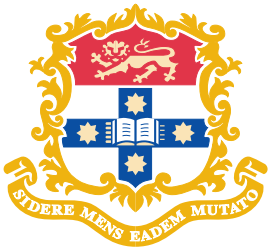
A Markov Chain Approach to IEEE 802.11 WLAN Performance Analysis

Lixiang Xiong

A thesis submitted in fulfilment of
requirements for the degree of

Doctor of Philosophy



School of Electrical & Information Engineering

The University of Sydney

October 2008

Abstract

Wireless communication always attracts extensive research interest, as it is a core part of modern communication technology. During my PhD study, I have focused on two research areas of wireless communication: IEEE 802.11 network performance analysis, and wireless cooperative retransmission.

The first part of this thesis focuses on IEEE 802.11 network performance analysis. Since IEEE 802.11 technology is the most popular wireless access technology, IEEE 802.11 network performance analysis is always an important research area. In this area, my work includes the development of three analytical models for various aspects of IEEE 802.11 network performance analysis.

First, a two-dimensional Markov chain model is proposed for analysing the performance of IEEE 802.11e EDCA (Enhanced Distributed Channel Access). With this analytical model, the saturated throughput is obtained. Compared with the existing analytical models of EDCA, the proposed model includes more correct details of EDCA, and accordingly its results are more accurate. This better accuracy is also proved by the simulation study.

Second, another two-dimensional Markov chain model is proposed for analysing the coexistence performance of IEEE 802.11 DCF (Distributed Coordination Function) and IEEE 802.11e EDCA wireless devices. The saturated throughput is obtained with the proposed analytical model. The simulation study verifies the proposed analytical model, and it shows that the channel access priority of DCF is similar to that of the best effort access category in EDCA in the coexistence environment.

The final work in this area is a hierarchical Markov chain model for investigating the impact of data-rate switching on the performance of IEEE 802.11 DCF. With this analytical model,

the saturated throughput can be obtained. The simulation study verifies the accuracy of the model and shows the impact of the data-rate switching under different network conditions. A series of threshold values for the channel condition as well as the number of stations are obtained to decide whether the data-rate switching should be active or not.

The second part of this thesis focuses on wireless cooperative retransmission. In this thesis, two uncoordinated distributed wireless cooperative retransmission strategies for single-hop connection are presented. In the proposed strategies, each uncoordinated cooperative neighbour randomly decide whether it should transmit to help the frame delivery depending on some pre-calculated optimal transmission probabilities. In Strategy 1, the source only transmits once in the first slot, and only the neighbours are involved in the retransmission attempts in the subsequent slots. In Strategy 2, both the source and the neighbours participate in the retransmission attempts. Both strategies are first analysed with a simple memoryless channel model, and the results show the superior performance of Strategy 2. With the elementary results for the memoryless channel model, a more realistic two-state Markov fading channel model is used to investigate the performance of Strategy 2. The simulation study verifies the accuracy of our analysis and indicates the superior performance of Strategy 2 compared with the simple retransmission strategy and the traditional two-hop strategy.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my principle supervisor Dr. Guoqiang Mao for his kind encouragement, advice, and support during the whole procedure of my PhD study.

Also, I would like to express my sincerest gratitude to Dr. Lavy Libman from NICTA, for his kind encouragement, advice, and support during the final year of my PhD study in the area of wireless cooperative communication.

I am graceful to my parents Weibang Xiong and Kunjun Zhu, and my younger brother Likun Xiong for their love and encouragement.

This work would not have been possible without the financial support of the following organisations:

- The University of Sydney, through the Australian Postgraduate Awards (APA) scholarship.
- The school of Electrical and Information Engineering at the University of Sydney, through the Norman I Price scholarship.
- NICTA, through the NICTA top-up scholarship.
- Australian Research Council, through the Australian Research Council discovery grant (grant number:DP0559248).

Finally, I would like to acknowledge that this thesis has been proof-read by Pat Skinner for English writing before submission.

Statement of Originality

I hereby declare that this thesis, submitted in fulfillment of the requirement for the award of Doctor of Philosophy, in the school of Electrical and Information Engineering, University of Sydney, is my own work unless otherwise referenced or acknowledged. The document has not been accepted for the award of any other qualification at any educational institution.

Signed

Lixiang Xiong

Date:

Related Publications

- Journal publications:

  1. Lixiang Xiong and Guoqiang Mao, “Performance Analysis of IEEE 802.11 DCF with Data Rate Switching”, IEEE Communication Letters, Vol 11, Issue 9, pp 759-761, September 2007.
  2. Lixiang Xiong, Guoqiang Mao, “Saturation Throughput Analysis for IEEE 802.11e EDCA”, Computer Networks, Vol. 51, Issue 11, pp. 3047-3068, 8 August 2007.

- Conference publications:

  1. Lixiang Xiong, Lavy Libman, and Guoqiang Mao, “Optimal Strategies for Cooperative MAC-Layer Retransmission in Wireless Networks”, accepted by IEEE Wireless Communications and Networking Conference (WCNC), Las Vegas, USA, March 2008.
  2. Lixiang Xiong and Guoqiang Mao, “An Analysis of the Coexistence of 802.11 DCF and 802.11e EDCA”, IEEE Wireless Communications and Networking Conference (WCNC), Hong Kong, March 2007.
  3. Lixiang Xiong and Guoqiang Mao, “Saturated throughput analysis of IEEE 802.11e using two-dimensional Markov chain model”, The Third International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine), Waterloo, Ontario, Canada, August 2006.
  4. Lixiang Xiong, Guoqiang Mao, Yanqiang Luan. “Enhancement to IEEE 802.11e EDCA with QoS Differentiation Based on Station Priority”, Australian Telecommunication Networks and Applications Conference (ATNAC), Sydney, Australia, December 2004.

Contents

List of Figures

List of Tables

List of Acronyms

AC	access category
ACK	acknowledgement
AIFS	arbitration IFS
AIFSN	AIFS number
AWGN	additive white Gaussian noise
ARQ	automatic repeat-request
CDMA	code-division multiple access
CNR	carrier to noise ratio
CSMA/CA	carrier sense multiple access with collision avoidance
CTS	clear-to-send
CW	contention window
DCF	distribution coordination function
DIFS	DCF IFS
DQPSK	differential quadrature phase shift keying
DRSA	data-rate switching active
DRSI	data-rate switching inactive
DSSS	direct sequence speared spectrum
EDCA	enhanced distributed channel access
EIFS	extended IFS
FER	frame error rate
IFS	inter frame space
MAC	medium access control
MIMO	multiple-input and multiple-output
NCSW	node-cooperative stop-and-wait
non-QSTA	non QoS station, or station using DCF
QoS	quality of service

QSTA	QoS station, or station using EDCA
RTS	request-to-send
SIFS	the shortest IFS
SNR	signal to noise ratio
TXOP	transmission opportunity
WLAN	wireless local area network

List of Notations/Symbols

The notations/symbols used in IEEE 802.11 network performance analysis

$AC\ A$	access category A
$AC\ B$	access category B
ACK	the time to transmit an ACK frame
$AIFS[A]$	$AIFS$ value for AC A
$AIFS[B]$	$AIFS$ value for AC B
$AIFS_{min}$	the minimum $AIFS$ used in a system
$AIFSN$	AIFS number
$b(i)$	the state probability for the state (i) in the one-dimensional Markov chain model in Fig 5.1(b)
$b_{A(r,k)}$	the state probability for the state (r, k) in the two-dimensional Markov chain model in Fig 3.2 for AC A
$b_{B(r,k)}$	the state probability for the state (r, k) in the two-dimensional Markov chain model in Fig 3.3 for AC B
$c(k)$	a function to indicate whether a value k is within a given value range
CW	contention window size
$CW(i)$	the CW size of the i^{th} backoff stage
CW_{max}	the maximum CW size
CW_{maxA}	the maximum CW size for $AC\ A$
CW_{maxB}	the maximum CW size for $AC\ B$
CW_{maxD}	the maximum CW size in DCF
CW_{maxE}	the maximum CW size in EDCA
CW_{min}	minimum CW size

$d(k)$	the state probability for the state (k) in the one-dimensional Markov chain model in Fig 4.5
$e(t)$	the stochastic process modeling the number of consecutive successful and unsuccessful transmissions experienced by a given station
$E[A]$	the average effective payload for $AC\ A$
$E[B]$	the average effective payload for $AC\ B$
$E[DCF]$	the average effective payload for non-QSTAs
$E[EDCA]$	the average effective payload for QSTAs
$E[P]$	the effective payload in a data frame
$EIFS$	the duration of EIFS
$EIFS[A]$	$EIFS$ for $AC\ A$
$EIFS_{min}$	the value of $SIFS + ACK + AIFS_{min}$
EL	the average time duration between two successive transmissions
FER_{11}	the frame error rate at data rate 11 $Mbps$
$FER_{5.5}$	the frame error rate at data rate 5.5 $Mbps$
FER_1	the frame error rate at data rate R_1
FER_2	the frame error rate at data rate R_2
H	the time required for transmitting the physical layer header and the MAC layer header of a frame"
IFS_D	IFS in DCF
IFS_E	IFS in EDCA
n	the total number of stations in a system
n_A	the number of $AC\ A$ stations
n_B	the number of $AC\ B$ stations
N_D	the number of non-QSTAs in the system
N_E	the number of QSTAs in the system
$o(i)$	the state probability for the state (i) in the Markov chain model in Fig 4.5
P	the time required for transmitting the data payload of a frame
P_{col_1}	the probability that an unsuccessful transmission occurs at data rate $R1$ due to a collision

P_{col_2}	the probability that an unsuccessful transmission occurs at data rate $R2$ due to a collision
P_{event}	the probability that a channel event occurs
P_{FER_1}	the probability that an unsuccessful transmission occurs at data rate $R1$ due to frame error
P_{suc_1}	the probability that a successful transmission occurs at data rate $R1$
P_{suc_2}	the probability that a successful transmission occurs at data rate $R2$
p_1	the unsuccessful probability for a station at data rate $R1$
p_2	the unsuccessful probability for a station at data rate $R2$
P_{bA}	the probability that an $AC\ A$ station suffers a busy channel immediately after it finishes an $IFS[A]$ since the end of the last busy channel
P_{bB}	the probability that an $AC\ B$ station suffers a busy channel immediately after it finishes an $IFS[A]$ since the end of the last busy channel
p_c	the collision probability for a given station
$P(i)\|(j)$	the transition probability from state (j) to state (i) in a one-dimensional Markov chain model
$P(i,j)\|(r,k)$	the transition probability from state (r,k) to state (i,j) in a two-dimensional Markov chain model
p_A	the average collision probability for an $AC\ A$ station
$p_{A:zone(i)}$	the collision probability for an $AC\ A$ station in a time slot in zone i
p_B	the average collision probability for an $AC\ B$ station
$p_{B(i)}$	the probability that an AC B station suffers a busy channel in the time slot $AIFS[A]+i$ since the end of the last busy channel
$P_{col:zone(i)}$	the probability that a collision occurs in a time slot in zone i
P_{idleA}	the average probability that an $AC\ A$ station experiences an idle time slot
$P_{idleA:zone(i)}$	the probability that an $AC\ A$ station experiences an idle time slot in zone i

P_{idleB}	the average probability that an $AC\ B$ station experiences an idle time slot
$Pr(r)$	the probability that a station starts a new backoff procedure with an initial counter value r
$Pr\_A(r)$	the probability that an $AC\ A$ station starts a backoff procedure with an initial counter value r
$Pr\_B(r)$	the probability that an $AC\ B$ station starts a backoff procedure with an initial counter value r
P_{sB}	the probability that an $AC\ B$ station suffers a busy channel in a time slot within the $IFS[B]$ period since the end of the last busy channel
$P_{sucA:zone(i)}$	the probability that a successful transmission from an AC A station occurs in a time slot in zone i
$P_{sucB:zone(i)}$	the probability that a successful transmission from an AC B station occurs in a time slot in zone i
$P_{tr:zone(i)}$	the probability that a station transmits in a time slot in zone i
$q(t)$	the stochastic process modeling the number of the transmission attempts involved in sending a single data frame"
$R1$	data rate $R1$
$R2$	data rate $R2$
$s(i)$	the state probability for the state (i) in the one-dimensional Markov chain models in Fig 3.5, Fig 4.3, and Fig 4.7
$s(i,j,k)$	the state probability for the state (i,j,k) in the three-dimentional Markov chain model in Fig 5.1(a)
$SIFS$	the duration of $SIFS$
T_{event}	the time for a channel event
Tc	the time required for a collision
$Throughput$	the overall throughput
$Throughput_{DCF}$	the throughput for each non-QSTA
	the throughput for each QSTA
$Throughput_{EDCA}$	
$Throughput_A$	the throughput for each station of $AC\ A$
$Throughput_B$	the throughput for each station of $AC\ B$

$timeslot$	the length of one IEEE 802.11 time slot
Ts	the time required for a successful transmission
$u(t)$	the stochastic process modeling the data rate switching of a given station
$v(t)$	the stochastic process modeling the backoff stage of a given station
$w(t)$	the stochastic process modeling the backoff counter decrement of a given station
$Z(i)$	the probability that a system reaches zone i, used in the one-dimensional Markov chain models in Fig 4.3 and Fig 4.7
$\rho_E(i)$	the collision probability for a QSTA in zone i
$\overline{\rho_E}$	the average collision probability for a QSTA
$\varrho(k)$	the probability of no transmission in a time slot in zone k
τ	the transmission probability of a given station
τ_1	the transmission probability of a given station at data rate R1
τ_2	the transmission probability of a given station at data rate R2
τ_A	the transmission probability of an AC A station
τ_B	the transmission probability of an AC B station
τ_D	the transmission probability of a non-QSTA
τ_E	the transmission probability of a QSTA
$\phi(i)$	the probability that i out of the N_D non-QSTAs become involved in the previous transmission
$\omega_D(i)$	the probability that a non-QSTA detects an idle time slot in zone i
$\overline{\omega_D}$	the average probability that a non-QSTA detects an idle time slot
$\omega_E(i)$	the probability that a QSTA detects an idle time slot in zone i
$\overline{\omega_E}$	the average probability that a QSTA detects an idle time slot

$\psi_D(i)$	the probability that a successful transmission from a non-QSTA occurs in a time slot in zone i
$\psi_E(i)$	the probability that a successful transmission from a QSTA occurs in a time slot in zone i

The notations/symbols used in wireless cooperative retransmission

k	the number of neighbours that successfully receive a frame from the source
K	the total number of neighbours
E	the expected number of slots until success
L_{nd}	the distance between the neighbours and the destinations
L_{sd}	the distance between the source and the destinations
L_{sn}	the distance between the source and the neighbours
m	the number of iterations for Strategy 1
$\underset{int[i]}{P}\{c\|k\}$	the conditional probability, given that k neighbours have the frame, that c out of the remaining $K-k$ neighbours have an interim channel in the "on" state in slot i
$\underset{rel[i]}{P}\{r,d\|k\}$	the conditional probability, given that k neighbours have the frame, that exactly r out of them have their relay channel in the "on" state and the direct channel is "off" (d=0) or "on" (d=1), respectively, in slot i
$\underset{rel[i]}{P}^{ap}\{r,d\|k\}$	revised $\underset{rel[i]}{P}\{r,d\|k\}$
$P^{ap}_{[i]}\{k\}$	revised $P_{[i]}\{k\}$
P^{suc*}	the successful retransmission probability after using the optimal transmission probability τ^*
P^{suc}	the successful retransmission probability
$P^{suc}(\tau_s,\tau_n\|k)$	the success probability, provided with that k neighbours have the frame, and τ_s and τ_n are used
$P^{suc}(\tau_s,\tau_n\|k,r,d)$	the conditional success probability in slot i, provided that the system is in a particular state (k,r,c,d), and τ_s, τ_n in that slot

$P^{suc}(k)$	the probability that k neighbours can receives a frame from the source and retransmits the frame to the destination successfully
$P^{suc}_{[i]}$	the total probability of success in slot i
$P_{[i]}\{k\}$	the probability that k neighbours have received a frame from the source before slot i
$P_i(k)$	the probability that k neighbours have received a frame from the source in slot i
$P^{rev}_i(k)$	revised $P_i(k)$in slot i
P^{suc}_{k,τ^*}	the retransmission success probability when k neighbours have received a frame from the source and they use the optimal transmission probability τ^*
P_{bg_nd}	the transit probability from "bad" status to "good" status for a radio channel between a neighbour and the destination
P_{bg_sd}	the transit probability from "bad" status to "good" status for a radio channel between the source and the destination
P_{bg_sn}	the transit probability from "bad" status to "good" status for a radio channel between the source and a neighbour
P_{gb_nd}	the transit probability from " good" status to "bad" status for a radio channel between a neighbour and the destination
P_{gb_sd}	the transit probability from " good" status to "bad" status for a radio channel between the source and the destination
P_{gb_sn}	the transit probability from " good" status to "bad" status for a radio channel between the source and a neighbour
SNR	signal to noise ratio in dB
τ	the transmission probability of a neighbour in Strategy 1

τ^*	the optimal transmission probability of a neighbour in Strategy 1
τ_s	the transmission probability of the source in Strategy 2
τ_n	the transmission probability of a neighbour in Strategy 2
τ_s^*	the optimal transmission probability of the source in Strategy 2
τ_n^*	the optimal transmission probability of a neighbour in Strategy 2
$\tau_{s[i]}^*$	the optimal transmission probability of the source in Strategy 2 in slot i
$\tau_{n[i]}^*$	the optimal transmission probability of a neighbour in Strategy 2 in slot i
$\Pi_{[i]}\{k, \hat{k}\}$	the probability of the system to have $\hat{k}$ frame copies in slot $i+1$ if it had k of them in slot i
$\prod_{old}\{r, r'\|k\}$	the probability to have r' relay channels (out of the "old" k) in the "on" state after that number was r in the previous slot
$\prod_{rel}\{r, \hat{r}\|k, \hat{k}\}$	the probability to have $\hat{r}$ relay channels (out of $\hat{k}$) in the "on" state after that number was r out of k in the previous slot
$\prod_{int}\{c, \hat{c}\|\hat{k}\}$	the probability of $\hat{c}$ interim channels (out of $K-\hat{k}$) to be "on" after that number was c in the previous slot

Chapter 1

Introduction

In the last decade, the telecommunication industry has been experiencing a revolution sparked by modern wireless communication technologies. More and more users are enjoying the convenience of today's wireless communication. In 2002, a milestone event occurred when the number of mobile telephone users in the world first exceeded that of traditional wired telephone users [1]. The great success of modern wireless communication also motivates researchers to develop better wireless communication systems. During my PhD study, I have been working on two research areas of wireless communication: IEEE 802.11 network performance analysis, and wireless cooperative retransmission.

1.1 Background: IEEE 802.11 Network Performance Analysis

In recent years, a widespread deployment of wireless hot spots (Wi-Fi spots) has been witnessed, and it is predicted that this trend will continue in the near future. In [2], the authors state that "Market estimates indicate that approximately 4.5 million Wi-Fi APs were sold during the 3^{rd} quarter of 2004 alone and that the sales of Wi-Fi equipment will triple by 2009" based on two online market reports [3, 4].

Among various wireless access technologies that have been used for the Wi-Fi networks, the

IEEE 802.11 access technology [5, 6] is widely considered as the most popular Wi-Fi access technology. In fact, the terms “802.11” and “Wi-Fi” are often used interchangeably, due to the dominant position of the IEEE 802.11 technology in the Wi-Fi equipment market.

In addition to its extensive application in the Wi-Fi networks for offering wireless Internet service, the IEEE 802.11 technology has been widely used in other areas, such as wireless sensor networks and wireless mesh networks. Also, the IEEE 802.11 network has been considered as an important part in the future 4G telecommunication network, where customers may use voice or even video communication over the IEEE 802.11 network [7, 8]. The great success of the IEEE 802.11 technology also motivates researchers to devote themselves to improving the performance of the existing IEEE 802.11 technology [9–12].

Technically, the term “IEEE 802.11” should be referred to as a set of IEEE standards for wireless local area network (WLAN), such as IEEE 802.11a [13], IEEE 802.11b [14], IEEE 802.11g [15], and IEEE 802.11e [16]. A core part of these IEEE 802.11 standards is the technical definition of its MAC (medium access control) layer access function. The fundamental MAC layer access mechanism of the IEEE 802.11 technology is DCF (Distributed Coordination Function). DCF is based on CSMA/CA (carrier sense multiple access with collision avoidance) backoff mechanism for channel access control, and each station implements its own backoff procedure for channel access in DCF. With DCF, the IEEE 802.11 technology offers a simple distributed approach for sharing radio channel.

However, DCF can only offer a best-effort channel access service, where all stations statistically share the channel fairly. It cannot support QoS (quality of service) differentiation. This shortcoming has attracted considerable research attention, and it finally resulted in the publication of QoS supported IEEE 802.11e standard [16] in 2005. In IEEE 802.11e, EDCA (Enhanced Distributed Channel Access) replaces DCF as the fundamental MAC layer access function in the IEEE 802.11 technology. In EDCA, stations perform backoff mechanism with different parameters based on the category of traffic they carry, and stations carrying higher priority traffic can have a better chance for channel access.

The technical details of DCF and EDCA will be introduced as follows. First, the details of CSMA/CA is introduced, as it is a fundamental access mechanism used in both DCF and EDCA; Second, DCF and EDCA are briefly introduced; Third, some subtle differences between DCF and EDCA are introduced, because they will be analysed in this thesis; Finally,

the data-rate switching in IEEE 802.11 is introduced, because later it will be analysed in this thesis.

1.1.1 An Introduction to CSMA/CA

The CSMA/CA mechanism is the fundamental access mechanism in the IEEE 802.11 technology. The detailed procedure of CSMA/CA can be described as follows:

When a station is ready for a transmission, it must sense the channel as being idle for a complete IFS (inter frame space) before it can start the next step. In the case that the channel becomes busy before the completion of an idle IFS, the station must wait through another complete IFS after the channel returns to the idle state. As long as the station cannot detect the channel as idle for a complete IFS, it cannot start the next step. Once the station finishes a complete idle IFS interval, it needs to complete a backoff procedure before it can start a transmission. During the backoff procedure, an initial backoff counter is drawn randomly and uniformly from a contention window (CW)range of [0, CW]. The station decreases its backoff counter by one after every idle time slot. If the channel becomes busy during the backoff procedure because of transmission activity from other stations transmits, the station will suspend its backoff procedure and freeze the backoff counter decrement until the channel returns to the idle state. After the channel returns idle, the station must wait through a complete idle IFS before it may resume its routine backoff procedure. As long as the channel does not remain idle for a complete IFS, the station keeps suspending its backoff procedure. Once the backoff counter is decreased to zero, the station will start its transmission.

If the transmission is successful, the receiving station will return an acknowledgment (ACK) frame after waiting through another IFS. If the transmission fails, the station will retransmit the failed frame following the aforementioned backoff procedure. The value of CW for each transmission attempt may vary: for the first transmission attempt, the CW value is assigned to be CW_{min}, and it will be doubled after every unsuccessful transmissions until it reaches CW_{max}. Once the station reaches the maximum retransmission limit, it will drop the frame.

1.1.2 An Introduction to IEEE 802.11 DCF and IEEE 802.11e EDCA

DCF and EDCA are the fundamental access mechanisms in the IEEE 802.11 standard and IEEE 802.11e standard respectively, and both of them are based on the aforementioned CSMA/CA. In DCF, all stations use an identical parameter setting (including IFS size and CW size), and they compete for channel access fairly in a best effort manner. On the contrary, stations in EDCA uses different parameter settings according to the category of traffic they carry. In EDCA, traffic is classified into four access categories (ACs), including voice, video, best effort, and background. A distinct parameter setting is used by each AC, so that AC based service differentiation can be implemented.

1.1.2.1 IFS Differences between DCF and EDCA

In IEEE 802.11 DCF, every time a station starts or resumes its backoff procedure for a transmission, it must sense the channel being idle for an $DIFS$ (DCF IFS) or $EIFS$ (Extended IFS) duration depending on the result of the previous transmission on the radio channel. If the previous transmission is a successful transmission, $DIFS$ is used, otherwise $EIFS$ is used.

The duration of $DIFS$ is defined as

$$DIFS = SIFS + 2 \times timeslot, \tag{1.1}$$

Here $SIFS$ is the shortest IFS, and $timeslot$ is the duration of a time slot, and their values depend on the specific physical layer access mechanism used in the IEEE 802.11 network.

The duration of $EIFS$ is defined as

$$EIFS = SIFS + ACK + DIFS, \tag{1.2}$$

where ACK is the time for transmitting an ACK frame. Accidently, "SIFS+ACK" is equal to $ACKTimeOut$. It is the duration starting from the end of the last busy channel, and ending at the time point where an ACK frame should have been received. If no ACK frame is received and the channel remains idle during this $ACKTimeOut$ period, the transmission is considered unsuccessful.

Compared with the single $DIFS$ used in DCF, four AC based $AIFS$ (arbitration IFS) values are used in IEEE 802.11e EDCA, given by

$$AIFS = SIFS + AIFSN \times timeslot, \tag{1.3}$$

where $AIFSN$ is the AC based $AIFS$ number. Accordingly, there are four AC based $EIFS$ values, since $EIFS$ are defined as

$$EIFS = ACKTimeOut + AIFS. \tag{1.4}$$

In this thesis, the term"IFS" is used as a generic term representing various IFS values used in both DCF and EDCA. In the case that it is necessary to specify whether DCF or EDCA is being used, two terms, IFS_D and IFS_E are used. IFS_D represents the IFS used in DCF, and it can be either $DIFS$ or $EIFS$ depending on whether the previous transmission is successful or not. Also, the term IFS_E represents the IFS used in EDCA, and it can be either $AIFS$ or $EIFS$ depending on whether the previous transmission is successful or not.

Table 1.1 presents a summary of the differences between DCF and EDCA in IFS size [16, Table 20df, p.49]. According to Table 1.1, when a station starts or resumes a backoff procedure, DCF, AC video and AC voice in EDCA wait the shortest IFS duration, AC best effort waits a longer IFS duration, and AC background waits the longest IFS duration.

Table 1.1: IFS values in DCF and EDCA

Traffic category	IFS
DCF	DIFS=SIFS+ 2 $\times$ timeslot
EDCA voice	AIFS=SIFS+ 2 $\times$ timeslot,
EDCA video	AIFS=SIFS+ 2 $\times$ timeslot
EDCA best effort	AIFS=SIFS+ 3 $\times$ timeslot
EDCA background	AIFS=SIFS+ 7 $\times$ timeslot

1.1.2.2 CW Size Differences between DCF and EDCA

The differences between DCF and EDCA in CW size are summarised in Table 1.2 [16, Table 20df, p.49]. As shown in Table 1.2, in the coexistence environment, DCF, EDCA AC best

effort, and EDCA AC background use the same CW size, while EDCA AC voice and EDCA AC video use a smaller CW size.

Table 1.2: CW sizes in DCF and EDCA

Traffic category	CW_{min}	CW_{max}
DCF	CW_{min}	CW_{max}
EDCA voice	(CW_{min}+1)/4-1	(CW_{min}+1)/2-1
EDCA video	(CW_{min}+1)/2-1	CW_{min}
EDCA best effort	CW_{min}	CW_{max}
EDCA background	CW_{min}	CW_{max}

According to Tables 1.1 and 1.2, EDCA AC voice and AC video should have higher priority to channel access over DCF, because of smaller CW size used by these ACs, while DCF has higher priority over EDCA AC best effort and AC background because IFS_D is smaller than their IFS_E values.

1.1.3 Other Subtle Differences between DCF and EDCA

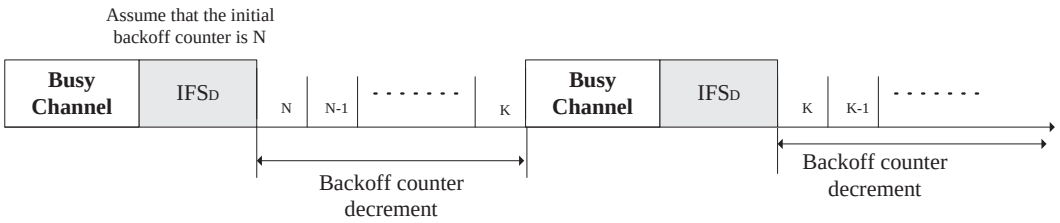
In addition to the differences in IFS value and CW size, there are some other subtle differences between DCF and EDCA.

1.1.3.1 Backoff Counter Decrement Rule

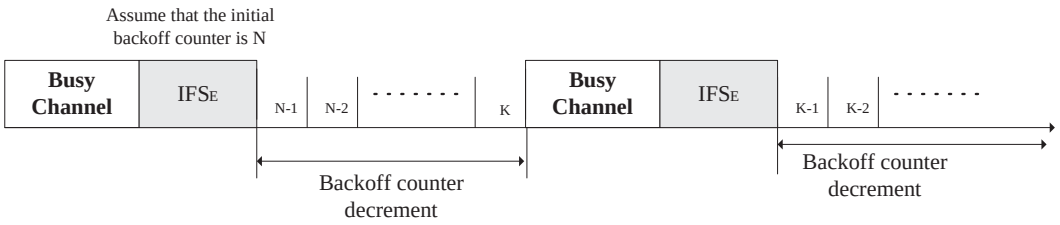
Although in both DCF or EDCA, a station shall decrease its backoff counter by one after every idle time slot, some subtle differences exist between their backoff decrement rules, which is shown in Fig 1.1. To assist the demonstration, two new terms are defined: QoS station (QSTA) and non QoS station (non-QSTA). These refer to stations using DCF and stations using EDCA, respectively.

In DCF, every time a non-QSTA waits through a complete idle IFS_D and starts a new backoff procedure or resumes a suspended backoff procedure, it must sense the channel as being

idle for an extra time slot following the IFS_D in order to decrease its backoff counter by one. This means that a non-QSTA actually must sense the channel as being idle for a complete $(IFS_D + timeslot)$ interval before it may decrease its backoff counter. In comparison, a QSTA in EDCA can decrease its backoff counter by one immediately following a complete idle IFS_E interval, and the decrement is independent of the channel status in the immediately following time slot. Fig 1.1 illustrates this difference.



(a) The backoff counter decrement rule in DCF



(b) The backoff counter decrement rule in EDCA

Figure 1.1: The difference between the backoff counter decrement rules in DCF and EDCA

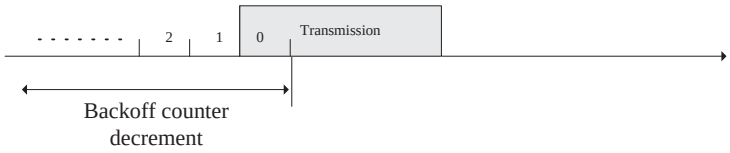
This difference in the backoff counter decrement rules may affect the performance when the number of contending stations in a WLAN (wireless local area network) system is large, because the large number of contending stations can cause a station's backoff procedure be frequently interrupted by any transmission activity from other stations. In that case, QSTAs may obtain higher priority over non-QTSAs, because they do not need to wait through the extra time slot required for non-QSTAs and can decrease their backoff counter more quickly than non-QSTAs can.

In addition, a special case should be noted. That is, if a QSTA or a non-QSTA starts a new backoff procedure with an initial backoff counter at zero, both of them can start a transmis-

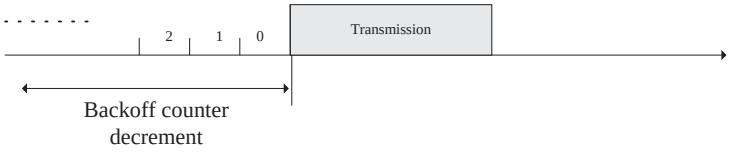
sion immediately after completing their respective IFS_E or IFS_D. It is the only scenario in which a non-QSTA does not need to wait through one extra time slot after the IFS_D.

1.1.3.2 The Time Instant for Starting a Transmission when the Backoff Counter Reaches Zero

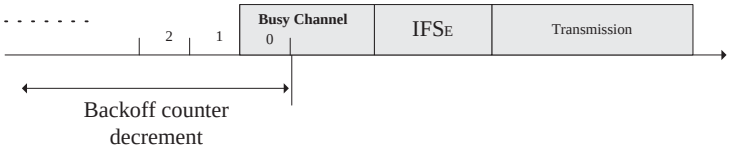
Subtle difference exists between non-QSTAs and QSTAs on the time instant to start a transmission when the backoff counter reaches zero. This is illustrated in Fig 1.2.



(a) The time instant for starting a transmission in DCF



(b) The time instant for starting a transmission in EDCA, case 1



(c) The time instant for starting a transmission in EDCA, case 2

Figure 1.2: The difference between DCF and EDCA on the time instant for starting a transmission when the backoff counter reaches zero.

As shown in Fig 1.2(a), a non-QSTA will start its transmission at the beginning of the time slot in which its backoff counter is decreased to zero. In contrast, the transmission of a QSTA will depend on the channel status in the time slot in which its backoff counter is

decreased zero. If the channel is idle in that time slot, the QSTA will start a transmission at the beginning of the immediately-following time slot, as illustrated in Fig 1.2(b). If the channel is busy in that time slot, the QSTA will wait through a complete IFS_E interval after the channel returns idle and start its transmission at the beginning of the time slot immediately following the completed IFS_E interval, as shown in Fig 1.2(c).

This difference may give non-QSTAs a slightly higher priority over QSTAs because QSTAs need to wait through an extra time slot in order to start a transmission. If channel becomes busy in that time slot, the QSTAs must wait until the channel returns to the idle state and remains idle for an IFS_E before they can start the transmission.

1.1.4 Data-rate Switching

Supporting multiple data rates for transmitting data frame has been included in IEEE 802.11 standards. This is independent of whether DCF or EDCA is used on the MAC layer. For example, four data rates are supported in IEEE 802.11b standard, including 1 $Mbps$, 2 $Mbps$, 5.5 $Mbps$, and 11 $Mbps$ [14, p. 10].The general concept for supporting multiple data rates is that a higher data rate is always preferred but a higher data rate requires a higher carrier to noise ratio (CNR) on the radio channel so that the bit error rate (BER) can remain at a relatively low level. Therefore, when the channel quality is good, a higher data rate is used so that the bandwidth can be fully used. On the contrary, when the channel quality is poor, a lower data rate is used so that the low BER, and accordingly, the low frame error rate (FER), can be guaranteed. However, there is no definition about the detailed data-rate switching mechanism in IEEE 802.11 standards. The mechanism through which a station switches its data rate is left to the IEEE 802.11 product vendors.

1.2 Background: Wireless Cooperative Retransmission

Wireless multi-hop networks have been widely used in many areas, such as wireless sensor networks, wireless mesh networks, and vehicular ad hoc networks [17]. Due to its great application potential, the wireless multi-hop network has been a research focus for more than a decade. The concept of the wireless multi-hop network is that traffic is transferred in a

hop-by-hop manner: each relay hop receives the traffic frame or packet from the previous hop and then forwards it to the next hop. With this approach, the signal strength can remain at a satisfactory level along the entire multi-hop path against the hostile character of the radio channel (compared with the wired channel). However, this approach has its own shortcomings:

- Its performance heavily depends on the selection of interim hops. An ideal interim hop should have good radio channels with its preceding hop as well as its proceeding hop. However, the availability of such perfect interim hops cannot be always guaranteed in a practical wireless network.
- The network resource may not be fully used in the wireless multi-hop network because only one pre-determined interim node is involved in the traffic transfer within a particular segment along the entire multi-hop path . For example, in a two-hop segment, only the node acting as the interim hop can forward the traffic to the next hop. While other nodes may also have good or even better radio channels, they cannot contribute to the traffic delivery.

Recently, the research interest in the wireless cooperative communication approach [18, 19] has increased as it may overcome the aforementioned problems of the traditional wireless multi-hop network. With the wireless cooperative communication approach, multiple forwarders within each segment along the entire multi-hop path may contribute to help the traffic delivery to the next hop. The idea of wireless cooperative communication is similar to MIMO (multiple-input and multiple-output) technology [20, 21]. MIMO technology uses multiple antennas at the transmitter and the receiver, and communication quality can be significantly improved due to the spatial diversity of the multiple-antenna system by using space time coding technology [22–25]. But different from the original MIMO ideas where the multiple antennas are installed on a single transceiver, the term “wireless cooperative communication” is referred to as a wireless communication system where users share and coordinate their resources to enhance the transmission quality. That is, the forwarders at different locations in a wireless cooperative network act as “antennas” in the MIMO system. Therefore, wireless cooperative communication can achieve a similar spatial diversity gain to that of MIMO system by using the multiple cooperative forwarders at different locations.

Wireless cooperative retransmission is a key concept in wireless cooperative communication. Its basic idea is that the retransmission, if the first transmission from the source fails, is handled by cooperative neighbours rather than the original source.

1.3 Motivations and Research Problems

1.3.1 IEEE 802.11 network performance analysis

Extensive attention has been attracted to developing analytical models of DCF and EDCA, because an accurate analytical model can be a fundamental base for analysing and improving their performance. However, there is still some room left for us to develop better analytical models:

- Most of the existing studies on EDCA performance analysis were finished before the publication of the IEEE 802.11e standard, and accordingly they were based on some draft proposals of EDCA, which are not fully consistent with the EDCA details in the IEEE 802.11e standard. This inconsistency may result in inaccuracy in their analytical models.
- With the publication of the IEEE 802.11e standard, a proliferation of IEEE 802.11e capable products is expected. Meanwhile, the traditional IEEE 802.11 capable products will exist for a considerably long period. Thus, there is significance for practice to investigate the network performance under the coexistence of IEEE 802.11 and IEEE 802.11e products. However, this problem has not attracted enough research attention, as most of the existing studies focus on IEEE 802.11 and IEEE 802.11e separately.
- Multiple data rates are defined in the IEEE 802.11 standard. However, the specific mechanism for data-rate switching is not defined in the standard and it is left for the vendors. It also results in that the majority of the existing studies on the IEEE 802.11 network performance analysis ignore the data-rate switching in IEEE 802.11.

In this thesis, several models are presented for analysing the IEEE 802.11 network performance, which address the aforementioned inadequacy in the existing studies.

1.3.2 Wireless Cooperative Retransmission

The majority of existing works in the area of wireless cooperative retransmission tackle this issue from physical layer perspective. That is, they focus on the use of some physical layer technologies, such as space time coding technology, so that they can fully use the spatial diversity gain of wireless cooperative retransmission without considering collision issue on higher layers (such as MAC layer). That is, multiple simultaneous transmissions can improve the transmission quality rather than mutually collide.

However, such a physical layer wireless cooperative retransmission approach requires additional equipments, such as MIMO transceivers supporting space time coding. It is more feasible to develop wireless cooperative retransmission strategies from the MAC layer perspective. Such MAC layer oriented wireless cooperative retransmission strategies can be implemented with simple and cheap equipments, such as traditional IEEE 802.11 adaptors.

Plenty of existing studies consider wireless cooperative retransmission from the MAC layer perspective, and majority of them use an $opportunistic forwarding$ approach. Such an approach uses some local coordination mechanism to choose one sole forwarder from several potential candidate nodes. Therefore, only one node is allowed to transmit and the collision event can be avoided. Such an approach may work well for the multi-hop scenarios, but it may not be suitable for the single-hop scenarios which are still common in the wireless cooperative networks. The local coordination mechanism may appear complex for the single-hop scenarios and it may cause extra retransmission delay.

Compared with the $opportunistic forwarding$ approach, the uncoordinated distributed wireless approach appears more suitable for the single-hop scenarios. In such an approach, all cooperative nodes may transmit and they do not agree on that one of them should be chosen as the sole forwarder. Such an approach does not need a local coordination system and accordingly it may avoid the related retransmission delay.

Only limited existing studies consider the uncoordinated distributed cooperative retransmission strategies. However, they ignore the collision issue by assuming the use of some physical layer technology, such as space time coding technology. On this point, these studies are still tackling wireless cooperative retransmission from the physical layer perspective. In this thesis, some uncoordinated distributed cooperative retransmission strategic are presented, in

which the collision issue on the MAC layer is carefully considered.

1.4 Organisation of The thesis

The rest of this thesis is organised as follows.

- Chapter 2 reviews literature on the performance analysis of DCF and EDCA.
- Chapter 3 presents an analytical model of EDCA. A two-dimensional Markov chain model is used to analyse the performance of EDCA under the saturated traffic load. Compared with the existing analytical models of EDCA, the proposed model incorporates more features of EDCA into the analysis. Based on the proposed model, saturated throughput of EDCA is analysed. Simulation study is performed, which demonstrates that the proposed model has better accuracy than those in the literature.
- Chapter 4 presents an analytical model for the coexistence of DCF and EDCA. A three-dimensional Markov chain model is used to investigate the coexistence of 802.11 DCF and 802.11e EDCA stations. The performance impact of the differences between 802.11 DCF and 802.11e EDCA is carefully analysed. Based on the proposed model, the saturated throughput is analysed. Simulation study is carried out to evaluate the accuracy of the proposed model.
- Chapter 5 presents an analytical model of DCF using data-rate switching. A hierarchial Markov chain model is used to analyse the performance of IEEE 802.11 DCF, considering a commonly used data rate switching mechanism. In this analysis, the switching between multiple data rates may be triggered by either collisions or transmission errors. Simulation results are presented which verify the accuracy of the proposed model and demonstrate the effect of the data rate switching mechanism.
- Chapter 6 reviews literature on wireless cooperative retransmission.
- Chapter 7 presents two uncoordinated distributed wireless cooperative retransmission strategies and analyses them in a memoryless channel model and a two-state Markov fading channel model respectively. The numerical study and the simulation study are

carried out to evaluate the superior performance of the proposed retransmission strategies over the retransmission by the source or by one relay.

- Finally, Chapter 8 concludes this thesis.

Chapter 2

Literature Review: IEEE 802.11 Network Performance Analysis

The previous chapter has shown that DCF and EDCA use a distributed network management mode: each station performs its own backoff procedure independently for channel access. This distributed mode can significantly reduce the system complexity, because there is no need to set a powerful central controller in the system. However, this distributed mode also results in difficulty regarding the performance analysis of DCF and EDCA, due to the independence of each station. The introduction of AIFS difference in EDCA especially complicates the analysis. In DCF, all stations use an identical $DIFS$ and no station can transmit within the IFS_D (that is, $DIFS$ or $EIFS$) duration, and any interruption (that is, the transmissions from other stations) to their backoff procedure cannot occur within the IFS_D duration. On the contrary, in EDCA, stations use different AIFS values. Some stations may finish their shorter IFS_E (that is, $AIFS$ or $EIFS$) duration and transmit while other stations still wait through their longer IFS_E duration. Therefore, the backoff procedure of some stations in EDCA may be interrupted during their longer IFS_E duration.

2.1 Related Work about IEEE 802.11 Network Performance Analysis

Extensive work has been undertaken to analyse DCF and EDCA separately [26–45]. The majority of these existing studies use Markov chain models to analyse the IEEE 802.11 network performance, including those in [26–41]. In [26], Bianchi proposes a two-dimensional multiple-layer Markov chain model for modelling the backoff procedure in DCF. Each layer in the Markov chain model represents a backoff procedure of a transmission attempt, and each state in a layer represents a specific backoff counter value in the corresponding backoff procedure. Based on the work in [26], many Markov chain based analytical models have developed for EDCA [27–41]. The effect of using different $AIFS$s and CW sizes are analysed in those models for EDCA, but some limitations exist among them, which leaves room for us to develop a better model to achieve more accurate analytical results.

In [27–33], some Markov chain models are developed based on that in [26]. Different contributions are made to develop these Markov chain models so that they can be used for EDCA performance analysis, such as the zone specific transmission probability analysis presented in [27, 28], which considers the effect of using different $AIFS$s, the delay analysis in [30], and the Z-transform approach in [31].The zone specific transmission probability analysis presented in [27, 28] should empathized, as it is widely used in the models proposed in this thesis. As aforementioned in Chapter 1, stations using different AIFS values will start or resume their backoff procedure at different time slots after the busy channel. Therefore, the time slots after the busy channel can be classified into different zones, where different set of stations may transmit. A common problem exists among the work in [27–33]: the possibility that a station's backoff procedure may be interrupted by the transmissions from other stations is ignored or not clearly analysed in their Markov chain models. As will be shown later in our analysis in Chapter 4, this will have a significant impact on the accuracy of the Markov chain model.

Compared with those in [27–33], the models presented in [34–36] consider the above backoff interruption possibility. In [34, 35], the backoff interruption possibility is considered by adding a transition for each backoff state, and this transition starts and ends in the same state. This represents that the possible backoff interruption in the corresponding backoff stage. In [36], the backoff interruption possibility is considered by using some extra states

in addition to each backoff state to represent this possibility. However, some potential flaws may exist, which results in the effect of using different $AIFS$s not being correctly analysed. First, the self-transition in [34, 35] will cause inaccuracy, because this self-transition does not consider the difference between the backoff decrement procedures in the backoff interruption scenario and in the normal backoff scenario. When a station's backoff procedure is interrupted by the transmissions from other stations, this station must wait through a complete idle IFS before it can decrease its backoff counter. On the contrary, in the normal backoff scenario, a station only needs to wait an idle time slot in order to decrease its backoff counter. Such scenario specific difference will result in the different probabilities whether the station can decrease its backoff counter or not. In the backoff interruption scenario, such a probability is the probability that no other stations transmit during a complete IFS. In the normal backoff scenario, it is the probability that no other stations transmit in one time slot only. However, this difference is not considered in [34, 35] as they use the same probability for both scenarios. Second, it is considered in [36] that all stations using different $AIFS$s may transmit in any time slot after the busy channel, but in fact some stations using larger IFS cannot transmit in some time slot because they are still waiting through their IFS duration.

Finally, some Markov chain models consider both the effect of backoff interruption possibility and the effect of using different $AIFS$s [37–41]. In [37–40], a three-dimensional Markov chain model is used for the lower priority traffic flow with a larger IFS, where the third dimension is a stochastic process representing the possible backoff interruption. In [41], an extra stochastic process is used in its three-dimensional Markov chain model to represent the number of time slots that have been passed since the end of a transmission. The three-dimensional Markov chain models used in [37–41] have some extra states representing the possible backoff interruption, and the effect of using different $AIFS$s is considered when analysing the transition probabilities among those states. However, some limitations exist among them in addition to a complex Markov chain architecture being used. Firstly, it is assumed in [37, 41] that a station will keep retransmitting until the frame has been successfully transmitted. The possibility that the frame may be dropped after reaching the maximum retransmission limit is not considered. Secondly, the two-dimensional Markov models for high-priority traffic flow in [37, 38] do not consider the possibility that the backoff procedure of a station with high-priority traffic flow may also be interrupted by transmissions from other stations. Thirdly, a problem exists for defining the transition probabilities between different backoff stages in the Markov chain model in [41]. That is, a station will obtain a

random initial value among the range $[0, CW]$ for its backoff counter when it starts a new backoff procedure, and the probability that it obtains a specific value within this range should be $\frac{1}{CW+1}$. However, this probability is considered to be 1 in [41]. Finally, an approximation has been made in [39, 40] to simplify the analysis on the AIFS difference, and it will cause inaccurate results as shown in later chapter. The Markov chain models in [39, 40] cannot accurately trace the zone specific difference defined in [27, 28] for each idle time slot after the busy channel. Therefore it has been approximated in [39, 40] that such idle time slots are located at the same zone where all other stations may transmit.

The use of Markov chain model has the advantage that a well-designed Markov chain can easily model and fully capture the complexity of the backoff procedure. However, using Markov chain models results in a complex non-linear equation system. It is hard to obtain the closed-form solution of the equation system, and the equations can be numerically solved only.

Comparatively, some researchers try to analyse DCF or EDCA with a non-Markov approach (that is, they do not use the Markov chain models to model the backoff procedure of stations), and its advantage is that a simpler equation system, or even a closed-form solution may be obtained. In [42], Venkatesh $et\ al.$ propose a so-called fixed-point approach for analysing DCF and EDCA separately, and this approach can generate a much simpler non-linear equation system. The results from this approach are very close to those from the Markov chain approaches. In [43, 44], the authors obtain a closed-form solution for the saturated throughput for EDCA, using elementary probability theory directly. The disadvantage of the non-Markov approach is that it is difficult to fully capture the complexity of the backoff procedure. For example, the backoff interruption possibility is not considered in [42–44]. Additionally, the work in [45] should be mentioned. A Markov chain model is used in [45] to model the number of stations at different backoff stages, compared with all the aforementioned Markov chain models that model the backoff procedure. Like the work in [42–44], it can significantly simplify the analysis but it has to ignore some details of the backoff procedure, such as the possibility of backoff interruption.

Furthermore, it should be noted that the coexistence of DCF and EDCA has not been well considered among the existing studies in this area. In [46, 47], some detailed differences between DCF and EDCA are discussed, and the simulation or experimental results about the coexistence of DCF and EDCA are demonstrated, but an analytical model has not been

presented. In [48], an analytical model is proposed to analyse the coexistence of DCF and EDCA with elementary probability theory (that is, no Markov chain model is used). However, only two $AIFS$ values of EDCA are considered in [48] for simplicity of analysis, that is, $AIFS = SIFS + 2 \times timeslot$ and $AIFS = SIFS + 3 \times timeslot$. However, as mentioned in Table 1.1 in the previous chapter, the third $AIFS$ value exists in EDCA, that is, $AIFS = SIFS + 7 \times timeslot$ for background traffic. The analysis in [48] cannot be easily modified to include this $AIFS$ value.

Finally, it should be noted that most of the aforementioned existing studies do not use the accurate EDCA parameter setting defined in the IEEE 802.11e standard. This could be caused by the fact that most of them were finished before the final publication of IEEE 802.11e standard.

Additionally, there is a lack of analytical work to investigate the impact of data-rate switching mechanism on the IEEE 802.11 network performance. The aforementioned studies consider a single data rate only. Some studies investigate the impact of multiple data rates on the performance of the IEEE 802.11 networks, such as those in [49–54]. However these existing studies only consider the situation that each station uses a fixed data rate. The possibility that stations can switch their data rates dynamically by using some data-rate switching mechanism is not considered in [49–54]. Most research attention about the data-rate switching mechanism in the IEEE 802.11 networks focus on proposing various data-rate switching mechanisms, such as those in [55–62]. The performance of these proposed data-rate switching mechanisms is evaluated using simulations or experiments only, and an analytical model is lacking. Such a lack of analytical work in this area may be attributable to the fact that no data-rate switching mechanism is defined in the IEEE 802.11 standards. Since most existing IEEE 802.11 products can support multiple data rates and implement some kind of data-rate switching mechanism, it is practice of significance to investigate the impact of the data-rate switching mechanism.

2.2 Most Recent Work on IEEE 802.11 Network Performance Analysis

Since the completion of my analytical work on IEEE 802.11 network performance analysis in early 2007, which has been published in [63–67], a considerable amount of studies have been published in the same area, such as those in [68–80]. We have included these studies in the literature review for completeness.

Compared with the previous studies that usually only consider simple analytical scenarios with saturated traffic load and single-hop connection, the majority of these most recent studies focus on more complex analytical scenarios. In [68, 69], the performance of the IEEE 802.11 networks for video traffic transmission is analysed. In [70], the authors consider the performance of the IEEE 802.11 networks under TCP (transmission control protocol) protocol. In [71], the impact of multiple data rates is analysed. In IEEE 802.11 DCF, all stations contend for the channel access fairly, despite that stations using a lower data rate may occupy the channel for a longer time once they obtain the channel access. Therefore, such channel access fairness for stations using different data rates may result in a negative impact on the system capacity. In [71], the authors propose that stations using a higher data rate should use an optimal set of MAC layer parameters so that they have a higher priority for the channel access to achieve a larger system capacity. In [72], a Markov chain model is proposed for analysing the performance of IEEE 802.11 DCF under a Poisson traffic load. In [73–80], the performance of the multi-hop IEEE 802.11 networks are analysed. Such an analysis appears significantly interesting and challenging compared with that for the single-hop scenarios. In the multi-hop scenarios, each hop along the multi-hop path receives the traffic passed by its preceding hops, and it will relay the traffic to its proceeding hops. Therefore, the traffic load on each hop is mutually related. Such relationships add extra complexity to the analysis. In [73], the traffic patterns of a multi-hop path is investigated. The authors observes that only the first few hops along the multi-hop path have large traffic queues, while the traffic queues in the last few hops are very small. In [74, 75], the channel capacity of a multi-hop path is analysed. In [76], the performance of EDCA on a multi-hop wireless vehicular ad hoc network is investigated. In [77], the impact of the traffic sending rate at the source node along a multi-hop path is analysed. The authors observes that the sending rate should be adjusted appropriately to achieve the maximum end-to-end throughput. In [78, 79], the fair-

ness issue among hops along a multi-hop path is studied. If each hop can fairly occupy the channel, the performance of the multi-hop path can be maximised. However, to achieve such a fairness is not easy if each hop uses a contention based channel access mechanism, such as DCF or EDCA. In [80], the authors propose some optimal setting for the contention window (CW) size in order to maximise the performance of a multi-hop path under TCP protocol.

It should be mentioned that usually some simplifications must be made for the analysis on these more complex scenarios, otherwise the analysis work may appear significantly difficult. For example, the analysis in [72] has been considerably simplified with a so-called system approximation technique [81]. This system approximation technique may approximate the system as a versatile queueing model which is easier to be analysed.

2.3 Summary

In this chapter, a literature review about IEEE 802.11 network performance analysis has been performed. Some previous publications in this area have been discussed. It can be summarised as follows:

1. The existing analytical models for IEEE 802.11 EDCA can still be improved.
2. There is a lack of analytical models for the coexistence of DCF and EDCA.
3. There is a lack of analytical models to investigate the performance of the IEEE 802.11 network with data-rate switching.

Chapter 3

Saturated Throughput Analysis of EDCA

In this chapter, an analytical model is proposed for investigating the performance of EDCA. Compared with the existing work mentioned in the previous chapter, the proposed analytical model includes more details of EDCA. Consequently, its analytical results are more accurate than those ignoring these details.

This chapter uses the following structure: in Section 3.1, details of EDCA are introduced, and these details are considered in the proposed analytical model; in Section 3.2, the fundamental two-dimensional Markov chain model is proposed; in Section 3.3, the saturated throughput performance is analysed based on the proposed Markov chain model; in Section 3.4, the simulation study is performed; finally, this chapter is summarised in Section 3.5.

3.1 Details of EDCA Considered in the Proposed Analytical Model

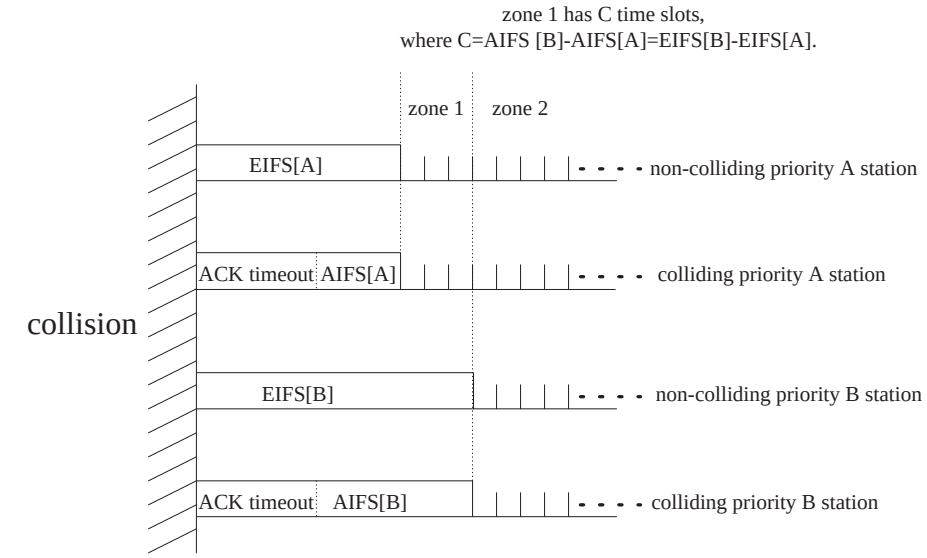
To investigate the performance of EDCA, an accurate analytical model is necessary. In addition to the effect of using different CW sizes that has been well investigated in the existing studies mentioned in the previous chapter, some other important factors should be carefully considered for an accurate analysis of EDCA performance:

Firstly, the effect of using different $AIFS$s should be carefully considered. Fig 3.1 indicates that AC A stations with a smaller $IFS[A]$ may begin their backoff procedure and transmit after $IFS[A]$, while AC B stations with a larger $IFS[B]$ are still in the backoff suspension procedure and can not transmit. When $IFS[B]$ is completed, both sets of stations can begin their backoff procedure and transmit. Therefore the time period from the end of the busy channel can be classified into different intervals, referred to as contention zones in this chapter, depending on the different transmission probabilities of different sets of stations in each zone caused by the use of different $AIFSs$.

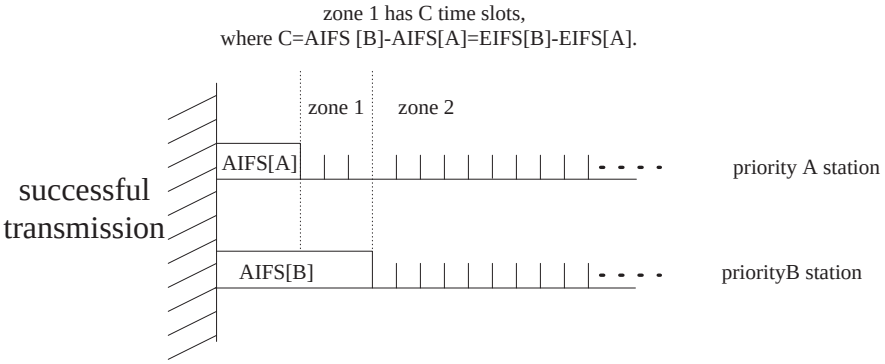
Secondly, the possibility of backoff suspension should be analysed. As mentioned earlier in the previous chapter, before the start of a new backoff procedure, as well as every time the channel becomes busy during the backoff procedure, the station may experience a backoff suspension procedure. The occurrence of backoff suspension depends on the channel status, which is affected by the activities of other stations. Moreover, while a station is in the backoff suspension procedure, the transmission activity from other stations may occur before the station waits through an idle IFS. In this case the station must wait through another complete idle IFS after the channel returns to the idle state. Therefore the exact duration of each backoff suspension procedure is uncertain since it is affected by the transmission activity from other stations. It is obvious that the occurrence and the duration of the backoff suspension procedure can affect the performance of EDCA.

Moreover, some other details of EDCA are also considered by the proposed analytical model:

- In the case that a collision happens, colliding stations (that is, stations involved in the collision) will wait through an $ACK\ timeout$ duration to detect the collision, and then they will wait an $AIFS$ before starting another backoff procedure. According to [82], the sum of the $ACK\ timeout$ duration and an $AIFS$ is equal to an $EIFS$. Non-colliding stations (that is, stations not involved in the collision) also wait an $EIFS$ after a collision [6, clause 9.2.5.2, pp.77-79]. Fig 3.1(a) depicts this situation. Therefore, all stations wait an $AIFS$ from the end of the busy channel after a successful transmission, and wait an $EIFS$ (or an equivalently $AIFS + ACK\ timeout$) from the end of the busy channel after a collision. As mentioned previously (Chapter 1, Section 1.1.2.1), we still use the term "IFS" to represent both $AIFS$ and $EIFS$ in this chapter, when there is no need to specify their difference.



(a) after a collision



(b) after a successful transmission

Figure 3.1: The contention zone specific transmission probability.

- As mentioned previously (Chapter 1, Section 1.1.3.1), a station decreases its backoff counter by one at the beginning of a time slot during its backoff procedure. This means whether the backoff counter is decreased or not depends on the channel status in the previous time slot. This backoff counter decrement is independent of whether the channel is busy or not in the current time slot. Furthermore, every time the station leaves a backoff suspension procedure after completing an IFS, its non-zero backoff counter will be decreased by one at the beginning of the immediately following time slot, and this decrement is independent of the channel status in that time slot [16, clause 9.9.1.3, pp.81-83], [46].

- When the backoff counter is decreased to zero at the beginning of a time slot, the station will start its transmission at the beginning of the next time slot, provided that there is no transmission from other stations in the current time slot. Otherwise the station will enter into a backoff suspension state to wait through a complete idle IFS and start its transmission at the beginning of the immediately following time slot [16, clause 9.9.1.3, pp.81-83], [46].

3.2 A Markov Chain Based Analytical Model

In this section, we present the proposed analytical model of EDCA using Markov chain. Firstly, the basic Markov chain models are proposed. Secondly, the transition probabilities for the proposed Markov chain models are analysed, where the contention zone specific transmission probability caused by using different $AIFS$s is analysed following the method in [27]. Finally, a solution for the Markov chain models is obtained. The following assumptions are made in our analysis.

- Traffic load is saturated. That is, traffic is always backlogged at each station.

- Only two ACs are considered: AC A and AC B. AC A has higher priority than AC B and $AIFS[A] < AIFS[B]$. However, our analysis can be easily extended to include more than two ACs.

- Each station carries traffic from one AC only. Thus a station may be referred to as an AC A station or an AC B station, depending on the AC of the traffic it carries.

- Only one frame is transmitted in each $TXOP$ (Transmission opportunity).
- A WLAN system with a fixed number of stations is considered in our analysis. The number of stations for AC A and AC B is denoted by n_A and n_B respectively. n_A and n_B are known numbers.
- The transmission probability of a station in a generic time slot is a constant, which is determined by its AC only. This is an assumption widely adopted in the area [27, 31–36, 38]. The transmission probabilities of an AC A station and an AC B station in a generic time slot are represented by τ_A and τ_B respectively. The values of τ_A and τ_B are unknown and need to be solved. Here the term "generic time slot" refers to as the time slot following an idle IFS because it is not possible that a transmission occurs within IFS.
- The wireless channel is ideal. That is, there is no noise, no external interference and hidden station problems. Moreover, the channel is perfectly synchronised, and the propagation delay can be ignored. That is, all stations can immediately sense the channel busy or idle, and they can perform their backoff procedure synchronously. Unless otherwise specified, such ideal wireless channel assumption is applied to all Markov chain models for IEEE 802.11 network performance analysis presented in this thesis.

3.2.1 Two Discrete Time Two-dimensional Markov Chain Models

3.2.1.1 The Basic Markov Chain Models

Fig 3.2 and Fig 3.3 illustrates two discrete time two-dimensional Markov chain models for an AC A station and an AC B station respectively. Each Markov chain model represents the channel contention procedure for a station of a specific AC. For ease of illustration, we use the symbol "C" to represent $AIFS[B] - AIFS[A]$. There are two stochastic processes within the Markov chain model. The first process, denoted by $w(t)$, is used to model the decrement of the backoff counter during the backoff procedure of the station. Here a special value of $w(t) = -1$ is used to represent the station's own transmission, which includes the idle $IFS[A]$ immediately following the end of the busy channel as no frame transmission is

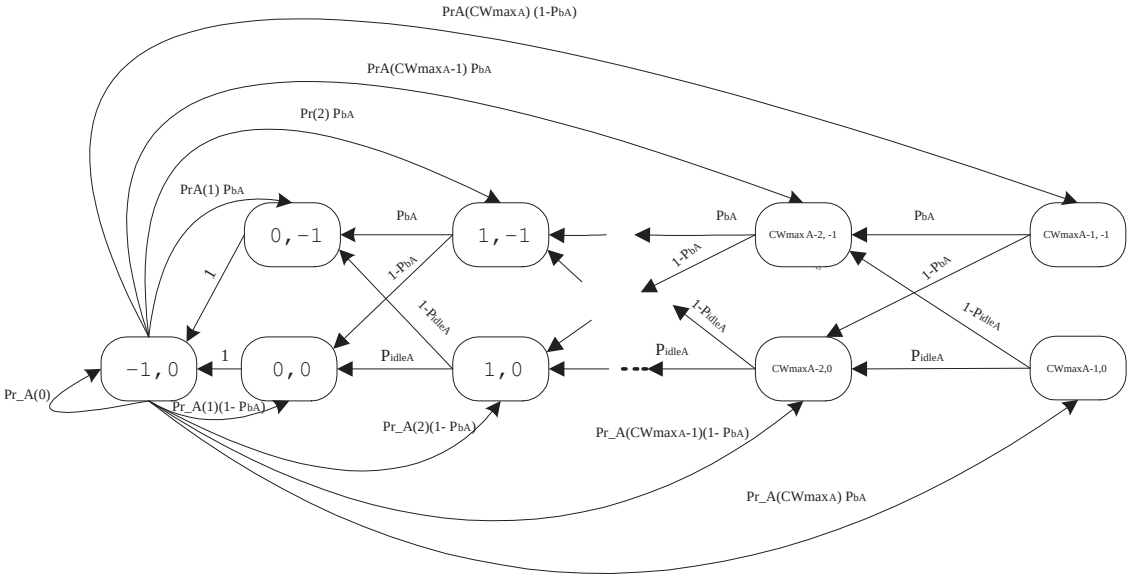


Figure 3.2: The Markov chain model for AC A.

possible during this interval. The second process, denoted by $v(t)$, is used to model the backoff suspension procedure. $v(t) = 0$ indicates the station is in the normal backoff procedure or is transmitting its own frame. When the station is in the backoff suspension procedure, $v(t)$ is non-zero and its value represents the number of idle time slots after the idle $IFS[A]$ following the end of the busy channel. Here we use a special value of $v(t) = -1$ to represent a frame transmission from other stations, which also includes the idle $IFS[A]$ immediately following the end of the busy channel.

In both Markov chain models, states $(r, 0)$, $0 \leq r \leq CW_{max} - 1$ represent an idle time slot in the normal backoff procedure, where r represents the value of the backoff counter. States $(r, -1)$, $0 \leq r \leq CW_{max} - 1$ represent a transmission activity (that is, it may be either a successful transmission from one station, or a collision caused by multiple transmissions from multiple stations) from other stations, which includes the idle $IFS[A]$ following the end of the busy channel, and r represents the corresponding value of the backoff counter. The special state $(-1, 0)$ is used to represent the station's own transmission, which includes the idle $IFS[A]$ following the end of the busy channel.

Another special state $(-1, -1)$ is used to represent a transmission activity from other stations, which occurs before the completion of the IFS immediately following the end of the

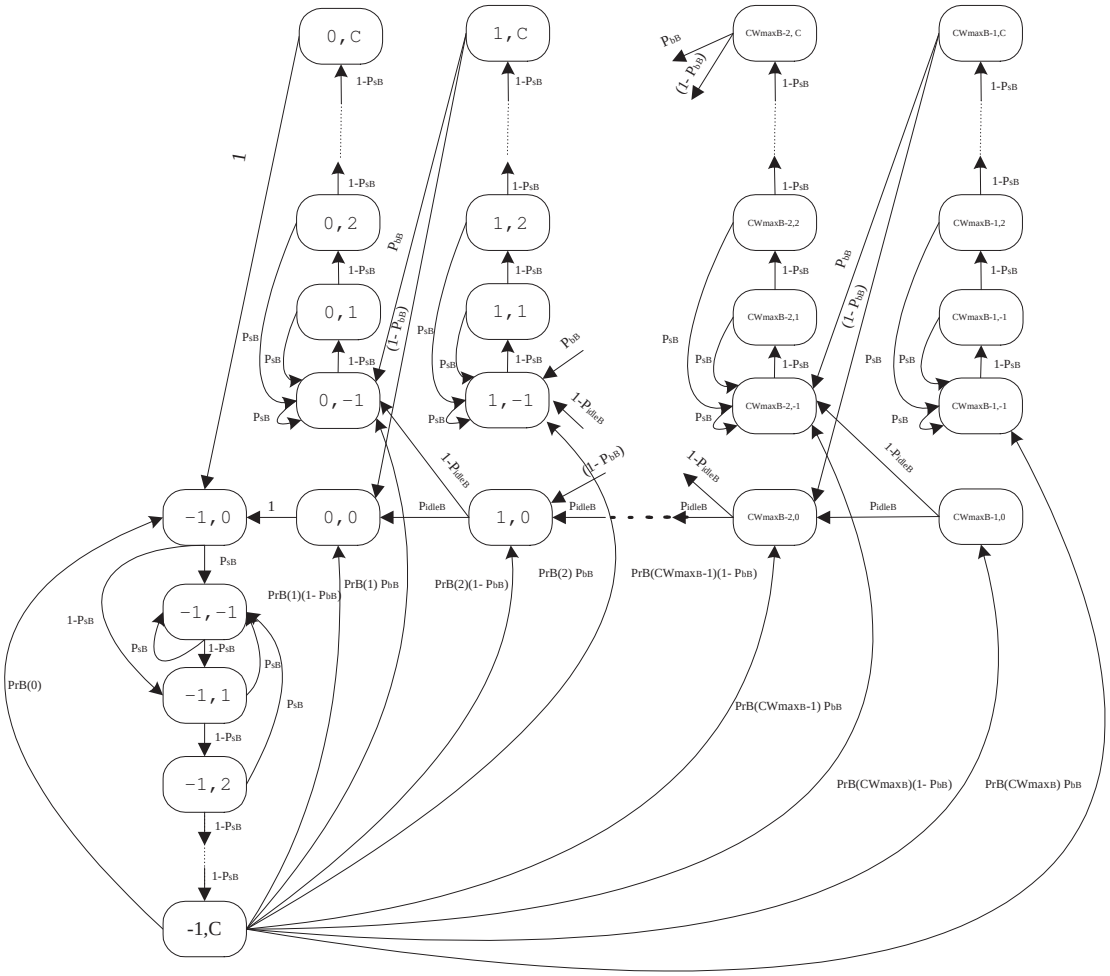


Figure 3.3: The Markov chain model for AC B.

busy channel caused by the station's own transmission. As usual, this state also include the idle $IFS[A]$ following the end of the busy channel. This special state only exists for AC B stations, because the transmission activity from other AC A stations is possible before an AC B station completes the idle $IFS[B]$ immediately following its own transmission.

After leaving the state $(-1, -1)$, an AC B station may traverse each state $(-1, k)$, $1 \leq k \leq C$ if the channel remains idle. The state $(-1, k)$, $1 \leq k \leq C$ represents an idle time slot in the backoff suspension procedure, where k indicates the number of idle time slots after the idle $IFS[A]$ following the end of the last busy channel. If the channel becomes busy due to the transmission activity from other AC A stations before the state $(-1, C)$ is reached, the station will move back to the state $(-1, -1)$. After reaching the state $(-1, C)$, the station will start a backoff procedure with a random initial backoff counter. Similarly, states $(r, -1)$ and (r, k), $0 \leq r \leq CW_{maxB} - 1, 1 \leq k \leq C$ are used to model the backoff suspension procedure, which occurs when the normal backoff procedure has been started. An AC B station in state (r, C), $1 \leq r \leq CW_{maxB} - 1$ may transit to either $(r - 1, 0)$ or $(r - 1, -1)$, depending on whether there is a transmission activity from other stations.

For an AC A station, since no transmission is possible during the $IFS[A]$ following the end of the busy channel, the states (-1, -1), (-1,k), (r,-1), and (r,k), $0 \leq r \leq CW_{maxA} - 1$, $1 \leq k \leq C$ do not exist for AC A stations.

The embedding points of the Markov chain models can be readily determined from the earlier definition of the states. Fig 3.4 depicts an example of the embedding points used in the Markov chain models.

In this example, the channel turns busy because of a transmission activity at time point t. After the busy status ends, the channel will remain idle until $C + 1$ time slots following the idle $IFS[A]$ have elapsed. The following time points, t+k, $1 \leq k \leq C + 2$, are located in the time slot boundary, as shown in Fig 3.4. We describe AC A and AC B stations that transmit during the transmission activity starting at time point t as transmitting AC A and AC B stations. Accordingly, we describe AC A and AC B stations that do not transmit during this transmission activity as non-transmitting AC A and AC B stations. At time point A, all transmitting AC A and AC B stations will enter the state $(-1, 0)$, and all non-transmitting AC A or AC B stations will suspend their backoff procedure and enter the state $(r, -1)$, where the value of r is station-specific. At time point $t + 1$, following the completion of

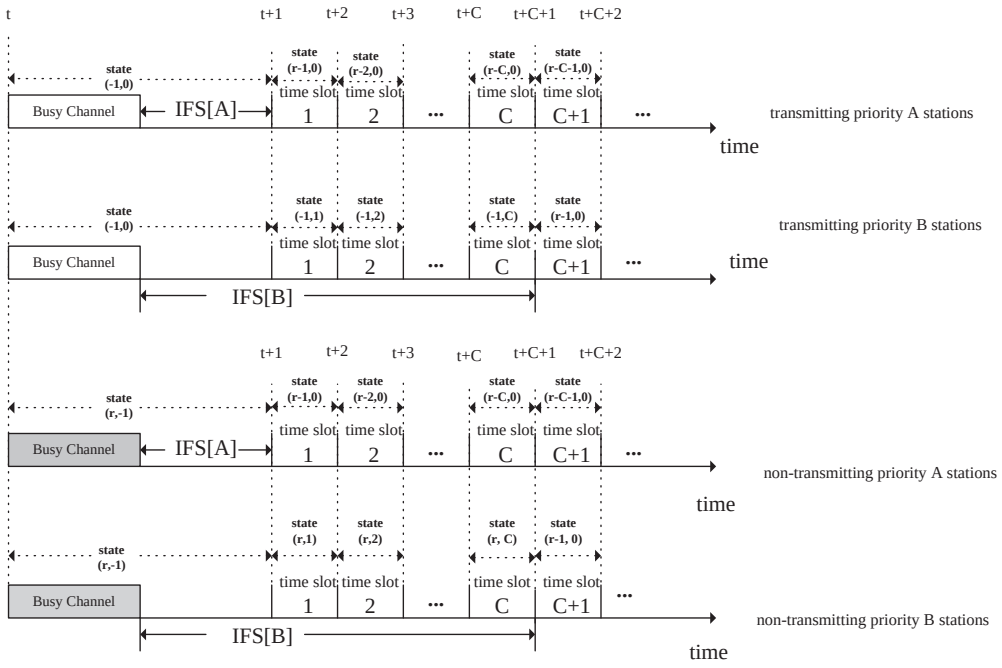


Figure 3.4: An example of the embedding points used in the proposed model.

the idle $IFS[A]$ from the end of the busy channel, all AC A stations start or resume their backoff procedure. For AC A stations , they will start a new backoff procedure with a station-specific random initial backoff counter r. As mentioned earlier in Section3.1 in this chapter, a station will decrease its backoff counter by one at the end of the idle IFS. Therefore, all AC A stations will enter the state $(r-1,0)$ at time point $t+1$, and their backoff counter will be decreased by one following each idle time slot. For AC B stations, they will start to traverse a series of states $(-1,k)$, (for transmitting AC B stations), or (r,k) (for non-transmitting AC B stations), $1 \leq k \leq C$ at time point $t+1$, and they will leave the state (r,C) at time point $t+C+1$ and enter the state $(r-1,0)$ to begin or resume a normal backoff procedure. Here $r-1$ also represents that their backoff counter is decreased by one following the completion of the idle $IFS[B]$ following the end of the busy channel. Then all AC B stations can also decrease their backoff counter by one following each idle time slot by entering the corresponding state.

It should be noted that some special scenarios are not included in the aforementioned example for ease of illustration. For example, non-transmitting AC B stations may enter the state $(-1,-1)$ at time point t, or at least one station has decreased its backoff counter to zero before the time point $t+C+1$ is reached. They are explained more clearly in the following

description for the one-step transition probabilities.

3.2.1.2 Transition probabilities for the Markov chain model of AC A

The one-step transition probabilities for the Markov chain model in Fig 3.2 are explained in the following. It should be noted that some unknown parameters, including P_{idleA}, P_{bA}, and $Pr\_A(r)$, are used here. They will be analysed later in Section 3.2.2.

1. When a specific AC A station finishes a transmission and completes the following idle $IFS[A]$, the station will leave the corresponding state $(-1, 0)$ and move into the next state to start a new backoff procedure with an initial backoff counter r at the beginning of the immediately following time slot. As previously described in Section 3.1, the backoff counter will be decreased by one following the end of the $IFS[A]$. Therefore, the backoff counter will be decreased to $r - 1$ as the station reaches the next state. Moreover, the channel status at this moment decides the next state in the Markov chain model: the state $(r - 1, -1)$ (if the channel turns busy with a probability of P_{bA}), or the state $(r - 1, 0)$ (if the channel remains idle with a probability of $1 - P_{bA}$).

$$\begin{cases} P\{(r-1,1)|(-1,0)\} = P_{bA} Pr\_A(r), \\ P\{(r-1,0)|(-1,0)\} = (1 - P_{bA}) Pr\_A(r), \end{cases} \tag{3.1}$$

   where $Pr\_A(r)$ is the probability that the AC A station starts a new backoff procedure with a random initial backoff counter r. For the special case that the initial backoff counter is zero, the station may start a transmission at the beginning of the immediately following time slot, independent of the channel status in this time slot:

$$P\{(-1,0)|(-1,0)\} = Pr\_A(0). \tag{3.2}$$

2. If the station reaches the state $(r, 0)$, it will reside in this state for an idle time slot. Then the station will decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. The channel status at this moment decides the next state: the state $(r-1, -1)$ (if the channel becomes busy with a probability of $1 - P_{idleA}$) or the state $(r - 1, 0)$ (if the channel remains idle with a probability of P_{idleA}).

$$\begin{cases} P\{(r-1,-1)|(r,0)\} = 1 - P_{idleA}, \\ P\{(r-1,0)|(r,0)\} = P_{idleA}. \end{cases} \tag{3.3}$$

For the special case that r equals zero, a station will remain in the state $(0,0)$ for an idle time slot and start the frame transmission at the beginning of the immediately following time slot with a probability of 1:

$$P\{(-1,0)|(0,0)\} = 1. \tag{3.4}$$

3. If the station reaches the state $(r,-1)$, it will remain in this state until the idle $IFS[A]$ following the end of the busy channel is completed. Then it will decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. The channel status at this moment decides the next state: the state $(r-1,-1)$(if the channel becomes busy with a probability of P_{bA}) or the state $(r-1,0)$ (if the channel remains idle for a probability of $1-P_{bA}$).

$$\begin{cases} P\{(r-1,-1)|(r,-1)\} = P_{bA}, \\ P\{(r-1,0)|(r,-1)\} = 1-P_{bA}. \end{cases} \tag{3.5}$$

For the special case that r equals to zero, a station shall stay in the state $(0,-1)$ until the idle $IFS[A]$ following the end of the busy channel is completed, and the station will start a transmission at the beginning of the immediately following time slot with a probability 1:

$$P\{(-1,0)|(0,-1)\} = 1. \tag{3.6}$$

3.2.1.3 Transition probabilities for the Markov chain model of AC B

As for the Markov chain model in Fig 3.3, its one-step transition probabilities are slightly different from those for the Markov chain model in Fig 3.2, because extra states $(-1,-1)$, $(-1,k)$, $(r,-1)$ and (r,k), $0 \le r \le CW_{maxB}-1$, $1 \le k \le C$ are used to represent the C (that is, $AIFS[B]-AIFS[A]$) idle time slots remaining in the $IFS[B]$ and the possible transmission activity from AC A stations during this time interval. The details of its one-step transition probabilities are explained in the following. Also, some unknown parameters, including P_{idleB}, P_{sB}, P_{bB}, and $Pr\_B(r)$, are used here. They will be analysed later in Section 3.2.2.

1. When a specific AC B station finishes its frame transmission including the idle $IFS[A]$ following the end of the busy channel, it will leave the corresponding state $(-1,0)$.

The station still needs to complete the C idle time slots remaining in its $IFS[B]$ before it can start a new backoff procedure. The station will move into the next state at the beginning of the immediately following time slot, and the channel status at this moment decides the next state: the state $(-1,-1)$ (if the channel becomes busy with a probability of P_{sB}), or the state $(-1,1)$ which represents that the first idle time slot following the $IFS[A]$ can be elapsed (if the channel remains idle with a probability of $1-P_{sB}$).

$$\begin{cases} P\{(-1,-1)|(-1,0)\} = P_{sB}, \\ P\{(-1,1)|(-1,0)\} = 1-P_{sB}. \end{cases} \tag{3.7}$$

2. If the station enters the state $(-1,-1)$, it will remain in this state until the idle $IFS[A]$ following the end of the busy channel is completed. At the beginning of the immediately following time slot, the station will move into the next state. If the channel remains idle with a probability of $1-P_{sB}$, the station will move into the state $(-1,1)$.

$$P\{(-1,1)|(-1,-1)\} = 1-P_{sB}. \tag{3.8}$$

   If the channel becomes busy with a probability of P_{sB}, the station will remain in the state $(-1,-1)$ to wait through the transmission activity from AC A stations.

$$P\{(-1,-1)|(-1,-1)\} = P_{sB}. \tag{3.9}$$

3. When the station moves into the state $(-1,k)$, $1 \le k \le C-1$ and completes an idle time slot, it will move into the next state at the beginning of the immediately following time slot. If the channel becomes busy with a probability of P_{sB}, the station will move back to the state $(-1,-1)$ to wait through another transmission activity from AC A stations.

$$P\{(-1,-1)|(-1,k)\} = P_{sB}. \tag{3.10}$$

   If the channel remains idle with a probability of $1-P_{sB}$, the station will move into the next state $(-1,k+1)$.

$$P\{(-1,k+1)|(-1,k)\} = 1-P_{sB}. \tag{3.11}$$

4. When the AC B station moves into the state $(-1,C)$, it will wait through the final idle time slot remaining in the $IFS[B]$ and start a new backoff procedure with an initial backoff counter r at the beginning of the immediately following time slot. Similar to

the Markov chain model in Fig 3.2, the station will decrease its backoff counter by one following the end of the $IFS[B]$ and move into the next state at the beginning of the immediately following time slot. If the channel becomes busy with a probability of P_{bB}, the station will move into the state $(r-1,-1)$.

$$P\{(r-1,-1)|(-1,C)\} = Pr\_B(r)P_{sB}, \tag{3.12}$$

where $Pr\_B(r)$ is the probability that the AC B station gets an initial backoff counter value r. If the channel remains idle with a probability of $1-P_{sB}$, the station will move into the state $(r-1,0)$.

$$P\{(r-1,0)|(-1,C)\} = Pr\_B(r)(1-P_{sB}). \tag{3.13}$$

For the special case that r equals zero, the station will start a transmission immediately, independent of the channel status,

$$P\{(0,-1)|(-1,C)\} = Pr\_B(0). \tag{3.14}$$

5. If the station enters the state $(r,0)$, it will remain in this state for an idle time slot, decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. If the channel becomes busy with a probability of $1-P_{idleB}$, it will move into the state $(r-1,-1)$.

$$P\{(r-1,-1)|(r,0)\} = 1-P_{idleB}. \tag{3.15}$$

If the channel remains idle with a probability of P_{idleB}, it will move into the state (r-1,0).

$$P\{(r-1,0)|(r,0)\} = P_{idleB}. \tag{3.16}$$

For the special case that r equals zero, the station will remain in the state $(0,0)$ for an idle time slot, and start a transmission at the beginning of the immediately following time slot with a probability of 1.

$$P\{(-1,0)|(0,0)\} = 1. \tag{3.17}$$

6. If the station enters the state (r,-1), $0 \leq r \leq CW_{maxB-1}$, the one-step transition probabilities between the state $(r,-1)$ and the states (r,k), $1 \leq k \leq C-1$ are similar

to those between the state $(-1,-1)$ and the states $(-1,k)$, $1 \leq k \leq C-1$:

$$\begin{cases} P\{(r,1)|(r,-1)\} = 1 - P_{sB}, \\ P\{(r,-1)|(r,-1)\} = P_{sB}, \\ P\{(r,k+1)|(r,k)\} = 1 - P_{sB}, \\ P\{(r,-1)|(r,k)\} = P_{sB}. \end{cases} \tag{3.18}$$

7. When the station reaches the state (r,C), it will remain in this state for the final idle time slot in the $IFS[B]$, decrease its backoff counter by one, and move into the next state at the beginning of the immediately following backoff slot. If the channel becomes busy at this moment with a probability of P_{bB}, the station will move into the state $(r-1,-1)$.

$$P\{(r-1,-1)|(r,C)\} = P_{bB}. \tag{3.19}$$

If the channel remains idle with a probability of $1-P_{bB}$, the station will move into the state $(r-1,0)$.

$$P\{(r-1,0)|(r,C)\} = 1 - P_{bB}. \tag{3.20}$$

For the special case that r equals zero, the station will wait through an idle time slot in the state $(0,C)$ and start a transmission at the beginning of the immediately following backoff slot with a probability of 1.

$$P\{(-1,0)|(0,C)\} = 1. \tag{3.21}$$

3.2.1.4 System Equations

Let $b_{A(r,k)}$ be the steady probability of state (r,k) in the Markov chain model in Fig 3.2. The following system equations for this Markov chain model can be obtained due to the

regularity of the Markov chain:

$$\begin{cases} b_{A(CW_{maxA}-1,0)} = b_{A(-1,0)} Pr\_A(CW_{maxA})(1-P_{bA}), \\ b_{A(CW_{maxA}-1,-1)} = b_{A(-1,0)} Pr\_A(CW_{maxA}) P_{bA}, \\ \\ b_{A(r,0)} = b_{A(-1,0)} Pr\_A(r+1)(1-P_{bA}) \\ \qquad +b_{A(r+1,0)} P_{idleA} + b_{A(r+1,-1)}(1-P_{bA}), \\ \text{for } 0 \le r \le CW_{max}-1, \\ \\ b_{A(r,-1)} = b_{A(-1,0)} Pr\_A(r+1) P_{bA} \\ \qquad +b_{A(r+1,0)}(1-P_{idleA}) + b_{A(r+1,-1)} P_{bA}, \\ \text{for } 1 \le r \le CW_{max}-2, \end{cases} \tag{3.22}$$

and

$$\sum b_{A(r,k)} = 1. \tag{3.23}$$

Since the state $(0,-1)$ represents the transmission procedure of the station, the corresponding steady probability $b_{A(-1,0)}$ should be equal to its transmission probability τ_A:

$$b_{A(-1,0)} = \tau_A, \tag{3.24}$$

where τ_A is the unknown probability to be solved.

Similarly, the system equations for the Markov chain model in Fig. 3.3 can be obtained.

$$\begin{cases} b_{B(-1,-1)} = \frac{b_{B(-1,0)}[1-(1-P_{sB})^C]}{P_{sB}+(1-P_{sB})^C}, \\ b_{B(-1,1)} = (1-P_{sB})(b_{B(-1,-1)} + b_{B(-1,0)}), \\ b_{B(-1,k)} = (1-P_{sB})b_{B(-1,k-1)}, \\ \text{for } 2 \leq k \leq C, \\ \\ b_{B(CW_{maxB-1},0)} = b_{B(-1,C)}Pr\_B(CW_{maxB})(1-P_{bB}), \\ b_{B(CW_{maxB-1},-1)} = \frac{b_{B(-1,C)}Pr\_B(CW_{maxB})P_{bB}}{P_{sB}+(1-P_{sB})^C}, \\ \\ b_{B(r,0)} = b_{B(-1,C)}Pr\_B(r+1)(1-P_{bB}) \\ +b_{B(r+1,C)}(1-P_{bB}) + b_{B(r+1,0)}P_{idleB}, \\ b_{B(r,-1)} = [b_{B(-1,C)}Pr\_B(r+1)P_{bB} + b_{B(r+1,C)}P_{bB} \\ +b_{B(r+1,0)}(1-P_{idleB})]/[P_{sB} + (1-P_{sB})^C], \\ \text{for } 0 \leq r \leq CW_{maxB} - 2, \\ \\ b_{B(r,1)} = (1-P_{sB})b_{B(r,-1)}, \\ b_{B(r,k)} = (1-P_{sB})b_{B(r,k-1)}, \\ \text{for } 0 \leq r \leq CW_{maxB} - 1 \text{ and } 2 \leq k \leq C. \end{cases} \tag{3.25}$$

$$\sum b_{B(r,k)} = 1. \tag{3.26}$$

and

$$b_{B(-1,0)} = \tau_B, \tag{3.27}$$

where τ_B is the unknown probability to be solved.

3.2.2 Derivation of Key System Parameters

In this section, we analyse the unknown parameters in the transition probability equations shown in the last section, including P_{idleA}, P_{idleB},P_{sB}, P_{bA}, P_{bB}, $Pr\_A(r)$, and $Pr\_B(r)$. This section is organised as follows. Firstly, a new Markov chain model is used for analysing the contention zone specific transmission probabilities ($P_{tr:zone(1)}$ and $P_{tr:zone(2)}$), which results from the effect of using different $AIFSs$. Secondly, using the new Markov chain

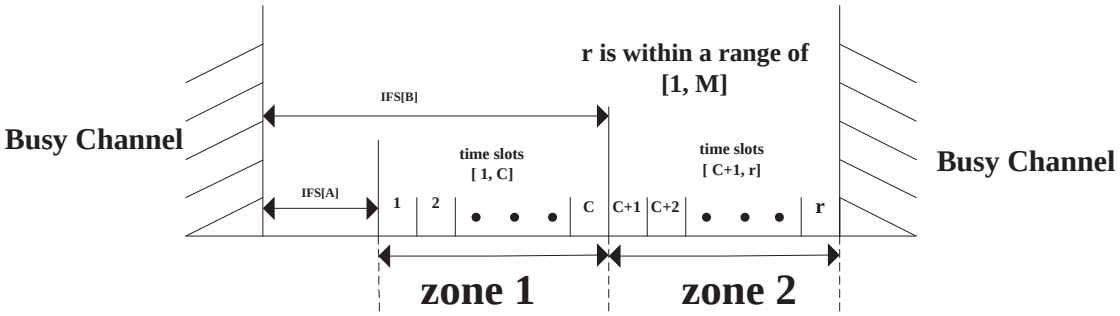


Figure 3.5: Time slot distribution between two successive transmissions in the system.

model, the AC specific average collision probabilities p_A and p_B are obtained. Also, the AC specific probabilities that the channel remains idle in a time slot during the normal backoff procedure, P_{idleA} and P_{idleB}, are obtained. Thirdly, the transition probability that the channel becomes busy in a time slot within the $IFS[B]$, P_{sB}, is obtained. Fourthly, the AC-specific probabilities that the channel becomes busy after IFS, P_{bA} and P_{bB}, are obtained. Finally, the AC specific transition probabilities $Pr\_A(r)$ and $Pr\_B(r)$ are analysed by using another new Markov chain model.

3.2.2.1 A Markov Chain Model for analysing the Effect of the Contention Zone-specific Transmission Probability

Fig 3.5 depicts the number of consecutive time slots between two successive transmissions in the WLAN system. In Fig 3.5, no station can transmit during the first $IFS[A]$ time interval from the end of the busy channel. During the time slots in the range of [1, C] after the $IFS[A]$, referred to as zone 1, AC A stations that have completed their $IFS[A]$ may begin their backoff procedure and transmit, while AC B stations are still waiting for the completion of their $IFS[B]$ and cannot transmit. During the time slots in the range of [$C+1$, r], referred to as zone 2, AC B stations also begin their backoff procedure and may transmit by contending with AC A stations. Here r is bounded by M, which is the maximum number of possible consecutive time slots between two successive transmissions in the WLAN:

$$M = min(CW_{maxA}, C + CW_{maxB}). \tag{3.28}$$

From Fig 3.5, a new discrete time one-dimensional Markov chain model can be created,

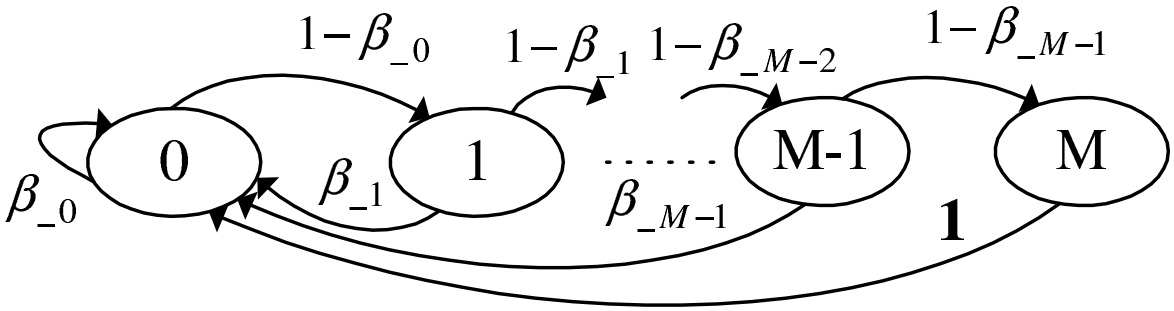


Figure 3.6: The Markov chain model for modeling the number of consecutive idle time slots between two successive transmissions in the WLAN.

which is shown in Fig 3.6. The stochastic process in this Markov chain model represents the number of consecutive idle time slots between two successive transmissions in the WLAN. The state (r) in the Markov chain model represents the r^{th} consecutive idle time slot starting from the end of the last transmission in the WLAN, which includes the idle $IFS[A]$ following the end of the busy channel. The transition events following the states (r), $0 \leq r \leq C-1$ represent the possible channel activity in zone 1, and the transition events following the states (r), $C \leq r \leq M$ represent the possible channel activity in zone 2.

The activity of this Markov chain is described by its one-step transition probabilities in the following.

1. In zone 1, if the channel status becomes busy following the end of the r^{th} idle time slot, the system will move from state (r) to state (0):

$$P\{(0)|(r)\} = P_{tr:zone(1)}, \text{for } 0 \leq r \leq C-1, \tag{3.29}$$

   where $P_{tr:zone(1)}$ is the probability that at least one priority A station starts the frame transmission at the beginning of a time slot in zone 1, given by

$$P_{tr:zone(1)} = 1 - (1-\tau_A)^{n_A}. \tag{3.30}$$

2. If no transmission occurs, the system will move from state (r) to state $(r+1)$ with a probability of $1 - P_{tr:zone(1)}$:

$$P\{(r+1)|(r)\} = 1 - P_{tr:zone(1)}, \text{for } 1 \leq r \leq C-1. \tag{3.31}$$

3. In zone 2, both AC A stations and AC B stations begin their backoff procedure and may transmit. A transmission from either AC A or AC B stations can cause the system to return to the state (0):

$$P\{(0)|(r)\} = P_{tr:zone(2)}, \text{for } C \leq r \leq M-1, \tag{3.32}$$

where $P_{tr:zone(2)}$ is the probability that at least one station starts the frame transmission in a time slot in zone 2, given by

$$P_{tr:zone(2)} = 1 - (1-\tau_A)^{n_A}(1-\tau_B)^{n_B}. \tag{3.33}$$

4. If no transmission occurs, the system will move from state (r) to state (r+1) with a probability of $1 - P_{tr:zone(2)}$:

$$P\{(r+1)|(r)\} = 1 - P_{tr:zone(2)}, \text{for } C \leq r \leq\ M-1. \tag{3.34}$$

5. When the system reaches the last state (M), a frame transmission will definitely occur after the corresponding time slot. Thus the system will return to the state (0) with a probability of 1:

$$P\{(0)|(M)\} = 1. \tag{3.35}$$

Using the above transition probability equations and the regularity of the Markov chain, the relations between the steady probability $s_{(r)}$ for the Markov chain model can be obtained by

$$\begin{cases} s_{(r+1)} = (1 - P_{tr:zone(1)})s_{(r)}, \\ \text{for } 0 \leq r \leq C-1, \\ \\ s_{(r+1)} = (1 - P_{tr:zone(2)})s_{(r)}, \\ \text{for } C \leq r \leq M-1, \end{cases} \tag{3.36}$$

and

$$\sum_{r=0}^{M} s_{(r)} = 1. \tag{3.37}$$

Finally, the steady probability $s_{(r)}$ can be solved as

$$\begin{cases} s_{(0)} = [\frac{1-(1-P_{tr:zone(1)})^{C+1}}{P_{tr:zone(1)}} + (1-P_{tr:zone(1)})^{C+1}(1-P_{tr:zone(2)})\frac{1-(1-P_{tr:zone(2)})^{M-C}}{P_{tr:zone(2)}}]^{-1}, \\ s_{(r)} = (1-P_{tr:zone(1)})^{r} s_{(0)}, \text{ for } 1 \leq r \leq C, \\ s_{(r)} = (1-P_{tr:zone(2)})^{r-C} s_{(0)} (1-P_{tr:zone(1)})^{C}, \text{ for } C+1 \leq r \leq M. \end{cases} \tag{3.38}$$

3.2.2.2 p_A, p_B, P_{idleA}, and P_{idleB},

For a specific station transmitting its frame, collision may occur if one or more other stations start a transmission in the same time slot. The corresponding collision probability is determined by the composition of contending stations. In zone 1, only AC A stations can transmit and cause collisions. In zone 2, both AC A stations and AC B stations can transmit and collide with each other. Thus the collision probability for an AC A station should be contention zone specific, which can be obtained by

$$\begin{cases} p_{A:zone(1)} = 1-(1-\tau_A)^{n_A-1}, \\ p_{A:zone(2)} = 1-(1-\tau_A)^{n_A-1}(1-\tau_B)^{n_B}, \end{cases} \tag{3.39}$$

For an AC A station in the backoff counter count-down procedure, it sees an "idle" time slot when no other stations start a transmission in the same time slot. Considering the contention zone specific transmission probability, the contention zone specific probability that an AC A station sees an idle time slot can be obtained by

$$\begin{cases} P_{idleA:zone(1)} = (1-\tau_A)^{n_A-1}, \\ P_{idleA:zone(2)} = (1-\tau_A)^{n_A-1}(1-\tau_B)^{n_B}. \end{cases} \tag{3.40}$$

Thus, the average collision probability for a specific AC A station can be obtained as the sum of the weighted contention zone specific collision probability:

$$p_A = \sum_{r=1}^{M} s_{(r)} p_{A:zone_r}, \tag{3.41}$$

where $p_{A:zone_r}$ is the contention zone specific collision probability in the r^{th} time slot. Depending on whether the r^{th} time slot belongs to zone 1 or zone 2, $P_{idleA:zone(1)}$ or $P_{idleA:zone(2)}$

should be used for $p_{A:zone_r}$. $s_{(r)}$ is the steady probability of the state (r), which is obtained from equation (3.38).

Similarly, the average probability P_{idleA} that a specific AC A station in the backoff procedure sees an idle time slot can be obtained by

$$P_{idleA} = \sum_{r=1}^{M} s_{(r)} P_{idleA:zone_r}, \tag{3.42}$$

where $P_{idleA:zone_r}$ is the contention zone specific probability for an AC A station that the channel is idle in the r^{th} time slot. Depending on whether the r^{th} slot belongs to zone 1 or zone 2, $P_{idleA:zone(1)}$ or $P_{idleA:zone(2)}$ should be used for $P_{idleA:zone_r}$.

For a specific AC B station, all of its time slots are located in zone 2, where all stations may transmit. Thus its average collision probability can be simply obtained by

$$p_B = 1 - (1 - \tau_A)^{n_A}(1 - \tau_B)^{n_B - 1}, \tag{3.43}$$

and the average probability that a specific AC B station has an idle time slot can be expressed as:

$$P_{idleB} = (1 - \tau_A)^{n_A}(1 - \tau_B)^{n_B - 1}. \tag{3.44}$$

3.2.2.3 P_{sB}

As described earlier in Section 3.1 in this chapter, a station suspending its backoff procedure may leave the backoff suspension procedure if the channel remains idle for an AC specific IFS interval from the end of the last busy channel. Any transmission from other stations during this time interval can stop the station from leaving the backoff suspension procedure.

An AC A station needs to wait through an idle $IFS[A]$ from the end of the last busy channel to leave the backoff suspension procedure. No transmission is possible during the $IFS[A]$ interval. Thus an AC A station can remain in the backoff suspension procedure for the duration of a single frame transmission only, and it will leave for the next state at the beginning of the immediately following time slot.

An AC B station needs to wait through an idle $IFS[B]$ from the end of the last busy channel to leave the backoff suspension procedure. According to Fig 3.5, the C time slots in zone

1 are part of the $IFS[B]$, where transmission from AC A stations is possible. Thus, the probability PsB that the channel turns busy in a time slot in zone 1 for a specific AC B station can be obtained by

$$P_{sB} = 1 - (1 - \tau_A)^{n_A}. \tag{3.45}$$

3.2.2.4 P_{bA} and P_{bB}

According to Fig 3.5, the time slot immediately following the $IFS[A]$ is located in zone 1, where only priority A station may transmit. Thus, the probability that the channel becomes busy at the beginning of this time slot for a specific AC A station can be obtained by

$$P_{bA} = 1 - (1 - \tau_A)^{(n_A - 1)}. \tag{3.46}$$

Also according to Fig 3.5, the time slot immediately following the $IFS[B]$ is located in zone 2, where all other stations may transmit. Thus, the probability that the channel turns busy at the beginning of this time slot for a specific AC B station can be obtained by

$$P_{bB} = 1 - (1 - \tau_A)^{n_A}(1 - \tau_B)^{(n_B - 1)}. \tag{3.47}$$

3.2.2.5 $Pr\_A(r)$ and $Pr\_B(r)$

As described previously (Chapter 1, Section 1.1), the backoff counter is drawn randomly from the range [0, CW] and the parameter CW is determined by the AC specific CW_{min} and CW_{max} values, as well as the number of previous consecutive retransmissions. Therefore the probability of obtaining a specific backoff counter value r is related to the number of previous consecutive retransmissions. The Markov chain models in Fig 3.2 and Fig 3.3 do not explicitly consider the effect of consecutive retransmissions. Instead, their effect is considered in the probability $Pr\_A(r)$ or $Pr\_B(r)$ of obtaining a specific backoff counter r by weighting the probability of the number of consecutive retransmissions. To simplify the presetation, we use the generic terms $Pr(r)$, p, CW_{min}, and CW_{max} in this section instead of the AC specific terms.

In order to obtain the probability that an AC specific station performs a specific number of consecutive retransmissions, a discrete time one-dimensional Markov chain model is created,

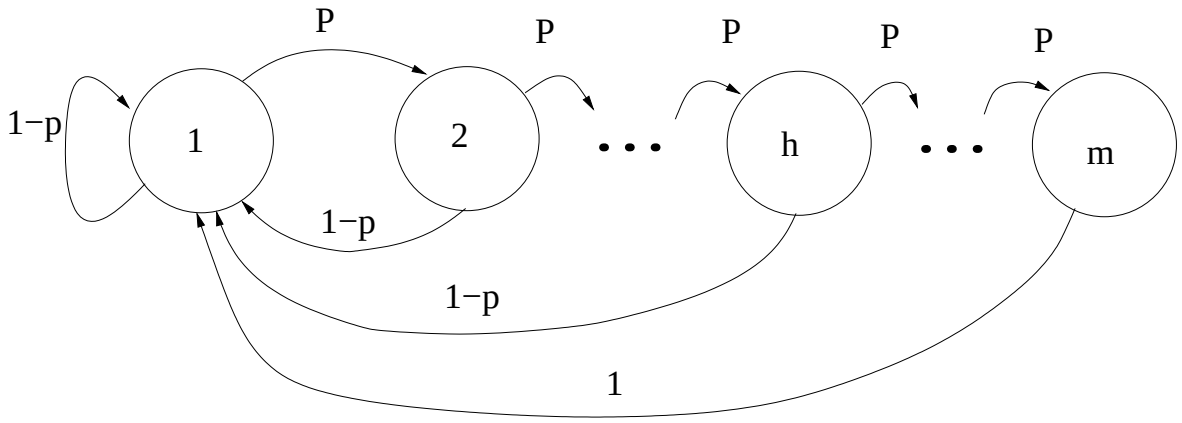


Figure 3.7: The Markov chain model for modeling the number of the consecutive retransmissions of a station.

as shown in Fig 4.5. The stochastic process in this Markov chain model represents the number of consecutive retransmissions (including the first transmission of the frame) for a station at time t. Thus state (k) represents that the station is performing the k^{th} consecutive retransmission. In this Markov chain, state (h) represents the h^{th} consecutive retransmission in which the CW value reaches CW_{max} for the first time, and state (m) represents the m^{th} consecutive retransmission, which is the maximum retransmission limit. Both h and m are constants determined by the IEEE 802.11 standard.

The activity of the Markov chain in Fig 4.5 is governed by its one-step transition probabilities as follows:

1. If the k^{th} retransmission is unsuccessful, the system will move from state (k) to state (k+1) with a probability of p:

$$P\{(k+1)|(k)\} = p, \text{for } 1 \leq k \leq m-1, \tag{3.48}$$

   where p is the AC specific average collision probability, which can be obtained from (3.41) or (3.43).

2. If the k^{th} consecutive retransmission is successful, the system will move from state (k) to state (1) with a probability of $1-p$ and the station will start transmitting a new frame:

$$P\{(1)|(k)\} = 1-p, \text{for } 1 \leq k \leq m. \tag{3.49}$$

3. when the maximum retransmission limit m is reached, the station will begin the first transmission of a new frame regardless of whether the m^{th} consecutive retransmission is successful or not. Thus the system will return to state (1) with a probability of 1:

$$P\{(1)|(m)\} = 1. \tag{3.50}$$

From (3.48), the relationship between two adjacent states can be obtained by

$$d_{(k+1)} = d_{(k)}p, \tag{3.51}$$

where $d_{(k)}$ is the corresponding steady probability for state (k).

Also, due to the regularity of the Markov chain, the following relationship can be obtained:

$$\sum_{k=1}^{m} d_{(k)} = 1. \tag{3.52}$$

Thus, the steady probability $d_{(k)}$ can be obtained:

$$d_{(k)} = p^{k-1}(1-p)/(1-p^m), \text{ for } 1 \leq k \leq m. \tag{3.53}$$

Since the backoff counter is a random integer uniformly distributed in the range $[0, CW]$, the probability of obtaining a specific backoff counter value from this range should be $\frac{1}{1+CW}$. Thus, the AC specific probability $Pr(r)$ of obtaining a specific backoff counter r can be obtained as the sum of the probability of obtaining a specific initial backoff counter r in the k^{th} consecutive retransmission, weighted with the probability of the occurrence of the k^{th} consecutive retransmission:

$$Pr(r) = \sum_{k=1}^{m} \frac{d_{(k)}c_{(r)}}{CW(k)+1}, \tag{3.54}$$

where $d_{(k)}$ is the steady probability of performing the k^{th} consecutive retransmission, which is obtained from (3.53); $CW(k)$ is the corresponding CW size in the k^{th} consecutive retransmission; and $c_{(r)}$ indicates whether the specific value r is included in the range [0, $CW(k)$] or not (if yes, $c_{(r)}$ is 1, otherwise it is zero).

Based on the earlier analysis, an expression for the AC specific probability $Pr(r)$ can be obtained:

$$\Pr(r) = \begin{cases} \sum_{k=1}^{h-1} \frac{d_{(k)}}{2^{k-1}CW_{min}+1} + \sum_{k=h}^{m} \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 0 \leq r \leq CW_{min}, \\ \\ \sum_{k=j}^{h-1} \frac{d_{(k)}}{2^{k}CW_{min}+1} + \sum_{k=h}^{m} \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 2^{j-1}CW_{min}+1 \leq r \leq 2^{j}CW_{min} \\ \quad \text{and } 1 \leq j \leq h-1, \\ \\ \sum_{k=h}^{m} \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 2^{h-1}CW_{min}+1 \leq r \leq CW_{max}, \end{cases} \tag{3.55}$$

where CW_{min} and CW_{max} are AC specific and known.

3.2.2.6 Summary of Analysis

Finally, this section presents a summary of the relationships between earlier analysis.

1. In Section 3.2.1, two novel Markov chain models have been illustrated for each AC in the WLAN, that are shown in Fig 3.2 and Fig 3.3 respectively. The system equations for each Markov chain model are also obtained, as shown in equations (3.22)-(3.27). Those equations show the state probability $b_{(r,k)}$ in each Markov chain model can be expressed in the form of the AC specific transition probabilities, including P_{idleA}, P_{idleB}, P_{sB}, P_{bA}, P_{bB}, $Pr\_A(r)$, and $Pr\_B(r)$.

2. The above AC specific transition probabilities for the Markov chain models in Fig 3.2 and Fig 3.3 have been analysed in Section 3.2.2 and they can be expressed in terms of τ_A and τ_B.

3. By using the system equations in Section 3.2.1 and the transition probabilities expressed in terms of τ_A and τ_B in Section 3.2.2, the steady state probability $b_{(r,k)}$ for both Markov chain models shown in Fig 3.2 and Fig 3.3 can be obtained in terms of τ_A and τ_B.

4. Finally, two non-linear equations about τ_A and τ_B based on equations (3.23) and (3.26) have been constructed for the AC specific Markov chain models presented in Fig 3.2 and Fig 3.3. The values of τ_A and τ_B can be numerically obtained from the equations.

The numerical calculation tool to solve non-linear equation systems is $fsolve$ function from the optimisation toolbox in MATLAB [83]. By using this function, some medium-scale optimisation algorithms (such as those in [84–87]) or some large-scale optimisation algorithms (such as those in [88, 89]) can be applied to solve non-linear equation systems. Since these numerical techniques are not the focus of this work, we will not go further to investigate them in details.

3.3 Saturated Throughput Analysis

In this section, we shall analyse the saturated throughput of EDCA. We consider that the throughput is equal to the ratio of the effective payload to the time required for successfully transmitting the effective payload. Here the effective payload is referred to as the size of the data field within a data frame, excluding the physical layer and MAC layer headers. The Markov chain model presented in Fig 3.6 is used to obtain the throughput, and its state probabilities can be obtained after τ_A and τ_B are solved. This Markov chain model represents the time slot distribution between two successive transmissions in the WLAN. Two possible events may occur in a time slot:

1. At least one transmission occurs in the time slot. Depending on whether the time slot is in zone 1 or zone 2 a transmission may occur with a zone specific probability of $P_{tr:zone(1)}$ or $P_{tr:zone(2)}$. $P_{tr:zone(1)}$ and $P_{tr:zone(2)}$ have been defined in (3.30) and (3.33) respectively. Furthermore, depending on whether the transmission is successful or not, two possibilities may occur:

   (a) A successful transmission. That is, only one transmission from either an AC A station or an AC B station occurs in the time slot. The corresponding contention zone probability for a successful transmission can be obtained by

$$\begin{cases} P_{sucA:zone(1)} = n_A\tau_A(1-\tau_A)^{n_A-1}, \\ P_{sucA:zone(2)} = n_A\tau_A(1-\tau_A)^{n_A-1}(1-\tau_B)^{n_B}, \\ P_{sucB:zone(1)} = 0, \\ P_{sucB:zone(2)} = n_B\tau_B(1-\tau_B)^{n_B-1}(1-\tau_A)^{n_A}. \end{cases} \tag{3.56}$$

(b) A collision. That is, two or more stations start transmitting in the same time slot. The corresponding contention zone specific collision probability can be obtained by

$$\begin{cases} P_{col:zone(1)} = P_{tr:zone(1)} - P_{sucA:zone(1)} \\ \qquad\qquad -P_{sucB:zone(1)}, \\ P_{col:zone(2)} = P_{tr:zone(2)} - P_{sucA:zone(2)} \\ \qquad\qquad -P_{sucB:zone(2)}. \end{cases} \tag{3.57}$$

2. No transmission occurs in the time slot. The corresponding contention zone specific probability for an idle time slot can be obtained by

$$\begin{cases} P_{idle:zone(1)} = 1 - P_{tr:zone(1)}, \\ P_{idle:zone(2)} = 1 - P_{tr:zone(2)}. \end{cases} \tag{3.58}$$

Therefore, the average effective payload for AC A stations can be obtained as:

$$E[A] = \sum_{r=1}^{M} P_{sucA:zone(r)} s_{(r)} E[P], \tag{3.59}$$

where $E[P]$ is the aforementioned effective payload (that is, the size of the data field within a data frame), and $s_{(r)}$ can be obtained from (3.38). $E[P]$ is considered as a known constant. The effective payload for AC A station measures the effective amount of AC A traffic that is transmitted between two successive transmissions.

Similarly, the average effective payload for AC B stations can be obtained by

$$E[B] = \sum_{r=1}^{M} P_{sucB:zone(r)} s_{(r)} E[P]. \tag{3.60}$$

The average time duration between two successive transmissions can be obtained as:

$$EL = \sum_{r=1}^{M} s_{(r)} [(P_{sucB:zone(r)} + P_{sucA:zone(r)}) Ts + P_{col:zone(r)} Tc + P_{idle:zone(r)} aTimeSlot], \tag{3.61}$$

where Ts and Tc are the time required for a successful transmission and a collision respectively. They are illustrated in Fig 3.8 and can be obtained by

$$Ts = H + P + SIFS + ACK + AIFS_{min}, \tag{3.62}$$

(a) a collision (b) a successful transmission

Figure 3.8: Transmission duration.

and

$$Tc = H + P + EIFS_{min}, \tag{3.63}$$

where H is the time required for transmitting the physical layer header and the MAC layer header of a frame, P is the time required for transmitting the data payload of a frame, ACK is the duration for transmitting an ACK frame, $AIFS_{min}$ is the minimum $AIFS$ used in the WLAN, and $EIFS_{min}$ equals to $SIFS + ACK + AIFS_{min}$. Here a basic access data rate determined by the WLAN physical layer is used for transmitting the physical layer header and ACK frame, while the payload data rate of sending the MAC layer header and payload can be higher [90, p. 11].

Finally, the throughput for each station of each AC can be obtained by

$$\begin{cases} Throughput_A = E[A]/EL/n_A, \\ Throughput_B = E[B]/EL/n_B. \end{cases} \tag{3.64}$$

3.4 Simulation Study

In this section, the theoretical analysis presented in the earlier sections is validated using simulation. Simulation is conducted using OPNET [91]. The impact of using different $AIFSs$ and different CW sizes on network performance is analysed. Finally, a comparison is performed between theoretical results obtained using the proposed model and those in [36–38, 41], which demonstrates that the proposed model has better accuracy.

The parameters used in the simulation are shown in Table 3.1. Four ACs are used in the simulation and their parameters are consistent with those defined in [16, Table 20df, p.49]. Two scenarios are simulated. In the first scenario, two ACs, i.e., voice and video, are used.

This scenario is designed to investigate the effect of using different CW sizes since a common $AIFS$ but different CW sizes are used by AC[voice] and AC[video] respectively. In the second scenario, two ACs, i.e., best effort and background, are used. The purpose of this scenario is to investigate the effect of using different $AIFS$s, since a common CW size but different $AIFS$s are used by AC[best effort] and AC[background] respectively. In both scenarios, there are equal number of stations in each AC.

Table 3.1: WLAN simulation parameter setting in EDCA performance analysis

PHY header	192 bits
MAC header	224 bits
Frame payload size	8000 bits
ACK frame size	PHY header+112 bits
Physical layer	IEEE 802.11b DSSS [14]
Basic access data rate	1Mbp/s
Payload data rate	1Mbp/s
Time slot	20 μs
$SIFS$	10 μs
Maximum retransmission limit	7
$AIFSN$	$AIFSN[voice] = 2$ $AIFSN[video] = 2$ $AIFSN[besteffort] = 3$ $AIFSN[background] = 7$
CW[voice]	$CW_{min} = 7, CW_{max} = 15$
CW[video]	$CW_{min} = 15, CW_{max} = 31$
CW[best effort]	$CW_{min} = 31, CW_{max} = 1023$
CW[background]	$CW_{min} = 31, CW_{max} = 1023$

Fig 3.9 shows the simulation results as well as theoretical results obtained from the proposed model for the first scenario. The throughput of a station in a specific AC under different number of stations is shown. It is shown in the figure that theoretical results obtained from the proposed model generally agree very well with simulation results. As shown in the figure, by using different CW_{min} and CW_{max}, traffic is successfully classified into two different

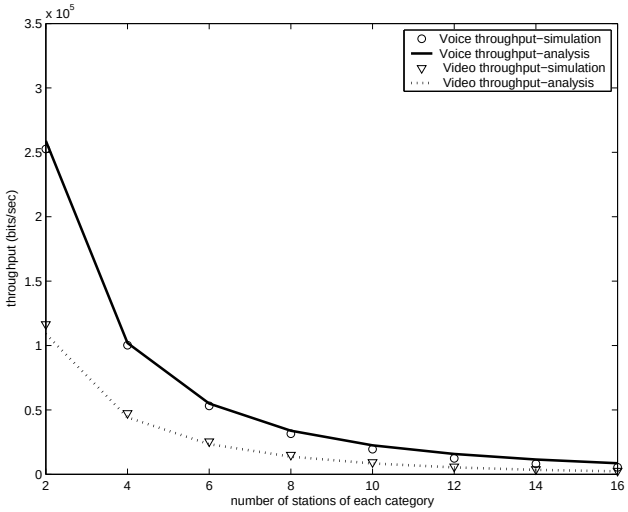


Figure 3.9: Simulation and analysis results for voice and video traffic.

classes. Traffic with a smaller CW_{min} and CW_{max} can have a better quality of service. When the number of stations in each AC is small, the difference in throughput for each AC is significant. When the number of stations in each AC increases, the difference in throughput decreases. Also, the throughput of both ACs decreases significantly due to more stations contending for bandwidth.

Fig 3.10 shows the simulation result as well as theoretical results obtained from the proposed model for the second scenario. As presented in the figure, by using different $AIFSs$, traffic is successfully classified into two different classes, and this difference is more significant than that in the first scenario. Traffic with a smaller $AIFS$ can have a better quality of service. It should be noticed that when the number of stations in each AC increases, the lower priority traffic belonging to AC[background] may be starved.

The effects of $AIFS$ and CW size on traffic prioritisation observed in the simulation results as well as theoretical results can be easily explained. Use of different $AIFSs$ introduces the contention zone specific transmission probability. Lower priority stations may be excluded for being allowed to transmit in some contention zone, which results in the possibility that some higher priority stations monopolize transmission opportunities and bandwidth. However, use of different CW sizes will only result in longer delay for lower priority stations and lower priority stations can still get the opportunity to transmit. Moreover, as shown in Fig 3.9, when the number of voice and video stations increases, the throughput of both ACs drops severely. The reason is that both AC[voice] and AC[video] have small $AIFS$ and

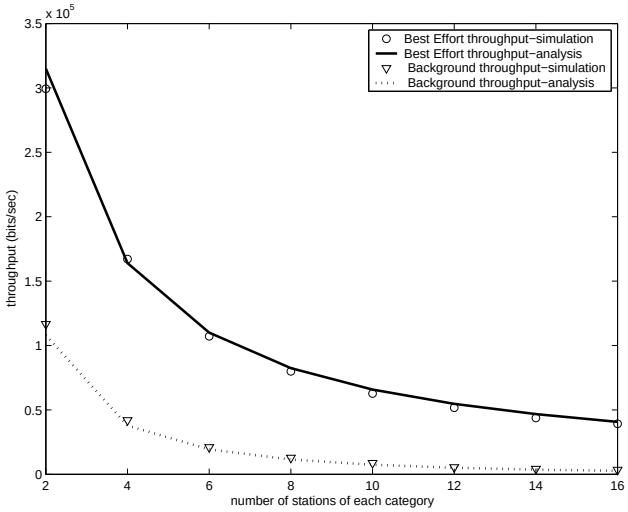


Figure 3.10: Simulation and analysis results for best effort and background traffic.

CW values. This enables stations to have a high transmission probability in a time slot, and accordingly their transmission will suffer a high collision probability when the number of stations is large. Therefore the majority of the available bandwidth is wasted on collision instead of successful transmission.

Finally, a larger discrepancy between theoretical and simulation results for a smaller number of stations is observed. It results from the assumption used in the model, that is, the transmission probability in a generic time slot is constant. As pointed out in [26], this assumption is more accurate when the number of stations is larger.

3.4.1 Comparison

The results obtained in this chapter have been compared with those in [36–39, 41]. For the sake of fair performance comparison, some existing analytical models are slightly modified with realistic system parameters. Firstly, equation (17) in [37] has been revised as

$$p1 = 1 - (1 - \tau_1)^{n_1 - 1} \left[P_{hold} + (1 - P_{hold})(1 - \tau_2)^{n_2 - 1} \right], \tag{3.65}$$

because an incorrect term P_{temp} instead of P_{hold} has been used in [37]. This typo error has been confirmed by personal communication with the authors. Secondly, equation (2) in [41] considers that the probability of allocating a random initial backoff counter within a range

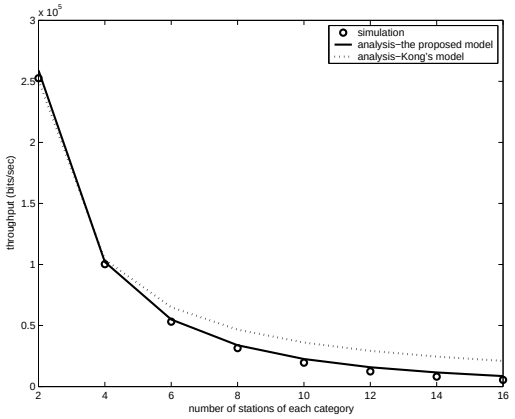
of $[0, CW]$ is 1, which is apparently incorrect and can lead to a solution not being obtained. We revise the probability to $\frac{1}{CW+1}$, and the revised equation (2) is given by

$$\begin{cases} P^{(l)}(0,0,k|i,j,0) = \frac{1-P_{b,i}^{(l)}}{CW_0^{(l)}+1} & k \in [0, CW_0^{(l)}] \\ P^{(l)}(0,j+1,k|i,j,0) = \frac{P_{b,i}^{(l)}}{CW_{j+1}^{(l)}+1} & k \in [0, CW_{j+1}^{(l)}] \\ P^{(l)}(0,m,k|i,m,0) = \frac{P_{b,i}^{(l)}}{CW_m^{(l)}+1} & k \in [0, CW_m^{(l)}] \end{cases} \tag{3.66}$$

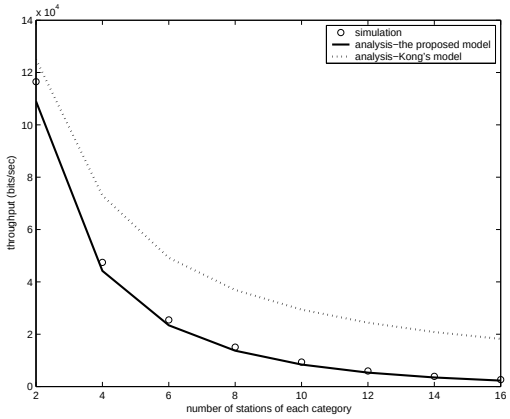
The results of the comparison are displayed in Fig 3.12-3.15. As shown in the results, the proposed model can achieve better accuracy than those in [36–39, 41]. These results are expected, as the proposed model captures the complexity of EDCA and removes some potential problems in [36–39, 41]. These potential problems have been explained in detail in the previous chapter.

3.5 Summary

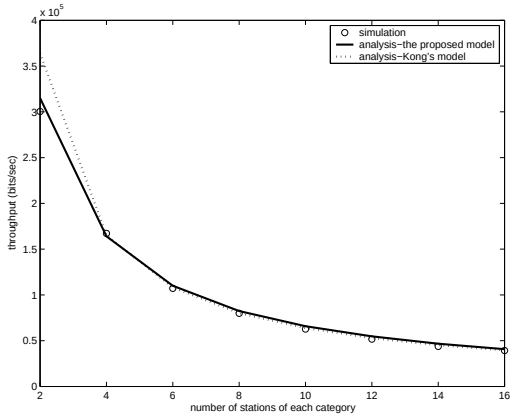
In this chapter, a novel Markov chain model for EDCA performance analysis under the saturated traffic load has been presented. Compared with the existing analytical models of EDCA, the proposed model incorporates more features of EDCA into the analysis and therefore more accurate. Both the effect of the contention zone specific transmission probability differentiation caused by using different $AIFSs$ and the effect of backoff suspension caused by transmission from other stations have been considered. Based on the proposed model, the saturated throughput of EDCA has been analysed. Simulation study using OPNET was performed, demonstrating that the theoretical results obtained from the proposed model closely match the simulation results, and the proposed model has better accuracy than those in the literature.



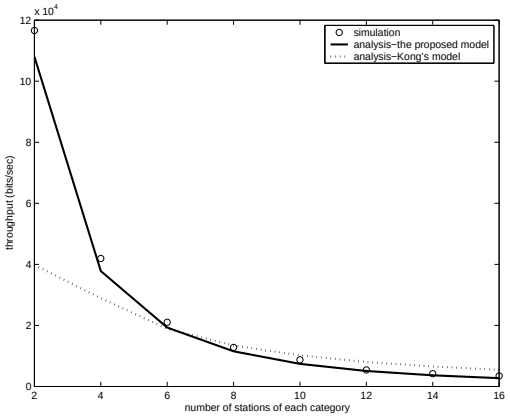
(a) voice.



(b) video.

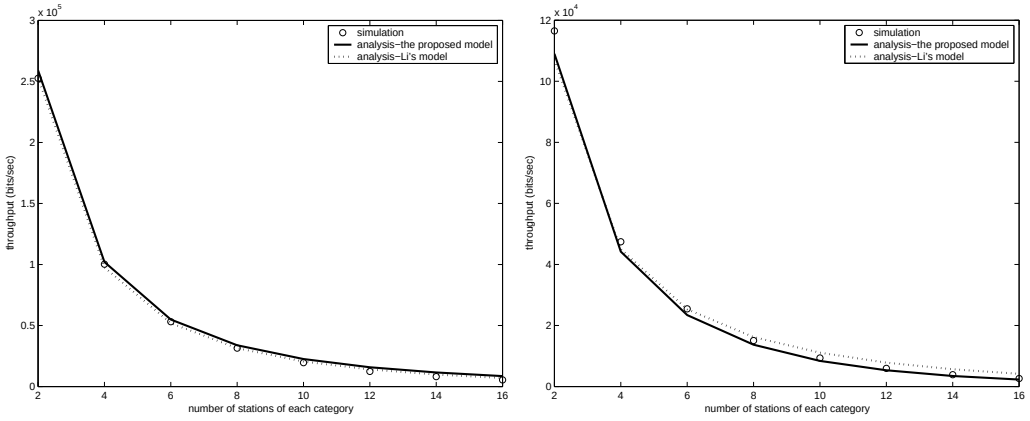


(c) best effort.

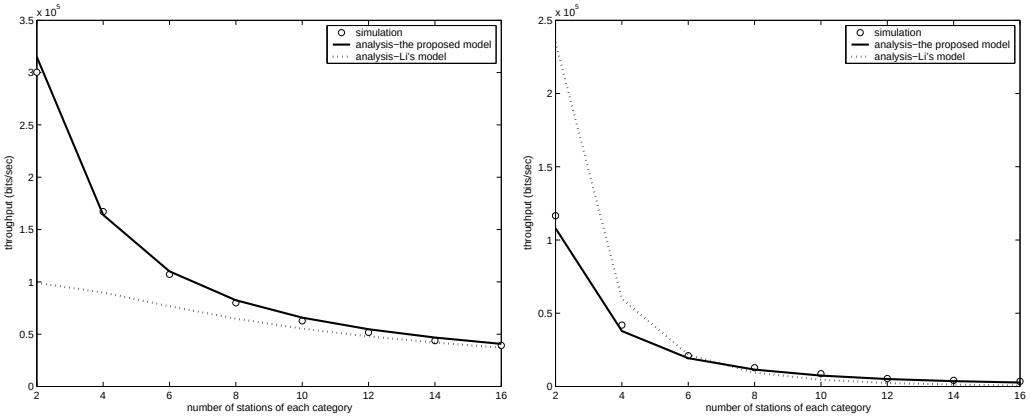


(d) background.

Figure 3.11: A comparison of the proposed model with the model in [36].

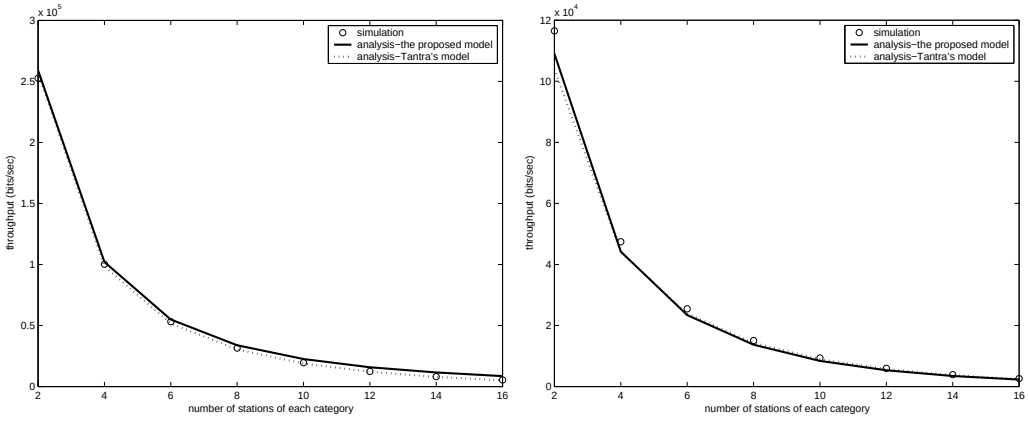


(a) voice (b) video

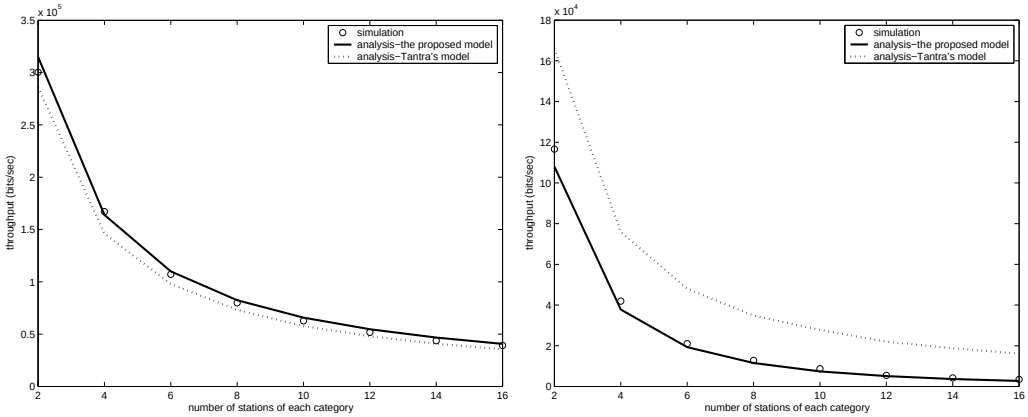


(c) best effort (d) background

Figure 3.12: A comparison of the proposed model with the model in [37].

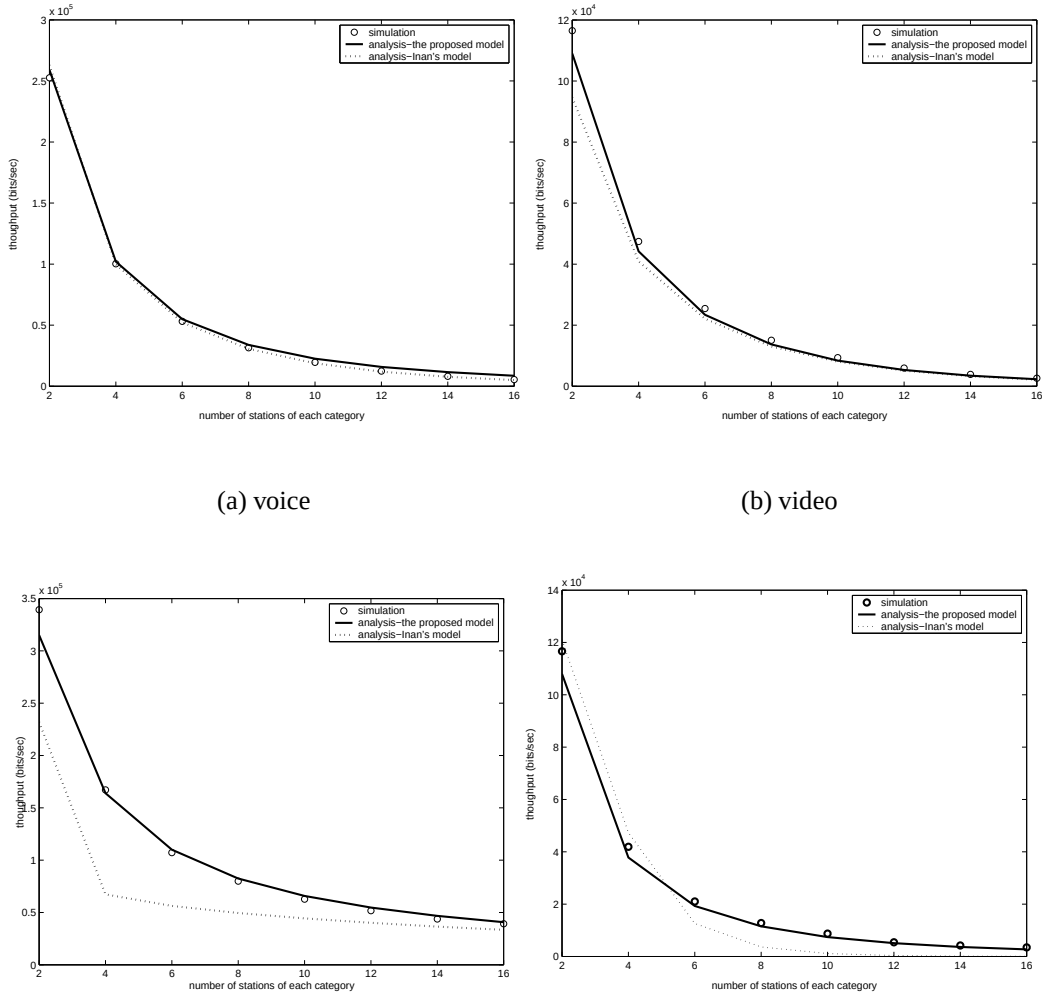


(a) voice (b) video



(c) best effort (d) background

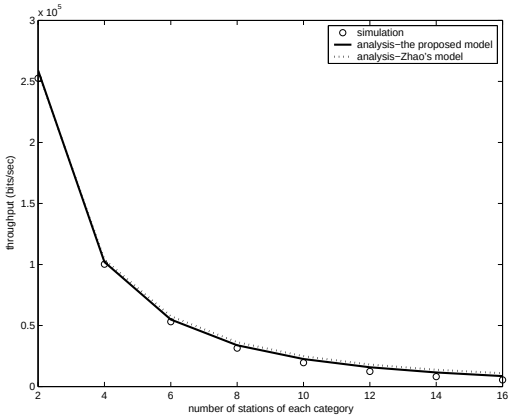
Figure 3.13: A comparison of the proposed model with the model in [38].



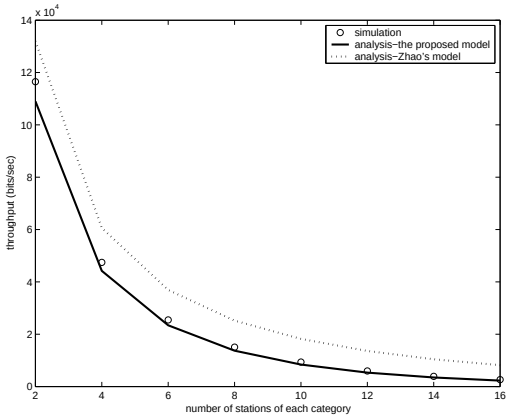
(a) voice (b) video

(c) best effort (d) background

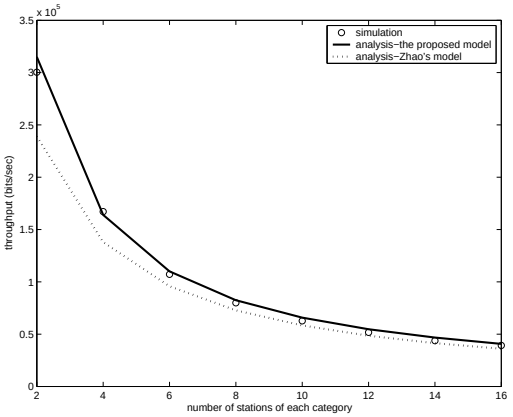
Figure 3.14: A comparison of the proposed model with the model in [39].



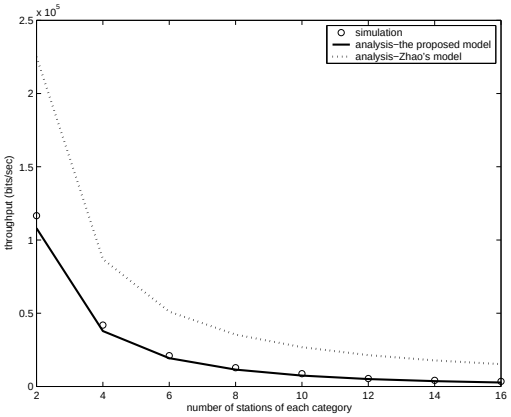
(a) voice



(b) video



(c) best effort



(d) background

Figure 3.15: A comparison of the proposed model with the model in [41].

Chapter 4

Saturated Throughput Analysis under the Coexistence of DCF and EDCA

Different from the previous chapter in which the analysis considered IEEE 802.11e capable stations only, the coexistence of traditional IEEE 802.11 stations (non-QSTAs) and QoS supported IEEE 802.11e stations (QSTAs) is considered in the analytical work in this chapter.

Following the publication of IEEE 802.11e standard in 2005, a proliferation of IEEE 802.11e capable products is expected. In the meantime, the traditional IEEE 802.11 products will exist for a considerably long period. Therefore, this work has a significance for practice.

DCF and EDCA are the fundamental access mechanisms for IEEE 802.11 and IEEE 802.11e respectively. Therefore, the focus of this work is the coexistence of DCF and EDCA. The literature review in Chapter 2 has demonstrated a lack of analytical work in this area, and the work discussed in this chapter fills this gap.

This chapter is structured as follow: in Section 4.1, the detailed difference between DCF and EDCA is discussed; in Section 4.2, the Markov chain based analytical model is proposed; in Section 4.3, throughput is analysed; in Section 4.4, simulation study is performed; finally, the chapter is summarised in Section 4.5.

4.1 The Detailed Differences between DCF and EDCA

DCF and EDCA are the fundamental access mechanisms for 802.11 and 802.11e respectively. The major difference between DCF and EDCA is that DCF uses the same backoff parameter set for all stations, while EDCA classifies traffic into four access categories (ACs), that is, voice, video, best effort, and background, each AC uses a distinct set of parameter sets, including CW, IFS, and $TXOP$ limit.

As mentioned previously (Chapter 1, Section 1.1.2.1), the term "IFS_E" is used to represent both $AIFS$ and $EIFS$ for QSTAs (stations using DCF) when it is not necessary to specify the difference between them. Similarly, the term "IFS_D" is used to represent both $DIFS$ and $EIFS$ for non-QSTAs (stations using EDCA) when there is no need to specify the difference between them.

In addition, some other detailed differences between DCF and EDCA exist [46]:

1. Each time a station starts a new backoff procedure or resumes a suspended backoff procedure, it must sense the channel as being idle for a complete IFS interval from the end of the last busy channel. A QSTA will decrease its backoff counter by one at the beginning of the time slot immediately following the IFS_E, irrespective of the channel status in that time slot. In comparison, a non-QSTA must sense the channel as being idle in the time slot immediately following the IFS_D too, in order to decrease its backoff counter by one at the beginning of the next following time slot. That is, a non-QSTA needs to wait through an extra idle time slot. A special case should be noted, when a non-QSTA or a QSTA starts a backoff procedure with an initial backoff counter of zero, both of them can start a transmission immediately after the respective IFS. This is the only case in which a non-QSTA does not need to wait through the extra time slot after the idle IFS_D in its channel contention procedure.

2. When a non-QSTA decreases its backoff counter to zero at the beginning of a time slot, it will start a transmission immediately, which is independent of the channel status in this time slot. In contrast, a QSTA will not transmit immediately when its backoff counter is decreased to zero at the beginning of a time slot. It can only start a transmission at the beginning of the next time slot, provided that the channel remains idle in the current time slot. If the channel becomes busy in the current time slot, the QSTA must

wait through a complete idle IFS_E after the busy channel and start the transmission after the IFS_E.

4.2 A Markov Chain Model for Coexistence Analysis

In this section, we introduce the Markov model for analysing the network performance when QSTAs and non-QSTAs coexist in the same base station set. The following assumptions are used in the analysis

- Traffic at each station is saturated;
- Each QSTA carries traffic of one AC only;
- The transmission probability of a specific QSTA or non-QSTA in a generic time slot is a constant, which is represented by "τ_E" or "τ_D"respectively. They are unknown variables to be solved.
- The number of non-QSTAs ("N_D"), and QSTAs ("N_E") are fixed and known;
- For simplicity, we consider the coexistence of non-QSTAs, and QSTAs carrying the traffic of one AC only.
- Only one fixed-size data frame is transmitted in each TXOP.
- The radio channel is ideal.

This section is divided into two parts: Section 4.2.1 analyses the performance of the system where non-QSTAs and QSTAs carrying best effort or background traffic coexist in the same base station set; Section 4.2.2 analyses the performance of the system where non-QSTAs and QSTAs carrying voice or video traffic coexist in the same base station set.

4.2.1 Coexistence of non-QSTAs and QSTAs with Best Effort or Background Traffic

This section is organised as follow. Section 4.2.1.1 illustrates the basic Markov chain models; Section 4.2.1.2 analyses the zone specific transmission probability, a concept that will be explained shortly later; Section 4.2.1.3- 4.2.1.6 analyse the transition probabilities in the basic Markov chain models; finally Section 4.2.1.7 summaries the analysis and obtains the final solution.

4.2.1.1 Discrete time two-dimensional Markov chain models

Fig 4.1 and Fig 4.2 illustrate the proposed two-dimensional discrete time Markov chain models. The model in Fig 4.1 is used to model the channel contention procedure of a non-QSTA; and the model in Fig 4.2 is for a QSTA carrying best effort or background traffic whose IFS_E is larger than IFS_D.

There are two stochastic processes in each Markov chain model. The first process, $u(t)$, represents the value of the backoff counter. Here a special value of $u(t) = -1$ is used to represent a transmission from the station, which includes the transmission and the idle IFS_D interval after the transmission. Here IFS_D is included in the "transmission" state because it is the smallest IFS in the system, and no transmission is possible in this interval. This definition of the transmission state is used throughout this chapter. The second process, $v(t)$, indicates the station's status. Here $v(t) = 0$ represents that the station is either in a normal backoff procedure or in a data transmission state. $v(t) = -1$ represents that the station's backoff procedure is being interrupted by transmission from other stations. $v(t) > 0$ represents that the station is in the $v(t)^{th}$ idle time slot after the IFS_D.

In Fig 4.1, the state $(k, 0)$, $1 \leq k \leq CW_{max\_D}$ represents an idle time slot in which the backoff counter of a non-QSTA is decreased to k. Here CW_{max_D} is the CW_{max} value for a non-QSTA. State $(k, -1)$, $1 \leq k \leq CW_{max\_D}$ represents a transmission from other stations which interrupts the non-QSTA's backoff procedure. The state $(k, 1)$, $1 \leq k \leq CW_{max\_D}$ represents an idle time slot immediately following a transmission activity (that is , a successful transmission from one station, or a collision caused by multiple transmissions from multiple stations) from other stations. The state $(-1, 0)$ represents a transmission from

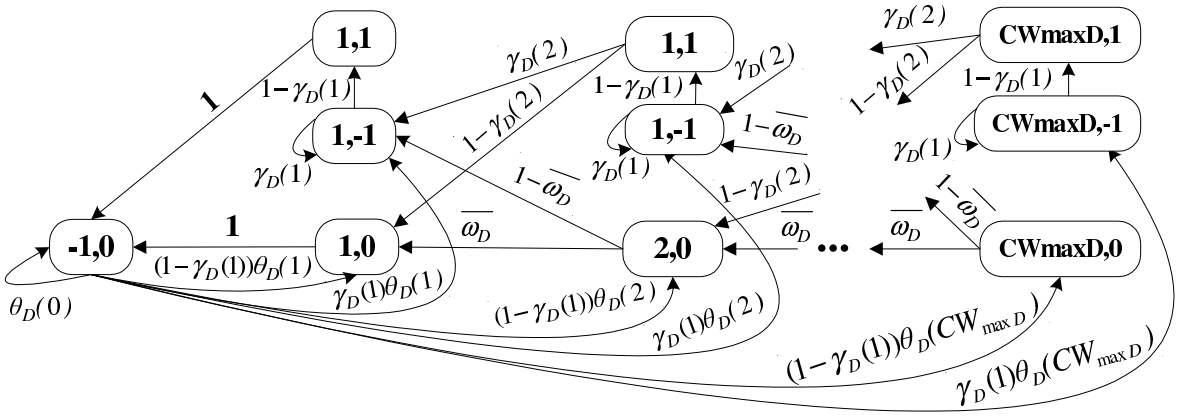


Figure 4.1: The Markov chain model for a non-QSTA.

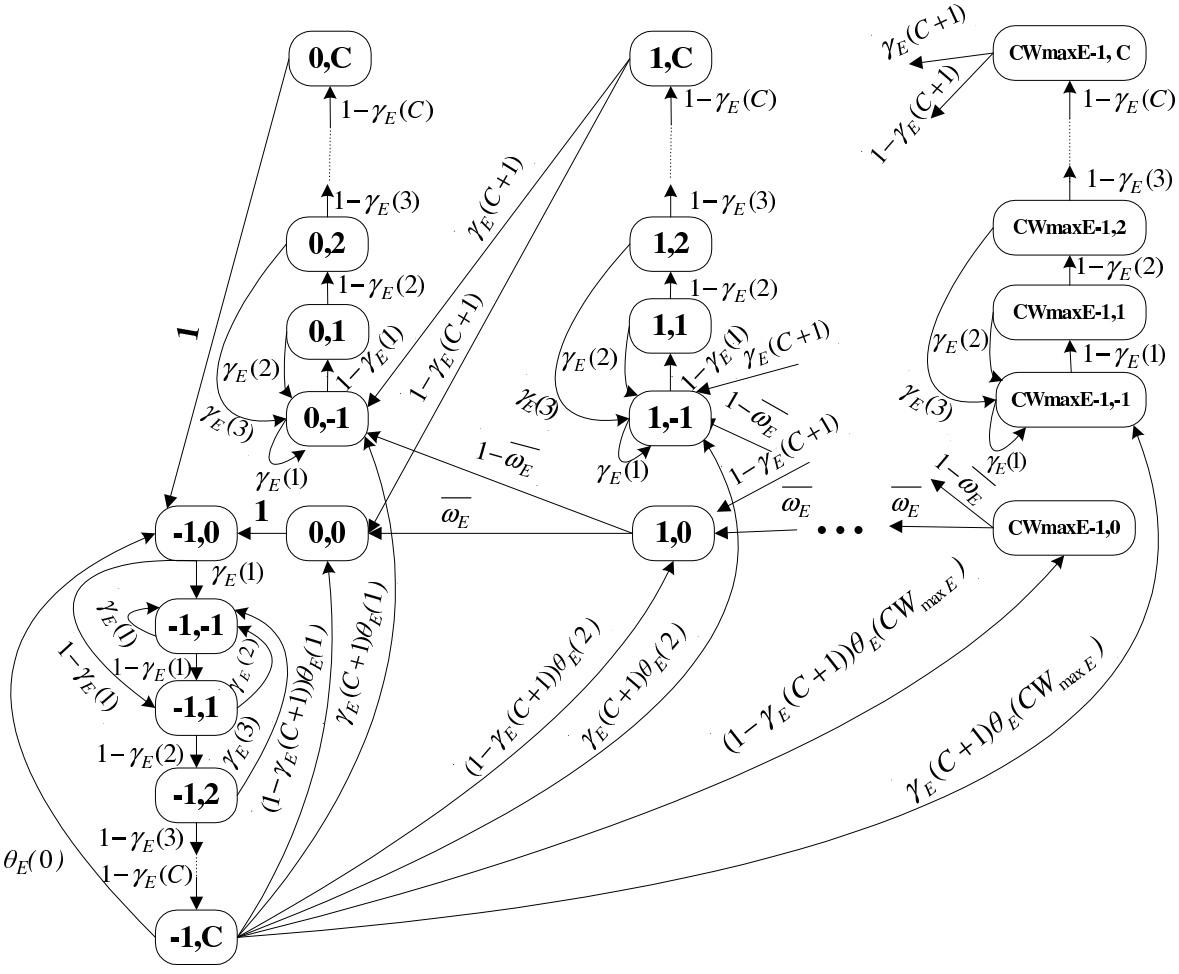


Figure 4.2: The Markov chain model for a QSTA carrying best effort or background traffic.

the non-QSTA itself. The transition probability $\theta_D(k)$, $0 \leq k \leq CW_{max\_D}$ represents the probability that a non-QSTA starts a new backoff procedure with an initial backoff counter k. The probabilities $\gamma_D(d)$ and $1-\gamma_D(d)$, $d = 1, 2$ represent the transition probabilities of the channel in the d^{th} time slot after a transmission event. Here $\gamma_D(d)$ is the probability that the channel turns busy in the d^{th} time slot after a transmission event. Accordingly, $1-\gamma_D(d)$ is the probability that the channel remains idle in the d^{th} time slot after a transmission event. The transition probability $\overline{\omega_D}$ is the probability that the channel remains idle in a time slot for a non-QSTA in its backoff procedure; and $1-\overline{\omega_D}$ is the probability that the channel becomes busy. State $(1, 0)$ or $(1, 1)$ represents the idle time slot immediately before the backoff counter of a non-QSTA is decreased to zero and the non-QSTA is ready to start a transmission. After leaving the state $(1, 0)$ or $(1, 1)$, the non-QSTA will enter into state $(-1, 0)$ to start a transmission with a probability of 1.

Slight differences exist in the Markov chain shown in Fig 4.2, which are caused by the differences between DCF and EDCA described in Section 4.1. The parameters $\theta_E(k)$, $\gamma_E(d)$, CW_{max_E}, and $\overline{\omega_E}$ have similar meanings as the corresponding terms $\theta_D(k)$, $\gamma_D(k)$, CW_{max_D}, and $\overline{\omega_D}$ in Fig 4.1. The symbol "C" represents the difference between IFS_E and IFS_D, given by $C = AIFS - DIFS$. For ease of illustration, we assume that $C > 2$. The differences between the model for DCF and that for EDCA are explained in the following.

First, following the end of IFS_E, a QSTA will decrease its backoff counter by one, while a non-QSTA must wait through an extra time slot after the IFS_D. Therefore, no special state is required in Fig 4.2 to represent the time slot after the IFS_E. Moreover, after completing the IFS_E following the QSTA's own transmission, the QSTA will reach the state $(k-1, 0)$ or $(k-1, -1)$ if it starts a new backoff procedure with a non-zero initial backoff counter k, while a non-QSTA will reach state $(k, 0)$ or $(k, -1)$ if it has a non-zero initial backoff counter k.

Second, state $(0, n)$, $-1 \leq n \leq C$ is used to represent the channel contention procedure of the QSTA when its backoff counter has been decreased to zero. These states do not exist in the Markov chain model shown in Fig 4.1 because a non-QSTA will start a transmission immediately after its backoff counter is decreased to zero.

Finally, because IFS_E is larger than IFS_D, the state (k, n), $-1 \leq k \leq CW_{max\_E}, 1 \leq n \leq C$ in the chain is used to represent the n^{th} idle time slot following a transmission, which is

still within the IFS_E interval. In the time interval corresponding to these states, a non-QSTA can start a transmission or decrease its backoff counter but a QSTA (carrying best effort or background traffic) cannot start a transmission or decrease its backoff counter.

As the state $(-1, 0)$ in each Markov chain model represents the station's own transmission, its state probability is equal to τ_D (in Fig 4.1) or τ_E (in Fig 4.2) respectively. Both τ_D and τ_E are unknown parameters which need to be solved.

It should be noted, within the two Markov chain models, several unknown parameters exist, including $\theta_D(k)$, $\theta_E(k)$ $\gamma_D(d)$, $\gamma_E(d)$, $\overline{\omega_D}$, and $\overline{\omega_E}$. In our later analysis, we will analyse these unknown parameters with unknown values τ_D and τ_E.

4.2.1.2 The zone specific transmission probability

Before we delve into detailed analysis of the earlier Markov chain models, here we first analyse the so-called zone specific transmission probability [27]. The result obtained in this section will be used to solve unknown parameters in the the earlier Markov chain models.

Fig 4.3 illustrates the time slots between two adjacent transmissions in the system. Here the maximum number of the possible consecutive idle time slots between two successive transmissions in the system is bounded by M, where $M = min(CW_{max\_D}, C + CW_{max\_E})$. As shown in Fig 4.3, no transmission is possible in the IFS_D interval immediately following the busy channel. In the first time slot after the IFS_D, referred to as zone 1, only non-QSTAs involved in the previous transmission with an initial backoff counter zero and start a transmission. In the time slots $[2, C]$, referred to as zone 2, all non-QSTAs may start a transmission. In the remaining time slots, referred to as zone 3, both non-QSTAs and QSTAs may transmit. The above statement shows that the transmission probabilities of non-QSTAs and QSTAs are different in each zone. The corresponding zone specific transmission probability can be obtained by

$$\begin{cases} \beta(1) = \sum_{i=0}^{N_D} \left\{ [1 - (1 - \tau_D)^i] \, \phi(i) \right\}, \\ \beta(2) = 1 - (1 - \tau_D)^{N_D}, \\ \beta(3) = 1 - (1 - \tau_D)^{N_D} (1 - \tau_E)^{N_E}, \end{cases} \tag{4.1}$$

where $\beta(i)$ represents the probability that there is a transmission in a time slot in zone i, and $\phi(i)$ represents the probability that i out of the N_D non-QSTAs become involved in the

previous transmission, given by

$$\phi(i) = \begin{pmatrix} N_D \\ i \end{pmatrix} \tau_D^i (1-\tau_D)^{N_D - i}. \tag{4.2}$$

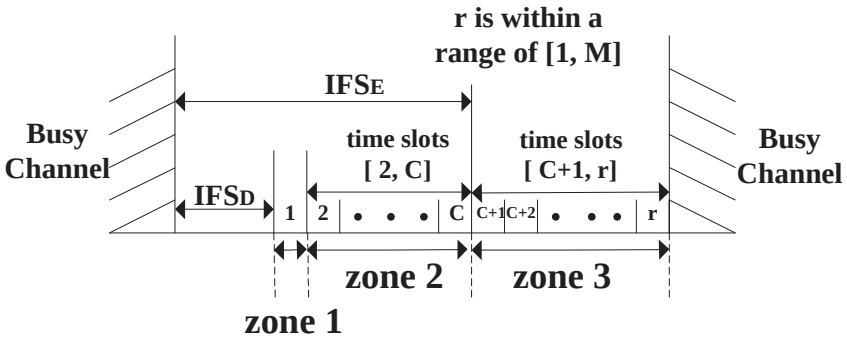


Figure 4.3: Time slots between two successive transmissions in the system.

From Fig 4.3, a new discrete time one-dimensional Markov chain model can be created, which is presented in Fig 4.4. The stochastic process in this Markov chain model represents the number of consecutive idle time slots between two successive transmissions in the system. The state $(r), 0 \leq r \leq M$ in the Markov chain model represents the r^{th} consecutive idle time slot from the end of the last transmission in the system. The transition probabilities $\beta(i)$ and $1-\beta(i), i = 1, 2, 3,$ represents the possible channel activity after the r^{th} idle time slot. Here $\beta(i)$ has the same meaning as that in the last paragraph. State (M) represents the M^{th} idle time slot, and a transmission will definitely occur after it. Therefore the system will move from state (M) to state (0) with a probability of 1.

The state probability $s(r)$ for this Markov chain model can be readily obtained, given by

$$\begin{cases} s(0) = \frac{1}{1+[1-\beta(1)]\frac{1-[1-\beta(2)]^C}{\beta(2)}+[1-\beta(1)][1-\beta(2)]^{C-1}[1-\beta(3)]\frac{1-[1-\beta(3)]^{M-C}}{\beta(3)}}, \\ s(r) = s(0)\left[1-\beta(1)\right]\left[1-\beta(2)\right]^{r-1}, \text{for } 1 \leq r \leq C, \\ s(r) = s(C)\left[1-\beta(3)\right]^{r-C}, \text{for } C+1 \leq r \leq M, \end{cases} \tag{4.3}$$

where we can observe that $s(r)$ can be expressed in terms of $\beta(k)$, $k = 1, 2, 3$. According to (4.1), $\beta(k)$, $k = 1, 2, 3$ can be expressed in terms of τ_D and τ_E, therefore we can say that $s(r)$ can also be expressed in terms of τ_D and τ_E.

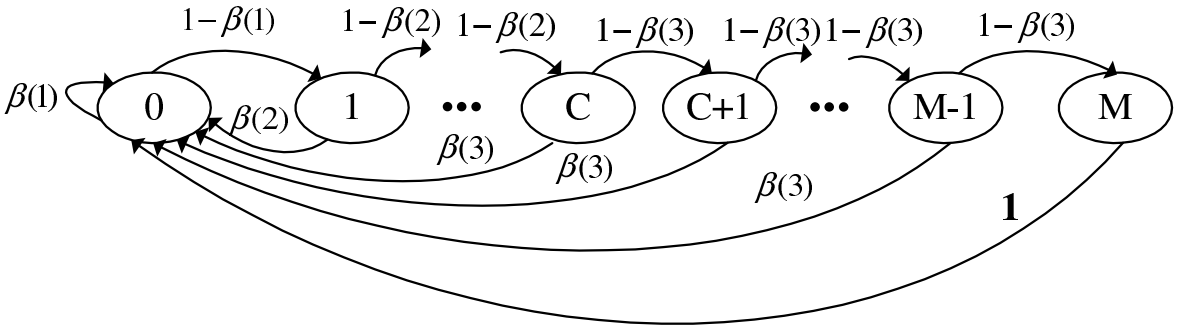


Figure 4.4: The Markov chain model for the number of consecutive idle time slots between two successive transmissions.

With the solution of $s(r)$, we may obtain the probabilities that the system reaches zone 1, zone 2, and zone 3 respectively, given by

$$\begin{cases} Z(1) = s(0), \\ Z(2) = \sum_{i=1}^{C-1} s(i), \\ Z(3) = \sum_{i=C}^{M} s(i). \end{cases} \tag{4.4}$$

4.2.1.3 Average collision probabilities $\overline{\rho_D}$ and $\overline{\rho_E}$

Here we begin to analyse the average collision probabilities $\overline{\rho_D}$ and $\overline{\rho_E}$ for non-QSTAs and QSTAs carrying best effort or background traffic respectively. These average collision probabilities are not used in the proposed Markov chain models directly and they are used to obtain transition probabilities in the Markov chains. According to Fig 4.3, in zone 1, for a transmission from a non-QSTA, only other non-QSTAs involved in the previous transmission may transmit and cause a collision. In zone 2, all other non-QSTAs may transmit and cause a collision. In zone 3, all other non-QSTAs and QSTAs may transmit and cause a collision. Thus the collision probability for a non-QSTA should be zone specific, which can be obtained as

$$\begin{cases} \rho_D(1) = \sum_{i=0}^{N_D-1} \left\{ [1-(1-\tau_D)^i]\, \xi(i) \right\}, \\ \rho_D(2) = 1-(1-\tau_D)^{N_D-1}, \\ \rho_D(3) = 1-(1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \end{cases} \tag{4.5}$$

where $\xi(i)$ represents the probability that i out of the remaining $N_D - 1$ non-QSTAs get involved in the previous transmission, given by

$$\xi(i) = \begin{pmatrix} N_D - 1 \\ i \end{pmatrix} \tau_D^i (1 - \tau_D)^{N_D - 1 - i}. \tag{4.6}$$

The corresponding average collision probabilities can be obtained as the sum of the weighted contention zone specific collision probabilities:

$$\overline{\rho_D} = \sum_{i=1}^{3} Z(i) \rho_D(i). \tag{4.7}$$

A QSTA can transmit in zone 3 only, where all other QSTAs and all non-QSTAs may transmit and cause a collision. Therefore the average collision probability for a QSTA can be obtained by

$$\overline{\rho_E} = 1 - (1 - \tau_D)^{N_D} (1 - \tau_E)^{N_E - 1}. \tag{4.8}$$

4.2.1.4 The average probabilities that a station decreasing its backoff counter detects an idle time slot $\overline{\omega_D}$ and $\overline{\omega_E}$

A non-QSTA in the backoff state detects an idle time slot when no other stations start a transmission in the same time slot. The zone specific probability that a non-QSTA detects an idle time slot is then given by

$$\begin{cases} \omega_D(1) = \sum_{i=0}^{N_D - 1} \left[(1 - \tau_D)^i \xi(i) \right], \\ \omega_D(2) = (1 - \tau_D)^{N_D - 1}, \\ \omega_D(3) = (1 - \tau_D)^{N_D - 1} (1 - \tau_E)^{N_E}, \end{cases} \tag{4.9}$$

where $\xi(i)$ has been given in (4.6).

We may obtain the corresponding average probability by

$$\overline{\omega_D} = \sum_{i=1}^{3} Z(i) \omega_D(i).$$

A QSTA can only decrease its backoff counter in zone 3, where all other QSTAs and all non-QSTAs may transmit. Therefore, we can simply obtain the corresponding average probability as

$$\overline{\omega_E} = (1 - \tau_D)^{N_D} (1 - \tau_E)^{N_E - 1}. \tag{4.10}$$

4.2.1.5 The probabilities that the channel remains idle in the d^{th} time slot after a transmission $\gamma_D(d)$ and $\gamma_E(d)$

In the Markov chain models depicted in Fig 4.1 and Fig 4.2 respectively, $\gamma_D(d)$, d=1, 2 and $\gamma_E(k)$, $1 \leq k \leq C+1$ are used. In this section, we obtain an analytical expression for the two parameters.

According to Fig 4.3, the first time slot after the IFS_D is located in zone 1. For a non-QSTA, the channel will remain idle in this time slot if other non-QSTAs which get involved in the previous transmission do not transmit, and the corresponding probability $\gamma_D(1)$ is given by

$$\gamma_D(1) = \sum_{i=0}^{N_D-1} \left[1-(1-\tau_D)^i\right] \xi(i), \tag{4.11}$$

where $\xi(i)$ has been given in (4.6).

In the second time slot after the IFS_D, which is located in zone 2 according to Fig 4.3, the channel will remain idle for a non-QSTA if all other non-QSTAs do not transmit, and the corresponding probability $\gamma_D(2)$ is given by

$$\gamma_D(2) = 1-(1-\tau_D)^{N_D-1}. \tag{4.12}$$

For a QSTA, the channel will remain idle in the first time slot after the IFS_D in zone 1 if non-QSTA which get involved in the previous transmission do not transmit, and the corresponding probability $\gamma_E(1)$ is given by

$$\gamma_E(1) = \sum_{i=0}^{N_D} \left\{\left[1-(1-\tau_D)^i\right] \phi(i)\right\}. \tag{4.13}$$

The channel will remain idle for a QSTA in time slots [2 C] in zone 2 if all non-QSTAs do not transmit, and the corresponding probabilities $\gamma_E(k)$, $2 \leq k \leq C$ are given by

$$\gamma_E(k) = 1-(1-\tau_D)^{N_D}, 2 \leq k \leq C. \tag{4.14}$$

In the $C+1^{th}$ time slot after the IFS_D, which is located in zone 3 according to Fig 4.3, the channel will remain idle for a QSTA in this time slot if all non-QSTAs and other QSTAs do not transmit, and the corresponding probability $\gamma_E(C+1)$ is given by

$$\gamma_E(C+1) = 1-(1-\tau_D)^{N_D}(1-\tau_E)^{N_E-1}. \tag{4.15}$$

4.2.1.6 The probabilities that a station obtains an initial backoff counter of k, $\theta_D(k)$ and $\theta_E(k)$

Both non-QSTAs and QSTAs use the same procedure to choose a random initial backoff counter and the only difference is in the respective CW_{min} and CW_{max} values. Therefore we only present the analysis on $\theta_D(k)$ in this chapter. The analytical expression for $\theta_E(k)$ can be obtained analogously.

A new Markov chain is created to model the number of transmission attempts by a non-QSTA for sending the same data frame, which is shown in Fig 4.5.

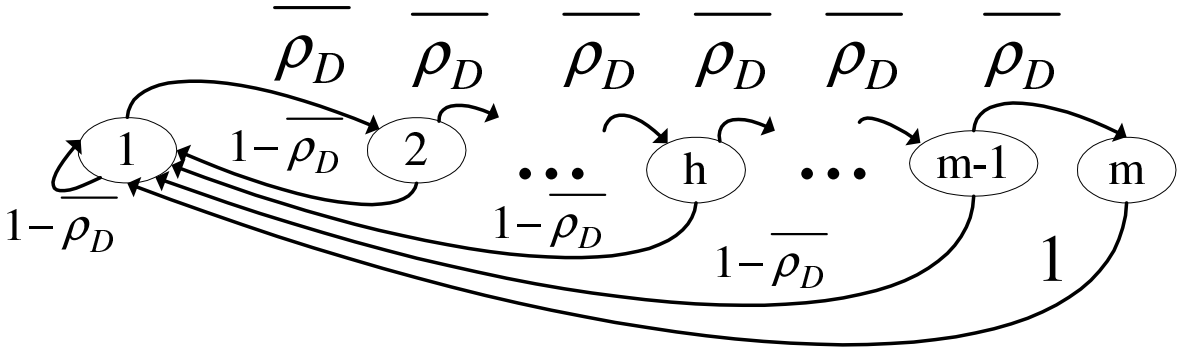


Figure 4.5: The Markov chain for modelling the number of transmission attempts of a non-QSTA for sending a data frame.

In this Markov chain model, h is the number of transmission attempts at which CW_{max_D} is first reached, m is the maximum number of transmission attempts for sending a data frame, and $\overline{\rho_D}$ is the average collision probability for a non-QSTA, which was obtained in (4.7). Each state (I) represents the I^{th} transmission attempt of a non-QSTA. We can readily obtain the state probability $o(I)$ as:

$$\begin{cases} o(1) = \frac{1-\overline{\rho_D}}{1-\overline{\rho_D}^m}, \\ o(I) = o(1)\overline{\rho_D}^{I-1}, \text{for } 2 \leq I \leq m. \end{cases} \tag{4.16}$$

Therefore, the probability $\theta_D(k)$ that a non-QSTA station obtains an initial backoff counter value k is given by

$$\theta_D(k) = \sum_{I=1}^{m} \frac{o(I)c(k)}{CW(I)+1}, \tag{4.17}$$

where $CW(I)$ is the CW size for the I^{th} transmission attempt for sending a data frame. $c(k)$ is an indicator function. $c(k) = 1$ if k in [0 $CW(I)$]; otherwise $c(k) = 0$.

4.2.1.7 Summary of Analysis

Here the relations among earlier analysis are summarised.

1. In Section 4.2.1.1, two novel Markov chain models, presented in Fig 4.1 and Fig 4.2 respectively, have been created for each station category. In addition to τ_D and τ_E, other unknown transition probabilities include $\theta_D(k)$, $\theta_E(k)$ $\gamma_D(d)$, $\gamma_E(d)$, $\overline{\omega_D}$, and $\overline{\omega_E}$.

2. In Section 4.2.1.2, the zone specific transmission probability $\beta(k)$ has been obtained in terms of τ_D and τ_E. A new Markov chain model, shown in Fig 4.4, is created for modelling the number of consecutive idle time slots between two successive transmissions, and its state probability $s(r)$ has been obtained in terms of $\beta(k)$. Thus, $s(r)$ can also be obtained in terms of τ_D and τ_E.

3. In Section 4.2.1.3 - 4.2.1.6, based on the results obtained in Section 4.2.1.2, the unknown transition probabilities $\theta_D(k)$, $\theta_E(k)$ $\gamma_D(d)$, $\gamma_E(d)$, $\overline{\omega_D}$, and $\overline{\omega_E}$, have also been obtained in terms of τ_D and τ_E.

4. All unknown parameters have been expressed in terms of τ_D and τ_E. By considering the relationships between the states in the Markov chains shown in Fig 4.1 and Fig 4.2, each state probability can be expressed in terms of τ_D and τ_E. We may denote the sum of the state probabilities in the Markov chain model in Fig 4.1 as $\sum s_D$. $\sum s_D$ can also be expressed as in terms of τ_D and τ_E. Similarly, we may denote the sum of the state probabilities in the Markov chain model in Fig 4.2 as $\sum s_E$, which can also be expressed in terms of τ_D and τ_E. If we considering that the sum of a Markov chain's state probability should be equal to 1, we may obtain two independent non-linear equations, that is, $\sum s_D = 1$ and $\sum s_E = 1$. Both equations depends on τ_D and τ_E. Therefore, a non-linear equation system is finally constructed. This can result in the values of of τ_D or τ_E can be numerically solved. The numerical calculation tool we used to solve non-linear equation systems is $fsolve$ function from the optimisation toolbox in MATLAB [83].

4.2.2 Coexistence of Non-QSTAs and QSTAs with Voice or Video Traffic

4.2.2.1 Basic Markov Chains

In this section, we analyse the performance when non-QSTAs and QSTAs carrying either voice or video traffic coexist in the same base station set. Two Markov chain models are required for modeling non-QSTAs and QSTAs respectively. The Markov chain model shown in Fig 4.1 can be used for modelling the non-QSTA, and a new Markov chain model is created for the QSTA, which is shown in Fig 4.6.

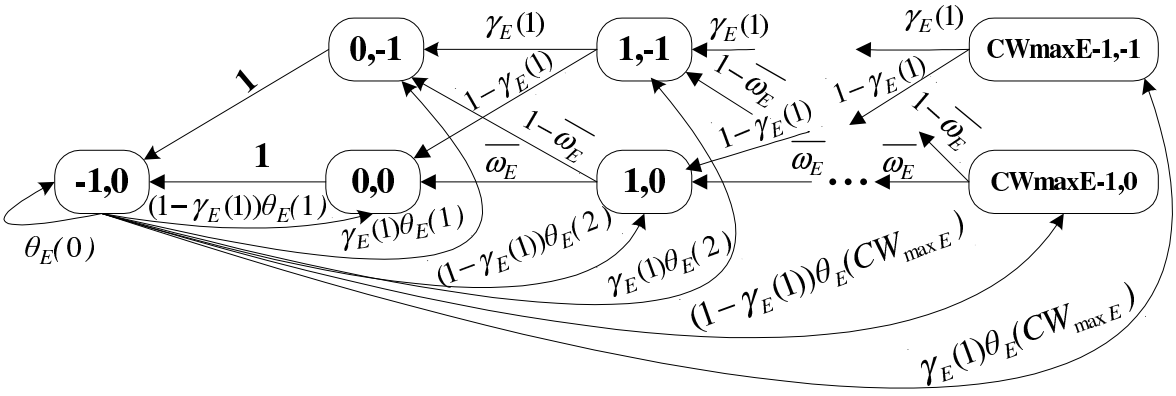


Figure 4.6: The Markov chain model for a QSTA carrying voice or video traffic.

Because $IFS_E = IFS_D$ in this system, there is no transmission during the IFS_E interval after the busy channel. Therefore, states $(k, l), 0 \leq k \leq CW_{max}, 1 \leq l \leq C$ are not required.

4.2.2.2 The zone specific transmission probability

The time slot distribution between two successive transmissions in the system is illustrated in Fig 4.7. Here $M = min(CW_{max\_D}, CW_{max\_E})$.

According to Fig 4.7, two zones exist. In the first time slot after the IFS_D, referred to as zone 1, non-QSTAs involved in the previous transmission and QSTAs may transmit. In the remaining time slots, referred to as zone 2, all non-QSTAs may start a transmission. The

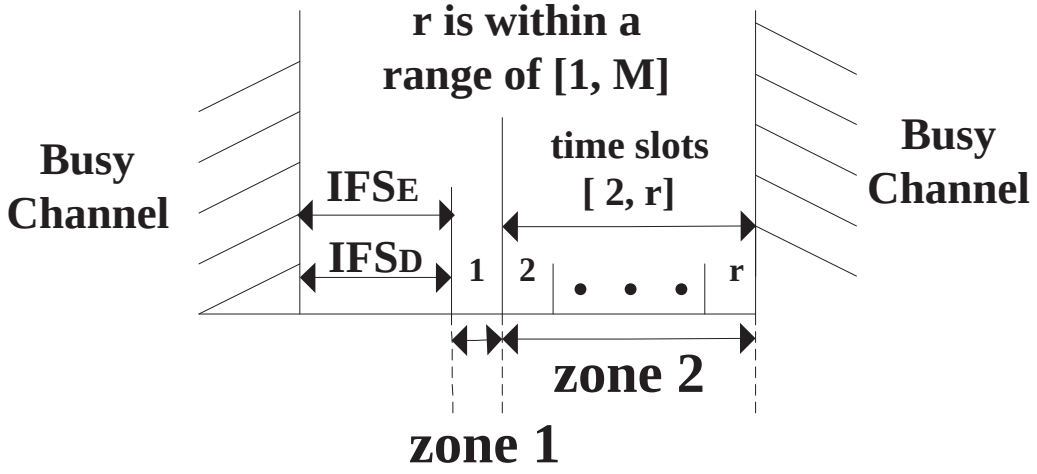


Figure 4.7: Time slots between two successive transmissions in the system.

corresponding zone specific transmission probabilities can be obtained by

$$\begin{cases} \beta(1) = \sum_{i=0}^{N_D} \left\{ \left[1 - (1-\tau_D)^i (1-\tau_E)^{N_E} \right] \phi(i) \right\}, \\ \beta(2) = 1 - (1-\tau_D)^{N_D} (1-\tau_E)^{N_E}. \end{cases} \tag{4.18}$$

Also from Fig 4.7, a new discrete time one-dimensional Markov chain can be created to model the number of the idle time slots between two successive transmissions in the system, as shown in Fig 4.8.

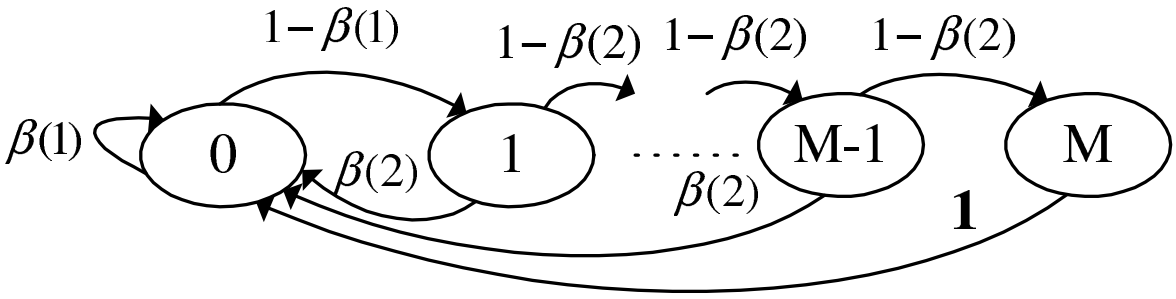


Figure 4.8: The Markov chain model for modeling the number of consecutive idle time slots between two successive transmissions in the system.

The state probability $s(r)$ for this Markov chain can be easily obtained as

$$\begin{cases} s(0) = \frac{1}{1+[1-\beta(1)]\frac{1-[1-\beta(2)]^M}{\beta(2)}}, \\ s(r) = s(0)\left[1-\beta(1)\right]\left[1-\beta(2)\right]^{r-1}, \text{for } 1 \leq r \leq M. \end{cases} \tag{4.19}$$

With the solution of s(r), we may simply obtain the probability that the system reaches a specific zone, given by

$$\begin{cases} Z(1) = s(0), \\ Z(2) = \sum_{i=1}^{M} s(i). \end{cases} \tag{4.20}$$

4.2.2.3 Transition probabilities

Other transition probabilities can be obtained analogously as those in Section 4.2.1.

$$\begin{cases} \rho_D(1) = \sum_{i=0}^{N_D-1}\left\{\left[1-(1-\tau_D)^i(1-\tau_E)^{N_E}\right]\xi(i)\right\}, \\ \rho_D(2) = 1-(1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \\ \\ \rho_E(1) = \sum_{i=0}^{N_D}\left\{\left[1-(1-\tau_D)^i(1-\tau_E)^{N_E-1}\right]\phi(i)\right\}, \\ \rho_E(2) = 1-(1-\tau_D)^{N_D}(1-\tau_E)^{N_E-1} \\ \\ \overline{\rho_D} = \sum_{i=1}^{2}\left[Z(i)\rho_D(i)\right], \\ \overline{\rho_E} == \sum_{i=1}^{2}\left[Z(i)\rho_E(i)\right], \end{cases} \tag{4.21}$$

$$\begin{cases} \omega_D(1) = \sum_{i=0}^{N_D-1}\left[(1-\tau_D)^i(1-\tau_E)^{N_E}\xi(i)\right], \\ \omega_D(2) = (1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \\ \\ \omega_E(1) = \sum_{i=0}^{N_D}\left[(1-\tau_D)^i(1-\tau_E)^{N_E-1}\phi(i)\right], \\ \omega_E(2) = (1-\tau_D)^{N_D}(1-\tau_E)^{N_E-1}, \\ \\ \overline{\omega_D} = \sum_{i=1}^{2}\left[Z(i)\omega_D(i)\right], \\ \overline{\omega_E} = \sum_{i=1}^{2}\left[Z(i)\omega_E(i)r\right], \end{cases} \tag{4.22}$$

and

$$\begin{cases} \gamma_D(1) = \sum_{i=0}^{N_D-1}\left\{\left[1-(1-\tau_D)^i(1-\tau_E)^{N_E}\right]\xi(i)\right\}, \\ \gamma_D(2) = 1-(1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \\ \gamma_E(1) = \sum_{i=0}^{N_D}\left\{\left[1-(1-\tau_D)^i(1-\tau_E)^{N_E-1}\right]\phi(i)\right\}. \end{cases} \tag{4.23}$$

Because all stations in the system use the identical procedure to choose a random initial backoff counter, we may obtain the transition probabilities, $\theta_D(k)$ and $\theta_E(k)$, with the same approach described in Section 4.2.1.6.

4.2.2.4 Summary of Analysis

As a brief summary, first, two basic Markov chain models have been presented in Section 4.2.2.1. Second, the zone specific transmission probabilities is analysed in Section 4.2.2.2, and transition probabilities have been analysed in Section 4.2.2.3. Now two independent non-linear equations obtained from the two Markov chain models shown in Fig 4.1 and Fig 4.2 respectively may give the solutions of τ_D and τ_E.

4.3 Saturated Throughput Analysis

In this section, we analyse the saturated throughput for QSTAs and non-QSTAs. According to Fig 4.3 and Fig 4.7, a maximum of M time slots may exist between two successive transmissions, and those time slots are located in different zones. Four events may occur in each time slot: (i) a successful transmission from a non-QSTA; (ii) a successful transmission from a QSTA; (iii) a collision; (iv) an idle time slot. The corresponding zone specific probabilities for the system of non-QSTAs and QSTAs with best effort or background traffic can be obtained by

$$\begin{cases} \psi_D(1) = \sum_{i=0}^{N_D} \left\{ [i\tau_D(1-\tau_D)^{i-1}]\, \phi(i) \right\}, \\ \psi_D(2) = N_D\tau_D(1-\tau_D)^{N_D-1}, \\ \psi_D(3) = N_D\tau_D(1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \\ \psi_E(1) = 0, \\ \psi_E(2) = 0, \\ \psi_E(3) = N_E\tau_E(1-\tau_D)^{N_D}(1-\tau_E)^{N_E-1}, \\ \\ \epsilon(k) = \beta(k) - \psi_D(k) - \psi_E(k), \\ \varrho(k) = 1 - \beta(k), k = 1, 2, 3, \end{cases} \tag{4.24}$$

where $\psi_D(k)$ and $\psi_E(k)$ are the probabilities of a successful transmission from a non-QSTA and a QSTA respectively, following a time slot located in zone k, $\epsilon(k)$ is the probability of a collision following a time slot located in zone k, $\varrho(k)$ is the probability of no transmission following a time slot in zone k, and $\beta(k)$ is the zone specific transmission probability given in Equation (4.1).

The corresponding zone specific probabilities for the system of non-QSTAs and QSTAs with voice or video traffic can be obtained by

$$\begin{cases} \psi_D(1) = \sum_{i=0}^{N_D} \left[i\tau_D(1-\tau_D)^{i-1}(1-\tau_E)^{N_E}\phi(i) \right], \\ \psi_D(2) = N_D\tau_D(1-\tau_D)^{N_D-1}(1-\tau_E)^{N_E}, \\ \psi_E(1) = \sum_{i=0}^{N_D} \left[N_E\tau_E(1-\tau_D)^{i}(1-\tau_E)^{N_E-1}\phi(i) \right], \\ \psi_E(2) = N_E\tau_B(1-\tau_E)^{N_E-1}(1-\tau_D)^{N_D}, \\ \\ \epsilon(k) = \beta(k) - \psi_D(k) - \psi_E(k), \\ \varrho(k) = 1 - \beta(k), k = 1, 2, \end{cases} \tag{4.25}$$

where $\beta(k)$ is given in Equation (4.18).

Therefore, the average effective payload for non-QSTAs or QSTAs can be obtained by

$$\begin{cases} \text{For the system of non-QSTAs and QSTAs with best effort or background traffic:} \\ E[DCF] = \sum_{i=1}^{3} \left[Z(i)\psi_D(i)P \right], \\ E[EDCA] = \sum_{i=1}^{3} \left[Z(i)\psi_E(i)P \right], \\ \\ \text{For the system of non-QSTAs and QSTAs with voice or video traffic:} \\ E[DCF] = \sum_{i=1}^{2} \left[Z(i)\psi_D(i)P \right], \\ E[EDCA] = \sum_{i=1}^{2} \left[Z(i)\psi_E(i)P \right], \end{cases} \tag{4.26}$$

where P is the payload size of a data frame, which is considered as a known constant, and Z(i) is the probability that the system resides in zone i, given in Equation (4.4) and Equation (4.20) respectively.

The average time between two successive transmissions can be obtained as:

$$\begin{cases} \text{For the system of non-QSTAs and QSTAs with best effort or background traffic:} \\ EL = \sum_{i=1}^{3} \{Z(i)[(\psi_D(i) + \psi_E(i))Ts + \epsilon(i)Tc + \varrho(i)TimeSlot]\}, \\ \\ \text{For the system of non-QSTAs and QSTAs with voice or video traffic:} \\ EL = \sum_{i=1}^{2} \{Z(i)[(\psi_D(i) + \psi_E(i))Ts + \epsilon(i)Tc + \varrho(i)TimeSlot]\}, \end{cases} \tag{4.27}$$

where Ts and Tc are the time required for a successful transmission and a collision respectively. They can be obtained by

$$Ts = H + P + SIFS + ACK + DIFS, \tag{4.28}$$

and

$$Tc = H + P + DIFS + ACKtimeout, \tag{4.29}$$

where H is the time required for transmitting the physical layer header and the MAC layer header of a frame, P is the time required for transmitting the data payload of a frame, ACK is the time required for transmitting an ACK frame, $ACKtimeout$ is time required for a sending station to detect an unsuccessful transmission.

Finally, the throughput for each station of each category can be obtained by

$$\begin{cases} Throughput_{DCF} = E[DCF]/EL/N_D, \\ Throughput_{EDCA} = E[EDCA]/EL/N_E. \end{cases} \tag{4.30}$$

4.4 Simulation Study

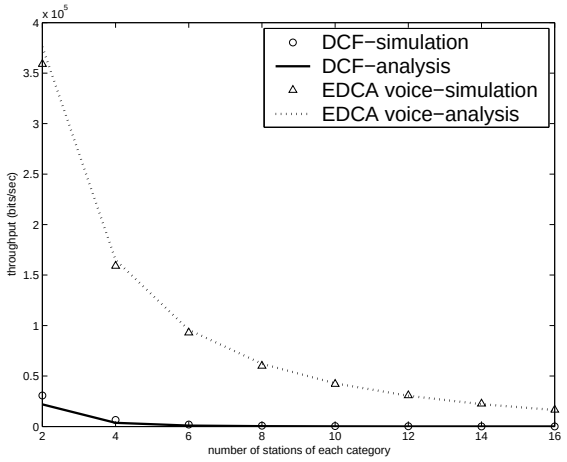
The simulation study is carried out using OPNET [91]. The parameters of DCF and EDCA are listed in Table-4.1, consistent with those defined in [16, Table 20df, p.49].

Four scenarios are simulated. Each of them contains an equal number of non-QSTAs and QSTAs carrying traffic from one AC. The results are shown in Fig 4.9.

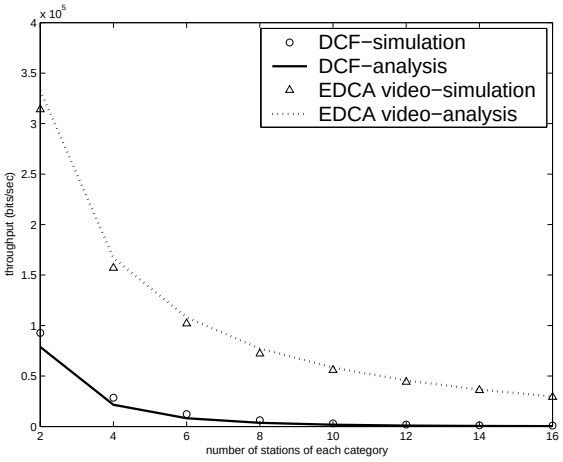
As illustrated in Fig 4.9, the analytical results from the proposed model generally agree well the simulation results, especially when the number of stations is large. However, a larger discrepancy between the analytical and the simulation results is observed when the number

Table 4.1: WLAN simulation parameter setting in DCF and EDCA coexistence performance analysis

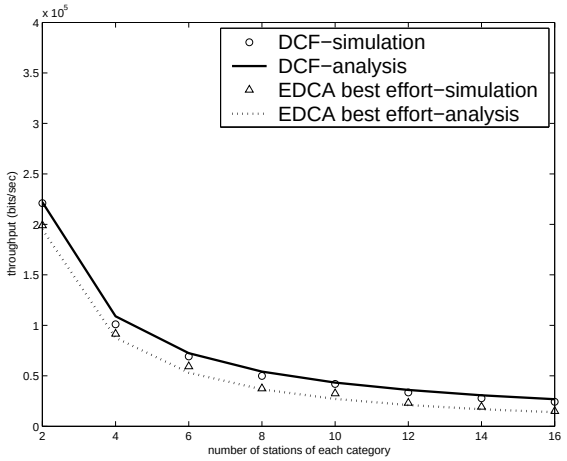
Frame payload size	8000 bits
data rate	1Mbps
Maximum retransmission limit	7
DCF parameter set	$CW_{min} = 31$, $CW_{max} = 1023$, DIFS=SIFS+2×TimeSlot
EDCA voice parameter set	$CW_{min} = 7, CW_{max} = 15$, AIFS=DIFS
EDCA video parameter set	$CW_{min} = 15, CW_{max} = 31$, AIFS=DIFS
EDCA best effort parameter set	$CW_{min} = 31, CW_{max} = 1023$, AIFS=DIFS+TimeSlot
EDCA background parameter set	$CW_{min} = 31, CW_{max} = 1023$, AIFS=DIFS+5×TimeSlot



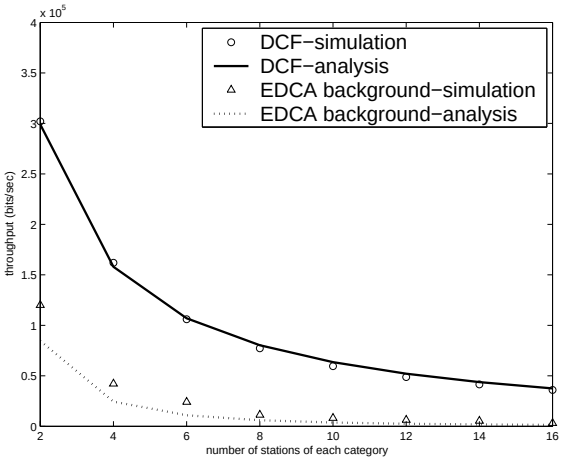
(a) DCF + EDCA voice



(b) DCF + EDCA video



(c) DCF + EDCA best effort



(d) DCF + EDCA background

Figure 4.9: Simulation and analytical results

of stations is small. It is caused by an assumption used in the model, that is, the transmission probability at a generic time slot is constant. This assumption is more accurate when the number of stations is larger [26].

It is observed that EDCA ACs voice and video have higher priority over DCF, and DCF has higher priority over EDCA AC background. This is caused by the large differences between their CW sizes and IFSs. It is also observed tha DCF has a marginal priority over EDCA AC best effort, which is caused by IFS_D being only one time slot shorter than IFS_E of AC background. It is obvious that traffic prioritisation can still be implemented effectively in the coexistence condition. However, the results also imply that non-QSTAs may suffer a serious service starvation if they coexist with QSTAs carrying voice or video traffic.

4.5 Summary

In this chapter, we have proposed novel Markov chain models for analysing the coexistence of DCF and EDCA. Some important factors were considered in our analysis, including the CW size, the IFS, the backoff counter decrement rule, and the start of a transmission when the backoff counter reaches zero. Saturated throughput for QSTAs and non-QSTAs has been obtained using the proposed model. Simulation study has verified the accuracy of the proposed model. The results we observed has indicated that traffic prioritisation can be effectively implemented in the coexistence environment, but non-QSTAs may suffer a serious service starvation when they coexists with QSTAs carrying high-priority traffic.

Chapter 5

Performance Analysis of DCF Using Data-rate Switching

In contrast to Chapter 5 and Chapter 6 which have focused on DCF and EDCA, this chapter focuses on the impact of data-rate switching on the IEEE 802.11 network performance. Due to the fact that the detailed data-rate switching mechanism is not defined in IEEE 802.11 standard, the majority of the existing studies about IEEE 802.11 network performance analysis have simply ignored it. The literature review in Chapter 3 has shown a lack of analytical work in this area. In this chapter, an analytical model is proposed to investigate the performance of DCF using data-rate switching. The results will demonstrate the impact of data rate switching on the network performance, and these results can be applied straightforwardly to EDCA.

The rest of this chapter is organised as follow. In Section 5.1, a commonly used data-rate mechanism mentioned in [92] is briefly introduced; In Section 5.2, the proposed model is presented; In Section 5.3, the saturated throughput is analysed; Simulation study is carried out in Section 5.4; Finally Section 5.5 concludes this chapter.

5.1 Data-rate Switching in IEEE 802.11

Most IEEE 802.11 products support multiple data rates to cater for different channel and traffic conditions. Support for multiple data rates has also been included in the IEEE 802.11 standard [6] although the details of the multiple data rate switching mechanism have been left for the equipment manufacturers. In [92], Inoue $et\ al.$ discovered that most commercial IEEE 802.11 products use a simple mechanism to implement data-rate switching. That is, if a station has a predetermined U ($U \geq 1$) number of consecutive successful transmissions, it will increase its data rate to a higher data rate until the highest data rate has been reached. If the station suffers a predetermined D ($D \geq 1$) number of consecutive unsuccessful transmissions, it will decrease its data rate to a lower data rate until the lowest data rate has been reached.

5.2 The Markov Chain Model

In this section, we will present the Markov chain model considering the data rate switching mechanism introduced in [92]. First the basic Markov chain model will be introduced. Secondly we will analyse each state of the basic Markov chain model in further detail, which will relate the state probability in the basic model to the transmission probability of a station. Finally we will summarise this section and obtain the final solution.

The following assumptions are used in the model.

- Traffic load is saturated.
- The number of stations, n, is fixed and known.
- The transmission probability of a station in a generic time slot is a constant, denoted by τ. The value of τ is unknown and to be solved.
- Only two data rates, R2 and R1 (R2>R1), are considered for simplicity. The maximum retransmission limit for sending a data frame, m, is set to be 7 [6, p. 361]. In this chapter, we consider $m > D$. The proposed model can be easily revised for $m \leq D$.

- The transmission error is measured in frame error rate (FER). FERs at the data rates of R2 and R1 are set to be known constants, denoted by FER_2 and FER_1 respectively. The transmission error occurs on data frame only. FER_2 and FER_1 are the same for all stations.

5.2.1 The Basic Markov Chain Model

The proposed Markov chain model is shown in Fig 5.1(a). There are three state variables in the model, i.e., $u(t)$, $e(t)$, and $q(t)$. The first state variable, $u(t)$, models the data rate switching of a given station, and $u(t) = 2$ (representing R2) or $u(t) = 1$ (representing R1). The second state variable, $e(t)$, models the number of consecutive successful and unsuccessful transmissions experienced by the station, which is explained in the following.

1. $e(t) = -i$, $i \geq 1$ represents that the station has suffered i consecutive unsuccessful transmissions before the current transmission.

2. $e(t) = i$, $i \geq 1$ represents that the station has experienced i consecutive successful transmissions before the current transmission.

3. $e(t)$ will be reset to 0 or it will remain 0 in three occasions: i) The station experiences a rate switching; ii) The station experiences an unsuccessful transmission at R1; iii) The station experiences a successful transmission at R2. With such definition, we may avoid unnecessary Markov states to record the number of consecutive successful transmissions at R2 and the number of consecutive unsuccessful transmissions at R1, and it will simplify the Markov chain.

Finally, the third state variable, $q(t)$, models the number of the transmission attempts involved in sending a single data frame. $q(t) = j$, $j \geq 1$ means that the station is performing the j^{th} transmission attempt. When the maximum retransmission limit, m is reached, the frame will be dropped and $q(t)$ will be reset to 1, which means a new data frame will be transmitted.

Denote the state probability of the Markov chain by $s(u(t), e(t), q(t))$, the transition equations are given by:

$$s(2,0,1) = s(2,-j+1,j)(1-p_2), 1 \leq j \leq D \quad (5.1)$$

$$s(2,-j,j+1) = s(2,-j+1,j)p_2, 1 \leq j \leq D-1 \quad (5.2)$$

$$s(1,0,D+1) = s(2,-D+1,D)p_2, \quad (5.3)$$

$$s(1,0,j+1) = s(1,0,j)p_1, 1 \leq j \leq m-1, \quad (5.4)$$

$$s(1,0,1) = s(1,0,m)p_1, \quad (5.5)$$

$$s(1,1,1) = s(1,0,j)(1-p_1), 1 \leq j \leq m, \quad (5.6)$$

$$s(1,j,1) = s(1,j-1,1)(1-p_1), 2 \leq j \leq U-1, \quad (5.7)$$

$$s(2,0,1) = s(1,U-1,1)(1-p_1), \quad (5.8)$$

where p_2 and p_1 are the probabilities that a transmission from the station is unsuccessful at the data rates R2 and R1 respectively:

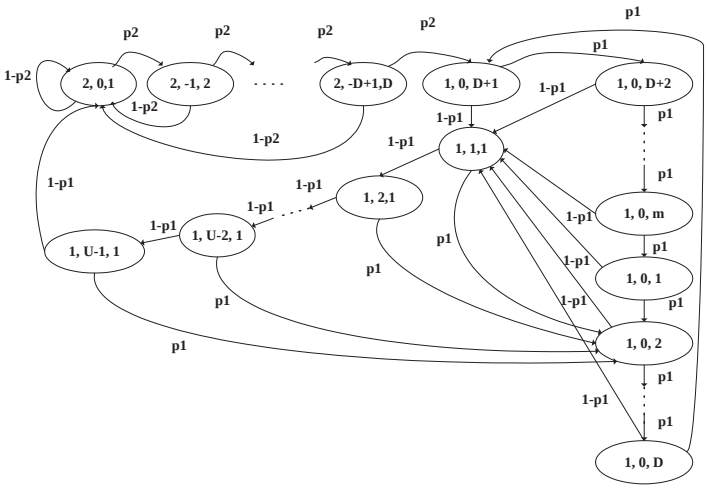
$$p_i = p_c + (1-p_c)FER_i, i = 1, 2. \quad (5.9)$$

Here p_c is the probability that a transmission from the station collides with transmissions from other stations, given by $p_c = 1 - (1-\tau)^{n-1}$.

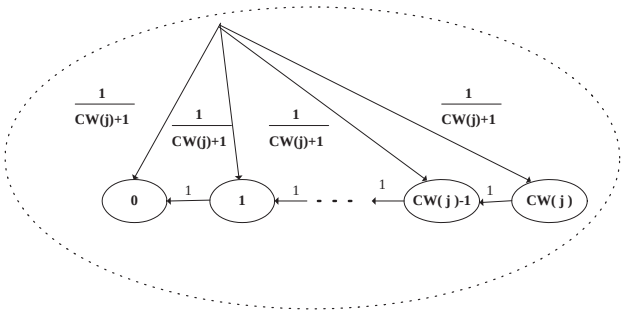
The transition equations are explained as follows: Equation (5.1) represents a successful transmission at R2; Equation (5.2) represents an unsuccessful transmission at R2 and the next transmission should be at R2 because the limit of D consecutive unsuccessful transmissions at R2 is not reached yet; Equation (5.3) represents a decrement of the data rate from R2 to R1 when the station experiences D consecutive unsuccessful transmissions at R2; Equation (5.4) and Equation (5.5) represent unsuccessful transmissions at R1; Equation (5.6) and Equation (5.7) represent successful transmissions at R1; finally Equation (5.8) represents an increase of the data rate from R1 to R2 when the station experiences U consecutive successful transmissions at R1.

5.2.2 The Transmission Probability, τ

The Markov chain model in Fig 5.1(a), however, does not allow us to relate the state probabilities to the transmission probability of a given station, τ, which must be found in order to



(a) The basic Markov chain model.



(b) The sub-states of a state in the basic Markov chain.

Figure 5.1: The Markov chain models used in this work.

determine the collision probability, p_c, in Eqaution (5.9). To solve the problem, the evolution of the backoff counter in each state of the Markov model in Fig 5.1(a) is modeled and shown in Fig 5.1(b). In Fig 5.1(b), the sub-state represents the value of the backoff counter of the station [26]. It varies between 0 and $CW(j)$, where $CW(j)$ is the contention window size corresponding to the j^{th} transmission attempt from the station for sending a frame. The value of j is determined by the state variable $q(t)$ of the Markov model. Details about how $CW(j)$ varies with j can be found in [6] or the previous part in this thesis (Chapter 1,Section 1.1). When the sub-state (0) is reached, a transmission will occur.

Let $s(i,k,j)$ be the state probability of a given state in the Markov chain model presented in Fig 5.1(a), and $b(r)$ be the state probability of its sub-state (r), $0 \leq r \leq CW(j)$. Based on Fig 5.1(b), it can be readily obtained that

$$b(0) = s(i,k,j)\frac{2}{CW(j)+2}. \tag{5.10}$$

When the backoff counter reaches zero, a transmission will occur. The sum of $b(0)$s for all the states in the Markov chain model shown in Fig 5.1(a) should be equal to the transmission probability τ:

$$\tau = \sum_{i,k,j} s(i,k,j)\frac{2}{CW(j)+2}. \tag{5.11}$$

5.2.3 Summary of Analysis

In this section, a basic Markov chain model has been created in Fig 5.1(a) which models a station's backoff stage, data rate, and the number of the consecutive successful or unsuccessful transmissions. Each state in this basic Markov chain model represents a transmission event (including the related backoff procedure). Based on this basic Markov chain model, system equations (5.1)-(5.11) are created.

Each state in the basic Markov chain has been evolved into a fixed number of sub-states, as shown in Fig 5.1(b). Similar to Bianchi's work in [26], these sub-states represent the backoff counter decrement procedure. Accordingly, the time scale in this hierarchical Markov chain model is per slot scale, identical to that used in [26].

Finally, considering that the sum of the state probabilities of a Markov chain is 1, we may

obtain

$$\sum_{i,k,j} s(i,k,j) = 1. \tag{5.12}$$

With (5.12) and (5.1)-(5.11), a non-linear equation systems about τ can be obtained. This will lead to the numerical solution of τ, p_c, p_1, p_2, and eventually $s(i,k,j)$. The numerical calculation tool we used to solve non-linear equation systems is $fsolve$ function from the optimisation toolbox in MATLAB [83].

5.3 Saturated Throughput

Within a generic time slot, one of the following four events may occur: (i) the channel remains idle; (ii) a successful transmission starts; (iii) an unsuccessful transmission occurs due to transmission error; (iv) a collision occurs. The probability that the channel remains idle is given by

$$P_{idle} = (1-\tau)^n, \tag{5.13}$$

where τ has been solved in the last section.

Because a transmission will occur at either R2 or R1 which takes different amount of time, we calculate the conditional probabilities τ_2 and τ_1, representing that a transmission occurs at R2 and R1 respectively:

$$\tau_2 = \left[\sum s(2,i,j)\tfrac{2}{CW(j)+2}\right]/\tau, \tag{5.14}$$

$$\tau_1 = \left[\sum s(1,i,j)\tfrac{2}{CW(j)+2}\right]/\tau, \tag{5.15}$$

where parameters $s(2,i,j)$ and $s(1,i,j)$ have been solved in the last section.

Accordingly, we may obtain the following probabilities:

$$P_{suc\_2} = n\tau(1-\tau)^{n-1}\tau_2(1-FER_2), \tag{5.16}$$

$$P_{suc\_1} = n\tau(1-\tau)^{n-1}\tau_1(1-FER_1), \tag{5.17}$$

$$P_{FER\_2} = n\tau(1-\tau)^{n-1}FER_2\tau_2, \tag{5.18}$$

$$P_{FER\_1} = n\tau(1-\tau)^{n-1}FER_1\tau_1, \tag{5.19}$$

$$P_{col\_2} = \sum_{i=2}^{n}\binom{n}{i}\tau^i(1-\tau)^{n-i}\tau_2^i, \tag{5.20}$$

$$P_{col\_1} = \sum_{i=2}^{n}\binom{n}{i}\tau^i(1-\tau)^{n-i}(1-\tau_2^i). \tag{5.21}$$

Equation (5.16) and Equation (5.17) calculate the probabilities that a successful transmission occurs at R2 and R1 respectively; Equation (5.18) and Equation (5.19) calculate the probabilities that an unsuccessful transmission caused by transmission error occurs at R2 and R1 respectively; Equation (5.20) calculates the probability that a collision occurs and all stations involved transmit at R2; finally Equation (5.21) calculates the probability that a collision occurs and at least one station involved transmits at R1, which will result in a longer duration for the collision than that in Equation (5.20).

Finally, the overall throughput can be obtained:

$$Throughput = \left[(P_{suc\_1} + P_{suc\_2})E[P]\right]/EL, \tag{5.22}$$

where $E[P]$ is payload size of the data frame, and EL is the average time duration required for the four possible events, given by $EL = \sum P_{event}T_{event}$. Here P_{event} is the probability for the four aforementioned events, given in Equations (5.13), (5.16)-(5.21), and T_{event} is the duration for each event. The related calculations for T_{event} can be found in the previous part of this thesis (Chapter 3,Section 3.3, and Chapter 4, Section 4.3).

5.4 Simulation Study

In our simulation using OPNET [91], IEEE 820.11b DSSS (direct sequence spread spectrum) physical layer is used. The payload size of the data frame is 4000 bits. Stations will

transmit the payload at two data rates: $11\ Mbps$ and $5.5\ Mbps$. In our simulation, DQPSK (Differential Quadrature Phase Shift Keying) is used at both data rates [6, pp. 195-223], and all stations use the identical transmission power under the same additive white Gaussian noise (AWGN) channel condition at both data rates. Therefore the carrier to noise ratio (CNR) is the same at both data rates. Accordingly, bit error rate (BER) at both data rates can be obtained [93]:

$$BER = 0.5 \exp(-10^{\frac{CNR}{10}}), \tag{5.23}$$

where CNR is the value of carrier to noise ratio in dB.

With BER, FER for data frame can be obtained by $FER = 1 - (1 - BER)^{PL}$. We set $U = 8$, $D = 3$ following the setting in [92]. Frame header and ACK frame are always transmitted at $1\ Mbps$.

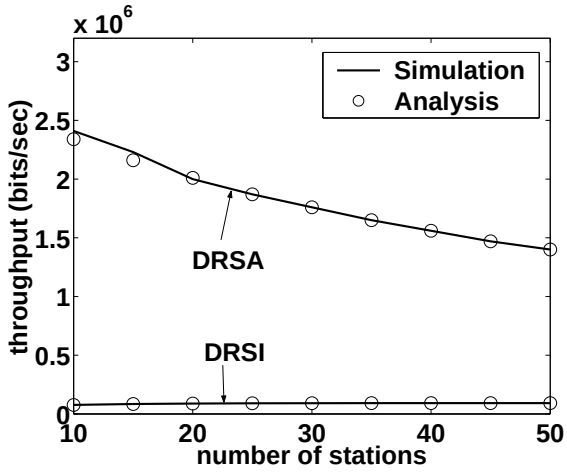
The results for different channel conditions are shown in Fig 5.2(a)- 5.2(c), where we observe that the analytical results generally agree very well with the simulation results. For comparison, we also simulate the scenarios that the data rate is fixed at 11Mbps, shown in Fig 5.2(a)- 5.2(c) as well[1].

Because IEEE 802.11 standard does not differentiate whether an unsuccessful transmission is caused by either transmission error or collision, the effect of using data-rate switching is determined by both CNR and the number of competing stations. When CNR$\leq$5dB, the transmission error is large and has a dominant impact, consequently using data-rate switching always results in an improved throughput compared with the scenario not using data-rate switching, as illustrated in Fig 5.2(a). When CNR$\geq$7dB, the transmission error is small, and most transmission failures are caused by collision. It is always beneficial for stations to transmit at a higher data rate, and using data-rate switching will result in a reduced throughput, as shown in Fig 5.2(c). When 5dB$<$CNR$<$7dB, the impacts of transmission error and collision are close. When the number of stations is large, collision will have a dominant impact and it is beneficial for stations to transmit at a higher data rate and the converse, as shown in Fig 5.2(b). In Fig 5.2(d), two regions are marked according to CNR and the number of stations: in region 2, using data-rate switching can increase the throughput, and in region 1, using data-rate switching will reduce the throughput.

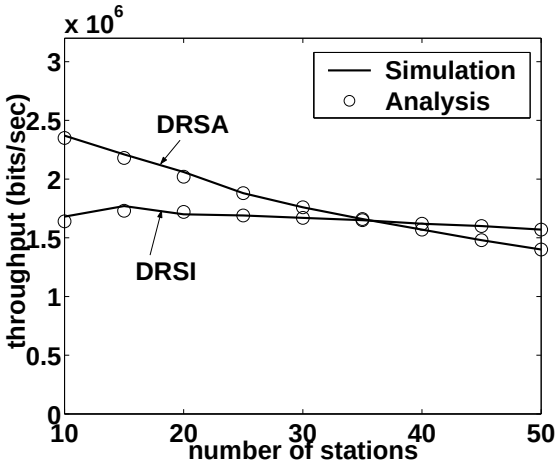
[1]The analytical results for the scenarios without using data-rate switching are obtained based on the work in [94].

5.5 Summary

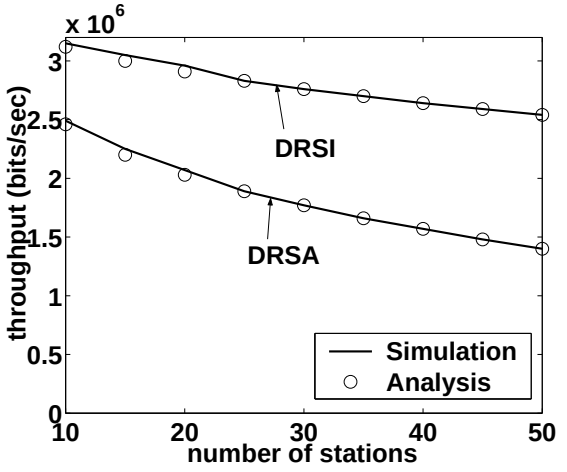
In this chapter, a Markov chain model for IEEE 802.11 DCF has been presented, considering a data-rate switching mechanism used by the most commercial IEEE 802.11 products. Using the proposed model, the saturated throughput has been obtained. The accuracy of the proposed model has been validated using simulation study. The results have shown different impacts of the data-rate switching on the network performance under different network conditions. When the channel condition is poor and the number of stations is small, using data-rate switching can significantly improve the network performance, otherwise the improvement is ignorable or it may even degrade the network performance. A series of threshold values for the channel condition as well as the number of stations have been obtained. The analytical model developed in this chapter will be helpful for designing guidelines assisting the decision on whether or not to use data-rate switching in a specific wireless environment, without resorting to lengthy simulations and experimentation.



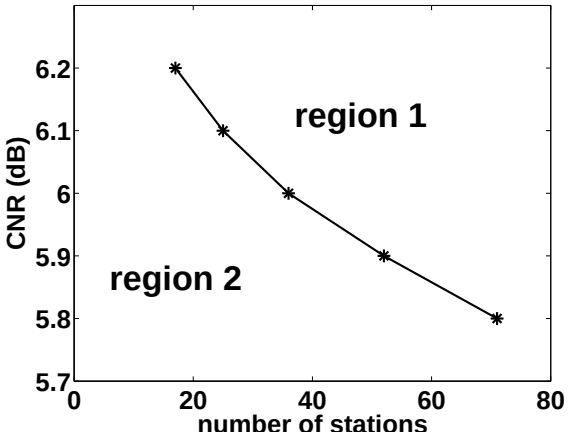
(a) CNR=5dB, FER_{11}=0.9737, $FER_{5.5}$=0.0068.



(b) CNR=6dB, FER_{11}=0.5084, $FER_{5.5}$=2.6174e-004.



(c) CNR=7dB, FER_{11}=0.0868, $FER_{5.5}$=4.3223e-006.



(d) Threshold values for CNR and number of stations.

Figure 5.2: The simulation and analysis results ("DRSA" means "data-rate switching active", and "DRSI" means "data-rate switching inactive").

Chapter 6

Literature Review: Wireless Cooperative Retransmission

Beginning with this chapter, the remaining part of this thesis focuses on wireless cooperative network performance optimisation rather than IEEE 802.11 network performance analysis. Two uncoordinated distributed wireless cooperative retransmission strategies will be presented in this part.

First, the related work on wireless cooperative retransmission is discussed in this chapter,

6.1 Related Work

The majority of existing studies on wireless cooperative retransmission have taken place in the physical layer context, with a growing literature on *cooperative diversity* methods, such as those in [95–103]. Essentially, this can be seen as an extension of the *spatial diversity* concept of MIMO (multiple-input multiple-output), where the multiple antennas are located at the cooperative nodes. Such cooperative-diversity methods require the support of complex physical layer technologies so that the receivers must be able to combine the cooperative signals and decode them jointly. The work in [95–103] focuses on developing such physical layer technologies. They can be classified into two categories: coding based cooperation and

non-coding based cooperation. The approach of coding based cooperation has attracted significant research attention, and most of the existing studies use this approach [95–100]. This approach integrates cooperation into channel coding by using space time coding technology such that each cooperative neighbour encodes the data from the source with space-time coding technology and transmits the encoded data to the destination. This allows the destination to combine the signals from multiple nodes, including both the cooperative neighbors and the source, in the physical layer and decode the combined signal to retrieve the original data from the source. In [95–97], various coding based cooperation strategies are presented. In [98–100], some space time coding schemes are proposed to realize the aforementioned coding based cooperation strategies. In contract, the approach of non-coding based cooperation uses technologies other than the space time coding technology, and the number of the existing studies using non-coding based cooperation approach is limited, such as those in [101–103]. In [101, 102], a simple code-division multiple access (CDMA) system is presented that implements the decode-and-forward cooperative communication procedure, in which each cooperative neighbour uses a distinct spreading code. In [103], a cooperative communication system using time-division channel is proposed, in which a separate time slot is assigned to each cooperative neighbour.

The collision issue caused by multiple simultaneous transmissions, which presents a major challenge in the MAC layer, becomes trivial for such physical layer cooperative retransmission strategies, because these strategies can exploit the space diversity gain of multiple simultaneous transmissions. However, such physical layer cooperative retransmission strategies requires additional hardware equipments. For example, the transceivers may be required to be equipped with multiple antennas and multiple decoders to implement space-time coding. In contrast, wireless cooperative retransmission methods on higher layers (for example, MAC layer) can be used with simple physical layer technologies and can be implemented with simple transceivers with a traditional single-antenna/single-user decoder, like popular IEEE 802.11 adaptors. However, the MAC layer cooperative retransmission strategies need to consider the collision caused by the mutual interference between multiple simultaneous transmissions.

To solve the collision problem in the MAC layer cooperative retransmission strategies, some researchers use the technique of *opportunistic forwarding* [104–113], where the forwarder (or the next hop) of each frame is determined on-the-fly (rather than pre-selected by a routing protocol), through local coordination among the neighbours that overhear the frame. This co-

ordination is usually achieved using acknowledgment (ACK) or clear-to-send (CTS) frames returned by neighbours.

In [104], the source first broadcast the data frame to the first and the second nearest neighbours that are located between the source and the intended destination. Compared with the first nearest neighbour, the second nearest neighbour has a longer distance to the source but a shorter distance to the destination. If the second nearest neighbour receives the frame, it will take over the first nearest neighbor and becomes the forwarder of the frame. In [105], the source first transmits the data frame, and some neighbours may overhear this transmission and store the transmitted frame in their buffers. Once the source's transmissions ends, those neighbours with the data frame will return an ACK frame to the source to acknowledge their reception of the data frame. The source will select one of them as the forwarder. The details of how to coordinate the transmissions of the source and the multiple neighbours are not explicitly specified in [104, 105]. But it is mentioned in [105] that such a coordination mechanism can cause extra retransmission overhead. Such a coordination mechanism performs two functions: first, it decides how multiple neighbours acknowledge the source about their intention as the forwarder and/or their reception of the frame; second, it decides how the source selects one of them as the forwarder while keeping other neighbours informed about this selection so that these neighbors will abort their retransmission attempt.

Other cooperative retransmission strategies based on *opportunistic forwarding* propose some detailed coordination mechanisms [106–113]. These coordination mechanisms can be classified into two categories. In the first category, a set of neighbours are pre-selected as the potential forwarders, and each of them is assigned with a distinct priority. Such a priority can be defined based on various factors, such as a neighbour's distance to the intended destination or its battery energy level. A neighbour with a shorter distance to the intended destination or a higher battery energy level usually has a higher priority to be selected as the forwarder. Each of these potential forwarders responds to the source with an ACK or CTS frame to indicate whether it can act as the forwarder or not. The transmissions of these ACK or CTS frames are staggered in time in the order of the pre-defined priorities so that they will not collide. The source will select one of the neighbours that respond as the forwarder (usually it is the neighbour that first makes the response). The work in [106, 107] belongs to such a category. In [106], a handshake procedure is implemented before the data frame is transmitted from the source to the neighbours. The source first broadcasts a request-to-send (RTS) frame, several pre-selected neighbours will respond by broadcasting a CTS frame separately,

if they hears the RTS frame from the source. These CTS transmissions are staggered in time in the order of a pre-defined priorities. As mentioned in [106], such priorities may depend on the neighbour's channel quality or any other specific factors. The source node will start the transmission of the data frame immediately after it gets a CTS response from the neighbour with the highest priority. Other neighbours will abort their retransmission attempt, because the neighbor with the highest priority transmits its CTS frame ahead of all other neighbors. The neighbor with the highest priority will become the forwarder of the frame. In [107], several potential forwarders are pre-chosen by the source, and each potential forwarder is assigned with a distinct priority depending on its distance to the intended destination. After the source transmits the data frame, each potential forwarders is given a transmission time slot in the order of their priorities. If a potential forwarder overhears the frame, it will transmit in its time slot unless another potential forwarder with a higher priority already transmits. If this potential forwarder does not overhear the frame, it will pass the chance to the next potential forwarder with a lower priority. After finally hearing the transmission from one forwarder, all other potential forwarders as well as the source will stop their retransmission attempt.

In the second category, a set of neighbours may be pre-selected as potential forwarders, or all neighbours are considered as potential forwarders. Different from the first category in which a pre-defined priority order is followed, these potential forwarders will contend with each other to become the forwarder. Each of them waits a random time before it responds to the source with an ACK or CTS frame. The neighbour that responds first will be selected as the forwarder, and other neighbours will abort their retransmission attempt after they hear the ACK or CTS frame from that neighbour. The random waiting time of each neighbour may be drawn from a range specific to that neighbor. This range may depend on various factors, such as a neighbour's distance to the intended destination or its battery energy level. For example, neighbour 1 has a shorter distance to the destination compared with neighbour 2. A range $[0, a_1]$ is applied to neighbour 1, and another range $[0, a_2]$ is applied to neighbour 2. Here $0 < a_1 < a_2$ such that neighbour 1 has a higher probability to wait a shorter time. Due to this randomness of waiting time, the transmissions of these ACK or CTS frames are also random and the chance of collision is small. The source can decide the forwarder only when it receives a response from one neighbour. The work in [108–113] belongs to such a category. In [108, 109], each potential forwarder waits a random time before it responds a CTS frame to the RTS frame from the source. The major difference between the work in [108, 109] is that the set of cooperative neighbours is pre-selected in [108] according

to the energy consumption estimation, whereas there is no such limitation for cooperative neighbours in [109]. No explicit definition is given in [108, 109] about how to decide this random waiting time. In [110, 111], an IEEE 802.11 based backoff mechanism and a busy tone mechanism [114] are used to help the avoidance of collision between CTS frames from multiple neighbours. In [112, 113], such a random waiting time is drawn from a pre-defined range, and this range depends on various factors. In [112], this range depends the neighbour's residual battery energy level, its geographical position, and its channel quality. In [113], this range simply depends on the neighbour's channel quality.

Such *opportunistic forwarding* methods work well in multi-hop (and especially dense) network settings, but the excessive overhead introduced by their coordination process for every frame render them unsuitable for delay-critical applications in a single-hop setting, which is still popular and practical in the wireless cooperative networks.

In comparison, the approach of uncoordinated distributed wireless cooperative retransmission appears promising for the sing-hop connection. With such an approach, multiple uncoordinated neighbours may participate in the retransmission attempt, and they do not attempt to agree on that just one forwarder is allowed to transmit. Therefore, such an approach does not need a coordination mechanism to pre-choose a sole forwarder. Accordingly, it may avoid the extra coordination overhead. Most of the existing studies using this approach focus on the development of cooperative ARQ (Automatic Repeat-reQuest) methods, such as those in [115–120]. In [115], a fixed TDMA scheme is used, so that any neighbour node overhearing the source node's unsuccessful frame may retransmit it in its own allocated slot. In [116–118], the system is assumed to operate in a stop-and-wait regime with neighbours continuously retransmitting overheard frames until the destination returns an ACK frame. In [119], an error-tolerant cooperative ARQ system is proposed. In this system, each cooperative neighbour is able to encode the frame from the source even if the frame is erroneous, and retransmit it to the destination. The destination is able to decode these "erroneous" frames from multiple neighbours and recover the original frame from the source. In [120], the retransmission successful probability is analysed for a simple cooperative ARQ system, in which the source and a neighbour continuously retransmitting until the destination returns an ACK frame.

However, the work in [115–120] sidesteps the possibility of collision among the cooperative retransmissions. In [115], such collisions cannot occur by virtue of the fixed TDMA

allocation. In [116, 119], it is implicitly assumed that the frame can be recovered from multiple simultaneous retransmissions, thereby requiring a cooperative diversity-enabled receiver, such as MIMO transceivers. In [117, 118], it is assumed that a separated "sub-channel" at physical layer is assigned to each cooperative neighbour, so that multiple simultaneous retransmissions will not collide. In [120], the use of the space-time coding technology is assumed. On this point, they still tackle wireless cooperative retransmission from the physical layer perspective rather than MAC layer perspective, and the collision issue form the MAC layer perspective is not considered.

Additionally, the work in [121, 122] should be noted. In addition to using the aforementioned opportunistic forwarding approach, the authors of [121, 122] also suggest a contention scheme using a transmission probability to avoid the collision among multiple transmitting nodes. If a neighbour receives the frame from the source, and it may retransmit this frame to the destination or remain silent according to the predefined transmission probability. The value of the transmission probability should be optimally set such that the successful probability of the retransmission is maximised. Although this contention scheme is similar to the retransmission strategies later presented in this thesis, it is based on the assumption that the system is aware about the number of neighbours that receive the frame from the source. Such awareness must be obtained with the coordination from the neighbours. For example, neighbour must return an ACK frame to indicate their reception of the frame.

6.2 Summary

The literature review has demonstrated that the majority of the existing studies in this area tackle wireless cooperative retransmission from the physical layer perspective, where the collision issue can be ignored. Meanwhile some existing studies consider the collision issue from the MAC layer perspective, but their opportunistic forwarding approach may cause excessive coordination overhead. Only limited existing studies focus on the approach of uncoordinated distributed wireless cooperative retransmission, but they still ignore the collision issue by using some complex physical layer technologies. There is a lack of work on uncoordinated distributed wireless cooperative retransmission that carefully considers the collision issue from the MAC layer perspective. Such work can be of more practical significance, as it can be easily implemented by simple transceivers with single antenna and single decoder,

such as popular IEEE 802.11 adaptors.

Chapter 7

Uncoordinated Wireless Cooperative Retransmission Strategies

The literature review in the previous chapter has shown that only a limited number of existing studies have attempted to consider uncoordinated distributed wireless cooperative retransmission, and they still tackle it from a physical layer perspective rather than a MAC layer perspective. However, the cooperative methods considering from the MAC layer perspective can be easily implemented with simple transceivers of a traditional single-antenna/single-user decoder (such as popular IEEE 802.11 adaptors). Therefore, it is of practical significance to consider wireless cooperative retransmission from the MAC layer perspective. In this chapter, two uncoordinated wireless cooperative retransmission strategies are proposed, and the collision issue from the MAC layer perspective is considered.

First, both strategies are analysed with a simple memoryless channel model for the purpose of elementary investigation. Second, based on the elementary investigation results, the proposed strategies are analysed with a more realistic two-state Markov fading channel model [123–130].

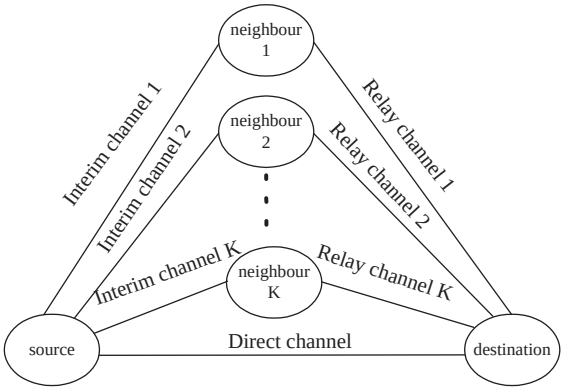


Figure 7.1: The wireless co-operative network considered in this work.

7.1 System Model and Assumptions

The system being considered is a wireless network consisting of a source node, a destination node, and a fixed number K of cooperative neighbour nodes in their vicinity (Fig 7.1). A frame transmitted by the source may or may not arrive at the destination successfully over the direct channel, and may also be overheard by some of the neighbour nodes via the interim channels. Only the intended destination returns an acknowledgment (ACK) upon successful reception; neither the source nor any neighbour can tell which other neighbours, if any, have obtained a copy of the frame. A *slot* is defined to be the duration of a frame transmission plus the time of waiting for an ACK; it is assumed that slots are of fixed duration and synchronised among the nodes. In the subsequent slots after a frame's first transmission, any node possessing a copy of the frame may decide to make a cooperative retransmission. For simplicity, it is assumed that the feedback (ACK) channels to the source and all neighbours are error-free, and that only one frame is active at a time, that is, no other frames are handled, since a frame is first transmitted by the source until it is eventually acknowledged by the destination (or, possibly, dropped after reaching a maximum number of retransmission attempts). For successful reception, the destination must receive exactly one collision-free transmission in a slot. Thus, our system model is similar to the node-cooperative stop-and-wait (NCSW) setting of [116], with the notable difference in our case being the possibility of collision if multiple retransmissions occur at the same time.

To simplify later analysis, we assume that any channel can be in one of two states: either "on" (no fade), in which the transmitted signal arrives with sufficient power to be decoded

without error (barring a collision), or "off" (deep fade), in which a transmitted signal does not arrive at all. A collision occurs if and only if a node receives two or more transmissions simultaneously with the respective channels being "on" (in other words, a transmission over an "off" channel does not cause any interference).

Based on the assumption of the on-off channel states, two channels models are used. The first channel model is a simple memoryless channel model which is used for the elementary performance investigation of the proposed retransmission strategies. In this channel model, whether a channel is "on" or "off" at any slot simply depends on a constant channel probability, denoted as P_{xx}. Here "P_{xx}" can be P_{sd}, P_{sn}, and P_{nd}. They are used to denote the channel probability on direct channel, interim channel, and relay channel, respectively.

The second channel model is a two-state Markov channel fading model (also known as an *order-1* Markov model or *Gilbert* model), which has been shown by numerous studies to provide an adequate description of the bursty frame loss process in practical wireless fading channels [123–130]. In particular, a comprehensive experimental study for typical IEEE 802.11 channels [130] has confirmed that, while more complex models are required for an accurate representation of the *bit-level* error process, a two-state model is quite sufficient at time scales of frames. In this channel model, the transition between "on" and "off" states is not memoryless, and the transition probability from the "off" (*bad*) state to the "on" (*good*) state in every slot (and vice versa) is denoted by P_{bg_sd} (respectively P_{gb_sd}) for the direct channel (between *source* and *destination*); P_{bg_sn}, P_{gb_sn} for any of the interim channels (between *source* and *neighbour*); and P_{bg_nd}, P_{gb_nd} for any of the relay channels (between *neighbour* and *destination*). It is assumed that initially (that is, before the first transmission), the states of all channels are sampled according to their respective *steady-state* probabilities:

$$P_{ss\_xx} \triangleq \frac{P_{bg\_xx}}{P_{bg\_xx} + P_{gb\_xx}}, \tag{7.1}$$

where "xx" $\in \{$"sd","sn","nd"$\}$ is substituted for the corresponding channel type.

Furthermore, in both channel models, the channel states and transitions are assumed to be mutually independent among different node pairs, which is realistic in most practical scenarios where nodes are spaced sufficiently far apart. It should be pointed out that, for reasons of tractability, our analysis assumes a symmetric system, where all neighbours are equivalent *a priori*. This should not be interpreted as a requirement that the channel quality of all neighbours, or their underlying physical characteristics (for example, their distances from

the source and destination), must be identical; the symmetry only means that the neighbours are indistinguishable *a priori* for the purpose of computing the cooperation strategy.

In reality, the transition between the two states does not occur instantly, and an intermediate state exists wherein the received signal power is insufficient for correct decoding, but enough to cause interference with other transmissions. For simplicity, this intermediate state is ignored in this work and it is assumed that the time any channel spends in it is negligible; however, it should be pointed out that our analysis can be extended from a two-state to a three-state model in a straightforward manner.

7.2 Analysis with Memoryless Channel Model

7.2.1 The First Retransmission

The analysis begins by considering the optimal cooperative retransmission strategy for a single time slot with an uncoordinated manner. Thus, the strategy simply boils down to a single number, namely the probability of retransmission for any node that had successfully overheard the frame; this probability is denoted by τ. The optimal value of τ is that can maximises the probability of successful delivery. In order to find it, we first consider the probability distribution of the number of neighbours k that have successfully overheard the frame from the source's original transmission. Since the interim channels are symmetric, this distribution is binomial:

$$P\{k\} = \left(\begin{array}{c} K \\ k \end{array} \right) P_{sn}^{k} \left(1 - P_{sn}\right)^{K-k}. \tag{7.2}$$

For a successful delivery, out of these k nodes, there must be exactly one that both makes a retransmission *and* has a good channel. Consequently,

$$P^{suc} = \sum_{k=1}^{K} P\{k\} \cdot k\tau P_{nd}(1 - \tau P_{nd})^{k-1}, \tag{7.3}$$

which, after a straightforward simplification, becomes

$$P^{suc} = K P_{sn} \tau P_{nd}(1 - P_{sn}\tau P_{nd})^{K-1}. \tag{7.4}$$

Alternatively, P^{suc} can be obtained with another approach. The probability that k out of K neighbours correctly receive and successfully retransmit the frame can be calculated as

$$P^{suc}(k) = \begin{pmatrix} K \\ k \end{pmatrix} (P_{sn}\tau P_{nd})^k (1 - P_{sn}\tau P_{nd})^{K-k}. \tag{7.5}$$

P^{suc} is achieved when $k = 1$ for (7.5), and it means that only one among K neigbhbours correctly receives the frame and successfully transmits it. The result is identical to that in (7.4).

The optimal τ^* is now obtained by equating the first derivative of (7.4) to zero, which yields

$$\tau^* = \frac{1}{KP_{sn}P_{nd}}. \tag{7.6}$$

The above expression for τ^*, of course, is only valid if $\frac{1}{KP_{sn}P_{nd}} \leq 1$. Otherwise, that is, if $P_{sn}P_{nd} < \frac{1}{K}$, the probability of successful delivery (7.4) is monotonically increasing in τ, and its optimum is then achieved with $\tau^* = 1$.

Now the optimal vaule τ^* is assigned back into (7.4), to evaluate the maximum success probability that can be obtained after one cooperative retransmission slot. If $K > \frac{1}{P_{sn}P_{nd}}$, then τ^* is given by (7.6), and

$$P^{suc^*} = \left(1 - \frac{1}{K}\right)^{K-1}. \tag{7.7}$$

Curiously, it can be observed that this expression is *decreasing* in K (it tends to $\frac{1}{e}$ for $K \to \infty$); in other words, having too many neighbours in the cooperation group may, in fact, degrade the performance of cooperative retranmission. It is easily verified that, in the range $1 \leq K < \frac{1}{P_{sn}P_{nd}}$ (such that $\tau^* = 1$), the probability of successful delivery is increasing in K, as intuitively expected. Hence, it may be concluded that the best size of the cooperation group is around $\frac{1}{P_{sn}P_{nd}}$; if the number of neighbour nodes is larger than that, it is better to voluntarily choose a smaller cooperation group (and thereby keep τ^* close to 1), rather than use all the available neighbours with a smaller retransmission probability.[1]

[1] Since $\frac{1}{P_{sn}P_{nd}}$ is, in general, not a whole number, the optimal cooperation group size may be the integer to either side of it.

7.2.2 Subsequent Retransmissions

7.2.2.1 Strategy 1: neighbours repeat attempts with silent source

We now consider the cooperation strategies beyond the first retransmission slot. We first focus on the strategy where, if the first retransmission attempt fails, the neighbours continue to make additional retransmission attempts while the source remains silent. Since the source does not transmit again, the number of neighbours with a copy of the frame k, and its distribution, remains unchanged from the first slot; consequently, the optimal τ^* that maximises the probability of exactly one neighbour transmitting with a good relay channel is the same as for the first slot, derived in Section 7.2.1.[2] Since it has been assumed that the channel states are independent between slots, the retransmission attempts will form a Bernoulli process with a success probability of

$$P_{k,\tau^*}^{suc} \triangleq k\tau^* P_{nd}(1-\tau^* P_{nd})^{k-1} \tag{7.8}$$

in each slot.

However, due to the possibility of the case $k = 0$ (that is, all interim channels were "off" during the original transmission and no neighbour overheard the frame, in which case $P_{k,\tau^*}^{suc} = 0$), the above strategy is not guaranteed to succeed after a finite number of attempts. Therefore, we define a maximum number of cooperative attempts before the retransmission process restarts again with the original source node, and denote it by $m-1$. Thus, we are considering a periodic strategy with a period of m slots, where each period starts with a transmission by the source, followed by $m-1$ cooperative retransmissions by the neighbours. The choice of m reflects a tradeoff between the time wasted on cooperative attempts in the case of $k = 0$ and that wasted on a source retransmission otherwise. Generally, the better the interim channels (P_{sn}) and the worse the relay channels (P_{nd}), the higher the optimal value of m.

To find the optimal m analytically, the following recursive expression is written for the ex-

[2] In the analysis of Strategy 1, the possibility of overhearing the frame from another neighbour's transmission is ignored; this possibility is considered later in the discussion of Strategy 2.

pected number of slots until success, E:

$$E = P_{sd} \cdot 1 + (1 - P_{sd}) \sum_{k=0}^{K} P\{k\} \cdot \left[\sum_{i=1}^{m-1} P_{k,\tau^*}^{suc} \left(1 - P_{k,\tau^*}^{suc}\right)^{i-1} \cdot (i+1) + \left(1 - P_{k,\tau^*}^{suc}\right)^{m-1} \cdot (E+m) \right]. \quad (7.9)$$

This expression accounts for the probability of success in the direct transmission over the primary channel, or after $i \in \{1, \ldots, m-1\}$ attempts of cooperative retransmission in the first period, or, if all $m-1$ such attempts prove unsuccessful, the entire first period of m slots is wasted and the expected number of additional slots until success is the same as originally. Using the geometric-sum formula $\sum_{i=1}^{N} i r^{i-1} = \frac{1-r^{N+1}}{(1-r)^2} - \frac{(N+1)r^N}{1-r}$ and grouping together the coefficients of E, the following formula arrives:

$$E = \frac{P_{sd} + (1 - P_{sd}) \sum_{k=0}^{K} P\{k\} \left[\frac{1-\left(1-P_{k,\tau^*}^{suc}\right)^{m-1}}{P_{k,\tau^*}^{suc}} + 1 \right]}{1 - (1 - P_{sd}) \sum_{k=0}^{K} P\{k\} \left(1 - P_{k,\tau^*}^{suc}\right)^{m-1}}, \quad (7.10)$$

where $P\{k\}$ is given by (7.2), and the expression in brackets in the numerator for $k = 0$ (that is, $P_{k,\tau^*}^{suc} = 0$) should be taken as equal to m. In the subsequent performance evaluation in Section 7.2.3, (7.10) can be used to manually find the optimal period m for this strategy, for any instance of P_{sd}, P_{sn}, P_{nd}, and K.

7.2.2.2 Strategy 2: simultaneous source+neighbour transmissions

In the strategy described in the previous subsection, the parameter m reflected the tradeoff between extending the chance to retransmissions by the cooperative neighbours (which normally have better channels to the receiver), and wasting the time in case no neighbours had overheard the frame (which, as a direct consequence of the fact that the sender is silent, is not rectified during the entire $m-1$ slots). In order to overcome this disadvantage, we now describe a heuristic strategy in which the transmission probabilities of both the sender (τ_s) and the neighbours (τ_n) are allowed to be greater than zero simultaneously. As a result, the number of neighbours with a copy of the frame continues to grow over time (up to K).

The heuristic is based on a greedy approach that attempts to maximise the probability of successful reception in each slot in turn. To assist in the calculation, we maintain and update

the distribution $P\{k\}$, that is, the number of neighbours with a copy of the frame so far. Thus, in every slot i, the following calculation steps are made:

1. the optimal τ_s and τ_n are solved numerically to maximise the probability of success in this slot, given by the expression

$$P^{suc} = \sum_{k=0}^{K} P_i\{k\} \cdot P^{suc}_{k,\tau_s,\tau_n}, \tag{7.11}$$

   where $P_i\{k\}$ denotes the distribution of k *before* the start of slot i, and

$$P^{suc}_{k,\tau_s,\tau_n} \triangleq (1-\tau_s P_{sd}) k \tau_n P_{nd} (1-\tau_n P_{nd})^{k-1} + \tau_s P_{sd} (1-\tau_n P_{nd})^k \tag{7.12}$$

2. assuming that the slot nevertheless results in a failure, the distribution $P_i\{k\}$ is revised *a posteriori* using Bayes' formula, as follows:

$$P^{rev}_i\{k\} = \frac{P_i\{k\}(1-P^{suc}_{k,\tau^*_s,\tau^*_n})}{\sum_{k'=0}^{K} P_i\{k'\}(1-P^{suc}_{k',\tau^*_s,\tau^*_n})}, \tag{7.13}$$

   where τ^*_s, τ^*_n are the optimal values obtained in step 1;

3. finally, $P_i\{k\}$ is updated to account for the new neighbours that overhear the frame from the source in this slot, yielding

$$P_{i+1}\{k\} = (1-\tau^*_s) P^{rev}_i\{k\} + \tau^*_s \sum_{k'=0}^{k} P^{rev}_i\{k'\} \cdot \binom{K-k'}{k-k'} P^{k-k'}_{sn} (1-P_{sn})^{K-k+k'}. \tag{7.14}$$

   Note that expression (7.14) considers only the possibility of overhearing a transmission from the source, not from another neighbour. One can also consider the case in which a channel between two neighbours can be "on" with a probability $P_{nn} > 0$; then, a new neighbour may overhear the frame from either the source or another neighbour, provided there is no collision. The extension of (7.14) to this case is straightforward and omitted here.

Example: Consider a network with only $K = 1$ cooperating neighbour, $P_{sd} = 0.5$, $P_{sn} = 0.99$, $P_{nd} = 1$. If the first transmission by the source fails, the neighbour has a probability of $P_2\{k = 1\} = 0.99$ to have the frame at the start of the second slot. Therefore, clearly, the

optimal uncoordinated strategy in this slot is to allow it to transmit the frame uninterrupted ($\tau_s^* = 0, \tau_n^* = 1$). Indeed, a simultaneous transmission by the source would interfere with the neighbour's one with a probability of $0.99 \cdot 0.5$, and would only be helpful with a probability of $0.01 \cdot 0.5$. However, if this strategy is applied and still fails, then $P_2^{rev}\{k = 1\} = P_3\{k = 1\} = 0$ can be obtained, as failure can only occur if the neighbour does not have the frame after all. Accordingly, the optimal strategy in the third slot is $\tau_s^* = 1$ (the strategy of the neighbour is immaterial).

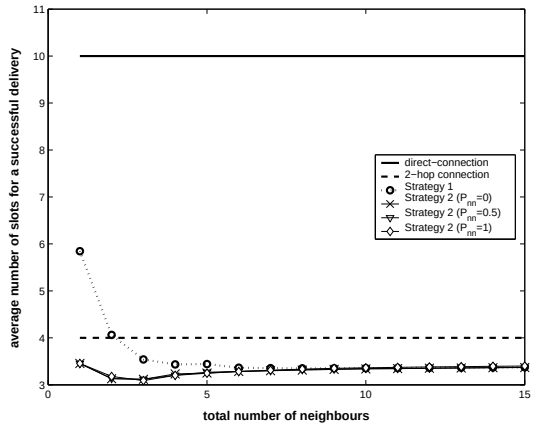
However, in the same system but with $P_{sd} = 0$, the optimal uncoordinated strategy trivially becomes $\tau_s^* = 1, \tau_n^* = 1$ in every slot. The source does not have a channel to the receiver and therefore cannot interfere with the neighbour; meanwhile, the simultaneous transmission by the source saves time if the neighbour still has not heard the frame.

7.2.3 Evaluation of the Retransmission Strategies

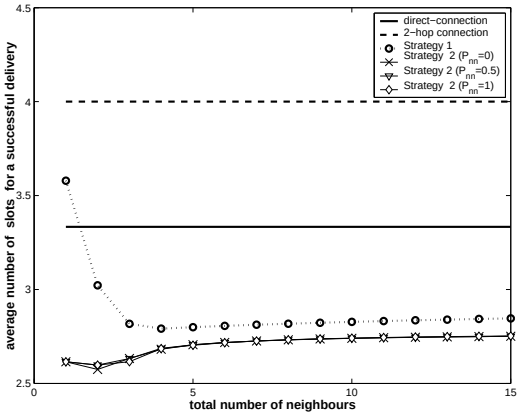
In this section, the performance of the proposed strategies is investigated numerically, under various combinations of channel quality for the direct, interim and relay channels. The numerical calculation tool to obtain the optimal transmission probabilities is $fmincon$ function from the optimisation toolbox in MATLAB [131]. In each scenario, the impact of the number of cooperative neighbours is examined on the expected latency, and it is compared with traditional one-hop and two-hop routing as well.

The evaluation begins with an arguably typical cooperative retransmission scenario: a poor primary channel, with better interim and relay channels. Accordingly, the performance of our strategies is demonstrated with $P_{sd} = 0.1$, $P_{sn} = P_{nd} = 0.5$. In this scenario, retransmissions over the direct hop require on average $\frac{1}{P_{sd}} = 10$ slots until success, while two-hop routing over any of the neighbours (with retransmission in each hop) achieves an average latency of $\frac{1}{P_{sn}} + \frac{1}{P_{nd}} = 4$ slots.

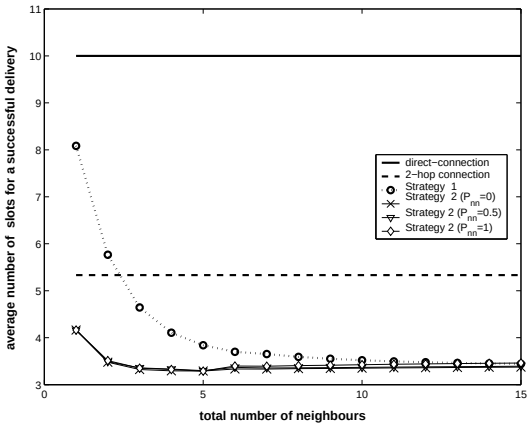
The results are shown in Fig 7.2. The performance of Strategy 1 is obtained with the optimal period m (manually found using expression (7.10) for each K). In addition, the performance of Strategy 2 is tested under three values of P_{nn}, that is, the neighbour-to-neighbour channel quality (see the comment following expression (7.14)). These range from $P_{nn} = 0$ (neighbours cannot overhear each other at all), to $P_{nn} = 1$ (neighbours always overhear



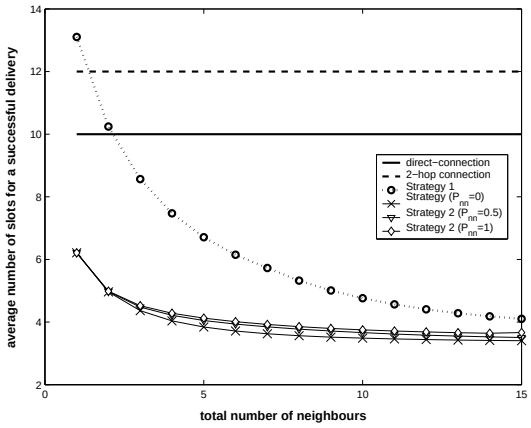
(a) $P_{sd} = 0.1, P_{sn} = 0.5, P_{nd} = 0.5$.



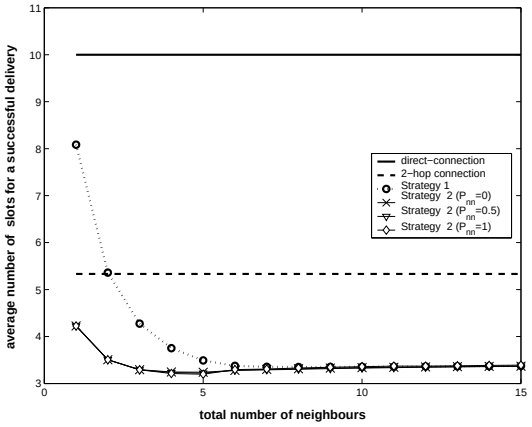
(b) $P_{sd} = 0.3, P_{sn} = 0.5, P_{nd} = 0.5$.



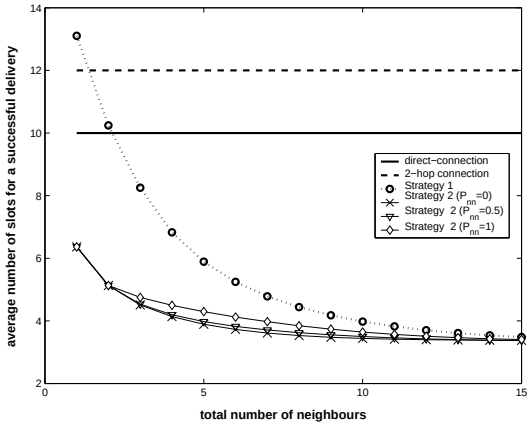
(c) $P_{sd} = 0.1, P_{sn} = 0.3, P_{nd} = 0.5$.



(d) $P_{sd} = 0.1, P_{sn} = 0.1, P_{nd} = 0.5$.



(e) $P_{sd} = 0.1, P_{sn} = 0.5, P_{nd} = 0.3$.



(f) $P_{sd} = 0.1, P_{sn} = 0.5, P_{nd} = 0.1$.

Figure 7.2: Numerical results.

transmissions from their peers). As Fig 7.2(a) clearly shows, even though the performance of Strategy 1 improves as the number of cooperative nodes increases and it attains a lower expected latency than two-hop routing for $K \geq 3$, Strategy 2 outperforms it consistently. It is also observed that the impact of the neighbour-to-neighbour channel quality on the overall performance is negligible; intuitively, the ability of a neighbour to overhear the frame from other neighbours (and not just from the source) is counter-balanced by the additional collisions that occur when the source and another neighbour transmit together.

Fig 7.2(b) presents the results for the case in which the quality of the direct channel is improved to $P_{sd} = 0.3$. Now, the direct channel is better than the two-hop route, requiring only $\frac{1}{0.3} \approx 3.33 < 4$ slots. Nevertheless, our cooperative retransmission strategies are still able to significantly improve the expected latency, mainly because they take advantage of the better relay channel when the original transmission over the direct channel fails. Of course, as the direct channel becomes better, this effect diminishes; once $P_{sd} = 0.5$ on par with the interim and relay channels, the optimal uncoordinated strategy trivially reduces to retransmission over the direct channel, ignoring the neighbours.

In Fig 7.2(c) and Fig 7.2(d), the effect of reducing the interim channel quality to $P_{sn} = 0.3$ and $P_{sn} = 0.1$ is examined, respectively; Fig 7.2(e) and Fig 7.2(f) do the same for the relay channel. As expected, the performance of all strategies deteriorates monotonically with the channel quality; nevertheless, there is merit in using cooperative retransmission as long as either the interim or relay channel is better than the direct one. Interestingly, these figures also demonstrate that the same effect that has been observed in the analysis of the first slot — namely, that the optimal number of cooperating neighbours increases as the quality of the channels deteriorates — holds for the overall performance of the greedy heuristic strategy as well.

The observations on Fig 7.2 show that Strategy 1 improves with increasing neighbour population, that Strategy 2 consistently outperforms it (and is never worse than one-hop or two-hop routing), and that the neighbour-to-neighbour channel quality has a very minor impact on the performance — occurring consistently throughout our evaluations.

7.3 Analysis with a Two-state Markov Fading Channel Model

The aforementioned analysis with the memoryless channel model has demonstrated the superior performance of Strategy 2. That is, the strategy in which both the source and the neighbours participate in the retransmission attempts. In this section we will investigate its performance under a more realistic two-state Markov Fading Channel Model.

7.3.1 Definitions and Preliminary Analysis

The analysis begins by defining the notion of *system state*. Clearly, the system state should include all the quantities that impact its future dynamics, and ultimately the strategy performance. In our case, these are: for each of the K neighbours, a binary value indicating whether it has already got a copy of the frame; and for each of the $2K+1$ channels, a binary value indicating whether it is "on" or "off". This implies that, in principle, the system state consists of a vector of $3K+1$ binary elements.

Fortunately, the assumption of symmetry among the neighbours allows the relevant state information to be considerably reduced. Accordingly, instead of tracking the system state to the granularity of every individual neighbour and channel, our main idea is to focus on the following probability distributions, henceforth referred to as the *state distributions*:

- $P_{[i]}\{k\}$ is the distribution of the number of neighbours, k, that have overheard a copy of the frame before slot i;
- $\underset{rel_{[i]}}{P}\{r, d|k\}$, where $d \in \{0, 1\}$, is the conditional probability, given that k neighbours have the frame, that exactly r out of them have relay channels in the "on" state *and* the direct channel is "off" ($d = 0$) or "on" ($d = 1$), respectively, in slot i;
- $\underset{int_{[i]}}{P}\{c|k\}$ is the conditional probability, given that k neighbours have the frame, that c out of the *remaining* $K - k$ neighbours have an interim channel in the "on" state in slot i.

Indeed, we are not interested in the states of interim channels of neighbours that have already received the frame, as they do not impact on the future system dynamics in any way. Simi-

larly, the state of the relay channel of any neighbour not yet in possession of the frame does not impact on anything else in the system; hence, at the moment the neighbour eventually overhears the frame, its relay channel state is still governed by the steady-state probability. Therefore, it may be argued that the state distributions defined above are sufficient to capture all the relevant information for the future system dynamics. Thus, the probability of the system in slot i to have k neighbour nodes with the frame, r of them to have relay channels that are "on", c of the *remaining* $K-k$ neighbours to have interim channels that are "on", and the direct channel to be in state d, is $P_{[i]}\{k\}\underset{rel_{[i]}}{P}\{r,d|k\}\underset{int_{[i]}}{P}\{c|k\}$. Henceforth the tuple (k,r,c,d) is referred to as the *system state vector*. As will become apparent below, the decoupling of the system state into the above separate distribution functions is undertaken to facilitate the calculations involved.

Now the analysis proceeds to derive the target function of the strategy optimisation problem. To that end, first the conditional success probability in a generic slot is computed, provided that the system is in a particular state (k,r,c,d), and the strategy values (that is, retransmission probabilities of the source and neighbours) are τ_s, τ_n in that slot. To be successful, the slot must have one and only one transmission by a node (source or neighbour) whose channel to the destination is "on"; thus,

$$P^{suc}(\tau_s,\tau_n|k,r,d) \triangleq \begin{cases} r\tau_n(1-\tau_n)^{r-1} & d=0 \\ (1-\tau_s)r\tau_n(1-\tau_n)^{r-1}+ & \\ \quad \tau_s(1-\tau_n)^r & d=1 \end{cases} \tag{7.15}$$

It is observed that the success probability in expression (7.15) is unaffected by the state of interim channels, which is why it is denoted $P^{suc}(\tau_s,\tau_n|k,r,d)$ rather than $P^{suc}(\tau_s,\tau_n|k,r,c,d)$. Consequently, the total probability of success in slot i is

$$P^{suc}_{[i]} \triangleq \sum_{k=0}^{K} \sum_{\substack{0\le r\le k \\ d\in\{0,1\}}} P^{suc}(\tau_{s[i]},\tau_{n[i]}|k,r,d)\cdot P_{[i]}\{k\}\underset{rel_{[i]}}{P}\{r,d|k\}, \tag{7.16}$$

and, finally, the expected frame latency (that is, the optimisation target) is

$$\sum_{i=1}^{\infty} i\cdot P^{suc}_{[i]}\cdot\prod_{j=1}^{i-1}\left(1-P^{suc}_{[j]}\right). \tag{7.17}$$

The deceptively simple form of expression (7.17) may lead to the wrong conclusion that, in order to minimise the expected frame latency, one must simply find the values of $\tau_{s[i]}$, $\tau_{n[i]}$

that maximise (7.16) for every slot i separately. Generally, this may not yield the optimal uncoordinated strategy, since it ignores the impact of the strategy choice on the system state distributions in future slots. For instance, the strategy of the source in any slot affects the number of neighbours that overhear the frame for the first time in that slot, and, consequently, the future distribution of k. Due to this dependency, a straightforward minimisation of (7.17) is unfeasible. In the following, a heuristic solution approach is described and analysed, based on an iterative greedy maximisation of (7.16) for each slot in turn, ignoring the impact of the strategy choice on the future dynamics.

7.3.2 The Proposed Heuristic Solution Method

The proposed heuristic solution approach operates in each slot iteratively. First, the state distributions are initialised before the first slot as follows:

$$P_{[1]}\{k=0\}=1, \quad P_{[1]}\{k>0\}=0;$$
$$\underset{int_{[1]}}{P}\{c|k=0\}=\binom{K}{c}(P_{ss\_sn})^c(1-P_{ss\_sn})^{K-c};$$
$$\underset{rel_{[1]}}{P}\{r=0,d=1|k=0\}=P_{ss\_sd},$$
$$\underset{rel_{[1]}}{P}\{r=0,d=0|k=0\}=1-P_{ss\_sd}.$$

Indeed, before the first slot, the number of neighbours with the frame is obviously zero, and the number of interim channels that are "on" during the first transmission is distributed binomially, with a parameter that is the interim channels' steady-state probability. Since $k=0$ with probability 1 during the first slot, it is not necessary to initialise the interim and relay state distributions for other possible values of k.

After the initialisation, the solution method proceeds for each slot i (starting from $i=1$) iteratively. Thus, it is assumed that the system state distributions for slot i are given; the calculations for that slot then yield the strategy elements $\tau_{s[i]}$, $\tau_{n[i]}$, as well as the state distributions for slot $i+1$. More specifically, the following calculation steps are performed in slot i:

1. the optimal $\tau^*_{s[i]}$ and $\tau^*_{n[i]}$ are solved numerically to maximise expression (7.16) (except for slot $i=1$, where $\tau^*_{s[1]}=1$ and $\tau^*_{n[0]}=0$ are required by definition);

2. assuming that the slot nevertheless results in a failure, the system state distributions for slot i are revised *a posteriori* (this is denoted by attaching a superscript of 'ap' to the respective distributions);

3. finally, the state distributions for slot $i+1$ are computed, accounting for the new neighbours that overhear the frame in slot i and the transitions in the channel states.

The need for an *a posteriori* revision in calculation step 2, which may not be readily apparent, is explained by the following example.

Example: Consider a network with only $K = 1$ cooperating neighbour, $P_{bg\_sd} = P_{gb\_sd} = 0.5$, $P_{bg\_sn} = 0.99$, $P_{gb\_sn} = 0.01$, $P_{bg\_nd} = 1$, $P_{gb\_nd} = 0$, i.e. a direct channel that is "on" half the time, interim channel "on" 99% of the time, and relay channel that is always perfect. Assuming the first transmission by the source fails, the neighbour has got the frame at the start of the second slot with a probability of $P_{[2]}\{k=1\} = 0.99$. Therefore, clearly, the optimal uncoordinated strategy in the second slot is to allow the neighbour to transmit the frame uninterrupted ($\tau^*_{s[2]} = 0$, $\tau^*_{n[2]} = 1$); indeed, a simultaneous retransmission by the source would far more likely cause a collision than result in a delivery. However, if this strategy is applied and the slot still ends up in a failure, then the distribution must be revised *a posteriori* to $P^{ap}_{[2]}\{k=1\} = 0$, as failure can only occur if the neighbour did not overhear the frame after all. This implies $P_{[3]} = 0$ as well (as the source did not transmit in slot 2), and, therefore, the optimal uncoordinated strategy in the third slot is $\tau^*_{s[3]} = 1$ (the strategy of the neighbour is immaterial). Clearly, ignoring the *a posteriori* revision step and attempting to calculate the strategy in slot 3 independently of the outcome of slot 2 would result in a wasted slot, for retransmission by a neighbour that cannot possibly have the frame if the slot is reached at all.

Similarly, consider the case of $K = 1$, $P_{bg\_sd} = P_{gb\_sd} = 0.5$, $P_{bg\_sn} = 1$, $P_{gb\_sn} = 0$, $P_{bg\_nd} = 0.09$, $P_{gb\_nd} = 0.01$. Here, the interim channel is perfect, while the relay channel's steady-state distribution is to be "on" 90% of the time. Now, the probability of the neighbour having the frame in the second slot is $P_{[2]}\{k=1\} = 1$; also, $\underset{rel_{[2]}}{P}\{r=1, d=0|k=1\} = \underset{rel_{[2]}}{P}\{r=1, d=1|k=1\} = 0.45$ while $\underset{rel_{[2]}}{P}\{r=0, d=0|k=1\} = \underset{rel_{[2]}}{P}\{r=0, d=1|k=1\} = 0.05$. Hence, again, the best strategy in this slot is a retransmission by the neighbour only. In this case, however, a failure in slot 2 will not change the distribution of $P_{[2]}\{k\}$ *a posteriori*, since it is known with certainty that the neighbour has the frame. Rather,

a failure implies that the relay channel must have been "off" after all, leading to a revision of the relay state distribution to $\underset{rel_{[2]}}{P^{ap}}\{r=1,d=0|k=1\} = \underset{rel_{[2]}}{P^{ap}}\{r=1,d=1|k=1\} = 0$, $\underset{rel_{[2]}}{P^{ap}}\{r=0,d=0|k=1\} = \underset{rel_{[2]}}{P^{ap}}\{r=0,d=1|k=1\} = 0.5$. After accounting for a single Markov transition step, the relay channel has a probability of only 0.09 to be "on" by slot 3, and it can be verified that the optimal value setting in slot 3 is $\tau^*_{s[3]} = 1$, $\tau^*_{n[3]} = 0$.

Finally, it should be mentioned that if the direct channel parameters are set at $P_{bg\_sd} = 0$, $P_{gb\_sd} = 1$, then the optimal strategy trivially becomes $\tau^*_s = 1, \tau^*_n = 1$ in every slot, for any setting of the interim and relay channel parameters. Indeed, if the direct channel is always "off", that is, there is no channel between the source and destination, then a simultaneous transmission by the source can never interfere with the neighbour, yet it may save time if the neighbour has not yet heard the frame. This case shows that, in general, the optimal value setting may specify τ_s and τ_n that are both greater than zero in the same slot. In fact, the earlier work on memoryless channel model in the previous section has shown that, even for the special case of memoryless channels, the best performance that can be achieved if the source and the cooperative neighbours avoid retransmitting simultaneously is strongly suboptimal.

Now the analysis proceeds to elaborate the details of the calculation steps in slot i, outlined above. The implementation of step 1 is not considered any further; it is beyond the scope of the current work to suggest a specific solution method for the respective maximisation problem. It is merely pointed out that, since expression (7.16) is in general non-concave, care must be taken to avoid choosing $\tau^*_{s[i]}$, $\tau^*_{n[i]}$, which only attain a local maximum.

In step 2, the revision of the state distribution *a posteriori* is achieved using Bayes' formula, that is, by scaling the probability of every possible state by the likelihood that the strategy (τ^*_s, τ^*_n) would have failed in that state.[3] For convenience, we define, in a similar fashion to (7.15),

$$P^{suc}(\tau_s, \tau_n|k) = \sum_{\substack{0\le r\le k \\ d\in\{0,1\}}} P^{suc}(\tau_s, \tau_n|k, r, d) \tag{7.18}$$

[3]Since there is no possible ambiguity, henceforth τ^*_s and τ^*_n are used to denote the strategy chosen in step 1 for slot i, without explicitly mentioning the index i.

and obtain

$$P^{ap}_{[i]}\{k\} = \frac{P_{[i]}\{k\}(1 - P^{suc}(\tau^*_s, \tau^*_n|k))}{\sum_{k'=0}^{K} P_{[i]}\{k'\}(1 - P^{suc}(\tau^*_s, \tau^*_n|k'))} \tag{7.19}$$

$$P^{ap}_{rel_{[i]}}\{r, d|k\} = \frac{P_{rel_{[i]}}\{r, d|k\}(1 - P^{suc}(\tau^*_s, \tau^*_n|k, r, d))}{\sum_{\substack{0\le r\le k \\ d\in\{0,1\}}} (1 - P^{suc}(\tau^*_s, \tau^*_n|k, r, d))} \tag{7.20}$$

Clearly, there is no *a posteriori* revision of the interim channel state distribution, as it has no impact on the success probability in the slot.

Remark: The observant reader may notice that expressions (7.19)–(7.20) do not yield a perfectly precise *a posteriori* system state distribution, because, strictly speaking, the revision should be performed on the entire history of the system, not just the distributions of the current slot. Indeed, a failure in slot i impacts on the *a posteriori* probabilities of transmission by the source and neighbours in previous slots as well, and through that, indirectly, the state distribution in the current slot – a second-order effect that is ignored in (7.19)–(7.20). As will be shown in Section 7.3.3 through comparison to simulation results for a wide variety of scenarios, the approximation introduced by ignoring the above effect is negligible in practice.

Finally, the analysis proceeds to consider calculation step 3, namely, finding the system state distributions that are in effect at the beginning of slot $i + 1$. This is the least straightforward step, due to the various interactions between the number of new neighbours overhearing the frame in slot i and the channel state transitions, which require a detailed and careful consideration. For simplicity, this analysis begins by assuming that neighbours are unable to overhear their peers' transmissions, and can only overhear the frame from the original source. Subsequently, the impact of this assumption is considered and it is partially alleviated. Throughout this subsection, $\hat{k}$, $\hat{r}$, $\hat{c}$, and $\hat{d}$ are used to denote the system state in slot $i + 1$, reserving k, r, c, and d to denote the state variables in slot i.

7.3.2.1 The distribution $P_{[i+1]}\{\hat{k}\}$

A new neighbour will overhear the frame in slot i if and only if the source has transmitted in that slot, and the corresponding interim channel is "on". Therefore, the probability of the

system to have $\hat{k}$ frame copies in slot $i+1$ if it had k of them in slot i can be defined:

$$\Pi_{[i]}\{k,\hat{k}\} = \begin{cases} \tau_s^* \cdot \underset{int_{[i]}}{P}\{\hat{k}-k|k\} & k < \hat{k} \\ \tau_s^* \cdot \underset{int_{[i]}}{P}\{0|k\} + (1-\tau_s^*) & k = \hat{k} < K \end{cases} \tag{7.21}$$

and, consequently,

$$P_{[i+1]}\{\hat{k}\} = \sum_{k=0}^{\hat{k}} P_{[i]}^{ap}\{k\} \cdot \Pi_{[i]}\{k,\hat{k}\}. \tag{7.22}$$

7.3.2.2 The distribution $\underset{rel_{[i+1]}}{P}\{\hat{r},\hat{d}|\hat{k}\}$

To calculate the relay channel state distribution in slot $i+1$, we distinguish between the k "old" channels corresponding to nodes that already had the frame in slot i (whose state distribution is given by $\underset{rel_{[i]}}{P^{ap}}\{r,d|k\}$), and the $\hat{k}-k$ "new" channels of nodes that obtained the frame copy in slot i for the first time. As the state of these "new" channels is independent of the system's history so far, they are still governed by their steady-state distribution; thus, the number thereof that are "on" in slot $i+1$ will be distributed binomially with a parameter of P_{ss_nd}.

To capture the state transitions in the "old" channels, an auxiliary function $\underset{old}{\Pi}\{r,r'|k\}$ is defined, which is the probability to have r' relay channels (out of the "old" k) in the "on" state after that number was r in the previous slot. This requires some j out of the r channels to remain "on", plus $r'-j$ additional channels to have a transition from "off" to "on". Thus,

$$\underset{old}{\Pi}\{r,r'|k\} = \sum_{j=\max[0,r'-(k-r)]}^{\min(r,r')} \binom{r}{j}(1-P_{gb\_nd})^j(P_{gb\_nd})^{r-j} \cdot \binom{k-r}{r'-j}(P_{bg\_nd})^{r'-j}(1-P_{bg\_nd})^{k-r-(r'-j)} \tag{7.23}$$

To combine this with the "new" channels, another auxiliary function $\underset{rel}{\Pi}\{r,\hat{r}|k,\hat{k}\}$ is defined, which is the probability to have $\hat{r}$ relay channels (out of $\hat{k}$) in the "on" state after that number was r out of k in the previous slot:

$$\underset{rel}{\Pi}\{r,\hat{r}|k,\hat{k}\} = \sum_{r'=\max[0,\hat{r}-(\hat{k}-k)]}^{\min(\hat{r},k)} \underset{old}{\Pi}\{r,r'|k\}\binom{\hat{k}-k}{\hat{r}-r'} \cdot (P_{ss\_nd})^{\hat{r}-r'}(1-P_{ss\_nd})^{\hat{k}-k-(\hat{r}-r')}. \tag{7.24}$$

Expression (7.24) is a conditional probability, given that the number of frame copies in slot i was k. Summing the total probability and taking into account the state transition of the direct channel, we finally obtain

$$\underset{rel_{[i+1]}}{P}\{\hat{r},0|\hat{k}\} = \sum_{k=0}^{\hat{k}} \left(\frac{\Pi_{[i]}\{k,\hat{k}\}}{\sum_{k'=0}^{\hat{k}} \Pi_{[i]}\{k',\hat{k}\}} \right) \cdot \sum_{r=0}^{k} \underset{rel}{\Pi}\{r,\hat{r}|k,\hat{k}\} \cdot \left[(1-P_{bg\_sd}) \underset{rel_{[i]}}{P^{ap}}\{r,0|k\} + P_{gb\_sd} \underset{rel_{[i]}}{P^{ap}}\{r,1|k\} \right] \quad (7.25)$$

and

$$\underset{rel_{[i+1]}}{P}\{\hat{r},1|\hat{k}\} = \sum_{k=0}^{\hat{k}} \left(\frac{\Pi_{[i]}\{k,\hat{k}\}}{\sum_{k'=0}^{\hat{k}} \Pi_{[i]}\{k',\hat{k}\}} \right) \cdot \sum_{r=0}^{k} \underset{rel}{\Pi}\{r,\hat{r}|k,\hat{k}\} \cdot \left[P_{bg\_sd} \underset{rel_{[i]}}{P^{ap}}\{r,0|k\} + (1-P_{gb\_sd}) \underset{rel_{[i]}}{P^{ap}}\{r,1|k\} \right]. \quad (7.26)$$

7.3.2.3 The distribution $\underset{int_{[i+1]}}{P}\{\hat{c}|\hat{k}\}$

From the assumption that neighbours do not overhear each other's transmissions, it follows that if the source transmitted in slot i, all interim channels of neighbours that do not have the frame by the end of that slot must be "off"; consequently, in slot $i+1$ each such channel has a probability of P_{bg_sn} to be "on", and their total number is distributed binomially. On the other hand, if the source was silent, the number of neighbours without a frame copy does not change, and the distribution of their interim channel states makes a single Markov transition. Accordingly, an auxiliary function $\underset{int}{\Pi}\{c,\hat{c}|\hat{k}\}$ is defined, which is the probability of $\hat{c}$ interim channels (out of $K-\hat{k}$) to be "on" after that number was c in the previous slot (in a similar manner to (7.23)):

$$\underset{int}{\Pi}\{c,\hat{c}|\hat{k}\} = \sum_{j=\max[0,\hat{c}-(K-\hat{k}-c)]}^{\min(c,\hat{c})} \binom{c}{j} (1-P_{gb\_sn})^{j} (P_{gb\_sn})^{c-j} \cdot \binom{K-\hat{k}-c}{\hat{c}-j} (P_{bg\_sn})^{\hat{c}-j} (1-P_{bg\_sn})^{K-\hat{k}-c-(\hat{c}-j)} \quad (7.27)$$

With the help of this auxiliary function, we obtain

$$\underset{int_{[i+1]}}{P}\{\hat{c}|\hat{k}\} = \left[\tau_s^* \sum_{k=0}^{\hat{k}} P_{[i]}^{ap}\{k\} \underset{int_{[i]}}{P}\{\hat{k}-k|k\} \underset{int}{\Pi}\{0,\hat{c}|\hat{k}\} + \right.$$
$$\left. (1-\tau_s^*) P_{[i]}^{ap}\{\hat{k}\} \sum_{c=0}^{\hat{k}} \underset{int_{[i]}}{P}\{c|\hat{k}\} \underset{int}{\Pi}\{c,\hat{c}|\hat{k}\}\right] / P_{[i+1]}\{\hat{k}\}, \quad (7.28)$$

where $P_{[i+1]}\{\hat{k}\}$ has been calculated in (7.22). Expression (7.28) obtains the required conditional probability of $\hat{c}|\hat{k}$ by dividing the total probability of moving into state $\hat{c},\hat{k}$ in slot $i+1$ (in brackets) by the probability of having $\hat{k}$ copies in that slot. The total probability in the brackets is a sum of two terms. The first term corresponds to the cases where the source transmitted in slot i; thus, the interim channels of those neighbours that still do not possess a copy must have been "off" in slot i. The second term corresponds to the case where the source was silent and no additional neighbours received the frame (that is, $\hat{k}=k$), regardless of the state of their interim channels. In each of these terms, the auxiliary function $\underset{int}{\Pi}(\cdot)$ is then used to capture a single Markov transition of the interim channel states.

7.3.2.4 Extension to overhearing neighbours

The analysis so far has assumed that no neighbour is able to overhear its peers' retransmissions, and can only obtain a copy of a frame directly from the original source. In principle, our method could be extended to allow for inter-neighbour channels of arbitrary quality, by introducing the parameters P_{bg_nn} and P_{gb_nn} for neighbour-to-neighbour channels and considering the state distribution of channels from neighbours with a frame copy to those still without. Due to the complexity of this extension, we do not pursue it comprehensively in this chapter. Rather, only the opposite extreme case is considered, namely, where all inter-neighbour channels are perfect (that is, always "on"), the analysis of which is relatively simple. Our reasoning is that if the difference between the strategy performance in the two extreme cases is found to be small, then one may conjecture that the performance will remain similar for all other non-extreme inter-neighbour channel parameters as well. Our motivation comes from our previous work on the simple memoryless channel model, where it was shown (albeit for the simpler case of memoryless channels) that the inter-neighbour channel quality has only a minor impact on the overall performance of the cooperation strategy

(intuitively, the ability of a neighbour to overhear the frame from a peer, and not just from the source, is balanced by the additional collisions that occur when the source and another neighbour transmit together). It can be seen in Section 7.3.3 that, indeed, the same holds for the two-state Markov channel model as well, in all the evaluation scenarios considered.

Clearly, overhearing among neighbours only impacts on calculation step 3, as it has no effect on the probability of successful delivery in the slot. In the case of always-"on" inter-neighbour channels, a neighbour will overhear the frame if either: (a) its interim channel is "on" and the source transmits, *or* (b) another single neighbour transmits; however, both may not occur simultaneously, as that results in a collision. Consequently, the following changes from the previous analysis are used.

The probability of moving from k frame copies in slot i to $\hat{k}$ copies in slot $i+1$ (expression (7.21)) now becomes

$$\Pi_{[i]}\{k,\hat{k}\} = \begin{cases} \tau_s^*(1-\tau_n^*)^k \underset{int_{[i]}}{P}\{\hat{k}-k|k\}+ & \\ \quad \tau_s^* k\tau_n^*(1-\tau_n^*)^{k-1} \underset{int_{[i]}}{P}\{K-\hat{k}|k\} & k<\hat{k}<K; \\ \tau_s^*(1-\tau_n^*)^k \underset{int_{[i]}}{P}\{\hat{k}-k|k\}+ & \\ \quad \tau_s^* k\tau_n^*(1-\tau_n^*)^{k-1} \underset{int_{[i]}}{P}\{K-\hat{k}|k\}+ & \\ \quad (1-\tau_s^*)k\tau_n^*(1-\tau_n^*)^{k-1} & k<\hat{k}=K; \\ \left[\tau_s^* \underset{int_{[i]}}{P}\{0|k\}+(1-\tau_s)\right](1-\tau_n^*)^k+ & \\ \quad \tau_s^* k\tau_n^*(1-\tau_n^*)^{k-1} \underset{int_{[i]}}{P}\{K-\hat{k}|k\}+ & \\ \quad \left[1-(1-\tau_n^*)^k-k\tau_n^*(1-\tau_n^*)^{k-1}\right] & k=\hat{k}<K; \\ 1 & k=\hat{k}=K. \end{cases} \tag{7.29}$$

This alternative expression for $\Pi_{[i]}\{k,\hat{k}\}$ is then used inside expressions (7.22) and (7.25)–(7.26), which by themselves remain otherwise unchanged.

The other change from the previous analysis relates to the interim channel state distribution. Unlike expression (7.28), where all relevant channels must have been "off" in slot i unless the source was silent, the case of perfect inter-neighbour channels allows a wider range of possibilities, since a transmission by the source over an "on" interim channel can be

destroyed by a simultaneous transmission by one or more neighbours. Accordingly, in this case, expression (7.28) is replaced by

$$\begin{aligned}
\underset{int_{[i+1]}}{P}\{\hat{c}|\hat{k}\} = & \\
& \left[\sum_{k=0}^{\hat{k}} \tau_s^*(1-\tau_n^*)^k P_{[i]}^{ap}\{k\} \underset{int_{[i]}}{P}\{\hat{k}-k|k\} \underset{int}{\Pi}\{0,\hat{c}|\hat{k}\} + \right. \\
& \sum_{k=0}^{\hat{k}} \tau_s^* k \tau_n^* (1-\tau_n^*)^{k-1} P_{[i]}^{ap}\{k\} \underset{int_{[i]}}{P}\{K-\hat{k}|k\} \underset{int}{\Pi}\{K-\hat{k},\hat{c}|\hat{k}\} + \\
& \left[(1-\tau_s^*)(1-\tau_n^*)^{\hat{k}} + 1 - (1-\tau_n^*)^{\hat{k}} - \hat{k}\tau_n^*(1-\tau_n^*)^{\hat{k}-1}\right] \cdot \\
& \left. P_{[i]}^{ap}\{\hat{k}\} \sum_{c=0}^{K-\hat{k}} \underset{int_{[i]}}{P}\{c|\hat{k}\} \underset{int}{\Pi}\{c,\hat{c}|\hat{k}\} \right] / P_{[i+1]}\{\hat{k}\}. \quad (7.30)
\end{aligned}$$

This expression is structured similarly to (7.28), except that the total probability in the brackets now consists of three terms. The first term includes the cases in which the source transmitted in slot i while all neighbours were silent (and therefore the interim channels of those neighbours that still do not possess a copy were "off" in slot i). The second term corresponds to a simultaneous transmission by the source and exactly one peer neighbour; thus, the nodes that still do not have the frame are those whose channels were "on" in slot i. The third term considers the remaining possibilities, where no additional neighbours obtain the frame regardless of the state of their interim channels. In other words, either the source and all neighbours were silent, or more than one neighbour transmitted simultaneously.

7.3.3 Evaluation of the Cooperation Strategies

As mentioned, it is generally accepted that the popular two-state Markov fading channel model is adequate in most practical scenarios [116, 129], and in particular with typical 802.11 channels [130, 132]. For the purpose of evaluating the cooperation strategy performance, we follow the study in [132], which explored the correspondence between a channel's average signal-to-noise ratio (SNR) and the Markov model parameters (average duration in "good" and "bad" states, which is readily converted to P_{bg}, P_{gb}), for various combinations of frame size and transmission rate. Accordingly, the parameter values are set based on the results of [132] for IEEE 802.11 b/g channels with 1500-byte frames transmitted at $11\ Mbps$. The

average SNR of a channel are set to be $SNR[dB] = -20 \log L$, where L is the distance between its endpoint nodes (this corresponds to free-space propagation).

Since a typical cooperative retransmission scenario arguably consists of a poor direct channel with better interim and relay channels, L_{sd} is set such that the direct channel has $\mathrm{SNR}_{sd} = 21.5$dB, which corresponds to $P_{bg\_sd} = 0.11$ and $P_{gb\_sd} = 0.99$; thus, in this case, the direct channel is "on" one-tenth of the time on average. For the neighbours, the channel parameters corresponding to several possible locations are considered, as follows:

- $L_{sn} = L_{nd} = \frac{1}{2} L_{sd}$ (mid-way between source and destination along the straight line);
- $L_{sn} = L_{nd} = \frac{\sqrt{3}}{3} L_{sd}$ (equidistant from source and destination, 30° from the straight line);
- $L_{sn} = L_{nd} = \frac{\sqrt{2}}{2} L_{sd}$ (equidistant from source and destination, 45° from the straight line);
- $L_{sn} = L_{nd} = L_{sd}$ (equidistant from source and destination, 60° from the straight line);
- $L_{sn} = \frac{1}{2} L_{sd}$, $L_{nd} = \frac{\sqrt{3}}{2} L_{sd}$ (triangle of $30^\circ, 60^\circ, 90^\circ$ with the right angle at the neighbour, closer to the source);
- $L_{sn} = \frac{\sqrt{3}}{2} L_{sd}$, $L_{nd} = \frac{1}{2} L_{sd}$ (mirror image of above, closer to destination).

These scenarios, along with the corresponding channel model parameters, are summarised in Table 7.1. The evaluation results are shown in Fig 7.3. Each graph shows the expected latency as a function of the number of cooperative neighbours K. For each scenario, the performance of our strategy is evaluated for both of the inter-neighbour channel quality extremes, that is, "always-off" and "always-on" (denoted in the figure as $P_{nn} = 0$ and $P_{nn} = 1$, respectively), as well as that of simple retransmissions over a direct connection and over a two-hop connection via one of the neighbours.[4]

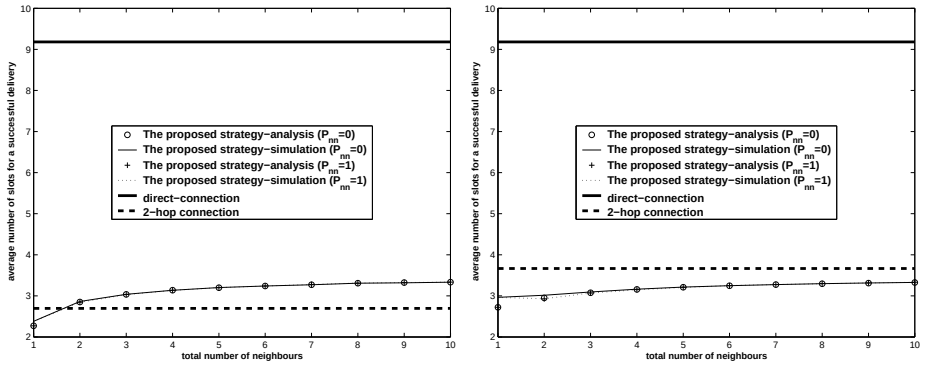
[4]For simple retransmissions, the expected latency of successful delivery of a frame over a link with a two-state Markov channel is

$$\frac{P_{bg}}{P_{bg} + P_{gb}} \cdot 1 + \frac{P_{gb}}{P_{bg} + P_{gb}} \cdot \left(\frac{1}{P_{bg}} + 1 \right), \tag{7.31}$$

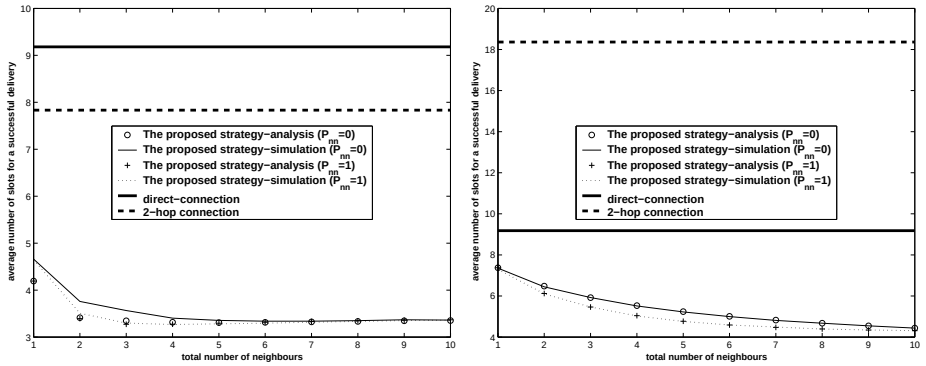
since $\frac{1}{P_{bg}}$ is the expected number of slots for the channel to turn "on" if it was "off" during the initial transmission.

Table 7.1: Channel parameter settings for numerical evaluation on the proposed uncoordinated retransmission strategies.

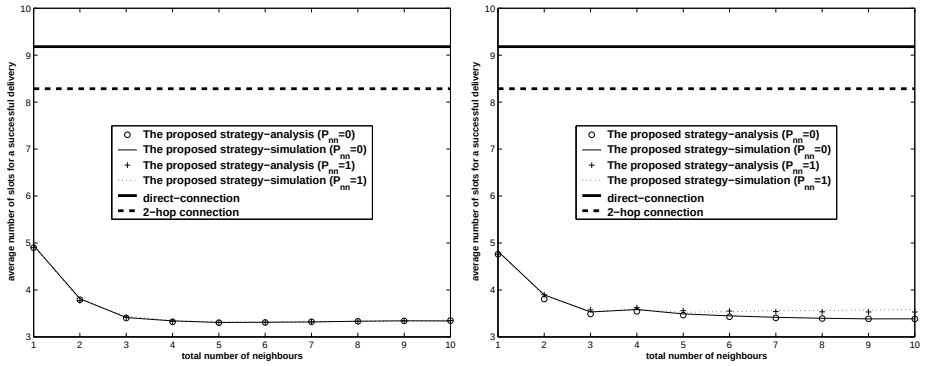
Scenario	L_{sn}	L_{nd}	Interim channel	Relay channel
1	$\frac{1}{2}L_{sd}$	$\frac{1}{2}L_{sd}$	SNR=27.5dB, $P_{bg\_sn} = 0.23$, $P_{gb\_sn} = 0.02$	SNR=27.5dB, $P_{bg\_sn} = 0.23$, $P_{gb\_sn} = 0.02$
2	$\frac{\sqrt{3}}{3}L_{sd}$	$\frac{\sqrt{3}}{3}L_{sd}$	SNR=26.3dB, $P_{bg\_sn} = 0.20$, $P_{gb\_sn} = 0.04$	SNR=26.3dB, $P_{bg\_sn} = 0.20$, $P_{gb\_sn} = 0.04$
3	$\frac{\sqrt{2}}{2}L_{sd}$	$\frac{\sqrt{2}}{2}L_{sd}$	SNR=24.5dB, $P_{bg\_sn} = 0.16$, $P_{gb\_sn} = 0.13$	SNR=24.5dB, $P_{bg\_sn} = 0.16$, $P_{gb\_sn} = 0.13$
4	$\frac{\sqrt{3}}{3}L_{sd}$	$\frac{\sqrt{3}}{3}L_{sd}$	SNR=21.5dB, $P_{bg\_sn} = 0.11$, $P_{gb\_sn} = 0.99$	SNR=21.5dB, $P_{bg\_sn} = 0.11$, $P_{gb\_sn} = 0.99$
5	$\frac{1}{2}L_{sd}$	$\frac{\sqrt{3}}{2}L_{sd}$	SNR=27.5dB, $P_{bg\_sn} = 0.23$, $P_{gb\_sn} = 0.02$	SNR=22.7dB, $P_{bg\_sn} = 0.13$, $P_{gb\_sn} = 0.44$
6	$\frac{\sqrt{3}}{2}L_{sd}$	$\frac{1}{2}L_{sd}$	SNR=22.7dB, $P_{bg\_sn} = 0.13$, $P_{gb\_sn} = 0.44$	SNR=27.5dB, $P_{bg\_sn} = 0.23$, $P_{gb\_sn} = 0.02$



(a) Scenario 1 (b) Scenario 2



(c) Scenario 3 (d) Scenario 4



(e) Scenario 5 (f) Scenario 6

Figure 7.3: Performance of uncoordinated cooperation strategy: expected latency as a function of the number of cooperating neighbours.

These results lead us to observe several important insights. First, it should be noted that the simulation results match the analytical results well, despite some approximations made in the analysis (recall the remark following expressions (7.19)–(7.20) in Section 7.3.2). Furthermore, the difference in performance between the two inter-neighbour channel quality extremes is indeed negligible, leading us to conjecture that the performance will be similar for any non-extreme setting as well.

Most importantly, in all scenarios, the cooperation strategy obtained via our heuristic method (with a proper choice of K) achieves a substantially better latency performance than both the direct connection and the two-hop routing alternatives. In scenarios 1 and 2, where the interim and relay channels are of very good quality, two-hop routing already comes close to the best possible latency, and one cooperative neighbour is best (more uncoordinated neighbours merely increase the rate of collisions). As the interim and relay channels become worse, the optimal neighbour retransmission probabilities increase, and, furthermore, it becomes better to involve a larger number of cooperative neighbours; this is similar to the effect observed when the simple memoryless channel model was used. Thus, the potential benefits of uncoordinated, simultaneous cooperation by multiple neighbours, which is the subject of this chapter, are clearly demonstrated, especially for wireless environments with low-quality channels.

7.4 Summary

In this chapter, the problem of optimization for two uncoordinated MAC layer wireless cooperative retransmission strategies in single-hop setting has been studied. They employ uncoordinated probabilistic retransmission by neighbours overhearing the original frame to minimise the expected frame latency. Two strategies have been presented: (i) Strategy 1: the cooperative neighbours retransmit exclusively while the original sender is silent; (ii) Strategy 2: both may have a non-zero transmission probability simultaneously. Both strategies have been analysed with a simple memoryless channel model. The best results for Strategy 1 have been analytically derived. For Strategy 2, a heuristic approach has been considered that combines a greedy maximisation of successful probability in each slot with a Bayesian re-estimation of the distribution of the number of neighbours with a copy the frame following each slot. It has been demonstrated that, in general, Strategy 2 achieves a superior

performance and considerably reduces the frame latency.

From the results with the memoryless channel model, Strategy 2 has been further analysed with a more realistic two-state Markov fading channel model. A similar but more complex heuristic approach has been still used, which also combines a greedy maximisation of successful probability in each attempt with a Bayesian re-estimation of the system state probability distribution after each failure. It has been demonstrated that the strategy computed by the proposed method, though perhaps not perfectly optimal, still achieves a superior performance and significantly reduces the expected frame delivery latency compared to traditional methods, including simple retransmission and two-hop routing.

Chapter 8

Conclusion of This Thesis

8.1 Contributions of This Thesis

8.1.1 Contributions to IEEE 802.11 Network Performance Analysis

First, the details of the IEEE 802.11 network MAC layer access functions have been introduced in Chapter 2, including CSMA/CA mechanism, IEEE 802.11 DCF, IEEE 802.11e EDCA, and data-rate switching. This introduction has demonstrated the complexity of the IEEE 802.11 network access mechanism and implies that accurately predicting the performance of IEEE 802.11 network is challenging.

Second, a literature review has been performed to study the existing works on IEEE 802.11 network performance analysis. The literature review has shown that the majority of the existing analytical models on IEEE 802.11 EDCA use Markov chain models, and they have some potential limitations. Also, the literature review has indicated a lack of analytical work on investigating the coexistence of DCF and EDCA stations and the impact of data-rate switching.

Third, three Markov chain based analytical models are proposed to investigate the performance of IEEE 802.11 network in this thesis. In Chapter 3, an analytical model has been proposed to analyse the saturated throughput of EDCA. The proposed model overcomes a

number of potential limitations of some exiting analytical models, and it shows better accuracy. In Chapter 4, an analytical model has been proposed to investigate the performance when DCF and EDCA stations coexist in the same base station set, and the saturated throughput has been obtained with the proposed model. The results have indicated that DCF may have a priority similar to that of the best effort traffic in EDCA. In Chapter 5, an analytical model has been proposed to investigate the impact of data-rate switching mechanism on the performance of DCF. A commonly used data-rate switching mechanism has been considered in this model, and the saturated throughput has been analysed. The results have indicated that some threshold values exist for channel condition as well as the number of stations to decide whether data-rate switching should be active or not.

8.1.2 Contributions to Wireless Cooperative Retransmission

First, a literature review has been performed in Chapter 7 to investigate the existing works in this area. The result has shown a lack of uncoordinated distributed wireless cooperative retransmission methods that considers the collision issue on the MAC layer.

Second, two wireless cooperative retransmission strategies have been proposed in Chapter 8 to consider the collision issue from the MAC layer perspective: Strategy 1 where the cooperative neighbours retransmit exclusively while the original sender is silent, and Strategy 2 where both may have a non-zero transmission probability simultaneously. Both strategies have been analysed with a simple memoryless channel model. It has been demonstrated that, in general, Strategy 2 achieves a superior performance by greatly reducing latency. Then Strategy 2 has been analysed with a more realistic two-state Markov fading channel model, where it has shown a superior performance compared to traditional methods, including simple retransmission and two-hop routing.

8.2 Future Work

8.2.1 Future Work on IEEE 802.11 Network Performance Analysis

The analytical models for the IEEE 802.11 network performance analysis proposed in this thesis all consider saturated traffic load and single-hop connection (that is, any pair of stations can be directly connected). The following future work can be suggested to extend them to a more practical network environment:

- **The traffic load is non-saturated.** In this case, some stations may have no queueing traffic in some stages, and they will not join channel access competition. In this thesis, only the saturated traffic load has been considered, because the network performance under such a scenario can be considered as a lower bound. It is predicted that the network performance can be improved with the non-saturated traffic load. The Markov chain models presented in this thesis can be easily extended by adding extra states to model the post-backoff stage[1] used by a station without traffic, similar to that in [35].

- **Multiple data frame payload sizes are used.** In this thesis, only a fixed data frame payload size is used. In reality, this size may be various. A larger payload size can result in that the channel will be occupied for a longer period to transmit it. For the individual station, using a larger payload size implies that it has some priority for channel access over other stations using a smaller data frame size and the same backoff parameters. For the overall network performance, a larger payload size may result in the performance degrade, especially when the network is seriously congested. The reason for such a degrade is that a larger payload size can result in a longer period for the busy channel caused by a collision, and accordingly the channel bandwidth is less efficiently utilised. The models in this thesis can be easily modified such that various data frame sizes can be used.

- **The impact of physical layers should be investigated.** Various physical layer tech-

[1] After a station successfully transmits a frame or discards it, it will immediately start a new backoff stage with the minimum contention window CW_{min}. Such a backoff stage is defined as a post-backoff stage. If this station has no traffic after the backoff counter reaches zero in this post-backoff stage, the backoff counter will remain zero and the station may immediately start a transmission once it has traffic.

nologies have been developed for IEEE 802.11, such as IEEE 802.11a [13], IEEE 802.11b [14], and IEEE 802.11g [15]. Accordingly, they may result in some differences on the physical layer, such as different set of data rates being supported and different transmission coverage. However, the same CSMA/CA based DCF or EDCA are used on the MAC layer. Therefore, the Markov chain models for DCF and EDCA presented in this thesis can be used independent of the physical layer technology being used.

- **Multi-hop scenario should be considered.** In this case, traffic should travel in a hop-by-hop manner. These hops can be mutually affected. For example, the traffic load at an interim hop will depend on the performance of all other hops, because its incoming traffic is from the previous hops, and its ongoing traffic is to the next hops. Also, the channel access of the interim hop will be affected by the channel access activities of its adjacent hops.

8.2.2 Future Work on Wireless Cooperative Retransmission

The wireless cooperative retransmission strategies proposed in this thesis have been well investigated with a symmetric channel system. However, such a a symmetric channel system may appear unrealistic in a real wireless cooperative communication environment. In a typical mobile communication system, it is more realistic that the cooperative neighbours are randomly distributed around the source and the destination, and the quality of their channels is random accordingly. Therefore, such a random network environment will definitely be included in our future work, and we are considering Poisson node distribution model, which is popularly used to model this randomness in this area [133–137].

In addition to the above, the following future work can be suggested:

- **A more realistic channel model should be used** We have assumed that the channel is either “on” or “off” in our existing work. Moreover, we have assumed that a deep fading channel condition is applied to the “off” state such that an unsuccessful transmission has no impact on the receiver. Such assumptions greatly simplify our analysis, but they are not consistent with the facts that unsuccessful transmissions may

still possibly cause interference to the receiver in a realistic radio channel environment. Therefore, a more realistic channel model should be considered in the future work.

- **The impact of wireless cooperative retransmission on the overall network performance should be investigated.** In wireless cooperative retransmission, the cooperative neighbours may help the delivery of traffic from the source to the destination. However, the activities of these cooperative neighbours may cause negative impacts on the overall network performance. For example, the cooperative neighbours's own traffic may be delayed, or its cooperative transmission may interfere with other traffic delivery activities. Therefore, a trade-off issue between the overall network performance and the individual traffic delivery arises for wireless cooperative retransmission. Such an issue has been raised by some recent studies, such as that in [138].

- **Multi-hop scenario should be considered:** When the source is far away from the destination and the traffic delivery between them is quite impossible (even with the help of wireless cooperative neighbours), a practicable solution is to use multi-hop approach. However, different from traditional multi-hop scenario, wireless cooperative neighbours will participate in the traffic delivery between adjacent hops. Thus, the mutual impact of hops will be more complex because cooperative neighbours also become involved.

Bibliography

[1] International Telecommunication Union (ITU), "World telecommunication indicators databased 2003," Geneva, Switzerland, ITU, 2003.

[2] A. Akella, G. Judd, S. Seshan, and P. Steenkiste, "Self-management in chaotic wireless deployments," in *Proceeding of The* $11st$ *Annual Internaional Conference on Mobile Computing and Networking*, Cologne, Germany, 2005, pp. 185–199.

[3] Instat/MDR, "3Q 2004 WLAN Market Share Report"., 2004. [Online]. Available: http://www.instat.com/r/nrep/2004/IN0401429WL.htm

[4] Datacomm Research,"New Datacomm Research Report: Wireless LAN Equipment Shipments to Triple Within Five Years", 2005. [Online]. Available: http://www.tmcnet.com/usubmit/2005/Feb/1120138.htm

[5] "Information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," 1999.

[6] "Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," pp. i–513, 2003.

[7] J. Pereira, "Fourth generation: now, it is personal!" in *Proceeding of the* $11st$ *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2000*, vol. 2, pp. 1009–1016.

[8] S. Hui and K. Yeung, "Challenges in the migration to 4G mobile systems," *Communications Magazine, IEEE*, vol. 41, no. 12, pp. 54–59, December 2003.

[9] Y. Ge and J. Hou, "An analytical model for service differentiation in IEEE 802.11," in *Proceeding of IEEE International Conference on Communications (ICC), 2003*, vol. 2, pp. 1157–1162 vol.2.

[10] A. Banchs and X. Perez, "Providing throughput guarantees in IEEE 802.11 wireless LAN," in *Proceeding of IEEE Wireless Communications and Networking Conference (WCNC), 2002*, vol. 1, pp. 130–138 vol.1.

[11] ——, "Distributed weighted fair queuing in 802.11 wireless LAN," in *Proceeding of IEEE International Conference on Communications (ICC), 2002*, vol. 5, pp. 3121–3127 vol.5.

[12] N. Vaidya, A. Dugar, S. Gupta, and P. Bahl, "Distributed fair scheduling in a wireless LAN," *IEEE Transactions on Mobile Computing*, vol. 4, no. 6, pp. 616–629, November-December 2005.

[13] "Supplement to IEEE standard for information technology telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: high-speed physical layer in the 5 GHz band," *IEEE Standard 802.11a-1999*, 1999.

[14] "Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: Higher-speed Physical Layer Extension In The 2.4 GHz Band," *IEEE Standard 802.11b-1999*, pp. i–90, 2000.

[15] "Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band," *IEEE Standard 802.11g-2003*, pp. c1–68, 2006.

[16] "IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," 2005.

[17] M. Conti and S. Giordano, "Multihop ad hoc networking: The reality," *IEEE Communications Magazine*, vol. 45, no. 4, pp. 88–95, April 2007.

[18] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, 2004.

[19] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.

[20] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antenna," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[21] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *Bell Lab, Internal Technical Memorandum*, October 1995.

[22] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1121–1128, May 1999.

[23] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, March 1999.

[24] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 199–207, Feburary 1999.

[25] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1169–1174, July 2000.

[26] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, March 2000.

[27] J. Robinson and T. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 5, pp. 917–928, June 2004.

[28] H. Wu, X. Wang, Q. Zhang, and X. Shen, "IEEE 802.11e Enhanced Distributed Channel Access (EDCA) Throughput Analysis," in *Proceeding of IEEE International Conference on Communications (ICC), 2006*, pp. 223–228.

[29] K. Xu, Q. Wang, and H. Hassanein, "Performance analysis of differentiated QoS supported by IEEE 802.11e enhanced distributed coordination function (EDCF) in WLAN," in *Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), 2003*, vol. 2, pp. 1048–1053.

[30] A. Banchs and L. Vollero, "A delay model for IEEE 802.11e EDCA," *IEEE Communications Letters*, vol. 9, no. 6, pp. 508–510, June 2005.

[31] X. Chen, H. Zhai, and Y. Fang, "Enhancing the IEEE 802.11e in QoS Support: Analysis and Mechanisms," in *Proceeding of the* $2nd$ *International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, 2005*.

[32] H. Zhu and I. Chlamtac, "Performance analysis for IEEE 802.11e EDCF service differentiation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1779–1788, July 2005.

[33] J. Hui and M. Devetsikiotis, "A Unified Model for the Performance Analysis of IEEE 802.11e EDCA," *IEEE Transactions on Communications*, vol. 53, no. 9, pp. 1498–1510, September 2005.

[34] Y. Xiao, "Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1506–1515, July 2005.

[35] P. Engelstad and O. Osterbo, "Delay and Throughput Analysis of IEEE 802.11e EDCA with Starvation Prediction," in *Proceeding of the* $30th$ *IEEE Conference on Local Computer Networks, 2005*, pp. 647–655.

[36] Z.-N. Kong, D. Tsang, B. Bensaou, and D. Gao, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 10, pp. 2095–2106, December 2004.

[37] B. Li and R. Battiti, "Supporting Service Differentiation with Enhancements of the IEEE 802.11 MAC Protocol: Models and Analysis," *Technical Report of Department of Computer Science and Telecommunications of University of Trento, no. [DIT-03-024]*, 2003.

[38] J. Tantra, C.-H. Foh, and A. Mnaouer, "Throughput and delay analysis of the IEEE 802.11e EDCA saturation," in *Proceeding of IEEE International Conference on Communications (ICC), 2005*, vol. 5, pp. 3450–3454.

[39] I. Inan, F. Keceli, and E. Ayanoglu, "Saturation Throughput Analysis of the 802.11e Enhanced Distributed Channel Access Function," in *Proceeding of IEEE International Conference on Communications (ICC), 2007*, pp. 409–414.

[40] K.-S. Lee, Yutae. Lee and J. Jang, "Saturation Throughput Analysis of IEEE 802.11e EDCA ," *Lecture Notes in Computer Science, Spring*, vol. 4682/2007, pp. 1223–1232, July 2007.

[41] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, "Performance study of MAC for service differentiation in IEEE 802.11," in *Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), 2002*, vol. 1, pp. 778–782.

[42] V. Ramaiyan, A. Kumar, and E. Altman, "Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness, multistability and throughput differentiation," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, pp. 109–120, June 2005.

[43] Y.-L. Kuo, C.-H. Lu, E. H.-K. Wu, G.-H. Chen, and Y.-H. Tseng, "Performance analysis of the enhanced distributed coordination function in the IEEE 802.11e," in *Proceeding of the* $58th$ *IEEE Vehicular Technology Conference (VTC), Fall, 2003*, vol. 5, pp. 3488–3492.

[44] Y. Chen, Q.-A. Zeng, and D. Agrawal, "Performance analysis of IEEE 802.11e enhanced distributed coordination function," in *Proceeding of the* $11st$ *IEEE International Conference on Networks (ICON), 2003*, pp. 573–578.

[45] G. Sharma, A. Ganesh, and P. Key, "Performance Analysis of Contention Based Medium Access Control Protocols," in *Proceeding of the* $25th$ *IEEE International Conference on Computer Communications (INFOCOM), 2006*, pp. 1–12.

[46] G. Bianchi, I. Tinnirello, and L. Scalia, "Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," *IEEE Network*, vol. 19, no. 4, pp. 28–34, July-August 2005.

[47] J. Majkowski and F. Palacio, "Coexistence of IEEE 802.11B and IEEE 802.11E Stations in QoS Enabled Wireless Local Area Network," in *Proceeding of Communication Systems and Applications (CSA) conference, 2006*.

[48] G.-H. Hwang and D.-H. Cho, "Performance Analysis on Coexistence of EDCA and Legacy DCF Stations in IEEE 802.11 Wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3355–3359, December 2006.

[49] O. Abu-Sharkh and A. Tewfik, "Multi-rate 802.11 WLANs," in *Proceeeding of IEEE Global Telecommunications Conference (GLOBECOM), 2005*, vol. 5, pp. 3128–3133.

[50] ——, "Performance analysis of multi-rate 802.11 WLANs under finite load and saturation conditions," in *Proceeeding of the* $62nd$ *IEEE Vehicular Technology Conference (VTC), Fall 2005*, vol. 4, pp. 2652–2657.

[51] Y.-L. Kuo, K.-W. Lai, F. Y.-S. Lin, Y.-F. Wen, E. H.-K. Wu, and G.-H. Chen, "Multi-rate throughput optimization for wireless local area network anomaly problem," in *Proceeding of the* $2nd$ *International Conference on Broadband Networks, 2005*, vol. 1, pp. 591–601.

[52] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proceeding of the* $22nd$ *IEEE Annual Joint Conference of Computer and Communications Societies (INFOCOM), 2003*, vol. 2, pp. 836–843.

[53] Y. Li, X. Wang, and S. Mujtaba, "Impact of physical layer parameters on the MAC throughput of IEEE 802.11 wireless LANs," in *Conference Record of the* $38th$ *Asilomar Conference on Signals, Systems and Computers, 2004*, vol. 2, pp. 1468–1472.

[54] M. Ergen and P. Varaiya, "Throughput formulation and WLAN optimization in mixed data rates for IEEE 802.11 DCF mode," in *Proceeding of IEEE Global Telecommunications Conference Workshops (GlobeCom Workshops), 2004*, pp. 266–269.

[55] Y.-J. Kim and Y.-J. Suh, "An efficient rate switching scheme for IEEE 802.11 wireless LANs," in *Proceeding of the* $61st$ *IEEE Vehicular Technology Conference (VTC), Spring 2005*, vol. 4, pp. 2364–2368.

[56] A. Kamerman and L. Monteban, "WaveLAN II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, pp. 118–133, Summer 1997.

[57] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-Hop wireless networks," in *Proceedings of the* $7th$ *Annual International Conference on Mobile Computing and Networking, Rome Italy, 2001*, pp. 236 – 251.

[58] K. Saitou, Y. Inoue, T. Kumagai, M. Iizuka, S. Aikawa, and M. Morikura, "An Effective Data Transfer Method for IEEE 802.11Wireless LANs," *IEICE Ttransactions on Communications*, vol. E88CB, no. 3, pp. 1266–1270, March 2005.

[59] K. Saitoh, Y. Inoue, A. Iiuka, and M. Morikura, "An effective data transfer method by integrating priority control into multirate mechanisms for IEEE 802.11 wireless LANs," in *Proceeding of the* $55th$ *IEEE Vehicular Technology Conference (VTC), Spring 2002*, vol. 4, pp. 2589 – 2593.

[60] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," in *Proceeding of ACM MobiCom, 2002*.

[61] R. Mahadevappa and S. ten Brink, "Rate-feedback schemes for MIMO-OFDM wireless LANs," in *Proceeding of the* $60th$ *IEEE Vehicular Technology Conference (VTC), Fall 2004*, vol. 1, pp. 563–567.

[62] E. Kim and Y.-J. Suh, "ATXOP: an adaptive TXOP based on the data rate to guarantee fairness for IEEE 802.11e wireless LANs," in *Proceeding of the* $60th$ *IEEE Vehicular Technology Conference (VTC), Fall 2004*, vol. 4, pp. 2678–2682.

[63] L. Xiong, G. Mao, and Y. Luan, "Enhancement to IEEE 802.11e EDCA with QoS Differentiation Based on Station Priority," in *Proceeding of Australian Telecommunication Networks and Applications Conference (ATNAC), 2004*.

[64] L. Xiong and G. Mao, "Saturated throughput analysis of IEEE 802.11e using two-dimensional Markov chain model," in *Proceeding of the* $3rd$ *International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, 2006*.

[65] ——, “An Analysis of the Coexistence of IEEE 802.11 DCF and IEEE 802.11e EDCA,” in *Proceeding of IEEE Wireless Communications and Networking Conference (WCNC), 2007*, pp. 2264–2269.

[66] ——, “Saturation Throughput Analysis for IEEE 802.11e EDCA,” *Computer Networks*, vol. 51, no. 11, pp. 3047–3068, August 2007.

[67] ——, “Performance Analysis of IEEE 802.11 DCF with Data Rate Switching,” *IEEE Communications Letters,*, vol. 11, no. 9, pp. 759–761, September 2007.

[68] N. Shankar and M. van der Schaar, “Performance Analysis of Video Transmission Over IEEE 802.11a/e WLANs,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 2346–2362, July 2007.

[69] C.-H. Foh, Y. Zhang, Z. Ni, J. Cai, and K.-N. Ngan, “Optimized Cross-Layer Design for Scalable Video Transmission Over the IEEE 802.11e Networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1665–1678, December 2007.

[70] N. Blefari-Melazzi, A. Detti, I. Habib, A. Ordine, and S. Salsano, “TCP Fairness Issues in IEEE 802.11 Networks: Problem Analysis and Solutions Based on Rate Control,” *IEEE Transactions on Wireless Communications,* vol. 6, no. 4, pp. 1346–1355, April 2007.

[71] A. Babu and L. Jacob, “Fairness Analysis of IEEE 802.11 Multirate Wireless LANs,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3073–3088, September 2007.

[72] C.-H. Foh, M. Zukerman, and J. Tantra, “A Markovian Framework for Performance Evaluation of IEEE 802.11,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1276–1265, April 2007.

[73] C. Que, X. Zhang, Y. Liu, and Y. Huang, “Performance Analysis of Chain Topology in IEEE 802.11 Multi-hop Ad hoc Networks,” in *Proceeding of International Conference on Convergence Information Technology, 2007*, pp. 1880–1883.

[74] K. Wang, F. Yang, Q. Zhang, and Y. Xu, “Modeling path capacity in multi-hop IEEE 802.11 networks for QoS services,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 738–749, Feburary 2007.

[75] J.-Y. Yoo and J. Kim, "Maximum End-to-End Throughput of Chain-Topology Wireless Multi-Hop Networks," in *Proceeding of IEEE Wireless Communications and Networking Conference (WCNC), 2007*, pp. 4279–4283.

[76] Z. Niu, W. Yao, Q. Ni, and Y. Song, "Study on QoS Support in 802.11e-based Multihop Vehicular Wireless Ad Hoc Networks," in *Proceeding of IEEE International Conference on Networking, Sensing and Control (ICNSC), 2007*, pp. 705–710.

[77] P.-C. Ng and S.-C. Liew, "Throughput Analysis of IEEE802.11 Multi-Hop Ad Hoc Networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 309–322, April 2007.

[78] M. Deekshitulu, S. Nandi, and A. Chowdhury, "Improving MAC Layer Fairness in Multi-Hop 802.11 Networks," in *Proceeding of the* $2nd$ *International Conference on Communication Systems Software and Middleware (COMSWARE), 2007*, pp. 1–8.

[79] L. Li and P. A. Ward, "Structural unfairness in 802.11-basedwireless mesh networks," in *Proceeding of the* $5th$ *Annual Conference on Communication Networks and Services Research (CNSR), 2007*, pp. 213–220.

[80] X. Li, P.-Y. Kong, and K.-C. Chua, "Finding an Optimum Maximum Congestion Window for TCP Reno over 802.11 Based Ad Hoc Networks," in *Proceeding of IEEE Wireless Communications and Networking Conference (WCNC), 2007*, pp. 3619–3624.

[81] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. New York: Wiley, 1998.

[82] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF performance analysis," *IEEE Communications Letters*, vol. 9, no. 8, pp. 765–767, August 2005.

[83] MATLAB online documentation. Optimization Toobox: $fsolve$. [Online]. Available: http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/fsolve.html

[84] J. Dennis, "Nonlinear least-squares," in *The State of the Art of Numerical Analysis, D. Jacobs, Ed. Academic Press*, Orlando, Fla, 1977, pp. 269–312.

[85] K. Levenberg, "A method for the solution of certain problems in least-squares," *Quarterly Applied Mathematics 2*, pp. 164–168, 1944.

[86] D. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *SIAM Journal Applied Mathematics*, vol. 11, pp. 431–441, 1963.

[87] M. Powell, “A fortran subroutine for solving systems of nonlinear algebraic equations,” *Numerical Methods for Nonlinear Algebraic Equations, P. Rabinowitz, ed., Ch.7*, vol. 11, 1970.

[88] T. Coleman and Y. Li, “On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds,” *Mathematical Programming*, vol. 67, no. 2, pp. 189–224, 1994.

[89] ——, “An interior, trust region approach for nonlinear minimization subject to bounds,” *SIAM Journal on Optimization*, vol. 6, pp. 418–445, 1996.

[90] “Supplement to IEEE standard for information Technology -Telecommunications and information exchange between systems -local and metropolitan area networks- specific requirements- part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications: Higher-speed physical layer extension in the 2.4 GHz band,” pp. i–90, 2000.

[91] “OPNET university program.” [Online]. Available: http://www.opnet.com/services/university/

[92] Y. Inoue , K. Saitoh, T. Sakata, M. Morikura, and H. Matsue,, “A Study on the Rate Swiching Algorithm for IEEE 802.11 Wireless LANs,” *The Institute of Electrical Engineers of Japan (IEEJ) Transactions on Electronics, Information and Systems*, vol. 24-C, pp. 33–40, January 2004.

[93] M. Roden, “Analog and Digital Communication Systems,” the 4th edition, Prentice Hall, 1996.

[94] P. Chatzimisios, A. Boucouvalas, and V. Vitsas, “Performance analysis of IEEE 802.11 DCF in presence of transmission errors,” in *Proceeding of IEEE International Conference on Communications (ICC), 2004*, vol. 7, pp. 3854–3858.

[95] T. Hunter and A. Nosratinia, “Cooperation diversity through coding,” in *Proceeding of IEEE International Symposium on Information Theory, 2002*, pp. 220–.

[96] B. Aazhang, R. Blum, J. Laneman, K. Liu, W. Su, and A. Wittneben, "Guest editorial: Cooperative communications and networking," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 241–244, Feburary 2007.

[97] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 283–289, Feburary 2006.

[98] J. Laneman and G. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, October 2003.

[99] M. Janani, A. Hedayat, T. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing, [see also IEEE Transactions on Acoustics, Speech, and Signal Processing]*, vol. 52, no. 2, pp. 362–371, Feburary 2004.

[100] A. Murugan, K. Azarian, and H. E. Gamal, "Cooperative lattice coding and decoding in half-duplex channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 268–279, February 2007.

[101] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity — part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, November 2003.

[102] ——, "User cooperation diversity — part II. implementation aspects and performance evaluation," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, November 2003.

[103] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[104] N.-S. Kim, B. An, D.-H. Kim, and Y.-H. Lee, "Wireless Ad-hoc Networks Using Cooperative Diversity-based Routing in Fading Channel," in *Proceeding of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2007*, pp. 70 – 73.

[105] P. Larsson, “Selection diversity forwarding in a multihop packet radio network with fading channel and capture,” *ACM Mobile Computing and Communications Review*, vol. 5, no. 4, pp. 47–54, October 2001.

[106] S. Jain and S. Das, “Exploiting path diversity in the link layer in wireless ad hoc networks,” in *Proceeding of the* $6th$ *IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM), 2005*, pp. 22–30.

[107] ——, “ExOR: Opportunistic Multi-Hop Routing for Wireless Networks,” *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 4, pp. 133–144, October 2005.

[108] M. Tacca, P. Monti, and A. Fumagalli, “Cooperative and Reliable ARQ Protocols for Energy Harvesting Wireless Sensor Nodes,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2519–2529, July 2007.

[109] H. Hu and Z. Yang, “The study of power control based cooperative opportunistic routing in wireless sensor networks,” in *Proceeding of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2007*, pp. 345 – 348.

[110] M. Zorzi and R. Rao, “Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Energy and Latency Performance,” *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, October-December 2003.

[111] ——, “Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance,” *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 337–348, October-December 2003.

[112] P. Coronel, R. Doss, and W. Schott, “Geographic routing with cooperative relaying and leapfrogging in wireless sensor networks,” in *Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), 2007*, pp. 646–651.

[113] F. Librino, M. Levorato, and M. Zorzi, “Distributed Cooperative Routing and Hybrid ARQ in MIMO-BLAST Ad Hoc Networks,” in *Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), 2007*, pp. 657–662.

[114] F. Tobagi and L. Kleinrock, “Packet switching in radio channels: Part ii–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution,” *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1417–1433, December 1975.

[115] I. Cerutti, A. Fumagalli, and G. Ho, "Saturation throughput gain in fixed multiplexing radio networks with cooperative retransmission protocols," in *Proceeding of IEEE International Conference on Communications (ICC), 2006*, vol. 10, pp. 4489–4494.

[116] M. Dianati, X. Ling, K. Naik, and X. Shen, "A node-cooperative ARQ scheme for wireless ad-hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 1032–1044, May 2006.

[117] H. Choi and J. H. Lee, "Cooperative ARQ with Phase Pre-Compensation," in *Proceeding of the* $66th$ *IEEE Vehicular Technology Conference (VTC) , Fall 2006*, pp. 215 – 219.

[118] D. Lee, Y. S. Jung, and J. H. Lee, "Amplify-and-forward cooperative transmission with multiple relays using phase feedback," in *Proceeding of the* $64th$ *IEEE Vehicular Technology Conference (VTC), Fall 2006*, pp. 1–5.

[119] J. D. Morillo-Pozo and J. Garcia-Vidal, "A Low Coordination Overhead C-ARQ Protocol with Frame Combining," in *Proceeding of the* $18th$ *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2006*, pp. 1 – 5.

[120] G. Yu, Z. Zhang, and P. Qiu, "Cooperative ARQ in Wireless Networks: Protocols Description and Performance Analysis," in *Proceeding of IEEE International Conference on Communications (ICC), 2006*, vol. 8, pp. 3608 – 3614.

[121] B. Zhao and M. Valenti, "Practical relay networks: a generalization of hybrid-ARQ," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 7–18, January 2005.

[122] ——, "Position-based relaying with hybrid-ARQ for efficient ad hoc networking," *EURASIP Journal on Wireless Communications and Networking*, vol. 5, no. 5, pp. 610 – 624, October 2005.

[123] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 2, pp. 353–357, May 1996.

[124] M. Zorzi and R. Rao, "On the statistics of block errors in bursty channels," *IEEE Transactions on Communications*, vol. 45, no. 6, pp. 660–667, June 1997.

[125] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, Feburary 1995.

[126] M. Zorzi, R. Racd, and L. Milstein, "A Markov model for block errors on fading channels," in *Proceeding of the* $7th$ *IEEE International Symposium onPersonal, Indoor and Mobile Radio Communications (PIMRC), 1996*, vol. 3, pp. 1074–1078.

[127] M. Zorzi, R. Rao, and L. Milstein, "Error statistics in data transmission over fading channels," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1468–1477, November 1998.

[128] R. Rao, "Perspectives on the impact of fading on protocols for wireless networks," in *Proceeding of IEEE International Conference onPersonal Wireless Communications, 1997*, pp. 489–493.

[129] M. Zorzi, R. Rao, and L. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 2, pp. 445–455, May 1997.

[130] S. Khayam and H. Radha, "Markov-based modeling of wireless local area networks," in *Proceeding of ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM),San Diego, CA, 2003*, pp. 100–107.

[131] MATLAB online documentation. Optimization Toobox:$fmincon$. [Online]. Available: http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/fmincon.html

[132] J. Aráuz and P. Krishnamurthy, "Markov modeling of 802.11 channels," in *Proceeding of IEEE Vehicular Technology Conference (VTC), Fall 2003*, pp. 771–775.

[133] L. Song and D. Hatzinakos, "Cooperative transmission in poisson distributed wireless sensor networks: protocol and outage probability," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 10, pp. 2834–2843, October 2006.

[134] P.-J. Wan and C.-W. Yi, "Coverage by randomly deployed wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2658–2669, June 2006.

[135] X. Ta, G. Mao, and B. Anderson, “On the probability of k-hop connection in wireless sensor networks,” *IEEE Communications Letters*, vol. 11, no. 8, pp. 662–664, August 2007.

[136] S. S. Ram, D. Manjunath, S. K. Iyer, and D. Yogeshwaran, “On the path coverage properties of random sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 6, no. 5, pp. 446–458, May 2007.

[137] Y.-R. Tsai, “Sensing coverage for randomly distributed wireless sensor networks in shadowed environments,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 556–564, January 2008.

[138] A. Nigara, M. Qin, and R. Blum, “On the Performance of Wireless Ad Hoc Networks Using Amplify-and-Forward Cooperative Diversity,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3204–3214, November 2006.