FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Departamento de Engenharia Electrotécnica e de Computadores

MAC Regenerative Analysis Of Wireless Ad-Hoc Networks

Por
Luís Filipe Dias de Sousa

Dissertação apresentada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para obtenção do grau de Mestre em Engenharia Electrotécnica e de Computadores.

Orientador: Professor Rodolfo Oliveira

Lisboa
2010

Preface

I would like to express my gratitude to my advisors Professor Rodolfo Oliveira, Professor Yevgeni Koucheryavy and Researcher Sergey Andreev. It was a pleasure to work on such an interesting topic and to learn new things from them all. I really appreciate Professor Rodolfo availability and patience to answer all my questions. It was very kind of him to accept being my supervisor under my exchange studies period. I feel very grateful for the opportunity that Professor Yevgeni Koucheryavy gave me to work in the communications research group and for supporting all technical material needed. I really appreciate the patience that Researcher Sergey Andreev had with me in all this learning process and for all the hours that he spent with me explaining the contents needed to this thesis. I would also like to thank Professor Jarmo Harju for all his support.
I dedicate this thesis to my parents and my sister for their support and encouragement all over this process of making this work abroad. A special thanks to all my Portuguese and international friends that always support me.

Resumo

O IEEE 802.11 é uma tecnologia em expansão por todo o mundo, sendo hoje em dia utilizada por centenas de milhões de utilizadores. Apesar do elevado número de utilizadores, nem todos precisam da mesma Qualidade de Serviço (QoS). Deste modo, a diferenciação de serviço é um factor importante e, por essa razão, deve ser considerada em modelos matemáticos que modelam o desempenho da rede. Além disso, os utilizadores comunicam tipicamente através de conexões ponto-a-ponto (esquema de transmissão unicast) e conexões ponto-multiponto (esquema de transmissão broadcast). A co-existência de tráfego unicast e broadcast influencia o desempenho da rede e a sua importância não deve ser negligenciada. Estes factos motivam o trabalho apresentado nesta dissertação, o qual afere a sua importância no desempenho da rede.

Esta tese descreve um modelo que modela o comportamento do MAC (Medium Access Control) usado nas redes baseadas na tecnologia IEEE 802.11. Este é o primeiro passo para desenvolver um modelo que considera grupos de utilizadores que usam diferentes parâmetros MAC e a coexistência de dois diferentes esquemas de transmissão (unicast e broadcast).

Por fim, o modelo apresentado é validado através de simulações, caracterizando-se a sua precisão e algumas propriedades das redes são discutidas.

Abstract

The IEEE 802.11 is a fast growing technology all over the world. This growth is essentially due to the increasing number of users in the network. Despite the increasing number of users, not all of them need the same quality of service. Thus, service differentiation is an important aspect that shall be considered in mathematical models that describe network performance. Moreover, users typically communicate using point-to-point connections (unicast transmission scheme) and point-to-multipoint connections (broadcast transmission scheme). The co-existence of unicast and broadcast traffic impacts the network performance and its importance cannot be neglected in the network performance evaluation. This motivates the work presented in this thesis, which characterizes the network accounting for these important parameters.

This thesis formulates a model to describe the behavior of the medium access control used in IEEE 802.11-based networks. This is the first step to develop a model that considers both different groups of users configured with different medium access control parameters and the co-existence of two different transmission schemes (unicast and broadcast). The model also assumes a finite number of retransmissions for unicast packets and it is confirmed that several models already proposed in other works are especial cases of the proposed model.

Finally, a theoretical validation of the model is done as well as some simulations to assess its accuracy and, some realistic network features are discussed.

List of Abbreviations

ACK	–	Acknowledgement;
AIFS	–	Arbitration Interframe Space;
BA	–	Block Acknowledgment;
BEB	–	Binary Exponential Backoff;
BC	–	Binary Exponential Backoff Counter;
CSMA/CA	–	Carrier Sense Multiple Access/Collision Avoidance;
CS	–	Carrier Sense;
CTS	–	Clear to Send;
CW	–	Contention Window;
DCF	–	Distribution Coordination Function;
DIFS	–	DCF Interframe Space;
ECDA	–	Enhanced Distributed Channel Access;
EIFS	–	Extended Interframe Space;
FDMA	–	Frequency Division Multiple Access;
FCS	–	Frame Check Sequence;
GA	–	Geometric ALOHA;
GA-IT	–	Geometric ALOHA with Immediate Transmission;
GA-NIT	–	Geometric ALOHA with Non-Immediate Transmission;
IFS	–	Interframe Space;
MAC	–	Medium Access Control;
MANET	–	Mobile Ad Hoc Network;
NAV	–	Network Allocation Vector;
ns-2	–	Network Simulator - 2;
PHY	–	Physical Layer;
QoS	–	Quality of Service;
RCP	–	Retransmission Control Procedure;
RMA	–	Random Multiple Access;
RTS	–	Request to Send;
SIFS	–	Short Interframe Space;
TXOP	–	Transmission Opportunity;
TDMA	–	Time Division Multiple Access;
UA	–	Uniform ALOHA;
WLAN	–	Wireless Local Area Network;

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The *Wireless Local Area Networks* (WLANs) based on IEEE 802.11 standard [1] are a fast growing technology all over the world. This kind of networks is easy to deploy and it provides an effective low cost way to achieve wireless data connectivity between devices. There are special cases of WLANs such as *Ad Hoc* networks and *Mobile Ad Hoc Networks* (MANETs). An Ad Hoc network is a self-configuring network of devices connected by a wireless link. What differentiates MANETs from Ad Hoc networks is the mobility of devices in the network. These emerging and fast propagation technologies bring the need to evaluate the network performance in order to support more users and to exploit limited wireless bandwidth resources more efficiently. The lack of infrastructure of these wireless technologies makes the *Medium Access Control* (MAC) more complex.

Typically, Ad Hoc networks rely on two kinds of traffic: Unicast and Broadcast. Unicast traffic is used in point-to-point connections whereas that broadcast traffic is used in point-to-multipoint connections. As these two types of traffic serve different purposes and generally coexist within a network, they shall not be neglected.

The number of users in WLANs is increasing, however not all of them have the same necessities in terms of channel resources consumption. For this reason, a traffic differentiation is needed in nowadays networks. This introduces the notion of priority in the access to the radio channel. Thus, a group of users can have the possibility to access the channel with higher probability than others. For instance, real-time traffic needs a higher priority to access the channel than other types of traffic.

Due to the traffic differentiation requirement, *Enhanced Distributed Channel Access* (EDCA) [2] extended the *Distributed Coordination Function* (DCF) of legacy IEEE 802.11 [1] in order to provide *Quality of Service* (QoS) for different types of users. This extended version of the protocol provides QoS levels for different types of traffic and it allows for the optimization of network resources.

The focus of this work is to propose a model that studies the IEEE 802.11 MAC performance (under the conditions presented in section 2.3). The model shall account for the mixture of traffic, groups of heterogeneous users with different QoS requirements and a finite number of (re)transmissions. Consequently, a simplified EDCA model is also studied. The simplified EDCA model cannot fully capture the complex behavior of EDCA, however it gives a lower bound on the network performance that takes into account some network parameters that arbitrate users' channel access priority.

1.1 Problem Statement and Motivation

In [3] it is shown that broadcast traffic is a critical parameter that strongly affects the network throughput. According to Oliveira, this effect is essentially due to two main reasons. Firstly, broadcast and unicast traffic rely on different transmission schemes. What really differentiates these schemes is the unicast dependency on receiver's acknowledgement. This acknowledgement-based transmission scheme allows the data frame to be retransmitted when an attempt fails (due to noisy channel, interference between simultaneous transmissions, etc). On the contrary, broadcast transmission scheme is unreliable. The sender has no way to know whether the data frame was received by the recipients or not. Figure 1.1[1] depicts the probability to successfully transmit a frame when different amounts of broadcast traffic are used in the network. The consequences of this traffic coexistence is that with the increasing broadcast traffic generation, the aggregated network data frame success probability of transmission drops.

Secondly, broadcast transmission frames are typically sent with a lower transmission rate. This difference of transmission rates influences the channel throughput because the broadcast data frames need a larger portion of time to be sent. When more than one user trans-

[1]Like it is explained in [3], p_b stands for the probability of broadcast traffic generated in the network.

mits at the same time, the channel utilization is limited by the frame that is transmitted at a slower transmission rate.

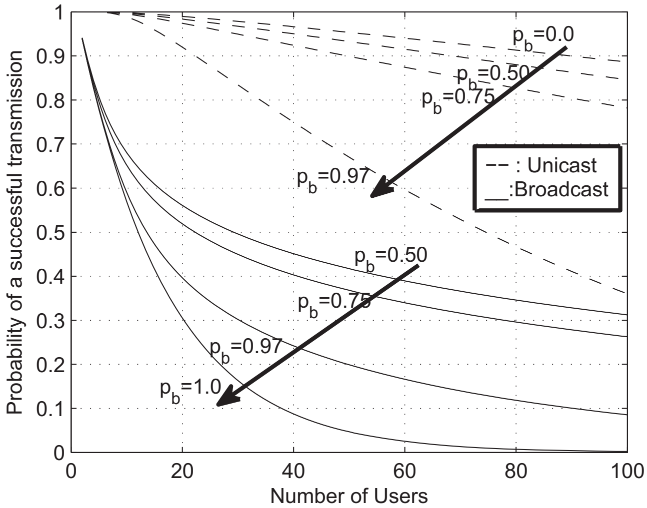


Figure 1.1: Consequences in success probability for IEEE 802.11b considering the co-existence of traffic in the network.

In the literature, the influence of broadcast traffic on different groups of users (heterogeneity) was never studied. There is a need to come up with a model capable of evaluating the MAC throughput performance of the traffic co-existence with groups of heterogeneous users.

The amount of broadcast traffic in the network is a critical parameter that can degrade the channel throughput and it shall be studied in more detail. This study is even more important when the radio channel bandwidth is limited and the number of users is growing. Thus a novel model[2] capable of describing the system is essential, since it will be helpful to enhance the network throughput. Another relevant issue is the need of traffic differentiation, because the users do not necessarily have the same needs in terms of channel utilization. Hence, new studies shall account for QoS and user differentiation.

There are many models for evaluating the MAC behavior but most of the works done so far are done based on the network timings such as the idle time duration and the time duration of a packet transmission (for successful and unsuccessful transmissions). These timings increase the complexity of the models. Therefore, the focus of this work is to first

[2]In particular, this research work focus is the saturation conditions. A more detailed description of saturated and unsaturated traffic is given in the next chapters.

develop a model based on the network probabilities and then extend the model to account for the throughput.

1.2 Objectives

The main goal of this thesis is to provide a simple model to analyze the network performance under saturated traffic conditions for the recent IEEE 802.11 standard. The simple model shall also support heterogeneity between groups of users and consequently present a simplified approach for EDCA evaluation.
The simple model is a tool that can be used whether to describe the network behavior or for study possible ways to enhance the network throughput. The proposed model is meant to be easy to compute due to its focus on channel probabilities. Moreover, channel timings are taken into account for simpler cases. Finally, the simple model is tested against simulations in order to study its accuracy.

1.3 Thesis Outline

The thesis is structured in seven chapters, including the introduction and the conclusions chapter.
Second chapter introduces a brief state-of-the-art description of what was done in this area and what is going to be done in this thesis. It describes the most important aspects of IEEE 802.11 and the main related research works. This chapter also defines the main set of assumptions that are adopted in this work. The foundations of the model explored throughout this work are also described in this chapter.
The ALOHA protocol is analyzed in the third chapter. This is a very important chapter since it explains how ALOHA protocol and *Binary Exponential Backoff* (BEB) mechanism are related.
Chapter four describes the BEB mechanism. It formally introduces the *Regeneration Cycle Concept*. This is the key concept used in this work and it is applied not only to the proposed model, but also to some older works in literature. It is important to explain that this chapter deals with an equally-slotted system, that is the reason why the chapter title

reads *"Simple Analysis"*. Here, the first part of the proposed model is also deduced.
The fifth chapter is concerned with the *"Advanced Analysis"* of the BEB model. The word "Advanced" in this case means that the chapter deals with unequally-slotted system and heterogeneous groups of users. Moreover, a simplified EDCA model is also proposed. The chapter ends with a list of hypothetical practical scenarios where the proposed model can be used.
The model validation is done in chapter sixth through ns-2 (*Network Simulator - 2*). The main reason for the proposed model mismatch are discussed.
Finally in the last chapter, the conclusions of this thesis are presented. This chapter also describes future research directions in which it is interesting to continue.

Chapter 2

Related Work

This chapter provides the necessary information that the reader needs to understand the next chapters. Firstly, the IEEE 802.11 standard aspects are introduced. Secondly, the background section presents related research work. Next the general set of model assumptions is described. This set of assumptions is needed because it represents a unified view of the system. For each model studied through this thesis the relevant changes to this set of assumptions are presented in a list to make the model scope more clear. This has an extreme importance since the models are only valid under some certain conditions. After the basic set of assumptions, the protocols classification is done. This is important because this thesis deals with several different protocols, helping to clarify the type of *Random Multiple Access* (RMA) being described. A RMA protocol is a set of rules that allows users to access the channel without a centralized point of coordination. Finally the proposed model of this thesis is introduced and its characteristics are discussed in detail.

2.1 IEEE 802.11 MAC-Layer Protocols

The most important MAC layer aspects of IEEE 802.11 standard are described in this subsection. For a more detailed information the reader shall refer to [1] and [2]. The scope of this section is the overview of DCF and EDCA mechanisms that allow users to access the channel.

2.1.1 Distribution Coordination Function

In 1999, the IEEE 802.11 standard [1] was proposed to provide the features for the Physical (PHY) layer and Media Access Control (MAC) layer. It describes the information that these two layers provide to each other and its tasks. In what follows, little attention is given to PHY since it is not the scope of this research work. The DCF is a medium access protocol that allows a group of users to share the same wireless channel through the use of a mechanism called *Carrier Sense Multiple Access/Collision Avoidance* (CSMA/CA) and a Binary Exponential Backoff (BEB) mechanism.
The *Carrier Sense* (CS) of DCF is done though physical and virtual means. The virtual CS is typically implemented with the exchange of two special reservation frames, *Request to Send* (RTS) and *Clear to Send* (CTS), before the exchange of the data frames. These two control frames have a duration field in the frame header that contains the period of time that the medium will be busy transmitting the data frame and for the recipient user to return the *Acknowledgement* (ACK) frame. The duration field is useful to prevent the hidden users from accessing the channel while a transmission is in process. A hidden user is a user that cannot sense the sender but can influence the receiver if it starts to transmit a data frame. In a wireless environment a user can receive a frame that was not addressed to it. As such, this node shall use the duration field information contained in the header of the received frame to update the *Network Allocation Vector* (NAV). This timer is just a prediction of how long the medium is expected to be busy, so the user can defer its transmission until this timer expires. The physical CS is not a MAC function because it is the PHY that deals with it through the analysis of the channel *Signal to Interference-plus-Noise Ratio* (SINR). Finally, the user shall never transmit when weather the NAV timer is active or the PHY senses the medium busy.

The time interval between frames is called the *Interframe Space* (IFS). Different IFS values allow for different kinds of packets to access the medium in a different way. For instance, an ACK frame or a CTS frame has always priority when compared with a regular data frame. Figure 2.1 shows the difference between these xIFS times.

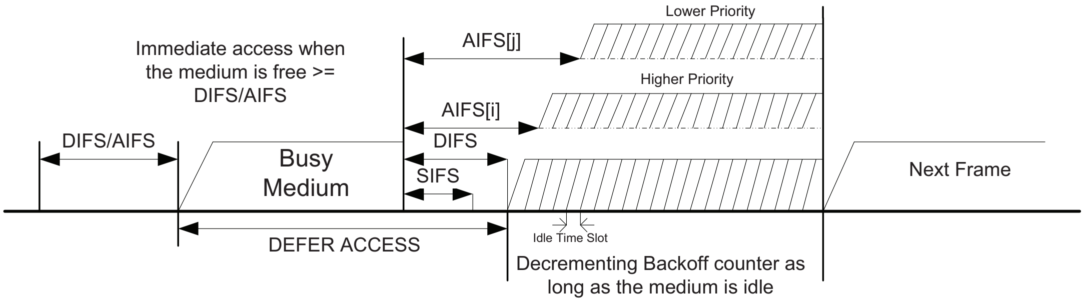


Figure 2.1: Time intervals relationships of DCF/EDCA.

The different xIFS description can be found below:

- *Short Interframe Space* (SIFS): This is the shortest IFS time and it is essentially used between the following frames: RTS-CTS, CTS-DATA and DATA-ACK. Using SIFS between the frame exchanges prevents other users from attempting to use the medium, which is required to wait for the medium to be idle for a longer time.

- *DCF Interframe Space* (DIFS): Each user having a data frame to transmit shall defer at least a DIFS time before starting transmission.

- *Extended Interframe Space* (EIFS): This is the longest IFS time and it is used when a user has received a frame that contains errors. It is possible to detect errors in a frame each time that the MAC *Frame Check Sequence* (FCS) is not correct. This IFS time is useful since the user could not understand the last received frame, and consequently could not update the NAV timer. This avoids the user from colliding with a future frame belonging to the current dialog. This means that all users that received the corrupted frame wait for an EIFS while the sender waits for the expiration of the ACK time-out.

The *Arbitration Interframe Space* (AIFS) time duration will be explained in the next subsection since only EDCA uses it.

The BEB (mentioned in Figure 2.1) has a counter, often called *Backoff Counter* (BC), and it is initiated with a random integer between zero and an initial *Contention Window* (CW) minus one. For each idle time slot, BC is decremented by one unit. When the BC reaches zero the user is allowed to start the transmission. Notice that the BC is only decremented when there is an idle slot. This means that the CS mechanism has to indicate that the medium is idle. When a collision occurs the BEB doubles its previously used CW and all the process is repeated again. The CW is doubled as soon as the maximum CW is not reached. A formal description of BEB mechanism is discussed and its mathematical formulation (including the equation for the growth of CW) is done in chapter 4.

The DCF has two different acknowledgement based transmission schemes for unicast and one acknowledgement-free based transmission scheme for broadcast. All these schemes assume immediate transmission after the DIFS or EIFS, conditioning on if a collision in a channel was sensed respectively. The BEB shall be started if the user has a packet to send but the CS indicates that the medium is busy or after a non-successful packet transmission.

The standard also describes the post-BEB in which the BEB shall be started after each transmission, no matter whether or not the user has packets to send. The post-BEB was imposed in IEEE 802.11 implementations to avoid a user to capture the channel. A user shall send a packet immediately after a DIFS, however, after a successful transmission, if there are packets in the queue the BEB is started. Hence, by controlling the timing with which a packet is inserted in the queue it is possible to enable immediate transmission to all packets.

A recipient does not receive a packet correctly if it was corrupted by channel noise, channel interference or due to multiple channel transmissions in the same time slot. Thus, a user is ready to send a packet if the medium is idle at least for a period of time equal to DIFS, the NAV timer shows zero, the PHY indicates that the medium is idle and the BC reached zero.

As mentioned above, there are two mechanisms to send a unicast data frame: the Basic Access and the RTS/CTS mechanism. The basic access mechanism and the RTS/CTS are depicted in Figures 2.2 and 2.3, respectively. The basic access does not use the RTS /CTS

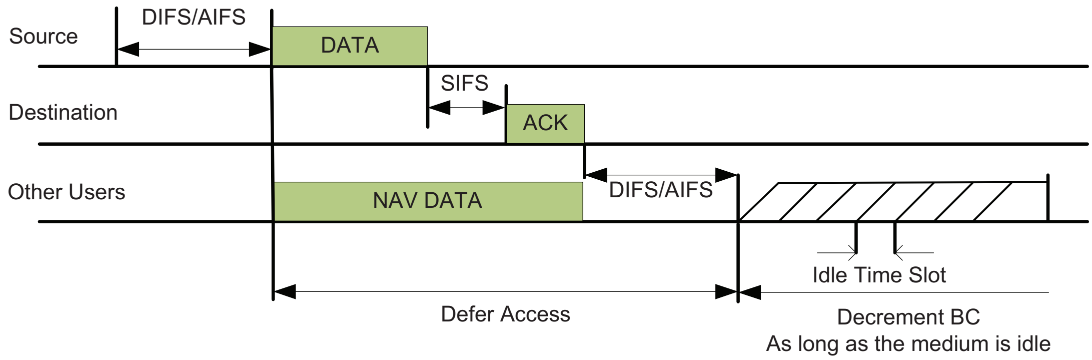


Figure 2.2: Unicast basic access scheme mechanism.

reservation frames. This transmission scheme is used whenever the data frame length is equal or below a given threshold[1]. After sensing the medium idle for a period of time equal to DIFS the user sends the data frame. If the recipient acquires the packet correctly, it prepares the ACK frame. This is a 2-way handshake mechanism. The remaining users in the network sense that the medium is busy due to PHY CS or due to NAV timer. Figure 2.2 also shows why the basic access is not always a good option for multi-hop networks. The reason is that typically only the sender's neighbors update the NAV timer. Hence, if there are hidden users ready to transmit, the data frames are more likely to be corrupted because of the simultaneous transmissions.

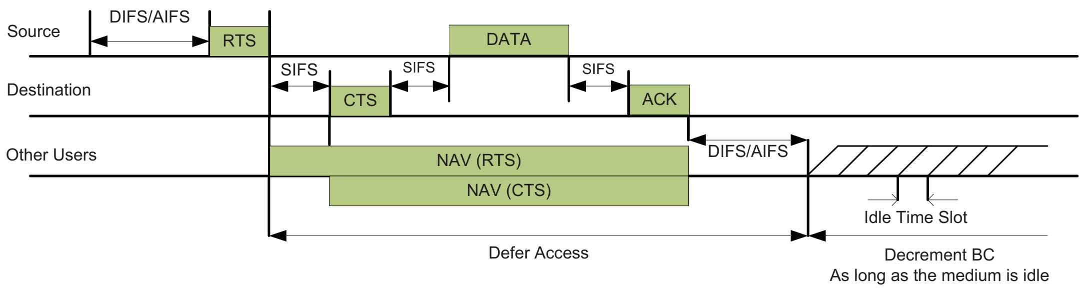


Figure 2.3: Unicast RTS/CTS access scheme mechanism and NAV setting.

Figure 2.3 shows the RTS/CTS mechanism. After the medium is sensed to be idle for a DIFS, a reservation frame RTS is sent. This frame contains in the information about the total time needed to finalize the transmission. When the recipient acquires the RTS frame, it relies on its CS function to figure it out whether there is a transmission going on

[1]This threshold is called *dot11RTSThreshold* in the IEEE 802.11 standard.

in its neighborhood. If the recipient's CS indicates that the medium is idle, then it answers the sender with a CTS. Notice that between the RTS and CTS there is a SIFS time. This time is shorter than a DIFS in order to prevent other users from interfering with this 4-way-handshake. Once the sender receives the answer to its RTS, it prepares itself to send the data frame. After receiving the data frame the recipient prepares the ACK frame to acknowledge the correct reception of the packet. Typically the time duration of RTS updates the NAV timer of the sender's neighbors and the time duration of CTS updates the NAV timer of the recipient's neighbors.
The broadcast transmission scheme is a special case of the basic access mechanism. In the transmission of a broadcast data frame there is no acknowledgement because there are many candidates to send it. Thus, a packet may be lost and in this case it will never be retransmitted. This is the reason why broadcast traffic is unreliable.
The retransmission mechanism is a characteristic of unicast traffic only. Retransmissions attempts are done when:

- The sender transmits an RTS and does not receive the CTS as a response.
- The sender transmits a data frame and the recipient does not acquire it due to bad channel conditions or multiple transmissions at the same time.
- The receiver sends an ACK but the sender does not receive it.

There is a maximum number of retransmissions[2] that a packet can suffer. If the packet is not successfully sent after those retransmitions, it is discarded.

2.1.2 Enhanced Distributed Channel Access

The EDCA enhances the IEEE 802.11 DFC by introducing the traffic differentiation (or QoS). The QoS is handled by introducing four separate queues for different types of traffic. Each one of the queues, or Access Categories (AC), has a separate BEB instance to control the access priority.

[2]The 802.11 specification allows for different retry limits, dot11ShortRetryLimit and dot11LongRetryLimit, for packets that are shorter than and longer than the dot11RTSThreshold, respectively.

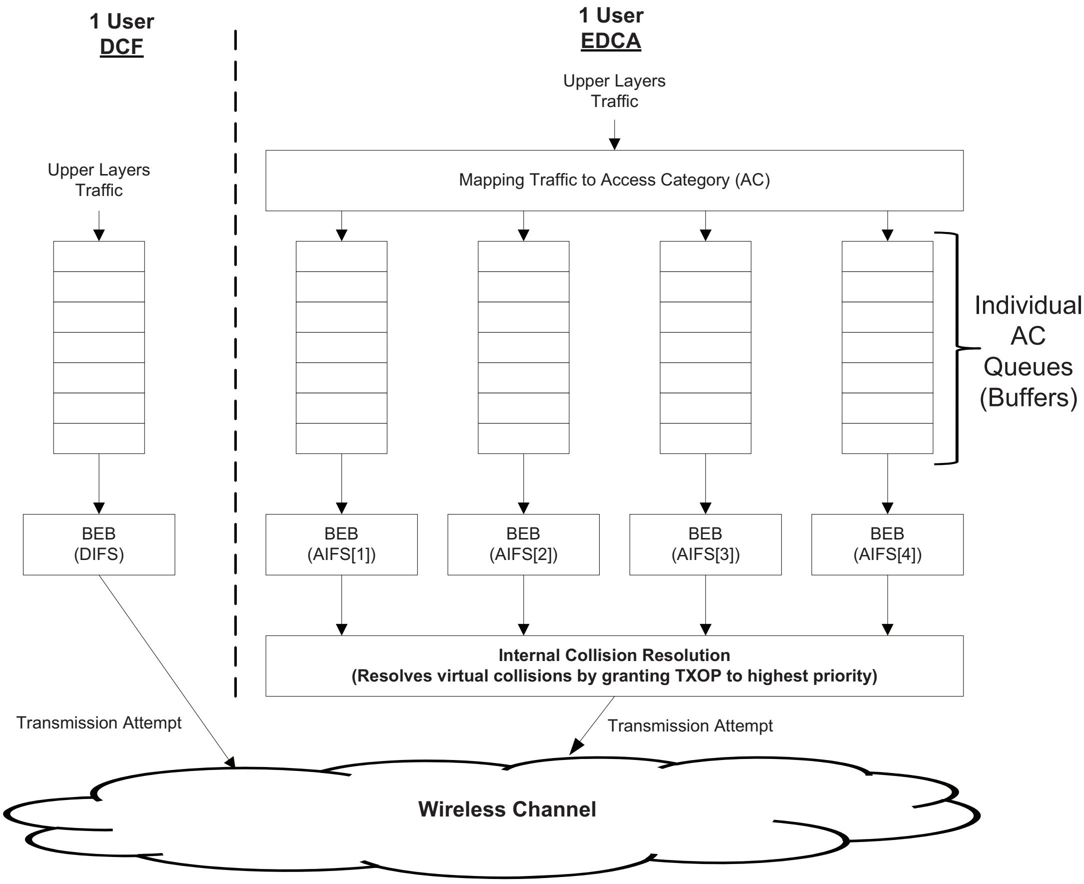


Figure 2.4: EDCA reference model.

The four different ACs are: *Video*, *Voice*, *Background* and *Best Effort*. Basically the multimedia AC (Audio and Video) have higher priority to access the channel. Figure (2.4) shows the basic model of EDCA and its difference to DCF. In EDCA the upper layer traffic if mapped onto the right AC. Each AC has its own BEB with separate parameters to control the access to the *Transmission Opportunity* (TXOP). The idea is that different ACs contend for a TXOP instead of contending to a channel transmission. This is needed because internal collisions can occur and, they are not real collisions but virtual collisions. For that reason, an *Internal Collision Resolution* is needed. When a virtual collision occurs, the internal collision resolution attributes the TXOP to the AC with higher priority. Thus, the AC that could not obtain the TXOP behaves like if a channel collision had occurred.

Each AC has to ensure that the medium is idle for a specific AIFS before a packet transmission. Hence, each AC has its own AIFS time where the highest priority AIFS has the same duration as DIFS. The AIFS times are described by equation (2.1), where AIFSN denotes the arbitration inter-frame space number, which is different for each one of the ACs. Each different AIFS provides an extra way to define priorities between ACs by making the low-priority ACs deferring for a longer time than the high priority ACs.

$$AIFS[AC] = AIFSN[AC] * (IdleSlotDuration) + SIFS \tag{2.1}$$

The transmission schemes for unicast and broadcast of EDCA are basically the same as DCF. However EDCA introduces a mechanism called *Block Acknowledgment* (BA). The idea of BA is similar to the fragmentation burst in legacy IEEE 802.11, but instead of sending several fragments of a frame the sender transmits a burst of data frames belonging to a specific AC. There is no need for any xIFS space between the frames. The last frame of the burst is the BA request. After receiving this BA request the sender answers with a BA frame where it acknowledges all the successfully-received frames.

Besides an improvement to the legacy IEEE 802.11, EDCA is a contemporary protocol, suitable for heterogeneous Ad-Hoc networks.

2.2 Background

ALOHA Protocol. The first *Random Multi-Access* (RMA) protocol used for wireless communications was proposed in 1970 by Abramson [4] and it is often called *Pure ALOHA*. The main idea of pure ALOHA is that whenever a user has a packet to send it shall be transmitted. If the transmitted packet was not successfully received by the recipient, then it is retransmitted in a future opportunity. In pure ALOHA, two or more users can transmit at the same time corrupting each other signals. This is a serious problem especially when the transmission of a packet interferes with the end of a packet already being transmitted. For this reason, it was also proved in [4] that the maximum channel capacity[3] is approximately 18% when fixed packet size is used for pure ALOHA. Later, Gaarder [5] in 1972 proved that the pure ALOHA channel capacity is always superior when fixed packet size is used when compared with variable packet sizes. In 1973, Roberts [6] proposed a different version of ALOHA that requires users to be synchronized. The synchronization means that users coincide the edges of their packet transmissions with an imaginary equal time slot boundary. This version is often called *Slotted ALOHA* and it provides twice the channel capacity of pure ALOHA (see section 3.1.1 for the proof). The increasing popularity of slotted ALOHA arose the necessity of studying models to analyze the protocol performance and stability. Kleinrock [7] introduced in 1975 an important model for slotted ALOHA. This model introduces a 1-D Markov Chain where users are considered to have a single packet buffer. This work studies the throughput-delay tradeoff by showing that the higher is the throughput, the higher is the delay. It also discusses the protocol operation point as well as the stability points. These points are shown through the relationship between the number of packets arriving to all users' buffers (*Channel Arrival*) and the number of packets sent per slot (*Channel Departure*). More detail about this work is shown in chapter 3.

BEB Protocol. Lam [8] presented in 1975 a heuristic called *Retransmission Control Procedure* (RCP) that he claims to be the first draft of BEB. The BEB protocol is able to change the probability with which users access the channel by changing the CW. This characteristic is very important because the protocol adjusts to the number of active users

[3]The channel capacity is the maximum possible number of packets that can be successfully sent in the channel per a time unit.

in the network. The BEB importance made several researchers studying its performance by means of mathematical models. Research works like [9] and [10] describe Markov Chains that are used to find the stationary equilibrium point only for saturated network. The reason why these authors present saturated models is that it is easier when compared to unsaturated ones. In 2003, Kwak et al.[4] propose an infinite 1-D Markov Chain, where each state of the chain represents a growth of the BEB window (often called BEB stage). Hence, this work assumes an infinite number of BEB stages. With this approach it is possible to calculate the saturation throughput by expressing the transmission probability of a packet as a function of CW and the medium access delay can also be analyzed. The simplified BEB model of [10] helps to understand some BEB proprieties (better explained in chapter 4), but in a real system implementation there is a fixed number of BEB stages. Back in 2000, Bianchi [9] uses an approach with 2-D Markov Chain [5]. One of the Markov Chain's dimensions is the BC and the other dimension is the BEB stage. Bianchi assumes that in the stationary equilibrium point, the *conditional collision probability* of each user is constant. This assumption was controversial, however Bordenave et al. [11] prove that it is valid for a large number of users in the network. The assumption leads to an extremely accurate model. Under the same set of assumptions as Bianchi, Tobagi and Medepalli [12] propose a model based on the *Average Cycle Time* approach. On each cycle it is assumed that a user performs a successful transmission. The network throughput is calculated and it is proved by simulations that this model is more accurate than Bianchi's. This difference of accuracy between the models is due to the timings simplifications that Bianchi uses. Unlike Tobagi, Bianchi does not consider EIFS times when a collision occurs in his throughput formula.

Neither Bianchi nor Tobagi take into account several features of the BEB access algorithm for IEEE 802.11, such as finite number of retransmissions. Markov Chains can be used to study the effect of finite number of retransmissions however, few works present closed form transmission probability functions with this technique. In 2009 Andreev et al. [13] introduced the *Regeneration Cycle Concept* to extend [9] in order to account for a finite number of transmission attempts in the analytical model. This approach is a powerful tool

[4]This research work dated from 2003 was later in 2005 extended by the same authors in [10]. From now on only the 2005 research work will be mentioned.

[5]Note that this text presents the research works in order of complexity instead of chronologically.

because it lies on simple mathematical series, which offers an alternative solution to easily extend future models. There is also a research work proposed by Wu [14] that evaluates a finite number of retransmissions in the same way as it is shown in [13] but using a 2-D Markov Chain. This model is not reliable since it does not converge to Bianchi's model when the number of retransmission is infinite (See chapter 4 for the proof). Kwak uses a truncated version of his 1-D Markov Chain in [10] to analyze the effect of finite transmission attempts. This is not the only author using this technique to evaluate influence of the retransmissions number that a packet can suffer, Oliveira et al. [15] also used this technique in their 2-D Markov Chain.

Broadcast Traffic. All the research works for BEB described so far (with an exception of [15]) only account for unicast traffic. However, in a real system there are a co-existence of unicast and broadcast traffic in the network. There are a few research works in the area where broadcast traffic is analyzed. In 2007, Ma [16] studies saturation throughput using only broadcast traffic by means of a 1-D Markov Chain. It was the first research work dealing with the *Consecutive Freeze Process* (CFP) in broadcast. The CFP has different effects on the network performance when unicast or broadcast scheme is considered. When using unicast scheme under saturation, after a busy period, only a user that has finished a successful transmission may access immediately to the channel if the chosen value for BC is zero. However, when using broadcast scheme, CFP happens more often because it does not rely on receiver acknowledgement (or retransmission mechanism). Because of this CFP problem, models that deal only with unicast cannot simply be applied to broadcast. Thus, [16] addresses this problem by dividing the BC into two sub-processes. One process is the *Sequential BEB Process* (SBP) which describes the general BEB procedure without zero initial BC and another process where CFP is modeled by involving consecutive transmissions as a result of zero BC. Later, Ma and Chen [17] extend [16] to include average delay time and the packet delivery ratio[6]. Finally, this research group combines [17] and [16] to present in 2008 an extended work [18]. The novelty of this last work is the analysis of the initial CW value importance for the network performance. They found that CFP effect is neglected when $CW \geq M$, where M is the total number of users.

[6]Defined as the ratio between the number of packets successfully transmitted and the total number of transmitted packets

Another work dealing with saturated broadcast traffic, using similar 1-D Markov Chain approach as in [16] but without the CFP problem, is done by Wang et al. [19] in 2008. They found that for one-hop networks, broadcast reliability does not depend on the frame size but it depends of the initial CW. Moreover, [19] concludes that broadcast has a reliability-throughput tradeoff in which high reliability and maximum throughput cannot be achieved simultaneously. What differentiates [16] from [19] is that later [19] accounts for the freezing of BEB counter when the channel is busy.

Co-existence of Unicast and Broadcast Traffic. Another independent research group had a different research direction by evaluating the mixture of traffic in a network. The first research work proposed in 2006 by Oliveira et al. [15] is an extension of [9] where the percentage of generated broadcast traffic is accounted for. The proposed model describes the aggregated one-hop network dynamics in a saturated network without hidden terminals through a 2-D Markov Chain. Later, Oliveira et al. proposed in 2007 a model based on their previous work to account for non-saturated traffic [20]. This model provides a tool for the total frame delay analysis. The combined results of [15] and [20] are shown in [3] where two important conclusions are made. Firstly, the throughput performance degradation is essentially due to the difference of the transmission schemes used for unicast and broadcast traffic. Secondly, the increase of broadcast traffic in a network does not necessarily degrade significantly the performance when the same bit rate is used for the transmission of these two schemes.

Wang et al. [21] in 2009, proposed a 1-D model that extends [18] and [19] by considering saturated and unsaturated traffic and the freezing process of BEB counter when the channel is busy. Wang concludes that when compared to unicast traffic, broadcast achieves a higher optimal throughput under low traffic conditions, when the bit rate is the same for both transmission schemes and for a few number of users. However, as the load increases the throughput of broadcast deteriorates much faster than in unicast. This work also concludes that [18] is not very accurate as the number of nodes in the network increases because it does not account for the freezing process of BEB. Another research work of Wang et al. [22] was proposed in 2008 to extend and improve [15] and [20] by considering the freezing process of BEB and unsaturated traffic. This work evaluates on a separate basis the unicast and broadcast throughput. They conclude that the differential performance of unicast and broadcast traffic under heavy load conditions results in an unfair division of the available channel resources. This is essentially because BEB only beneficiates unicast traffic under high load conditions.

Heterogeneity. All works presented so far only account for one group of homogeneous users. However, nowadays networks are made of different kinds of users that need service differentiation. Li and Battiti [23] presented in 2003 an extension to Bianchi's work where groups of heterogeneous users are considered. This work deals with saturated traffic conditions and it is proved by simulations that the presented model has an acceptable accuracy when compared with practical results. Later in 2005, Bellalta et al. [24] presented a model for unsaturated traffic heterogeneous network. This work considers two types of traffic flows: elastic and streaming flow. This model is a tool for evaluating the aggregate throughput and queue utilization in heterogeneous networks. In 2007, Malone et al. [25] also proposed a model for non-saturation traffic conditions. This model is more complete than [24] because it evaluates the aggregate throughput and the per-node throughput, queuing mean delay, the influence of the initial CW on throughput and fairness.

Sumary. So far, there are accurate models that evaluate saturated traffic conditions in the network when the mixture of unicast and broadcast is considered like [3] and [22]. However, it would be interesting to have a model with backward compatibility with previous well-known models, which could describe the retransmission mechanism of BEB. Moreover, there is no model able to analyze the influence of the traffic mixture when groups of heterogeneous users are considered. These are the steps that this thesis intends to fulfill in the proposed model of section 2.5.

2.3 System Model Assumptions

This subsection provides a set of general assumptions that restrict the mathematical models of the next chapters. Some parts of this list can be changed to better describe each individual mathematical model. To be more specific, some assumptions are re-defined for a better system conditions understanding. The list of general assumptions was made based on the well-known model by Bertsekas and Gallager [26] and is shown below.

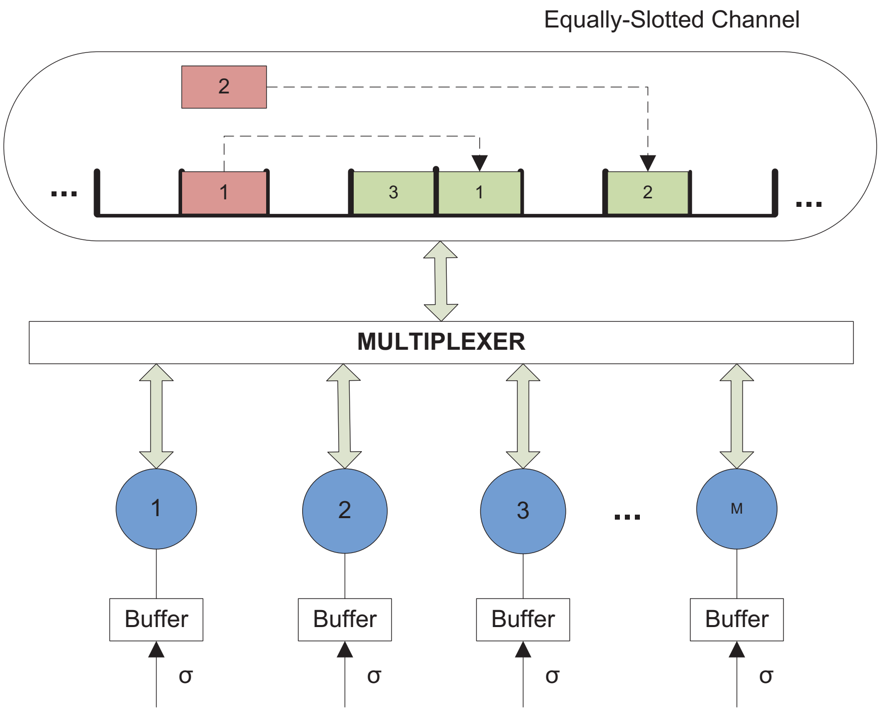


Figure 2.5: Generic network topology.

1. *Communication system*

   (a) *Synchronization*: The channel is divided into equal time intervals (also described as *Equally-Slotted System*). All users are aware when the slots start and end.

   (b) *Fixed Network Topology*: The general network topology is depicted in Figure 2.5. The multiplexer in this figure does not represent a physical device. However, the multiplexer represents in a good way the behavior that users shall have in order to use the channel efficiently. This means that when the multiplexer "rules" are not applied by users because multiple (re)transmissions are done by different users in the same slot, the channel will not be used efficiently.

(c) *Link*: It is assumed one-hop network where there are no hidden users. All users can communicate with each other.

(d) *Data Packets*: All the transmitted packets in the channel are assumed to have the same size. A data packet transmission takes exactly one time slot.

2. *Transmission Channel*

   (a) *Channel Events*: There are only three possible channel events: a successful transmission, an unsuccessful transmission and an idle slot.

      i. *Successful Transmission*: Occurs when exactly one user transmit in a given time slot. This event is often called *Success*.

      ii. *Unsuccessful Transmission*: Occurs when at least two users transmit in a given time slot. This event is also called *Collision*.

      iii. *Idle slot*: Occurs when no user transmits in a given time slot.

   (b) *Perfect Channel Conditions*: It is assumed ideal channel conditions. This means that the only reason why a user does not receive a packet successfully is due to a collision.

   (c) *Channel Input*: The channel input is a variable that represents the sum of all new packets arriving to users' buffers per slot.

   (d) *Channel Output*: The channel output is a variable that represents the number of packets successfully sent per slot.

   (e) *Channel Load*: Defines the possible channel input amounts of traffic.

      i. *Saturated Traffic*: A channel is said to be saturated when the channel input is equal or greater than one packet per slot.

      ii. *Unsaturated Traffic*: A channel is said to be unsaturated when the channel input is less than one packet per slot.

3. *Feedback Information*: This is the information that the sender gets after the transmission. Thus a user always knows if its unicast packets were successfully transmitted or not.

(a) *Actuality*: The feedback information is available at the end of the time slot in which the transmission takes place.

(b) *Reliability*: All the feedback information is error-free.

4. *Users*

   (a) *Buffer Length*: Each user has an individual buffer of fixed size. Typically the buffer can hold one packet at a time. All packets arriving to a full buffer are discarded.

   (b) *Incoming Traffic*: The paket inter-arrival times to users' buffers are independent and identically distributed (i.i.d.). For simplicity reasons, it is assumed that the number of packets arriving to users' queues is Bernoulli distributed in time. This means that a user generates traffic according with σ probability.

   (c) *Users Operation*: For each time slot a user may receive a packet, transmit a packet or stay idle.

   (d) *Users State*: A user is called *backlogged* when it has packets in the buffer ready to be sent and is called a *thinker* if its buffer is empty.

5. *Retransmissions*: After a collision a packet shall be retransmitted. There are two possible different systems: *Lossless System* and *Lossy System*.

   (a) *Lossless System*: After a collision a packet is retransmitted until it is successfully received.

   (b) *Lossy System*: The packet can only be transmitted k number of times (or in other words it can be retransmitted $(k-1)$ times). After k unsuccessful transmission attempts the packet is discarded from the user buffer.

2.4 Classification of Protocols

There are various types of RMA protocols. For this reason, it is important to classify the protocols that are analyzed in this thesis. Protocol classification can be done in different ways, however we adopt the classification from Rom and Sidi [27]. Their Classification is shown in Figure 2.6.

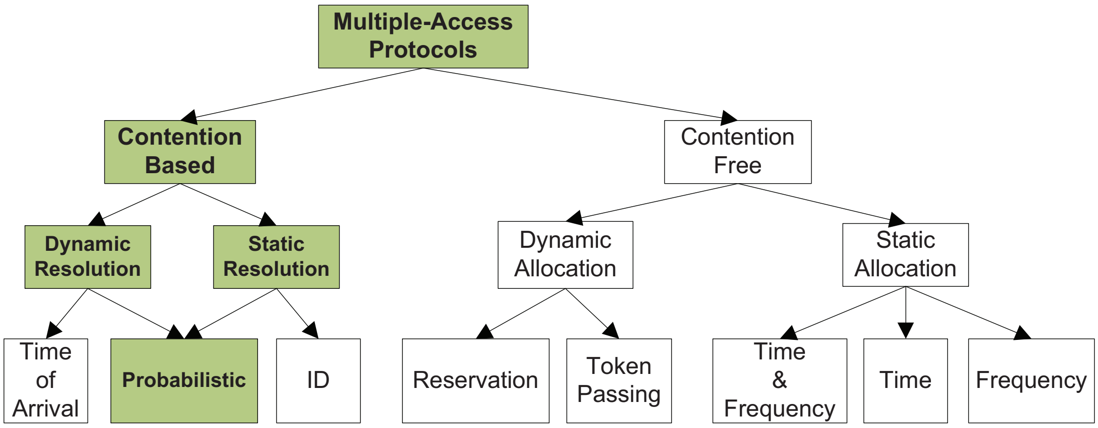


Figure 2.6: Classification of multi-access protocols.

This thesis is focused in analyzing the RMA protocols in which there is no guarantee that the user packet transmission will be successful. These protocols may resolve a collision when it occurs in order to make the network more reliable by using a retransmission mechanism, so that eventually the data packets are transmitted successfully. While solving collisions the protocol does not consume channel resources because there is not a pre-allocation of it for any user. Even idle users do not consume channel resources either, because they do not transmit.

The two ways of solving collisions are: Static and Dynamic resolution. The static resolution means that the channel dynamics does not influence the way that collisions are handled. On the contrary, dynamic resolution relies on channel behavior in order to resolve the collision.

Both, static and dynamic resolution can be probabilistic. For instance in static resolution, the transmission schedule for the users involved in the collision is chosen according with a fixed distribution probability and it is independent of the number of users involved in the collision. Dynamic resolution takes advantage of the system changes so that it changes the access probability to better adjust to the channel conditions.

The protocols on which this thesis is focused are marked with a different colour in Figure 2.6. All ALOHA protocols discussed in the next chapter (with an exception of Optimal ALOHA) belong to Probabilistic Static Resolution with Contention group. Additionally, all BEB protocols (and Optimal ALOHA) belong to Probabilistic Dynamic Resolution with Contention group. The remaining protocols are contained in the category of Contention Free. Here, all transmissions are successful because there is a pre-allocation of resources. This pre-allocation can be whether Static of Dynamic. Static pre-allocation are done by means of *Time Division Multi-Access* (TDMA), *Frequency Division Multi-Access* (FDMA) or a combination of both. Dynamic pre-allocation is done on demand when a user announces the intention of transmitting and reserves the channel resources or by passing a token between users (in which only the token holder is allowed to transmit). No more attention will be paid to these multi-access categories since it is not the scope of the thesis.

Due to the high number of ALOHA versions studied, correspondent classification is done in chapter 3.

2.5 Analytical Model Requirements

The analytical model requirements of this research work are introduced in this section. The first part of the model considered in this work is described in section 4.3.2 and it is concluded in chapter 5. The requirements of the model considered in this work are as follows:

1. Considers one-hop saturated network without hidden users.
2. Convergence to the previous well-known models presented in [10] and [9].
3. It is an extension of [15] and [23], by improving the finite number of (re)transmissions capability and considering groups of heterogeneous users.
4. It uses the Regeneration Cycle Concept approach.
5. It analyzes the network throughput.
6. It allows up to G number of heterogeneous groups. Each heterogeneous group has its own probability for accessing the channel, however inside each group, users are homogeneous.
7. It describes BEB behavior through: the initial CW, the number of BEB states, *the maximum number of packet transmissions* and the amount of broadcast traffic generated.
8. The model shall provide the channel statistics that later on are used to extend the model for an unequally-slotted system.
9. Provides a simplified tool for EDCA evaluation.

Chapter 3

Slotted ALOHA Protocol: Simple Analysis

This chapter is dedicated to a simple analysis of Slotted ALOHA protocol. The simple analysis means that in the analysis, a group of homogeneous users is considered. The reasons why this protocol is being studied are shown below.

- ALOHA is the antecessor of BEB.
- The BEB evolution can be better understood when ALOHA protocol is studied in detail.
- The ALOHA is a special case of BEB.

ALOHA is a simple protocol and it works as follows: when a user is ready for a packet transmission, it transmits immediately the packet in the next possible time slot. Since all users share the same radio channel, when two or more users start their transmissions at the same time, a collision occurs such that all the collided packet transmissions cannot be detected and decoded correctly. The channel collisions shall be handled in such a way that the packets transmissions are scheduled for a random future time.
There are two major groups in this chapter: the *Geometric ALOHA* (GA) and the *Uniform ALOHA* (UA). In GA version users transmit according to a given transmission probability p_t. Each user generate a random number a between 0 and 1, If $p_t \leq a$ the user transmits the packet, otherwise the user defers the packet transmission. UA is different in the sense

that it uses a fixed CW to control the channel access. Each user selects a random integer number W between 0 and $(W_0 - 1)$[1]. This number W is decreased each time that an idle time slot is sensed in the channel. When W reaches 0, the user transmits the packet.
Due to the amount of different versions presented in this section and to make the analysis simpler, a list of acronyms for these GS versions is defined and explained below.

- *Geometric ALOHA - Immediate Transmission* (GA-IT): This version assumes that whenever the user has a new packet, it is sent in the next possible time slot regardless of p_t. If a collision occurs, the packet is retransmitted in some future slot according with p_t. Hence, the immediate transmission is only true for the first packet transmission attempt.

- *Geometric ALOHA - Non-Immediate Transmission* (GA-NIT): Contrarily to what happens in GA-IT, there is no immediate transmission in this version. The packet is always transmitted according with the user's p_t probability. Two types of system are available with this system and both are explained later.

- *Geometric ALOHA - Optimal Transmission* (GA-OT): It is characterized by adapting the users' p_t to the number of backlogged users in the channel. This technique can be applied to ALOHA versions with and without immediate transmission.

Notice that UA admits the same versions as GA. Moreover, later in this chapter GA with Optimal Transmission is introduced and it also assumes the same set of versions as above. Figure 3.1[2] shows the basic user diagram. This diagram has three states: *Idle*, *waiting* and *data frame transmission*. A user is in the idle state whenever it is a thinker (there are no packets in the buffer), this is the state where the packet generation occurs too. The waiting state is reached by the event E1. This event means that the user became backlogged. Hence, in waiting state the user expects the start of the boundary slot so that it can start the transmission. The last state is the data frame transmission and it is reached by the event E2. This event occurs when the start of the slot is detected meaning that the transmission can start. In state number three the user transmits the packet and

[1]To make matters clear CW is referred to be the Contention Window and W_0 is the initial value that CW takes.

[2]This diagram was borrowed from Foh [28]

receives the feedback from the receiver. If the feedback is positive then the user returns to state one, otherwise state 2 is reached again.

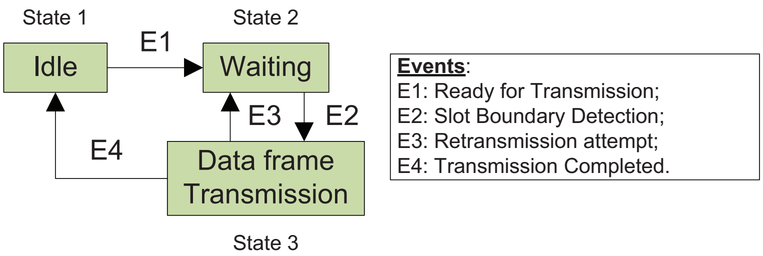


Figure 3.1: Basic user sate diagram.

One shall notice that all this process is a cycle that describes the user medium access process. More attention shall be given to ALOHA versions. Especially, in non-immediate transmission versions there are two possible implementations. These implementations represent the same system, however it has a different mathematical formalism. To make matters more clear a diagram is shown in Figure 3.2. Here, it is shown where this difference of implementations is. Basically, the difference is based on which order users generate and transmit packets in the time slot boundary.

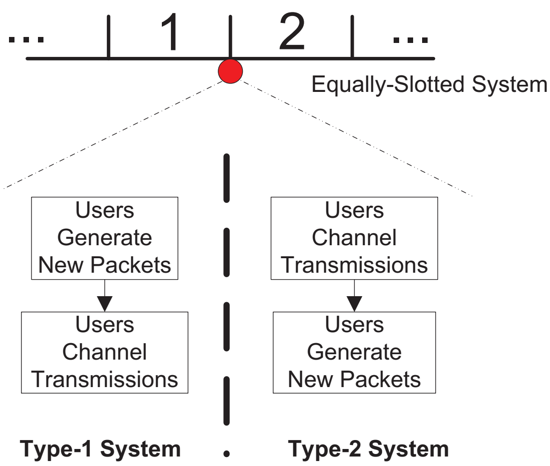


Figure 3.2: System implementation for simulation purposes.

The picture above shows what differentiates the two possible implementation versions for simulation purposes of non-immediate transmission versions of UA and GA. In that figure those different implementations are called *Type-1* and *Type-2* system. Type-1 system is considered when for each beginning boundary of a slot, users generate traffic in first place and then the transmission takes place. The Type-2 system is the opposite behavior

of Type-1: first the transmission takes place and then users generate traffic. No matter each type of system is considered because both of them describe the same system, however the mathematical formalism for each one of the types is different.
In this chapter first the GA is described by means of saturated and unsaturated traffic. Then, only saturated UA is described analytically (because of the mathematical analysis complexity) and finally the chapters' conclusions are made.

3.1 Geometric ALOHA

This subsection presents the GA models for saturated and unsaturated traffic. For unsaturated traffic several models are presented: the immediate transmissions versions, the two types of non-immediate versions and the optimal ALOHA.

3.1.1 Saturated Traffic Conditions

The saturated model is important because it represents the extreme load conditions that a channel can experience with a given number of users. This saturated model is usually easier to analyze mathematically because all users are backlogged thus, it is simpler to deduce the mathematical formulas that describe the MAC behavior. The changes to the general system assumption are:

- 2(e)i: It is assumed saturated traffic conditions.
- 4b: Users buffers' are always full, this means $\sigma = 1$, and consequently all users always have a packet to send.
- 5a: It is considered a lossless system. Users retransmit a given packets until it is successfully transmitted.

All users transmit a packet in a given time slot with the same p_t probability and defer the transmission with $(1 - p_t)$.
Under these saturation conditions, every user has always a packet ready to be sent. A successfully transmitted packet in the channel is observed when exactly 1 of M users transmits and, all the remaining users do not transmit with probability.

$$S_{out} = \begin{pmatrix} M \\ 1 \end{pmatrix} p_t(1-p_t)^{M-1} = Mp_t(1-p_t)^{M-1}. \qquad (3.1)$$

In equation (3.1)[3], p_t represents the probability in which the successful user transmits and $(1-p_t)^{M-1}$ are the remaining users that do not transmit.

Since only one packet can be successfully sent in each time slot, equation (3.1) can also be interpreted as the system departure (S_{out}) in a time slot. To maximize the system departure, it is needed to find the maximum of the equation S_{out} like is shown below.

$$\frac{\partial S_{out}}{\partial p_t} = 0 \leftrightarrow M(p_t \frac{\partial}{\partial p_t}(1-p_t)^{M-1} + (1-p_t)^{M-1}) = 0 \leftrightarrow$$
$$\leftrightarrow M(1-p_t)^{M-1}[1-(M-1)p_t(1-p_t)^{-1}] = 0 \leftrightarrow p_t = 1 \vee p_t = \frac{1}{M}. \qquad (3.2)$$

The probability that maximizes the system departure (p_t^{opt}) under saturation conditions is $\frac{1}{M}$. The other solution of the equation (3.2) does not make sense because if all users transmit with probability one ($p_t = 1$) in each time slot under saturation conditions, they would collide in every transmission attempt.

The maximum possible system departure ($S_{sat}^{\max}$) is given by replacing the optimum transmission probability p_t^{opt} in (3.1), for an infinite channel input ($M \to \infty$).

$$S_{sat}^{\max} = \lim_{M\to\infty} S_{out}(p_t^{opt}) = \lim_{M\to\infty} (1-\frac{1}{M})^{M-1} = \frac{1}{e} \cong 0.36. \qquad (3.3)$$

Equation (3.3) proves that the maximum possible system departure for an infinite arrival rate in probabilistic slotted Aloha is $\cong 36\%$. One might notice that $S_{sat}^{\max}$ can only be archived when all users use the same optimum transmission probability p_t^{opt} under saturation conditions.

3.1.2 Non-Saturated Traffic Conditions

In this section, different GA models for unsaturated traffic are presented. These models describe the system departure, delay and stability analysis. The unsaturated

[3]It is worth to mention that $\begin{pmatrix} M \\ 1 \end{pmatrix}$ is the number of possible combinations (1 user out of M).

channel traffic condition is important because it is the most common traffic condition in networks. In this subsection, the *Geometric ALOHA with Immediate Transmission* (GA-IT), *Geometric ALOHA non-Immediate transmission* (GA-NIT) and the *Geometrical ALOHA - Optimal Transmission* (GA-OT) are described. In GA-NIT, the Type-1 and Type-2 systems are compared in order to make it clear that both versions describe the same system. The changes to the general system assumption are:

- 2(e)ii: It is assumed unsaturated traffic conditions. The traffic generation is assumed to be Bernoulli distributed with σ probability.

- 5a: It is considered a lossless system. Users retransmit a given packets until it is successfully transmitted.

A general Markovian model is formulated based on Kleinrock research [7], for a network with a population of M users. The number of users in the network is considered to be large and finite.

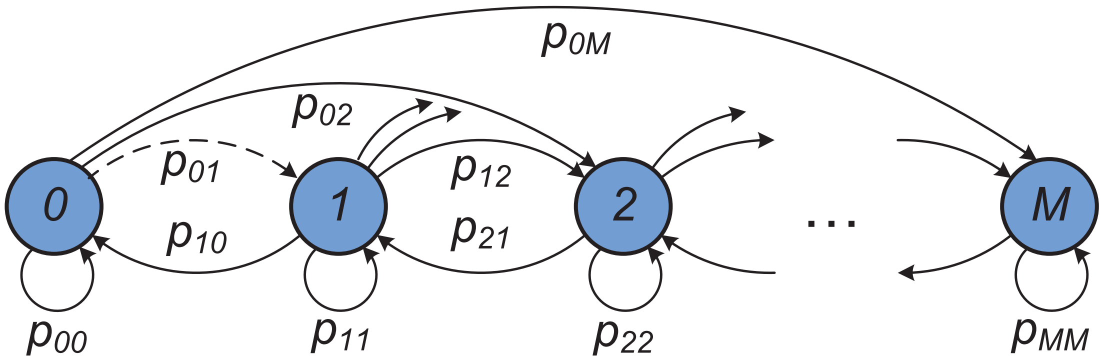


Figure 3.3: Generic Markov chain for ALOHA protocol.

The states in the Markov chain depicted in Figure 3.3 represent number of network backlogged users. There is $M + 1$ chain states representing the $0, 1, 2, \ldots, M$ possible backlogged users. Each version of the GA has its own transition probability matrix.
The dashed arrow in the general Markov chain is only meant to be considered the GA versions without immediate transmission. Moreover, in the GA-IT it is not possible to make the transition from state 0 to state 1 because only one user can transmit successfully in a given time slot. This means that whether only one user transmits successfully and the network remains in state 0 or k users transmit (and collide) at the same time and the

network suffers a k chain forward hop transmission. Note that this property also explains the first equation of all Markov Chains system equations of all GA versions explained later on in this chapter.
All presented versions of GA assume a fixed one size buffer. Like it is explained in [26], these models rely on multiple access channels with large number of users, small arrival rate λ and consequently small delay. With these assumptions, the new arrivals to backlogged users are almost negligible. Thus, this one size buffer assumption provides a lower bound on the delay and system departure. Moreover, it makes the model easier to analyze because only 1-D Markov chain is needed.

Geometric ALOHA - Immediate Transmission (GA-IT)

Here, the GA-IT mathematical model is presented. Notice that in this version of GA, the dashed arrow of Figure 3.3 is not considered due to the immediate transmission feature. Assuming M as the total number of users and N^t as the number of backlogged users at time t, the channel input at time t is $S^t = (M - N^t)\sigma$. Once that a user can only be either a thinker of a backlogged, the difference $M - N^t$ represents the number of thinker users in the network, the ones whose actually can generate new packets. The transition matrix that represents the GA-IT version is shown bellow, where i and j represents the state number (or number of backlogged users) $0, 1, 2, \ldots, M$ at different discrete time instants.

$$p_{i,j} = \Pr\{N^{(t+1)} = j | N^{(t)} = i\} = \tag{3.4}$$
$$= \begin{cases} 0, & \text{if } j \leq i-2, \\ ip_t(1-p_t)^{i-1}(1-\sigma)^{M-i}, & \text{if } j = i-1, \\ (1-p_t)^i(M-i)\sigma(1-\sigma)^{M-i-1}+ & \\ +[1-ip_t(1-p_t)^{i-1}](1-\sigma)^{M-i}, & \text{if } j = i, \\ (M-i)\sigma(1-\sigma)^{M-i-1}[1-(1-p_t)^i], & \text{if } j = i+1, \\ \begin{pmatrix} M-i \\ j-i \end{pmatrix} \sigma^{j-i}(1-\sigma)^{M-j}, & \text{if } j \geq i+2. \end{cases} .$$

The equations presented in (3.4) represent the chain transitions and have their meaning explained in Table 3.1. Notice that regardless of ALOHA version it is not possible to occur more than one hop backwards because only one packet can be successfully transmitted in a given time slot. The expecting channel departure is conditioned on the number of backlogged users in the network for each instant ($N^t = n$).

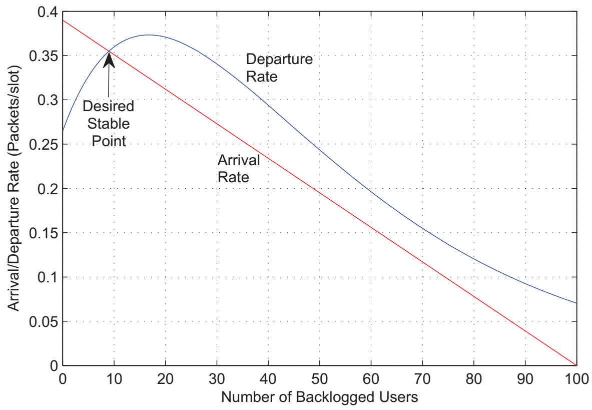
Chain transition	Function description
$j = i - 1$	One backwards hop is achieved when only one of the i backlogged users transmits and no thinker generates a packet.
$j = i$	The network remains in state i if no backlogged user transmit and only one thinker user generates and transmits a packet or, if no backlogged users transmit successfully and none of the thinker users generate a packet.
$j = i + 1$	The chain has one forward hop when only one of the thinker users generates and transmits a packet while the medium is busy (not idle).
$j \geq i + 2$	More than one forward transition hop in the chain can happen when $j - i$ out of $M - i$ thinker users generate and transmit a new packet.

Table 3.1: GA-IT transition equations description.

$$S_{out}(n, \sigma) = (1 - p_t)^n (M - n)\sigma(1 - \sigma)^{M-n-1} + np_t(1 - p_t)^{n-1}(1 - \sigma)^{M-n}. \tag{3.5}$$

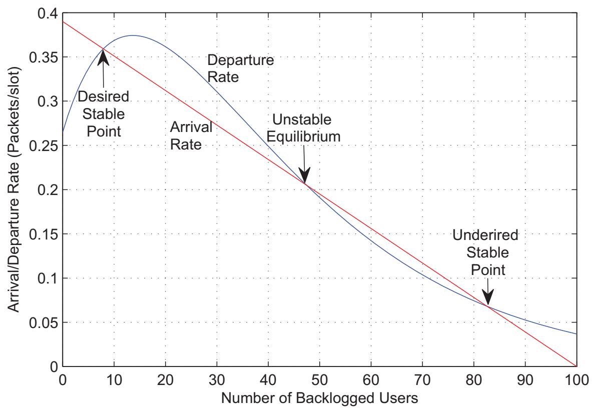
Equation (3.5) is in reality a vector of $M+1$ size. Each index n of the vector $S_{out}(n, \sigma)$ represents the system departure when the network contains n backlogged users. Moreover, a packet is successfully sent in one of the two following cases: only one of the $(M - n)$ thinker users generates and transmits a new packet while none of the n backlogged users transmit or, only one of the n backlogged users transmit while none of the $M - n$ thinker users generate a packet. Notice that equation (3.1) is a particular case of $S_{out}(n, \sigma)$ when $n = M$. This result means that the saturated model is a particular case of the non-saturated model.

According to Kleinrock [7], a channel is defined to be stable if the channel arrivals (S^t) intercept (nontangentially) the channel departure $S_{out}(n, \sigma)$. Otherwise a channel is unstable. Notice that the channel arrival rate is equal to the channel departure rate under equilibrium assumption. In Figure 3.4 two examples of channel stability are depicted. Each figure has two curves, $S_{out}(n, \sigma)$ and the channel arrivals (the linear curve). Both figures were done using $\sigma = 0.0039$, 1 million time slots, 100 users, $p_t = 0.04$ and $p_t = 0.049$ for Figure 3.4 a and b respectively[4].



(a) Stable Channel

Figure 3.4: Arrival and departure rate for GA-IT.



(b) Unstable Channel

Figure 3.4: Arrival and departure rate for GA-IT.

Figure 3.4 shows where the stability points for a stable and an unstable system are. Kleinrock claims in his work that the channel arrivals may have one or more equilibrium points. The author justifies the stability points in his work using fluid analysis of equation

[4]Notice that these 2 figures could also be achieved with different values of p_t, σ and number of users M.

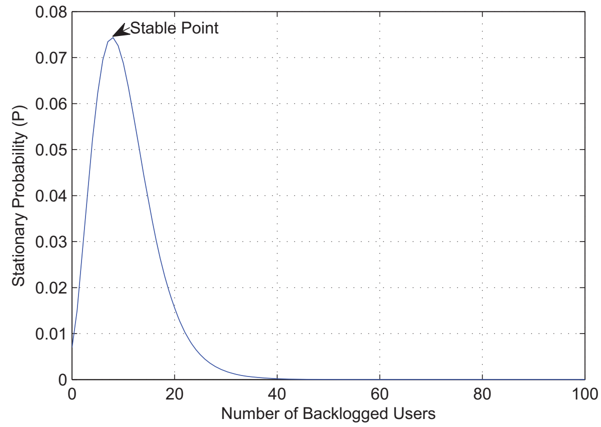
$S_{out}(n, \sigma)$. However, a simpler explanation can be obtained for these stability points if we became aware of the curve's dynamics. For instance an equilibrium point is a point of intersection between the channel arrival curve and the channel departure curve when, from that point on, the arrival rate is always smaller that the departure rate. Otherwise the point of intersection is considered to be unstable. This rationale makes sense because if the thinker users are generating more packets than the channel can handle, it represents an unstable condition of the channel. For a stable system, only one equilibrium point exists. However in an unstable system two equilibrium points exist, which sometimes can also be referred to as desirable and undesirable points like in [26]. The desirable point leads to a high departure rate while the undesirable one leads the system to a departure rate closer to zero. An unstable system alternates between these desirable and undesirable points with time, due to stochastic fluctuations in the system. The desirable and undesirable points are represented in Figure 3.4.
Since the Markov chain has a finite state space and is irreducible, a stationary probability P with $M + 1$ elements always exists[5]. The vector P can be calculated is by solving the system equation in (3.6).

$$P = P.p_{i,j}. \tag{3.6}$$

Like it was explained before, the vector P contains the probability for which a channel is going to have $0, 1, 2, \ldots, M$ backlogged users. The sum of all P elements is always equal to 1. However, under saturated traffic conditions, the last element of P vector $(M + 1)$ takes the value of 1 and, for remaining states it takes the value of 0. The channel contains as many equilibrium points as the number of maximum values in P.

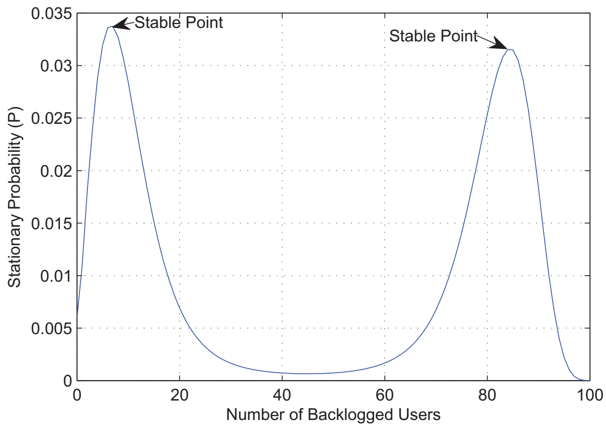
In Figure 3.5 a and b, it is shown the P vector for each one of the two cases depicted in Figure 3.4 respectively. In both cases there is a correspondence between the number of equilibrium points and the number of maximum values in the P vector. As it will be proved further in this section, the steady-state departure rate is closely approximated by the equilibrium point in a stable channel. However, in an unstable channel, the steady-state

[5]Notice that P is only possible if the transition matrix is ergodic, meaning that the sum of each individual row is equal to 1. One shall be aware that this stationary probability P is a vector of $M + 1$ positions. Each position of the vector represents the probability of having n backlogged users in steady state, where n also represents the P vector index.



(a) Stable Channel

Figure 3.5: Stationary probability vector for GA-IT.



(b) Unstable Channel

Figure 3.5: Stationary probability vector for GA-IT.

departure rate cannot be calculated as precisely as in a stable channel. Moreover, in an unstable channel the steady-state departure rate depends on the initial system condition, which is related to the number of initial backlogged users. Using the P vector it is easy to calculate the steady-state channel departure and the average number of backlogged users like it is shown in (3.7) and (3.8) respectively.

$$S_{out} = S_{out}(n, \sigma) \cdot P. \tag{3.7}$$

$$\overline{N} = n \cdot P. \tag{3.8}$$

Equations (3.7) and (3.8) are obtained by applying the internal product between two vectors. This means that in equation (3.8), n represents the vector of $M+1$ size that takes the values of $0, 1, 2, \ldots, M$. According to Little [29], the channel delay can be calculated by dividing the average backlogged users by the steady-state departure:

$$\overline{D} = \frac{\overline{N}}{S_{out}}. \tag{3.9}$$

For all GA protocol versions presented in this section $\overline{D}$, $\overline{N}$ and S_{out} is calculated with the same technique presented in the equations above.

Geometric ALOHA - Non-Immediate Transmission (GA-NIT)

The mathematical formalism of Type-1 system is studied in first place and then the Type-2 system. In Figure 3.3 the generic Markov chain for GA was presented however, this time the dash arrow is considered in the analysis. The transition from state 0 to state 1 is possible whenever the network contains no backlogged users and one thinker user generate a new packet with σ probability. Now, the packet is not transmitted immediately and, the user becomes backlogged. One of the most important differences between GA-IT and GA-NIT is that now σ is only referred to as the probability with which a user generates a packet. The transition equations for GA-IT Type-1 is given in (3.10).

$$p_{i,j} = \Pr\{N^{(t+1)} = j | N^{(t)} = i\} = \tag{3.10}$$
$$= \begin{cases} 0, & \text{if } j \leq i-2, \\ (1-\sigma)^{M-i} i p_t (1-p_t)^{i-1}, & \text{if } j = i-1, \\ (1-\sigma)^{M-i}[1 - i p_t (1-p_t)^{i-1}] + & \\ +(M-i)\sigma(1-\sigma)^{M-i-1}(i+1)p_t(1-p_t)^{i}, & \text{if } j = i, \\ \begin{pmatrix} M-i \\ j-i \end{pmatrix} \sigma^{j-i}(1-\sigma)^{M-j}[1 - j p_t (1-p_t)^{j-1}] + & \\ + \begin{pmatrix} M-i \\ j-i+1 \end{pmatrix} \sigma^{j-i+1}(1-\sigma)^{M-j-1}(j+1)p_t(1-p_t)^{j}, & \text{if } j \geq i+1. \end{cases}$$

One shall remember that this type of system assumes that users generate new packets in first place and then, they try to send them later depending on p_t probability. The meaning of each individual equation is explained in the Table 3.2.

Chain transition	Function description
$j = i - 1$	In order to have only one hop backwards in the chain, any thinker user can generate a new packet and only one backlogged user can transmit.
$j = i$	The network remains in state i in two cases. First, at least 2 backlogged users transmit and collide while none of the thinker users generate a packet. Second, one thinker user generate a new packet while one backlogged user transmits successfully.
$j \geq i + 1$	More than one hop forward in chain happens with one of the two possible events. First, $j - i$ thinker users generate a new packet while any backlogged user transmit. Second, $j - i + 1$ thinker uses generate a new packet while only one backlogged user transmit.

Table 3.2: GA-NIT transition equations description (Considering Type-1 system).

The channel departure is conditioned on σ and the instantaneous number of backlogged users in the network. However, as it is assumed that there is no immediate transmission, equation (3.5) cannot be applied. The $S_{out}(n, \sigma)$ for this type of system considers in first place how many thinker users generated a new packet before the users contend for the channel. This means that a successful transmission occurs whether 1, 2, … or $M - i$ users generate a new packet while only one backlogged user transmits:

$$S_{out}(n, \sigma) = \sum_{k=0}^{M-n} \begin{pmatrix} M - n \\ k \end{pmatrix} \sigma^k (1 - \sigma)^{M-n-k} (n + k) p_t (1 - p_t)^{n-1+k}. \tag{3.11}$$

Now, it is considered the Type-2 system. This type of system is characterized by its mathematical simplicity when compared with GA-NIT type-1 $S_{out}(n, \sigma)$ calculation. In this system it is assumed that users transmit the packets in first place and then, they generate the new packets. The transition equations are shown in (3.12).

$$p_{i,j} = \Pr\{N^{(t+1)} = j | N^{(t)} = i\} = \tag{3.12}$$

$$= \begin{cases} 0, & \text{if } j \leq i-2, \\ ip_t(1-p_t)^{i-1}(1-\sigma)^{M-(i-1)}, & \text{if } j = i-1, \\ [1-ip_t(1-p_t)^{i-1}](1-\sigma)^{M-i}+ & \\ +ip_t(1-p_t)^{i-1}(M-(i-1))\sigma(1-\sigma)^{M-i}, & \text{if } j = i, \\ [1-ip_t(1-p_t)^{i-1}]\begin{pmatrix} M-i \\ j-i \end{pmatrix}\sigma^{j-i}(1-\sigma)^{M-j}+ & \\ +ip_t(1-p_t)^{i-1}\begin{pmatrix} M-(i-1) \\ j-(i-1) \end{pmatrix}\sigma^{j-(i-1)}(1-\sigma)^{M-j}, & \text{if } j \geq i+1. \end{cases}.$$

Since in this system the transmissions are performed in first place, the meaning of the equations (3.12) changes when compared with Type-1 system. The explanations of these transition equations are presented in Table 3.3.

Chain transition	Function description
$j = i-1$	One hop backwards in the chain is done when only one backlogged user transmit and any of the $M-(i-1)$ thinker users generate a new packet.
$j = i$	The system maintains its state in two situations. When at least 2 backlogged users transmit and collide while none of the thinker users generate a new packet. When only one backlogged user transmits successfully a packet while only 1 out of the $M-(i-1)$ users generates a new packet.
$j \geq i+1$	More than one hop forward in chain happens with one of the two possible events. Firstly, $j-i$ thinker users generate a new packet while any backlogged user transmit successfully. Secondly, $j-(i-1)$ thinker uses generate a new packet while only one backlogged user transmit.

Table 3.3: GA-NIT transition equation description (Considering Type-2 system).

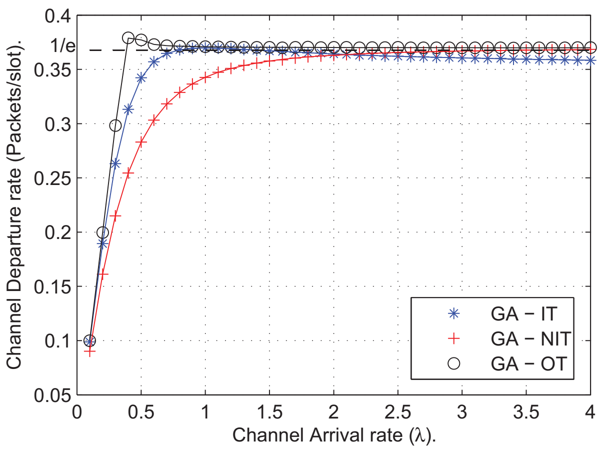
In Type-1 system, $S_{out}(n,\sigma)$ depends on the number of thinker users that generated a packet before the users contend for the channel. However in this system, $S_{out}(n,\sigma)$ does not depend on the number of thinker users that generated a new packet because they only generate a packet after the transmissions attempts. This is the reason why S_{out} takes the form shown in (3.13).

$$S_{out}(n, \sigma) = S_{out}(n) = np_t(1 - p_t)^{n-1}. \tag{3.13}$$

Note how simple equation (3.13) is when compared to (3.11). This difference depends on whether it is assumed that users generate new packets in first place and then transmit it or vice-versa.

Geometric ALOHA - Optimal Transmission (GA-OT)

There is an optimization that can be applied to all GA versions presented before. This optimization is also known as *Geometric ALOHA with Optimal Transmission* (GA-OT) and it is achieved by assuming a probability $p_t = \frac{1}{n}$ for all backlogged users in the network in each time slot[6]. Figure 3.6 compares all versions of ALOHA discussed so far.

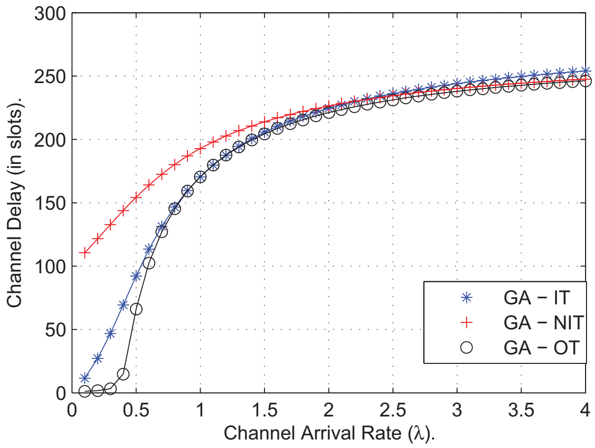


(a) Channel Departure

Figure 3.6: Performance of different protocol implementations.

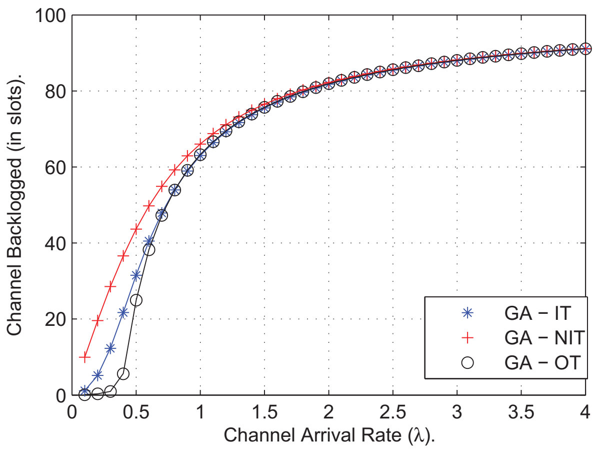
Figure 3.6 shows the GA versions performance for the S_{out}, $\overline{N}$ and $\overline{D}$. The figure was produced assuming 100 equal users and by changing the channel arrival rate λ from 0.01 to 4 packets per slot. It was used the optimal p_t calculated in equation (3.2) for GA-IT and GA-NIT. In the GA-OT used was assumed immediate transmission. Since in GA-NIT, Type-1 and Type-2 systems produce exactly the same results in terms of S_{out}, $\overline{N}$ and $\overline{D}$,

[6]Optimal ALOHA can be analyzed by a mathematical point of view, by doing $p_t = \frac{1}{i}$ in the Markov chain transition matrix for any of the discussed protocol versions.



(b) Average Delay

Figure 3.6: Performance of different protocol implementations.



(c) Average Backlogged

Figure 3.6: Performance of different protocol implementations.

only Type-2 was considered due its mathematical simplicity.

Like it was expected, these three different systems produce different results under unsaturated traffic conditions. However, one might notice that as the arrival rate increases these different versions converge to the same asymptotic values.

3.2 Uniform ALOHA

Like in GA, the *Uniform ALOHA* (UA) also has many versions. All versions of GA can also be considered in UA. The UA constitute a more realistic system because it is based on the window concept. It is also more difficult to analyze mathematically but at the same time is easier to implement in a real system. The changes to the general system assumptions are:

- 2(e)i: It is assumed saturated traffic conditions.
- 4b: Users buffers' are always full, this means $\sigma = 1$, and consequently no traffic distribution is used.
- 5a: It is considered a lossless system.

The UA is a function of an initial window W_0 thus, the transmission probability (p_t) is also a function of W_0, and can be obtained by the regeneration cycle concept.

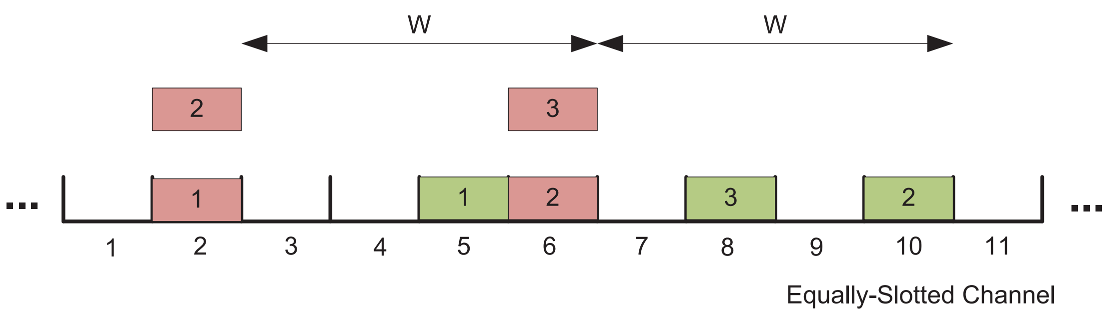


Figure 3.7: Multi-access using uniform ALOHA for $W^{max} = 4$.

A regeneration cycle under saturation conditions is the time interval from which a user generates a random number between 0 and $(W_0 - 1)$ until the transmission finishes (successfully or unsuccessfully). In Figure 3.7, two cycles of user 2 are shown. The first cycle starts in slot number 2 and ends in slot number 6. The second cycle starts in the slot where the collision occurred (slot number 6) and finishes in slot number 10. Notice that the maximum cycle duration is $W_0 - 1$ because a user can only generate a random window within the range of 0 and $(W_0 - 1)$.

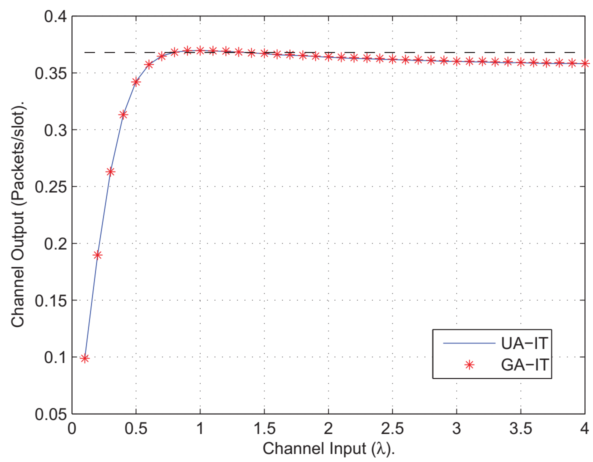
The mathematical definition for p_t is the quotient between the number of transmissions (T_x) in a cycle and the mathematical expectation for the cycle duration[7], like is shown in Equation (3.14).

$$p_t = \frac{\sum\limits_{cycle} T_x}{E[cycle]} = \frac{1}{\sum\limits_{i=1}^{W_0} i \frac{1}{W_0}} = \frac{1}{\frac{1}{W_0}\frac{(1+W_0)W_0}{2}} = \frac{2}{W_0+1} \tag{3.14}$$

Like it was found before in equation (3.2), the optimal transmission probability for saturation is $p_t^{opt} = \frac{1}{M}$. Hence, the optimal window (W^{opt}) for a saturated traffic scenario is calculated by the following doing $p_t = p_t^{opt}$.

$$p_t = {p_t}^{opt} \leftrightarrow \frac{2}{W^{opt}+1} = \frac{1}{M} \leftrightarrow W^{opt} = 2M-1 \tag{3.15}$$

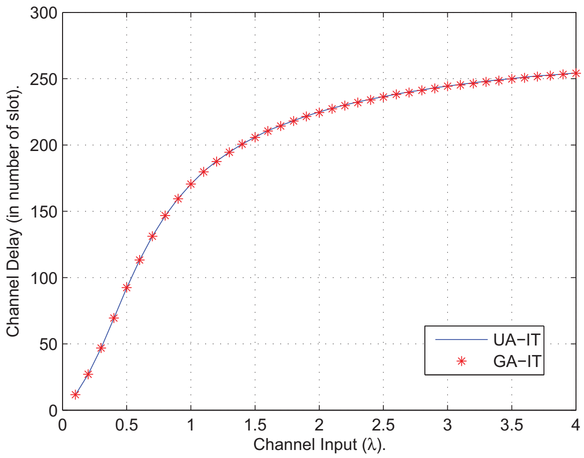
Figure 3.8 shows the difference between the GA-IT and *Uniform ALOHA with Immediate Transmission* (UA-IT) in terms of S_{out}, $\overline{D}$ and $\overline{N}$. This figure was done with 100 users and changing the channel arrival rate in order to cover saturation and unsaturated traffic scenario. For GA-IT and UA-IT was used p_t^{opt} and W^{opt} respectively.



(a) Channel Departure

Figure 3.8: Difference between UA and GA considering immediate transmission.

[7]In this case, the transmission probability p_t can also be calculated by the inverse of the expected number of contending slots ($W = \frac{W_0-1}{2} + 1 = \frac{W_0+1}{2}$).



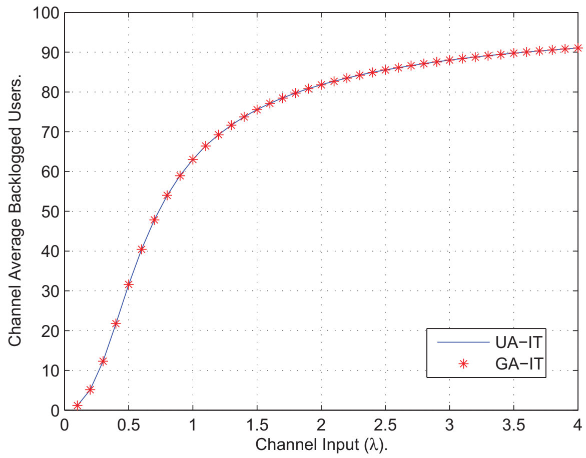
(b) Average delay

Figure 3.8: Difference between UA and GA considering immediate transmission.

Like is shown in Figure 3.8, both versions of ALOHA protocol produce approximately the same results. These results were expected under saturation because the optimum p_t was used. However, for the unsaturated scenario it was not clear if the results would match because the difference between these two statistical distribution functions of both access modes. The difference between this two different statistical distribution functions are shown in Figure 3.9.

By looking at the different distribution functions some conclusions arise. In UA, a user chooses any time slot in the interval between 0 and $W_0 - 1$ with the same $\frac{1}{W_0}$ probability. The maximum number of slots that a user waits is $W_0 - 1$ time slots. However, in GA, each user has higher probability of transmitting with small delay, but on contrary it has also a lower probability to transmit with a high delay.

In Figure 3.10, the difference between the GA-NIT and UA-NIT is shown. Just like what happens with immediate transmission versions, all the analyzed parameters match quite well for saturated and unsaturated traffic without assuming immediate transmission. As expected, the average delay in non-immediate transmission versions is higher than for the versions where immediate transmission is considered. Consequently it increases the average number of backlogged users. Throughput wise, the non-immediate versions appear to be more stable in saturation conditions, when compared with immediate versions, in the sense that the asymptotic limit for throughput is reached and maintained steady as



(c) Average Backlogged

Figure 3.8: Difference between UA and GA considering immediate transmission.

the arrival rate increases. This difference of throughput stability is because in immediate transmission mechanism, collisions are more likely to occur when new packets arrive to their queues.

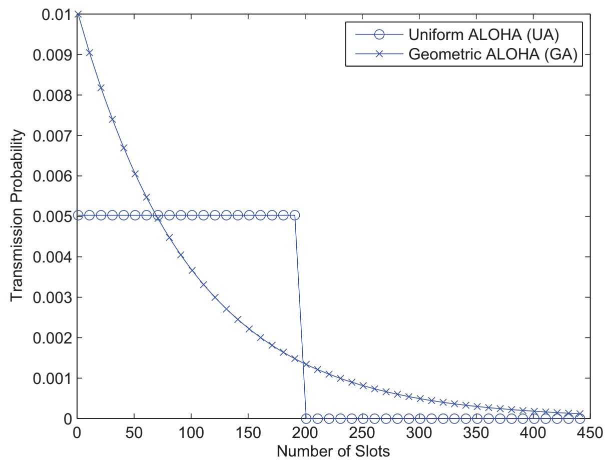
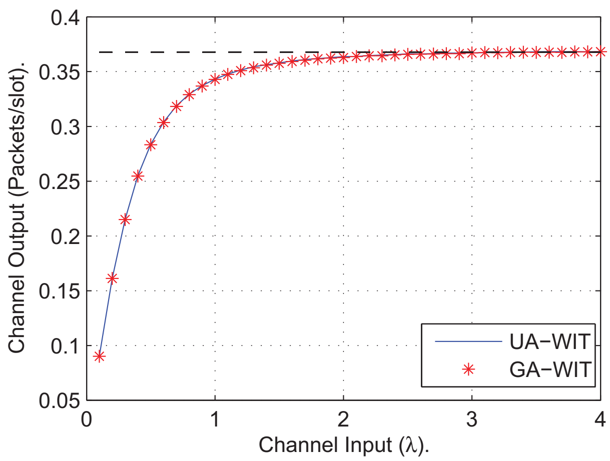
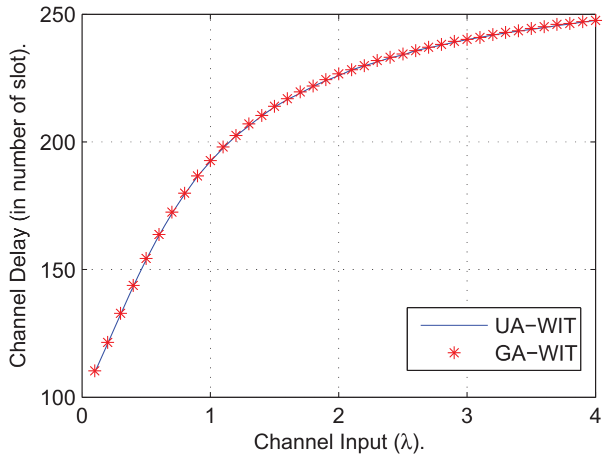


Figure 3.9: Difference between uniform distribution function and geometric distribution function.



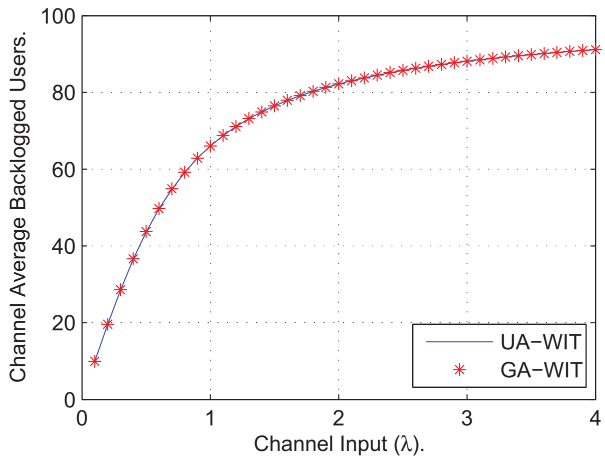
(a) Channel Departure

Figure 3.10: Difference between UA and GA considering non-immediate transmission.



(b) Average delay

Figure 3.10: Difference between UA and GA considering non-immediate transmission.



(c) Average Backlogged

Figure 3.10: Difference between UA and GA considering non-immediate transmission.

3.3 Conclusions

The main conclusion of this chapter is presented below. Notice that all GA versions converge to the same asymptotic limits for high channel arrival rate (saturation). GA-OT, with its adaptive characteristics, is better than other implementations for S_{out}, $\overline{N}$ and $\overline{D}$ under non-saturation traffic conditions. However, GA-OT has the unrealistic assumption that all users know in each time slot the number of backlogged users in the network. Still, the channel departure is higher because p_t is adapted to the number of backlogged users. If just one user is backlogged, p_t is 1 and the immediate transmission is activated. For a large number of backlogged users, p_t is small, decreasing the channel access probability. This adaptive p_t also has some implications for $\overline{D}$ because for small number of backlogged users p_t is higher reducing the average delay of the packet due to the higher channel access probability. The consequence of having small packet delays is that a user spends less time in the backlogged state, decreasing the average number of backlogged users $\overline{N}$.
GA-NIT has a higher $\overline{D}$, as expected because the users' packets are not immediately transmitted after its generation. The consequence of not transmitting a packet immediately is that it stays longer in the buffer causing $\overline{N}$ to increase. Notice that GA-IT's departure rate is typically higher for unsaturated traffic conditions than GA-NIT. This is easier explained because the number of thinker users is also higher and consequently, the users generate packets more often since they do not have any packet in the buffer waiting for transmission. Another important issue that might arise some questions is the fact that for $\lambda = 1$ the number of backlogged users is not 100. However one might start noticing that the channel saturation is different from the user buffer saturation. Hence, on one hand the channel is considered to be saturated when the total number of users generating packets exceeds (or it is equal to) the maximum channel departure rate capacity (typically $\lambda = 1$ packet per slot). On other hand, the saturation of each user's buffer is achieved whenever the packet generation rate is high enough to cause the buffer to be always full (typically for $\sigma = 1$ packet per slot). This is the reason why the number of backlogged users for $\lambda = 1$ is different from 100, because there will only be such a number of backlogged users when $\lambda = 100$ or in other words when $\sigma = 1$.
The general analysis of the $\overline{N}$ and $\overline{D}$ are out of the scope of this thesis. The reason why

it was studied in this chapter was to introduce the throughput-delay tradeoff. Different definitions for throughput can be found in the literature but all of them are somehow related to the definition of system departure described previously. Since $\overline{N}$ and $\overline{D}$ are related to each other, as depicted in Figure 3.6, it is possible to say that there is also a tradeoff between throughput and the average number of backlogged users. The main idea that should not be forgotten is that higher throughput implies higher delay. This is easier explained if one thinks about what happens when throughput increases. When the throughput increases the collision probability in the channel also increases, causing more users to have packets in their buffer waiting for a retransmission and, consequently, increasing the number of backlogged users in the network. No matter which version of ALOHA is considered, because all of them converge approximately to the same asymptotic value under saturation conditions. This means that from this section on, small attention will be paid to the protocol version because the focus of this thesis is the saturated traffic. Notice that when channel saturation occurs $\lambda \geq 1$, GA-IT and GA-NIT do not exactly converge to the same asymptotic throughput limit. This is because the buffers considered only have capacity to enqueue a single frame. If users have buffer lengths greater than one packet, GA-IT converges to the $1/e$. In GA-IT, a packet is only immediately sent if it is the first one to arrive to the queue. The remaining packets in the queue are transmitted according with p_t. Hence, all packets in this model are treated like the first packet arriving to the buffer. This increases the collision probability for high data rates arrivals and consequently the decrease in the throughput.

Considering UA, it does not necessarily have advantages in terms of system departure and mean delay analysis over GA. In fact, UA and GA have similar results in terms of these parameters under saturated and unsaturated traffic conditions. This work focuses on saturated traffic conditions and, it is important to mention that GA and UA converge to the same asymptotic values in terms of all parameters analyzed.

Chapter 4

BEB Protocol: Simple Analysis

This chapter introduces a simple analysis for *Binary Exponential Backoff* (BEB) access mechanism. Lam [8] introduced the Retransmission Control Procedure (RCP) mechanism. This author claims that his work shows that the BEB is a special case of RCP when the function that controls the CW growth is exponential. The simple analysis of BEB protocol means that an equally-slotted channel is considered and all users are homogeneous. One of the reasons why BEB is used instead of ALOHA is because this protocol can be better adapted to the number of users in the network. This is recommended when the network is dynamic in terms of users that are coming in and out. Another reason is that BEB tries to increase p_t in order to achieve a higher system departure rate. By Figure 3.4a, one shall realize that if the channel input increases slightly the system departure rate also increases. The BEB protocol can increase this system departure probability by starting with a small initial window (W_0).
The rules of BEB are as follows:

- On the first packet transmission, the user chooses randomly and uniformly an integer number between 0 and the initial contention window $W_0 - 1$. This is often called the BEB counter;
- For each idle slot, the BEB counter shall be decremented by one;
- Every time the medium is sensed to be busy the BEB counter is frozen;
- After a collision the BEB doubles the contention window according to the adaptation

defined in equation (4.1):

$$\begin{cases} W_i = 2^i W_0, & \text{if } 0 < i < m \\ W_i = 2^m W_0, & \text{if } i \geq m \end{cases}. \tag{4.1}$$

In equation (4.1), m is called the maximum BEB stage and i is the number of unicast packets retransmissions. The stage is the number of times that the window is doubled.

- The window is reset when a packet is successfully sent or when a packet is discarded due to the maximum number of retransmissions.

This chapter starts to introduce the regeneration cycle concept in a formal way. Then, this concept is applied to the following works: [10], [9] and [13]. The proposed model of this thesis is then formulated by extending [13] to account for the co-existence of different types of traffic and finally the backwards compatibility of the model is discussed.

4.1 Regeneration Cycle Approach

In section 3.2, a regeneration cycle under saturation conditions was defined as being the time interval from which a user generates a random number between 0 and $W_0 - 1$, until the transmission finishes (successfully or unsuccessfully). However this is only true if the CW does not change. When CW changes the regeneration process depends on the CW growth and on the maximum number of retransmissions that a packet can suffer. Hence, this thesis approaches three types of a regeneration cycle:

- A regeneration cycle for a system with fixed CW (already presented in section 3.2);
- A regeneration cycle for a variable CW and infinite transmission attempts;
- A regeneration cycle for variable CW and finite transmission attempts.

A regeneration cycle for an infinite BEB stages and infinite transmission attempts is shown in Figure 4.1.

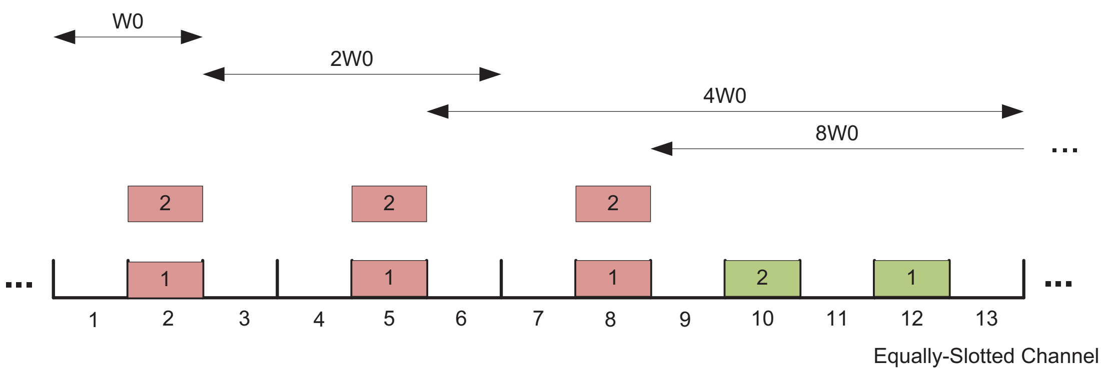


Figure 4.1: Regeneration cycle concept for an infinite number of BEB stages and an infinite number of (re)transmissions.

Figure 4.1 represents a lossless system in which a packet is sent until it is successfully received. Here, the regeneration cycle starts when a user generates a random number between 0 and $W_0 - 1$, until the transmission finishes successfully. The regeneration cycle starts in slot 1 for users 1 and 2 and it ends in slots 10 and 12 for users 2 and 1, respectively. All BEB stages shall be accounted for the regeneration cycle because these stages influence the number of slots that a user has to wait until finishing the transmission. In this Figure, the packets from users 1 and 2 are successfully sent after three packet transmissions (or two packet retransmissions) and three BEB stages.

Figure 4.2 shows the regeneration cycle concept for a finite number of BEB stages and a finite number of (re)transmissions.

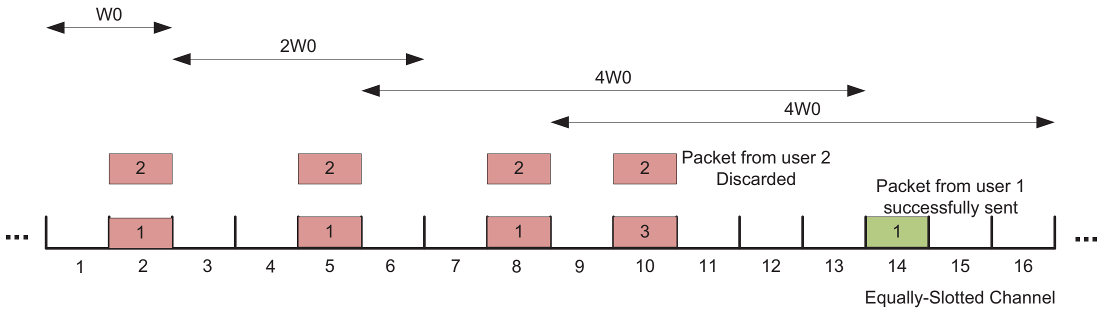


Figure 4.2: Regeneration cycle concept for a finite number of BEB stages and a finite number of (re)transmissions.

Figure 4.2, shows the regeneration cycle concept for a finite number of BEB stages m, and a finite number of transmissions k[1]. In this Figure it is considered that $m = 2$ and $k = 4$. This means that the maximum number of BEB stage growth is 2 and after 4 transmission attempts the packet is discarded no matter if it is successfully transmitted or not. Here, the regeneration starts when the user selects a random number between 0 and $W_0 - 1$ for the first packet transmission and it ends whether the packet is successfully sent or discarded. To be more specific, the regeneration cycles from users 1 and 2 starts in the slot 1. However the regeneration cycle ends at slots 10 and 14 for user 2 and 1, respectively. The regeneration cycle ended for user 2 due to a collision in the last retransmission opportunity and it ended for user 1 due to a successful transmission at the last retransmission.

Notice that the regeneration cycle concept does not consider the number of slots before the packet transmission. In fact this concept also accounts for the average number of packet transmission during the regeneration cycle. Hence, the equation to compute the probability with which a user access the medium (often called transmission probability) p_t is considered as follows:

[1] At this point is necessary to distinguish the difference between the number of transmission attempts k and the number of retransmissions. The number of retransmissions is equal to the number of transmissions minus one $(k - 1)$.

$$p_t = \lim_{n \to k} \frac{\sum_{i=1}^{n} B^{(i)}}{\sum_{i=1}^{n} D^{(i)}} = \frac{E[B]}{E[D]}. \tag{4.2}$$

Equation (4.2) represents the general expression for computing p_t in this particular regeneration process. In this equation, $B^{(i)}$ is the number of transmissions in a cycle and $D^{(i)}$ is the mean number of contending time slots for the i^{th} transmission attempt. Notice that $B^{(i)}$ and $D^{(i)}$ are independent and identically distributed with respect to i, but both are dependent on the conditional collision probability p_c. Here, $E[B]$ stands for the average number of transmissions in a cycle and $E[D]$ the average number of slots that a user shall wait until it starts a transmission. Moreover, the limit in this equation changes according to the system that one wishes to evaluate. Thus, for an infinite number of BEB stages and an infinite number of transmissions, $n \to \infty$. Notice that, equation (3.14) is a special case of equation (4.2) in which $E[B] = 1$ and $E[D] = (\frac{W_0+1}{2})$, which is the average number of slots that a user waits until it starts a transmission.

The regeneration cycle concept is presented in section 4.2 for an infinite number of BEB stages and an infinite number of transmission attempts k. The regeneration approach for a maximum number of BEB stages m and a fixed number of transmission attempts k is presented in section 4.3.

4.2 Lossless System Operation

Here the regeneration cycle concept is applied to works [9] and [10] because these works represent a lossless system. The list of changes to the general system assumptions are:

- 2(e)i: It is assumed channel saturation.
- 4b: Users buffers' are always full, this means $\sigma = 1$, and consequently all users always have a packet to send.
- 5a: It is considered a lossless system. Users retransmit a given packet until it is successfully transmitted.
- The conditional collision probability (p_c) is assumed to be constant, no matter what is the number of retransmissions that a packet suffers. This is the probability that is "seen" by a transmitted packet. This assumption was proved to be true in [11].

4.2.1 Model from Kwak et al.

To the simplified BEB model presented in [10], the regeneration cycle concept for an infinite number of BEB stages can be applied as follows. The stationary transmission probability is obtained through equation (4.2) when $n \to \infty$.
The mathematical expectation $E[B]$ is given by:

$$\begin{aligned} E[B] &= \sum_{i=1}^{\infty} i \Pr\{B = i\} = \sum_{i=1}^{\infty} i {p_c}^{i-1}(1-p_c) = (1-p_c)\sum_{i=1}^{\infty} i {p_c}^{i-1} = \\ &= (1-p_c)\frac{\partial}{\partial p_c}\left(\sum_{i=1}^{\infty} {p_c}^{i}\right) = (1-p_c)\frac{\partial}{\partial p_c}\left(\frac{p_c}{1-p_c}\right) = \frac{1}{1-p_c}. \end{aligned} \tag{4.3}$$

The mathematical expectation of the number of transmissions, $E[B]$, follows a geometric distribution with parameter $1 - p_c$. This expression is well-known and it is also referenced in [9], where p_c is considered constant for all users.

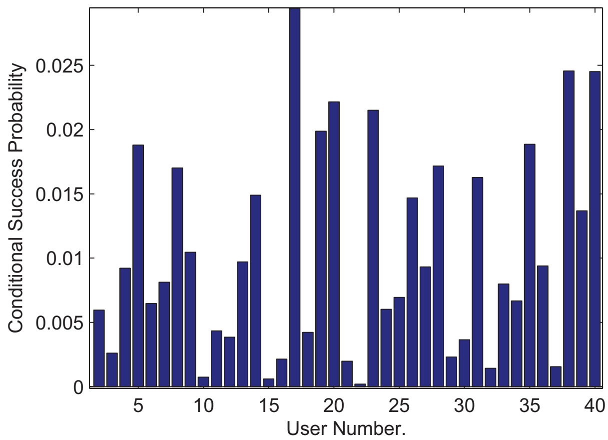
$$
\begin{aligned}
E[D] &= \sum_{i=1}^{\infty} D(i)\Pr\{B=i\} = \sum_{i=1}^{\infty} D(i)\left({p_c}^{i-1}(1-p_c)\right) = \\
&= \sum_{i=1}^{\infty}\left(\frac{(2^i-1)W_0+i}{2}\right)\left({p_c}^{i-1}(1-p_c)\right) = \left(\frac{1-p_c}{2}\right)\sum_{i=1}^{\infty}\left((2^i-1)W_0+i\right){p_c}^{i-1} = \\
&= \left(\frac{1-p_c}{2}\right)\left[\frac{W_0}{p_c}\sum_{i=1}^{\infty}(2p_c)^i - \frac{W_0}{p_c}\sum_{i=1}^{\infty}{p_c}^i + \frac{\partial}{\partial p_c}\left(\sum_{i=1}^{\infty}{p_c}^i\right)\right] = \\
&= \left(\frac{1-p_c}{2}\right)\left[\frac{2W_0}{1-2p_c} - \frac{W_0}{1-p_c} + \frac{1}{(1-p_c)^2}\right] = \frac{W_0(1-p_c)+(1-2p_c)}{2(1-p_c)(1-2p_c)}.
\end{aligned}
\tag{4.4}
$$

The mathematical expectation of the regeneration cycle duration, $E[D]$, is represented in equation (4.4). To obtain $E[D]$, it is necessary to multiply the mean duration of the cycle with i transmissions by the probability in which i transmission attempts are made ($Pr\{B=i\}$).

Finally, the (re)transmission probability p_t for an infinite number of backoff stages is:

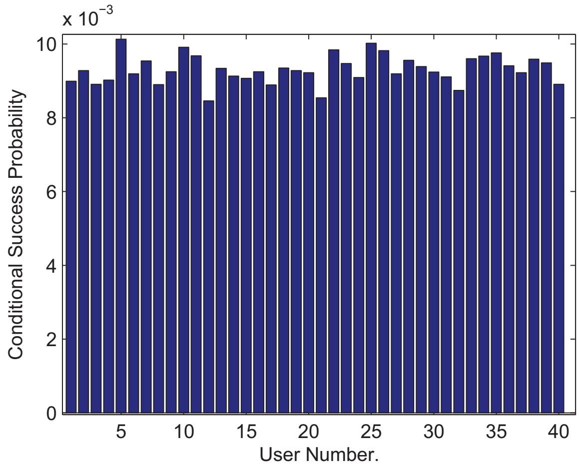
$$
p_t = \frac{E[B]}{E[D]} = \frac{\frac{1}{1-p_c}}{\frac{W_0(1-p_c)+(1-2p_c)}{2(1-p_c)(1-2p_c)}} = \frac{2(1-2p_c)}{W_0(1-p_c)+(1-2p_c)}. \tag{4.5}
$$

The equation (4.5) is the special case of the equation (17) in [10] when $r=2$ (BEB). Notice that the series $\sum_{i=1}^{\infty}(2p_c)^i$ of equation (4.4) imposes the constraint $p_c < 0.5$, this means that equation (4.5) is only valid for this constraint. The model described introduces the capture effect problem. This problem arises when the initial BEB window is too small, making only a few users to consume the whole transmission channel resources. With the capture effect the throughput increases because most of the users do not transmit for a long period of time. The reason behind the long periods without transmitting is due to the several collisions that users may have experienced. These collisions caused them to double the CW and consequently, users have to defer for longer periods of time. Once one of these users performs a successful transmission, the CW is reset and that user is more likely to transmit again in the channel. In this way, small groups of users tend to capture the channel in turns. The medium access becomes unfair because of this capture effect, like it is possible to observe in Figure 4.3.



(a) Network with capture effect ($W_0 = 8$).

Figure 4.3: Capture effect consequences on fairness for a network with 40 users.



(b) Network without capture effect ($W_0 = 32$).

Figure 4.3: Capture effect consequences on fairness for a network with 40 users.

The data contained in the Figure 4.3 was achieved for 40 users and W_0 equal to 8 and 32 were used for Figures (a) and (b), respectively. The Figure shows the consequences in terms of conditional success probability. This is the probability of success conditioned on the fact that a user transmits a packet successfully. Here it is clear how unfair the network can become if a small contention window is used when there is no limit on the number of BEB stages. This also introduces the tradeoff between the throughput and fairness, meaning that the higher the throughput is the lower is the fairness. For $W_0 = 1$ the capture is total because a single user captures the channel with probability 1. Due to the capture effect problem and its consequences, the IEEE 802.11 standard introduces a maximum number of BEB stages m to avoid it. This effect was studied also in [10].

4.2.2 Model from Bianchi

Assuming a more realistic BEB behavior, where there is a limit on the number of BEB stages (m), $E[B]$ does not need to be recalculated since equation (4.3) is still valid because it is assumed an infinite number of transmission attempts per packet. However, $E[D]$ takes a different value since after $m+1$ transmission attempts, CW remains constant and equals to $W^{\max}$. In fact, the regeneration cycle duration expression $D(i)$ has now two different expressions. An expression for the case when there are less than $m+2$ number of transmission attempts and, another expression for the subsequent transmission attempts.

$$D(i) = \begin{cases} 2^{i-1}W_0 - \frac{W_0-i}{2}, & \text{if } 1 \leq i \leq m+1 \\ 2^{m-1}W_0(i-m+1) - \frac{W_0-i}{2}, & \text{if } m+1 \leq i \leq +\infty \end{cases}. \tag{4.6}$$

$$\begin{aligned} E[D] &= \sum_{i=1}^{\infty} D(i) \Pr\{B=i\} = \\ &= (1-p_c)\left[\sum_{i=1}^{m+1}\left(2^{i-1}W_0 - \frac{W_0-i}{2}\right)p_c^{i-1} + \sum_{i=m+2}^{\infty}\left(2^{i-1}W_0(i-m+1) - \frac{W_0-i}{2}\right)p_c^{i-1}\right] = \\ &= (1-p_c)\left[\sum_{i=1}^{\infty}\left(\frac{i-W_0}{2}\right)p_c^{i-1} + \sum_{i=1}^{m+1}(2p_c)^{i-1}W_0 + \sum_{i=m+2}^{\infty}\left(2^{i-1}W_0\left(i-m+1\right)\right)p_c^{i-1}\right] = \\ &= \frac{1-W_0\left(1-p_c\right)}{2\left(1-p_c\right)} + \frac{W_0\left(1-p_c\right)\left(1-\left(2p_c\right)^{m+1}\right)}{\left(1-2p_c\right)} + 2^{m+1}W_0\left(\frac{p_c^{m+1}\left(3-2p_c\right)}{\left(1-p_c\right)}\right) = \\ &= \frac{1-2p_c+W_0-W_0p_c-\left(2p_c\right)^m W_0 p_c}{2\left(1-p_c\right)\left(1-2p_c\right)} = \frac{\left(1-2p_c\right)\left(W_0+1\right)+p_cW_0\left(1-\left(2p_c\right)^m\right)}{2\left(1-p_c\right)\left(1-2p_c\right)}. \end{aligned} \tag{4.7}$$

Using equations (4.3) and (4.7) the (re)transmission probability p_t is easily obtained for a system with infinite number of transmission attempts and a finite number of backoff stages.

$$
\begin{aligned}
p_t = \frac{E[B]}{E[D]} &= \frac{\frac{1}{1-p_c}}{\frac{(1-2p_c)(W_0+1)+p_cW_0(1-(2p_c)^m)}{2(1-p_c)(1-2p_c)}} = \\
&= \frac{2(1-2p_c)}{(1-2p_c)\left(W_0+1\right)+p_cW_0\left(1-(2p_c)^m\right)}.
\end{aligned}
\tag{4.8}
$$

Equation (4.8) is also introduced in [9], where a 2D-Markov Chain is used. Certainly, this equation is more complex than (4.5) since it considers a fixed number of BEB growths. Here, the capture effect is less likely because the maximum number of idle slots for which a user has to wait before it retransmits a packet after $m+1$ times is $W^{\max}-1$.

4.3 Lossy System Operation

In this section the formulation described in [13] is discussed. Then the first part of the proposed model of this thesis is deduced. Both models are lossy models in which, after k transmission attempts the packet is discarded. Moreover there is also a limit number of m BEB stages. The list of changes to the general system assumptions are essentially the same as in section 4.2 for the lossless system with an exception of:

- 5b: It is considered a lossy system. Users transmit a given packets at most for k times, after which it is discarded.

4.3.1 Model from Andreev et al.

A real system usually accounts for a maximum number m of CW growths and a maximum number k of transmissions attempts per packet. To compute the (re)transmission probability of such a system, $E[B]$ and $E[D]$ shall be recalculated.

$$
\begin{aligned}
E[B] &= \sum_{i=1}^{k} i \Pr\{B = i\} = \sum_{i=1}^{k} i p_c^{i-1}(1-p_c) + p_c^k k = \\
&= (1-p_c)\frac{\partial}{\partial p_c}\left(\sum_{i=1}^{k} p_c^i\right) + p_c^k k = \\
&= \frac{(1-(k+1)p_c^k)(1-p_c) + p_c - p_c^{k+1}}{(1-p_c)} + p_c^k k = \\
&= \frac{1-p_c^k}{(1-p_c)}.
\end{aligned}
\tag{4.9}
$$

The term $p_c^k k$ accounts for the fact that a packet is discarded. Notice that there are two possible expressions for $E[D]^2$, one in which $k \leq m+1$ and another one for $k > m+1$, where $k, m \in \mathbb{N}$ and both parameters are finite. These two expressions for $E[D]$ will result in two different expressions for p_t as it will be presented later.

[2]This two expressions are represented as $E^{(1)}[D]$ and $E^{(2)}[D]$. One shall not confuse this notation with the expected value of first and second order for $E^1[D]$ and $E^2[D]$, respectively.

$$
\begin{aligned}
E^{(1)}[D] &= \sum_{i=1}^{k} D^{(1)}(i)\Pr\{B=i\} + p_c^k D^{(1)}(k) = \\
&= \sum_{i=1}^{k}\left[\left(2^{i-1}W_0 - \frac{W_0-i}{2}\right)p_c^{i-1}(1-p_c)\right] + p_c^k\left(2^{k-1}W_0 - \frac{W_0-k}{2}\right) = \\
&= (1-p_c)\sum_{i=1}^{k}\left[\left((2p_c)^{i-1}W_0\right) - \left(\frac{W_0}{2}p_c^{i-1}\right) + \left(\frac{1}{2}\right)\frac{\partial}{\partial p_c}\left(p_c^i\right)\right] + \\
&+ p_c^k\left(2^{k-1}W_0 - \frac{W_0-k}{2}\right) = \\
&= \frac{(1-p_c)}{2}\left[\left(\frac{W_0}{p_c}\right)\left(\frac{2p_c-(2p_c)^{k+1}}{(1-2p_c)} - \frac{p_c-p_c^{k+1}}{(1-p_c)}\right) + \frac{\partial}{\partial p_c}\left(\frac{p_c-p_c^{k+1}}{(1-p_c)}\right)\right] + \\
&+ p_c^k\left(2^{k-1}W_0 - \frac{W_0-k}{2}\right) = \\
&= \frac{(1-p_c)}{2}\times\Big(\frac{W_0(1-p_c)^2\left(2-2^{k+1}p_c^k\right)}{2(1-2p_c)(1-p_c)} + \frac{W_0(1-2p_c)(1-p_c)\left(1-p_c^k\right)}{2(1-2p_c)(1-p_c)} + \\
&+ \frac{\left(1-kp_c^k-p_c^k+kp_c^{k+1}\right)(1-2p_c)}{2(1-2p_c)(1-p_c)}\Big) + \frac{\left((2p_c)^kW_0 - W_0p_c^k + kp_c^k\right)(1-2p_c)(1-p_c)}{2(1-2p_c)(1-p_c)} = \\
&= \frac{W_0(1-p_c)\left(1-(2p_c)^k\right) + (1-2p_c)\left(1-p_c^k\right)}{2(1-2p_c)(1-p_c)}.
\end{aligned}
\tag{4.10}
$$

Equation (4.10) is obtained by using the first row of equation (4.6). The second expression for $E[D]$ is deduced below.

$$
\begin{aligned}
E^{(2)}[D] &= \sum_{i=1}^{m+1} D^{(1)}(i)\Pr\{B=i\} + \sum_{i=m+2}^{k} D^{(2)}(i)\Pr\{B=i\} + p_c^k D^{(2)}(k) = \\
&= (1-p_c)\left[\sum_{i=1}^{k}\left(\frac{i-W_0}{2}\right)p_c^{i-1} + \sum_{i=1}^{m+1}(2p_c)^{i-1}W_0 + \sum_{i=m+2}^{k}\left(2^{m-1}W_0(i-m+1)\right)p_c^{i-1}\right] + \\
&+ p_c^k\left(2^{m-1}W_0(k-m+1) - \frac{W_0-k}{2}\right) = \\
&= \frac{\left(1-kp_c^k-p_c^k+kp_c^{k+1}\right)(1-2p_c) - W_0\left(1-p_c^k\right)(1-2p_c)(1-p_c)}{2(1-2p_c)(1-p_c)} + \\
&+ \frac{W_0(1-p_c)^2\left(2-2^{m+1}p_c^{m+1}\right)}{2(1-2p_c)(1-p_c)} + \\
&+ \frac{2^m W_0\left[\left((m+2)p_c^{m+1}-(k+1)p_c^k\right)(1-2p_c) + (m+1)\left(p_c^{m+1}-p_c^k\right)(1-2p_c)(1-p_c)\right]}{2(1-2p_c)(1-p_c)} + \\
&+ \frac{p_c^k\left(2^m W_0(k-m+1) - W_0 + k\right)(1-2p_c)(1-p_c)}{2(1-2p_c)(1-p_c)} = \\
&= \frac{(1-2p_c)\left(W_0\left(1-2^m p_c^k\right) + \left(1-p_c^k\right)\right) + p_c^k W_0\left(1-(2p_c)^m\right)}{2(1-2p_c)(1-p_c)}.
\end{aligned} \tag{4.11}
$$

Finally, with the equations (4.9), (4.10) and (4.11) it is possible to write the two expressions for p_t.

$$
\begin{aligned}
p_t &= \frac{E[B]}{E[D]} = \\
&= \begin{cases}
\dfrac{\frac{1-p_c^k}{(1-p_c)}}{\frac{W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)}{2(1-2p_c)(1-p_c)}}, k \le m+1 \\
\dfrac{\frac{1-p_c^k}{(1-p_c)}}{\frac{(1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c^k W_0(1-(2p_c)^m)}{2(1-2p_c)(1-p_c)}}, k > m+1
\end{cases} = \\
&= \begin{cases}
\dfrac{2(1-2p_c)\left(1-p_c^k\right)}{W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)}, k \le m+1 \\
\dfrac{2(1-2p_c)\left(1-p_c^k\right)}{(1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c^k W_0(1-(2p_c)^m)}, k > m+1
\end{cases}.
\end{aligned} \tag{4.12}
$$

Note that equation (4.12) is a special case of equation (13)[3] in [13], for $K = 1$. Moreover, the equation defined by the constraint $k \leqslant m+1$ was also deduced by Kwak

[3] In this equation Q represents the number of retransmissions.

et al. [10] and by Wu et al. [14][4]. The work in [10] describes that the constraint $p_c < 0.5$ does not apply for this lossy system. The authors explain this by considering $p_c \geqslant 0.5$ and by equation (4.12), it converges to 0 as k goes to infinity. However, if p_t converges to 0 this means that there are no transmissions and consequently p_c converges to 0 too, contradicting the first assumption.

4.3.2 Proposed Model

Even though describing a more realistic system, equation (4.12) only considers unicast traffic. However, users can typically transmit both broadcast and unicast traffic. As it is well known, broadcast relies on one transmission attempt only and, because of this equation (4.12) shall be changed in order to account for this kind of traffic. It is assumed that each user generate a broadcast packet with probability p_b, and a unicast packet with probability $p_u = 1 - p_b$. First, the mathematical expectation for a (re)transmissions attempts shown in (4.13), is divided in two parts; the mathematical expectation for an unicast (re)transmission packet attempts, weighted by p_u and the mathematical expectation for a broadcast transmission packet attempt, weight by p_b.

$$\begin{aligned}
E[B] &= \overbrace{E_u[B]p_u}^{Unicast} + \overbrace{E_b[B]p_b}^{Broadcast} = \sum_{i=1}^{k} i \Pr\{B = i\} = \\
&= \left(\sum_{i=1}^{k} i p_c^{i-1}(1 - p_c) + k p_c^k \right) p_u + ((1 - p_c) + p_c)\, p_b = \\
&= \left(\sum_{i=1}^{k} i p_c^{i-1}(1 - p_c) + k p_c^k \right) p_u + p_b = \frac{(1 - p_c^k)}{(1 - p_c)} p_u + p_b = \\
&= \frac{(1 - p_c^k)\, p_u + (1 - p_c)p_b}{(1 - p_c)}.
\end{aligned} \tag{4.13}$$

Likewise what happens with $E[B]$, the mathematical expectation for the cycle duration also changes due to the cycle duration difference between both transmission schemes. There are two different expressions for $E[D]$ like in equations (4.10) and (4.11). Below it

[4] See equation (42) in [10] and equations (8) and (9) in [14]

is presented $E^{(1)}[D]$ and $E^{(2)}[D]$.

$$\begin{aligned}
E^{(1)}[D] &= E_u^{(1)}[D]p_u + E_b^{(1)}[D]p_b = \sum_{i=1}^{k} D^{(1)}(i)\Pr\{B=i\} = \\
&= \left(\sum_{i=1}^{k}\left[\left(2^{i-1}W_0 - \frac{W_0-i}{2}\right)p_c^{i-1}(1-p_c)\right] + p_c^k\left(2^{k-1}W_0 - \frac{W_0-k}{2}\right)\right)p_u + \\
&+ \left(2^0 W_0 + \frac{W_0-1}{2}\right)p_b = \\
&= \left(\frac{W_0(1-p_c)\left(1-(2p_c)^k\right) + (1-2p_c)\left(1-p_c^k\right)}{2(1-2p_c)(1-p_c)}\right)p_u + \left(\frac{W_0+1}{2}\right)p_b + \\
&= \frac{\left(W_0(1-p_c)\left(1-(2p_c)^k\right) + (1-2p_c)\left(1-p_c^k\right)\right)p_u + (W_0+1)(1-2p_c)(1-p_c)p_b}{2(1-2p_c)(1-p_c)}.
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
E^{(2)}[D] &= E_u^{(2)}[D]p_u + E_b^{(2)}[D]p_b = \sum_{i=1}^{k} D^{(2)}(i)\Pr\{B=i\} = \\
&= (1-p_c)\left[\sum_{i=1}^{k}\left(\frac{i-W_0}{2}\right)p_c^{i-1} + \sum_{i=1}^{m+1}(2p_c)^{i-1}W_0 + \sum_{i=m+2}^{k}\left(2^{m-1}W_0(i-m+1)\right)p_c^{i-1}\right]p_u + \\
&+ p_c^k\left(2^{m-1}W_0(k-m+1) - \frac{W_0-k}{2}\right)p_u + \left(2^0W_0 + \frac{W_0-1}{2}\right)p_b = \\
&= \left(\frac{(1-2p_c)\left(W_0\left(1-2^m p_c^k\right) + \left(1-p_c^k\right)\right) + p_c^k W_0\left(1-(2p_c)^m\right)}{2(1-2p_c)(1-p_c)}\right)p_u + \left(\frac{W_0+1}{2}\right)p_b = \\
&= \frac{\left((1-2p_c)\left(W_0\left(1-2^m p_c^k\right) + \left(1-p_c^k\right)\right) + p_c^k W_0\left(1-(2p_c)^m\right)\right)p_u + (W_0+1)(1-2p_c)(1-p_c)p_b}{2(1-2p_c)(1-p_c)}.
\end{aligned} \tag{4.15}$$

The expression that describes the (re)transmission probability p_t, is given by (4.2) and it is adapted to both BEB mechanisms discussed so far. The transmission probability is described below:

$$
\begin{aligned}
p_t &= \frac{E[B]}{E[D]} = \\
&= \begin{cases}
\dfrac{\frac{\left(1-p_c^k\right)p_u+(1-p_c)p_b}{(1-p_c)}}{\frac{\left(W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}{2(1-2p_c)(1-p_c)}}, k \le m+1 \\
\dfrac{\frac{\left(1-p_c^k\right)p_u+(1-p_c)p_b}{(1-p_c)}}{\frac{\left((1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c W_0(1-(2p_c)^m)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}{2(1-2p_c)(1-p_c)}}, k > m+1
\end{cases} \\
&= \begin{cases}
\dfrac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left(W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}, k \le m+1 \\
\dfrac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left((1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c W_0(1-(2p_c)^m)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}, k > m+1
\end{cases} .
\end{aligned}
\tag{4.16}
$$

4.4 Conclusions

This section is dedicated to show how equation (4.16) converges to all the models presented previously. This can also be regarded as a theoretical model validation. There are two ways to achieve the broadcast transmission probability. The easier one is to force $p_u = 0$ and the other way is to impose the broadcast parameters. The broadcast parameters rely on $m = 0$ because there are no retransmissions of broadcast packets, $k = 1$ for the same reason and $p_u = 1$.

$$\lim_{\substack{p_b = 0 \\ k = 1 \\ m = 0}} p_t = \lim_{p_u=0} p_t = \begin{cases} \frac{2}{W_0+1}, k \leqslant m+1 \\ \frac{2}{W_0+1}, k > m+1 \end{cases}. \tag{4.17}$$

Considering from now on that $p_b = 0$, equation (4.16) converges to equation (4.12). Bianchi's formula [9] can be obtained when $p_b = 0$ and k is infinite.

$$\begin{aligned} &\lim_{\substack{p_b=0 \\ k\to\infty}} \left[\frac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left((1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c W_0(1-(2p_c)^m)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b} \right] = \\ &= \frac{2(1-2p_c)}{(1-2p_c)(W_0+1)+p_c W_0(1-(2p_c)^m)}. \end{aligned} \tag{4.18}$$

Note that only equation (4.16) with the constraint $k > m + 1$ was needed, since in Bianchi's approach k is infinite. Thus, $m + 1$ is always smaller than k. Finally, to prove the compatibility between the Kwak's work [10] and equation (4.16), it is only necessary to analyze the equation with $k \leq m + 1$ branch. This is the special case where k and m converge to infinity. However in this case k is always equal to $m + 1$.

$$\begin{aligned} &\lim_{\substack{p_{b=0} \\ k\to\infty \\ m\to\infty}} \left[\frac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left(W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b} \right] = \\ &= \frac{2(1-2p_c)}{W_0(1-p_c)+(1-2p_c)}. \end{aligned} \tag{4.19}$$

Notice that UA studied in section (3.2) is the special case of BEB when $m = 0$. It does not really matter the value of transmissions k because as it was explained before, a cycle in this case ends when whether the packet is sent successfully or not.

In [14] it is presented a model that also accounts for the complete BEB retransmission mechanism as in [13]. However, it is not very accurate because there is no backwards compatibility with previous well-known models. The p_t equation for $k \leq 1$ is the same as the one Kwak also deduced in [10] or Andreev in [13]. In the equation of p_t for $k > m$ if the number of retransmissions $k \to \infty$, the equation does not converge to Bianchi's formula. This is observed by the following limit:

$$\lim_{k\to\infty} \frac{2(1-2p_c)(1-p_c^k)}{W(1-(2p_c)^k)(1-p_c)+(1-2p_c)(1-p^{m+1})+W2^{k-1}p_c^k(1-2p_c)(1-p_c^{m-k+1})} = 0. \tag{4.20}$$

Chapter 5

BEB Protocol: Advanced Analysis

This chapter describes several advances to the model presented in chapter 4. The advanced analysis means that the model formulated in section 4.3.2, is extended. Firstly the equally-slotted homogeneous system is extended to heterogeneous scenario. Secondly, the unequally-slotted system for groups of heterogeneous users is formulated. As it is explained later in this chapter, the extension is not generalized to all cases. Finally, this chapter describes a simple way to get a lower bound on the EDCA throughput using the proposed model.

5.1 Network Heterogeneity

In a real network there are different kinds of users with different necessities. These different users might need different settings to access to the channel with different probabilities. To make matters clear, Table 5.1 shows the most relevant parameters of the proposed model and their meaning. The list of changes to the general system assumptions is essentially the same as the ones in section 4.3 with the following exception:

- 1b: It is considered fixed network topology. However this time the network is assumed to have G groups of heterogeneous users like it is depicted in Figure 5.1 (more detailed information later on in this text).

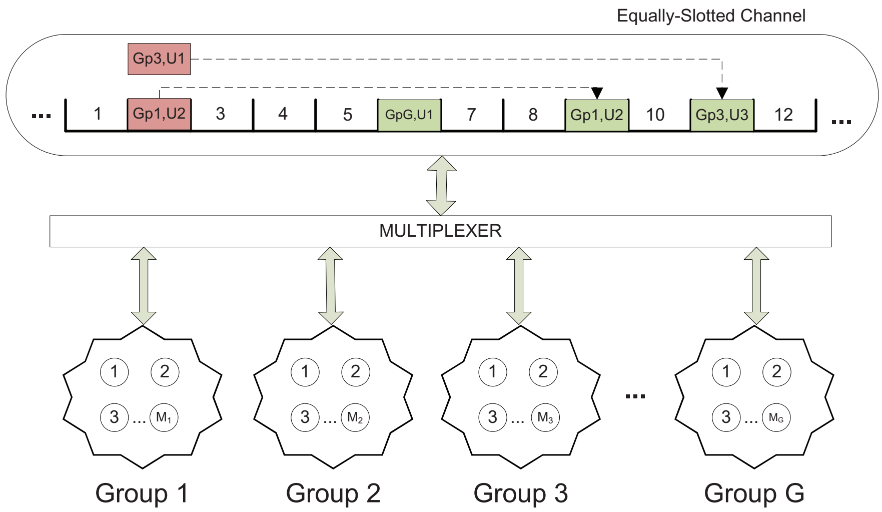


Figure 5.1: General topology for a network with groups of heterogeneous users.

In Figure 5.1, the stars represent the groups $1, 2, \ldots, G$ and, inside each star there are circles that represent the users number $1, 2, \ldots, M_j$, where M_j represents the number of users in the group number j. Moreover, the figure also shows packets collision and packets that are send successfully in the channel. A collision is shown in slot 2, where user 2 of group 1 transmits in the same slot as user 1 of group 3. Successful transmission are shown in slots 6, 9 and 11 where users 1, 2 and 3 of the groups G, 1 and 3, respectively, send a packet in different time slots. This kind of approach is not new and it is introduced in [23]. The multiplexer in this figure has the same meaning as the multiplexer in Figure 2.5.

Parameters	
W_0	This is the initial contention window for BEB protocol.
k	Represents the maximum number of transmissions that a unicast packet can suffer. Remember that a broadcast frame is only sent once. This parameter is an integer and it is finite.
m	The BEB stage is represented by this letter. In other words, m is the maximum number of possible growths for W_0. This parameter is also an integer and it is finite.
p_b	This is the probability in which a user generates a new broadcast packet.
p_u	This is the complementary of p_b probability. Hence, this is the probability of generating a unicast packet.

Table 5.1: Heterogeneous model parameters.

Next, it is assumed that the network is made of G different groups as depicted in Figure 5.1. Heterogeneity is referred to as a network with groups of users that have different access probabilities. These groups are heterogeneous between themselves however, all users belonging to a certain group j are homogeneous between themselves. Inside each group, users share the same parameters. The parameters are shown in Table 5.1.

Each group j has its own transmission probability $p_t^{(j)}$ and consequently observes a different conditional collision probability $\mathrm{p}_c^{(j)}$ in the channel.

$$\mathrm{p}_c^{(j)} = 1 - \left(1 - p_t^{(j)}\right)^{n_j - 1} \prod_{i=1; i \neq j}^{G} \left(1 - p_t^{(i)}\right)^{n_i}. \tag{5.1}$$

The equation (5.1) shows how the conditional collision probability can be obtained for a generic group j. Note that a collision can be intergroup, intragroup or both at the same time. An intergroup is a collision between users belonging to different groups. An intragroup is a collision between users of the same group. The transmission probability $p_t^{(j)}$ is calculated according with the parameters of group j.

$$p_t^{(j)}(W_0 = W_0^{(j)}, m = m^{(j)}, k = k^{(j)}, p_b = p_b^{(j)}, p_c = \mathrm{p}_c^{(j)}) =$$
$$= \begin{cases} \frac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left(W_0(1-p_c)\left(1-(2p_c)^k\right)+(1-2p_c)\left(1-p_c^k\right)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}, k \leq m+1 \\ \frac{2(1-2p_c)\left[\left(1-p_c^k\right)p_u+(1-p_c)p_b\right]}{\left((1-2p_c)\left(W_0\left(1-2^m p_c^k\right)+\left(1-p_c^k\right)\right)+p_c^k W_0(1-(2p_c)^m)\right)p_u+(W_0+1)(1-2p_c)(1-p_c)p_b}, k > m+1 \end{cases} . \tag{5.2}$$

Notice that the equations (5.1) and (5.2) form a system of equations with a numerical solution. This solution for this system of equations is unique for any $G \geq 1$ as Li and Battiti describe in [23].

Once the transmission probability for each group is established, the system wide probabilities can be studied. These probabilities are important because they describe the equally-slotted channel dynamics. In Figure 5.2 it is shown how the probabilities are classified in the channel.

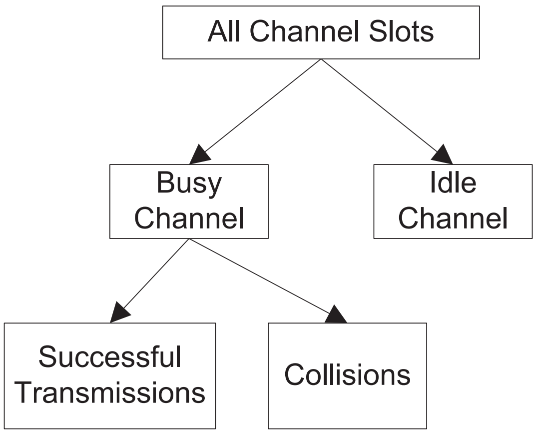


Figure 5.2: Channel probabilities.

The channel can be seen in terms of all the slots that it contains. From all channel slots there is a group of busy time slots with probability P_{busy} and a group of idle time slots with probability P_{idle}. The group of busy time slots can also be divided in the group of time slots where a successful transmission occurred with probability P_S^* (conditional system-wide success probability) and the group of time slots where a collisions occur with probability $1 - P_S^*$. All the probabilities are shown below.

$$P_{idle} = \prod_{j=1}^{G} \left(1 - p_t^{(j)}\right)^{n_j}. \tag{5.3}$$

$$P_{busy} = 1 - P_{idle}. \tag{5.4}$$

$$P_S^* = \frac{1}{P_{busy}} \left(\sum_{j=1}^{G} \left\{ n_j p_t^{(j)} \left(1 - p_t^{(j)}\right)^{n_j - 1} \prod_{i=1, i \neq j}^{G} \left(1 - p_t^{(i)}\right)^{n_i} \right\} \right). \tag{5.5}$$

$$P_c^* = 1 - P_S^*. \tag{5.6}$$

Notice that equations (5.5) and (5.6) are probabilities conditioning on the fact that the channel is busy. In order to obtain the system throughput (see next subsection) it is necessary to obtain the unconditional system wide probabilities. The equation (5.7) represents the unconditional success probability of a generic group j and the equation (5.8) represents the unconditional collision probability of the network.

$$P_{S,j} = n_j p_t^{(j)} \left(1 - p_t^{(j)}\right)^{n_j - 1} \prod_{i=1, i \neq j}^{G} \left(1 - p_t^{(i)}\right)^{n_i}. \tag{5.7}$$

$$P_c = P_{busy} - \sum_{j=1}^{G} P_{S,j}. \tag{5.8}$$

5.2 Unequally-Slotted System

So far, the mathematical analysis assumed that the radio channel was divided by equal slots in time. This assumption is not true when for IEEE 802.11 is considered because the channel is supposed to be divided by slots of different sizes. The unequally-sized time slots technique allows the increase of throughput because it increases the channel utilization. This channel utilization improvement is achieved because idle slots are normally shorter than the remaining slots. Thus, a user spends more time in a data packet transmission slot than in an idle slot. In order to make the proposed model of this project fit into a unequally-slotted system like IEEE 802.11, the slot rescaling is needed. This rescaling technique is done by some authors like in [9] or [23].

This section proposes an extension to the previous section. The unequally-slotted system is shown in Figure 5.3.

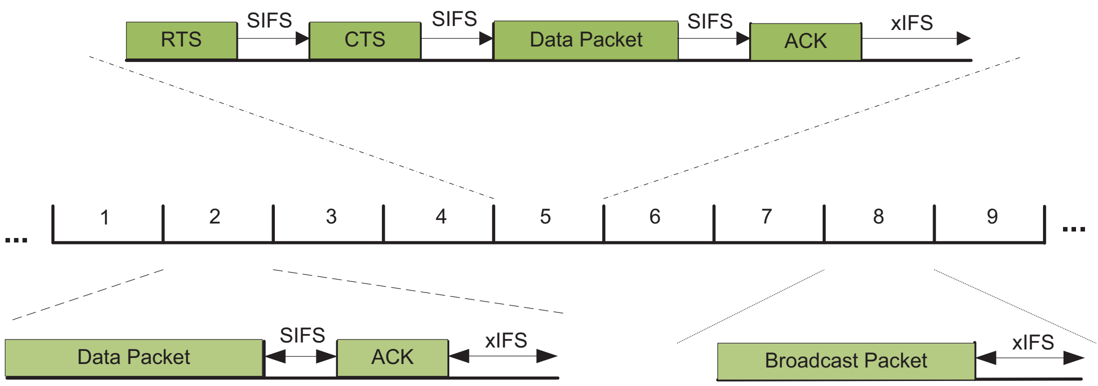


Figure 5.3: Unequally-slotted system in IEEE 802.11-based networks.

As explained before, there are essentially 3 kinds of packet transmissions in IEEE 802.11: The basic access transmission, the RTS/CTS mechanism and the transmission of a broadcast packet. These 3 different transmission techniques are shown in Figure 5.3. The slots number 2, 5 and 8 show the basic access mode, the RTS/CTS mechanism and the broadcast frame transmission respectively. All the remaining slots are considered idle time slots or time slots where a collision occurs. A collision can occur when 2 or more users transmit a packet using either the same transmission mechanism or not. Studying the network throughput is easy when all packets in the network are sent with the same transmission mechanism, transmission rate and, all packets have the same payload length. Otherwise the study of the throughput can be a challenge. Works like [9] and [23] consider a fixed payload packet size for analysis simplicity in the basic access validation, avoiding at the same time the problem related to the time duration of a collision. When different payload sizes of packets are used it is not very clear the amount of time that a collision takes. Moreover, different transmission mechanisms use different schemes and consequently exchange different kinds of packets. Different transmission mechanisms often use different bitrates like what happens with the basic access transmission and a transmission of a broadcast packet. The work presented in [3] solves these challenges for one group of homogeneous users. In a future work it will be interesting to apply the same technique as in [3], in order to solve these challenges to groups for heterogeneous users. With the system-wide probabilities deduced previously, it is possible to define an approximation for the unequally-slotted system network throughput $\widetilde{S}$.

$$\widetilde{S} = \frac{\sum_{j=1}^{G} [P_{S,j} E[P]]}{P_{idle} T_{idle} + \sum_{j=1}^{G} P_{S,j} T_S + P_c T_c}. \tag{5.9}$$

The actual rescaling of slots is done in equation (5.9) by the times T_{idle}, $T_{S,j}$ and T_c. Notice that the simple proposed model can only be applied in a finite number of cases. These cases are described below:

1. When users just transmit unicast and broadcast packets with fixed packet payload length, using the basic access mechanism only;

2. When users only transmit unicast packets with RTS/CTS mechanism and the packets have fixed payload lengths.

3. When users only transmit unicast packets with RTS/CTS mechanism and the packets have different payload lengths.

In order to better explain all the cases it is necessary at this point to introduce the slots timings. These timings were defined by Bi and Battiti [23] as follows:

$$
\begin{aligned}
T_S^{RTS/CTS} &= RTS + CTS + MAC_{hdr} + PHY_{hdr} + ACK + E[P] + \\
&+ DIFS/AIFS + 3SIFS + 4\delta \\
T_c^{RTS/CTS} &= RTS + EIFS + \delta \\
T_c^{bas} &= PHY_{hdr} + MAC_{hdr} + E[\overline{P_c}] + EIFS + \delta \\
T_S^{bas} &= PHY_{hdr} + MAC_{hdr} + E[P] + ACK + SIFS + DIFS/AIFS + 2\delta.
\end{aligned}
\tag{5.10}
$$

In equations (5.10), δ stands for the channel propagation delay, $E[P]$ is the average packet payload and $E[\overline{P_c}]$ is the average payload of the longest packets involved in a collision. Only case 1 can study the influence of broadcast in the network. However, because of the throughput definition, there is always an error associated with broadcast traffic. It is assumed that the timing to transmit a unicast packet is the same as the timing to transmit a broadcast packet. This assumption generally is not true because a broadcast packet transmission timing does not have the time to transmit an ACK. Moreover, this problem becomes even worse when the amount of broadcast traffic increases when compared to the unicast traffic. Notice that the packet length shall always be below the RTS/CTS threshold so that users use always the basic access mechanism.

An approximation for the throughput can also be captured with the proposed model for cases number 2 and 3. These cases only address unicast traffic, this means that $p_b = 0$ in equation (5.2). Moreover, independently of the payload length, it has always to be greater than RTS/CTS threshold to make sure that users transmit with this mechanism. In this case, the model shall be more accurate than in case one because it is well-known the amount of time spend in a collision and the amount of time spend in a successful transmission. Despite this accuracy improvement, all cases have the problem of disregard the users' timeout due to not receiving an ACK packet. Some problems described here allow are also discussed in Bianchi's work [9]. This work presents a technique to allow studying the network throughput using basic access for different payload lengths. In [3], it is also presented solutions for some of the problems discussed above. For simplicity reasons this thesis work is focused only three cases on the above but in future work it would be interesting to develop the mathematical formalism of equation (5.10) for extending the model to the remaining cases.

Maximizing equation (5.9) in order to find the optimal transmission probability is not easy due to its mathematical complexity. However, for one group of users this value is well-known and can be achieved like in [9].

$$\frac{\partial}{\partial p_t}\left[\lim_{G\to 1}\widetilde{S}\right] = \frac{\partial}{\partial p_t}\left[\frac{E[P]}{T_S - T_c + \frac{T_{idle}(1-P_{busy})/P_{busy}+T_c}{P_S^*}}\right]. \tag{5.11}$$

Since $E[P]$, T_S and T_c are constants, $\widetilde{S}$ can be maximized when the following expression yields:

$$\frac{\partial}{\partial p_t}\left[\frac{P_S^*}{(1-P_{busy})/P_{busy}+T_c/T_{idle}}\right] = \frac{\partial}{\partial p_t}\left[\frac{Mp_t(1-p_t)^{M-1}}{T_c^* - (1-p_t)^M(T_c^*-1)}\right] = 0. \tag{5.12}$$

Where $T_c^* = T_c/T_{idle}$ is the collision duration measured in idle time slot units. After some simplifications, Bianchi finds that the optimum transmission probability is approximately:

$$p_t \approx \frac{1}{M\sqrt{T_c^*/2}}. \tag{5.13}$$

Using equation (5.13), and imposing $m = 0$ the optimal BEB window can be found like it is shown below:

$$\frac{2}{W+1} = \frac{1}{M\sqrt{T_c^*/2}} \leftrightarrow W^{opt} = 2M\sqrt{T_c^*/2} - 1. \tag{5.14}$$

Equation (5.14) shows the optimal window version of the unequally-slotted for one group of users. Note that this equation is not very different from equation (3.15). In fact, if one considers that the collision slots are equal to the idle slots and successful slots, equation (5.14) takes the form of equation (3.15).

5.3 EDCA Simplified Model

The model proposed in this chapter can also be applied to calculate a lower bound limit on the EDCA channel throughput. To make matters clear, consider the same assumptions as in the previous section for a special case of four different groups like it is depicted in Figure 5.4.

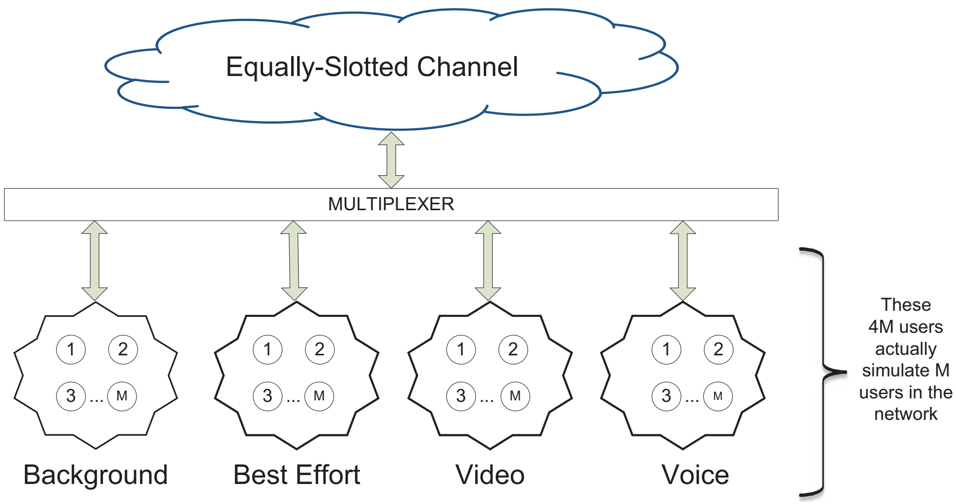


Figure 5.4: Network configuration for EDCA evaluation.

To simulate an EDCA scenario using this simplified model, one shall understand that four users are needed, each one belonging to different AC. This is the reason why in Figure 5.4 each AC has M virtual users. Each one of the 4 users is configured with the corresponding AC parameters (Background, Best Effort, Voice and Video). This idea is illustrated in Figure 5.4. Notice that, it is possible that packet collisions from different ACs of a specific EDCA user occur. This would never happen in a real system because different ACs inside an EDCA user compete for a TXOP. These collisions are solved inside each user when the AC of higher priority receives the TXOP making the packets from low priority AC behave like if a channel collision had occurred. Moreover, packets from different ACs belonging to different EDCA users can also occur. In a real system this would also never happen because of the AIFS times.

Finally all the mathematical formalisms introduced before in this chapter are still valid for the EDCA analysis when users transmit only unicast traffic with basic access or RTS/CTS mechanism. Remember that while using this model for EDCA evaluation, it is needed four different groups with the M number of users each. Each group emulates a different AC such that one EDCA user is in reality four independent users representing one of the different ACs. This approach is not new and can be found in more detail in [30].

Chapter 6

Model Validation

In this chapter we validate the proposed model through simulations. We used the well-known ns-2 simulator [31]. A special traffic generator was developed for ns-2 to generate a certain amount of broadcast/unicast traffic. This agent generates a specified percentage of broadcast traffic as well as unicast traffic. All validations were done using the basic access transmission scheme because the collision time duration between the packets is more critical in this case, as we discuss in section 6.2, allowing us to understand how far the model is from the simulation results. However the same validations can be repeated using RTS/CTS mechanism. The IEEE 802.11b ns-2 implementation was parameterized according to Table 6.1.

ns-2 Settings	
SIFS	10 μs
DIFS	50 μs
EIFS	364 μs
Idle slot time (T_e)	20 μs
ACK frame duration	27.7 μs
ACK timeout (aprox.)	304 μs
Propagation delay	1 μs
Data rate	11.0 Mbps
Frame Size	1500 bytes
Simulation Time	750 s

Table 6.1: Parameters used in all validation results.

In this chapter, only the transmission probability p_t is validated by simulations because it is the main result of chapter 4 and there is a direct relationship between it and the system-wide probabilities presented in section 5.1. Moreover, the throughput expression (equation 5.9) depends indirectly on the transmission probabilities because it relies on system wide probabilities. The practical results are presented by means of the average values with 95% confidence interval (notice that in some cases the confidence intervals are too small to be perceptible).

For all validations, three different scenarios were adopted:

- Homogeneous scenario: One group of users;
- Heterogeneous scenario: Three different groups of homogeneous users;
- An EDCA scenario in which there are four groups of users. Each group represents a different AC.

6.1 Transmission Probability Validation

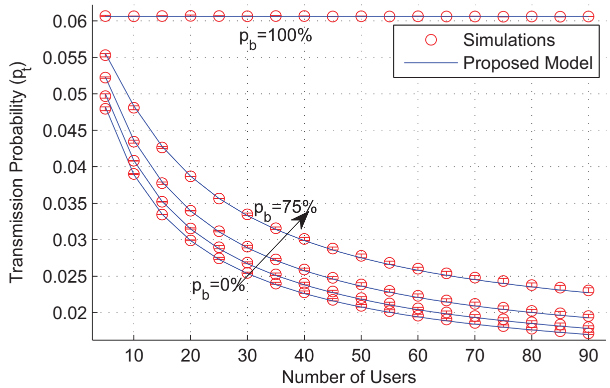
This section is dedicated to the transmission probability validation. For one group of users, two scenarios were simulated. These scenarios are shown in Table 6.2 in order to validate both branches of equation (4.16).

	Case 1	Case 2
W_0	32	32
m	5	3
k	5	5
p_b	0;0.25;0.5;0.75;1	

Table 6.2: Parameters used in ns-2 simulator for one homogeneous group p_t validation.

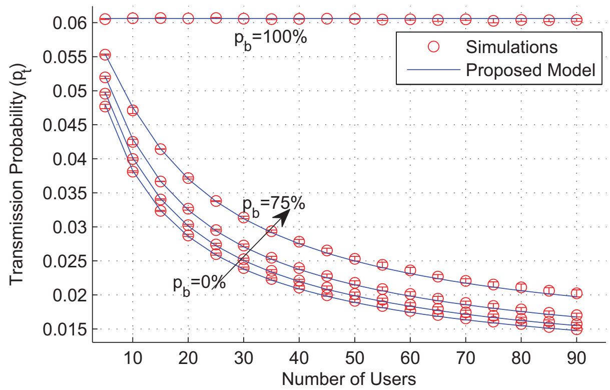
The case 1 and case 2 basically give the parameters to evaluate the first and the second branch of p_t, respectively. Figure 6.1 depicts the theoretical values of p_t against the simulation results.

The model predictions are accurate because the transmission probability is directly related with the initial contention window W_0, the number of BEB stages, the number of transmissions per packet, the number of users and the amount of broadcast traffic, as observed in equation (5.2). As the number of users in the network increases, the collision probability also increases and consequently the p_t decreases due to BEB CW management mechanism. In other words, a user transmits according to the number of times that its CW



(a) Case 1 validation.

Figure 6.1: Transmission probability simulations results.



(b) Case 2 validation.

Figure 6.1: Transmission probability simulations results.

was doubled and the number of attempts that a packet suffered. Note that this validation was made by collecting the instantaneous BEB counter values used for each transmission attempt. An average instantaneous BEB counter values was then computed using all the previous collected values from all the transmission attempts. Finally, p_t was computed using the average instantaneous BEB counter. By this way, the results are remarkably accurate. Note that in these simulations, the channel timings were not accounted for. Table 6.3 shows the parameters used for three heterogeneous groups. The first group only transmits unicast packets, the second group transmits 50% of broadcast traffic and the third group only transmits broadcast packets.

It is not easy to choose the topology for the heterogeneous scenario because there are multiple possible cases. Hence, we decided to validate the transmission probability

	Group 1	Group 2	Group 3
W_0	16	32	64
m	4	4	1
k	6	3	2
p_b	0	0.5	1

Table 6.3: Parameters used in ns-2 simulator for p_t validation, considering a heterogeneous scenario with three groups of homogeneous users.

for the heterogeneous case in the scenarios shown in Table 6.3, however the reader shall understand that any combination of the used parameters is possible, including different number of users in each group. The results of the simulations are shown in the Table 6.4.

M	p_t	Group 1	Group 2	Group 3
5	T	0.050724	0.043752	0.030769
	P	0.050706 ($\pm$ 2.54e-4)	0.04367 ($\pm$ 2.18e-4)	0.030846 ($\pm$ 1.54e-4)
10	T	0.031406	0.038367	0.030769
	P	0.031316 ($\pm$ 1.57e-4)	0.03847 ($\pm$ 1.92e-4)	0.030707 ($\pm$ 1.54e-4)
15	T	0.024285	0.035593	0.030769
	P	0.024136 ($\pm$ 1.21e-4)	0.035763 ($\pm$ 1.79e-4)	0.030729 ($\pm$ 1.54e-4)
20	T	0.02087	0.033937	0.030769
	P	0.020854 ($\pm$ 1.04e-4)	0.033976 ($\pm$ 1.70e-4)	0.030774 ($\pm$ 1.54e-4)

Table 6.4: Validations results for p_t considering a heterogeneous scenario with three groups of homogeneous users.

In Table 6.4, M stands for the number of users that each group has ($3 \times M =$ total number of users in the network). In this table, T and P stand for the Theoretical and Practical results, respectively. The theoretical results match very well with the simulation results like it is shown by the relative error. Using the value next to the practical result (in brackets), one can easily calculate the interval of confidence of 95%. In the heterogeneous case, the group 1 and the group 2 theoretical p_t is calculated with first and second branches of equation (4.16), respectively. For the third group theoretical p_t, it does not really matter each branch to use because all packets are sent according to equation (3.14). Notice that when only broadcast traffic is being used p_t takes the same value regardless of the number of users in the network. The reason for this effect is because when only broadcast traffic is used the BEB behaves like $m = 0$ and $k \to \infty$. Consequently, p_t remains the same as far as W_0 is not changed.

In order to validate a simplified EDCA scenario, four groups of heterogeneous users were considered. Each group represents one traffic class and its access parameters are shown in Table 6.5.

Parameter	Video	Voice	Background	Best Effort
W_0	8	16	16	32
m	1	1	6	5
k	4	4	7	6
p_b	0			

Table 6.5: BEB parameters used for the simplified EDCA approach.

The transmission probabilities for the simplified EDCA evaluation are simply a special case of four groups of heterogeneous users and the validation can be found in the Table 6.6.

Users	p_t	Voice	Video	Background	Best Effort
2	T	0.1650	0.0842	0.0402	0.0221
	P	0.1665 ($\pm$ 8.33e-4)	0.0840 ($\pm$ 4.20e-4)	0.0396 ($\pm$ 1.98e-4)	0.0221 ($\pm$ 1.11e-4)
4	T	0.1492	0.0767	0.0186	0.0125
	P	0.1496 ($\pm$ 7.48e-4)	0.0767 ($\pm$ 3.84e-4)	0.0178 ($\pm$ 8.90e-5)	0.0127 ($\pm$ 6.35e-5)
6	T	0.1423	0.0732	0.0123	0.0092
	P	0.1425 ($\pm$ 7.13e-4)	0.0738 ($\pm$ 3.69e-4)	0.0120 ($\pm$ 6.00e-5)	0.0097 ($\pm$ 4.85e-5)
8	T	0.1387	0.0716	0.0096	0.0078
	P	0.1386 ($\pm$ 6.93e-4)	0.0719 ($\pm$ 3.60e-4)	0.0100 ($\pm$ 5.00e-5)	0.0078 ($\pm$ 3.90e-5)
10	T	0.1366	0.0706	0.0085	0.0070
	P	0.1364 ($\pm$ 6.82e-4)	0.0707 ($\pm$ 3.54e-4)	0.0086 ($\pm$ 4.30e-5)	0.0071 ($\pm$ 3.55e-5)

Table 6.6: Validations results for p_t considering EDCA simplified scenario.

The transmission probabilities of Table 6.6 show that the Voice AC has the highest value, as expected. The Video traffic has the second greatest transmission probability, followed by Best Effort and Background. The relative error is small for all the ACs. The next section will discuss some of the model properties in more detail in order to understand possible reasons for the mismatch between the practical and the theoretical results.

6.2 Discussion

This section addresses the main reasons why the proposed model might present deviation from real systems. Remember that we propose a model for saturated traffic conditions.

Assumptions validity. Bordenave et al. [11] proved mathematically in 2005 that the constant steady-state conditional collision probability assumption holds, for networks with large number of users. In 2007, the heterogeneous Malone et al. [25] model assumptions validity, under saturation conditions, based on Bianchi's work [9] is discussed in [32]. This work shows that the model overestimates the collision probabilities by a few percent. Later in 2010, Malone et al. discuss in [33] the validity of IEEE 802.11 MAC modeling assumptions. The authors use real wireless cards, which implement the IEEE 802.11b/g protocol, to test the model assumptions from homogeneous Bianchi's work [9]. In [33], the authors conclude by observing the real measurements that for saturated traffic conditions, the decoupling assumption in which the steady-state conditional collision probability is constant holds even for a small number of users. Moreover, the authors prove that the sequence of transmission attempts consists of independent and identically distributed (i.i.d.) variable, which do not depend on past collision history. To be more specific, the conditional collision probability is similar for all BEB stages even for networks with a small number of users. These conclusions help to explain why works like [9], [10] or [25] are so precise. The work [33] also concludes that as the assumptions of these models are reasonable and, they shall be able to make accurate predictions regarding detailed QoS metrics.

Since we are operating in saturated traffic conditions the constant steady-state conditional collision probability assumption of our proposed model also holds. For the special case of heterogeneity, this assumption is still valid like it was concluded in [32], where the heterogeneous work [25] was tested against real network measurements. Note that [25] presents a heterogeneous model based on Bianchi's work. As proved before, Bianchi's model is a special case of our model, as such we conclude that our steady-state conditional collision probability assumption is right. Note that equations (5.1) (p_c) and (5.2) (p_t) form an equation system with a unique solution. All the probabilities discussed in the previous chapters, including throughput analysis, depend directly or indirectly on p_t. The transmission probability obtained by the proposed model was proved by simulations to be quite accurate even for the case of different amounts of broadcast traffic. Hence, this means that the proposed model shall predict with an acceptable error, the channel probabilities. Regarding the channel throughput, some problems might arise and they are explained later in this text.

Ignoring BEB counter freezing process. One of the problems that contribute to model mismatch when compared with a real scenario is the lack of BEB counter freezing process. Wang et al. also explain this process in [21]. Figure 6.2 depicts the model assumptions, and compares it with a ns-2 based user.

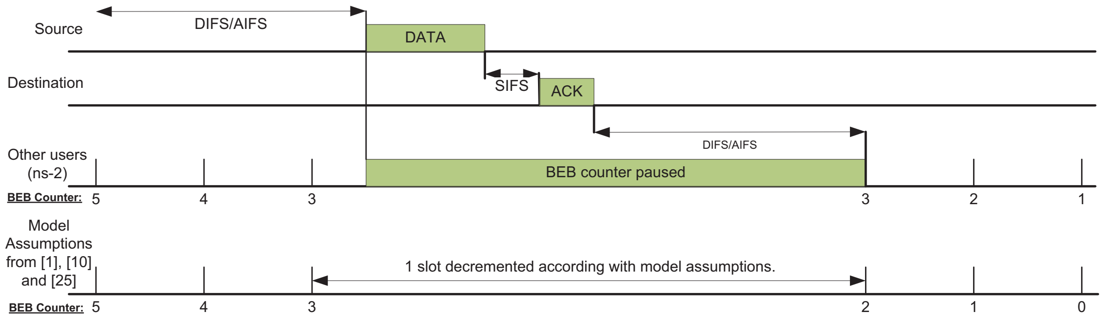


Figure 6.2: BEB freezing counter assumptions.

Observing Figure 6.2 it is much clearer why the lack of freezing the BEB counter will result in a model mismatch. In this example, both users (according with the model and ns-2) choose an instantaneous BEB counter equal to 5. When the medium is busy, ns-2 users pause the BEB counter due to the PHY carrier sense mechanism. However, the model assumptions treat the busy slot as an idle slot, decrementing the BEB counter. This increases the model system wide collision probability as more users are transmitting. Figure 6.3 depicts how the model assumptions can lead to a higher collision probability. Note that in Figure 6.3[1], due to the lack of freezing the BEB counter, a user transmits in a different slot when compared with the ns-2 user. If the busy periods occur in the same slots like is shown in this figure, there will be a packet collision due to the transmission in different slot. When broadcast traffic is considered this effect is worse because the BEB window is maintained constant at every transmission attempt and consequently p_t is higher than in a unicast scenario, making the users access the channel more often.

As the amount of broadcast traffic increases in the network, the medium becomes busier because under saturated traffic conditions a user has always a packet to send and it never doubles the window. This means that in ns-2, a user would freeze the counter several times and consequently this would decrease p_t. Thus, it is expected that the model accuracy diminishes as the broadcast traffic increases in the network.

[1]Remember that each one of the slots presented in this figure can represent a unicast packet transmission using basic access, a unicast packet transmission using RTS/CTS or a broadcast packet transmission like is shown in Figure 5.3.

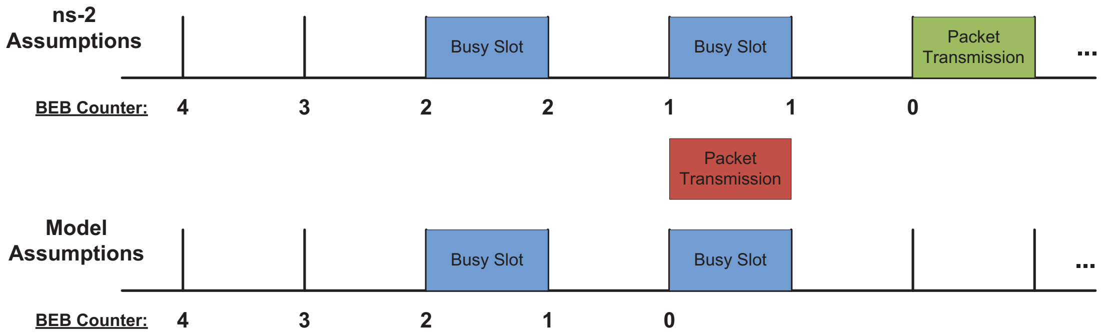


Figure 6.3: Behaviour of BEB counter in model and simulations assumptions.

The reason why the transmission probability of the previous section matches so well is because the freezing process was not accounted for the statistics collection. The collected statistics was obtained through the average window generated at the beginning of the BEB process. Note that the average window generated in the beginning of the BEB process does not have any information about the number of times that the BEB counter is frozen. Hence, the model gives accurate predictions of p_t because it also does not account for the PHY carrier sense mechanism.

Influence of timings on probabilities and throughput. The channel timings might also be responsible for some model mismatch. The main reason for this mismatch is the difference between the transmission schemes used for unicast and broadcast packet transmissions. A unicast packet transmission always relies on an ACK frame in order to increase this type of traffic reliability. However, a broadcast packet is less reliable due to the lack of the recipients' feedback. This makes sense since broadcast packets would have many candidates for acknowledging the transmission and consequently, many collisions could arise in this process.

There is a desynchronization between users due to the different channel timings. Let us consider a network with unicast traffic only. When a collision occurs, the users that transmitted the packets will wait for an ACK timeout (304 μs) while all the remaining users wait an EIFS time (364 μs), according with Table 6.1. There is a difference of 60 μs (3 idle slots) that will cause the senders to desynchronize. This means that the users involved in the collision will start to decrement the BEB counter for the next (re)transmission sooner than the users that sensed the collision but were not involved in the communication. Notice that the worst case of this problem is when only broadcast traffic is considered. When a collision occurs, all users that sent the broadcast traffic will wait for a DIFS/AIFS and then start to decrement the BEB counter for sending the next packet. However, the remaining users sensed the collision and will defer an EIFS before starting to decrement the BEB counter. A DIFS is 7.28 shorter than an EIFS (15.7 idle slots of difference). If the BEB initial window W_0 is equal to 32, this timings difference means approximately half of the window. Thus, it is expected a model mismatch due to this difference in timings. Moreover, the mismatch shall increase as the broadcast traffic in the network increases because the initial BEB CW is never doubled. Figure 6.4 gives an example of a collision between a unicast and a broadcast packet of equal length.

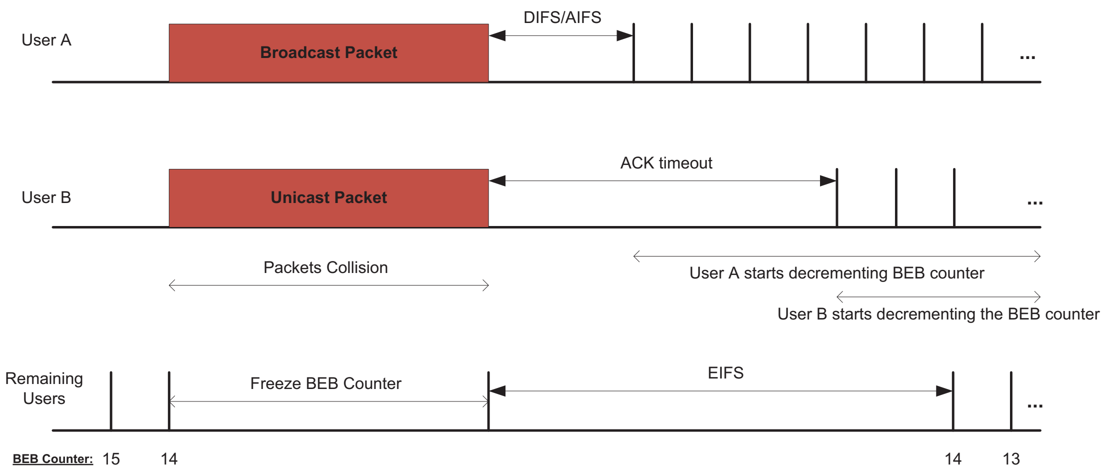


Figure 6.4: Definition of T_c using the basic access scheme for fixed-size packets.

Section 5.2 presents three cases for which the model is able to predict the throughput. To make matters clearer, the system throughput can be calculated through equation (5.9). With this equation we can only obtain an approximation because of the way the timings T_s and T_c are defined. Both unicast and broadcast packets are described in the same way in equation (5.10) for the basic access. This is not true because the broadcast success time does not include the time needed to send the ACK plus the SIFS. Hence, we expect that as the amount of broadcast traffic increases in the channel the accuracy decreases. Another problem is that T_c is defined as being the duration that a user observes the collision plus the EIFS time. Figure 6.4 makes it clearer since T_c is described as "Remaining users" see the collision.

For future work it would be interesting to come up with the individual success probabilities for broadcast and unicast traffic, such as it would be possible to have different timing definitions for both types of traffic. When only unicast traffic is considered by means of RTS/CTS mechanism, T_c is simpler because the packets involved in the collision have always the same length. For this case the timings problem would be the difference between the ACK timeout and the EIFS (3 idle slots).

Summary. Most of the models presented in section 2.2 demonstrate the same problems. However, despite all these problems, the homogeneous works [9] or the heterogeneous model from Malone et al. [25] predict with little error the unicast network throughput under certain assumptions. Oliveira et al. [3] present a good solution to deal with channel timings under the assumption of co-existence of both unicast and broadcast traffic that shall be adopted for the proposed model in a future work.

Chapter 7

Conclusions

The current chapter discusses the thesis' conclusions. Section 7.1 is a small synthesis where the main contents of each chapter is highlighted and section 7.2 sumarizes the main conclusions. Section 7.3 enumerates a few topics to address by future investigation.

7.1 Synthesis

This section briefly describes each chapter's content. In Chapter 1, it was explained that the thesis scope is the performance evaluation in heterogeneous networks considering a mixture of traffic. It also highlights the importance of the QoS differentiation in nowadays networks.

Chapter 2 discusses the IEEE 802.11 protocol and some related works in network models that address BEB mechanism, mixture of traffic and groups of heterogeneous users in Ad-Hoc networks based on IEEE 802.11. Moreover, we present the general system assumptions that are used in the following chapters.

Chapter 3 overviews ALOHA protocols. This is an important chapter because ALOHA is a special case of BEB which makes it easier to understand the basic concepts of the RMA models described in this work.

The relationship between ALOHA protocol and BEB is discussed in Chapter 4. Here, the regeneration cycle concept is applied to several proposed BEB models. The first part of the proposed model is deduced for a homogeneous network using equally-slotted system. The unequally-slotted system and heterogeneity are addressed in Chapter 5 by applying

the first part of the proposed model deduced in Chapter 4.
Chapter 6 validates the model's transmission probability. It also discusses the validity of some model assumption and it presents the main reasons for possible mismatch between practical and theoretical results.

7.2 Conclusions

This thesis accomplished a deep study of random multiple access protocols under saturated traffic conditions. It was discussed the first multiple access protocol (ALOHA) and its relations to BEB. The study of ALOHA protocol was important in several aspects. Firstly, it helped to understand why the random protocols based on contention windows are used instead of geometric-based protocols. Secondly, it introduced the tradeoff between throughput and delay (the higher is the throughput, the higher is the delay). This tradeoff is important to warn the reader that focusing only on throughput optimization can have severe consequences on delay. Finally, it was proved that uniform ALOHA is a special case of BEB for $m = 0$.
The regeneration cycle approach was successfully applied to previous well-known models under saturation conditions. It was also successfully applied to the proposed model where BEB is described by means of an initial contention window, a maximum number of BEB stages, a maximum limit of packet transmissions and the probability with which a broadcast packet is generated. This model can be applied to study both equally-slotted systems and unequally-slotted systems. The model has also the following characteristics:

- Support for different groups of users from different QoS classes;
- Support for different amounts of broadcast traffic for different groups of users;
- Rescaling of slots duration to allow unequally-slotted system analysis;
- Simplified EDCA analysis capabilities;

The model was validated theoretically by proving the backwards compatibility to previous well-known models. The ns-2 simulations proved that the model predicts very accurately the transmission probability for both the homogeneous and the heterogeneous scenarios. The model assumptions hold under saturation conditions and it was indentified the main reasons for possible mismatch between theoretical and practical results: The lack of BEB counter and the frame timings. The lack of freezing the BEB counter during busy periods makes the theoretical p_t higher than the practical one. Consequently a user transmits more often and the theoretical system wide collision probability increases. The frame timings influence the channel probabilities. The reason is that due to the difference in IFSs timings and ACK timeout, users "see" the channel differently. The consequence is that it influences the channel probabilities, especially with the increase of the amount of broadcast traffic because of the lack of acknowledgment. This gives higher channel access probability to the user that transmitted a broadcast packet. The throughput equation shall give a lower bound for the throughput because the collision and success time duration are given as the worst possible scenario in the network. Moreover, the definition of the collision and success time is defined accounting for the unicast packets only. Thus, as the broadcast traffic increases in the network it is expected that the throughput expression presents a higher margin of error.

The maximum channel departure for ALOHA-based protocols in an equally-slotted system is approximately 36% (see section 3.1.1) and, because of this small value, the slots rescaling was done to increase the throughput. However, it does not increase linearly with the increase in the transmission rate. The latest advances show that, the data rates are being increased (600 Mbps for IEEE 802.11n) and the data packets are sent faster in the network. However, the idle slots do not decrease in time as fast as the data rate increases due to physical issues. So, the high speed networks are coming towards the equally-slotted systems. This means that by increasing the data rate, it does not necessarily increase the channel utilization.

7.3 Future Work

Future work on this model shall address in first place the freezing BEB process when the medium is sensed idle. The success and collision probability for unicast and broadcast traffic shall be studied separately in order to increase the model's accuracy. Some simulations shall be done in order to evaluate the model precision. The model can be extended to unsaturated traffic conditions like it is done in [3]. With this extension, the model would be able to predict the channel average throughput and delay for groups of heterogeneous users accounting for the co-existence of broadcast and unicast traffic.

Bibliography

[1] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE standard 802.11, 1999 edition (revised 2003)."

[2] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE standard 802.11, 2007."

[3] R. Oliveira, L. Bernardo, and P. Pinto, "The influence of broadcast traffic on IEEE 802.11 DCF networks," *Elsevier*, vol. Computer Comunications 32, pp. 439–452, 2009.

[4] N. Abramson, "The ALOHA system - another alternative for computer communications," *Fall Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 37, pp. 281–285, 1970.

[5] N. Gaarder, "ARPANET satellite system," *AEPA Network Inform. Center, Stanford Res.Inst., Menlo Park, Calif.*, vol. Ass Note 3 (NIC 11285), 1972.

[6] L. Roberts, "Dynamic allocation of satellite capacity through packet reservation," *Nat. Comput. Conf., AFIPS Conf Proc.*, vol. 42, pp. 703–710, 1973.

[7] L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE transactions on communications*, vol. com-23, pp. 410–416, 1975.

[8] S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE transactions on communications*, vol. com-23, pp. 891–904, 1975.

[9] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.

[10] B.-J. Kwak, N. Song, and L. E. Miller, "Performance analysis of exponential backoff," *IEEE transactions on networking*, vol. 13, pp. 343–355, 2005.

[11] C. Bordenave, D. McDonald, and A. Proutière, "Random multi-access algorithms - a mean field analysis," *Rapport de Recherche*, vol. 5632, pp. 1–12, 2005.

[12] K. Medepalli and F. A. Tobagi, "Throughput analysis of IEEE 802.11 wireless LANs using an average cycle time approach," *IEEE Globecom*, pp. 3007–3011, 2005.

[13] S. Andreev and A. Turlikov, "Binary exponential backoff algorithm analysis in the lossy system with frames," *In the Proc. of the XII International Symposium on Problems of Redundancy in Information and Control Systems*, pp. 201–210, 2009.

[14] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement," *Proceedings of IEEE INFOCOM*, vol. Vol. 2, pp. 599–607, 2002.

[15] R. Oliveira, L. Bernardo, and P. Pinto, "Performance analysis of the IEEE 802.11 distributed coordination function with unicast and broadcast traffic," *The 17th Annual IEEE International Symposium*, vol. on Personal Indoor and Mobile Radio Communications, pp. 1–5, 2006.

[16] X. Ma and X. Chen, "Saturation performance of IEEE 802.11 broadcast networks," *IEEE Communications Letters*, vol. 11, pp. 686–688, 2007.

[17] X. Chen, H. H. Refai, and X. Ma, "Saturation performance of IEEE 802.11 broadcast scheme in ad hoc wireless LANs," *Vehicular Technology Conference*, vol. VTC-2007 Fall. IEEE 66th, pp. 1897–1901, 2007.

[18] X. Ma and X. Chen, "Performance analysis of IEEE 802.11 broadcast scheme in ad hoc wireless LANs," *IEEE Transactions on Vehicular Technology Conference Technology*, vol. 57, N. 6, pp. 3757–3768, 2008.

[19] Z. Wang and M. Hassan, "Analytical evaluation of the 802.11 wireless broadcast under saturated conditions.," *School of Computer Science and Engineering, University of New South Wales.*, vol. Tech. Rep. UNSW-CSETR-0801, pp. 1–21, 2008.

[20] R. Oliveira, L. Bernardo, and P. Pinto, "Modelling delay on IEEE 802.11 MAC protocol for unicast and broadcast nonsaturated traffic," *Wireless Communications and Networking Conference*, pp. 463–467, 2007.

[21] J. C.-P. Wang, D. R. Franklin, M. Abolhasan, and F. Safaei, "Characterising the behaviour of IEEE 802.11 broadcast transmissions in ad hoc wireless LANs," *Communications, 2009. ICC '09. IEEE International Conference on*, pp. 1–5, 2009.

[22] J. C.-P. Wang, D. R. Franklin, M. Abolhasan, and F. Safaei, "Characterising the interactions between unicast and broadcast in IEEE 802.11 ad hoc networks," *Telecommunication Networks and Applications Conference*, pp. 180–185, 2008.

[23] B. Li and R. Battiti, "Performance analysis of an enhanced IEEE 802.11 distributed coordination function supporting service diferentiation," *Springer-Verlag Berlin Heidelberg*, vol. LNCS 2811, pp. 152–161, 2003.

[24] B. Bellalta, M. Oliver, M. Meo, and M. Gerrero, "A simple model of the IEEE 802.11 MAC protocol with heterogeneous traffic flows," *IEEE Region 8 Eurocon.*, 2005.

[25] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordenation function in non-saturated heterogeneous conditions," *IEEE/ACM transactions on networks*, vol. 15, No.1, pp. 159–172, 2007.

[26] D. Bertsekas and R. Gallager, *Data Networks.* Prentice-Hall, 1987.

[27] R. Rom and M. Sidi, *Multiple Access Protocols Performance and analysis.* Springer-Verlag, 1989.

[28] C. H. Foh, *Performance Analysis and Enhancement of MAC Protocols.* PhD thesis, The University of Melbourne, 2002.

[29] J. Little, "A proof for the queueing formula: $l = \lambda \cdot w$," *Operations Research*, vol. 4, Issue 3, pp. 383–387, 1961.

[30] G. Sharma, A. Ganesh, P. Key, and R. Needham, "Performance analysis of contention based medium access control protocols," *25th IEEE International Conference on Computer Communications. Proceedings*, vol. 0743-166X, pp. 1 – 12, 2006.

[31] ns 2, "Network simulator 2," *see http://www.isi.edu/nsnam/ns/*, 2009.

[32] D. Malone, I. Dangerfield, and D. Leith, "Verification of common 802.11 MAC model assumptions," *Proceedings of the Passive and active network measurement 8th international conference*, pp. 1–10, 2007.

[33] K. D. Huang, K. R. Duffy, and D. Malone, "On the validity of IEEE 802.11 MAC modeling hypotheses," *Networking, IEEE/ACM Transactions*, vol. PP Issue:99, pp. 1–14, 2010.