AN ANALYTICAL MODEL FOR THE SERVICE DELAY DISTRIBUTION OF IEEE 802.11E ENHANCED DISTRIBUTED CHANNEL ACCESS

by

Jeffrey William Todd Robinson

B.A.Sc., Simon Fraser University, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
Summer 2005

APPROVAL

Name: Jeffrey William Todd Robinson

Degree: Masters of Applied Science

Title of thesis : An Analytical Model for the Service Delay Distribution of IEEE 802.11e Enhanced Distributed Channel Access

Examining Committee: Dr. A. H. Rawicz, Chairman

Dr. R. H. S. Hardy
Professor, Engineering Science, SFU
Senior Supervisor

Dr. T. S. Randhawa
Adjunct Professor, Engineering Science, SFU
Supervisor

Dr. J. Liang
Assistant Professor, Engineering Science, SFU
Examiner

Date Approved: July 29, 2005

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

Abstract

This thesis presents a new model for analytically determining the cumulative distribution function for service delay of the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) function under saturation traffic conditions. The model accommodates all service differentiation mechanisms of the EDCA function and is scalable to typical IEEE 802.11e network parameters. A two-dimensional discrete-time Markov chain is used to model the EDCA backoff process and serves as the basis for the signal flow graph which is used to generate the cumulative distribution function for service delay. The model is validated by a comparison with a simulation model developed by a member of the IEEE 802.11e Task Group.

Acknowledgements

I would like to thank first my parents, who have a hand in my every success. I wish to acknowledge Dr. Hardy, who with great wisdom has guided my research efforts. This thesis is a tribute to his steady encouragement and thoughtful counsel. I wish to thank next Dr. Randhawa, who first suggested that I return to Simon Fraser University to take the Masters degree. Whenever my progress stalled, Dr. Randhawa was ready with a technical insight or a prescient piece of advice that invariably moved my work forward. Finally I thank Tara, whose support and love I rely on constantly.

Contents

List of Tables

List of Figures

List of Symbols

$\bar{p}$	Average Conditional Collision Probability
n	Number of different Access Categories in a network
$\vec{N}$	n-tuple describing the number of CAF from each Access Category contending in a network
$b(t)$	Stochastic process representing the backoff time counter for a given CAF
$s(t)$	Stochastic process representing the number of times a CAF has transmitted its head of queue frame
m	Backoff stage at which $\mathrm{CW}_{\mathrm{max}}$ is reached
$m+f$	Maximum backoff stage
W_i	Contention Window size at backoff stage i
$b_{i,k}$	Stationary distribution of the $\{i,k\}$ backoff state
τ	Uniform probability of transmission
$\vec{q}$	Tagged station from whose point of view conditional probabilities are conceived
$\mathbb{P}$	Transition probability matrix for contention vectors
$p_{\mathrm{set}}(\vec{y}\|\vec{x},h)$	Probability of transmission for a set of CAFs $\vec{y}$ conditioned on being in contention zone h and contention vector $\vec{x}$ prevailing
$p_{\mathrm{tx}}(\vec{y}\|\vec{x},h)$	Probability that $\vec{q}$ transmits in tandem with a set of CAFs $\vec{y}$, conditioned on being in contention zone h and contention vector $\vec{x}$ prevailing

$p_{\text{obs}}(\vec{y}\|\vec{x}, h)$	Probability that $\vec{q}$ observes a transmission from a set of CAFs $\vec{y}$, conditioned on being in contention zone h and contention vector $\vec{x}$ prevailing
$p_{\text{No tx}}(h\|\vec{x})$	Probability that no transmission occurs on a slot edge in contention zone h, conditioned on contention vector $\vec{x}$ prevailing
$p_{\text{rz}}(h\|\vec{x})$	Probability that a given contention period reaches contention zone h, conditioned on contention vector $\vec{x}$ prevailing
$P_{\text{tx}}(\vec{y}, h\|\vec{x})$	Probability that a transmission event including $\vec{q}$ and $\vec{q}$ precipitates the end of a contention period in contention zone h, conditioned on contention vector $\vec{x}$ prevailing
$P_{\text{obs}}(\vec{y}, h\|\vec{x})$	Probability that a transmission event $\vec{y}$, not including $\vec{q}$, precipitates the end of a contention period in contention zone h, conditioned on contention vector $\vec{x}$ prevailing
$P_{\text{tx}}(\vec{y}\|\vec{x})$	Probability that a transmission event including $\vec{q}$ and $\vec{y}$ ends a contention period where contention vector $\vec{x}$ prevails
$P_{\text{obs}}(\vec{y}\|\vec{x})$	Probability that a transmission event $\vec{y}$, not including $\vec{q}$ ends a contention period where contention vector $\vec{x}$ prevails
$P(\text{timeout}\|\vec{x})$	Probability that the expiry of ACKTimeout/CTSTimeout ends a contention period where contention vector $\vec{x}$ prevails
$p_{\text{No tx post}}(h\|\vec{x})$	Probability than no transmission occurs on a slot edge in a contention zone h during a post-collision period, conditioned on contention vector $\vec{zp}$ prevailing
$p_{\text{rzp}}(h\|\vec{x})$	Probability that a given post-collision contention period reaches contention zone h, conditioned on contention vector $\vec{zp}$ prevailing
$P_{\text{post}}(\vec{yp}, h\|\vec{zp})$	Probability that a transmission event $\vec{yp}$ precipitates the end of a post-collision contention vector in contention zone h, conditioned on contention vector $\vec{zp}$ prevailing

$P_{\text{post}}(\vec{yp}\|\vec{zp})$	Probability that a transmission event $\vec{yp}$ ends a post-collision contention period where contention vector $\vec{zp}$ prevails
$P(\text{post timeout}\|\vec{zp})$	Probability that the expiry of ACKTimeout/CTSTimeout ends a post-collision contention period where contention vector $\vec{zp}$ prevails
π	Stationary distribution of contention vectors
$p(i\|h)$	Probability that a transmission from $\vec{q}$ experiences a collision when contention vector $\vec{x} \rightarrow i$ prevails, conditioned on the transmission occurring in contention zone h
$b_{\text{slot}}(i\|\vec{x})$	Stationary distribution of the i^{th} backoff slot in a contention period, conditioned on contention vector $\vec{x}$ prevailing
$b_{\text{zone}}(h\|\vec{x})$	Stationary distribution of the contention zone h in a contention period, conditioned on contention vector $\vec{x}$ prevailing
$p_{\text{slot}}(i,h)$	Probability that any slot on which $\vec{q}$ transmits is characterized by the contention vector $\vec{x} \rightarrow i$ and contention zone h
$p_{\text{event}}(\vec{y})$	Probability that the transmission event $\vec{y}$ occurs on any given slot edge
T_k^{suc}	Duration of a successful transmission from a CAF belonging to Access Category k
T_k^{col}	Duration of a collision in which the longest transmission is from a CAF belonging to Access Category k
σ	Duration of a backoff slot

List of Abbreviations

AIFS	Arbitrary Inter-Frame Space
AC	Access Category
BSS	Basic Service Set
CAF	Channel Access Function
CDF	Cumulative Distribution Function
CPO	Contention Period Outcome
CTS	Clear-To-Send
CW	Contention Window
DCF	Distributed Control Function
DIFS	DCF Inter-Frame Space
DSSS	Direct Sequence Spread Spectrum
EDCA	Enhanced Distributed Channel Access
EDCF	Enhanced Distributed Control Function
EIFS	Extended Inter-Frame Space
FCS	Frame Check Sequence
FHSS	Frequency Hopping Spread Spectrum
HCF	Hybrid Control Function
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ISO	International Standards Organization
MAC	Medium Access Control
MSDU	MAC Service Data Unit
OSI	Open Systems Interconnect
PC	Point Controller
PCF	Point Control Function

PLCP	Physical Layer Convergence Protocol
QSTA	EDCA station
RTS	Request-To-Send
SIFS	Short Inter-Frame Space
SDL	System Description Language
TGe	Task Group E
TXOP	Transmission Opportunity

Chapter 1

Introduction

Modeling of networking protocols has long been a focus of communications networks researchers. Presently, research into modeling the IEEE 802.11 wireless networking protocol is especially active. We attribute this activity to a number of factors: general research interest in all things wireless, close resemblance of the 802.11 protocol to the well understood IEEE 802.3 protocol, and the strictly limited bandwidth of the IEEE 802.11 physical medium.

This last point deserves special mention. As devices grow smaller and more powerful, there is a general consensus that bandwidth demands on wireless networks will increase. We are already seeing a push to migrate the transmission of multimedia content to the wireless medium. When bandwidth hungry, delay sensitive media applications were first introduced to wired networks, the obvious and trivial solution was to supply more bandwidth as required. But in a wireless medium, bandwidth is not easily added. Strict limits on frequency use and constraints on power consumption mean that efficient protocols, and a clear understanding of these protocols, are crucial to the provision of multimedia applications.

To the end of providing an efficient protocol for transmission of multimedia over the wireless medium, the IEEE 802.11 Working Group created Task Group E (TGe). Tasked with supplementing the original IEEE 802.11 MAC layer standard (IEEE 802.11 Working Group 1999) to support differentiated Quality of Service, TGe produced a redesigned centralized control mechanism (Hybrid Coordination Function – HCF) and a modified version of the distributed control mechanism (Enhanced Distributed Channel Access – EDCA). The IEEE 802.11e supplementary standard (IEEE

802.11 Task Group E 2005) is still under development; at its July 20, 2005 plenary meeting TGe intends to resolve comments from the last sponsor ballot, generate a new version of the draft and work towards a submission to the IEEE Standards Board Review Committee.

Modeling research on IEEE 802.11e has been underway since the development of the supplementary standard began. During the time in which this research has been conducted, a number of models for EDCA have appeared in the literature. These models have focused on finding results for collision probability, mean throughput or mean service delay. But considering the multimedia uses for which 802.11e and EDCA are intended, the probability distribution of service delay is a more compelling topic for analysis.

1.1 Thesis Goals

In this thesis we develop an accurate model for predicting the service delay distribution of IEEE 802.11e EDCA under saturation traffic conditions. An accurate delay performance model is important to an understanding of the operation of 802.11e differentiation mechanisms and is useful tool for numerous applications, among them network performance optimization and admission control. It is also important because it necessarily requires pioneering new methods for analyzing contention-based protocols where contending stations operate on different backoff timescales.

Our effort is defined by the following three goals:

1. The model captures all relevant quality of service differentiation features of IEEE 802.11e EDCA.

2. The model scales to typical IEEE 802.11e parameters.

3. The model is validated by a simulation model that is both correct and was developed independent of our analytical efforts.

1.2 Summary of Contributions

The work presented in this thesis is the first to:

- Identify two major errors prevalent in nearly all IEEE 802.11 and IEEE 802.11e simulation and modeling research.
- Develop a two dimensional discrete time Markov chain model for IEEE 802.11e EDCA backoff that recognizes that IEEE 802.11e EDCA Arbitrary Inter-Frame Spaces create *contention zones* among which the probability of collision is different.
- Develop a framework for modeling real time delay based on discrete time Markov chain backoff models that is capable of modeling typical IEEE 802.11e networks where stations using different backoff timescales contend for the same channel.
- Validate an analytical model for IEEE 802.11e EDCA against simulation results from an implementation designed by an IEEE 802.11 TGe member (Moreton 2005).

1.3 Thesis Outline

Chapter 2 presents an overview of the IEEE 802.11 MAC layer Distributed Coordination Function protocol and the IEEE 802.11e EDCA enhancement. Chapter 3 describes two common errors in IEEE 802.11 and IEEE 802.11e simulation and modeling research, and surveys the state of the art in IEEE 802.11 and IEEE 802.11e modeling research. Chapter 4 develops a two dimensional discrete time Markov chain model (DTMC) for the backoff behaviour of IEEE 802.11e channel access functions. Chapter 5 extends the backoff model to analytically determine the cumulative distribution function for EDCA transmission delay. The model is then validated against results from a simulation model prepared by a member of IEEE 802.11 Task Group E in chapter 6. The final chapter of the thesis reviews the contributions of the work and suggests directions for future research.

Chapter 2

IEEE 802.11 and IEEE 802.11e MAC Layer Overview

2.1 Legacy IEEE 802.11 MAC Layer Overview

The IEEE 802.11 standard defines physical and logical mechanisms to facilitate data communication over a shared wireless medium. To minimize the cost of IEEE 802.11 devices, the physical layer specifies a half-duplex transceiver; that is a transceiver that can transmit and receive, but may not do both simultaneously. Since multiple stations may seek to access the shared medium concurrently a Medium Access Control (MAC) protocol is necessary to regulate contention and verify transmission success. The original IEEE 802.11 MAC protocol specifies two mechanisms for allocating channel access: Point Control Function (PCF) and Distributed Control Function (DCF).

The Point Control Function uses a station designated as the Point Controller (PC) to provide deterministic, centralized control over the physical medium for stations belonging to a given Basic Service Set (BSS). The PC specifies a division of channel time into intervals for contention free and contention based channel access. During contention free intervals, stations may not transmit unless explicitly granted permission via polling by the PC. During contention based intervals, stations vie for access to the channel according to the DCF. Though PCF to support real-time services it has not been widely supported primarily due to complexity and inefficient operation.

As opposed to a single station rationing access to the channel, DCF distributes

control of the channel through a set of listen and wait procedures observed by every station. DCF is based on the Carrier Sense Multiple Access / Collision Avoidance protocol, which in turn has its roots in earlier contention based networking protocols like Appletalk, wired ethernet (IEEE 802.3) and slotted ALOHA. The following subsection presents an overview of the IEEE 802.11 DCF.

2.1.1 IEEE 802.11 DCF

Outbound data arriving to the DCF is queued for transmission as MAC Service Data Units (MSDUs). A station with no queued MSDUs may transmit an arriving MSDU after observing an idle channel medium for *DCF Inter-Frame Space* (DIFS). If an MSDU arrives while the channel is busy, the station waits until the end of the current transmission and then initiates the backoff procedure.

The backoff procedure begins with the station selecting a random backoff time from a range of discrete values known as the *Contention Window.* The contention window is the range $[0, CW]$, where CW is an integer within the range of the physical layer (PHY) specific attributes aCWmin and aCWmax. Backoff time is tracked by a backoff counter that is decremented at the end of each idle backoff slot. Backoff slots occur after an idle DIFS following a correctly received frame, or after an idle *Extended Inter-Frame Space* (EIFS) following detection of a frame that was received with an incorrect MAC Frame Check Sequence (FCS). Should the station detect activity on the channel during DIFS, EIFS or a backoff slot, the backoff procedure is suspended until the medium is next idle for DIFS or EIFS.

Whenever its backoff counter reaches zero, the station initiates transmission of the queued MSDU, using one of two transmission mechanisms. In the basic transmission mechanism, the MSDU is sent immediately. Figure 2.1 illustrates the DCF backoff procedure and basic transmission sequence.

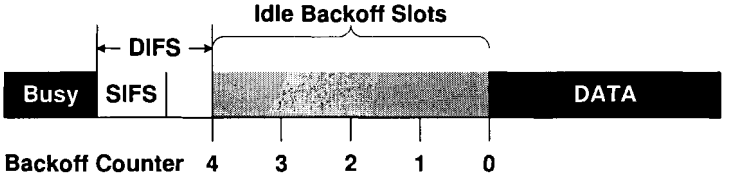


Figure 2.1: DCF backoff procedure, with basic transmission.

The alternative transmission mechanism, named Request-To-Send/Clear-To-Send (RTS/CTS), uses a handshaking mechanism whereby the sending station sends a short RTS control frame and waits to receive a CTS control frame from the destination before transmitting the MSDU. The RTS/CTS mechanism is a solution to the hidden node problem (Kleinrock and Tobagi 1975) but also has the effect of reducing bandwidth losses to transmission failure when MSDU frames are much larger than RTS frames. Figure 2.2 illustrates the DCF backoff procedure and RTS/CTS transmission sequence.

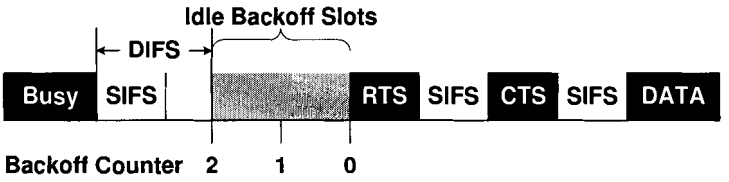


Figure 2.2: DCF backoff procedure, with RTS/CTS transmission.

Stations may also *fragment* MSDUs into a series of smaller transmissions to reduce both the probability of unsuccessful transmission and the penalty of retransmitting a large frame. When fragmentation is used, a station must contend for the channel when transmitting the first fragment; subsequent fragments are transmitted a Short IFS (SIFS) after the end of the ACK control frame sent by the destination to acknowledge the successful reception of the previous fragment.

Stations infer the success of unicast transmissions from the reception of acknowledgment (ACK) frames sent by the destination. Immediately after receiving an ACK, stations begin a new backoff procedure regardless of whether another MSDU is queued for transmission. This *post backoff* policy prevents stations from monopolizing access to the medium.

If a station does not start receiving an ACK in response to a unicast frame within ACKTimeout after its transmission completes, it infers that the transmission has failed and increments a retry counter for the MSDU. If the retry counter exceeds the retry limit the station will discard the MSDU, otherwise the station will restart the backoff procedure and attempt to retransmit the MSDU. In order to reduce congestion in heavily loaded networks, the size of the contention window from which the random backoff is drawn is doubled after each transmission failure, up to aCWMax.

It is important to note that while waiting to receive a CTS or ACK response, a station is effectively suspended: it does not draw or decrement a backoff counter and may not transmit. This behaviour has led to significant confusion in the simulation and modeling community, and is discussed in greater detail in section 3.2.2.

2.2 IEEE 802.11e MAC Layer Overview

The IEEE 802.11 Working Group created Task Group E to devise a supplementary standard for adding Quality of Service capabilities to the wildly popular IEEE 802.11 wireless networking standard. As part of this work, TGe has replaced PCF and DCF with new channel coordination mechanism called Hybrid Coordination Function (HCF). HCF apportions access to the channel by granting transmission opportunities (TXOPs) to stations. A station granted a TXOPs may use the medium for a particular time interval defined by a start time and maximum duration contained in the TXOP. A station acquire TXOPs through one of two access mechanisms specified by HCF: HCF controlled channel access (HCCA) and Enhanced Distributed Channel Access (EDCA).

2.2.1 HCF Controlled Channel Access

HCCA relies on a central hybrid coordinator (HC) to provision medium access. Because a single HC manages network activity, it can control contention on the medium and thereby increase efficiency. Like PCF, HCCA divides channel time into periods, now called the contention period (CP) and the controlled access period (CAP).

The CAP is the heart of HCCA, and consists of stations using TXOPs allocated by polling. The HC schedules these TXOPs by applying priority rules to the requests it receives from stations during polling. In contrast to PCF, polling may occur in the CP as well as the CAP, and polling data can be piggy-backed onto data and control frames.

2.2.2 Enhanced Distributed Channel Access

EDCA provides the distributed mechanism for channel access under HCF. It can be used as the only channel access mechanism or can be used during contention periods

allocated by a HC. EDCA was designed by TGe to be backwards compatible with legacy IEEE 802.11 devices, and so shares the same approach to channel access as DCF. In fact, the two protocols are so similar that EDCA is best described in terms of its differences from DCF.

EDCA stations (QSTAs) support up to four queues for incoming traffic. Each queue is associated with a *Channel Access Function* CAF, that contends for channel access using the EDCA parameters of its assigned *access category* (AC). Inside a QSTA, CAFs contend for the channel independent of each other; because of this our discussion of EDCA revolves around CAFs, not stations. When two or more CAFs within an QSTA realize TXOPs simultaneous, a *virtual collision* is said to occur. A virtual collision is resolved by permitting the higher priority CAF to transmit and forcing the lower priority CAF to perform a transmission failure response.

Different levels of service are provided to each AC through a combination of three service differentiation mechanisms:

- Arbitrary Inter-Frame Spaces (AIFS).
- contention window sizes, and
- TXOP duration limits

Arbitrary inter-frame spacing

In contrast to DCF where all backoff counters start decrementing DIFS after the end of the last indicated busy medium, EDCA backoff begins at different intervals according to which AC is assigned to the CAF. The duration of the inter-frame space (AIFSD[AC]) is given by:

$$\text{AIFSD[AC]} = \text{AIFS[AC]} \times \text{aSlotTime} + \text{aSIFSTime} \tag{2.1}$$

where AIFS[AC], aSlotTime and aSIFSTime are management information base (MIB) attributes. Figure 2.3 illustrates the use of AIFS for two CAFs with different ACs. AIFS furnishes higher priority CAFs better service in two ways. First, higher priority stations progress through backoff slots relatively faster since they may decrement their backoff counters while lower priority CAFs wait for the end of a longer AIFS. Consider the case where two transmissions from CAFs with different AIFS fail simultaneously

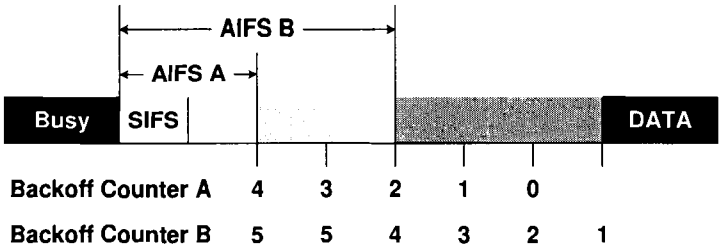


Figure 2.3: EDCA backoff after successful transmission.

and by chance choose identical backoff counter values: the CAFs with the smaller AIFS will transmit sooner since its backoff counter will be decremented more often in subsequent transmission periods.

Second, higher priority CAFs enjoy a lower average probability of transmission failure. In a network free of external interference, the only reason for transmission failure is *collision*: simultaneous, and therefore interfering, transmissions by competing CAFs. Since higher priority CAFs may transmit in backoff slots that lower priority CAFs still waiting in AIFS can not, higher priority CAFs have opportunities to transmit when there is less contention. This means that transmissions from higher priority CAFs are, on average, less likely to collide. Fewer collisions means that, on average, fewer backoff slots are traversed per successful transmission.

In this work we refer to the groups of slots where different sets of CAFs contend for access to the medium as *contention zones*. Referring again to Figure 2.3, the lighter backoff slots are the first contention zone where only CAFs from AC A contend and the darker backoff slots are the second contention zone where CAFs from both AC A and AC B contend.

Contention window sizes

In legacy DCF initial values for backoff counters are randomly selected from the interval [0,CW] where CW is a function of the PHY-specific aCWmin and aCWmax attributes. In contrast, EDCA contention windows are AC specific functions of aCWmin[AC] and aCWmax[AC] attributes. Higher priority CAFs receive superior service by having smaller aCWmin and aCWmax. Smaller window sizes correspond to fewer backoff slots being traversed per transmission, on average. This is clear

if one considers the expected values of two random variables drawn from uniform distributions of ranges $[0, X]$ and $[0, Y]$: if $X < Y$ then $E_X < E_Y$.

TXOP duration limits

EDCA places limits on medium occupancy using an AC specific transmission opportunity (TXOP) limit parameter, in contrast to a common limit for all stations in DCF. Upon gaining access to the medium, CAFs set a medium occupancy timer to the AC specific TXOP limit attribute. A CAF may continue to access the medium so long as the medium occupancy timer is greater than zero. This allows *TXOP bursting*, a procedure where a CAF sends several MSDUs without contending for the channel between transmissions. CAFs with longer TXOP limits have to contend for medium access less often than CAFs with similar traffic arrival rates but shorter TXOP limits. Less frequent contention means lower transmission overhead and fewer collisions per unit of payload, and thus superior service.

2.2.3 Backoff Decrement and Transmission Procedure

The operation of backoff counters in EDCA is significantly different from DCF. Where DCF backoff counters decrement at the *end* of idle backoff slots, EDCA backoff counters decrement on *backoff slot boundaries* (alternatively, *slot edges*). In DCF backoff slots occur DIFS after the medium becomes idle and CAFs transmit whenever their backoff counters reach zero. In EDCA the first backoff slot boundary corresponds to the end of AIFS, and CAFs initiate transmission on the first slot boundary *after* the backoff counter has reached zero. Figure 2.4 illustrates the difference between the two decrement and transmission procedures. Though this change is subtle, the effect is that EDCA backoff is more consistent, and easier to model.

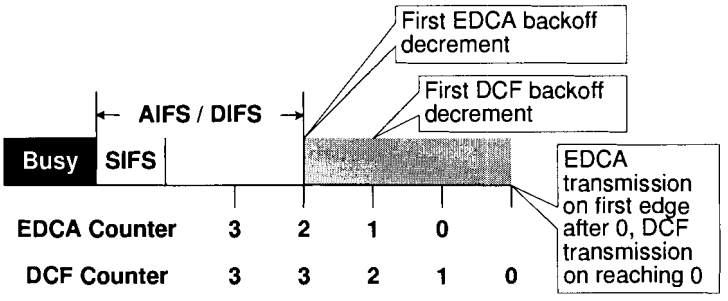


Figure 2.4: Contrast of backoff decrement and transmission conditions in DCF and EDCA

Chapter 3

Review of State of the Art

This literature review concentrates on efforts to develop analytical models for the performance of IEEE 802.11 DCF and 802.11e EDCA/EDCF under saturation conditions. The review starts by defining the commonly analyzed performance measures, and continues with a description of two common simulator problems that have plagued validation efforts and hindered modeling research. After that introduction the survey is divided into three main parts: 802.11 DCF models, 802.11e EDCA/EDCF models and models that analyze the probability distribution of service delay for either IEEE 802.11 DCF or IEEE 802.11e EDCA/EDCF.

3.1 Common Performance Measures

Modeling of network protocols is undertaken with mind to gain a better understanding of how a protocol operates and how best to utilize a network based on a particular protocol. When constructing a model it is important to focus on features that provide meaningful measures of network performance. What follows is a list of definitions for common performance measures of interest in the study of contention based networking protocols, like IEEE 802.11.

Probability of Collision is an important measure because it speaks to a protocol's ability to use the medium efficiently. Under very high loads it is expected that the probability of collision will increase.

Saturation Throughput is the measure of payload data transmitted successfully

per unit of time in a network where all stations have dispatch traffic queued. The condition of all stations having data to send is known as the saturation condition or asymptotic condition. Saturation throughput is useful measure of how well a network utilizes the channel; for protocols offering service differentiation (QoS) it also indicates how well one class of traffic is served relative to others.

Service Delay is the measure the time required for a frame to be received at its destination after it arrives at the head of the transmission queue. Service delay is alternately referred to as transmission delay. Service delay statistics, in particular the probability distribution of service delay, are useful because many applications, like voice and video over IP, are sensitive to transmission delays. Service delay statistics can help to determine whether a protocol is appropriate for a particular application and whether a network is likely to be able to support a given set of applications.

3.2 Collisions and Confusion

An understanding of how collisions occur is essential when modeling IEEE 802.11 DCF and EDCA. The reality is that every published work modeling IEEE 802.11 DCF or EDCA has an inadequate understanding of how collisions occur. The problem is compounded by the fact that the most popular simulator for IEEE 802.11 DCF and EDCA research implements collision behaviour incorrectly.

3.2.1 An IEEE 802.11 Crash Course

The IEEE 802.11 standard specifies behaviour at the Physical (PHY) and Medium Access Control (MAC) layers of the ISO OSI 7-layer network model (Zimmerman 1980). The PHY specification lays out how IEEE 802.11 stations transmit and receive radio signals, and how these signals decode into events and data that are indicated to the MAC layer. The MAC specification describes how IEEE 802.11 stations react to events and data from the PHY layer below and the Link-Layer above.

There are two PHY indications relevant to the IEEE 802.11 MAC response to colliding transmissions.

PHY-CCA.indication(STATE) Whenever a PHY receiver senses the status of transmission medium changing from channel idle to channel busy or from channel busy to channel idle, it sends an PHY-CCA.indication to the MAC indicating the current state of the medium. The STATE parameter can be one of two values: BUSY or IDLE.

PHY-RXSTART.indication Every frame transmission in IEEE 802.11 begins with a Physical Layer Convergence Protocol (PLCP) training sequence broadcast on a shared frequency. Receiving stations use the training sequence to determine the subsequent codes (DSSS or FHSS) required to lock-on to the following frame. A PHY receiver generates a PHY-RXSTART.indication when it has successfully validated the PLCP Header error check CRC at the start of a new PLCP PDU.

When two or more stations begin to transmit simultaneously, receiving PHYs signal PHY-CCA.indication(BUSY) to their respective MACs. Upon receiving the PHY-CCA.indication(BUSY), the MACs suspend backoff. Because the PLCP header transmissions mutually interfere, no receiving PHY validates the PLCP Header error check CRC, and no MAC receives a PHY-RXSTART.indication.

At the end of the colliding transmissions the medium becomes idle and the receiving PHYs signal PHY-CCA.indication(IDLE) to their respective MACs. At this point the MACs will resume the backoff procedure, and will decrement their backoff counters after DIFS or AIFS.

The colliding stations however, do not resume backoff immediately after transmitting. Instead they wait for a positive acknowledgement that their transmission has been successful. If a PHY-RXSTART.indication is not received within ACKTimeout (or CTSTimeout) after the end of transmission, the colliding stations infer transmission failure and resume backoff.

3.2.2 ACKTimeout and CTSTimeout

ACKTimeout and CTSTimeout values are not specified in the text of the IEEE 802.11 standard. They are however specified as SIFS + ACKLength $\times$ 8 and SIFS + CTSLength $\times$ 8 in the system description language (SDL) descriptions contained in the appendix to the standard. This detail has escaped designers of both simulation

and analytical models. Some ignore ACKTimeout altogether, while others assign an arbitrary value to ACKTimeout.

The problem with incorrectly specifying ACKTimeout is that it affects the set of stations that contend for channel access after a collision. When colliding stations wait for ACKTimeout, non-colliding stations may backoff and transmit. Thus these non-colliding stations are furnished opportunities to transmit where the probability of collision is lower vis-a-vis full contention. Figure 3.1 below illustrates this effect. When

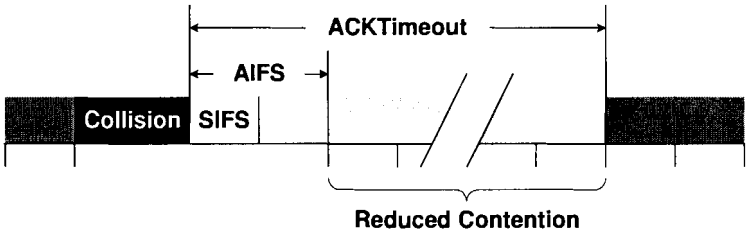


Figure 3.1: Reduced contention for non-colliding stations during ACKTimeout

ACKTimeout is effectively ignored, the modeled probability of collision is greater than the actual probability of collision because stations observing the timeouts are inactive and do not transmit.

3.2.3 Extended Inter-Frame Space

In the simulation and modeling communities there is confusion about the use of the Extended Inter-Frame Space (EIFS) in IEEE 802.11. This confusion stems partly from the description of EIFS in the IEEE 802.11 standard, but also from the laziness of MAC layer researchers. The IEEE 802.11 standard prescribes the use of EIFS as follows:

> 9.2.4.3 Extended IFS (EIFS) The EIFS shall be used by the DCF whenever the PHY has indicated to the MAC that a frame transmission was begun that did not result in the correct reception of the a complete MAC frame with a correct FCS value.

Some IEEE 802.11 simulation model designers have interpreted this to mean that any time a sent frame is not received correctly, the MAC receiving MACs defer for EIFS before resuming backoff. The reason that this interpretation is not correct is that the PHY section of the IEEE 802.11 defines the conditions for indicating that "a frame transmission was begun":

> 12.3.5.11 PHY-RXSTART.indication
> 12.3.5.11.1 Function
> This primitive is an indication by the PHY sublayer to the local MAC entity that the PLCP [Physical Layer Convergence Protocol] has received a valid start frame delimiter (SFD) and PLCP Header.
> ...
> 12.3.5.11.3 When generated
> This primitive is generated by the local PHY entity to the MAC sublayer whenever the PHY has successfully validated the PLCP Header error check CRC at the start of a new PLCP PDU.

In the case of a collision between two simultaneous transmissions, no station correctly receives a PLCP header and no MAC is ever alerted that a transmission has begun. Thus, after transmission failures caused by collision, receiving stations should defer for DIFS.

When EIFS and ACKTimeout are misinterpreted, colliding stations (who do not hear their own transmissions) will start to decrement their backoff counters and possibly transmit after DIFS while all other stations sit idle for EIFS. This is a benefit to the colliding stations, since any transmission during the EIFS period is less likely to collide. Figure 3.2 illustrates the effect of the combined misinterpretation of EIFS and ACKTimeout.

Because all DCF stations are identical, the benefit that this misinterpretation confers on transmitting stations is uniform and the mistake does not have a major impact on DCF simulation results. But the effects when simulating EDCA, especially when contention windows are on the order of EIFS, are severe. Because colliding CAFs may decrement their backoff counters and transmit while all other CAFs sit idle, higher priority CAFs benefit disproportionately since they transmit, and hence

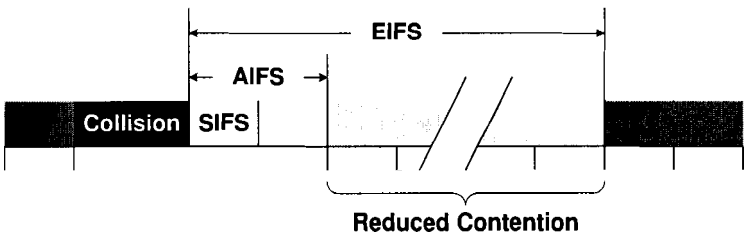


Figure 3.2: Incorrect interpretation: Reduced contention for colliding stations during EIFS

collide, more often and typically draw backoff counter values from smaller contention windows.

Small contention windows multiply the benefit of the reduced competition of the post-collision period. A CAF involved in a collision will be the next to transmit if it selects a backoff counter value that expires before the end of the EIFS and before the backoff counter values selected by the other colliding CAFs. CAFs choosing from smaller contention windows, especially those on the order of EIFS, have the best chances of successful transmitting during post-collision contention periods.

3.2.4 A Review of IEEE 802.11 and IEEE 802.11e EDCA/EDCF Simulation Models

In the IEEE 802.11 modeling literature the two primary simulation tools used to validate models are the open source ns-2 simulator (McCanne and Floyd 2005) and the OPNet simulator (Opnet Technologies Inc. 2005). Our experience with each of these has shown that the IEEE 802.11 DCF models fix ACKTimeout/CTSTimeout as (SIFS + aSlotTime) and misinterpret the operation of EIFS.

There are three publicly available ns-2 models for 802.11e EDCA/EDCF. The first to be released was prepared by TGe members Aman Singla and Greg Chesson (Singla and Chesson 2005). This model did not correct the underlying flaws in the ns-2 DCF implementation, and so suffers from the compound ACKTimeout/EIFS error.

The second 802.11e EDCA/EDCF model to be released was prepared by Mike

Moreton of STMicroelectronics (Moreton 2005). This model corrects both the ACK-Timeout and EIFS errors, and more closely models the interface between the PHY and MAC layers. We have investigated this model thoroughly and are convinced of its correctness.

The last 802.11e EDCA/EDCF model to be released was prepared by the Telecommunication Networks Group from the Technical University of Berlin (TKN TU Berlin 2005). Though this model corrects some minor errors from the ns-2 DCF implementation, it does not fix the ACKTimeout or the EIFS errors.

At the time of writing, no OPNet model for 802.11e EDCA/EDCF has been published.

An unfortunate consequence of the confusion over ACKTimeout and EIFS is that nearly all published analytical models have been validated with custom made simulators. Because readers cannot audit the custom simulation models it is difficult to have faith that the analytical model presented is accurate.

Another unfortunate consequence of this confusion is that modeling researchers looking for clarification on this issue may be persuaded to adopt the incorrect interpretation used by the simulators. This is true of our early work on this thesis, published as (Robinson and Randhawa 2004b) and (Robinson and Randhawa 2004a), for which we relied on the simulator prepared by TGe members Singla and Chesson. To date we know of no analytical model that treats ACKTimeout behaviour as it is defined in the IEEE 802.11 standard.

3.3 IEEE 802.11 DCF Modeling Research

The first effort to analyze IEEE 802.11 was based on an early draft of the protocol when unslotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was among the proposed methods for asynchronous data transfer (Chhaya and Gupta 1997). This work focused on geometric properties of IEEE 802.11 networks and the relative merits of the basic and RTS/CTS transmission mechanism. Because the model was actually of CSMA/CA over wireless links, it did not treat DCF as it came to be defined.

The first model to analyze the IEEE 802.11 DCF as it came to be defined in the standard was presented in (Cali, Conti, and Gregori 1998). This model estimates

protocol capacity under saturation conditions by evaluating the ratio between average message length and the average time the channel is occupied in transmitting a frame. Rather than model the details of the DCF backoff procedure, the authors analyze a p-persistent variant of DCF, that schedules transmission attempts using a persistence factor p (i.e. the backoff interval is sampled from a geometric distribution) rather than using binary exponential backoff. The persistence factor p is derived from the DCF contention window and it is shown that the p-persistent variant of DCF is roughly equivalent to the binary exponential backoff defined in the standard.

In (Kim and Hou 2003), the authors develop a method for estimating throughput that rests on treating the time between transmission attempts as random variable with an exponential distribution. Using Laplace transforms of of probability density functions for transmission, collision and success probability they develop expressions for the average time required to successfully transmit a frame. They apply their analysis to develop a new scheme for frame scheduling. In (Kim and Hou 2004) the authors repeat the analysis of (Kim and Hou 2003) and use the result as a basis for a simulation model that abstracts away from packet level events to queue level events for faster performance.

Bianchi (2000) presents simple and accurate model for the throughput performance of IEEE 802.11 DCF. The Bianchi model is based on a two dimensional discrete time Markov chain that closely follows the DCF backoff process. Due to its intuitive design and simplicity this model has spawned a significant body of literature devoted to its improvement and extension. It is worthwhile to explain this model in detail to provide context for the survey of subsequent modeling efforts.

The Bianchi model assumes an ideal channel, a finite number of stations in the BSS, and saturation traffic conditions. The model accounts for the details of the exponential backoff protocol and both basic and RTS/CTS access mechanisms. The model uses two discrete time Markov processes to track the backoff and transmission attempts of a single station. One process, $b(t)$, represents the value of the station's backoff counter. At the start of every idle backoff slot $b(t)$ is decremented; when $b(t)$ reaches zero the station initiates a transmission. Upon completion of the transmission $b(t)$ is assigned a new backoff value, regardless of whether the transmission was successful or not.

Because the IEEE 802.11 protocol uses a binary exponential backoff, the size of the

contention window from which the value of $b(t)$ is drawn after collision depends on the transmission history of the station. Thus a second Markov process, $s(t)$, is required to track the station's transmission history. The integer values of $s(t)$ are the number of times that the current MSDU has been transmitted. Because the value of $s(t)$ defines the contention window size, the value of $s(t)$ is referred to as the *backoff stage*. After each successful transmission $s(t)$ is reset to the initial backoff stage, and $b(t)$ is drawn from the initial contention window. After a failed transmission, $s(t)$ is incremented and $b(t)$ is drawn from the contention window corresponding to the updated $s(t)$. Once $s(t)$ reaches the stage m corresponding to the maximum contention window size, it persists at that same stage until the station transmits successfully.

It is important to note that the discrete timescale for backoff processes, $b(t)$ and $s(t)$, does not correspond to real time. The backoff processes change state only on the edges of idle transmission slots, and are suspended for the duration of transmissions, collisions and interframe spaces. Thus, the time between successive backoff counter decrements may be much longer than the duration of a single backoff slot (σ), since it may include frame transmission.

The combination of the two processes, $\{s(t), b(t)\}$, is governed by the following non-null one step transition probabilities

$$\begin{cases} P\{i,k|i,k+1\} = 1 & k \in [0, W_i - 2] \quad i \in [0,m] \\ P\{0,k|i,0\} = \dfrac{(1-p)}{W_0} & k \in [0, W_0 - 1] \quad i \in [0,m] \\ P\{i,k|i-1,0\} = \dfrac{p}{W_i} & k \in [0, W_i - 1] \quad i \in [1,m] \\ P\{m,k|m,0\} = \dfrac{p}{W_m} & k \in [0, W_m - 1]. \end{cases}$$

In the above equations, p is the probability of a transmission experiencing a collision and W_i is the contention window size in backoff stage i, defined in terms of the initial contention window W_0. These governing probabilities and their interaction with the backoff processes are neatly summarized in Figure 3.3, below.

A key approximation implicit in the one-step transition probabilities of the Markov chain is a constant probability of collision for all transmissions, regardless of transmission history. This approximation stems from the treatment of station transmission probability. Let $b_{i,k}$ denote the stationary distribution of the bi-dimensional Markov chain $\{s(t), b(t)\}$. Since stations transmit whenever $b(t)$ equals zero, and backoffs

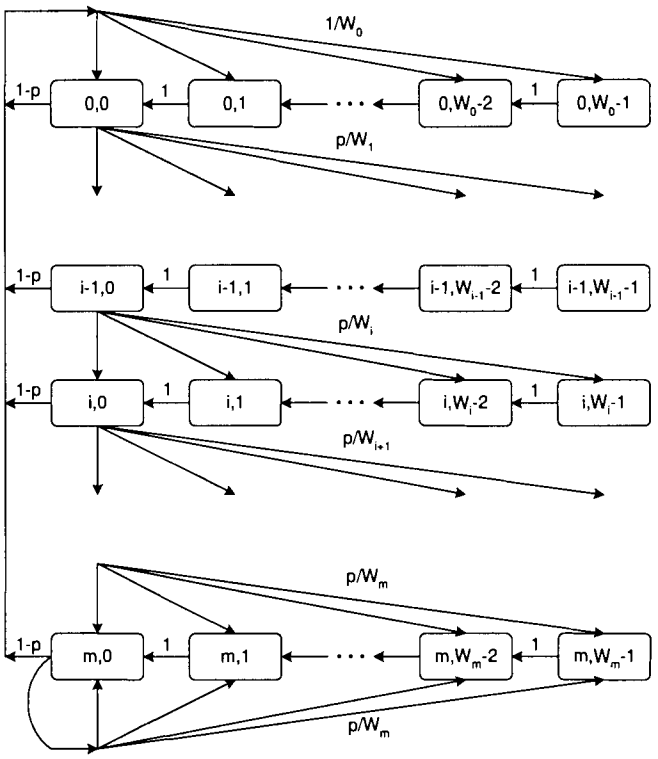


Figure 3.3: Bianchi Model Markov process for DCF backoff.

are chosen independently from uniform distributions, the probability that a station transmits in any given backoff slot, τ, is uniform and equal to the sum $\sum_{i=0}^{m} b_{i,0}$.

Because a transmission experiences a collision whenever one or more other stations transmit in the same backoff slot, the probability of collision p for a transmission in a network of n stations can be uniformly expressed in terms of τ

$$p = 1 - (1 - \tau)^{n-1}. \tag{3.1}$$

Combining equation (3.1) and the closed form solution for τ based on the stationary distribution of the Markov chain,

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{2(1-2p)}{(1-2p)(W+1) + pW\left(1-(2p)^{m}\right)}, \tag{3.2}$$

yields a two equation non-linear system with unknowns p and τ. Solving this system with numerical techniques yields a unique solution for p and τ. Having found τ, simple expressions in terms of τ for the probability of at least one station transmitting (P_{tr})

and a successful transmission occurring (P_s) in a given time slot are used to calculate throughput (S).

$$S = \frac{E\left[\text{Data transmitted in a time slot}\right]}{E\left[\text{Real-time duration of a time slot}\right]}$$
$$S = \frac{P_{tr}P_sE[P]}{(1-P_{tr})\,\sigma + P_sP_{tr}T_s s + P_{tr}\,(1-P_s)\,T_c} \tag{3.3}$$

where P_{tr} is the probability that at least one station transmits,

$$P_{tr} = 1 - (1-\tau)^n\,, \tag{3.4}$$

P_s is the probability of a successful transmission from any station,

$$P_s = \frac{n\tau\,(1-\tau)^{n-1}}{P_{tr}}, \tag{3.5}$$

σ is the duration of a backoff slot, T_s is the duration of a successful transmission, T_c is the duration of a colliding transmission and $E[P]$ is the expected amount of payload data delivered by a successful transmission.

Since the constants T_s and T_c are properties of the transmission mechanism, the Bianchi model accommodates both the basic and RTS/CTS transmission mechanisms. Bianchi also notes that the model can also be extended to accommodate variable sized payloads and collisions by substituting expectations for T_s and T_c instead of fixed values.

Validation of this model against simulation results for throughput versus number of saturated stations (load) confirms that model is accurate. However, the author remarks that the model is more accurate when there are greater numbers of stations without offering an explanation. A likely explanation is the failure of the model to account for stations idle during ACKTimeout and/or CTSTimeout, while the simulator had an incorrect, but still present ACKTimeout. When there are fewer stations, the effect of two stations being absent from contention on probability of collision is greater when there are many stations.

Another criticism of the paper is that the treatment of the backoff mechanism isn't strictly correct. In IEEE 802.11 backoff counters are decremented at the end of each idle backoff slot and upon the backoff counter reaching zero stations initiate transmission immediately. In all cases where stations choose initial backoff counter

values greater than zero, stations will wait at least (DIFS + aSlotTime) after the medium becomes idle before contending for access to the medium. This is in contrast to the case where a station chooses an initial backoff counter of 0 from the range [0,CW]. In this case, the station will transmit immediately after the end of the DIFS deferral. Since every other station must observe at least one backoff slot after DIFS before transmitting, this transmission at the end of DIFS cannot collide. Obviously this violates the approximation that the probability of collision is uniform for each transmission attempt, though the impressive accuracy of the model leads us to judge any corrections to the model to account for this behaviour as unworthy.

This model's simplicity and accuracy have made it a favourite starting point for improved IEEE 802.11 DCF models. In (Wu, Cheng, Peng, Long, and Ma 2002) the authors make a simple extension to model the finite retry counter limit by introducing a terminating backoff stage after which $s(t)$ is reset to zero regardless of the transmission outcome. In (Hadzi-Velkov and Spasenovski 2003) the model is extended to cover both finite retry limit and external frame-error probability (i.e. transmission failures not due to collision). Raptis, Vistas, Paprrizos, Chatzimisios, and Boucouvalas (2004) publish essentially the same work in the same conference, but a year later.

Chatzimisios, Vitsas, and Boucouvalas (2002) develop a complicated means for determining the expected service time for a packet arriving at the head of the queue in a saturated station. They present more polished versions of this work in (Chatzimisios, Boucouvalas, and Vitsas 2003b) and (Chatzimisios, Boucouvalas, and Vitsas 2003a). A very similar approach is used in (Wang, Shu, Zhang, and Yang 2003). Apparently these authors were unaware of Little's result (Little 1961), which makes calculation of mean service delay trivial once saturation throughput is known. In (Carvalho and Garcia-Luna-Aceves 2003) the authors linearize the non-linear expressions for transmission probability and collision probability from (Bianchi 2000) and use the simplified model to determine the first and second moments (average and variance) of service delay.

Alizadeh-Shabdiz and Subramaniam (2003) attempts to model the operation of IEEE 802.11 DCF under non-saturation conditions by wedding a continuous time arrival process to the bi-dimensional Markov backoff process developed by Bianchi (2000). The arrival process is accommodated by adding new one-step transition probabilities to the Markov processes between the completion of a packet transmission and

the next service of packet. The net result of the added states and transitions is a more complex system of equations for p and τ. The calculation for throughput is just that in (3.3). The authors validate the extended model by comparing analytical and simulation results for throughput versus applied load. The results indicate an agreement between the model and simulation, but this doesn't necessarily confirm the correctness of the system. In any situation where arrival traffic is below the saturation threshold network throughput is strictly limited by the arrival rate of the traffic. Hence, a steady state analysis of throughput versus load isn't very useful and doesn't demonstrate correctness of the model. A more useful and interesting result would be a calculation of mean service delay versus load.

3.4 802.11e EDCA/EDCF Modeling Research

Several complete analytical models have been published for IEEE 802.11e EDCA/EDCF. We say the first complete analytical model because there are a number of papers (He, Zheng, Yang, and Chou 2003) (Qiao and Shin 2002) (Hui and Devetsikiotis 2003) which trivially extend the model from (Bianchi 2000) to include heterogeneous contention window sizes. Because the extension to contention window differentiation is so simple, the following survey concentrates exclusively on the authors' treatment of AIFS.

In (Ge and Hou 2003) the authors extend the work of (Cali, Conti, and Gregori 1998) to treat multiple classes of traffic. The authors also demonstrate a method for tuning transmission probabilities to achieve a desired throughput ratio between classes while maximizing channel capacity. However, because the model is based on a p-persistent variant of EDCF it does not provide any insight into the operation of EDCF, and is of limited use in the studying of proper EDCF.

In (Zhao, Guo, and Zhu 2002) the authors present a three dimensional model for IEEE 802.11e EDCF. Starting from the Bianchi model, a third Markov process variable is added to track the number of slots elapsed since the DIFS boundary following the last transmission. One step transition probability are redefined to accommodate the third dimension and the stationary distribution if found by iterative application of the transition matrix to a state vector. The third dimension is used to condition the probability of collision so that it reflects the set of stations contending at

the time of transmission. Throughput is calculated by adapting Bianchi's method to accommodate the third dimension.

Yang (2003) attempts to model the effect of AIFS by conditioning each backoff process event in Bianchi's (Bianchi 2000) DCF model with the probability that the channel is idle in the slot prior. Nothing Yang proposes provides any differentiation based on AIFS. Since all ACs are conditioned by the same probability there is no advantage for having a smaller AIFS. Even the treatment of the probability of the channel being busy as uniform is incorrect. Since AIFS divides channel contention into periods where different sets of stations are active, the probability of the channel being busy must vary across the backoff period. Yang also makes the mistake of modifying the model so that it **does not** include post backoff, introducing a transition for stations transmitting to an idle medium immediately after a previous transmission.

Yang (2004b) presents essentially the same work as in (Yang 2003), but adds an analysis for mean service delay. The method is based on the expected number of transmission attempts, the expected number of total backoff slots drawn and the expected number of events seen during backoff. The author's next paper, (Yang 2004a), adds a trivial calculation for packet failure probability.

Zhu and Chlamtac (2003) tackle the effects of AIFS in EDCA by assigning the probability of backoff counter decrement in Bianchi's model (Bianchi 2000) values less than one. In contrast to (Yang 2003) the probability of backoff is unique to each access category. The intent is that lower priorty stations take longer than higher priority stations to complete the same backoff, and so have a lower probability of transmitting (τ) in any given backoff slot. However, the authors' formulation for the probability of decrement is clearly incorrect since it does not depend on the number of stations contending for access to the medium. The authors "validate" their model with an artful comparison to simulation results from ns-2. We call the comparison artful because (1) non-saturation traffic is used in the simulation, when comparison is to a saturation traffic model, (2) simulation and model results are presented in separate graphs (not overlaid) with absurdly imprecise scales and (3) results are only presented at very high loads, when AIFS differentiation has starved lower priority traffic to zero throughput.

In (Robinson and Randhawa 2004b) we introduced the concept of *average conditional collision probability* to capture the effect of contention zones in a two dimensional, Bianchi type model. Average conditional collision probability is calculated by modeling the evolution of contention zones within a contention period by a Markov chain. Using this approach meant that the timescales of CAFs belong to different ACs could be de-coupled, and the simplicity of the Bianchi model preserved. The model is used to predict throughput and collision probability, and is shown to be accurate in a comparison to the ns-2 simulator developed by TGe members Singla and Chesson. This model forms the basis for the work presented in this thesis, and its details are laid out later.

Hui and Devetsikiotis (2004) present a unified model for IEEE 802.11e EDCA based on the p-persistent model in (Cali, Conti, and Gregori 1998). The p-persistent adaptation to EDCA is similar to that developed in (Ge and Hou 2003). The work unifies this approach with a two-dimensional Markov process that is adapted to EDCA by reducing the contention window sizes by a function of AIFS.

In (Ramaiyan, Kumar, and Altman 2005) a fixed point analysis is applied to IEEE 802.11e EDCA throughput. The authors cite our work (Robinson and Randhawa 2004b) as the origin of the treatment they use for AIFS. The fixed point analysis is attractive for the simplicity of its expression and for its proof of the existence of a fixed point solution for the equation relating collision probability of a single station to the attempt probability of all other stations. The authors use the fixed point analysis to draw conclusions about the effectiveness of initial contention window size and AIFS for throughput differentiation.

A model for the average service delay of IEEE 802.11e EDCA recently appeared in (Banchs and Vollero 2005). The underlying backoff model is taken from (Wu, Cheng, Peng, Long, and Ma 2002) and extended to model the effect of AIFS by defining the probability of collision in terms of the probability that contention slots are empty. The average service delay is calculated using a method very similar to that in (Yang 2004b).

3.5 Service Delay Distribution Modeling Research

Because real-time services have strict delay requirements for application traffic and since delay variability in random access protocols may be high, the probability distribution of frame service delay is a much more useful metric that mean frame service delay. There have been several attempts to model the frame service delay distribution for 802.11 DCF and for 802.11e EDCA/EDCF.

In (Zhai and Fang 2003) the probability distribution of frame service delay for IEEE 802.11 is derived from the signal transfer function of a generalized state transition diagram based bi-dimensional Markov backoff process developed in (Bianchi 2000). Because this method is similar to the one we pursue in this thesis, we discuss it in detail here. [1]

In (Zhai and Fang 2003) the authors transform the Bianchi Markov backoff process model for IEEE 802.11 DCF to a signal flow graph in the z domain. The basic idea is that the signal in each node of the signal flow graph is the steady state probability of being in a Markov process state. Each state transition in the Markov process is tagged with a polynomial function of z denoting delay. The sum of the exponents in the polynomial function correspond to the the real-time time delay of the state transition. Treating the state transition diagram as a signal flow graph means that the transmissions of successive branches in a path are multiplied together. Thus for a given path from start to end, the probability is the product of the branches traversed and the delay is the sum of the z exponents tagging the branches.

Figure 3.4 shows the generalized state transition diagram that the authors proposed for IEEE 802.11 DCF.

In the above diagram p corresponds to the probability of a transmission colliding, and the time delay tags z^{σ}, $z^{T_{\text{suc}}}$ and $z^{T_{\text{col}}}$ correspond to backoff transition duration for an idle slot, a successful transmission and a collision, respectively. For correct manipulation of the signal flow graph, each delay is ultimately expressed in terms of a common unit of time σ.

[1] The presentation of derivations and results in this paper is poor, with several obvious errors. In the ensuing discussion we assume corrections for the typographic errors and pursue remedies for errors of substance. We also take the liberty of changing notation where consistency with previous discussions is desirable.

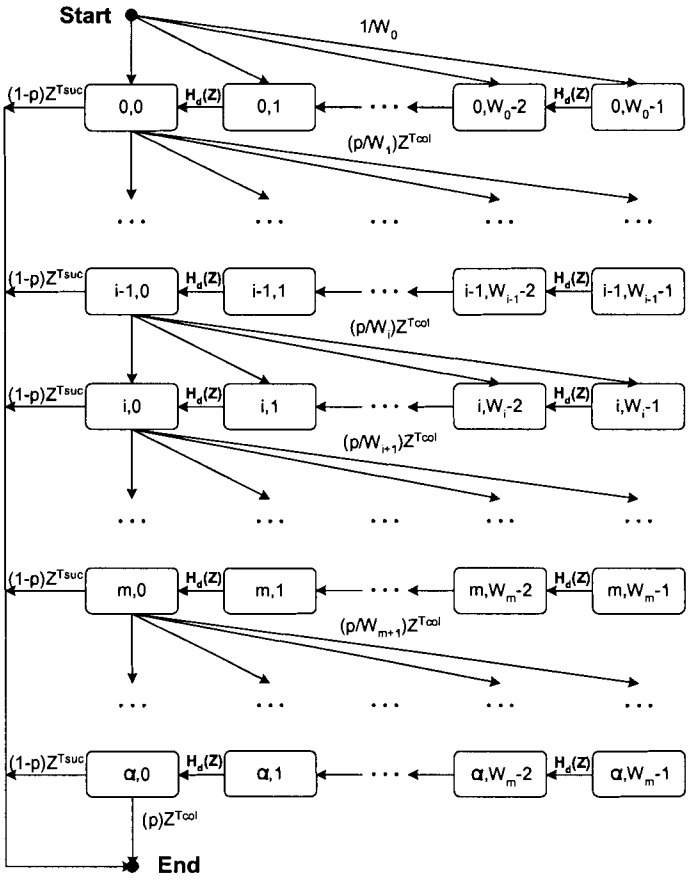


Figure 3.4: Generalized state transition diagram.

The transmission for the backoff decrement transition, $H_d(z)$ is given by the authors to be

$$H_d(z) = \frac{(1 - P_{\text{tr}})z^{\sigma}}{1 - P_{\text{s}}z^{T_{\text{suc}}} - (P_{\text{tr}} - P_{\text{s}})z^{T_{\text{col}}}}, \tag{3.6}$$

where P_{tr} is the probability of a transmission occurring and P_{s} is the probability of a transmission success occurring.

This transfer function corresponds to the following signal flow graph shown in Figure 3.5

The authors explain the function by saying the backoff timer decrements after an empty slot time σ with probability $(1 - P_{\text{tr}})$, stays in the same state after observing a transmission with time T_{suc} with probability P_{s} or stays in the same state after observing a collision of duration T_{col} with probability $(P_{\text{tr}} - P_{\text{s}})$.

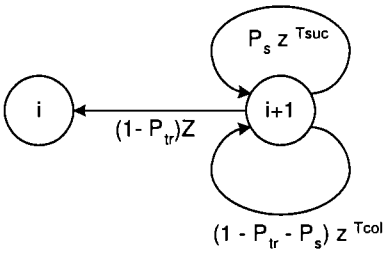


Figure 3.5: Signal flow graph for backoff in (zhai and Fang, 2003).

This explanation is at odds with the actual operation of the IEEE 802.11 protocol. Since stations transmit only when their backoff counter reaches zero, and backoff counters are only decremented *after* an idle slot, all transmissions must be preceded by an idle backoff slot. Therefore, stations in fact decrement their backoff counters on the slot edges where transmissions from other stations begin. The only exception to this rule, discounted by Bianchi in his analysis due to its rarity, is the case where a station selects 0 as its initial backoff counter value and transmits at the end of the first idle DIFS it observes.

Thus the correct signal flow graph looks as follows (Fig. 3.6): Branches in the

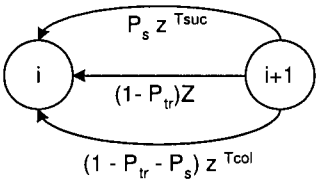


Figure 3.6: Correct signal flow graph for DCF backoff.

above graph can be interpreted as being tagged with the time a station must wait between successive decrement operations. With probability $1 - P_{\mathrm{tr}}$ a station decrements after observing only an idle slot σ since its last decrement (or backoff counter selection). The probability that a station observes an idle backoff slot followed by a successful transmission (T_{suc}) is P_{s}. And the probability that a station observes an idle backoff slot and a collision (T_{col}) between decrements is $(1 - P_{\mathrm{tr}} - P_{\mathrm{s}})$.

Note that Figure 3.4 clearly designates the start and end points of the signal transfer function for delay. Using common linear systems techniques the authors reduce the transition diagram into a signal transfer function from the start state to

the end state. This signal transfer function is the probability generating function for service delay, and can be expanded out in power series, that is,

$$B(z) = \sum_{i=0}^{\infty} P(T_s = i)z^i \tag{3.7}$$

From the PGF the authors obtain the discrete probability distribution for service time delay. The authors compare the analytical and simulation (ns-2) results for service delay distribution. The comparison shows that the service delay, estimate provided by the model is within 15% of the simulation values.

A further problem with the approach in (Zhai and Fang 2003) is that it will not scale to typical IEEE 802.11 DCF window sizes. Consider the contention window specification for the IEEE 802.11 Direct Sequence PHY. The minimum contention window is 31, the maximum 1023. When the generalized state transition graph, shown in Figure 3.4, is reduced, the path polynomial must be multiplied with $Hd(z)$ for each backoff state. Since $Hd(z)$ has three terms, the number of terms in the path polynomial for a given backoff stage grows as shown in Table 3.5. At this rate the path

$Hd(z)$	3
$Hd(z)^2$	6
$Hd(z)^3$	10
$Hd(z)^4$	15
$Hd(z)^5$	21
$Hd(z)^6$	28
$Hd(z)^n$	$15 + 7(n-4)$

Table 3.1: Number of terms in powers of the backoff polynomial $Hd(z)$

polynomial for the generalized state diagram will have at least $15+7(3033-4) = 21218$ terms! Since the best sparse algorithm for multiplication of input polynomials with s and t terms whose product has k terms uses $O(st)$ coefficient multiplications and $O(st-k)$ additions (Fateman 2003), it is clear that the brute force reduction of the generalized state diagram as proposed by Zhai is not possible on commodity computing hardware.

The work (Xu, Wang, and Hassanein 2003) models the state of an entire network consisting of m of EDCF stations as a single, discrete time m-dimensional Markov process where each dimension represents the backoff counter value of each station. The

authors adopt a discrete timescale denominated by transmission events. The process is regenerative in that it is anchored to transmission events where every station in the network collides and draws a new random backoff. Since all backoff counters are random after a full collision, the regenerative states are identical in that they restart probabilistically-identical operation.

The authors define a transition matrix for the probability of moving from each aggregate backoff counter state to the next aggregate backoff counter state in terms of the contention window size. However, the authors do not explain how model can possibly accommodate binary exponential backoff. Since only the backoff counter value of each station is tracked, there is no way for the model to know which backoff stage a station is in. Hence the model cannot know the size of the contention window from which backoff is to be drawn, and EDCF cannot be modeled realistically.

The work in (Robinson and Randhawa 2004a) constructs a CDF for the service delay distribution of IEEE 802.11e EDCA by calculating the probability that a frame sees a particular number of timeslots before successful transmission. The number of slots seen is multiplied by the average real time duration of each timeslot. The authors include a favourable comparison between results from the model and results from ns-2 simulations to illustrate the accuracy of their method. However, because the model averages slot duration, the service delay CDF is noticeably more angular than simulation results and so does not exhibit a uniform correspondence to the simulation curves. A further problem is that the method for determining the number of slots seen is based on combinatorial probability calculations that do not scale to large contention window sizes (consider the factorial of 1023!).

In (Ozdemir and McDonald 2004) the authors present a novel model for IEEE 802.11 DCF backoff, that while based on a two-dimensional discrete time Markov chain does not rely on the Bianchi assumptions of uniform transmission probability, and constant and independent collision probability. Instead of determining the probability of collision in terms of the uniform probability of transmission, the model calculates a backoff stage specific, pair-wise collision probability between the CAF being analyzed and each other stations. This calculation is based on the probability that the two stations have chosen backoff counters that will guarantee a collision between the stations. The pair-wise collision probability are assumed to be independent and the stage specific collision probability is calculated as the product of the pairwise

collision probabilities. The authors use this model to underpin an analysis of the service time distribution. The service time distribution is constructed using the same average slot duration method same method as (Robinson and Randhawa 2004a), and suffers from similar shortcomings.

The work in (Zhen-ning, Tsang, and Bensauo 2004) presents another model that attempts to characterize service delay for 802.11e EDCA. The model is based on an extension of the Markov chain from (Bianchi 2000) to three dimensions. To provide a bridge between model/event time and real time, the third dimension of the process tracks progress through both the slots that must be counted for AIFS and the through time that the backoff is frozen during transmissions. Because the time that backoff is frozen varies depending on the current transmission, the authors use the average duration of backoff pauses, denominated in units of aSlotTime, to fix the number of states in the third dimension.

Unfortunately, the authors do not discuss how they determine the average duration of backoff suspensions. In the examples used for validation, the basic transmission mechanism and homogeneous frame sizes ensure that all backoff suspensions have share the same duration. Since it is not possible to correctly predict the average duration of backoff suspensions *a priori*, and any computed value for average backoff duration would depend on initial guess of average backoff duration, the model would have to be run iteratively until convergence is achieved.

The authors use the three dimensional Markov model to produce results for throughput and service delay distribution. The throughput results are derived using a method similar to that proposed by Bianchi in (Bianchi 2000). The delay distribution is derived from a recursive characterization of the time delay from a given model state until the frame is transmitted successfully. The recursive definition identifies two possible delays between adjacent backoff states: a single backoff slot and a transmission. Though this definition does not accommodate heterogeneous transmission durations (use of the RTS/CTS mechanism, different frame sizes for each AC) it is suitable for uniformly sized frames transmitted using the basic mechanism.

In (Tao and Panwar 2004b) and (Tao and Panwar 2004a) Tao and Panwar present a saturation model for IEEE 802.11e EDCF based on a three dimensional Markov chain. The third dimension of the model tracks the progress of a station through slots within a contention period. The model accounts for the differentiated collision

probability of AIFS by defining the uniform transmission probability τ on a slot (third dimension) specific basis. The authors derive expressions for throughput based on (Bianchi 2000) and claim to be able to derive the delay distribution.

The method for deriving the delay distribution is described in a scant five sentence paragraph. It consists of multiplying a state probability vector 'with the transition matrix of the Markov chain, unil the resulting state probability vector converges', and 'counting the iterations and the corresponding probability that the packet is successfully distributed'. The authors do not mention of how model events are related to real-time delays. We do not see how the model can possibly estimate real-time service delay.

In (Raptis, Vistas, Paprrizos, Chatzimisios, and Boucouvalas 2005) the authors estimate the service delay distribution for IEEE 802.11 DCF by calculating the expected delay at each backoff stage and the probability that a transmission is successful at each particular stage. Though this does give some insight into the distribution of backoff delays, it is closer to the calculations for mean service delay than to a true CDF for service delay.

3.6 Context for Thesis Contributions

Based on the preceding literature review, we draw the following summary conclusions on the state of the art for modeling research on IEEE 802.11 DCF and IEEE 802.11e EDCA/EDCF:

1. There are no backoff models for IEEE 802.11 DCF or IEEE 802.11e EDCA that treat ACKTimeout as it is defined in the standard.

2. Most models for service delay distribution rely on averaging techniques, and provide only rough approximations of the service delay distribution.

3. The only model that attempts an exact analysis of the service delay distribution for IEEE 802.11 DCF is flawed and is not practical for estimating performance of typical IEEE 802.11 networks.

4. We have presented the only model for service delay distribution that can accommodate all the differentiation and transmission features of IEEE 802.11e EDCA as an earlier part of this research effort (Robinson and Randhawa 2004a).

5. There are no models for an exact analysis of the IEEE 802.11e service distribution.

Chapter 4

Backoff Model

This work extends the model from (Bianchi 2000) to accommodate the Quality of Service features provided by EDCA and to account for the effects of the ACK-Timeout/CTSTimeout. Sharing the approach of using p and τ to underlie a two-dimensional discrete-time Markov chain (DTMC), the following conjectures underpin the model:

- each queue in the system is modeled by a DTMC specific to the AC associated with the CAF,
- the probability that a CAF initiates a transmission in a given backoff slot (τ) is constant across all backoff slots for which it may transmit,
- the probability that any transmission experiences a collision (p) in a given contention zone among a given set of contending CAFs is constant, and independent of the number of transmissions attempts.

The reasoning behind the first point is obvious: since CAFs of different AC progress through backoff at different rates, they must be modeled by different DTMCs that may operate on different timescales.

The second conjecture responds to the random nature of pauses in the backoff procedure caused by randomly chosen backoff intervals. When multiple CAFs are contending for use of the medium, the backoff counter of a single CAF may be stalled several times while other CAFs access the medium. Given that the value of a backoff

counter at the beginning of any transmission period is random, the probability of the backoff counter being decremented to zero in a backoff slot is uniform.

The third conjecture is the parallel of Bianchi's approximation that all transmissions collide with constant and independent probability. The probability that a transmission collides is a function of the size and composition of the set of competing CAFs. Since ACKTimeout/CTSTimeout and AIFS affects the set of competing stations, the probability of collision must differ according to the set of contending CAFs and contention zones. Since our model does not track progress within a transmission period, nor the outcome of the previous contention period, our solution uses an average conditional collision probability $\bar{p}$ to govern the DTMC.

We present the development of the backoff model in two parts. First we present a DTMC backoff process to model backoff and transmission events. Then we introduce and develop the concept of average conditional collision probability.

4.1 The Discrete-Time Markov Chain for EDCA Backoff

Consider a network in which contending CAFs belong to one of n distinct ACs. In this network a fixed number N_k of CAFs from each AC k contend for access to the transmission medium. Under the assumption of saturation conditions, every CAF always has a frame to transmit and is therefore always either transmitting or engaged in the backoff procedure.

Let $b(t)$ be the stochastic process representing the backoff time counter for a given CAF. The backoff time $b(t)$ is an integer value denoting the number of slot boundaries that the CAF must observe before the backoff time counter expires. Since CAFs transmit on the first slot boundary after the backoff time counter expires, a CAF will transmit the first slot boundary after $b(t)$ is decremented to zero.

Because $b(t)$ operates on slot boundaries, it is a discrete time process. Note that the discrete time scale does not correspond to real time. As explained in section 2.1.1, backoff time counters are suspended whenever the transmission medium is sensed busy. Thus, the time interval between two consecutive slot boundaries depends on the state of the transmission medium.

CAFs draw new backoff time counter values each time a frame is successfully transmitted and each time a frame suffers a transmission failure. The range from which $b(t)$ is drawn after i transmission attempts is $[0, W_i - 1]$ where W_i is the contention window $\min(\text{aCWmax} + 1, 2^{i-1} \cdot (\text{aCWmin} + 1))$, and $i \in [0, m + f]$ is called the *backoff stage*. Because the range of values from which $b(t)$ are drawn depends on the number of times the current frame has been transmitted, the stochastic process $b(t)$ is non-Markovian.

Now consider a stochastic process $s(t)$ representing the number of times a CAF has transmitted the current frame. $s(t)$ is reset to zero whenever a frame is successfully transmitted or discarded for being transmitted without success $m + f$ times. Every time that a frame transmission fails $s(t)$ is incremented. Therefore $s(t + 1)$ depends only on $s(t)$ and the constant probability $\bar{p}$ that a frame transmission fails.

Since reference to $s(t)$ resolves the non-Markovian dependence of $b(t)$ on the backoff stage, the CAF backoff process can be described by a single bi-dimensional stochastic process $\{s(t), b(t)\}$. Once independence is assumed and $\bar{p}$ is supposed to be a constant value, $\{s(t), b(t)\}$ may be modeled by the discrete-time Markov chain shown in Fig. 4.1

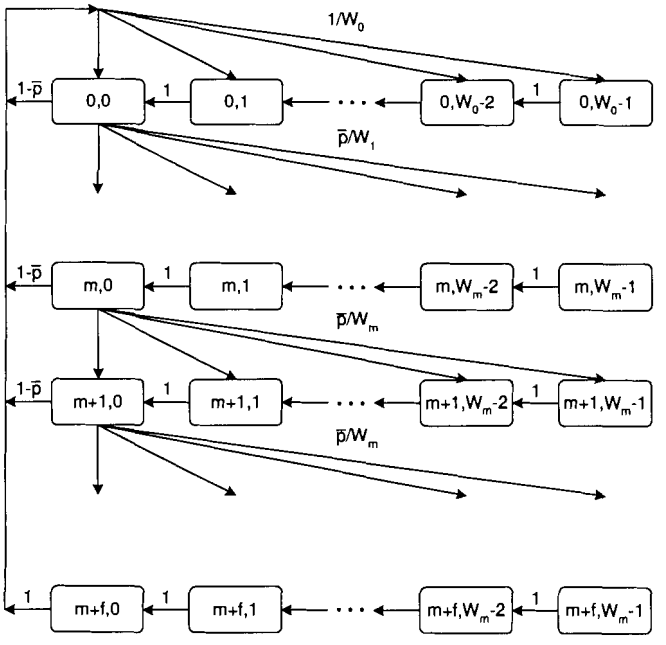


Figure 4.1: Two-dimensional DTMC for EDCA backoff

Assuming knowledge of the average collision probability ($\bar{p}$), and defining the range of possible backoff counter values for backoff stage i as $[0, W_i - 1]$, the following non-null one-step transition probabilities govern the activity of our DTMC for EDCA backoff:

$$\begin{cases} P\{i,k|i,k+1\} = 1 & k \in [0, W_i - 2] \quad i \in [0, m+f] \\ P\{i,k|i-1,0\} = \bar{p}/W_i & k \in [0, W_i - 1] \quad i \in [0, m] \\ P\{i,k|i-1,0\} = \bar{p}/W_m & k \in [0, W_i - 1] \quad i \in [m+1, m+f] \\ P\{0,k|i,0\} = (1-\bar{p})/W_0 & k \in [0, W_0 - 1] \quad i \in [0, m+f) \\ P\{0,k|m+f,0\} = 1/W_0 & k \in [0, W_0 - 1]. \end{cases}$$

The state transitions above correspond to (top to bottom): decrementing the backoff counter on a slot boundary, a failed transmission attempt before aCWmax is reached, a failed transmission attempt after aCWmax is reached, a successful transmission attempt, any transmission attempt in the last backoff stage.

Denoting the stationary distribution by $b_{i,k} = \lim_{t\to\infty} = P\{s(t) = i, b(t) = k\}, i \in [0, m+f], k \in [0, W_i - 1]$, we note the following relationship between backoff stages

$$b_{i-1,0} \cdot \bar{p} = b_{i,0} \qquad \rightarrow \qquad b_{i,0} = \bar{p}^i b_{0,0} \qquad i \in [1, m+f]. \tag{4.1}$$

Relationships between stationary distributions of neighboring backoff states are

$$b_{0,k} = b_{0,k+1} + \frac{1}{W_0}\left[(1-\bar{p})\sum_{i=0}^{m+f-1} b_{i,0} + b_{m+f,0}\right] \quad k \in [0, W_0 - 2], \tag{4.2}$$

$$b_{i,k} = b_{i,k+1} + \frac{1}{W_i} \cdot \bar{p} \cdot b_{i-1,0} \qquad k \in [0, W_0 - 2] \quad i \in [1, m+f]. \tag{4.3}$$

From (4.1) and (4.2) we deduce for all $k \in [0, W_i - 1]$,

$$b_{0,k} = b_{0,0} \cdot \frac{W_0 - k}{W_0}, \tag{4.4}$$

and for all $k \in [0, W_i - 1]$ and $i \in [1, m+f]$,

$$b_{i,k} = b_{0,0} \cdot \bar{p}^i \cdot \frac{W_i - k}{W_i}. \tag{4.5}$$

A solution for $b_{0,0}$ in terms of average conditional collision probability $\bar{p}$ is found by imposing the normalization condition on the DTMC,

$$\begin{aligned}
1 &= \sum_{i=0}^{m+f} \sum_{k=0}^{W_i-1} b_{i,k} \\
&= \sum_{i=0}^{m+f} b_{i,0} \sum_{k=0}^{W_i-1} \frac{W_i - k}{W_i} \\
&= \sum_{i=0}^{m+f} b_{i,0} \frac{W_i + 1}{2} \\
&= \sum_{i=0}^{m} b_{i,0} \frac{W_i + 1}{2} + \sum_{i=m+1}^{m+f} b_{i,0} \frac{W_m + 1}{2} \\
&= \frac{b_{0,0}}{2} \left\{ \sum_{i=0}^{m} (2\bar{p})^i W_0 + \sum_{i=0}^{m} \bar{p}^i + \sum_{i=m+1}^{m+f} \bar{p}^i 2^m W_0 + \sum_{i=m+1}^{m+f} \bar{p}^i \right\} \\
&= \frac{b_{0,0}}{2} \frac{1 - \bar{p}^{m+f+1}}{1 - \bar{p}} + \frac{b_{0,0}}{2} W_0 \left\{ \frac{1 - (2\bar{p})^{m+1}}{1 - 2\bar{p}} + 2^m W_0 \left(\frac{\bar{p}^{m+1} - \bar{p}^{m+f+1}}{1 - \bar{p}} \right) \right\} \\
&= \frac{b_{0,0}}{2} \frac{1 - \bar{p}^{m+f+1}}{1 - \bar{p}} + \frac{b_{0,0}}{2} W_0 \left\{ \frac{\left[1 - (2\bar{p})^{m+1}\right](1 - \bar{p}) + (1 - 2\bar{p})(2\bar{p})^m (\bar{p})(1 - \bar{p}^f)}{(1 - \bar{p})(1 - 2\bar{p})} \right\},
\end{aligned}$$

from which we deduce that

$$b_{0,0} = \frac{2(1 - 2\bar{p})(1 - \bar{p}^{m+f+1})}{(1 - \bar{p}^{m+f+1})(1 - 2\bar{p}) + W_0 \left\{ \left[1 - (2\bar{p})^{m+1}\right](1 - \bar{p}) + \bar{p}(2\bar{p})^m (1 - 2\bar{p})(1 - \bar{p}^f) \right\}}. \tag{4.6}$$

Recalling that CAFs transmit on the first backoff slot occurring after the backoff time counter, $b(t)$, reaches to zero, we find the probability τ that a CAF transmits on any backoff slot boundary is

$$\begin{aligned}
\tau &= \sum_{i=0}^{m} b_{i,0} \qquad (4.7) \\
&= b_{0,0} \sum_{i=0}^{m+f} \bar{p}^i \\
&= b_{0,0} \frac{1 - \bar{p}^{m+f+1}}{1 - \bar{p}} \\
&= \frac{2(1 - \bar{p})(1 - \bar{p}^{m+f+1})}{(1 - \bar{p}^{m+f+1})(1 - 2\bar{p}) + W_0 \left\{ 1 - \bar{p} - \bar{p}(2\bar{p})^m \left[1 + \bar{p}^f - 2\bar{p}^{f+1} \right] \right\}}
\end{aligned}$$

4.2 Average Conditional Collision Probability

The expression for τ developed in the previous section depends on the average probability of collision $\bar{p}$. This section develops an expression for average conditional collision probability $\bar{p}$ in terms of uniform transmission probability τ, the number of CAFs per AC in the network ($\vec{N}$) and the number of ACs active in the network (n). Once an expression for $\bar{p}$ is found, we can pair it with the expression for τ found previously to form a system of two non-linear equations.

In the development that follows, we are often at pains to distinguish one variable from another. To reduce confusion, we adopt the following conventions:

- lower case p is used for variables denoting probabilities for events occurring on slot edges
- upper case P is used for variables denoting probabilities for contention period outcomes
- plain text subscripts x_{plain} are use to distinguish among different variables
- plain text superscripts x^{plain} are use to distinguish among different constants
- vector arrows $\vec{x}$ are used to denote sets of CAFs
- carat characters $\hat{x}$ are used to indices for sets of CAFs

4.2.1 Static Conditional Collision Probability

A collision between transmissions occurs whenever more than one transmission is broadcast on the transmission medium at once. Since IEEE 802.11 operation is pleisosynchronous, collisions occur whenever more than one CAF initiates a transmission on a given slot boundary. The probability that any CAF's transmission collides is thus a function of the probability that any other CAF initiates a transmission at the same time. From the point of view of a single CAF, collision probability is conditional, that is conditioned on the knowledge that the CAF in question is transmitting on the particular slot boundary.

Consider a single CAF, assigned to AC j, which we will refer to as $\vec{q}$ (i.e. $q_j = 1$ and $q_k = 0$ for all $k \neq j$) and whose 'viewpoint' we adopt for the development that follows.

Now assume that $\vec{q}$ contends for channel access against a set of CAFs described by an n-tuple *contention state* $\vec{x} = [x_1 x_2 \dots x_k \dots x_n]$, where each element x_k in $\vec{x}$ specifies the number of contending CAFs belonging to AC $k \in [1, n]$. Note that the single CAF $\vec{q}$ is not included in the set of contending CAFs $\vec{x}$. If each CAF in $\vec{x}$ transmits on slot boundaries with uniform probability τ_k, the probability that a transmission from $\vec{q}$ collides is

$$p = 1 - \prod_{k=1}^{n} (1 - \tau_k)^{x_k}. \tag{4.8}$$

The expression above assumes that there are no lower priority CAFs collocated at the same QSTA as $\vec{q}$. We note that the virtual collision mechanism, described in Chapter 2, can be modeled by excluding from N_k all collocated CAFs belonging to lower priority ACs.

From the (4.8) it is obvious that collision probability depends on which contention state prevails at the time of transmission. Because the prevailing contention state changes from one contention period to the next, the stationary distribution of contention states across all transmission slots is essential to the expression for average collision probability.

4.2.2 Contention State Stationary Distribution

Determining the stationary distribution of contention states requires that we find the probability that each contention state is followed by any other. Let us define a *contention period* as a run of consecutive backoff slots, uninterrupted by transmission events, where the same contention state prevails. By this definition, the prevailing contention state only changes when either a transmission event terminates a contention period, or an ACKTimeout/CTSTimeout expires. We refer to the particular event that ends a contention period as a contention period outcome (CPO).

The probability that one contention state follows another is the net probability of CPOs that change the antecedent contention state to the subsequent contention state. This probability, like the collision probability in the previous section, is conditional in that contention period outcomes are 'seen' from the point of view of the single CAF $\vec{q}$ contending against other CAFs.

The probability relationship between antecedent and subsequent contention states can be expressed in the form of a two dimensional transition matrix $\mathbb{P}$. Since $\mathbb{P}$ describes a relationship between two n-tuple contention states, $\vec{x}$ and $\vec{y}$, a mapping is required to relate each n-tuple contention state to a unary index $\hat{i}$ for $\mathbb{P}$. Let the set of all CAFs active in the network be represented by the n-tuple $\vec{N}$. The set of CAFs against which $\vec{q}$ may contend is expressed by $\vec{M} = \vec{N} - \vec{q}$. The mapping from any contention state $\vec{x} \in \vec{M}$ to an index i is described by

$$\hat{i} = x_1 + \sum_{k=2}^{n} x_k \prod_{j=1}^{k-1} (M_j + 1)\,. \tag{4.9}$$

The reverse mapping is

$$x_1 = \hat{i} \bmod (M_1 + 1)\,, \tag{4.10}$$

$$x_k = \left\lfloor \frac{\hat{i}}{\prod_{j=1}^{k-1} (M_j + 1)} \right\rfloor \bmod (M_k + 1) \qquad k > 1. \tag{4.11}$$

The range $\hat{i} \in \left[0, \prod_{j=1}^{n} (M_i + 1) - 1\right]$ covers all possible $\vec{x} \in \vec{M}$; we shall refer to the upper limit on the range of $\hat{i}$ as $\hat{M}$. For convenience $\hat{i}$ and $\vec{x}$ notations will be used interchangeably.

Each row $\mathbb{P}_{\hat{i}}$ can be constructed by adding the probability of each CPO to the element $p_{\hat{i},\hat{j}}$ prescribed by that outcome. In the development that follows we first describe how to find the probability of CPOs, then describe how to relate these outcomes to subsequent contention states.

Contention Period Outcome Probabilities

Because the DTMC backoff model describes event probabilities on a slot boundary basis, CPO probabilities are a composite of slot boundary probabilities. As with the conditional collision probability, all event and outcome probabilities depend on the set of stations contending for access to the channel.

In networks using AIFS differentiation the set of contending CAFs may differ from one slot to the next within a contention period. In our previous discussion of the AIFS differentiation mechanism, we referred to groups of adjacent slots where the set of contending CAFs is fixed as *contention zones*. We do not treat the differences

in the set of contending CAFs between contention zones as changing the contention state. Rather, the effects of AIFS are accommodated by calculating event probabilities on a zone specific basis, conditioned on the contention state prevailing at the start of the contention period.

For slots in a given zone $h \in [1, n]$ the set of transmissions $\vec{y} \in \vec{x}$ that $\vec{q}$ may observe is indexed by $\hat{j} \in \left[1, \prod_{k=1}^{h} (M_k + 1) - 1\right]$. A transmission $\hat{j} \rightarrow \vec{y}$ is only possible in zone h if $y_k = 0$ for all $k > h$ and given a contention state $\vec{x}$ if $x_k - y_k \geq 0$ for all $k \in [1, h]$.

Because $\vec{q}$ may transmit, its behaviour affects the outcome of each contention period. For each set of CAFs $\vec{y}$ that may transmit there are in fact two possible transmission events: one where $\vec{q}$ joins the transmission and another where it observes idly. Events where $\vec{q}$ transmits may only occur in zones where $\vec{q}$ is active.

If the transmission $\vec{y}$ is possible, the probability that it occurs on a given slot boundary in zone h where the contention state $\vec{x}$ prevails is

$$p_{\text{set}}(\vec{y}|\vec{x}, h) = \prod_{k=1}^{h} \binom{x_k}{y_k} \tau_k^{y_k} (1 - \tau_k)^{x_k - y_k} . \tag{4.12}$$

The probability that $\vec{q}$ joins the transmission is

$$p_{\text{tx}}(\vec{y}|\vec{x}, h) = \begin{cases} \tau_{\vec{q}} \cdot p_{\text{set}}(\vec{y}|\vec{x}, h), & \text{if } h \geq j \\ 0, & \text{otherwise,} \end{cases} \tag{4.13}$$

and the probability that $\vec{q}$ observes the transmission is

$$p_{\text{obs}}(\vec{y}|\vec{x}, h) = \begin{cases} (1 - \tau_{\vec{q}}) \cdot p_{\text{set}}(\vec{y}|\vec{x}, h), & \text{if } h \geq j \\ p_{\text{set}}(\vec{y}|\vec{x}, h), & \text{otherwise,} \end{cases} \tag{4.14}$$

where j is the AC to which $\vec{q}$ belongs.

For a transmission event to be a CPO, that event must be the first event to occur in a contention period. That is, the probability that a given event is a CPO is the probability that zone h is reached in the contention period, multiplied by the sum of the probabilities that the outcome event occurs before any other event in that zone.

The probability of reaching a particular zone is the probability that there are not any transmissions in the preceding zones of the contention period. The probability that no stations transmit on a given slot boundary in zone h is

$$p_{\text{No tx}}(h|\vec{x}) = p_{\text{obs}}(\vec{0}|\vec{x}, h). \tag{4.15}$$

It follows that the probability of reaching zone h is

$$p_{\text{rz}}(1|\vec{x}) = 1, \tag{4.16}$$

$$p_{\text{rz}}(h|\vec{x}) = p_{\text{rz}}(h-1|\vec{x}) \cdot p_{\text{No tx}}(h|\vec{x})^{\text{AIFS}_h - \text{AIFS}_{h-1}}. \tag{4.17}$$

The probability that a particular transmission event precipitates the end of a contention period during zone h is then

$$\begin{aligned} P_{\text{tx}}(\vec{y}, h|\vec{x}) &= p_{\text{rz}}(h|\vec{x}) \cdot p_{\text{tx}}(\vec{y}|\vec{x}, h) \cdot \sum_{\ell=1}^{slots} p_{\text{No tx}}(h|\vec{x})^{\ell} \\ &= p_{\text{rz}}(h|\vec{x}) \cdot p_{\text{tx}}(\vec{y}|\vec{x}, h) \cdot \frac{\left(1 - p_{\text{No tx}}(h|\vec{x})^{slots}\right)}{\left(1 - p_{\text{No tx}}(h|\vec{x})\right)}, \text{ or} \end{aligned} \tag{4.18}$$

$$\begin{aligned} P_{\text{obs}}(\vec{y}, h|\vec{x}) &= p_{\text{rz}}(h|\vec{x}) \cdot p_{\text{obs}}(\vec{y}|\vec{x}, h) \cdot \sum_{\ell=1}^{slots} p_{\text{No tx}}(h|\vec{x})^{\ell} \\ &= p_{\text{rz}}(h|\vec{x}) \cdot p_{\text{obs}}(\vec{y}|\vec{x}, h) \cdot \frac{\left(1 - p_{\text{No tx}}(h|\vec{x})^{slots}\right)}{\left(1 - p_{\text{No tx}}(h|\vec{x})\right)}, \end{aligned} \tag{4.19}$$

where

$$\begin{aligned} slots &= \text{AIFS}_{h+1} - \text{AIFS}_h, && \text{if } h < \text{n} \\ slots &= \text{ACKTimeout/CTSTimeout} - \text{AIFS}_\text{n}, && \text{otherwise;} \end{aligned}$$

and ACKTimeout/CTSTimeout is $\lceil$(aSIFSTime + anACKTime + aSlotTime)/aSlotTime$\rceil$ (ACKTimeout and CTSTimeout are the same).

Once we have found the zone specific probability that a particular transmission event ends the contention period, the probability that the event is the contention period outcome is

$$P_{\text{tx}}(\vec{y}|\vec{x}) = \sum_{h=1}^{n} P_{\text{tx}}(\vec{y}, h|\vec{x}), \text{ or} \tag{4.20}$$

$$P_{\text{obs}}(\vec{y}|\vec{x}) = \sum_{h=1}^{n} P_{\text{obs}}(\vec{y}, h|\vec{x}), \tag{4.21}$$

where P_{tx} relates to CPOs where q joins the transmission, and P_{obs} relates to CPOS where q observes the transmission.

The probability that full contention is restored at the end of an ACKTimeout/CTSTimeout is equivalent to the probability that no transmissions occur before the timeout expires. Using (4.16) and (4.15) the probability of an ACKTimeout/CTSTimeout CPO for a given contention state $\vec{x}$ is

$$P(\text{timeout}|\vec{x}) = p_{\text{rz}}(n|\vec{x}) \cdot p_{\text{No Tx}}(n|\vec{x})^{\text{ACKTimeout}-\text{AIFS}_n}. \tag{4.22}$$

The relationship between contention period outcome and the next contention state

How the contention period outcome relates to the next contention state depends on whether $\vec{q}$ transmits as part of the outcome. If $\vec{q}$ merely observes the CPO then $\vec{y}$ completely determines the next contention state $\vec{z}$ seen by $\vec{q}$. When $\vec{y}$ contains only one CAF, that CAF's transmission is successful and all CAFs contend in the next contention period $\vec{z}$. When $\vec{y}$ contains more than one CAF, those CAFs' transmissions collide, and the transmitting CAFs observe ACKTimeout/CTSTimeout instead of contending in the next contention period. The foregoing rules are summarized below in (4.23).

$$\vec{z} = \begin{cases} \vec{M} & \text{if } \sum \vec{y} = 1, \\ \vec{M} - \vec{y} & \text{otherwise.} \end{cases} \tag{4.23}$$

Determining the next contention state is more complicated if $\vec{q}$ is part of the CPO. If the outcome is a transmission success for $\vec{q}$, then the next contention state will be full contention $\vec{M}$. But if the outcome is not a transmission success, $\vec{q}$ observes an ACKTimeout/CTSTimeout, and may observe transmissions during the *post-collision contention period* that occurs during that timeout, before returning to contention. Thus the next contention state seen by $\vec{q}$ is a result of both the original collision, and the outcome of the subsequent post-collision contention period.

The set of stations contending in the post-collision contention period is dictated by the set of stations involved in the precipitating collision, that is $\vec{zp} = \vec{M} - \vec{y}$.

Because $\vec{q}$ cannot transmit during the post-collision contention period, post-collision contention period event probabilities are prescribed by the transmission set probabilities $p_{\text{set}}(\vec{y}|\vec{x}, h)$ given by (4.12). Following the same analysis as for normal contention, the probability that no stations transmit on a given slot boundary in a post-collision

contention zone h is

$$p_{\text{No tx post}}(h|\vec{x}) = p_{\text{set}}(\vec{0}|\vec{x}, h), \tag{4.24}$$

and the probability of reaching post-collision zone h is

$$\begin{aligned} p_{\text{rzp}}(1|\vec{x}) &= 1, \\ p_{\text{rzp}}(h|\vec{x}) &= p_{\text{rzp}}(h-1|\vec{x}) \cdot p_{\text{No tx post}}(h|\vec{x})^{\text{AIFS}_h - \text{AIFS}_{h-1}}. \end{aligned} \tag{4.25}$$

The probability that a particular transmission event precipitates the end of a post-collision contention period during zone h is then

$$\begin{aligned} P_{\text{post}}(\vec{yp}, h|\vec{zp}) &= p_{\text{rz}}(h|\vec{zp}) \cdot p_{\text{set}}(\vec{yp}|\vec{zp}, h) \cdot \sum_{\ell=1}^{slots} p_{\text{No tx}}(h|\vec{zp})^{\ell} \\ &= p_{\text{rz}}(h|\vec{zp}) \cdot p_{\text{set}}(\vec{yp}|\vec{zp}, h) \cdot \frac{\left(1 - p_{\text{No tx}}(h|\vec{zp})^{slots}\right)}{(1 - p_{\text{No tx}}(h|\vec{zp}))}. \end{aligned} \tag{4.26}$$

where *slots* is defined as above.

The probability that a given event $\vec{yp}$ is a post-collision CPO is

$$P_{\text{post}}(\vec{yp}|\vec{zp}) = \sum_{h=1}^{n} P_{\text{post}}(\vec{yp}, h|\vec{zp}). \tag{4.27}$$

Similarly, the probability of a post-collision contention period ending with the expiry of ACKTimeout/CTSTimeout is

$$P(\text{post timeout}|\vec{zp}) = p_{\text{rz}}(n|\vec{zp}) \cdot p_{\text{No Tx}}(n|\vec{zp})^{\text{ACKTimeout} - \text{AIFS}_n}. \tag{4.28}$$

If a post-collision contention period is terminated by transmission $\vec{yp}$, the next contention state is

$$\vec{z} = \begin{cases} M, & \text{if } \sum_{k=1}^{n} \vec{yp}_k = 1 \\ M - \vec{yp}, & \text{otherwise.} \end{cases} \tag{4.29}$$

Each post-collision outcome contributes a probability of $P_{\text{tx}}(\vec{y}|\vec{x}) \cdot P_{\text{post}}(\vec{yp}|\vec{zp})$ to the element $p_{\hat{i},\hat{j}}$ of $\mathbb{P}$ for which $\hat{i}$ maps to $\vec{x}$ and $\hat{j}$ to $\vec{z}$, and outcome $\vec{y}$ gives rise to the post-collision contention state $\vec{zp}$.

Stationary distribution

After every element of $\mathbb{P}$ has been determined, the stationary distribution π can be found by repeatedly multiplying $\mathbb{P}$ by itself. Because $\sum \mathbb{P}_{\hat{i}} = 1$ the limit $\lim_{n\to\infty} \mathbb{P}^n$ must converge to a matrix satisfying $\pi = \pi\mathbb{P}$ where every identical row is π.

4.2.3 Using the Stationary Distribution of Contention States to find Average Conditional Collision Probability

Consider again the single CAF $\vec{q}$ from AC j contending for channel access in the context of a contention state $\vec{x}$. During the contention period for which $\vec{x}$ prevails, the subset of stations contending from $\vec{x}$ may differ from one contention zone to the next. If each CAF transmits on slot boundaries with uniform probability τ_k, the probability that a transmission in zone h from $\vec{q}$ collides is

$$p(\hat{i}|h) = p(\vec{x}|h) = 1 - \prod_{k=1}^{h} (1 - \tau_k)^{x_k} \tag{4.30}$$

From the above expression it is clear that the particular slot on which $\vec{q}$ transmits is crucial to collision probability. Every slot where $\vec{q}$ may transmit can be characterized by two properties: the prevailing contention state $\vec{x} \rightarrow \hat{i}$ and the zone h in which it occurs. Our calculation for the stationary distribution of contention states π gives the probability that a particular contention state prevails during a given slot. We now turn to finding the probability that a given slot occurs in a particular contention zone.

Contention zone stationary distribution

During a contention period, a slot boundary is reached only when no transmission occurs on the preceding slot boundary. Since the probability of passing through each slot is constant within each zone for a given contention state, we can use a Markov process to find the occupancy of backoff slots. Such a process is governed by the probability that no transmissions occur in a backoff slot, $p_{\text{No tx}}(h|\vec{x})$, and is illustrated by Fig. 4.2.

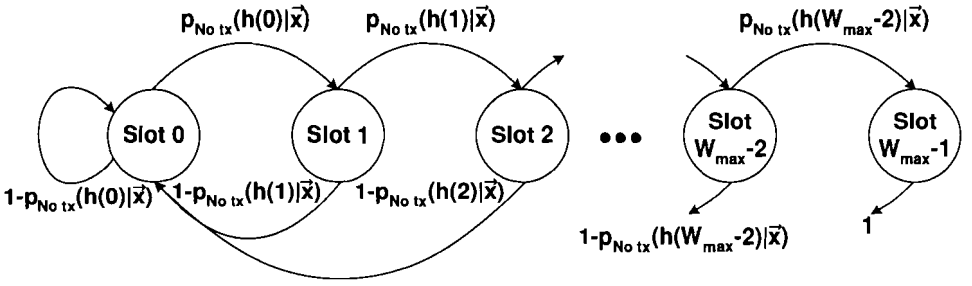


Figure 4.2: Relationship between adjacent timeslots in a transmission period.

The one-step transition probabilities for the backoff slot Markov process is

$$\begin{cases} P\{i|i-1\} = p_{\text{No tx}}(h(i-1)|\vec{x}) & i \in [1, \min(\mathrm{W}_m)] \\ P\{0|i\} = 1 - p_{\text{No tx}}(h(i)|\vec{x}) & i \in [0, \min(\mathrm{W}_m)] \end{cases}$$

The relationship between the occupancy of adjacent backoff slots when contention state $\vec{x}$ prevails is

$$b_{\text{slot}}(i|\vec{x}) = p_{\text{No tx}}(h(i-1)|\vec{x}) \cdot b_{\text{slot}}(i-1|\vec{x}) \qquad i \in [1, \min(\mathrm{W}_m)], \tag{4.31}$$

and

$$b_{\text{slot}}(0|\vec{x}) = \sum_{i=0}^{\min(\mathrm{W}_m)} (1 - p_{\text{No tx}}(h(i-1)|\vec{x})) \cdot b_{\text{slot}}(i|\vec{x}), \tag{4.32}$$

where $b_{\text{slot}}(i|\vec{x})$ is the occupancy the i^{th} backoff slot, $p_{\text{No tx}}(h(i-1)|\vec{x})$ is the probability that no transmission occurs in the zone h where the $i-1$ slot occurs, and $\min(\mathrm{W}_m)$ is the smallest maximum contention window size in the network.

Since the maximum number of backoff slots between successive transmissions is strictly bounded by the smallest contention window in the system we can fix the number of terms in the stationary distribution. Because the sizes of contention windows are increased after collisions, we select the smallest maximum window size ($\min(\mathrm{W}_m)$) in the system to bound the stationary distribution. With this approximation the solution to the stationary distribution yielded by imposing the normalization condition is

$$b_{\text{slot}}(0|\vec{x}) = \frac{1}{1 + \sum_{i=1}^{\min(\mathrm{W}_m)-1} \prod_{j=0}^{i} (1 - p_{\text{No tx}}(h(j)|\vec{x}))}. \tag{4.33}$$

Slot occupancy can be translated into zone occupancy by adding together the occupancy of the slots that comprise each zone, yielding

$$b_{\text{zone}}(1|\vec{x}) = \sum_{i=0}^{\text{AIFS}_2-\text{AIFS}_1-1} b_{\text{slot}}(i|\vec{x}), \tag{4.34}$$

$$b_{\text{zone}}(h|\vec{x}) = \sum_{i=\text{AIFS}_h-\text{AIFS}_1}^{\text{AIFS}_{h+1}-\text{AIFS}_1-1} b_{\text{slot}}(i|\vec{x}), \qquad \text{if } 1 < h < n, \tag{4.35}$$

$$b_{\text{zone}}(n|\vec{x}) = \sum_{i=\text{AIFS}_n-\text{AIFS}_1}^{\min(\text{W}_m)-\text{AIFS}_1-1} b_{\text{slot}}(i|\vec{x}). \tag{4.36}$$

$$\tag{4.37}$$

Using π and $b_{\text{zone}}(h|\vec{x})$ we can express the probability that a slot on which $\vec{q}$ transmits is characterized by the pair $(\hat{i}, h)$ as

$$p_{\text{slot}}(\hat{i}, h) = \pi_{\hat{i}} \cdot \frac{b_{\text{zone}}(h|\hat{i})}{\sum_{k=j}^{n} b_{\text{zone}}(k|\hat{i})}, \tag{4.38}$$

where $\pi_{\hat{i}}$ is the stationary distribution of contention states and

$$\frac{b_{\text{zone}}(h|\hat{i})}{\sum_{k=j}^{n} b_{\text{zone}}(k|\hat{i})}$$

is the normalized slot occupancy for a CAF $\vec{q}$ belonging to AC j for transmission slots when the contention state $\vec{x} \rightarrow \hat{i}$ prevails.

Average conditional collision probability can now be found using (4.30) and (4.38):

$$\bar{p} = \sum_{\hat{i}=1}^{\hat{M}} \sum_{h=1}^{n} p_{\text{slot}}(\hat{i}, h) \cdot p(i|h). \tag{4.39}$$

It is now possible to use numerical techniques to solve the set of non-linear equations defined by (4.7) and (4.39).

Chapter 5

Delay Model

In this chapter, we extend (and correct) the signal transfer function approach to delay analysis pioneered in (Zhai and Fang 2003), and reviewed in Chapter 3. We start by explaining how linear systems techniques can be used to generate delay probability functions from discrete-time Markov Chains. We then develop a signal transfer model for EDCA transmission delay. We end by explaining how to reduce the model to success and failure paths, and how to generate a cumulative distribution function for transmission delay.

5.1 The Analogy of Linear Systems to Markov Chains

Consider the diagram below (Figure 5.1 representing a simple CAF backoff, where each branch is labelled with a transition probability. The DTMC shown has just two backoff stages, and each stage has just two backoff states. Now imagine a frame starting at the top of the diagram and taking any one of a number of paths to eventual success or failure. The probability that a frame traverses a particular path from start to end is the product of each branch probability in the path.

Now consider the diagram in figure 5.2 as representing an electrical network, where each branch is labelled with a signal gain and the z operators that indicate that the signal is to be delayed by a unit time. In the above figure, we can trace the progress of a signal through the network, one step (z) at a time. The signal in any branch is the product of the branch gains feeding into that branch.

The parallels between Figures 5.1 and 5.2 are obvious. The signal gain of each

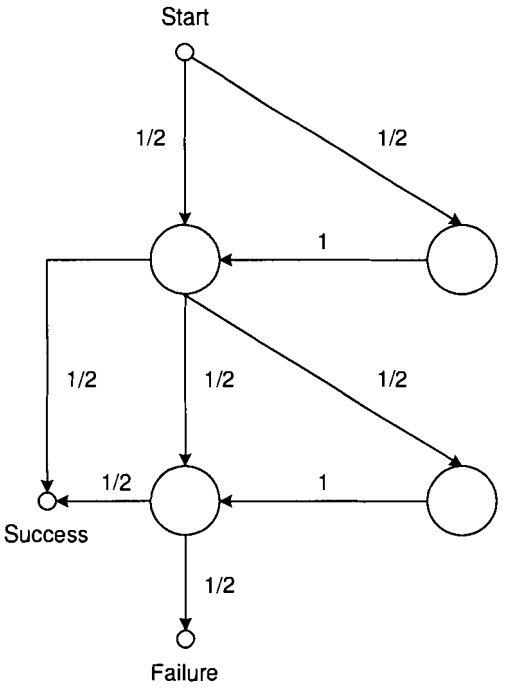


Figure 5.1: Simple CAF backoff

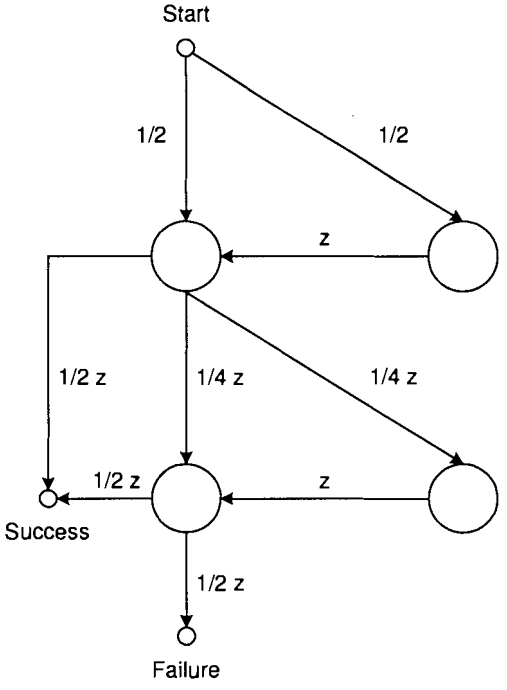


Figure 5.2: Simple electrical network

branch is analogous to transition probability and the signal at each node is analogous to the probability of finding the backoff process in that state. Each time a signal traverses a branch in the electrical network it suffers a unit delay z; in the Markov process, this unit delay is instead an observed slot edge. If we multiply signal gains along a path in the the electrical network, we find that the exponent of the z variable is equal to the number of unit delays, and analogously the number of slot edges observed, along the path.

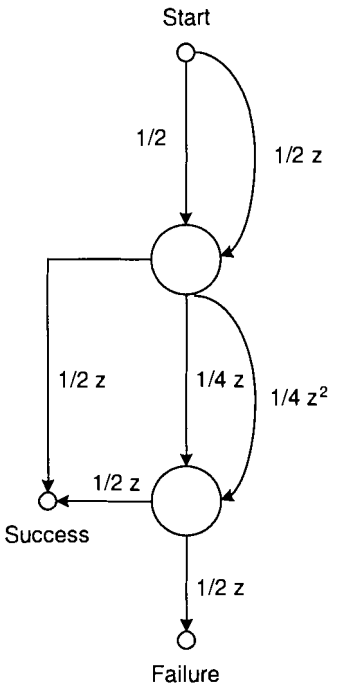
To find the probabilities of every possible transmission delay from start to success, we could traverse each possible path, tracking the total delay and net probability for each path. But this method is time consuming, and clearly would not scale to graphs of the size required to model IEEE 802.11e EDCA backoffs. Fortunately, signal flow graph reduction techniques can be used to reduce the graph to a single transfer function that describes this set of delays. By multiplying together the gains of series branches, and adding together the gains of parallel branches, the reduction shown in the Figures below results in the Figure shown in 5.3.

The transfer function for successful transmission in this case is $\frac{1}{4}z+\frac{5}{16}z^2+\frac{1}{8}z^3+\frac{1}{16}z^4$. This means that any given frame has a one in four chance of being transmitted successfully after one slot event, a 5 in 16 chance of being transmitted successfully after two slot events, a 1 in 8 chance of being successfully transmitted after three slot events, and a 1 in 15 chance of being successfully transmitted after four slot events.

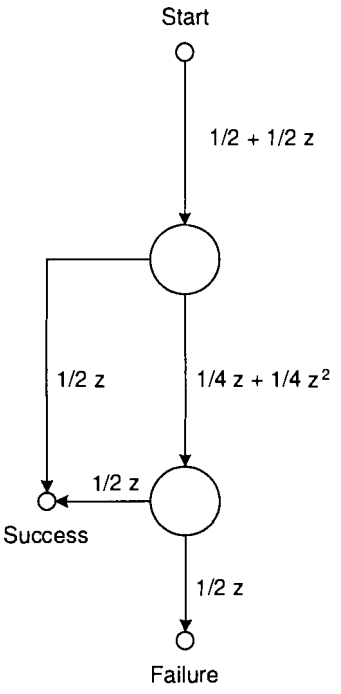
5.1.1 Non-Uniform Event Spacing

The transfer function in the example above expresses the transmission delay in terms of slot edges seen. From our work on the backoff process DTMC in chapter 4 we know that the real time between two successive slot edges varies according to the transmission event (or non-event) that occurs on the first slot edge. Thus a single unit of time z is inadequate to express the range of possible delays between adjacent backoff states.

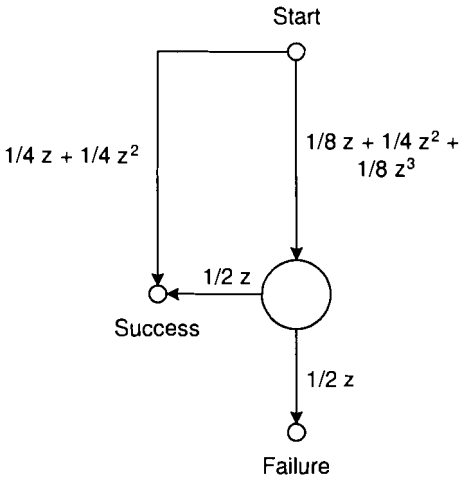
The solution to the problem of slot events occurring at irregular intervals is to tag each branch with a z polynomial that reflects the probabilities of the real world delays, denominated in terms of the base z, suffered when traversing that branch. For modeling delays in EDCA, the logical choice for the delay base z is the duration of a



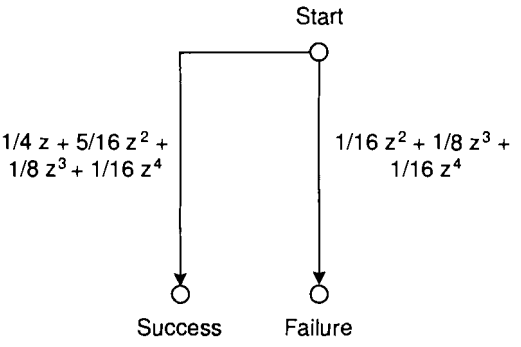
(a) Series branches multiplied together



(b) Parallel branches added together



(c) Series branches multiplied together



(d) Fully reduced signal flow graph

Figure 5.3: Signal flow graph reduction

single backoff slot, σ, since it is the smallest interval that may separate two successive slot edges. The duration exponent of all other events is then the ceiling of the quotient of the event duration divided by the slot duration: $\left\lceil \frac{T^{\text{event}}}{\sigma} \right\rceil$.

Consider a homogeneous network composed of N stations. A CAF $\vec{q}$ in this network contends for channel access against a set $\vec{x}$ of $N-1$ identical CAFs. On any slot edge where $\vec{q}$ is not transmitting, it will observe one of three events:

- an idle timeslot, duration σ
- a successful transmission, duration T^{suc}, or
- a collision between two transmissions, duration T^{col}.

Because the probability of each event is conditioned on $\vec{q}$ being in a non-transmission state, event probabilities are the same as the transmission set probabilities $p_{\text{set}}(\vec{y}|\vec{x},h)$ given in (4.12). Since all CAFs in this example are identical, the zone condition is irrelevant and event probability $p_{\text{set}}(\vec{y}|\vec{x})$ needs to be averaged over only contention vectors ($\vec{x} \rightarrow i$):

$$p_{\text{event}}(\vec{y}) = \pi \left[p_{\text{set}}(\vec{y}|0) \quad p_{\text{set}}(\vec{y}|1) \ldots p_{\text{set}}(\vec{y}|M_{\max}-1) \right]^T . \tag{5.1}$$

The z polynomial for branches between adjacent backoff states is then

$$H_d(z) = p_{\text{event}}([0])z^{\sigma} + p_{\text{event}}([1])z^{\left\lceil \frac{T^{\text{suc}}}{\sigma} \right\rceil} + \left(\sum_{k=2}^{N-1} p_{\text{event}}([k]) \right) z^{\left\lceil \frac{T^{\text{col}}}{\sigma} \right\rceil}, \tag{5.2}$$

where the terms represent an idle slot, a successful transmission and a collision, respectively.

5.1.2 Effect of AIFS on Delay

In both of the examples above the backoff process DTMC advances on every slot edge 'seen', so that only one event may occur between successive advances. But in networks using AIFS differentiation lower priority CAFs waiting out AIFS may observe transmission events from higher priority CAFs. This means that the lower priority CAFs may observe more than one event between successive backoff operations. Figure 5.4 illustrates this phenomenon.

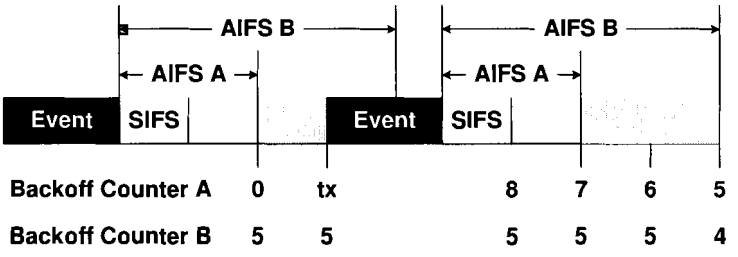


Figure 5.4: Lower priority AC observes event without backoff advance

In terms of the delay model, events that occur during AIFS add delay but do not affect the backoff state. The way to model this effect is to calculate event probability on a zone specific basis, and add those that occur before AIFS expiry to branch loops.

5.1.3 Effect of Different Transmission Lengths on Event Duration

In the homogeneous network example above, all collisions had the same duration. In heterogeneous networks where transmission durations differ among ACs the duration of a collision is the duration of the longest transmission in the collision. If the collision durations for all ACs is represented by the n-tuple $\vec{T^{\text{col}}}$, then the duration of a collision involving CAFs $\vec{y}$ is $\max\left(\vec{T^{\text{col}}} \circ (\vec{y} > 0)\right)$, where $(\vec{y} > 0)$ is a vector whose i^{th} element is one if $y_i > 0$ and zero otherwise.

5.2 Backoff State Signal Flow Graph

In this section we develop a backoff state signal flow graph to characterize the delay suffered by a CAF $\vec{q}$ between adjacent backoff states where the CAF does not transmit. Let $\vec{q}$ belong to AC $j \in n$.

5.2.1 Backoff State Event probabilities

Recall from the determination of average conditional collision probability the probability $p_{\text{slot}}(\hat{i}, h)$ that a given slot is characterized by a particular contention state and

contention zone pair $(\hat{i}, h)$. The probability that $\vec{q}$ observes a particular event $\vec{y}$ on any given slot is

$$p_{\text{event}}(\vec{y}, \hat{i}, h) = p_{\text{slot}}(\hat{i}, h) \cdot p_{\text{set}}(\vec{y}|\hat{i}, h). \tag{5.3}$$

For convenience we define the following probabilities for events with common durations:

$$p_{\text{idle}}(h) = \sum_{\hat{i}=0}^{\hat{M}} p_{\text{event}}\left(\vec{0}, i, h\right), \tag{5.4}$$

$$p_{\text{suc}}(k, h) = \sum_{\hat{i}=0}^{\hat{M}} p_{\text{event}}(\vec{y}, i, h), \qquad y_k = 1, y_j = 0 \forall j \neq k$$

$$p_{\text{col}}(k, h) = \sum_{\hat{i}=0}^{\hat{M}} \sum_{\mathcal{C}} p_{\text{event}}(\vec{y}, i, h),$$

where $\mathcal{C}$ is the set of observed collision events whose duration is determined by a transmission from AC k; that is

$$\mathcal{C} = \left\{ \vec{y} \middle| \sum_{i=1}^{n} y_i > 1, \max\left(\vec{T^{\text{col}}} \circ (\vec{y} > 0)\right) = T_k^{\text{col}} \right\} \tag{5.5}$$

The above probabilities correspond to an idle slot occurring during contention zone h, a successful transmission from a CAF belonging to AC k in zone h, and a collision whose duration is determined by a CAF belonging to AC k.

The backoff state signal flow graph has two branches. The loop branch is composed of all events that may occur before $\vec{q}$ finished observing AIFS. The forward branch is composed of all events that may occur after $\vec{q}$ has finished observing AIFS. Figure 5.5 depicts the backoff state signal flow graph.

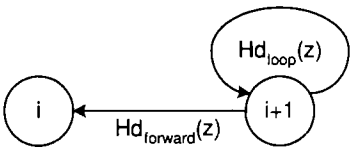


Figure 5.5: Backoff State Signal Flow Graph

In terms of the events defined in (5.4) the z polynomial for the loop branch is

$$Hd_{\text{loop}}(z) = \sum_{h=1}^{j-1} \left\{ p_{\text{idle}}(h) z^{\sigma} + \sum_{k=1}^{h} \left[p_{\text{suc}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{suc}}}{\sigma} \right\rceil} \right] + \sum_{k=1}^{h} \left[p_{\text{col}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{col}}}{\sigma} \right\rceil} \right] \right\}, \tag{5.6}$$

and the z polynomial for the forward branch connecting adjacent backoff states is

$$Hd_{\text{forward}}(z) = \sum_{h=j}^{n} \left\{ p_{\text{idle}}(h)\, z^{\sigma} + \sum_{k=1}^{h} \left[p_{\text{suc}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{suc}}}{\sigma} \right\rceil} \right] + \sum_{k=1}^{h} \left[p_{\text{col}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{col}}}{\sigma} \right\rceil} \right] \right\}. \tag{5.7}$$

The above backoff signal flow graph can be reduced to a single branch between adjacent states with transmission $Hd(z)$ given by

$$Hd(z) = \frac{Hd_{\text{forward}}(z)}{1 - Hd_{\text{loop}}(z)}. \tag{5.8}$$

5.3 Transmission State Signal Flow Graph

Once $\vec{q}$ reaches a transmit state, it will transmit on the first slot edge to occur after it finishes observing AIFS. This transmission event must result in either success or failure. On any slot edges that occur before AIFS expires $\vec{q}$ suffers the delays of those events without transmitting. Thus the transmit state signal flow graph has three branches, as shown in 5.6.

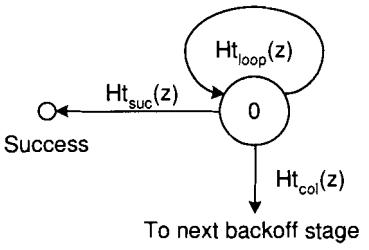


Figure 5.6: Transmit State Signal Flow Graph

The transmission for the loop branch tracking these events is defined in the same

way as the loop branch for the backoff state, that is

$$Ht_{\text{loop}}(z) = \sum_{h=1}^{j-1} \left\{ p_{\text{idle}}(h) z^{\sigma} + \sum_{k=1}^{h} \left[p_{\text{suc}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{suc}}}{\sigma} \right\rceil} \right] + \sum_{k=1}^{h} \left[p_{\text{col}}(k,h)\, z^{\left\lceil \frac{T_k^{\text{col}}}{\sigma} \right\rceil} \right] \right\}. \tag{5.9}$$

On the slot edge where $\vec{q}$ transmits, one of four outcomes is possible:

- $\vec{q}$ transmits successfully
- $\vec{q}$ experiences a collision, and observes a successful transmission during the post-collision contention period before returning to contention
- $\vec{q}$ experiences a collision, and observes a collision during the post-collision contention period before returning to contention
- $\vec{q}$ experiences a collision, and observes an idle medium for the duration of ACK-Timeout/CTSTimeout before returning to contention

Probabilities for each of the above events can be determined by combining the expressions for $p_{\text{event}}(\vec{y}, i, h)$ 5.3 and $P_{\text{post}}(\vec{yp}|\vec{zp})$ 4.27.

The probability that $\vec{q}$ transmits successfully is

$$p_{\text{tx suc}}(h) = \sum_{\hat{i}=0}^{\hat{M}} p_{\text{event}}\left(\vec{0}, \hat{i}, h\right). \tag{5.10}$$

The probability that $\vec{q}$ transmitting in zone h experiences a collision with duration limited by a transmission from a CAF belonging to AC k, and then during the post collision period observes a successful transmission from a CAF belonging to AC j is

$$p_{\text{tx col/suc}}(k,j,h) = \sum_{\hat{i}=0}^{\hat{M}} \sum_{\mathcal{F}} p_{\text{event}}\left(\vec{y}, \hat{i}, h\right) \cdot P_{\text{post}}\left(\vec{yp}, \vec{zp}\right), \tag{5.11}$$

where $\mathcal{F}$ is the set of transmission failure events whose duration is determined by a transmission from AC k; that is

$$\mathcal{F} = \left\{ \vec{y} \,\middle|\, \vec{y} \neq \vec{0}, \max\left(\vec{T^{\text{col}}} \circ (\vec{y} + \vec{q} > 0)\right) = T_k^{\text{col}} \right\}, \tag{5.12}$$

and $yp_j = 1, yp_\ell = 0 \forall \ell \neq j$ and $\vec{zp}$ is the post-collision contention vector corresponding to the particular transmission set $\vec{y}$.

The probability that $\vec{q}$ transmitting in zone h experiences a collision with duration limited by a transmission from a CAF belonging to AC k, and then during the post collision period observes a collision whose duration is determined by a transmission from a CAF belonging to AC j is

$$p_{\text{tx col/col}}(k, j, h) = \sum_{\hat{i}=0}^{\hat{M}} \sum_{\mathcal{F}} \sum_{\mathcal{C}} p_{\text{event}}\left(\vec{y}, \hat{i}, h\right) \cdot P_{\text{post}}\left(\vec{yp}, \vec{zp}\right), \tag{5.13}$$

where $\mathcal{F}$ is the set of transmission failure events whose duration is determined by a transmission from AC k as defined above, and $\mathcal{C}$ is the set of observed collision events whose duration is determined by a transmission from AC k; that is

$$\mathcal{C} = \left\{ \vec{yp} \middle| \sum_{i=1}^{n} yp_i > 1, \max\left(\vec{T^{\text{col}}} \circ (\vec{yp} > 0)\right) = T_j^{\text{col}} \right\}. \tag{5.14}$$

The probability that $\vec{q}$ transmitting in zone h experiences a collision with duration limited by a transmission from a CAF belonging to AC k, and then observes an idle the post collision period is

$$p_{\text{tx col/timeout}}(k, j, h) = \sum_{\hat{i}=0}^{\hat{M}} \sum_{\mathcal{F}} p_{\text{event}}\left(\vec{y}, \hat{i}, h\right) \cdot P\left(\text{post timeout} | \vec{zp}\right), \tag{5.15}$$

where $\mathcal{F}$ is the set of transmission failure events whose duration is determined by a transmission from AC k as defined above.

The transmission for the transmit success branch in terms of the above probabilities is

$$\widehat{Ht_{\text{suc}}(z)} = \sum_{h=j}^{n} \left\{ p_{\text{tx suc}}(h) z^{\left\lceil \frac{T_{\vec{q}}^{\text{suc}}}{\sigma} \right\rceil} \right\}, \tag{5.16}$$

and the transmission for the transmit collision branch in terms of the above probabilities is

$$\widehat{Ht_{\text{col}}(z)} = \sum_{h=j}^{n} \left\{ \sum_{k=1}^{h} \sum_{j=1}^{h} \left[p_{\text{tx col/suc}}(k,j,h)\, z^{\left\lceil \frac{T_k^{\text{col}} + T_j^{\text{suc}}}{\sigma} \right\rceil} \right] + \sum_{k=1}^{h} \sum_{j=1}^{h} \left[p_{\text{tx col/col}}(k,j,h)\, z^{\left\lceil \frac{T_k^{\text{col}} + T_j^{\text{col}}}{\sigma} \right\rceil} \right] + \sum_{k=1}^{h} \left[p_{\text{tx col/timeout}}(k,j,h)\, z^{\left\lceil \frac{T_k^{\text{col}} + \text{ACKTimeout}}{\sigma} \right\rceil} \right] \right\}. \quad (5.17)$$

In the above expressions backoff slots observed during the post-collision contention period are not included in the duration of post-collision events. Because the backoff is short relative to transmission duration this omission has a negligible effect on accuracy.

$Ht_{\text{loop}}(z)$ in the transmit state signal flow graph can be eliminated by incorporating it into the forward branches

$$Ht_{\text{suc}} = \frac{\widehat{Ht_{\text{suc}}}}{1 - Ht_{\text{loop}}}, \quad (5.18)$$

$$Ht_{\text{col}} = \frac{\widehat{Ht_{\text{col}}}}{1 - Ht_{\text{loop}}}. \quad (5.19)$$

5.4 Complete Signal Flow Graph

The complete signal flow graph shown in 5.7 is constructed by combining the signal flow graphs for backoff and transmission states. Figure 5.7 clearly designates the start and end points of the signal transfer function for delay. Using common linear systems techniques we can reduce the transition diagram into a signal transfer function from the start state to the end state.

5.4.1 Signal Flow Graph Reduction

First all backoff states within each stage are combined,

$$HW_i(z) = \begin{cases} \sum_{j=0}^{2^i W_0 - 1} \frac{H_d^j(z)}{2^i W_0}, & 0 \le i \le m \\ HW_m(z), & m < i \le f, \end{cases}$$

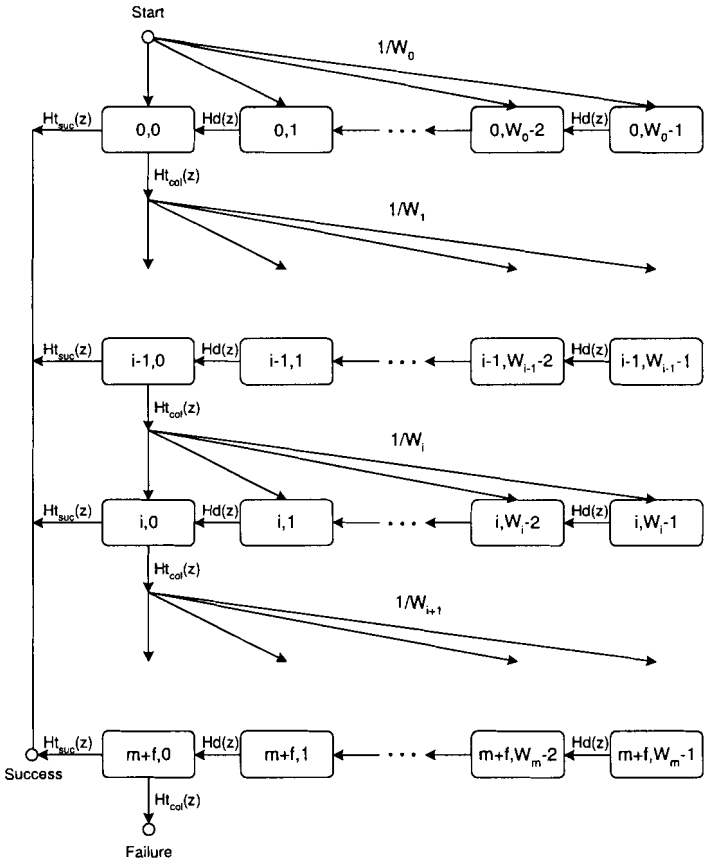


Figure 5.7: Generalized state transition diagram

then the backoff paths to leading to each stage,

$$H_i(z) = \prod_{j=0}^{i} HW_j(z), \qquad 0 \leq i \leq f$$

and finally the paths to successful transmission,

$$B(z) = Ht_{\text{suc}}(z) \sum_{i=0}^{m+f} \left(Ht_{\text{col}}(z)\right)^i H_i(z) + \left(Ht_{\text{col}}(z)\right)^{m+f+1} H_{m+f}(z), \tag{5.20}$$

to yield $B(z)$, the probability generating function for service delay. The CDF for service delay is constructed by expanding B(z) into a power series,

$$B(z) = \sum_{i=0}^{\infty} P(T^{\text{service}} = i) z^i.$$

5.4.2 Solutions to Practical Problems with Signal Flow Graph Reduction

Here we address two practical problems that arise when reducing a graph of this size. The first problem is that the polynomial division used in the reduction of transmission loops leads to infinitely long series of terms. Fortunately, these series are convergent and so can be terminated when the coefficients become small. But care must be taken not to terminate the series prematurely, so that the sum of the coefficients in the z-polynomial is not less than the probability of the corresponding branch in the Markov process.

The second problem is that the repeated polynomial multiplication used to reduce series branches gives rise to an arithmetically increasing number of terms (multiplying two univariate polynomials of degrees n and m results in a polynomial of degree $n+m$). At some point the number of terms must overwhelm the available computing resources available. Unlike the polynomial division the resulting series is not convergent and cannot simply be terminated.

To overcome this problem we turn to numerosity reduction techniques from statistics and database applications. The simple and effective method we adopt is equal frequency binning. After each polynomial multiplication operation, the intermediate transfer function polynomial is split into groups of adjacent terms, each group having roughly the same probability. The terms in each group are then represented by a single probability and delay value in a new reduced polynomial. In our implementation, we use a weighted average of the delays in a bin to represent the group. In generating results from our model we chose to terminate polynomial division at 10,000 terms, and reduce intermediate polynomials to 1000 terms.

This process is intuitive if we consider each intermediate polynomial as a probability density function for delay. Combining together delays which are similar reduces the number of terms, but hardly affects the distribution at a macroscopic level. Figure 5.8 illustrates the effect of equal frequency binning on a polynomial by showing before and after probability density plots. In this figure, the histogram bars represent individual terms in polynomial whose exponents are delay and whose coefficient are probability. We can see that the more dense the histogram bars are, that is the greater

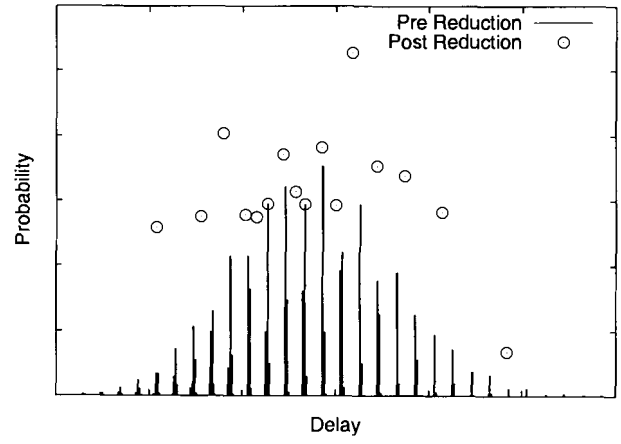


Figure 5.8: Probability density plots showing polynomial reduction using equal frequency binning

their magnitude and the smaller their spacing, the more dense the representative binning circles are. Figure 5.9 below shows the same polynomials as cumulative density plots. From this figure we can see how equal frequency binning removes the more detailed features of the CDF, but maintains the overall shape of the distribution.

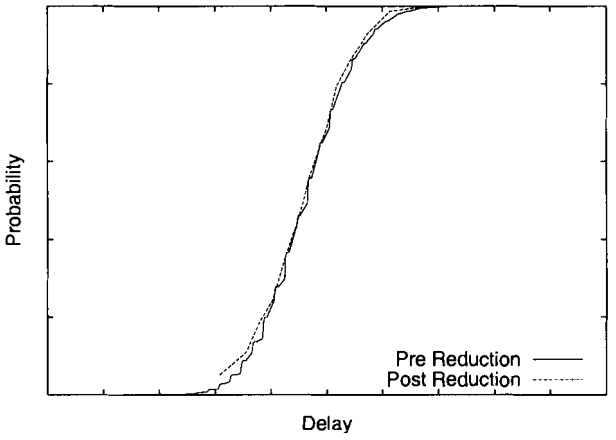


Figure 5.9: Cumulative density plots showing polynomial reduction using equal frequency binning

Chapter 6

Model Validation

The accuracy of the backoff and delay models presented here have been validated by comparisons with results of simulations conducted using the ns-2 simulator. The ns-2 EDCA implementation used was created by IEEE 802.11 TGe member Mike Moreton. We have examined this implementation of EDCA in minute detail, and are satisfied that it models EDCA correctly and has corrected the flaws of the legacy ns-2 802.11 implementation.

6.1 Simulation Parameters

In all simulations presented here, all transmitting stations contend to transmit fixed size UDP packets to a single non-transmitting station (i.e. an access point). Two sets of results are shown: one where stations use the basic transmission mechanism, the other where stations use the RTS/CTS transmission mechanism. Saturation conditions are created by high rate Constant Bit Rate traffic generators. All stations are configured according to DSSS PHY parameters and AC parameters suggested in the 802.11e draft standard. These parameters are listed for convenience in Table 6.1.

6.2 Model Parameters

We have implemented model developed in chapters 4 and 5 in MATLAB. The AC parameters used are the same as those listed in Table 6.1. The event duration constants

Table 6.1: DSSS System Parameters and Access Category Parameters Used In Simulation

Frame payload	8000 bits
MAC header	224 bits
PHY header	192 bits
ACK	112 bits + PHY header
RTS	160 bits + PHY header
CTS	112 bits + PHY header
Channel bitrate	1 Mbit/s
Propagation delay	1 μs
Slot time	20 μs
SIFS	10 μs
ShortRetryLimit	7
AIFSD[4]	SIFS + 2 $\times$ Slot time
AIFSD[3]	SIFS + 2 $\times$ Slot time
AIFSD[2]	SIFS + 3 $\times$ Slot time
AIFSD[1]	SIFS + 7 $\times$ Slot time
$CW_{min}[4]$	7
$CW_{min}[3]$	15
$CW_{min}[2]$	31
$CW_{min}[1]$	31
$CW_{max}[4]$	15
$CW_{max}[3]$	31
$CW_{max}[2]$	1023
$CW_{max}[1]$	1023

of sections 5.2 and 5.3 are defined as

$$T^{\text{suc}} = \min(\text{AIFSD}) + \text{SIFS} + \frac{\text{PHY header} + \text{MAC header} + \text{Frame payload} + \text{ACK}}{\text{Channel bitrate}} \tag{6.1}$$

$$T^{\text{col}} = \min(\text{AIFSD}) + \frac{\text{Frame payload}}{\text{Channel bitrate}} \tag{6.2}$$

for transmissions using the basic transmission mechanism, and as

$$T^{\text{suc}} = \min(\text{AIFSD}) + 3 \cdot \text{SIFS} + \frac{\text{RTS} + \text{CTS} + \text{PHY header} + \text{MAC header} + \text{Frame payload} + \text{ACK}}{\text{Channel bitrate}} \tag{6.3}$$

$$T^{\text{col}} = \min(\text{AIFSD}) + \frac{\text{RTS}}{\text{Channel bitrate}} \tag{6.4}$$

for transmissions using the RTS/CTS transmission mechanism. Figure 6.1 illustrates the event duration constants.

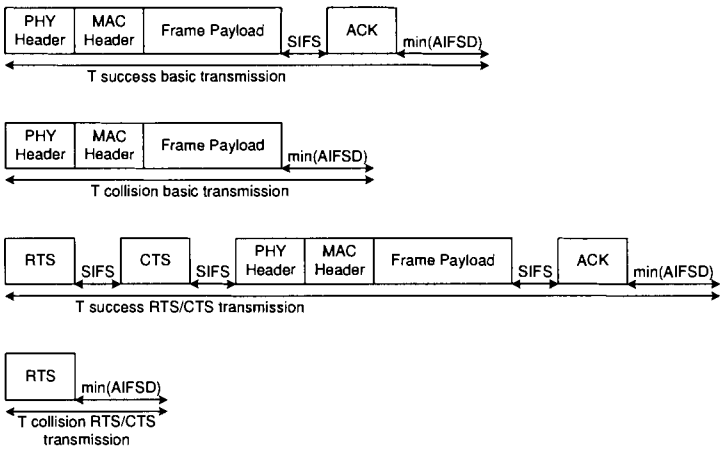


Figure 6.1: Transmission event timing

6.3 Statistical Methods

Statistical methods were used in the generation of collision probability results. Confidence intervals for collision probability were calculated using the formula provided in (Jain 1991). Given that n_1 of n observations are of one type, a confidence interval

for the proportion is obtained as follows:

$$\text{Sample proportion} = p = \frac{n_1}{n}$$

$$\text{Confidence interval for proportion} = p \mp z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \tag{6.5}$$

Here, $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of a unit normal variate. For a confidence interval at 95%, $\alpha = 0.05$ and $z_{1-\alpha/2} = 1.960$.

6.4 Collision Probability Results

Collision probability is the same no matter which transmission mechanism (basic or RTS/CTS) is used. Tables 6.2, 6.3 and 6.4 show comparisons of collision probability estimates from the model and results from simulation. Comparisons are shown for symmetrically increasing loads of CAFs from neighbouring ACs.

Table 6.2: Comparison of model and simulation collision probability for symmetrically increasing load of AC 4 and AC 3 stations

Load	AC 4 Model $\bar{p}$	AC 4 Simulation $\bar{p}$	AC 3 Model $\bar{p}$	AC 3 Simulation $\bar{p}$
5	0.60135	0.60012 ± 0.003814	0.62441	0.62436 ± 0.00509
10	0.83149	0.83235 ± 0.00736	0.84060	0.84140 ± 0.00969
15	0.92954	0.92956 ± 0.00564	0.93333	0.93322 ± 0.00744

Table 6.3: Comparison of model and simulation collision probability for symmetrically increasing load of AC 3 and AC 2 stations

Load	AC 3 Model $\bar{p}$	AC 3 Simulation $\bar{p}$	AC 2 Model $\bar{p}$	AC 2 Simulation $\bar{p}$
5	0.36241	0.36383 ± 0.00481	0.43001	0.43456 ± 0.00990
10	0.54721	0.54698 ± 0.00403	0.62824	0.63376 ± 0.00998
15	0.66584	0.66594 ± 0.00357	0.74908	0.75388 ± 0.00998

Table 6.4: Comparison of model and simulation collision probability for symmetrically increasing load of AC 2 and AC 1 stations

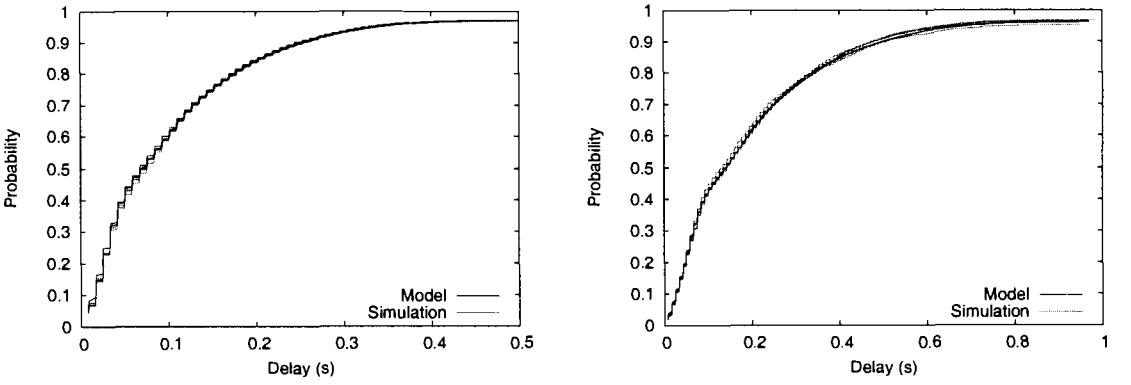
Load	AC 2 Model $\bar{p}$	AC 2 Simulation $\bar{p}$	AC 1 Model $\bar{p}$	AC 1 Simulation $\bar{p}$
5	0.21466	0.21347 ± 0.00404	0.31088	0.31205 ± 0.00976
10	0.32409	0.32390 ± 0.00372	0.44993	0.45249 ± 0.01341
15	0.40306	0.40278 ± 0.00421	0.53315	0.53136 ± 0.01993

The comparisons reveal that the model is impressively accurate; this not surprising, since it models the backoff process exactly.

6.5 Service Delay Distribution Results

Comparisons of model and simulation results for the service delay distribution are presented in two sections. The first section contains results for networks where stations use the basic transmission mechanism, and the second for networks where stations use the RTS/CTS transmission mechanism. Like the collision results comparison, the comparisons are shown for symmetrically increasing loads of CAFs from neighbouring ACs. As can be seen from the figures, the agreement between the model results and simulation results is remarkable.

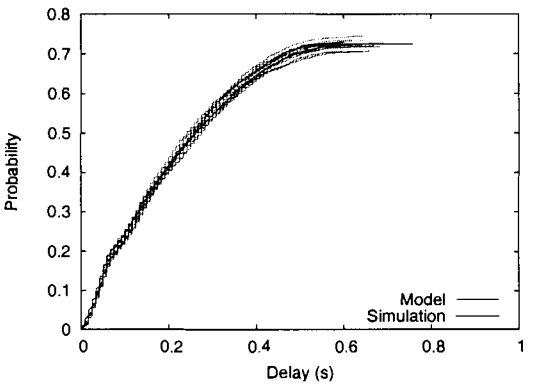
6.5.1 Delay Distribution for Results Using Basic Transmission Mechanism



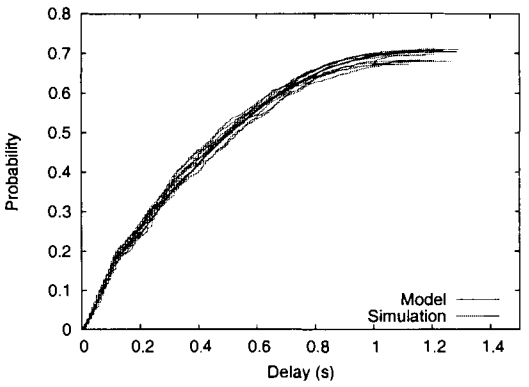
(a) Access Category 4 Delay Distribution

(b) Access Category 3 Delay Distribution

Figure 6.2: Simulation and Model Results for 5 stations using basic transmission mechanism from each of Access Categories 4 and 3

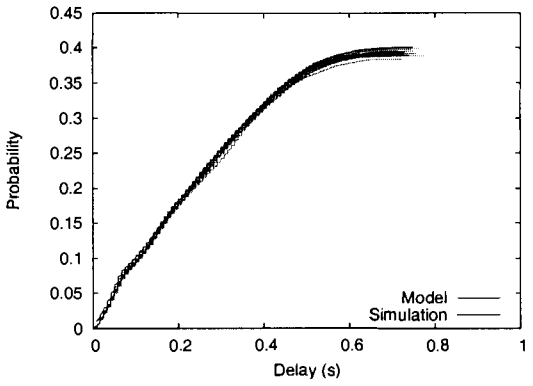


(a) Access Category 4 Delay Distribution

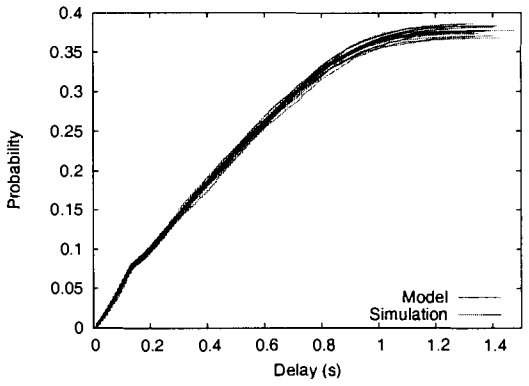


(b) Access Category 3 Delay Distribution

Figure 6.3: Simulation and Model Results for 10 stations using basic transmission mechanism from each of Access Categories 4 and 3

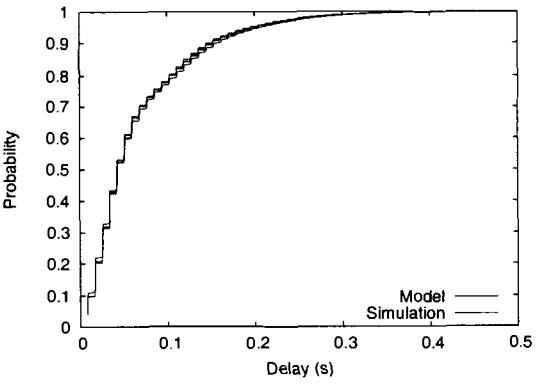


(a) Access Category 4 Delay Distribution

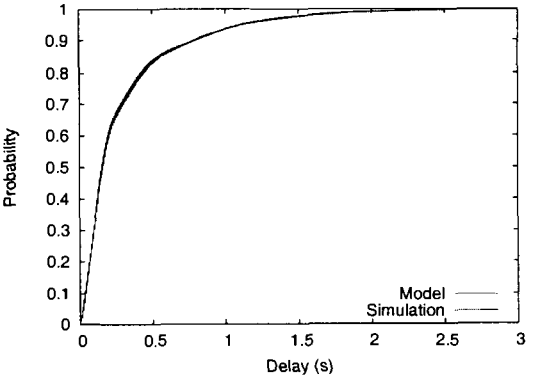


(b) Access Category 3 Delay Distribution

Figure 6.4: Simulation and Model Results for 15 stations using basic transmission mechanism from each of Access Categories 4 and 3

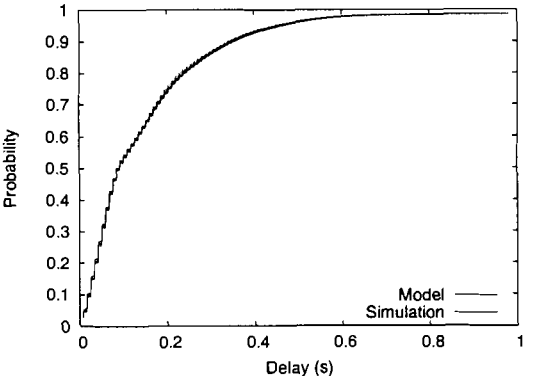


(a) Access Category 3 Delay Distribution

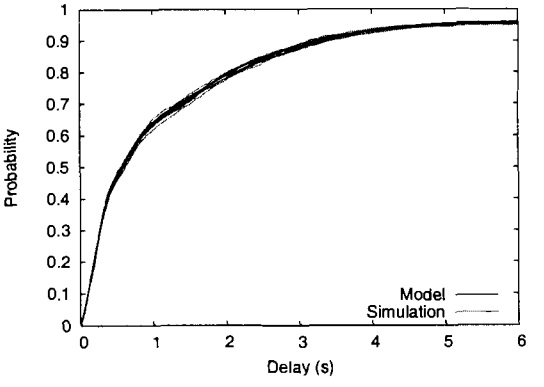


(b) Access Category 2 Delay Distribution

Figure 6.5: Simulation and Model Results for 5 stations using basic transmission mechanism from each of Access Categories 3 and 2

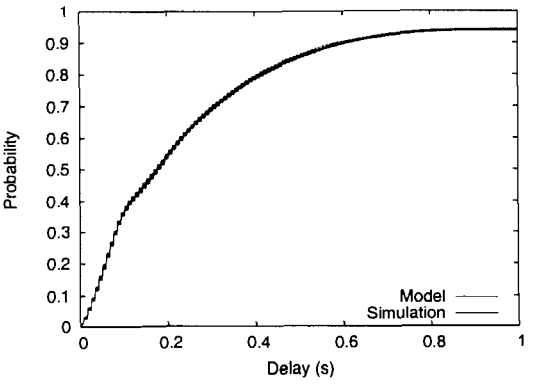


(a) Access Category 3 Delay Distribution

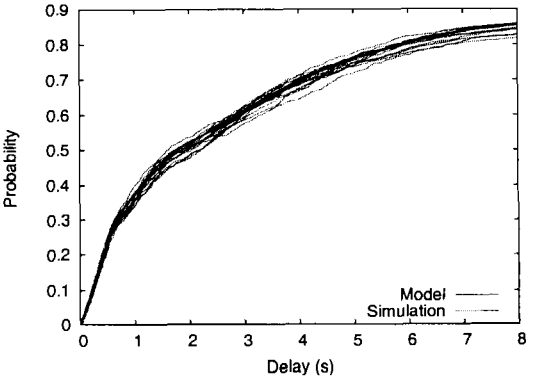


(b) Access Category 2 Delay Distribution

Figure 6.6: Simulation and Model Results for 10 stations using basic transmission mechanism from each of Access Categories 3 and 2

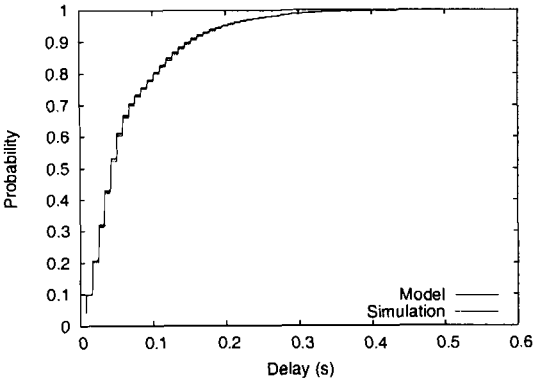


(a) Access Category 3 Delay Distribution

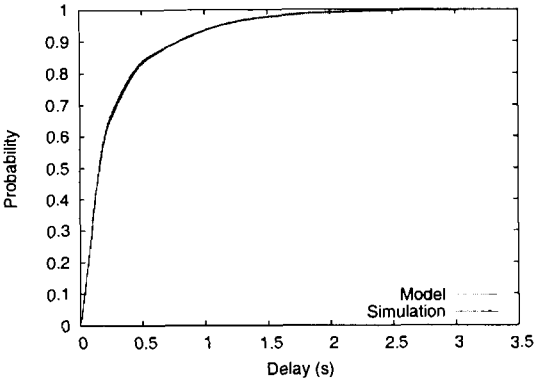


(b) Access Category 2 Delay Distribution

Figure 6.7: Simulation and Model Results for 15 stations using basic transmission mechanism from each of Access Categories 3 and 2

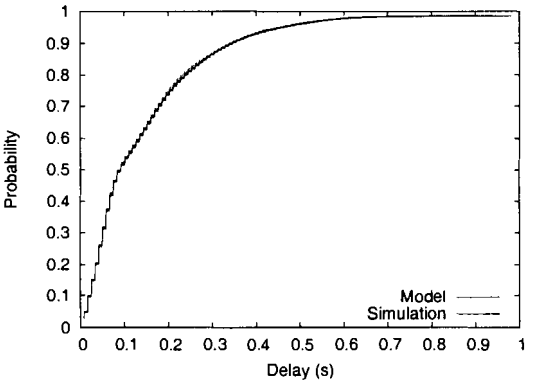


(a) Access Category 2 Delay Distribution

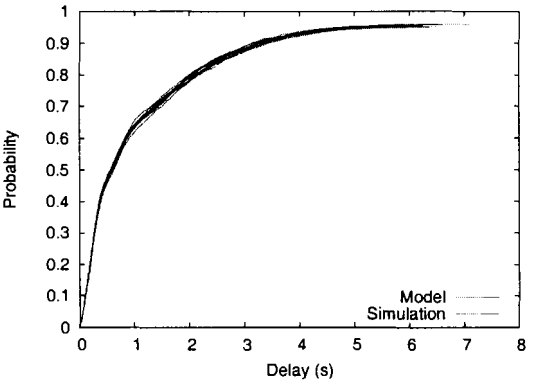


(b) Access Category 1 Delay Distribution

Figure 6.8: Simulation and Model Results for 5 stations using basic transmission mechanism from each of Access Categories 2 and 1

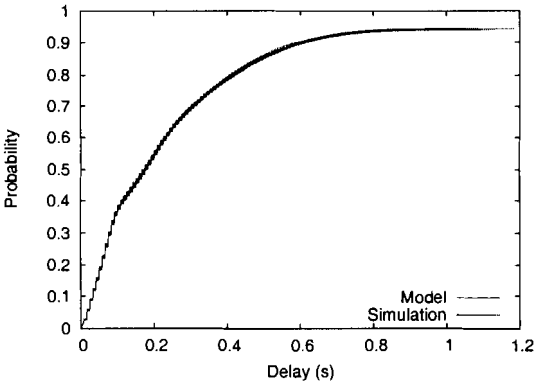


(a) Access Category 2 Delay Distribution

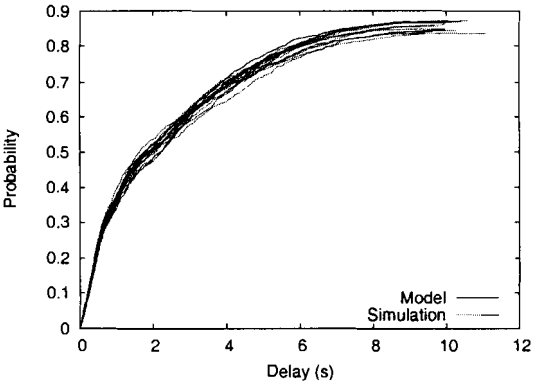


(b) Access Category 1 Delay Distribution

Figure 6.9: Simulation and Model Results for 10 stations using basic transmission mechanism from each of Access Categories 2 and 1



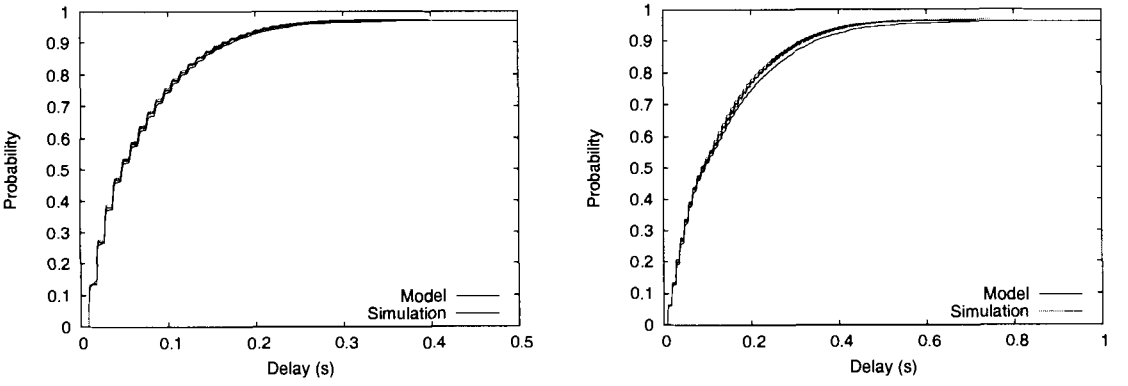
(a) Access Category 2 Delay Distribution



(b) Access Category 1 Delay Distribution

Figure 6.10: Simulation and Model Results for 15 stations using basic transmission mechanism from each of Access Categories 2 and 1

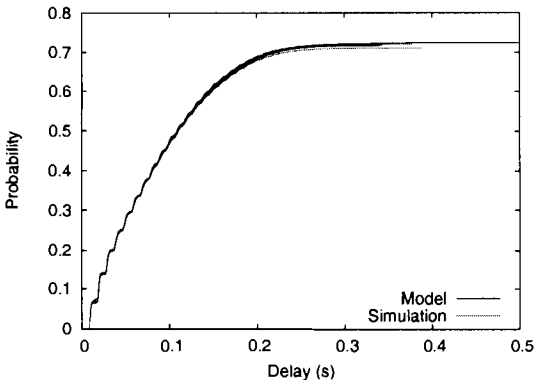
6.5.2 Delay Distribution for Results Using RTS/CTS Transmission Mechanism



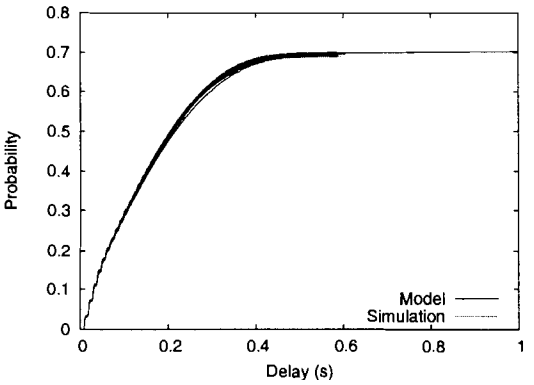
(a) Access Category 4 Delay Distribution

(b) Access Category 3 Delay Distribution

Figure 6.11: Simulation and Model Results for 5 stations using RTS/CTS transmission mechanism from each of Access Categories 4 and 3

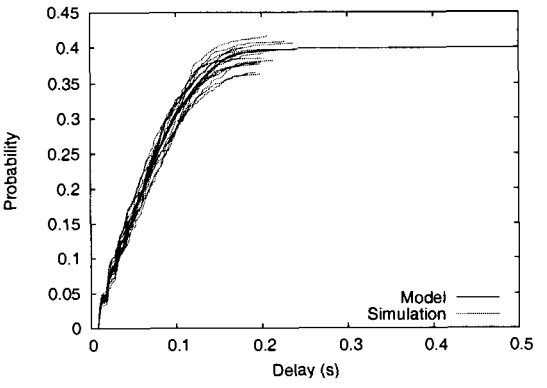


(a) Access Category 4 Delay Distribution

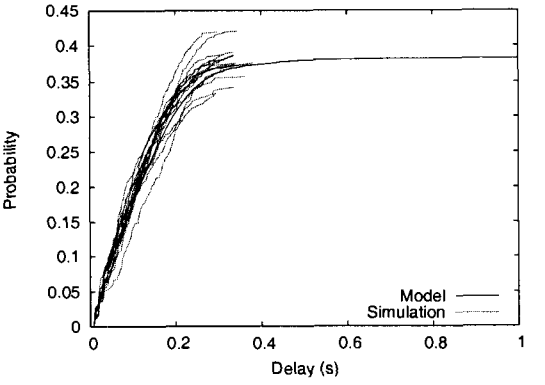


(b) Access Category 3 Delay Distribution

Figure 6.12: Simulation and Model Results for 10 stations using RTS/CTS transmission mechanism from each of Access Categories 4 and 3

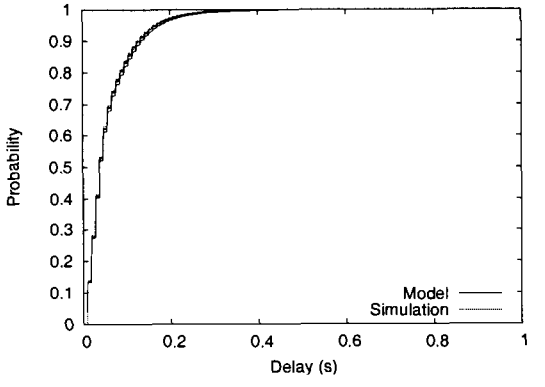


(a) Access Category 4 Delay Distribution

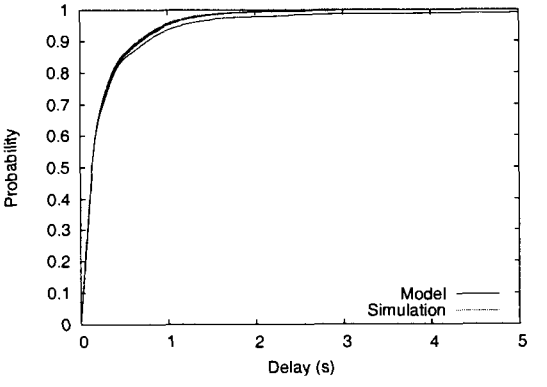


(b) Access Category 3 Delay Distribution

Figure 6.13: Simulation and Model Results for 15 stations using RTS/CTS transmission mechanism from each of Access Categories 4 and 3

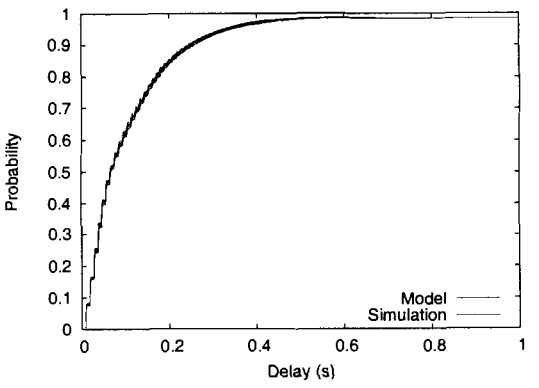


(a) Access Category 3 Delay Distribution

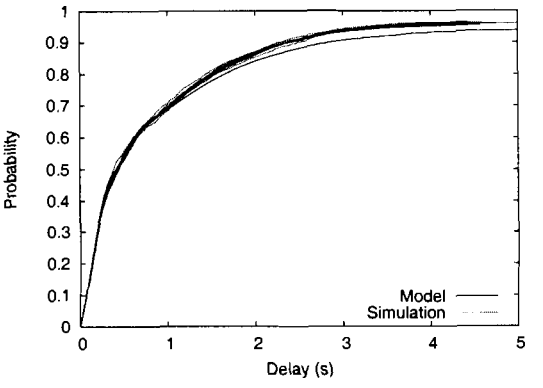


(b) Access Category 2 Delay Distribution

Figure 6.14: Simulation and Model Results for 5 stations using RTS/CTS transmission mechanism from each of Access Categories 3 and 2

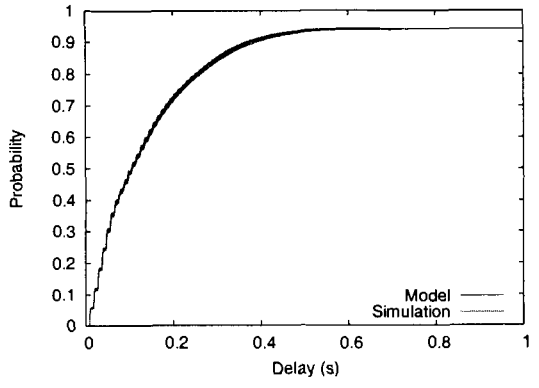


(a) Access Category 3 Delay Distribution

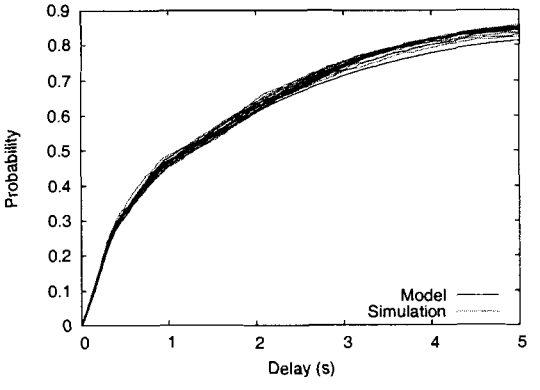


(b) Access Category 2 Delay Distribution

Figure 6.15: Simulation and Model Results for 10 stations using RTS/CTS transmission mechanism from each of Access Categories 3 and 2

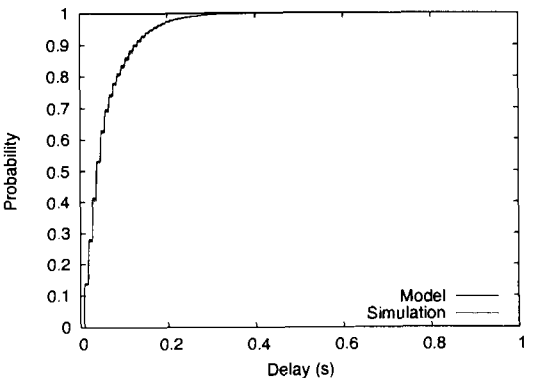


(a) Access Category 3 Delay Distribution

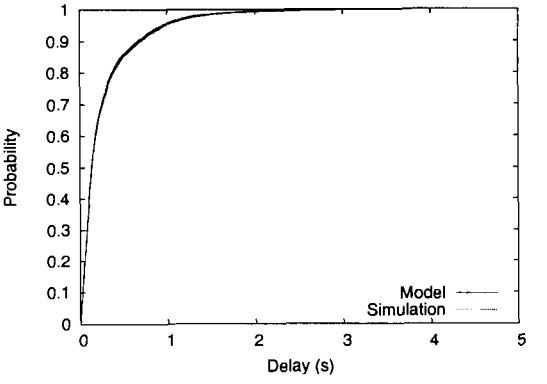


(b) Access Category 2 Delay Distribution

Figure 6.16: Simulation and Model Results for 15 stations using RTS/CTS transmission mechanism from each of Access Categories 3 and 2

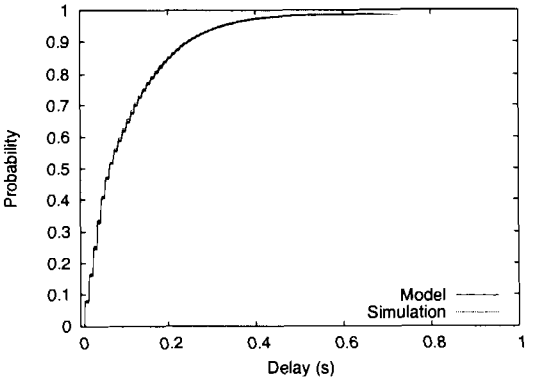


(a) Access Category 2 Delay Distribution

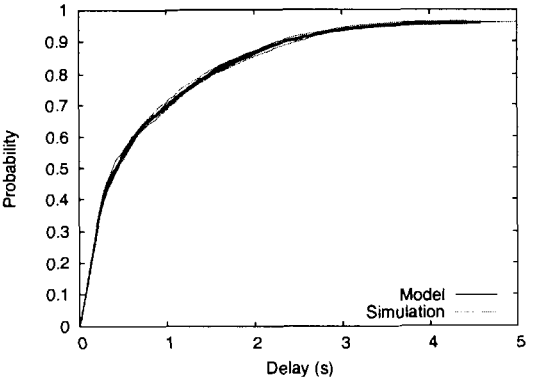


(b) Access Category 1 Delay Distribution

Figure 6.17: Simulation and Model Results for 5 stations using RTS/CTS transmission mechanism from each of Access Categories 2 and 1

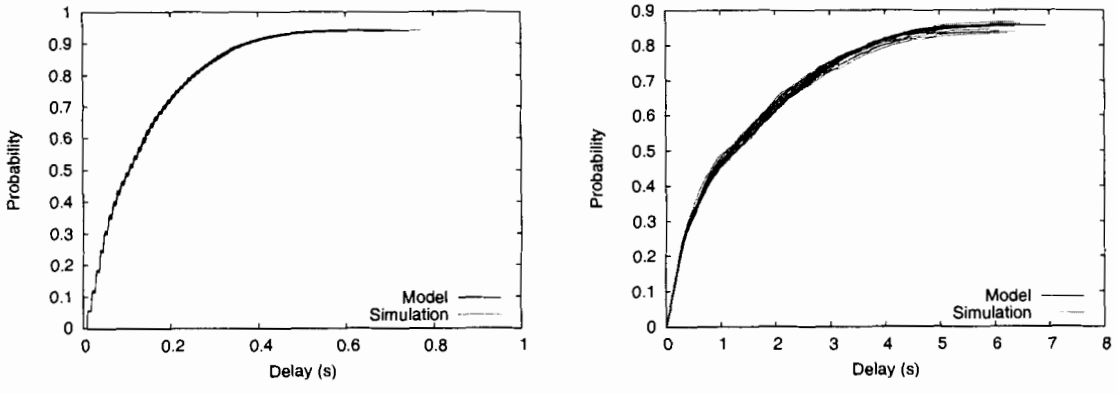


(a) Access Category 2 Delay Distribution



(b) Access Category 1 Delay Distribution

Figure 6.18: Simulation and Model Results for 10 stations using RTS/CTS transmission mechanism from each of Access Categories 2 and 1



(a) Access Category 2 Delay Distribution

(b) Access Category 1 Delay Distribution

Figure 6.19: Simulation and Model Results for 15 stations using RTS/CTS transmission mechanism from each of Access Categories 2 and 1

Chapter 7

Summary

7.1 Summary of Contributions

The research effort culminating in this thesis has made a number of important contributions. In the following subsections we enumerate these contributions, and briefly discuss their importance.

7.1.1 Correction of the ACKTimeout/CTSTimeout and EIFS errors

The modeling of IEEE 802.11 DCF and IEEE 802.11e EDCA/EDCF has been hindered by confusion over ACKTimeout and the use of EIFS. Until this work, no model had been developed that could accurately estimate the performance of a true IEEE 802.11e device. This work identifies the presence of these errors and develops an accurate model for a correct interpretation of the standard.

7.1.2 Development of the First Complete Two-Dimensional Discrete-Time Markov Chain Model for IEEE 802.11e EDCA Backoff

This work presents the first complete two-dimensional Markov process model for IEEE 802.11e EDCA backoff. In our opinion the two-dimensional model is superior to three-dimensional models. Because it tracks the backoff state and channel state separately it

is easier to understand. Because it treats collision probability as a single average value, its view of backoff is more intuitive. And because the Markov process for backoff has just two dimensions its stationary distribution has a closed form expression in terms of $\bar{p}$.

7.1.3 Development of a Scalable Model for Service Delay Distribution in Networks with Differentiated backoff Timescales

This work develops the first model to produce an exact estimate of the service delay distribution for IEEE 802.11e EDCA. The accuracy improvement relative to previous efforts based on averaging techniques is impressive. In contrast to the previous work based on signal flow graphs the model here can accommodate differentiated backoff timescales. Finally, the method shown here is also the first to demonstrate a technique for the reduction of the path polynomial. This enables the model to scale to typical IEEE 802.11 and IEEE 802.11e parameters, where other methods can not.

7.2 Future Research

There are a number of directions for future research with respect to the work presented in this thesis.

7.2.1 Improving the Model

The major performance bottleneck of this model is the calculation of the stationary distribution of contention vectors, π. The reason for this bottleneck is that the size of $\mathbb{P}$ grows according to the product of the number of stations from each AC in the network. Examination of π reveals that the occupancy of the more sparse contention vectors is very small. This corresponds to the intuition that collisions involving many stations have low probability. Future research could focus on eliminating these very rare contention states from $\mathbb{P}$.

Another area for improvement is the development of closed form expressions for

event probability where exhaustive summation is currently used. Aside from cleaning up the presentation of the model, these expressions would accelerate execution.

7.2.2 Extending the Model

Two extensions to this model stand out as worthwhile research objectives. The first is generalizing the signal flow graph technique to embrace the three dimensional models that have appeared in the literature. The other is extending the model to accommodate non-saturation traffic, perhaps by drawing on nascent techniques already appearing in the literature (Zhai and Fang 2003; Ozdemir and McDonald 2004).

7.2.3 Using of the Model to Study IEEE 802.11e Networks

Because of the model's impressive accuracy, it is an ideal platform for exploring the operation of IEEE 802.11e networks. In contrast to simulation, this model can produce accurate results rapidly, and can extract different performance metrics more easily. The model could be used to tune EDCA parameters for particular network applications, develop intelligent admission control frameworks and be incorporated in real time network control applications.

References

Alizadeh-Shabdiz, F. and S. Subramaniam (2003, March). A finite load analaytical model for the IEEE 802.11 distributed coordinated function MAC. In *Proc. WiOpt*, pp. 321–22.

Banchs, A. and L. Vollero (2005, June). A delay model for IEEE 802.11e EDCA. *IEEE Commun. Lett. 9*, 508–510.

Bianchi, G. (2000, March). Performance analysis of the IEEE 802.11 DCF. *IEEE J. Select. Areas Commun. 18*(3), 535–547.

Cali, F., M. Conti, and E. Gregori (1998, Mar). IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement. In *INFOCOM*, Volume 1, pp. 142–149.

Carvalho, M. and J. Garcia-Luna-Aceves (2003, November). Delay analysis of IEEE 802.11 in single-hop networks. In *Proc. IEEE International Conference on Network Protocols*, Atlanta, GA, pp. 146–155.

Chatzimisios, P., A. Boucouvalas, and V. Vitsas (2003a, Dec). IEEE 802.11 packet delay – a finite retry limit analysis. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Volume 2, San Francisco, USA, pp. 950–954.

Chatzimisios, P., A. Boucouvalas, and V. Vitsas (2003b, Sept.). Packet delay analysis of IEEE 802.11 MAC protocol. *Electronics Letters 39*(18), 1358–1359.

Chatzimisios, P., V. Vitsas, and A. Boucouvalas (2002, October). Throughput and delay analysis of IEEE 802.11 protocol. In *IEEE Networked Appliances*, pp. 168–174. International Workshop on.

Chhaya, H. S. and S. Gupta (1997). Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol. *Wirel. Netw. 3*(3), 217–234.

Fateman, R. (2003, March). Comparing the speed of programs for sparse polynomial multiplication. *ACM SIGSAM Bulletin 37*, 4–15.

Ge, Y. and J. Hou (2003, May). An analytical model for service differentiation in IEEE 802.11. In *Proc. IEEE International Conference on Computer Communication (INFOCOM)*, Volume 2, San Francisco, CA, pp. 1157–1162.

Hadzi-Velkov, Z. and B. Spasenovski (2003). Saturation throughput -delay analysis of IEEE 802.11 DCF in fading channel. In *Proc. IEEE International Communications Conference (ICC)*, Volume 1, pp. 121–126.

He, J., L. Zheng, Z. Yang, and C. T. Chou (2003, October). Performance analysis and service differentiation in IEEE 802.11 WLAN. In *Proc. IEEE Conference on Local Computer Networks*, pp. 691–697.

Hui, J. and M. Devetsikiotis (2003, December). Designing improved MAC packet schedulers for 802.11e wlan. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Volume 1, San Francisco, CA, pp. 184–189.

Hui, J. and M. Devetsikiotis (2004, December). Performance analysis of IEEE 802.111e EDCA by a unified model. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Volume 1, Dallas, TX, pp. 754–759.

IEEE 802.11 Task Group E (2005, January). Draft ammendment to standard for information technology – telecommunications and information exchange between systems – LAN/MAN specific requirements – part 11: Wireless medium access control (MAC) and physical layer (PHY) specifications: Medium access control (MAC)) quality of service (QoS) enhancements.

IEEE 802.11 Working Group (1999). IEEE 802.11, 1999 edition (ISO/IEC 8802-11: 1999) IEEE standards for information technology – telecommunications and information exchange between systems – local and metropolitan area network – specific requirements – part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications.

Jain, R. (1991). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation and modeling.* John Wiley & Sons, Inc.

Kim, H. and J. Hou (2004, June). A fast simulation framework for IEEE 802.11-operated wireless LANs. In *ACM SIGMETRICS 2004/PERFORMANCE 2004: Proc. joint international conference on measurement and modeling of computer systems*, Volume 1, New York, NY, pp. 143–154.

Kim, H. and J. C. Hou (2003). Improving protocol capacity with model-based frame scheduling in IEEE 802.11-operated wlans. In *Proc. international conference on Mobile computing and networking*, pp. 190–204. ACM Press.

Kleinrock, L. and F. Tobagi (1975, December). Packet switching in radio channels, Part II -The hidden terminal problem in carrier sense multiple access and the busy tone solution. *IEEE Trans. Commun. COM-23*(12), 1417–1433.

Little, J. (1961). A proof of the queuing formula it L = hW. *Oper. Res. J. 18*, 172–174.

McCanne, S. and S. Floyd (2005). The network simulator - ns-2. `http://www.isi.edu/nsnam/ns/`.

Moreton, M. (2005). Synad EDCA model. `http://sourceforge.net/projects/ns2-wlan-patch/`.

Opnet Technologies Inc. (2005). Opnet modeler. `http://www.opnet.com/products/modeler/home.html`.

Ozdemir, M. and A. B. McDonald (2004, October). An M/MMGI/1/K queuing model for IEEE 802.11 ad hoc networks. In *PE-WASUN '04: Proc. ACM international workshop on Performance evaluation of wireless ad hoc, sensor and ubiquitous networks*, Venezia, Italy, pp. 107–111.

Qiao, D. and K. G. Shin (2002, May). Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 wlan under the dcf. In *Proc. IWQoS'2000.*

Ramaiyan, V., A. Kumar, and E. Altman (2005, June). Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness, multistability and throughput differentiation.

In *ACM SIGMETRICS 2005: Proc. joint international conference on measurement and modeling of computer systems*, Banff, AB, pp. 109–120.

Raptis, P., V. Vistas, K. Paprrizos, P. Chatzimisios, and A. Boucouvalas (2004, June). Packet delay distribution of IEEE 802.11 dcf in presence of transmission errors. In *Proc. IEEE International Communications Conference (ICC)*, New Orleans, LA, pp. 3854–3858.

Raptis, P., V. Vistas, K. Paprrizos, P. Chatzimisios, and A. Boucouvalas (2005, June). Packet delay distribution of the IEEE 802.11 distributed coordination function. In *Proc. IEEE Symposium on a world of wireless mobile and multimedia networks*, pp. 299–304.

Robinson, J. and T. Randhawa (2004a, September). A practical model for transmission delay of IEEE 802.11e enhanced distributed channel access. In *Proc. IEEE International Symposium of Personal and Indoor Mobile Radio Communications (PIMRC)*, Volume 1, Barcelona, Spain, pp. 323–328.

Robinson, J. and T. Randhawa (2004b, June). Throughput analysis of IEEE 802.11e enhanced distributed coordination function. *IEEE J. Select. Areas Commun. 22*(5), 917–928.

Singla, A. and G. Chesson (2005). Atheros EDCF model. `ftp://ftp-sop.inria.fr/rodeo/qni/ns-edcf.tar.gz`.

Tao, Z. and S. Panwar (2004a, April). An analytical model for the IEEE 802.11e edcf. In *Proc. IEEE Workshop on local and metropolitan area networks (LANMAN)*, San Francisco, CA, pp. 39–44.

Tao, Z. and S. Panwar (2004b, June). An analytical model for the IEEE 802.11e enhanced distributed coordination function. In *Proc. IEEE International Communications Conference (ICC)*, Volume 7, Paris, France, pp. 4111–4117.

TKN TU Berlin (2005). TKN TU Berlin EDCA model. `www://tkn.tu-berlin.de/research/802.11e_ns2`.

Wang, G., Y. Shu, L. Zhang, and O. Yang (2003, Sept.). Delay analysis of the IEEE 802.11 DCF. In *Proc. IEEE International Symposium of Personal and Indoor Mobile Radio Communications (PIMRC)*, Volume 2, pp. 1737–1741.

Wu, H., S. Cheng, Y. Peng, K. Long, and J. Ma (2002, April). IEEE 802.11 distributed coordination function (dcf): Analysis and enhancement. In *Proc. IEEE International Communications Conference (ICC)*, New York, NY.

Xu, K., Q. Wang, and H. Hassanein (2003, December). Performance analysis of differentiated QoS supported by IEEE 802.11e enhanced distributed coordination function (EDCF) in WLAN. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Volume 2, San Francisco, USA, pp. 1048–1053. Global Telecommunications Conference.

Yang, X. (2003, March). IEEE 802.11e wireless LAN for quality of service. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Volume 2, New Orleans, LA, pp. 1291–1296.

Yang, X. (2004a, March). An analysis for differentiated services in IEEE 802.11 and IEEE 802.11e wireless LANs. In *Proc. IEEE International Conference on Distributed Computing Systems*, Tokyo, Japan, pp. 32–39.

Yang, X. (2004b, June). Performance analysis of IEEE 802.11e edcf under saturation condition. In *Proc. IEEE International Communications Conference (ICC)*, Volume 1, New Orleans, LA, pp. 170–174.

Zhai, H. and Y. Fang (2003, Sept.). Performance of wireless LANs based on IEEE 802.11 MAC protocols. In *Proc. IEEE International Symposium of Personal and Indoor Mobile Radio Communications (PIMRC)*, Beijing, China.

Zhao, J., Z. Guo, and W. Zhu (2002, November). Performance study of MAC for service differentiation in IEEE 802.11. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Volume 1, Taipei, Taiwan, pp. 778–782.

Zhen-ning, K., D. Tsang, and B. Bensauo (2004, December). Performance analysis of IEEE 802.11e contention-based channel access (EDCF). *IEEE J. Select. Areas Commun. 22*, 2095–2106.

Zhu, H. and I. Chlamtac (2003, October). An analytical model for IEEE 802.11e EDCF differential services. In *Proc. IEEE International Conference on Computer Communications and Networks (ICCCN)*, Volume 1, Dallas, TX, pp. 163–168.

Zimmerman, H. (1980, April). OSI reference model – the ISO model of architecture for open systems intercommunications. *IEEE Trans. Commun. COM-28*.