}$$

where the state vector $\mathbf{s}$ is given by

$$\mathbf{s} = \begin{bmatrix} s_i & s_t & s_c \end{bmatrix}^t \tag{4.3}$$

where t indicates transpose.

Let $\mathbf{s_1}$, $\mathbf{s_2}$, be the state vectors of a particular SS from Class 1 (High) and Class 2 (Low), respectively. Solving the set of linear equations for s_{i_1}, s_{t_1} and s_{c_1} and $\sum s_j = 1$ for Class 1 SSs we get

$$\begin{aligned} s_{i_1} &= \frac{\gamma x_1(1-a_1)}{D_1} \\ s_{t_1} &= \frac{a_1\gamma x_1}{D_1} \\ s_{c_1} &= \frac{a_1(1-x_1)}{D_1} \end{aligned} \tag{4.4}$$

where $D_1 = a_1(1-x_1) + \gamma x_1$

Similarly, for a Class 2 SS

$$\begin{aligned} s_{i_2} &= \frac{\gamma x_2(1-a_2)}{D_2} \\ s_{t_2} &= \frac{a_2\gamma x_2}{D_2} \\ s_{c_2} &= \frac{a_2(1-x_2)}{D_2} \end{aligned} \tag{4.5}$$

and $D_2 = a_2(1-x_2) + \gamma x_2(1+a_2)$

Define an active SS as either an idle SS (at state s_i) with a fresh request that arrived to be sent, or a collided SS (at state s_c) with a collided request to be retrans-

mitted. Let p_1 and p_2 be the probabilities that an SS is active from Class 1 and Class 2, respectively. Hence, for Class we have

$$p_1 = a_1 s_{i_1} + \gamma s_{c_1} \tag{4.6}$$

Similarly for Class 2 we have

$$p_2 = a_2 s_{i_2} + \gamma s_{c_2} \tag{4.7}$$

For a successful transmission, only one SS from either class accesses a particular slot in a frame duration. Equivalently, none of the remaining SSs chooses that particular slot in the same frame. Hence, the probability x_1 that a Class 1 SS successfully accesses a slot out of all K slots is given by

$$x_1 = \frac{K - K_2}{K}\left(1 - \frac{p_1}{K}\right)^{N_1 - 1} + \frac{K_2}{K}\left(1 - \frac{p_1}{K}\right)^{N_1 - 1}\left(1 - \frac{p_2}{K_2}\right)^{N_2} \tag{4.8}$$

The first term on the right-hand side of (4.8) represents the region where only Class 1 is contending. The second term represents the (overlapped) region, where both classes are contending.

Therefore, the probability x_2 that a Class 2 SS successfully accesses a slot out of the subset K_2 slots is given by

$$x_2 = \left(1 - \frac{p_2}{K_2}\right)^{N_2 - 1}\left(1 - \frac{p_1}{K}\right)^{N_1} \tag{4.9}$$

Substituting (4.6) in (4.8) and (4.9) and solving (4.4) and (4.5) numerically by minimizing the iteration error, the steady state vectors $\mathbf{s_1}$ and $\mathbf{s_2}$ are obtained.

The average system Th throughput is therefore defined by the average number of SSs that are in the transmitting state. Therefore Th_1, Th_2 for Class 1 and Class 2 are given by,

$$\begin{aligned} Th_1 &= N_1 s_{t_1} \\ Th_2 &= N_2 s_{t_2} \end{aligned} \tag{4.10}$$

respectively.

Although the system throughput given in (4.10) is a commonly used metric to measure the system performance. It is not the only metric that indicates the overall performance in all conditions. This is because intuitively, due to the differences in

the classes population of SSs, N_1 and N_2, low class (Class 2) SSs might produce a higher throughput reading than Class 1. This occurs if the contention level on their allocated slot region is lower than Class 1 (e.g. $N_1 << K$). Therefore, a more accurate interpretation than the system throughput in (4.10) to measure performance is provided by the *Acceptance probability* or *efficiency* P_a [21] (Ch.10).

P_a is defined as the ratio of a successful access to the average input traffic per time unit (frame) $P_a = \frac{Ns_t}{Na}$, which gives the probability of acceptance or the *efficiency* of a given SS in each class.

For the given Class 1 and Class 2 system we have

$$\begin{aligned} P_{a1} &= \frac{s_{t_1}}{a_1} \\ P_{a2} &= \frac{s_{t_2}}{a_2} \end{aligned} \tag{4.11}$$

respectively.

Since it is assumed that a collided SS retries to send its request till it is successfully transmitted and, with one transmission attempt occurs within a frame period, the average number retransmissions can directly be interpreted as the average *retransmission delay* in frames d_a an SS incurs till its request is successfully transmitted.

From [21], Ch.(10.5), the average retransmission delay is given by

$$\begin{aligned} d_a &= \sum_{i=0}^{\infty} i(1-P_a)^i P_a \\ &= \frac{1-P_a}{P_a} \end{aligned} \tag{4.12}$$

The term in (4.12) evaluates for Class 1 and Class 2 as

$$\begin{aligned} d_{a_1} &= \frac{1-P_{a_1}}{P_{a_1}} \\ d_{a_2} &= \frac{1-P_{a_2}}{P_{a_2}} \end{aligned} \tag{4.13}$$

respectively.

4.3 Numerical Results

In the following results, the total number of slots in the CI is set to $K = 16$ slots. The retransmission probability for Class 1 and Class 2 is $\gamma = .125$. Let R_k be the ratio of overlap in CI slots between K and K_2 such that $R_k = K_2/K$. The average input traffic from Class 1 and Class 2 are N_1a_1 and N_2a_2 requests/frame, respectively. Simulations to confirm anlytical results are perfomed using the OOP Matlab tool in Appendix A based on an average of $20,000$ trials.

In Figure 4.3, the impact of traffic interaction on the efficiency of both classes is investigated. The average input traffic is N_1a_1 and N_2a_2 for Class 1 and Class 2, respectively. $N_1 = 20$, $N_2 = 30$ and $K_2 = 12$.

In Figure 4.3 (a), a contour plot of the efficiency of Class 1, P_{a_1} is shown. It is noticed that, at a given value of Class 1 traffic e.g. $N_1a_1 = 10$ requests/frame, the efficiency is approximately 80% when $N_2a_2 = 2$ requests/frame from Class 2 is present.

Following the same value of traffic for Class 1 (moving vertically), the efficiency decreases to 40% and 20% at $N_2a_2 = 9$, $N_2a_2 = 20$ requests/frame, respectively.

On the other hand, in Figure 4.3 (b), a contour plot of the efficiency of Class 2, P_{a_2} is shown. The same input traffic for Class 2, $N_2a_2 = 10$ requests/frame is observed. It is noticed that, the 80% efficiency point for Class 2 occurs at approximately $N_1a_1 = 1.8$ requests/frame. Following the same point (moving horizontally), the efficiency decreases to 40% and 20% at $N_1a_1 = 5$ and $N_1a_1 = 9$ requests/frame, respectively.

Thus, a 60% loss in efficiency for Class 1 requires a 10 times increase in the interfering traffic from Class 2. On the other hand, only 5 times increase from Class 1 causes the same loss in efficiency for Class 2.

This observations shows that Class 1 traffic has a significant impact on the performance of Class 2 while Class 1 is considered relatively immune from the interference of Class 2.

Therefore, a significant difference in efficiency between classes to provide QoS can be achieved by having $R_k = 75\%$ (lower Class (2) is only deprived 1/4 the CI slots).

In Figure 4.4, the effect of varying the overlap ratio R_k is investigated by measuring the average retransmission delay d_{a1}, d_{a2} for Class 1 and Class 2, respectively. The population of both classes is fixed at $N_1 = N_2 = 50$. The measurement of the retransmission delay is performed for each class by varying the input traffic such that the other class is set at a saturation condition. This means that Class 1 performance

(solid) is shown with $a_2 = 1$ for Class 2, i.e., saturation condition. Similarly, Class 2 performance (dotted) is shown with $a_1 = 1$.

In Figure 4.4 (a) the retransmission delay d_a (subscript of a indicates the class) at $R_k = 0.75$ is shown. It is observed that at $Na \leq 20$ Class 2 incurs approximately twice the retransmission delay incurred on Class 1 d_{a_1}. At $30 \leq Na \leq 50$, d_{a_2} increase to approximately $2.5\, d_{a_1}$ due to the increase in generated retransmissions at higher traffic.

In Figure 4.4 (b), less slots are allocated for Class 2 by decreasing the slots overlap ratio to $R_k = 0.66$ ($K_2 = 10$). It is noticed that a 9% decrease in the allocated slots for Class 2 causes a significant increase in its retransmission delay but marginally affects Class 1.

At $20 \leq Na \leq 30$, $d_{a_2} \approx 2\, d_{a_1}$. At $30 \leq Na$, $d_{a_2} \approx 4\, d_{a_1}$. This indicates a 30% more delay occurs for Class 2 at $R_k = 0.66$ compared to $R_k = 0.75$ as the input traffic increases.

This shows that allocating less slots for the low Class (2) does not only cause an increase in retransmission delay due to higher collisions but the generated retransmissions has a minimal impact on the efficiency of the high Class (1).

Thus, the number of allocated slots in the CI dramatically effects performance. This proves that the proposed QoS technique provides a reliable differential QoS service mechanism.

4.4 Quality of Service with a Variable Radix

In this section, the effect of varying the the radix parameter r introduced in Chapter 2 is investigated. In the model in Section 2.3 for each service class, the Markov chain parameters are specified with a corresponding subscript. The radices are r_1, r_2 for Class 1 and Class 2, respectively. Equivalently, from (4.6), the corresponding probabilities of activity p_1 and p_2 for the model Section 2.3 corresponding to Class 1 and Class 2 are thus given by

$$\begin{aligned} p_1 &= a_1 \left(s_{i1} + s_{t1}\right) \left[1 + \frac{1 - x_1 - (1 - x_1)^{m+1}}{x_1}\right] \\ p_2 &= a_2 \left(s_{i2} + s_{t2}\right) \left[1 + \frac{1 - x_2 - (1 - x_2)^{m+1}}{x_2}\right] \end{aligned} \tag{4.14}$$

respectively.

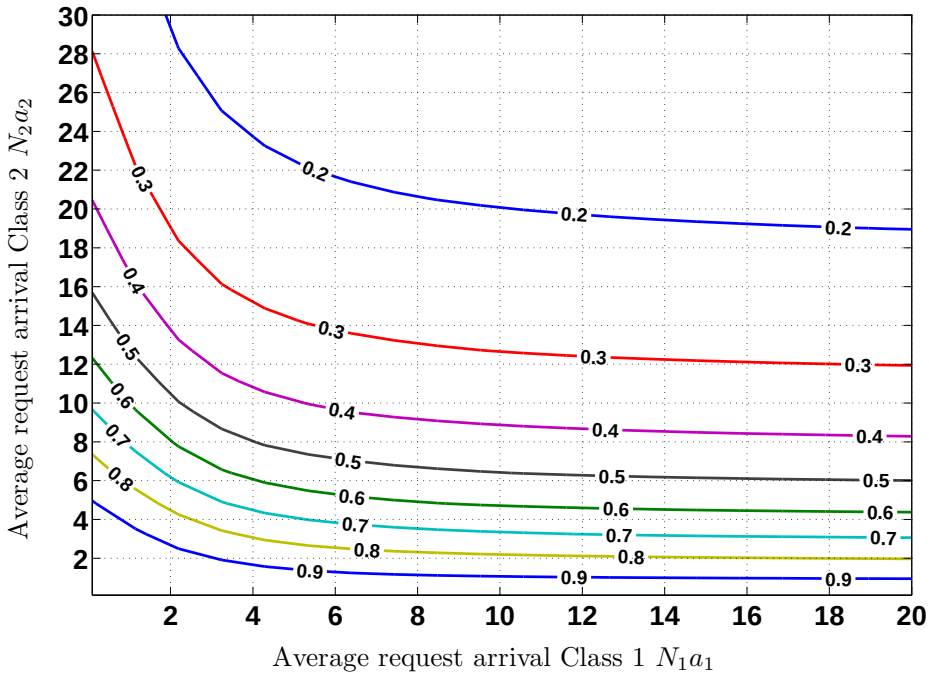
Using (4.8) and (4.9) in Section 4.2, the corresponding throughput measurements for Class 1 and Class 2 in (4.10) are obtained

Figure 4.5 shows the effect of a variable radix on the QoS strategy proposed in Chapter 2. Two classes of SSs (Class 1 and Class 2) with equal populations $N_1 = N_2 = 50$ are used. Class 1 has access to all the slots in the contention interval ($K = 16$). Class 2 has access to only a subset of the slots ($K_2 = 8$), as depicted in Figure 4.1. The initial backoff window sizes for Class 1 and Class 2 are W_1, W_2, respectively, with $W_1 = W_2 = 32$. The Class 1 performance in Figure 4.5 is shown with $a_2 = 1$ for Class 2, i.e., saturation condition. Similarly, Class 2 performance is shown with $a_1 = 1$. Two radix values for Class 1 are used, $r_1 = 2$ (BEB) and $r_1 = 1$, while the radix for Class 2 is $r_2 = 2$. Note that $r_1 = 1$ improves the performance of Class 1 SSs by 25%, while reducing that of Class 2 SSs by approximately the same amount. This indicates that a variable radix can be used for effective QoS control without changing the slots allocated to each class in the contention interval. Obviously reducing the low (Class 2) service quality will enhance the high (Class 1) service quality, but this is the tradeoff required in a QoS system. This is because decreasing the radix for Class 1, r_1, will increase Class 1 slot utilization over the entire contention interval K, thus improving the performance. This will have less of an effect on Class 2 SSs because of their access to fewer slots K_2. This is illustrated in Figure 4.6 with K increased to 32 while K_2 is still 8. Using a fixed backoff window for Class 1 ($r_1 = 1$), improves performance by 25% for Class 1 while degrading that of Class 2 by only 10%. This indicates that a variable radix can be used to enhance the QoS mechanism in a multichannel random access system.

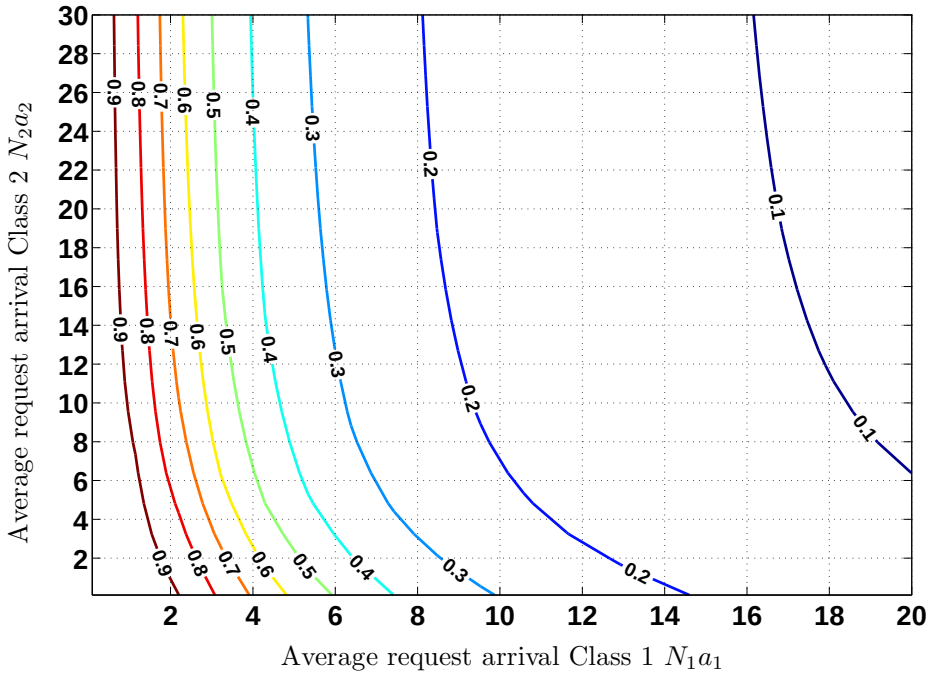
4.5 Summary

In this chapter, a two-class QoS model for random access request systems was introduced. The system was modeled by a Markov chain to determine the system transition and steady state probabilities. The random access success probabilities were derived as a function of average input traffic, allocated slots and steady state probabilities for each QoS class. Based on the model analysis, the vital performance metrics presented by system efficiency, average retransmission delay and request throughput for each Class were evaluated. The results presented show that this model can be used to provide an efficient and reliable QoS mechanism for random access by providing

higher class SSs a better access performance than lower class. The model also proved that high class SSs are relatively immune from performance degrade caused by overwhelming retransmissions generated by low class SSs due to high collisions. The model is appropriate for random access systems where slots are allocated differently to various applications. The proposed model can be employed for systems that utilize random access methodologies, e.g. the IEEE 802.16 random access channel used for resource (bandwidth) requests. The technique requires minimal overhead since the slot allocations can be assigned within the contention window of the uplink subframe. The adaptive radix strategy introduced in Chapter 8 was also studied. The adaptation of retransmission probabilities using the variable radix mechanism can be used to enhance differential services in a QoS mechanism using random access. A variable radix can dynamically provide a differential QoS technique without changing the amount or slots allocated for individual classes. This can provide a simple selective method of QoS strategy to contending SSs because the radix value can easily be conveyed to selected SSs on individual bases via the Uplink Channel Descriptor in the DL-MAP message (UCD) along with the contention window size. An adaptive radix can seamlessly be applied to most multichannel random access systems such as the contention-based bandwidth request mechanism in IEEE802.16. An object oriented simulation tool was designed to provide extensive simulations which verify the analysis of the proposed models. Simulation using the OOP toolbox developed in Appendix A is used to confirm the accuracy of the results.

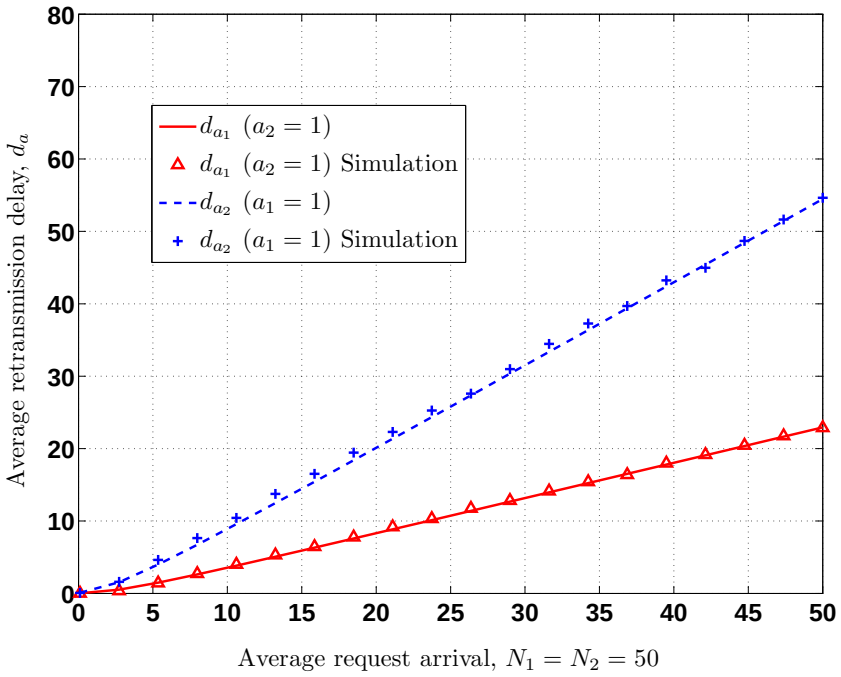


(a)

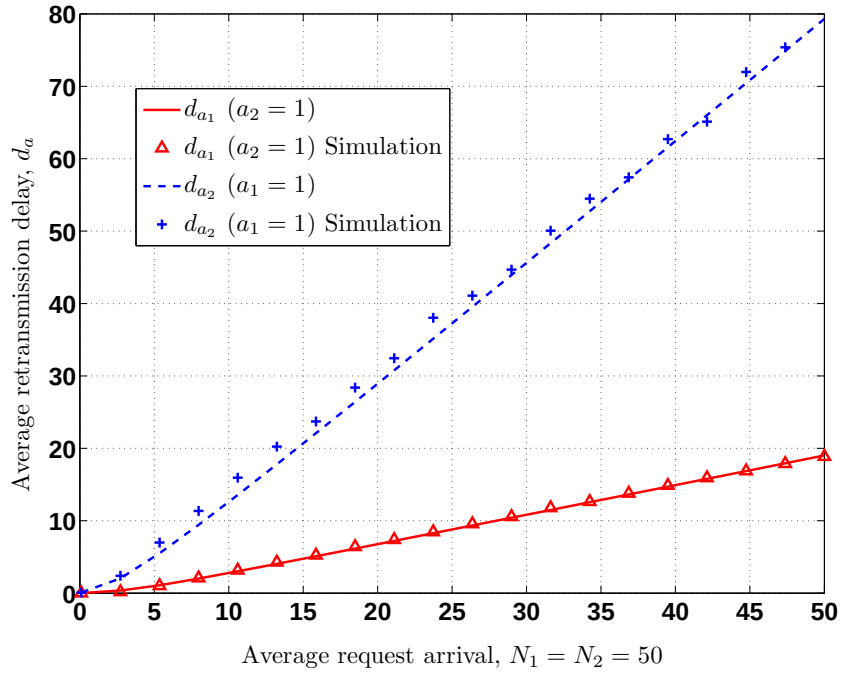


(b)

Figure 4.3: Efficiency P_a vs. average input traffic at $R_k = .75$, $N_1 = 20$, $N_2 = 30$, Class 1 (a), Class 2 (b).



(a)



(b)

Figure 4.4: Average retransmissions delay d_a, Class 1 (solid), Class 2 (dotted) $R_k = 0.75$(a), $R_k = 0.66$(b)

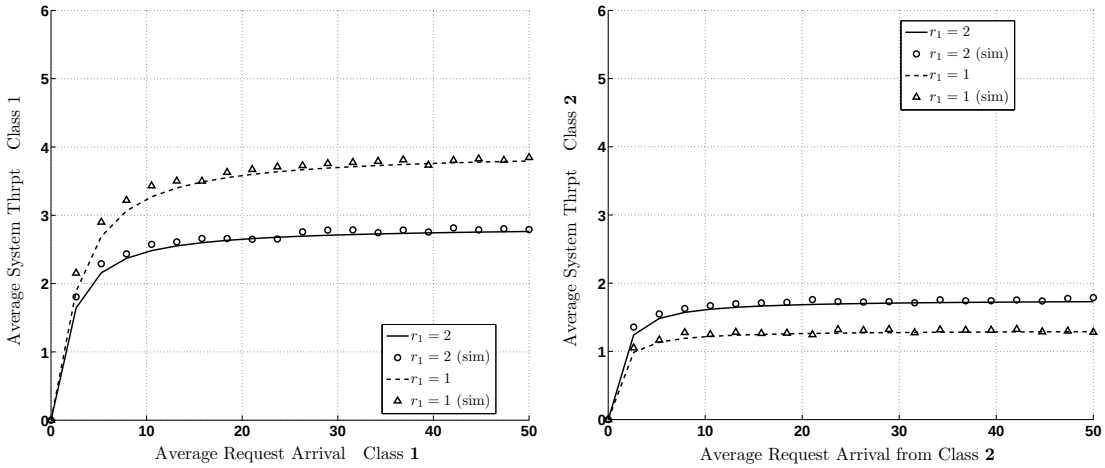


Figure 4.5: System throughput with variable radix for Class 1 and Class 2 traffic, $K = 16$, $K_2 = 8$.

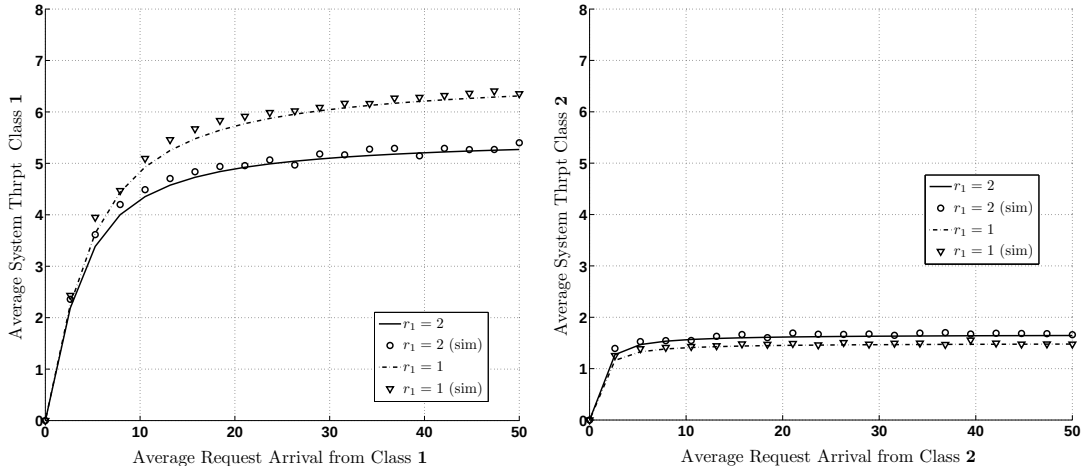


Figure 4.6: System throughput with variable radix for Class 1 and Class 2 traffic, $K = 32$, $K_2 = 8$.

Chapter 5

Modeling of Wireless Fading Channels

5.1 Introduction

5.1.1 Multipath Propagation

The radio channel is the standard transmission medium in wireless communications. When a wireless signal is transmitted from a transmit (Tx) antenna, it undergoes reflection and refraction effects due to intervening objects. This results in the arrival of numerous propagation paths at the receive (Rx) antenna. As the term dictates, *multipath* propagation is defined as the phenomena of superposition of different components of the transmitted wireless signal. Obviously, the resultant amplitude can be constructive or destructive depending on the phases of the multipath components. Multipath components are characterized by distinct amplitudes, delays and phase shifts, and these components add at the receiver.

Signal variations due to the distances between the Tx and Rx antennae impact signal quality. This defines the term, *large-scale fading*. Large scale fading characterizes a signal attenuation due to the fall of signal power exponentially by distance between the Tx and Rx antennae due to *pathloss*. An additional effect on the signal amplitude is due to the presence of large objects e.g. mountains, high-rise buildings, etc. These objects create a scattering effect on wireless signals which can significantly affect performance even if a line-of-sight (LOS) path exists between the Tx and Rx antennae. The LOS path can still be in the radio *shadow* of such objects even for large distances.

Unlike pathloss and shadowing that cause attenuation due to stationary objects over large distances, small-scale or *multipath fading* defines the multipath effects imposed on the received signal amplitude due to small distance variations or short periods of time. Typically, large scale fading effects are ignored when small scale fading is studied. Hence, the term *fading* commonly indicates small-scale or multipath fading. Figure 5.1 shows the resultant power gain at the Rx antenna with large and

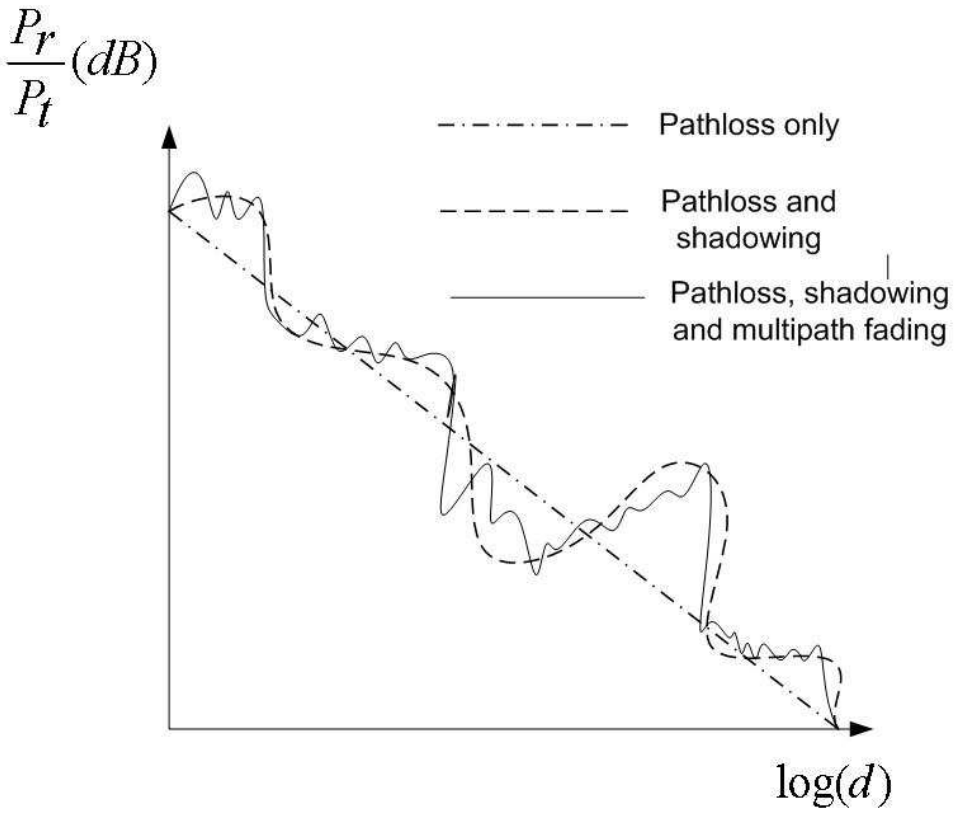


Figure 5.1: Ratio of received power to transmit power P_r/P_t (dB) in the presence of pathloss, shadowing and fading versus the logarithm of the distance d. [1].

small scale fading.

The amplitude of the received signal depends on the delay and phase of the multipath components at the receiver. If either the Tx or Rx antenna or even an intermediate object are moving, the relative distance covered by the multipath components will change with time. The resultant signal amplitude at the Rx antenna is therefore time-variant mainly because the signal components arrive from different directions with different propagation delays. The temporal variations in the received signal can be large if movements occur at high speeds.

A small change in the position of the Tx or Rx antenna can result in a significant change in the total amplitude of the received signal. For example, at a carrier frequency of 1900 MHz, a change of 20 cm can turn the received signal amplitudes from constructive to destructive and vice versa. A cell phone user can often move a few steps in order to improve his signal quality [34]. If a moving receiver stops at a point were the multipath components add destructively, the receiver is said to be in a *deep fade* state. This situation can occur frequently when a user in a fast vehicle passes through several deep fade points in a small period of time. Mobility is one of the the major advantages of wireless communications, thus fading is a significant challenge and an important aspect of modeling and designing a reliable wireless system.

The speed of a mobile user a very important factor affecting the fading phenomena. Mobility creates frequency modulation effects due to the *Doppler shifts* of the multipath components as explained below. Fading channels are classified in the next section, and the signal parameter changes due to mobility are discussed in more detail.

5.1.2 Classification of Fading Channels

Two main parameters are used to classify fading channels:

1. **Delay Spread**
   Multipath components combining at the receiver produce a distorted version of the transmitted signal and arrive at the Rx over a period of time. The delay spread is defined as the difference between the largest and the smallest of the delays of these components. This delay spread causes *time dispersion* and hence *frequency-selective fading*. If the bandwidth of a transmitted signal denoted by B_s is narrow enough, no distortion occurs. As B_s increases, the distortion becomes more significant. Define the channel *coherence bandwidth* as $B_c = T_d^{-1}$, where T_d denotes the channel delay spread. Beyond this bandwidth, the different frequency components of the signal undergo different fading effects. Thus, a signal with $B_s \geq B_c$ is classified as *frequency selective*. In other words, the *coherence bandwidth* B_c, is a measure of the maximum signal bandwoidth for which the received signal components are still strongly correlated in frequency. More precisely, the amplitude and phase of two signals will be different if their frequency difference exceeds Bc.

2. **Doppler Spread**

If the Rx antenna[1]. is moving slowly, channel variations can be negligible. As the velocity increases, or surrounding intermediate objects move faster, a time-varying Doppler effect occurs which cannot be ignored. This creates a *Doppler shift* which is a function of the signal carrier frequency and the velocity of the Rx antenna. Doppler shift causes a random frequency modulation of the received signal, and increases (spreads) the bandwidth of the signal. This *Doppler spread* causes *frequency dispersion* and hence *time selective fading*.

The Doppler spread is a measure of the largest and the smallest frequency components of the received signal. The *coherence time* denoted by T_c defines the maximum time duration for which the multipath components of the received signal are still strongly correlated in amplitude. We have $T_c = f_m^{-1}$, where f_m is the maximum Doppler shift. If the transmitted signal pulse width is $T_s \geq T_c$, then the received signal will be significantly distorted, and the channel is classified as *time selective*.

The classification of fading channels is illustrated in Figure 5.2 based on the following fading parameters [35].

- Coherence Bandwidth B_c: if $B_c >> B_s$, there is no time spread and thus the channel is categorized as non-frequency selective or *flat* in frequency.
- Coherence time T_c: if $T_c >> T_s$, there is no frequency spread and the channel is categorized as *flat* in time.

Channels that are flat in both time and frequency are not subject to significant multipath fading effects, however they are still subject to large scale fading and pathloss. Channels that are selective in both time and frequency are extremely unpredictable and very difficult to model or analyze [36].

Frequency selective channels are also classified as *wideband channels* and are subject to *Inter Symbol Interference* (ISI). Time selective (flat in frequency) or simply *flat fading* channels (FFC) represent a very wide range of terrestrial wireless systems. FFCs are also qualitatively known as *narrow-band* fading channels. FFCs provide a good model for many wireless communications systems and thus are widely used in the literature. FFCs are the focus of the work presented in this chapter.

[1]Although mobility indicates relative motion of the Tx or Rx antennae, or intermediate objects, mobility here refers to the Rx antenna only for simplicity. this is usually the case in real situations

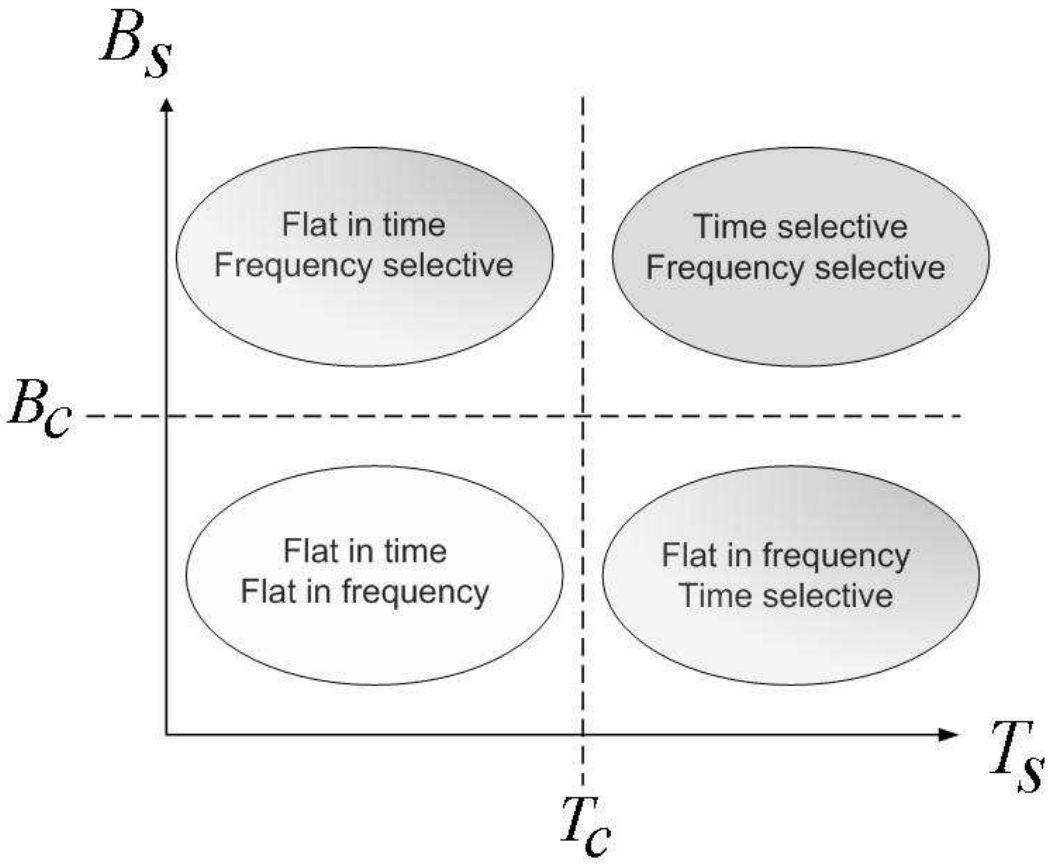


Figure 5.2: Classification of fading channels

5.2 Time-Varying Narrowband (Flat) Fading

As mentioned in Section 5.1.2, the focus in this dissertation is narrowband fading where the signal pulse or (symbol) duration T_s is small relative to the channel coherence time T_c. A flat fading channel is studied according to a quasi-static fading model. This quasi-static assumption means that each multipath component undergoes a fixed shadowing and pathloss where the Rx is moving in the presence of relatively static scatterers. The mobile receiver therefore travels through the crests and troughs of the signal amplitudes and undergoes a time varying received signal pattern. The troughs occur at a rate relate to the mobile speed. Thus, the fading rate (*fades/sec*) is determined by the receiver speed. For example, at a carrier frequency $f_c = 900$ MHz, fading troughs occur 16 cm apart or equivalently at half the wavelength ($\lambda/2$) [34].

Let the receiver move at a velocity v. This movement creates a Doppler shift in

the received signal which depends on the direction of movement relative to the transmitter. Let $z(t)$ be a single path passband signal modulated by a carrier frequency f_c and time-varying amplitude $\alpha(t)$. Thus, the received signal $z_r(t)$ is given by [37, Ch. 1]

$$z_r(t) = \alpha_r(t)\,\cos\left(2\pi(f_c - \frac{v}{\lambda})t - \phi_r(t)\right) \tag{5.1}$$

where $\alpha_r(t)$ and $\phi_r(t)$ are the time-varying amplitudes and phase shifts, respectively. Thus, the Doppler shift $f_d = f_c - v/\lambda$ decreases the received frequency by v/λ. Hence the Doppler shift created by a single multipath component (ray) and with an angle of incidence ν with the receiver is given by

$$\begin{aligned} f_d &= -\frac{v}{\lambda}\cos(\nu) \\ &= -f_c\,\frac{v}{c_l}\cos(\nu) \end{aligned} \tag{5.2}$$

where c_l is the free space velocity of light (3×10^8 m/s). Note that the Doppler shift f_d is negative or positive depending on the direction of movement of the receiver away from or towards the transmitter, respectively.

From (5.2), the Doppler shift may seem negligible being on the order of $1 - 1000$ Hz compared to the carrier frequency f_c. However, the presence of a large number of multipath components creates a large number of distinct Doppler shifts, i.e. fading troughs. As the velocity of the receiver increases, Doppler shifts for different multipath components vary considerably, creating a spectrum of frequencies where each according to (5.1) is a shifted version of the carrier frequency. Since the Doppler shift is a function of the receiver velocity, the faster it moves, the higher the rate of fading troughs it encounters. This severe fading due to high speed receiver movement is categorized as *fast* fading.

5.2.1 Statistical Modeling of Narrowband Fading

At the receiver, the quality of the signal is mainly determined by its amplitude. It has been found that as the number of multipath components increases, the statistics of the in-phase or real ($\Re$) and quadrature or imaginary ($\Im$) components of the signal approach a zero mean Gaussian distribution. This is a good consequence of the central

limit theorem. The received signal with M multipath components is given by

$$z_r(t) = \sum_{i=1}^{M} [\Re(t).\cos(2\pi f_c t) - \Im(t).\sin(2\pi f_c t)] \tag{5.3}$$

With the quasi-static assumption, a particular (i^{th}) multipath component within the time interval of interest has an amplitude $\alpha_i(t)$ and phase $\phi(t)$ which are very slowly varying so that $\alpha_i(t) \approx \alpha_i$ and $\phi_i(t) \approx \alpha_i$ [1]. Therefore, with a maximum Doppler shift of f_m we have

$$\begin{aligned} \Re(t) &= \sum_{i=1}^{M} [|\alpha_i| \cos(-2\pi f_m \cos(\nu_i)t + \phi_i)] \\ \Im(t) &= \sum_{i=1}^{M} [|\alpha_i| \sin(-2\pi f_m \sin(\nu_i)t + \phi_i)] \end{aligned} \tag{5.4}$$

Assuming a densely clustered scatterers, the phase ϕ is uniformly distributed in the range $[-\pi, \pi]$ [37].

It is well known that for two Gaussian random variables X and Y with mean zero and equal variance σ^2, $Z = \sqrt{X^2 + Y^2}$ is *Rayleigh* distributed and Z^2 is exponentially distributed [38], [1]. If we assume a variance of σ^2 for both time-varying in-phase and quadrature components, then the signal envelope is given by

$$z_r(t) = \sqrt{\Re(t)^2 + \Im(t)^2} \tag{5.5}$$

This envelope is Rayleigh distributed with *pdf*,

$$p_r(r) = \frac{2r}{\bar{P}_r} \exp\left(-\frac{-r}{\bar{P}_r}\right) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \qquad r \geq 0 \tag{5.6}$$

The received signal power is $z_r(t)^2 = |\Re(t)|^2 + |\Im(t)|^2$ with *pdf*

$$p_{r^2}(x) = \frac{1}{\bar{P}_r} \exp\left(-\frac{x}{\bar{P}_r}\right) = \frac{1}{2\sigma^2} \exp\left(-\frac{x}{2\sigma^2}\right) \qquad x = r^2 \geq 0 \tag{5.7}$$

The received signal is much better described by the received power because this power is proportional to channel gain, i.e., the received power is the transmitted power times the channel gain. The ratio of the signal power to the noise power at the receiver gives the *signal-to-noise ratio* which can be used to calculate the error

probability. Therefore, the received power is the metric of interest to represent channel quality. The received signal power fluctuates due to shadowing and pathloss as shown in Figure 5.1, and therefore $\bar{P}_r = \sum_{i=1}^{M} \mathbf{E}\left[\alpha_i^2\right] = 2\sigma^2$ [37].

Note that the above channel statistics are valid when there is no dominant in-phase and quadrature components and hence a Rayleigh distributed channel results. A Rayleigh channel is commonly used in the literature because the distribution has been shown to accurately model the behavior of some channels, and it is tractable mathematically since it only depends on the received signal power $\bar{P}_r$. This channel represents *worst case* fading, since there is no LOS path between the transmitter and receiver. When this is not the case, the channel has a LOS path so that $\Re(t)$ and $\Im(t)$ no longer have zero mean. The received signal in this case will be the superposition of the received multipath components as well as the the LOS component. In this case, the signal amplitude has a Rician distribution [1], [39]. Note that in this case the phase $\phi = \tan^{-1} \Re(t)/\Im(t)$ is also no longer uniformly distributed. This is due to the tendency for it to be equal to the phase of the dominant LOS component [34].

The large variety of fading situations led to the development of the Nakagami-m fading distribution [40]. The *pdf* of the amplitude and power corresponding to (5.6) and (5.7) for the Nakagami-m distribution are given by

$$p_r(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\bar{\gamma}^m} exp\left(\frac{-mr^2}{\bar{\gamma}}\right), \qquad m \geq .5, \tag{5.8}$$

and

$$p_{r^2}(x) = \left(\frac{m}{\bar{\gamma}}\right)^m \frac{x^{m-1}}{\Gamma(m)} exp\left(\frac{-mx}{\bar{\gamma}}\right) \tag{5.9}$$

respectively, where $\bar{\gamma}$ is the average received power and $\Gamma(.)$ is the Gamma function. The Nakagami-m distribution covers a wide range of practical (based on empirical data), fading channels. For $m = 1$, the distribution in (5.8) represents the Rayleigh fading in (5.6). At $m = \infty$, there is no fading and the channel is only subject to noise. The Nakagami-m model can be used to represent a wide range of fading conditions including the Rician fading channel which is described by the fading parameter K [39], [1]. In this case, $m = (K+1)^2/2K + 1$.

5.2.2 Temporal (Second Order) Statistics of Narrowband Fading

The fading rate is the number of fading troughs per second that a mobile reciever encounters. Jakes [37] proved that the shallower the fade the more likely it is to occur. Therefore, the rate and duration of fading play an important role in modeling the channel. The two main quantities used to interpret the rate and duration of the fading troughs are the *Level Crossing Rate* (LCR) and the *Average Fade Duration* (AFD), respectively. These quantities are termed as second order statistics parameters since they involve deriving the joint *pdf* of the channel envelope and its first order derivative with respect to time as shown in (5.10) and (5.11) below. They are defined as follows.

- Level Crossing Rate (LCR)

  The LCR is defined as the average (expected) number of times (crossings) per second the signal envelope drops below a specified level. The LCR depends on the signal strength as well as the signal level considered. In [37] a rigorous mathematical derivation of the complex multipath signal is introduced. The corresponding mathematical formulations for different channel distributions were subsequently obtained e.g. Rayleigh [34], [37] Rician [39] and Nakagami-m [41]. As will be seen later, the LCR is a crucial parameter in discrete modeling of the channel since it measures the rate at which the channel varies.

  Let N_z be the LCR for which a signal envelope z crosses (in the positive direction) a specified signal level Z. The LCR N_z is then given by [37]

$$N_z = p(z = Z) \times \int_0^\infty \dot{z} p(\dot{z}, z = Z) d\dot{z}. \tag{5.10}$$

- Average Fade Duration (AFD)

  The AFD indicates the average (expected) time the signal envelope is below a specified level. This level is typically based on a performance metric such as the received power required to obtain a target symbol/bit error rate. It can also be based on a given outage probability defined as the probability that the signal power falls below a given level.

  Thus the AFD is the expected time the envelope z stays below a target level Z,

given by

$$T_z = \frac{\Pr(z \leq Z)}{N_z} = \frac{1}{N_z} \int_0^Z p(z) dz \tag{5.11}$$

The outage probability is given by

$$P_{out} = cdf_Z(z_{min}) = \int_0^{z_{min}} pdf_Z(z) \tag{5.12}$$

Since the signal power is a random variable, reliable communications is not guaranteed at all times even if the average power is large. Thus, a strongly related metric to the LCR and AFD is the *fade margin*. It is defined as the maximum tolerable fading to maintain a reliable communication link as a percentage of time. In other words, given a minimum received signal power required for a reliable communication link z_{min}, how large does the average power have to be in order to maintain the link at this reliability level $x\%$ of time. In a fading channel, the SNR is a random variable and the bit error rate (BER) decreases monotonically with increasing SNR. Thus, accurate modeling of the fading channel is required so that appropriate measures can be taken to mitigate the effects of fading, as will be seen in the following Section.

The average fade duration determines the number of bits or symbols affected by a fading trough. Specifically, consider a system with bit time T_b. Suppose the probability of bit error is high when $z < Z$. In this case, if $T_b \approx T_z$ then the system will likely experience single error events, where bits that are received in error have the previous and subsequent bits received correctly (since $z > Z$ for these bits). On the other hand, if $T_b > T_z$ then many subsequent bits are received with $z < Z$, so large bursts of errors are likely. Finally, if $T_b \leq T_Z$, the fading can be neglected.

5.3 Finite State Markov Channels (FSMC)

5.3.1 Motivation

Most wireless systems are very complex with many components, therefore the received signals are time varying and have a highly non-linear behavior. Reasonable approximations and assumptions are unavoidable. Since wireless channels are randomly time-varying systems, they are well approximated using the concepts of stationarity and linearization. This implies that long time periods can be discretized into several short time intervals. Over such intervals, the system behavior can be considered to

change very slowly. Thus, the time-varying system can be modeled as several discrete linear time-invariant systems. Moreover, the performance of many wireless systems is measured via empirical methods, i.e., simulation. Time invariant simulation implies that during the simulation period, the signal properties do not change. Therefore, if the system parameters over this period change very slowly, the simulation results will be sufficiently accurate [42].

As explained in Section 5.1 if both the transmitter and receiver are stationary, and there is no object motion between them, the resulting BER is due to pathloss and shadowing only. If the channel remains constant during a symbol time or over a number of symbols, The channel is said to be *quasi-static* [42].

The performance of any communication system is influenced by the behavior of the underlying physical channel. In the digital world, a wireless system is composed of data which is channel encoded and modulated for transmission over the channel. Therefore, a discretized channel model can provide improved tractability at a suitable granularity level (symbol/bit). For example the number of bits that are corrupted by a deep fade is simpler to compute at a packet level rather than over a continuous waveform.

The most widely used discrete channel model is the Finite State Markov Channel (FSMC) model. What a FSMC model, the underlying time varying channel is represented by a Discrete Time Markov Chain (DTMF). The most commonly used FSMC models are *first-order*, referred to as *memoryless* channel models. In these models, transitions occur between channel states under the assumption that there is no temporal correlation in the transition mechanism. First order FSMC models are adequate to model a wide range of narrowband (flat) fading channels in which there is no ISI, thus these models are the focus of this chapter. Since this FSMC model is a constant (memoryless) Markov process which means that the state transition probabilities are independent of the time at which they occur, i.e., the current state only depends on the previous state. Most fading mitigation techniques such as diversity, adaptive modulation and coding require a discrete analysis of the underlying channel [36], which can be obtained using a FSMC channel model. .

5.3.2 FSMC Model Evolution

The FSMC concept started with the Gilbert-Eliott (GE) model [43], [44] in the early 1960s. The GE model describes a channel with two distinct states, *good* (S_0) and *bad*

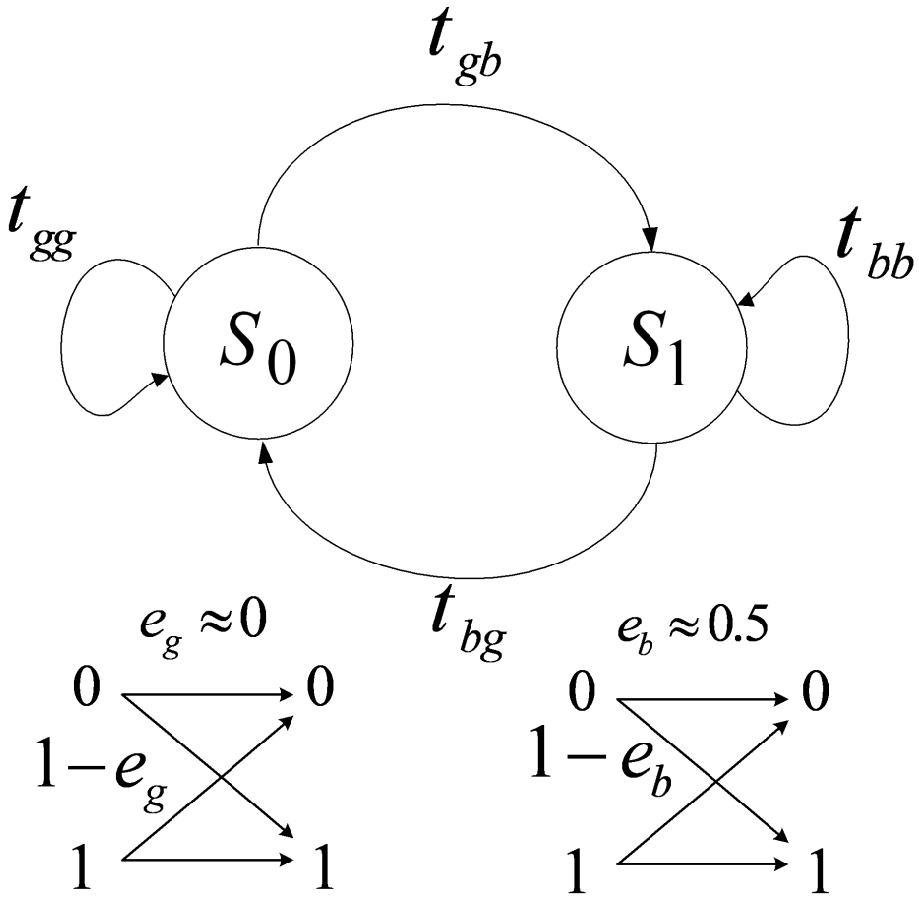


Figure 5.3: The Gilbert-Eliott channel showing the underlying discrete memoryless binary symmetric channels associated with each state.

(S_1). Each state is governed by an underlying discrete memoryless binary symmetric channel (BSC) as shown in Figure 5.3. The GE model is the first non-trivial model of burst-noise channels with memory. In Figure 5.3, the two states S_0 and S_1 refer to "noiseless" and "noisy" channels, respectively. The states transition probabilities are t_{gg}, t_{bb}, t_{gb} and t_{bg} where the subscripts indicate the corresponding state. The underlying BSCs demonstrate the error (crossover) probabilities associated with each state. During the good state, S_0, most bits transmitted through the channel arrive error free as $e_g \approx 0$. The bad state is assumed to be very noisy with a crossover probability of the associated binary symmetric channel $e_b \approx 0.5$. Gilbert calculated the capacity of the channel based on this assumption. Later, Elliott modified the model allow a non-zero probability of error for the noiseless channel ($e_g > 0$).

Sadeghi and Kennedy in [3], presented a chronological survey and evaluation of

both first and higher order FSMCs. Arauz in [2] considered only first order Rayleigh FSMCs. FMSC channels from an information theoretic point of view was developed by Goldsmith and Varayia. They examined the channel capacity and information rates of FSMC channels with iid inputs [45]. Almost concurrently, Wang and Chang in [46] developed a mutual information metric to validate the use of a first order Markov model. The main goal of the metric is to confirm that given the information about the previous symbol, the uncertainty of the current symbol is negligible. This uncertainty is measured in terms of average mutual information of the received amplitudes. Their goal was to compute the ratio of average mutual conditional information of two consecutive received symbols amplitudes, quantified as the average mutual information.

Zorzi and Rao in [47] examined the importance of the product $f_m T_s$[2] on the the first order assumption, where $1/T_s$ is the symbol rate. If $f_m T_s$ is in the range of 2×10^{-3} to 4×10^{-2}, the ratio of conditional mutual information between two consecutive symbols is less than 1%. This means that the information of a current symbol depends mainly on the previous symbol, which verifies the first order FSMC concept. On the other hand, Tan and Baulieu in [48] state that the mutual information of consecutive symbols is not sufficient to determine if the process is Markovian unless the symbols are highly correlated. They proposed a stochastic analysis approach using the autocorrelation function and suggested that a first order FSMC model is only appropriate for very slow fading applications.

For high symbol transmission rates, the product $f_m T_s < .01$ results in a similar result with [48] for distinct values of sample separations. Having described the accuracy of a FSMC model under very slowly fading conditions leads to conclude that an uncorrelated model is suitable under fast fading conditions. In this situation, a second or higher order FSMCs are required.

Wang and Moayeri [49] considered a FSMC model for radio communication in a flat Rayleigh fading channel. The novelty of the work in [50] lies in directly linking the Markov chain parameters, namely the state transitions, steady state probabilities and error probabilities, with a physical channel represented by the well known Clark flat fading model. They incorporated the time-varying nature of the fading channel by extending the GE model to a K state model. Each state constitutes a BSC with appropriate crossover probabilities. A state therefore represents a particular channel

[2]The product $f_m T_s$ is commonly known as the *fade speed* and is explained later when the K-state FSMC model is presented.

probability of error. This model is widely used in the literature and has been extended to non-flat (frequency selective) and fast fading channels. This model is examined in detail in the following sections where the parameters and methodology of the FSMC model are explained.

5.3.3 Methodology

Let the set $S^{(n)} = [s_0,\ s_1 \ldots, s_n]$, $n = 0, 1, \ldots,$ represent a constant (stationary) Markov process for discrete time step intervals t_n. The Markov process is said to be constant if it satisfies the property of *stationarity*. Synonymously, the Markov process is defined to be at *steady state* if it is independent of the time step index n, so that a transition to a new state s_n only depends on the present state s_{n-1}. We then have

$$\begin{aligned}\Pr\left[S(t_n) = s_n | S(t_{n-1}) = s_{n-1}, S(t_{n-2}) = s_{n-2}, \ldots, S(t_0) = s_0\right] \\ = \Pr\left[S(t_n) = s_n | S(t_{n-1}) = s_{n-1}\right] \qquad (5.13)\end{aligned}$$

This property is also referred to as the *first order* assumption and is the theory behind FSMCs [2]. As discussed above, this is a reasonable assumption for many practical wireless communication systems when the channel characteristics do not change rapidly over a number of symbols. Hence, the case of slow (narrowband) fading which is the focus of the work presented in this chapter.

A K-state FSMC represented by the state sequence $\mathbf{S} = \{S_0, S_1, \ldots, S_{K-1}\}$ is shown in Figure 5.4. The states have self-state and cross-state transition probabilities $t_{k,k}$ and $t_{k,k\pm1}$, respectively. From Figure 5.4, it is clear that the GE model shown in Figure 5.3 is a special case of a K-state FSMC. The connection between the time-varying nature of the real physical channel and its error process is encompassed in the elements of the vector $\mathbf{e} = [e_0, e_1, \ldots, e_{K-1}]$. Naturally, the FSMC concept is restricted to a discrete channel model. Figure 5.4 shows that a FMSC is a model of a continuous wireless physical channel that has been *partitioned* into a discrete finite set of states. The states correspond to the time varying amplitude of the received signal, or equivalently, the received power (SNR) γ [3].

The channel partitioning methodology to create the FSMC is shown in Fig 5.5. The continuous *pdf* waveform of the SNR γ, where $0 \leq \gamma < \infty$, is partitioned into a finite number of K contiguous states such that $S_0, S_1 \ldots S_{K-1}$. Each state has a

[3]The received power is typically denoted by the ratio of the received power P_r to the noise power spectral density $N_0/2$ within a bandwidth $2B$, so that $\gamma = P_r/N_0B$.

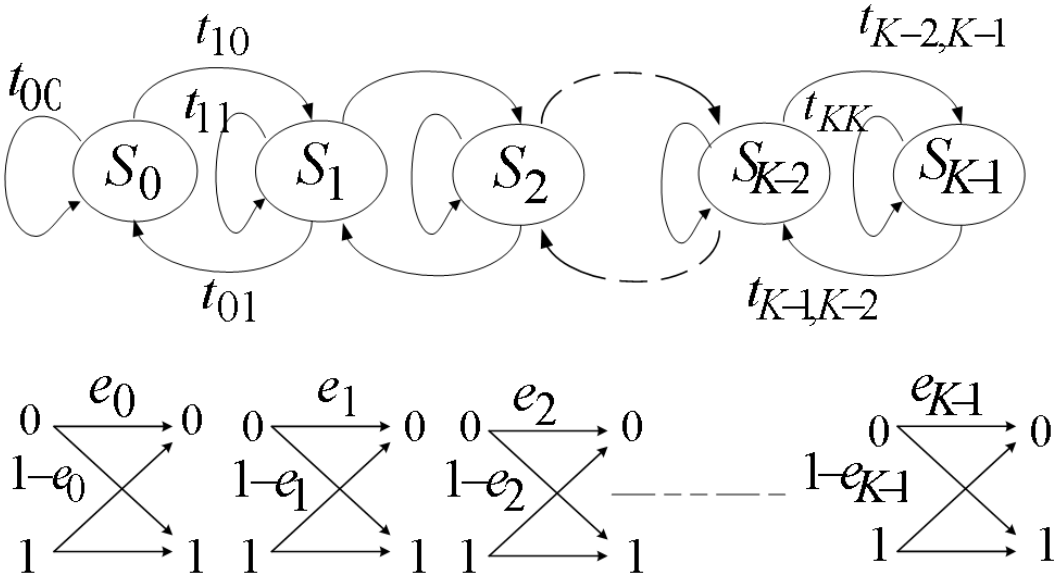


Figure 5.4: The set of discrete states S_k in the FSMC model where each state is associated with a a discrete memoryless BSC.[2]

steady state probability π_k and is bounded between two SNR thresholds, Γ_k and Γ_{k+1}, i.e., $\pi_k = \Pr(\Gamma_k \leq \gamma < \Gamma_{k+1})$ where $k = 0, 1, 2, \ldots, K-1$. The partitioned SNR *pdf* results in a series of discrete finite intervals which represents a finite state Markov chain. Therefore, while each state constitutes a range of SNR values, it is not characterized by the average SNR for that state. Instead, a state k is characterized by the corresponding BER e_k for a given modulation scheme. An advantage of discrete modeling of fading channels is the fact that each state of the FSMC represents a contiguous narrowband channel modeled as a simple BSC level.

Following the slow (narrowband) fading research in the literature, e.g. [50], [2], [51], [3], it is clear that a first order FSMC model should only allow transitions to occur to the same state or the immediate neighboring states. This dictates that a state s_k can only loop back to itself with a transition probability $t_{k,k}$ or its immediate neighbors $s_{k,k-1}$ or $s_{k,k+1}$ with probabilities $t_{k,k-1}$ and $t_{k,k+1}$, respectively. This assumption has been proven to be valid when the fading is slow (correlated) according to the

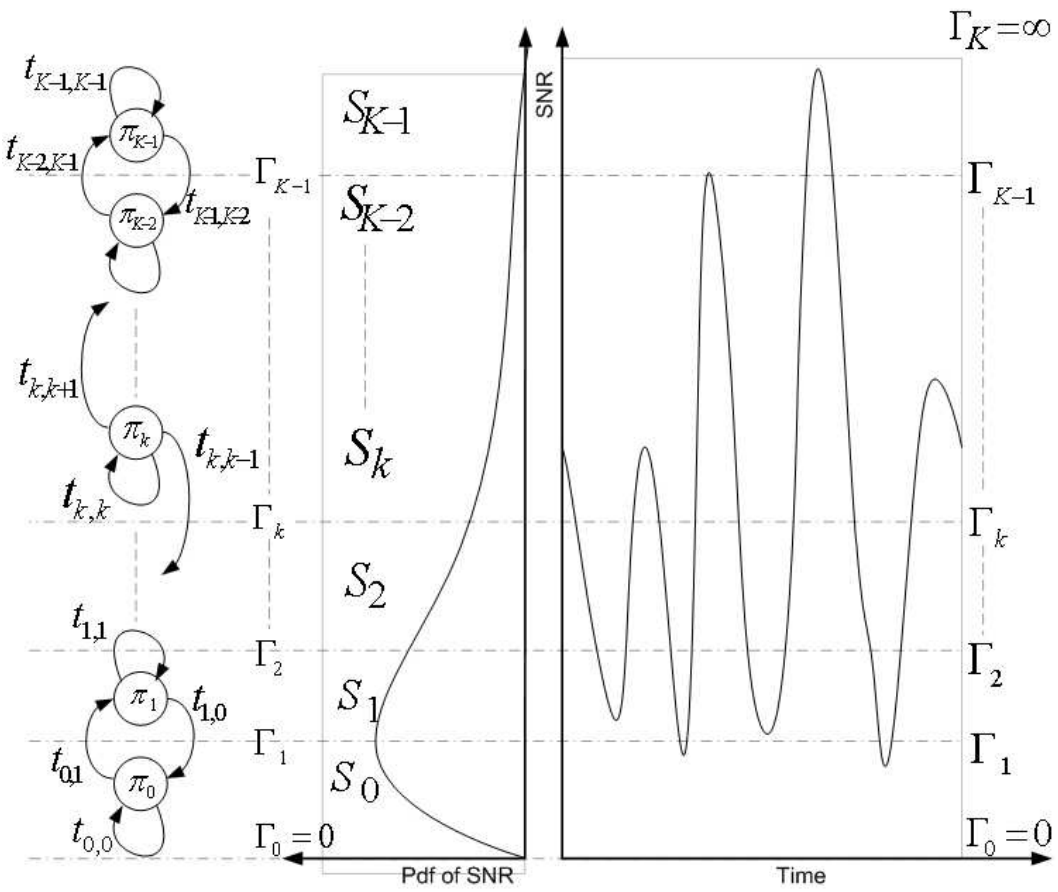


Figure 5.5: The relationship between the partitioned time varying signal to noise ratio (SNR) and the *pdf* of the SNR. The channel is in state π_k if the SNR lies between Γ_k and Γ_{k+1}.[3]

normalized fade speed factor $f_m T_s$. For example, Babich et al. in [52], [53] and Zorzi et el. in [47] showed that the fading can be accurately approximated by a first order FSMC model if $f_m T_s < 0.01$. However, the exact fade speed is subject to the number of states K and the partitioning methodology of the FSMC model, as will be explained in the next section. Sadegghi and Kennedy in [3] obtained the transition probabilities by integrating the joint channel *pdf* between two consecutive time indices over the state interval. Their results confirmed that the first order assumption is accurate when the temporal variations of the fading determined by the fade speed $f_m T_s$ is relatively small, provided that number of states is not too high, e.g. $K = 4 - 16$. With a first order FSMC, the state transition probabilities can be derived as a single simple function of the LCR N_k and the steady state probabilities π_k.

The components of the K-state FSMC are now described mathematically. We

then proceed to the Nakagami-m FSMC model.

5.3.4 Components of the FSMC

The three main components of a K-state FSMC [50] are derived as follows

1. Steady State probability vector π.

   A $K \times 1$ vector where each element π_k represents the steady state (stationary) probability of the FSMC being in state S_k at steady state as shown in Figure 5.5.

   Let $p(\gamma)$ and $F(\gamma)$ be the *pdf* and cumulative density function *cdf* of the instantaneous received SNR γ. The probability π_k that the SNR γ is in state s_k such that $\Gamma_k \geq \gamma < \Gamma_{k+1}$ is derived by integrating the channel *pdf* $p_\Gamma(\gamma)$ over the state boundaries Γ_k, Γ_{k+1}, given by

   $$\pi_k = \int_{\Gamma_k}^{\Gamma_{k+1}} p_\Gamma(\gamma) d\gamma = F_{\Gamma_{k+1}} - F_{\Gamma_k} \qquad k = 0, 1, 2, \ldots, K-1 \tag{5.14}$$

   and the vector must satisfy $\sum_{k=0}^{K-1} \pi_k = 1$

2. State transition probability matrix $\mathbf{T}$.

   A $K \times K$ transition matrix whose elements represent the transition probabilities between states (and to the same state). Hence, at time step n, the transition matrix element is given by

   $$t_{j,k} = \Pr\left[S_{n+1} = S_k | S_n = S_j\right] \qquad k, j \in \{0, 1, 2 \ldots K-1\} \tag{5.15}$$

   Note that $\mathbf{T}$ has a row (right) stochastic property given by $\sum_{k=0}^{K-1} t_{j,k} = 1$ and at steady state we have $\pi\mathbf{T} = \mathbf{T}$.

   As previously explained, γ stays within a given interval and only transits to the same interval or an adjacent one, so that $t_{j,k} = 0 \qquad \forall |k-j| > 1$

   According to [49], [50], the LCR N_k over an SNR partition k is much smaller compared to the average time spent in the state, therefore the state transition probabilities can be well approximated in terms of N_k and the steady state

probabilities π_k[4].

Let $R^{(k)}$ be the number of symbols transmitted per second when the system is in state s_k. Therefore the average number of symbols transmitted during the time spent in state k is $R^{(k)} = \pi_k/T_s$, where T_s is the symbol duration [2] .

The transition probabilities from state k to $k-1$, $(t_{k,k-1})$ are approximately given by

$$t_{k,k-1} \approx \frac{N_k}{R^{(k)}} \approx \frac{N_k}{\pi_k} T_s \qquad k = 1, 2, \ldots K-1 \tag{5.16}$$

Setting the boundary thresholds to $\Gamma_0 = 0$ and $\Gamma_K = \infty$, the transition probabilities from state k to $k+1$, $(t_{k,k+1})$ are approximately equal to

$$t_{k,k+1} \approx \frac{N_{k+1}}{R^{(k)}} \approx \frac{N_{k+1}}{\pi_k} T_s \qquad k = 0, 1, 2, \ldots K-2 \tag{5.17}$$

Clearly, the transition probability from a state k to itself is

$$t_{kk} = 1 - t_{k,k+1} - t_{k,k+1} \qquad k = 1, 2, \ldots K-2 \tag{5.18}$$

The remaining transition probabilities at the boundaries are then

$$\begin{aligned} t_{0,0} &= 1 - t_{0,1} \\ t_{K-1,K-1} &= 1 - t_{K-1,K-2} \end{aligned}$$

3. The Crossover probability vector **e**

   A $K \times 1$ vector with elements e_k which represent the crossover probabilities of the BSCs associated with the states S_k. Equivalently, e_k is the probability of a bit being in error in state K, $k = 0, 1, 2, \ldots, K-1$. The crossover probabilities must satisfy the constraint $0 \leq e_k \leq 0.5$ and the average error probability of the channel e over the K states is given by

$$e = \sum_{k=0}^{K-1} \pi_k e_k \tag{5.19}$$

[4]Wang and Moayeri in [50] verified this fact by simulation at reasonably slow fade speeds of $f_m T_s = .001$ and $f_m T_s = .0001$ with $K = 8$ partitions and equiprobable SNR states.

Following [49], [50], the error probability for each state is

$$e_k = \frac{1}{\pi_k} \int_{\Gamma_k}^{\Gamma_{k+1}} e(\gamma) p_\Gamma \gamma d(\gamma) \qquad k = 0, 1, \ldots K-1 \tag{5.20}$$

where $e(\gamma)$ is the average error probability of the continuous (non-partitioned) channel at a given instantaneous received SNR γ for a given modulation scheme.

The SNR thresholds define the intervals that represent the Markov chain states and the corresponding e_k. Therefore, a key to the accuracy and appropriateness of the FSMC model is the partitioning scheme [54]. In [50], the equal probability method (EPM) was introduced. This method assumes that the FSMC states are equiprobable and the SNR thresholds are chosen accordingly. Therefore for K states, the probability that the channel is in a particular state π_k in (5.14) is $\pi_0 = \pi_1 = \cdots = \pi_{K-1} = 1/K$. The EPM has been considered by many researchers for the Nakagami-m distribution [55], [56], because of its low computational complexity.

Another approach to partitioning which employs the Lloyd-Max quantizer was presented in [57] and [58]. The thresholds are chosen based on a central quantized point $\bar{\Gamma_k}$ represented by the average SNR for each state such that $\Gamma_k = \left(\bar{\gamma_{k+1}} + \bar{\gamma_k}\right)/2$ where $\bar{\gamma_k} = \left(\int_{\Gamma_k}^{\Gamma_{k=1}} \gamma p(\gamma) d\gamma\right)/\pi_k$. This method results in a minimum quantization MSE. However, the advantage of this method over the EPM may diminish as the number of states K increases according to [55] and [3].

Zhang and Kassam in [51] chose a partitioning strategy based on the average time duration of each state for the Rayleigh fading channel. The SNR thresholds and the number of states are determined based on a fixed state duration, hence the name equal duration method (EDM). It is assumed that the fading process is slow enough so that γ stays within the state thresholds Γ_k and Γ_{k+1} for a state duration. In this technique, the channel states defined by the SNR thresholds are specified according to the time a symbol is affected by fading. This is a much more efficient approach than the EPM since the EPM cannot model fading time variations as with the EDM [3]. Thus, the EDM is chosen in this chapter and the foundation of the future work presented in Chapter 6. The EDM is explained in detail in Section 5.4.

5.4 The Nakagami-m FSMC Model

The Nakagami-m distribution can be used to model a wide range of multipath fading conditions with different severity via the parameter m. In this chapter, the work in [51] is extended to present the analysis and performance of the EDM on the Nakagami-m fading channel. This approach can also be used to investigate the severity of fading and its duration.

At the network level, data transmission is packet-based, so the SNR thresholds bounding each state should be derived such that the average state duration (fade range) is an integer multiple of a number of symbols or packets. Since in most applications the packet duration is fixed, the model presented provides a versatile technique for linking the fade states to the packet duration.

The received SNR envelope is partitioned into discrete states based on the average duration of an SNR state. The EDM adopts an equal average fade range duration (AFRD) partitioning methodology to derive the state thresholds [1]. The AFRD statistical parameters for Nakagami-m fading are derived in the next section.

5.4.1 Statistical Parameters

The *pdf* and *cdf* of a Nakagami-m distributed signal SNR are

$$p_{\Gamma}(\gamma) = \left(\frac{m}{\bar{\gamma}}\right)^{m} \frac{\gamma^{m}}{\Gamma(m)} \exp\left(\frac{-m}{\bar{\gamma}}\gamma\right) \tag{5.21}$$

and

$$F_{\Gamma} = \frac{\gamma\left(m, \frac{-m}{\bar{\gamma}}\gamma\right)}{\Gamma(m)} \tag{5.22}$$

respectively where $\Gamma(.)$ and $\gamma(.,.)$ are the complete and upper incomplete Gamma functions given by

$$\begin{aligned} \Gamma(x) &= \int_{0}^{\infty} e^{-t} t^{x-1} dx \\ \gamma(x,a) &= \int_{0}^{a} e^{-t} t^{x-1} dx \end{aligned}$$

and t and a are arbitrary constants. Using (5.22), the steady state probabilities π_k for the Nakagami-m FSMC in (5.14) are given by

$$\begin{aligned} \pi_k &= F_\Gamma(\Gamma_{k+1}) - F_\Gamma(\Gamma_k) \\ &= \frac{\gamma\left(m, \frac{-m}{\bar{\gamma}}\Gamma_{k+1}\right) - \gamma\left(m, \frac{-m}{\bar{\gamma}}\Gamma_k\right)}{\Gamma(m)} \qquad k = 0, 1, 2, \ldots, K-1 \end{aligned} \tag{5.23}$$

where $\Gamma_0 = 0$. It is assumed that the fading is slow enough so that γ stays within the state thresholds Γ_k and Γ_{k+1} for a state duration. Transitions occur only to the current and adjacent states, i.e., $\Pr_{k,i} = 0$ if $|k-i| > 1\ \forall k$ [51].

Yacoub et al. in [41] showed that for a Nakagami-m distribution, the envelope z and its first order time derivative $\dot{z}$ are independent random variables. Moreover, the probability density function of the time derivative is also Gaussian distributed. Based on the joint distribution of z and $\dot{z}$ [37], the temporal (second order) statistics in Section 5.2.2 of the Nakagami-m fading characterized by the level crossing rate (LCR) and the average fade duration (AFD) can be obtained. Substituting the joint *pdf* of z and $\dot{z}$ in (5.10), the LCR for the Nakagami-m distribution is

$$N_z = \sqrt{2\pi} f_m \frac{m^{m-1/2}}{\Gamma(m)} \rho^{2m-1} \exp(-m\rho^2) \tag{5.24}$$

where $\rho = Z/\sqrt{\bar{\gamma}}$, f_m is the maximum Doppler shift and $\bar{\gamma}$ is the average SNR of the signal [55]. Hence, for state k with $\gamma_k = \Gamma_k$, the k^{th} LCR can be written as

$$N_k = \frac{\sqrt{2\pi} f_m}{\Gamma_m} \left(\frac{m}{\bar{\gamma}}\Gamma_k\right)^{m-1/2} \exp\left(-\frac{m}{\bar{\gamma}}\Gamma_k\right). \tag{5.25}$$

Define the average fade range duration (AFRD) $\bar{\tau_k}$ as the average time the received instantaneous SNR γ stays between two SNR levels defined by the thresholds Γ_k and Γ_{k+1} [1]. Following [37], [51], the AFRD $\bar{\tau_k}$ is ratio of the total time the received SNR stays in a given state τ_k measured over some arbitrary long time interval T multiplied by the expectation that the channel stays in that state, giving

$$\bar{\tau_k} = \frac{\sum \tau_k}{T} \times \frac{1}{t_{k,k+1} + t_{k,k-1}} \tag{5.26}$$

FMSC partitioning is typically based on the time that γ stays in the same state

or migrates to adjacent states. According to [50], [3] and [1] this time is typically a symbol period. However, in [51], the time is set to be a packet time T_p. Clearly, the latter assumption is valid only if $f_m T_p$ is sufficiently small.

In this dissertation, the partitioning time unit is set to be the IEEE 802.16 OFDM symbol period denoted henceforth by T_s where $T_s = 10\mu$S. This is a reasonable assumption since each burst in the IEEE 802.16 frame constitutes an integer number of symbols to form each subframe [10].

Using (5.26), (5.16) and (5.17) we have

$$\begin{aligned} \bar{\tau_k} &= \frac{T_s}{t_{k,k+1} + t_{k,k-1}} \\ &= \frac{\Pr[\Gamma_k \leq \gamma \leq \Gamma_{k+1}]}{N_k + N_{k+1}} \\ &= \frac{\pi_k}{N_k + N_{k+1}} \qquad k = 0, 1, 2, \ldots, K-1 \end{aligned} \tag{5.27}$$

Note that the AFRD is analogous to Jakes AFD in (5.11) [37] except that the latter defines the average duration where the instantaneous SNR γ stays above or below a certain fade level. Obviously, for the worst case when γ stays below a given fade level, there is only one fade region, i.e. $k = 0$. and hence the two definitions coincide [36].

From (5.27) it is shown that the value of $\bar{\tau_k}$ is a simple function of the channel Markov parameters presented by the steady state probability π_k and the k^{th} LCR N_k.

By substituting (5.23) and (5.24) in (5.27), the AFRD $\bar{\tau_k}$ is then given by

$$\bar{\tau_k} = \frac{\gamma\left(m, \frac{-m}{\bar{\gamma}}\Gamma_{k+1}\right) - \gamma\left(m, \frac{-m}{\bar{\gamma}}\Gamma_k\right)}{\Gamma(m)} \times \frac{1}{D} \tag{5.28}$$

with

$$D = \frac{\sqrt{2\pi} f_m}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^{m-1/2} \left[\Gamma_k^{m-1/2} \exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right) + \Gamma_{k+1}^{m-1/2} \exp\left(\frac{-m}{\bar{\gamma}}\Gamma_{k+1}\right)\right] \tag{5.29}$$

Based on (5.23), Γ_k and $\Gamma_{k\pm1}$ characterize the boundaries of partition k. Therefore, the set of Γ_k must be found for the K partitions where $\Gamma_0 = 0$, $\Gamma_K = \infty$ and $k = 0, 1, \ldots, K-1$. For K channel partitions, we obtain a set of K equations in $K-1$ unknowns in Γ_k.

Following [51], the continuous SNR envelope is partitioned using EDM such that the AFRD $\bar{\tau_k}$ is some multiple of the OFDMA symbol length T_s by introducing a factor c_k. Then for a fixed average fade duration for all K partitions, c_k becomes a constant value c. To obtain a system of K equations in K unknowns, the factor c is employed as the K^{th} variable. We then have

$$\bar{\tau_k} = c\,T_s. \qquad k = 1, 2, \ldots K-1 \tag{5.30}$$

where $\Gamma_0 = 0$ and $c \geq 1$.

5.4.2 The AFRD on a Nakagami-m FSMC model

It is clear that solving the K equations to obtain the set of Γ_k thresholds results in a unique value of c for each choice of K and $f_m T_s$. This requires that for a given fade speed $f_m T_s$, the number of partitions K must be carefully chosen so that the value of c obtained is reasonable. This means that the average state duration (AFRD) $\bar{\tau_k}$ in (5.26) must be large enough to guarantee that transitions are almost always to the same or adjacent states to maintain the first order assumption.

According to results in the literature, there is no clear solution to optimize the number of states K for an FSMC model [3]. This is because the number of states is a trade off between accuracy (partitioning granularity) and complexity. However, a range of $K = 4$ to 16 is widely used [46], [50], [57], [59], [56], [60], [61] and [55]. this gives a very good balance between accuracy and model complexity and tractability. The number of states K can be chosen to be aligned with the modulation signaling with adaptive modulation and coding techniques [1], [36].

The model presented in this chapter provides the freedom to choose the value of c for a given number of partitions K so that the state duration is a specific number of symbols, such as a packet. Of course, for a specific K, c must be chosen to adequately capture the fading characteristics.

The first task in this chapter is to develop the AFRD model for Nakagami-m fading according to the EDM in [51]. Next, the SNR thresholds are derived for a given c. To achieve these goals, the model is formulated as an optimization problem.

Rearranging (5.26), the equation for the k^{th} AFRD can be written as

$$\begin{aligned}\gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_{k+1}\right) &- \gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_{k}\right)\\ &= \bar{\tau_k}\sqrt{2\pi}f_m\left(\frac{-m}{\bar{\gamma}}\right)^{m-1/2}\\ &\cdot \left[\Gamma_k^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right)+\Gamma_{k+1}^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_{k+1}\right)\right] \quad (5.31)\end{aligned}$$

Let $\bar{\tau_k} = cT_s$ in (5.31), and move the left hand side to the right to produce a set of K equations equal to 0. Each equation is a system state function $f_k(\Gamma_k, \Gamma_{k+1})$, so we have

$$\begin{aligned}f_k(\Gamma_k,\Gamma_{k+1}) &= \gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_{k+1}\right)-\gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_{k}\right)\\ &- \sqrt{2\pi}f_m\,c\,T_s\left(\frac{-m}{\bar{\gamma}}\right)^{m-1/2}\\ &\cdot \left[\Gamma_k^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right)+\Gamma_{k+1}^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_{k+1}\right)\right] \quad (5.32)\end{aligned}$$

$\Gamma_0 = 0$ and Γ_K is chosen to be reasonably high such that when $\Gamma_K \geq \gamma$, the channel is considered one state with a negligible BER[5]. The set of K system functions is

$$\begin{aligned}f_1(\Gamma_1) &= \gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_1\right)\\ &- \sqrt{2\pi}\;c\,f_m\,T_s\left(\frac{-m}{\bar{\gamma}}\right)^{m-1/2}\left[\Gamma_1^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_1\right)\right]\\ &\vdots\\ f_{K-1}(\Gamma_{K-1},\Gamma_K) &= \gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_K\right)-\gamma\left(m,\frac{m}{\bar{\gamma}}\Gamma_{K-1}\right)\\ &- \sqrt{2\pi}\,c\,f_m T_s\left(\frac{-m}{\bar{\gamma}}\right)^{m-1/2}\\ &\cdot \left[\Gamma_K^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_K\right)+\Gamma_{K-1}^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_{K-1}\right)\right]\end{aligned}$$

[5]For example, Γ_K can obtained from (5.12) and (5.21) by computing the fade margin at $P_{out} = 5\%$, and solving numerically for $z_{min} = \Gamma_K$. $P_{out} = 5\%$, means that for 95% of time, $e < 10^{-4}$ [1].

which can be represented in vector form as

$$\mathbf{f} = \begin{bmatrix} f_1(\mathbf{\Gamma}) & f_1(\mathbf{\Gamma}) & \dots f_{K-1}(\mathbf{\Gamma}) \end{bmatrix}^T \tag{5.33}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_1 & \Gamma_1 & \dots \Gamma_{K-1} \end{bmatrix}$$

and T indicates transpose.

Following [62], the set of equations given by (5.33) can be solved as an optimization problem to obtain the values of Γ_k and c for a given K and $f_m T_s$. The solution is a vector $\mathbf{\Gamma}^*$ such that the functions in (5.33) are all zero. This problem can be formulated as the minimization of an objective function $\mathbb{F}$. Therefore, the following constrained optimization problem must be solved

$$\text{Minimize } \mathbb{F} = \sum_{k=0}^{K-1} |f_k(\mathbf{\Gamma})|^2 = \mathbf{f^T f} \tag{5.34}$$

subject to : $0 < \Gamma_1 < \Gamma_2 < \dots \Gamma_{K-1} < \Gamma_K$.

Since each function f_k depends on Γ_k, the gradient of the set can be represented by the Jacobian matrix. By differentiating (5.34) with respect to Γ_k we have

$$\frac{\partial \mathbb{F}}{\partial \Gamma_k} = \sum_{k=0}^{K-1} 2 f_k \frac{\partial f_k}{\partial \Gamma_k} \tag{5.35}$$

or in matrix form

$$\begin{bmatrix} \frac{\partial \mathbb{F}}{\partial \Gamma_1} \\ \frac{\partial \mathbb{F}}{\partial \Gamma_2} \\ \vdots \\ \frac{\partial \mathbb{F}}{\partial \Gamma_{K-1}} \end{bmatrix} = 2\mathbf{J}^T \begin{bmatrix} f_0(\mathbf{\Gamma}) \\ f_1(\mathbf{\Gamma}) \\ \vdots \\ f_{K-1}(\mathbf{\Gamma}) \end{bmatrix} \tag{5.36}$$

The derivative of the upper incomplete Gamma function exists, so the Jacobian matrix $\mathbf{J}$ for $\mathbf{f}$ can be derived for the optimization problem [62]. This can be obtained from the first derivative of the regularized Gamma function $P_x(x)$ which is $\frac{d}{dx} P_x(a, x) = \frac{e^{-x} x^{a-1}}{\Gamma(a)}$, where a is an arbitrary constant.

From (5.32) and (5.36), the Jacobian matrix is given by

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial \Gamma_1} & \frac{\partial f_1}{\partial \Gamma_2} & \cdots & \frac{\partial f_1}{\partial \Gamma_{K-1}} \\ \frac{\partial f_2}{\partial \Gamma_1} & \frac{\partial f_2}{\partial \Gamma_2} & \cdots & \frac{\partial f_2}{\partial \Gamma_{K-1}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_{K-1}}{\partial \Gamma_1} & \frac{\partial f_{K-1}}{\partial \Gamma_2} & \cdots & \frac{\partial f_{K-1}}{\partial \Gamma_{K-1}} \end{pmatrix} \tag{5.37}$$

The elements of the Jacobian $J_{l,j}$, $l, j = 1 \leq l \leq K-1$ are given by

$$\begin{aligned} \frac{\partial f_k}{\partial \Gamma_k} &= \left(\frac{m}{\bar{\gamma}}\right) \exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right) \left(\frac{m}{\bar{\gamma}}\Gamma_k\right)^{m-1} \\ &+ c\left[\Gamma_k^{m-3/2}(m-1/2)\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right) - m\,\Gamma_k^{m-1/2}\exp\left(\frac{-m}{\bar{\gamma}}\Gamma_k\right)\right] \end{aligned} \tag{5.38}$$

To account for the constraint $0 < \Gamma_1 < \Gamma_2 < \ldots \Gamma_{K-1} < \Gamma_K$, we apply the non-negativity bound variable transformation $x_i = y_i^2$ [62]. Therefore we have

$$\begin{aligned} \Gamma_1 &= y_1^2 \\ \Gamma_2 &= y_2^2 + y_1^2 \\ &\vdots \\ \Gamma_{K-1} &= \sum_{k=1}^{K-1} y_i^2 \end{aligned} \tag{5.39}$$

and following the chain rule, the derivatives of the transformed variables are obtained

$$\frac{\partial f_k}{\partial y_k} = 2y_k \sum_{k=1}^{K-1} \frac{\partial f_k}{\partial \Gamma_k} \tag{5.40}$$

The gradient from (5.36) is given by

$$\mathbf{g_F} = 2\mathbf{J^T f} \tag{5.41}$$

Following [62], $\mathbb{F}^{(i)}$ and $\mathbf{\Gamma}^{(i)}$ can be iteratively reduced to a solution in the neighborhood of $\mathbf{\Gamma}^*$. Therefore, $f_k(\mathbf{\Gamma}^{(i)})$ can be accurately represented as a linear approximation of the Taylor series. The second derivative of $f_k(\mathbf{\Gamma})$ is then neglected and the

Hessian of $\mathbb{F}$ can be approximated as

$$\mathbf{H_F} \approx 2\mathbf{J^T J} \tag{5.42}$$

The resulting recursive relation for the optimization is

$$\mathbf{\Gamma}^{(i+1)} = \mathbf{\Gamma}^{(i)} - \alpha\,(\mathbf{2J^T J})\mathbf{2J^T f} \tag{5.43}$$

where $\alpha = 10^{-7}$ is the step size that minimizes $\mathbb{F}(\mathbf{\Gamma}^{(i)} + \alpha \mathbf{g_F}^{(i)})$ [62]. $\mathbf{\Gamma}^*$ represents the thresholds $\Gamma_1, \Gamma_2, \ldots, \Gamma_{K-1}$ obtained through optimization.

5.5 Numerical Results

In this section, the accuracy of both the Zhang-Kassam [51] and the optimized AFRD models for Nakagami-m fading are investigated. Table 5.1 shows the effect of the fading severity of fade on the average fade time duration (AFRD) via the value of c at different fade speeds with $K = 4$. The values were obtained for a symbol time of $T_s = 10.6\mu$ s at different vehicular speeds spanning the fade speed range of $0.0009 \leq f_m T_s \leq 0.0021$. It is clear that, as the fade speed increases, the average fade time decreases. In other words, the signal has a smaller duration under a given fade level. This shows that as the fade intensity decreases, i.e. increasing m, the average fade duration also decreases. Thus in non-severe fading conditions, the signal is less likely to remain below a given fade level for a long time. This shows that the severity of the fading contributes to the average fade duration as well as the fade speed. Obviously, the number of partitions must me chosen so that the resulting value of c is reasonable [51].

Table 5.2 shows the same system parameters as above with a greater number of partitions $K = 8$. It is clear that as the number of partitions increases, the AFRD decreases. This is because the amount of time the signal spends between two closer thresholds decreases. Thus the fade intensity also has a significant effect on the AFRD for a higher number of partitions. For example, to achieve a reasonable value of $c = 90$ at 80 km/hr with Rician fading, $K = 4$ partitions are needed. On the other hand, $K = 8$ partitions are needed to maintain an approximately close value $c = 92$ at 20 km/hr with $m = 3$.

To confirm the accuracy of the AFRD Nakagami-m model, the BER of each state

m	5 km/hr	10 km/hr	20 km/hr	50 km/hr	80 km/hr	120 km/hr
1 (Rayleigh)	1461.7	781.09	390.54	156.21	92.63	78.42
1.6 (Rician)	1442.13	724.36	362.18	144.86	90.53	68.76
2	1346.41	673.36	336.67	136.21	88.47	59.98
3	712.88	356.41	178.21	71.27	44.62	31.26
4	245.47	122.74	61.37	24.52	15.52	10.41
5	63.033	31.51	15.74	6.28	4.11	2.76

Table 5.1: AFRD c values at different fade intensities for various m, $K = 4$ partitions and $\bar{\gamma} = 24$ dB

m	5 km/hr	10 km/hr	20 km/hr	50 km/hr	80 km/hr	120 km/hr
1 (Rayleigh)	770.31	385.19	192.59	77.03	39.02	32.65
1.7 (Rician)	748.86	374.43	187.22	74.88	35.05	22.26
2	714.15	357.75	178.53	71.43	34.77	15.36
3	395.05	197.52	92.76	39.51	24.70	3.51
4	139.89	69.95	34.95	14.03	8.81	-
5	36.68	18.32	9.15	3.76	2.37	-

Table 5.2: AFRD c values at different fade intensities for various m, $K = 8$ partitions and $\bar{\gamma} = 24$ dB

e_k for $K = 8$ states and QPSK modulation is shown in Figure 5.6. As given in (5.20), e_k depends on the state thresholds Γ_k, Γ_{k+1} and the modulation order. A Nakagami-m fading simulator was implemented using Matlab. A low average SNR $\bar{\gamma} = 7$ dB was used to show the range of e_k from $k = 1$ (very bad) to $k = 8$ (very good) at different fade intensities m. At $f_m T_s = 0.0088$, the figure shows that there is good agreement between the theoretical and simulation results at various fade intensities.

In Figure 5.7 the steady state vector π_k in (5.14) is shown for different values of m. An interesting observation is that at $m = 3$, the probability that the channel is in the higher index states $k \geq 4$ is greater than for $m = 1$. This is due to the fact that at the lower fade condition $(m - 3)$, it is more probable that the channel will be in a higher (better) state. For example, the probability of being in states $k = 5$ and 6 at $m = 3$ is twice the corresponding probabilities with Rayleigh fading ($m = 1$). This figure shows good agreement between the analytical and simulation results, which confirms the accuracy of the analysis.

Figures 5.8, 5.9 and 5.10 show the state transition probabilities $t_{k,k+1}$, $t_{k,k-1}$ and $t_{k,k}$ of the AFRD model, respectively, for $c = 8$ and 16. The values of $c = 8$ and $c = 16$ were chosen to approximate $c = 9.1981$ and $c = 13.6981$, respectively, computed

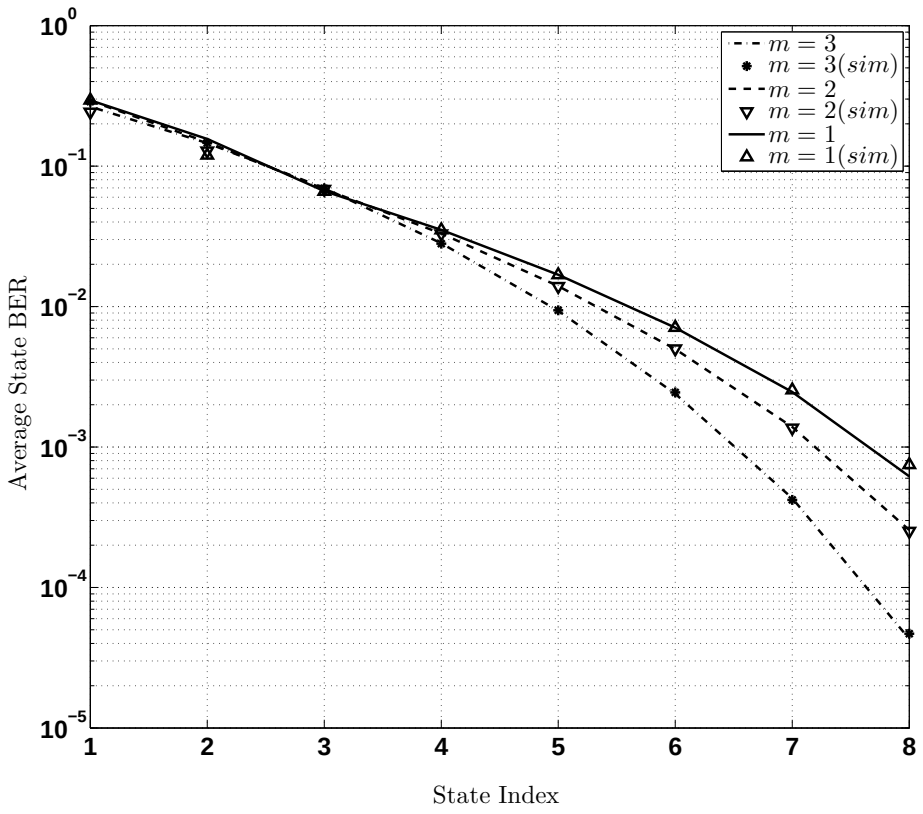


Figure 5.6: State BER e_k versus the state index for various fade intensities m, $\bar{\gamma} = 7$ dB, $f_m T_s = 0.0088$ and QPSK modulation.

using the Zhang-Kassam model at $f_m T_s = 0.0088$, $\bar{\gamma} = 20$ dB and $m = 2$. This approximation to integer c results in realistic values that represent IEEE 802.16 bursts containing 80 and 160 OFDM symbols, respectively [10]. These values of c provide transition probabilities $t_{k,k+1}$,$t_{k,k-1}$ which confirm the validity of a first order FSMC model as there is good agreement between the simulation and analysis. At $c = 8$, the transition probabilities $t_{k,k+1}$, $t_{k,k-1}$ are higher than at $c = 16$. The results given verify that the Zhang-Kassam model with a lower c results in a higher number of transitions, so a larger value of c cannot be used to represent a lower fade speed. Finally, the corresponding steady state vector values π_k are given in Figure 5.11. This further confirms the accuracy of the model.

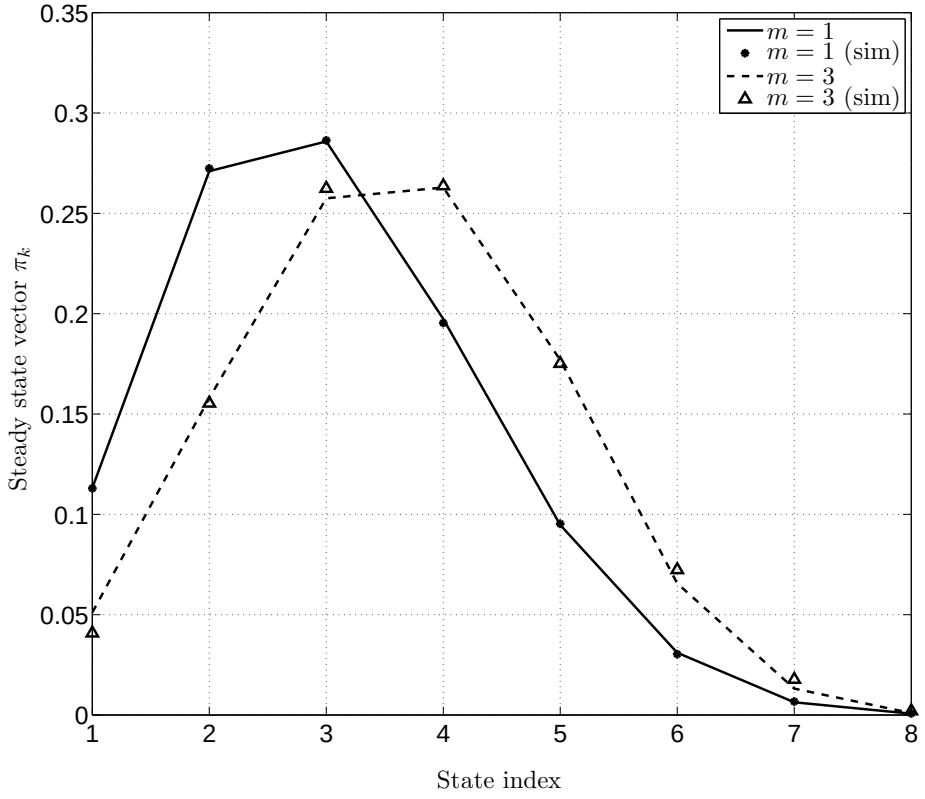


Figure 5.7: Steady state vector π_k vs state index, $f_m T_s = 0.088$, at $m = 1$ (Rayleigh) vs $m = 3$

5.6 Cross-Layer Markov Modeling for Random Access

In Figure 5.12, a cross-layer model is shown which links the FSMC model introduced previously and the random access model developed in Chapter 2. In this model, the effects of the physical channel derived in the FSMC model are conveyed to the MAC layer to provide a random access system involving channel errors. Knowledge of the channel state S_k dictates the BER and the steady state probability e_k and π_k in a time slot, respectively. For simplicity, the two-state (GE) FSMC shown in Figure 5.3 is used where $k = \{0, 1\}$, however the model can be extended to employ any K-state FSMC. Let π_0 and π_1 be the steady state probabilities of the FSMC being in states S_0 and S_1, respectively. The probability x indicates a non-collided request. If no collision occurs and the received SNR $\gamma > \Gamma_K$ (where the channel BER is low), the

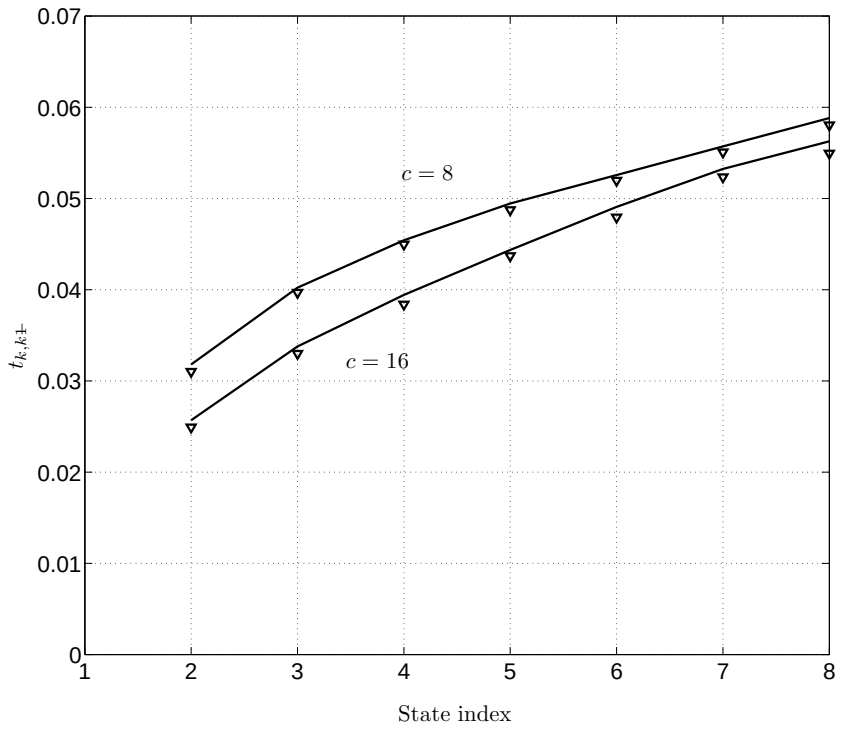


Figure 5.8: Transition probabilities $t_{k,k-1}$ for $f_m T_s = 0.008$, $\bar{\gamma} = 10$, $m = 2$, and $c = 8$ and 16.

request access process is a success and the system enters the transmitting state s_t. Assuming the bit errors in the request packet are independent, the packet error rate of a request packet of N_b bits e_p is given by

$$e_p = 1 - (1 - e_0)^{N_b} \tag{5.44}$$

where e_0 is the BER in state S_0 of the FSMC.

The model in Figure 5.12 is now described. Starting from an idle state, the request arrival per frame is Bernoulli distributed with probability a. The system is frame synchronized and thus the system state is determined by the acknowledgement in the downlink subframe. If no acknowledgement is received, the request is a failure, i.e., the request packet is assumed to either have been collided, or invalid due to the channel being in a deep fade (S_1), and the system enters the first collision state s_{c_1}. Similar to the mechanism of the Markov model in Chapter 2, if another failure occurs, i.e., no acknowledgement is received in the subsequent downlink subframe,

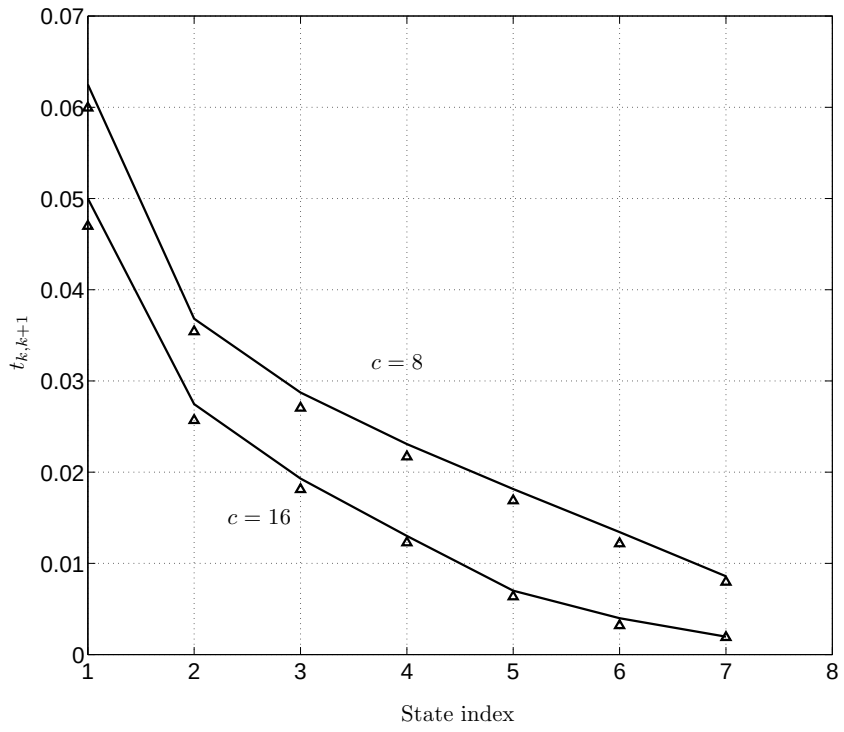


Figure 5.9: Transition probabilities $t_{k,k+1}$ for $f_m T_s = 0.008$, $\bar{\gamma} = 10$, $m = 2$, and $c = 8$ and 16.

the system enters a finite retrial phase of m stages $s_{c_{1\ldots m}}$. If a failure occurs after m retransmission attempts (state s_{c_m}), the request packet is discarded and the SS exits the contention process by returning to the idle state s_i. If a negative acknowledgement is received, the request sent is erroneous due to the channel being in error (S_0) and the system enters an error state s_e. The request can be retransmitted in a subsequent uplink subframe using error correction coding to correct the erroneous bits.

From Figure 5.12, the system transition matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{bmatrix} 1-a & 1-a & 0 & 0 & \cdots & \gamma_m(1-x)+\gamma_m x\pi_1 \\ ax\pi_0(1-e_p) & ax\pi_0(1-e_p) & x\pi_0(1-e_p) & \gamma_1 x\pi_0(1-e_p) & \cdots & \gamma_m x\pi_0(1-e_p) \\ ax\pi_0 e_p & ax\pi_0 e_p & x\pi_0 e_p & \gamma_1 x\pi_0 e_p & \cdots & \gamma_m x\pi_0 e_p \\ a(1-x)+ax\pi_1 & a(1-x)+ax\pi_1 & (1-x)+x\pi_1 & 1-\gamma_1 & \cdots & 0 \\ 0 & 0 & 0 & \gamma_1(1-x)+\gamma_1 x\pi_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-\gamma_m \end{bmatrix} \tag{5.45}$$

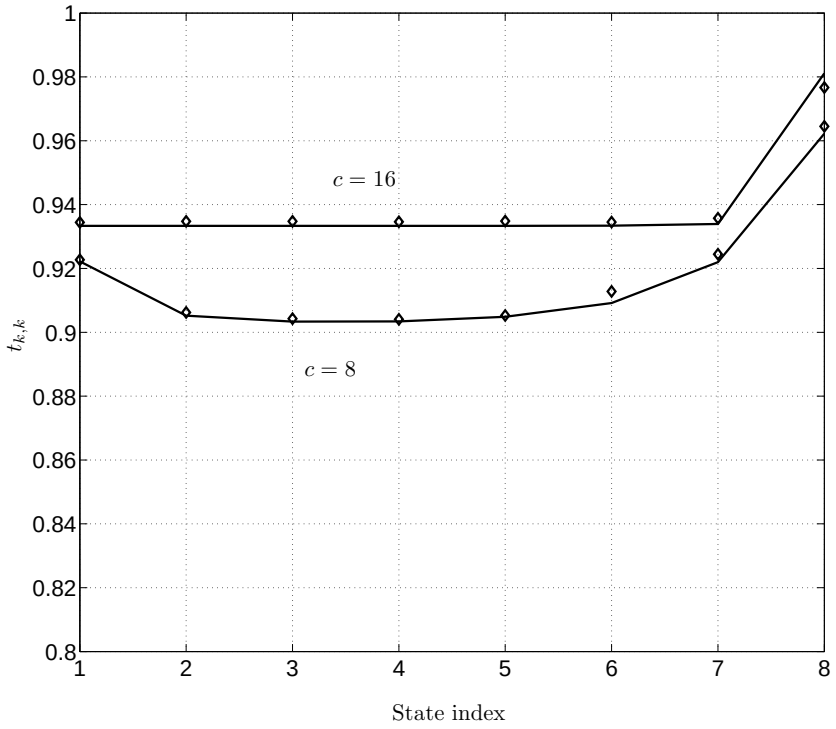


Figure 5.10: Transition probabilities $t_{k,k}$ for $f_m T_s = 0.008$, $\bar{\gamma} = 10$, $m = 2$, and $c = 8$ and 16.

The state distribution vector $\mathbf{s}$ is given by

$$\mathbf{s} = [\ s_i \quad s_t \quad s_e \quad s_{c_1} \ldots s_{c_m}\ ]^t \tag{5.46}$$

where t denotes transpose.

From the equilibrium equation (4.2), knowing that $\sum_{j=1}^{m} s_j = 1$, the unknown components s_j of the distribution vector $\mathbf{s}$ can be obtained [21]. Figure 5.12 shows that the state BERs (crossover probabilities) e_k of the FSMC contribute significantly to the system state transition probabilities of the random access process in the MAC layer. Further, this represents a cross-layer model linking the FSMC model with the random access multistage backoff model in Figure 2.2.

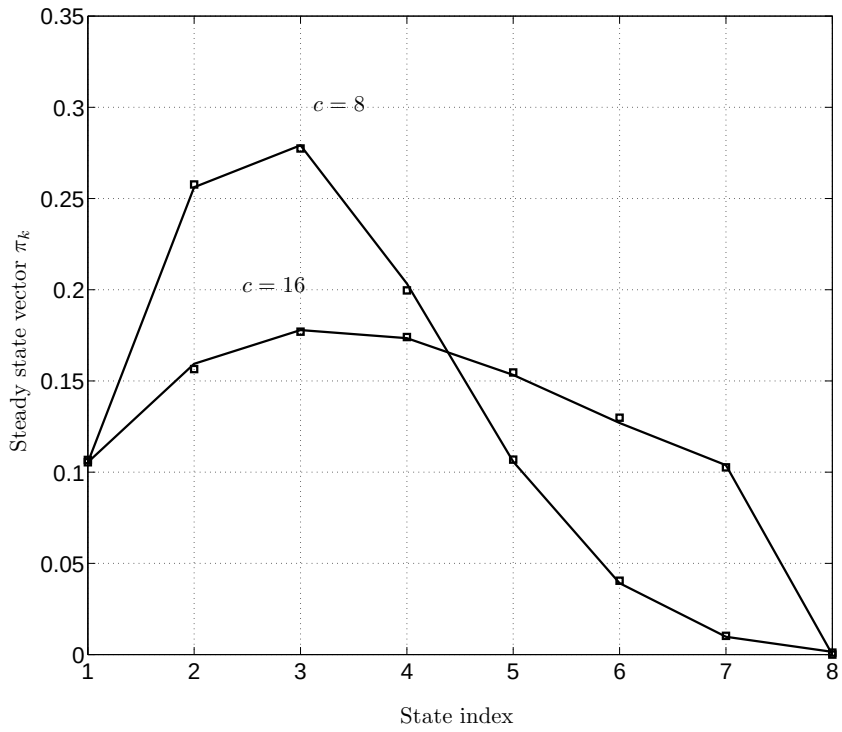


Figure 5.11: Steady state vector π_k for $f_m T_s = 0.008$, $\bar{\gamma} = 10$, $m = 2$, and $c = 8$ and 16.

5.7 Summary

In this chapter, an optimized finite state Markov channel (FSMC) model for Nakagami-m fading channels was presented. Based on the equal time duration method, the average fade range duration (AFRD) technique provides optimized SNR thresholds for the FSMC. The effect of the fade intensity on the AFRD was investigated by considering various Nakagami fading parameters m. It was shown that the fade intensity contributes to the average FSMC state time as well as the fade speed. The state duration was specified as an integer multiple of the symbol duration. Since data packets consists of integer multiples of symbols, this duration can be chosen to be a packet or a fraction of a packet. This model provides a practical and accurate partitioning methodology to develop a FSMC model for wireless channels. Finally, a cross-layer Markov chain model linking the FSMC model with the random access model developed in Chapter 2 was presented. It can be employed in cross layer designs to link higher layers with the properties and error characteristics of the physical channel.

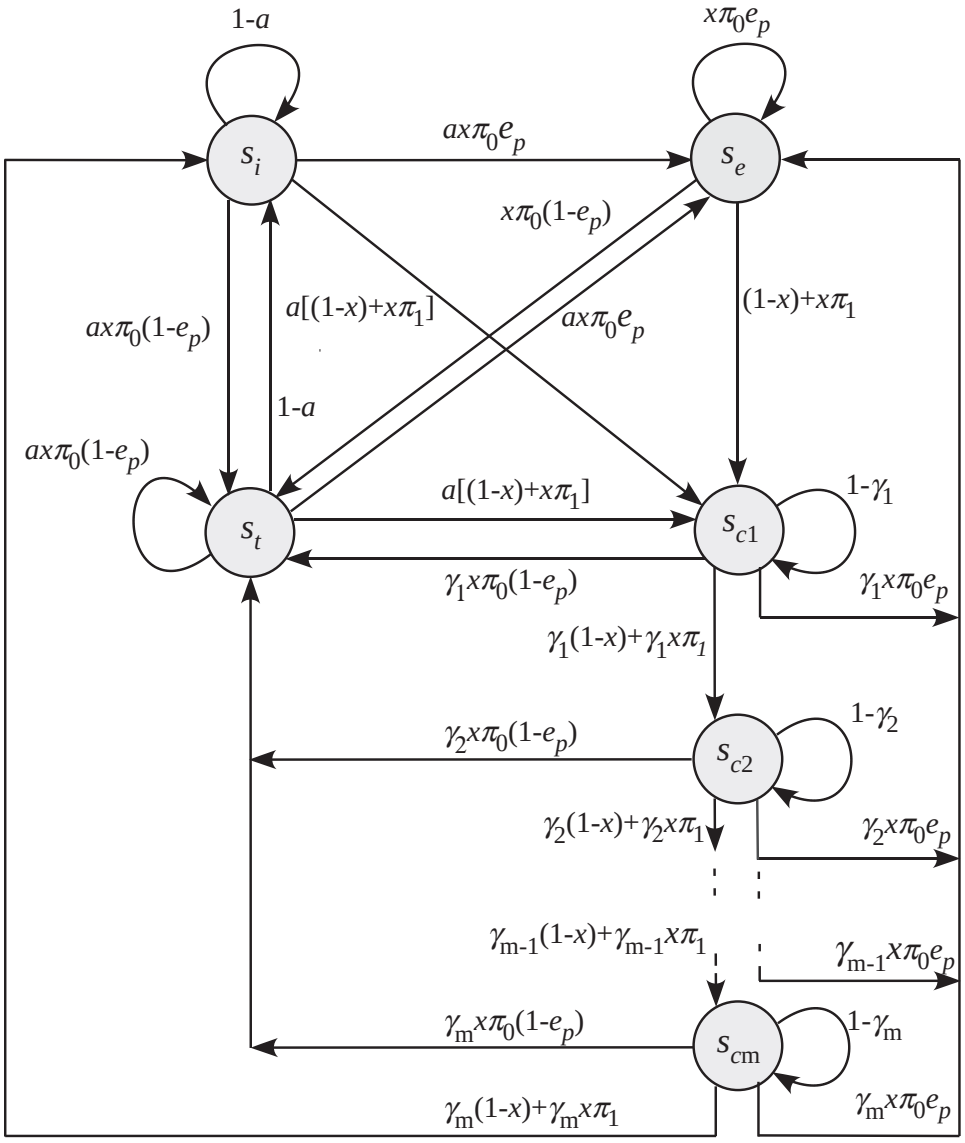


Figure 5.12: A cross-layer Markov chain model employing a two-state ($K = 2$) FSMC.

.

Chapter 6

Conclusions and Future Work

In this chapter a summary of the major research contributions of this dissertation is presented, as well as some suggestions for future work. The models presented in this dissertation are not meant to replace current medium access protocols, but rather to complement existing models. In fact, they represent an enhancement of the random access mechanism that is the backbone of medium access control in a myriad of wireless networks standards. The main goal of this dissertation was to improve this mechanism to achieve better medium access control. The results presented in the preceding chapters can be used to improve services to both wired and wireless users.

6.1 Conclusions

Chapter 2 presented a multistage model for multichannel random access systems. This model focuses on the random access process from the user perspective. The non-saturation condition with multichannel access results in a model which is general and applicable to a wide range of MAC protocols. The model can used to evaluate random access in many applications such as the contention phase in the IEEE 802.16 uplink subframe. A major contribution is the introduction of a variable radix backoff strategy. Performance results show that this strategy can significantly improve contention throughput compared with binary exponential backoff. Extensive simulation results confirmed the accuracy of the analysis.

In Chapter 3, the contention based BW request process in IEEE 802.16 was used to evaluate the variable radix strategy introduced in Chapter 2. A brief explanation of the IEEE 802.16 frame structure was presented. The BW request process was ana-

lyzed using a two-dimensional Markov model incorporating the variable radix backoff. Performance results showed that adjusting the radix with respect to the number of users (network contention level), can significantly improve the performance of the random access channel. Simulation results were provided to confirm the accuracy of the analysis.

In Chapter 4, a new QoS technique using random access was introduced. The novelty of the approach lies in exploiting multichannel contention systems to provide QoS for contending users. This was achieved by dividing the contention channel into distinct service classes. Two classes of users were considered, however the model can easily be applied to any number of classes. The system efficiency, average retransmission delay and request throughput were evaluated to show that this is a reliable QoS mechanism for random access systems. It was also shown that high class users are only marginally affected by the retransmissions generated by low class users due to high collisions on their part. The technique was first analyzed using a simple three-state Markov model and extended to a multistage model to utilize the variable radix backoff strategy introduced in Chapter 4. Results were presented which show that the variable radix strategy can be combined with a QoS mechanism.

In Chapter 5, a means of creating a cross-layer model linking the PHY and MAC collision models was introduced. An optimized finite state Markov channel (FSMC) model for flat fading channels was developed using the equal average fade range duration (AFRD) partitioning methodology. It can be employed in a packet based system to model the underlying physical channel by partitioning the *pdf* of the signal to noise ratio (SNR). A Nakagami-m fading channel model was considered as it is widely employed in the literature. The steady state and transition probabilities and state bit error rate vector were derived. The SNR thresholds to form the Markov model were then obtained based upon an integer approximation of the AFRD. This provides a practical and accurate partitioning methodology to develop a FSMC model for wireless channels. Simulation results were presented which confirm the accuracy of the model for different fade speeds and channel conditions.

Finally a cross-layer Markov model to link the FSMC model with a general random access Markov model was presented. The FSMC model is used to provide a realistic random access process by involving non ideal channel conditions.

In Appendix A, pseudo code for the simulation toolbox using object oriented programming (OOP) using Matlab is presented. The toolbox was developed to provide accurate simulation results for the models presented in this dissertation. It is

a robust simulation platform, and can easily be modified to add additional features and/or functionalities.

6.2 Suggestions for Future Work

Five topics which can be considered for future work are listed below.

- The extension of the cross-layer model introduced in Chapter 5 to more than two FSMC states will provide a cross-layer model for a a wide range of channel conditions. The impact of channel estimation via the FSMC transition portabilities on the random access process for different mobile speeds and modulation techniques can be performed. Knowing the duration of each state, backoff strategies for this duration can also be investigated.

- The work in Chapters 2-4 can be extended to other traffic distributions, e.g. heavy-tailed (such as a Pareto distribution) and self-similar traffic. This will allow traffic analysis for a wide range of practical wireless applications, e.g. third generation wireless systems. The design of a Markov model to accommodate bursty traffic within a frame period is of particular interest. Backoff and contention resolution strategies for this traffic can then be developed.

- The QoS mechanism in IEEE 802.16 has only been introduced through higher layers. In the MAC layer, QoS techniques are restricted to scheduling algorithms for Unsolicited Granted Services (UGS) and Real-time Polling Services. This motivations the extension of the random access QoS technique presented in Chapter 4 to contention-based bandwidth request mechanisms like non-Real-time Polling Services (nrtPS) and Best Effort Services. Minislot assignment in the random access channel can be categorized according to different classes of contending users. Moreover, this technique can be extended to OFDM subchannelization to support service differentiation through the allocation of subcarriers to distinct classes.

- Wireless Local Area Networks (WLANs) with QoS support through the IEEE 802.11(e) standard is gaining popularity. This standard uses the Enhanced Distributed Commutative Function (EDCF) as the MAC protocol. This protocol is similar to the DCF in conventional WLANs, but with QoS support. Services differentiation techniques are provided by variable inter-frame spacings

and contention window limits. The variable backoff window limits can be efficiently managed through the variable radix strategy introduced in Chapter 2. Furthermore, extending the EDCF to support multichannel access can enhance the performance of the IEEE 802.11(e) MAC.

- Considering the capture effect during backoff where a packet can be salvaged from multiple collisions by having a stronger SNR than the interference in the contention channel. The concept is known as Opportunistic ALOHA (O-ALOHA), and although it has recently been investigated, little research has been done from the user perspective. Thus, O-ALOHA can be considered in a cross-layer model which combines signal processing techniques with random access [14].

- With the dramatic increase in date rates, frequency selective channel modeling has gained popularity in the research community. This motivates investigation of FSMC modeling in the frequency domain. The approach can cover a wide range of applications such as Ultra-wide Band (UWB) communications.

Bibliography

[1] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[2] J. Arauz, "Discrete Rayleigh Fading Channel Modeling," *Wireless Communications and Mobile Computing*, vol. 4, pp. 413–425, Mar. 2004.

[3] P. Sadeghi, R. Kennedy, P. Rapajic, and R. Shams, "Finite-State Markov Modeling of Fading Channels - A Survey of Principles and Applications," *IEEE Signal Processing Magazine*, vol. 25, pp. 57 –80, Sep. 2008.

[4] S. Lam and L. Kleinrock, "Packet Switching in a Multiaccess Broadcast Channel: Dynamic Control Procedures," *IEEE Transactions on Communications*, vol. 23, pp. 410 – 423, Sep. 1975.

[5] S. Lam and L. Kleinrock, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Transactions on Communications*, vol. 23, pp. 891 – 904, Sep. 1975.

[6] S. Tasaka, *Performance Analysis of Multiple Access Protocols*. The MIT Press, 2008.

[7] G. Bianchi, "IEEE 802.11-Saturation Throughput Analysis," *IEEE Communications Letters*, vol. 2, pp. 318–320, Dec. 1998.

[8] G. Bianchi, L. Fratta, and M. Oliveri, "Performance Evaluation and Enhancement of the CSMA/CA MAC Protocol for 802.11 Wireless LANs," in *Seventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,*, vol. 2, pp. 392–396, Oct. 1996.

[9] M. Alam and A. Hossain, "Throughput Analysis of a Multichannel Slotted-Aloha Protocol in Short-Haul Communication Environment for an Exponential Backoff

Retransmission Scheme," in *Proceedings of the International Conference on Information, Communications and Signal Processing,*, vol. 2, pp. 1034 –1038, Sep. 1997.

[10] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Standard 802.16," 2004.

[11] C. Eklund, R. Marks, K. Stanwood, and S. Wang, "IEEE Standard 802.16: A Technical Overview of the Wirelessman Air Interface for Broadband Wireless Access," *IEEE Communications Magazine*, vol. 40, pp. 98–107, Jun. 2002.

[12] Y. Zhang, ed., *WiMAX Network Planning and Optimization*, ch. 4, pp. 73–91. Aurbach Publications, Taylor and Francis Group, 2009.

[13] L. Nuaymi, *WiMAX Technology for Broadband Wireless Access*. John Wiley & Sons Ltd, 2007.

[14] L. Tong, V. Naware, and P. Venkitasubramaniam, "Signal Processing in Random Access: A Cross Layer Perspective," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 29–39, 2004.

[15] Y. Abdel-Hamid, F. Gebali, and T. A. Gulliver, "A Reservation-Based Multiple Access Protocol for OFDMA Networks with Adaptive Backoff Strategy," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing,*, pp. 161–165, Aug. 2007.

[16] Y. Abdel-Hamid, F. Gebali, and T. Gulliver, "A Multi-channel QoS Model for Random Access Systems," in *IEEE Communication Networks and Services Research Conference,*, pp. 20 –24, May 2009.

[17] Y. Abdel-Hamid, F. Gebali, and T. Gulliver, "A QoS Model for Random Access Systems with Multiple Request Classes," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing,*, pp. 331 –335, Aug. 2009.

[18] "Wireless LAN Medium Access Control (MAC) and Physical Payer (PHY) Specification, IEEE Standard 802.11," 1997.

[19] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, Mar. 2000.

[20] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential," *IEEE Communications Magazine*, vol. 43, pp. 129 –136, Feb. 2005.

[21] F. Gebali, *Analysis of Computer Communication Networks.* Springer, 2008.

[22] Y. Fallah, F. Agharebparast, M. Minhas, H. Alnuweiri, and V. Leung, "Analytical Modeling of Contention-Based Bandwidth Request Mechanism in IEEE 802.16 Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 3094–3107, Sep. 2008.

[23] D. Bertsekas and R. Gallager, *Data Networks, 2nd ed.* Prentice Hall Inc., 1992.

[24] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, Mar. 2000.

[25] A. Vinel, Y. Zhang, M. Lott, and A. Tiurlikov, "Performance Analysis of the Random Access in IEEE 802.16," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 1596–1600, Sep. 2005.

[26] Q. Ni, A. Vinel, Y. Xiao, A. Turlikov, and T. Jiang, "Wireless Broadband Access: WiMAX and Beyond - Investigation of Bandwidth Request Mechanisms under Point-to-Multipoint Mode of Wimax Networks," *Communications Magazine, IEEE*, vol. 45, May 2007.

[27] A. Vinel, Y. Zhang, Q. Ni, and A. Lyakhov, "Efficient Request Mechanism Usage in IEEE 802.16," in *Global Telecommunications Conference*, pp. 1–5, Dec. 2006.

[28] H. Fattah and H. Alnuweiri, "Performance Evaluation of Contention-Based Access in IEEE 802.16 Networks with Subchannelization," in *IEEE International Conference on Communications*, vol. 4, pp. 1–6, Jun. 2009.

[29] R. Iyengar, V. Sharma, K. Kar, and B. Sikdar, "Analysis of Multi-channel Wireless MAC for Point-to-Multipoint Networks," tech. rep., RPI Networks, Apr. 2006.

[30] D. Staehle and R. Pries, "Comparative Study of the IEEE 802.16 Random Access Mechanisms," in *The International Conference on Next Generation Mobile Applications, Services and Technologies*, pp. 334–339, Sep. 2007.

[31] S.-M. Oh and J.-H. Kim, "The Analysis of the Optimal Contention Period for Broadband Wireless Access Network," in *Third IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 215–219, Mar. 2005.

[32] M. Hawa and D. Petr, "Quality of Service Scheduling in Cable and Broadband Wireless Access Systems," in *Tenth IEEE International Workshop on Quality of Service, 2002*, pp. 247 – 255, Aug 2002.

[33] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of Service Support in IEEE 802.16 Networks," *IEEE Network*, vol. 20, pp. 50–55, Mar 2006.

[34] Andrew Molisch, *Wireless Communications, 2nd ed.* Wiley-IEEE Press, 2011.

[35] E. Biglieri, G. Caire, and G. Taricco, "Coding for the Fading Channel: a Survey," *Signal Processing, Elsevier North-Holland, Inc.*, vol. 80, pp. 1135–1148, Jul. 2000.

[36] A. Goldsmith and S.-G. Chua, "Variable-Rate variable-power MQAM for Fading Channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218 –1230, Oct 1997.

[37] W. C. Jakes, *Microwave Mobile Communications, 2nd ed.* Wiley-IEEE Press, New York, 1992.

[38] J. G. Proakis, *Digital Communications, 4th ed.* McGraw Hill higher Education, Mc.Graw Hill Inc. New York, 2001.

[39] G. L. Stuber, *Principles of Mobile Communications, 2nd ed.* Kluwer Academic Publishers, 2002.

[40] M. Nakagami, "The m-dsitribution - a General Formula of Intensity Distribution of Rapid Fading," in *Pergamon Press Symposium on Statistical Methods of Radio Wave Propagation*, pp. 3–36, Jun. 1960.

[41] M. D. Yacoub, J. E. V. Bautistu, and L. Guerra de Rezende Guedes, "On Higher Order Statistics of the Nakagami-m Distribution," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 790–794, May 1999.

[42] W. H. T. K. S. S. T. S. Rappaport and K. L. Kosbar, *Principles of Communication Systems Simulation with Wireless Applications*. Prentice Hall; 1st ed., 2004.

[43] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J*, vol. 39, pp. 1253–1256, Sep. 1960.

[44] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J*, vol. 42, pp. 1977–1997, Sep. 1963.

[45] A. J. Goldsmith and P. P. Varaiya, "Capacity, Mutual Information, and Coding for Finite-State Markov Channels," *IEEE Transactions on Information Theory*, vol. 42, pp. 868–886, May 1996.

[46] H. S. Wang and P.-C. Chang, "On Verifying the First-order Markovian Assumption for a Rayleigh Fading Channel Model," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 353–357, May 1996.

[47] M. Zorzi, R. Rao, and L. Milstein, "On the Accuracy of a First-Order Markov Model for Data Transmission on Fading Channels," in *Fourth IEEE International Conference on Universal Personal Communications*, vol. 10, pp. 211 –215, Nov 1995.

[48] C. C. Tan and N. C. Beaulieu, "On first-order Markov Modeling for the Rayleigh Fading Channel," *IEEE Transactions on Communications*, vol. 48, pp. 2032–2040, Dec. 2000.

[49] H. S. Wang and N. Moayeri, "Modeling, Capacity, and Joint Source/Channel Coding for Rayleigh Fading Channels," in *IEEE 43rd Conference on Vehicular Technology Conference*, pp. 473–479, May 1993.

[50] H. S. Wang and N. Moayeri, "Finite-State Markov Channel. A Useful Model for Radio Communication Channels," *IEEE Trans. Veh. Tech.*, vol. 44, pp. 163–171, Feb. 1995.

[51] Q. Zhang and S. A. Kassam, "Finite-State Markov Model for Rayleigh Fading Channels," *IEEE Transactions on Communications.*, vol. 47, pp. 1688–1691, Nov. 1999.

[52] F. Babich, O. E. Kelly, and G. Lombardi, "Generalized Markov Modeling for Flat Fading," *IEEE Transactions on Communications*, vol. 48, pp. 547–551, Apr. 2000.

[53] F. Babich and G. Lombardi, "A Markov Model for the Mobile Propagation Channel," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 63–73, Jan. 2000.

[54] J. Arauz and P. Krishnamurthy, "A Study of Different Partitioning Schemes in First Order Markovian Models for Rayleigh Fading Channels," *The 5th International Symposium on Wireless Personal Multimedia Communications*, vol. 1, pp. 277–281, Oct. 2002.

[55] C. D. Iskander and P. T. Mathiopoulos, "Fast simulation of Diversity Nakagami Fading Channels Using Finite-state Markov Models," *IEEE Transactions on Broadcasting*, vol. 49, pp. 269–277, Sept. 2003.

[56] R. Zhang and L. Cai, "Packet-Level Channel Model for Wireless OFDM Systems," in *IEEE Global Telecommunications Conference*, pp. 5215–5219, Nov. 2007.

[57] M. Chu, D. Goeckel, and W. Stark, "On the Design of Markov Models for Fading Fhannels," in *IEEE VTS 50th Vehicular Technology Conference*, vol. 4, pp. 2372 –2376, Mar. 1999.

[58] H. Liu and M. El Zarki, "Performance of H.263 Video Transmission over Wireless Channels Using Hybrid ARQ," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1775 –1786, Dec. 1997.

[59] J. M. Park and G. Hwang, "Mathematical Modeling of Rayleigh Fading Channels Based on Finite State Markov Chains," *IEEE Communications Letters*, vol. 13, pp. 764 –766, Oct. 2009.

[60] Y. L. Guan and L. F. Turner, "Generalised FSMC Model for Radio Channels with Correlated Fading," in *IEE Communications*, vol. 146, pp. 133–137, Apr. 1999.

[61] W. Turin and R. Van Nobelen, "Hidden Markov Modeling of Fading Channels," in *The 48th IEEE Vehicular Technology Conference*, vol. 2, pp. 1234–1238, May 1998.

[62] A. Antoniou and W.-S. Lu, *Practical Optimization, Algorithms and Engineering Applications*. Springer, 2007.

[63] "Introduction to Object Oriented Programming Concepts (OOP) and More." http://www.codeproject.com/Articles/22769/Introduction-to-Object-Oriented-Programming-Concep. Accessed: 30/09/2011.

Appendix A

Model Simulation with Object Oriented Programming

To confirm the validity of the models in this dissertation, a simulation tool using Object Oriented Programming (OOP) was developed. In contrast with conventional programming, OOP provides a clear modularization technique to define the model entities such as the contention interval slot (channel) and the contending SS (node). These entities are defined as classes where each class comprises several methods (objects). This provides the advantage of programming transparency where the details of each class are hidden while maintaining a clearly defined interface. This abstraction results in an easy structural approach to maintain and modify the existing code by simply adding new methods to the class such as channel noise or error control with little or no changes to the existing classes. This significantly reduces complexity and improves modularity [63].

The OOP simulation tool defines the models presented in this chapter using two entities (classes). The **channel** class defines each contention slot in the contention interval and the **node** class defines each SS. In Algorithm A.1, some of the channel properties are shown with their associated methods. It is clear that more properties (states) such as channel fading with associated mitigation methods, e.g. forward error control, can easily be added to the class. Algorithm A.2 defines the node class with some of the associated properties and methods such as the transmit count and the delay metric. Other properties that can be added to the node class include buffer occupancy which can be associated with a corresponding method such as a scheduling algorithm. A descriptions of the methods associated with the channel and node classes

are shown in Tables A.1 and A.2, respectively. In Table A.3 a brief description of the calling functions is given.

Algorithm A.1 The channel class and associated methods

classdef **ChannelClass** → handle //construct the channel class

properties
id; //channel ID
state; // channel state, 0: idle, 1: one SS access, 2: two SS access ...
idleCount; // idle count
xmtCount; // transmitting
collideCount; // collision count
⋮
end //properties

methods
function methodInitializeChannel (channel, k)
function methodIncrementIdle (channel)
function methodIncrementCollide (channel)
function methodIncrementXmt (channel)
end // methods

end // classdef

Algorithm A.2 The node class and associated methods

classdef **NodeClass** → handle //construct the channel class

properties
id; // node ID
state; // node state, 0: idle, 1: 1^{st} attempt, 2: 2^{nd} attempt ...
bufferSize; // node input buffer size (= 1 in this model)
packetsInBuffer; // count packets in the input buffer
xmtCount; // count successful node transmissions
collideCount; // count node collisions
totalDelay; // count the total delay
⋮
end //properties

methods
function methodInitializeNode (node, i, b)
function methodfAddPacket (node, frame)
function methodPacketSelect (node, K)
function methodPacketXmt (node, frame)
function methodPacketCollide(node, m, K)
end
end // methods

end // classdef

channel (slot) method	Description
methodInitializeChannel	Initializes the channel properties at the beginning of each frame. Assigns the index k to each channel, where $k \in K$.
methodIncrementIdle	Increments the idle channels property. Sets channel.state $\rightarrow 0$ indicating no access in the current frame.
methodIncrementXmt	Increments the transmitting channels property. Sets channel.state $\rightarrow 1$ indicating only one access in the current frame.
methodIncrementCollide	Increments collided channels property. Sets channel.state ≥ 2 indicating more than one access in the current frame.

Table A.1: Methods associated with the channel class

node (SS) method	Description
methodInitializeNode	Initializes node properties at the beginning of each frame. Sets node.state $\rightarrow 0$ (idle); sets node ID to an index i, where $i \in N$. Sets maximum buffer size b (in this model, $b = 1$).
methodAddPacket	Checks the occupancy status of the node input buffer. Adds/discards requests according to the input buffer size b. Stamps the buffered request position with the current frame index to record the access time for delay computation.
methodPacketSelect	Checks the existence of a request in the head of the line (HOL) of the input buffer. Inserts the request into the xmt buffer. Randomly accesses a channel from the channel set K and stores its index k, where $k \in$ rand(K, 1). Assigns the chosen channel index k to the request in the xmt buffer. Sets node.state $\rightarrow 1$ indicating a first transmission attempt. Decrements and updates the request positions in the input buffer. Sets the xmt flag.
methodPacketXmt	Increments transmitting channel property. Increments xmt property. Computes delay (delay = frame index $-$ input buffer position index). Resets node properties for the next frame.
methodPacketCollide	If state $> m$, the limit is reached, so discards the request and reset all properties. If state $\leq m$, increments node state, Thus, node.state ≥ 2, i.e., 2^{nd} attempt, 3^{rd} attempt, ... Increments collided node property. Re-randomizes the channel set K and assigns a new set of channel indices, $k \in$ rand(K, 1). Resets the xmt flag.

Table A.2: Methods associated with the node class

Function	**Description**
functionMain()	Defines and initializes the **channel** and **node** classes. Collects and averages the statistics at the end of each frame.
functionGenerateTraffic()	Generates the request traffic according to the specified distribution (this model uses Bernoulli traffic). Outputs a traffic matrix where each element represents a TRUE/FALSE request arrival event (row) per frame (column).
functionAddPacket()	Checks the entries of the traffic matrix of every node and calls methodAddPacket.
functionPacketSelect()	Checks the HOL of every node and calls methodPacketSelect.
functionCheckActivity()	Checks the activity of every node with respect to each channel. This is done by inspecting the state of every node and confirming that this particular node satisfies the conditions to transmit in the current frame. This concludes with either a fresh (first) transmission attempt (node.state == 1) or a collided node retransmitting in the current frame according to γ_i, i.e., node.state ≥ 2 AND rand $\leq \gamma_i$, where $i = 1, 2, \ldots m$. Sets an active channel flag and a node xmt flag accordingly.
functionUpdateChannel()	Updates channel.state (active channel flag) and calls the appropriate channel method (methodIncrementIdle, methodIncrementXmt or methodIncrementCollide).
functionUpdateNode()	Updates the node.state (xmt buffer status) for every channel and calls the appropriate node method (methodPacketXmt, methodPacketCollide).
functionTraceChannel()	Collects channel statistics.
functionTraceNode ()	Collects node statistics.
functionIterateFrames()	Averages statistics over the specified number of frames.

Table A.3: Calling functions and the corresponding description