Design and Analysis of Medium Access Control Protocols for Broadband Wireless Networks

by

Lin Cai

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The next-generation wireless networks are expected to integrate diverse network architectures and various wireless access technologies to provide a robust solution for ubiquitous broadband wireless access, such as wireless local area networks (WLANs), Ultra-Wideband (UWB), and millimeter-wave (mmWave) based wireless personal area networks (WPANs), etc. To enhance the spectral efficiency and link reliability, smart antenna systems have been proposed as a promising candidate for future broadband access networks. To effectively exploit the increased capabilities of the emerging wireless networks, the different network characteristics and the underlying physical layer features need to be considered in the medium access control (MAC) design, which plays a critical role in providing efficient and fair resource sharing among multiple users.

In this thesis, we comprehensively investigate the MAC design in both single- and multi-hop broadband wireless networks, with and without infrastructure support. We first develop mathematical models to identify the performance bottlenecks and constraints in the design and operation of existing MAC. We then use a cross-layer approach to mitigate the identified bottleneck problems. Finally, by evaluating the performance of the proposed protocols with analytical models and extensive simulations, we determine the optimal protocol parameters to maximize the network performance.

In specific, a generic analytical framework is developed for capacity study of an IEEE 802.11 WLAN in support of non-persistent asymmetric traffic flows. The analysis can be applied for effective admission control to guarantee the quality of service (QoS) performance of multimedia applications. As the access point (AP) becomes the bottleneck in an infrastructure based WLAN, we explore the multiple-input multiple-output (MIMO) capability in the future IEEE 802.11n WLANs and propose a MIMO-aware multi-user (MU) MAC. By exploiting the multi-user degree of freedom in a MIMO

system to allow the AP to communicate with multiple users in the downlink simultaneously, the proposed MU MAC can minimize the AP-bottleneck effect and significantly improve the network capacity. Other enhanced MAC mechanisms, e.g., frame aggregation and bidirectional transmissions, are also studied.

Furthermore, different from a narrowband system where simultaneous transmissions by nearby neighbors collide with each other, wideband system can support multiple concurrent transmissions if the multi-user interference can be properly managed. Taking advantage of the salient features of UWB and mmWave communications, we propose an exclusive region (ER) based MAC protocol to exploit the spatial multiplexing gain of centralized UWB and mmWave based wireless networks. Moreover, instead of studying the asymptotic capacity bounds of arbitrary networks which may be too loose to be useful in realistic networks, we derive the expected capacity or transport capacity of UWB and mmWave based networks with random topology. The analysis reveals the main factors affecting the network (transport) capacity, and how to determine the best protocol parameters to maximize the network capacity. In addition, due to limited transmission range, multi-hop relay is necessary to extend the communication coverage of UWB networks. A simple, scalable, and distributed UWB MAC protocol is crucial for efficiently utilizing the large bandwidth of UWB channels and enabling numerous new applications cost-effectively. To address this issue, we further design a distributed asynchronous ER based MAC for multi-hop UWB networks and derive the optimal ER size towards the maximum network throughput. The proposed MAC can significantly improve both network throughput and fairness performance, while the throughput and fairness are usually treated as a tradeoff in other MAC protocols.

Acknowledgments

I would like to express my deep gratitude to my supervisors, Professor Xuemin (Sherman) Shen and Professor Jon W. Mark, for their continuous guidance, encouragement and support during my graduate studies at the University of Waterloo. Without their inspiration, deep insight, and invaluable advice, this dissertation would not have been possible. I have benefited tremendously from their extensive knowledge, professional experience, and strong commitment towards the excellence of research. More importantly, I have learned from them the highly positive attitude when facing challenges. They are not only advisors, but also role models of professionalism, integrity, respect, and responsibility. I would like to thank them sincerely for everything I have achieved.

I would also like to thank Professor Terence D. Todd, Professor Xinzhi Liu, Professor Sagar Naik, and Professor Liang-Liang Xie for serving on my dissertation committee. They contributed their precious time in reading my thesis and providing constructive comments and valuable suggestions that helped to improve this dissertation.

I am also very thankful to Professor Weihua Zhuang. The knowledge of stochastic processes I learned from her course E&CE 604 is the foundation of the research work in this dissertation.

My deep appreciation goes to Drs. Lin Cai, Fen Hou, Xinhua Ling, Yu Cheng, Ho Young Hwang, Hangguan Shan, Mr. Tom Luan, Mr. Yuanguo Bi and Mr. Jian Qiao for their fruitful discussions and professional collaborations. I am also grateful to Drs. Ping Wang, Wei Song, Minghui Shi, Mr. Hao Liang, Mr. Yongkang Liu, Mr. Wei Cai, Ms. Clare Li, Xiaoxia Zhang, and other friends and colleagues in the Broadband Communications Research Group (BBCR) for their warm friendship and invaluable comments and advice.

The research grants of Natural Science and Engineering Research Council (NSERC) of Canada, the Ontario Graduate Scholarship (OGS), and numerous scholarships from the University of Waterloo are gratefully acknowledged. Many thanks to the administrative staff: Wendy Boles, Lisa Hendel, Annette Dietrich, and Karen Schooley.

Finally, I wish to express my deepest gratitude to my husband, my parents and parents-in-law. Without their unconditional love, understanding, and support, I would never fulfill my dream and complete my Ph.D. program. Special thanks to my son Justin, who brings me the greatest joy and happiness in life and makes a unique contribution to my Ph.D. study.

To my husband Dong and our son Justin

To my parents

To my parents-in-law

Contents

List of Tables

List of Figures

List of Abbreviations

AP	Access Point
ARQ	automatic repeat request
AWGN	additive white Gaussian noise
BC	backoff counter
BIFS	backoff interframe space
BP	beacon period
CAP	contention access period
CBR	constant bit rate
CDMA	code division multiple access
CSMA/CA	carrier sense multiple access with collision avoidance
CSI	channel state information
CTAP	channel time allocation period
CTS	clear to send
CW	contention window
C-UWB	continuous wave Ultra Wideband
CCI	co-channel interference
DCF	distributed coordination function
DEX	distributed exclusive region based MAC

DIFS DCF interframe space

DRP distributed reservation protocol

DSSS direct sequence spread spectrum

DS-UWB direct sequence Ultra Wideband

DTP data transfer period

EDCA enhanced distributed channel access

ER exclusive region

FCS frame check sequence

FDMA frequency division multiple access

HCS header check sequence

HDV high definition video

IE information element

IPTV Internet Protocol TV

ISI intersymbol interference

ITU International Telecommunication Union

MAC medium access control

MAI multiple access interference

MAS medium access slot

MB-OFDM multi-band orthogonal frequency division multiplexing

MC-CDMA multi carrier CDMA

MIMO multiple input multiple output

mmWave millimeter wave

MN mobile node

MSI maximum sustainable interference

MTU	maximum transmission unit
MU-CTS	multi-user clear to send
MU-RTS	multi-user request to send
NAV	network allocation vector
OFDM	Orthogonal frequency-division multiplexing
PCA	prioritized channel access
PCF	point coordination function
PLCP	physical layer convergence protocol
PLR	packet loss rate
PNC	piconet coordinator
PQ	priority queue
QoS	quality of service
REX	randomized exclusive region based scheduling scheme
RTS	request to send
SIFS	short interframe space
SINR	signal to interference plus noise ratio
SLNR	signal to leakage noise ratio
TDMA	time division multiple access
TTL	time to live
TXOP	transmission opportunity
UWB	Ultra Wideband
VBR	variable bit rate
VoIP	voice over IP
WBAN	wireless body area network

WLAN wireless local area network

WMAN wireless metropolitan area network

WPAN wireless personal area network

List of Notations

λ_i	The mean traffic arrival rate of user i
μ_i	The mean frame service rate of user i
ρ_i	The queue utilization ratio or traffic intensity of user i
τ_i	The transmission probability of user i at a slot
$\tau\|k$	The conditional collision probability of a user at a slot given k contending users
η	Tranceiver efficiency
$\omega_i(t)$	The waiting time of the head-of-the-line frame in the queue to user i at time t
α	Path loss exponent
β	A SINR threshold
γ_i	The set of flows concurrently transmitting with flow f_i
ς	Antenna radiation efficiency
A	The area of the exclusive region
d_i	The transmission distance of flow i
$E[A_i]$	The average transmission attempts of user i
$E[CT]$	The expected number of concurrent transmissions
$E[N_{busy}]$	The average number of busy slots
$E[N_{idle}]$	The average number of idle slots
$E[S]$	The average throughput of a WLAN under the stable state
$E[T]$	The capacity of the network
$E[Tr]$	The transport capacity of the network
$E[Tr_S]$	The average transport throughput of a single flow when it is scheduled

$E[Tr_S\|k]$	The average bit-meter product of a single flow under $k-1$ interferers
$E[T_S\|k]$	The average transmission rate of a single flow under $k-1$ interferers
$E[T_S]$	The average throughput of a single flow when it is scheduled
$E[W_i]$	The average backoff time of user i
$f_S(s)$	The pdf of the Nakagami channel gain
$f_X(x)$	The pdf of the distance of two randomly deployed user in a two dimensional space
$f_Z(z)$	The pdf of the distance of two randomly deployed user in a three dimensional space
G_0	The cross correlation between two concurrent flows
$G_R(i)$	The antenna gain of the receiver of flow f_i
G_{R_M}	The receiver antenna gain within the beamwidth
G_{R_S}	The receiver antenna gain outside of the beamwidth
$G_T(i)$	The antenna gain of the sender of flow f_i
G_{T_M}	The transmitter antenna gain within the beamwidth
G_{T_S}	The transmitter antenna gain outside of the beamwidth
H_i	MIMO channel matrix
$h_i^{p,q}$	The channel gain from the q-th antenna at the AP to the p-th antenna at user i
$I_{j,i}$	the interference power of the sender of flow f_j to the receiver of flow f_i
K	The number of users that AP communicates in the downlink simultaneously
l	The average number of contending users
L	Room length
L_i	The payload of a frame
M	The number of antennas employed by the AP
M_i	The number of antennas employed by user i
m	Retranmission limit
N	The number of users in a wireless network
N_B	The maximum number of busy slots
N_f	The number of flows in a wireless network
N_0	One-sided power spectral density of additive white Gaussian noise
n_f	The number of failed requests

N_{min}	The lower bound of the number of flows
n_s	The number of successful requests
p_c	The probability of a collision
p_i	The probability of a collision seen by a frame being transmitted by user i
P_i	The transmission power level of node X_i
$p_i[T]$	The probability that user i transmits a frame in a randomly chosen slot
$p\|k$	The transmission probability of a user at a slot given k contending users
$P(k,n)$	The probability that k flows can concurrently transmit, after checking the first n flows
P_n	The ambient noise power level
P_{OH}	The transmission time of PHY overheads
$P_R(i)$	The received signal power of flow f_i
p_s	The probability of a successful transmission
P_T	The maximum transmission power
Q	The probability of a transmitter lying outside an ER of a receiver
$q_i(t)$	The number of frames to user i at time t
R_0	Basic transmission rate
$r(i)$	Exclusive radius of the receiver of flow f_i
$R_{i,1}^U$	The transmission rate of user i with a spatial diversity scheme
R_i^C	Achieved throughput if all flows transmit simultaneously
R_i^D	The average transmission rate of the AP to user i
R_i^T	Achieved throughput if all flows transmit in a TDMA fashion
$R_{i,2}^U$	The transmission rate of user i with a spatial multiplexing scheme
S	The area of the room
s_i	The data intended for user i
S_i	Fast fading effect over the transmission link of flow f_i
$SINR\|k$	The receiver SINR of a single flow given k concurrent flows
$S\{N_f\}$	The set of scheduled flows
T_a	Transmission time of a preamble

$T_a(j)$	A scheduling weight associate with flow j
T_{ACK}	The transmission time of an ACK frame
T_{beacon}	The transmission time of a beacon frame
T_{busy}	The duration of a busy slot
T_{CAP}	The duration of the CAP
T_{CD}	The collision time of a downlink frame
T_{c_i}	A collision time of a frame transmitted by user i
$\overline{T_{c_i}}$	The average collision time of a frame transmitted by user i
T_{CTA}	The time unit for a channel time allocation
T_{CU}	The collision times of an uplink frame
T_D	The transmission time of a downlink frame
T_{MF}	The transmission time of MAC payload
T_{sf}	The duration of a superframe
T_{s_i}	A frame transmission time of user i
T_{slot}	The duration of an idle slot
T_U	The transmission time of an uplink frame
V_i	the interference distance between the ith interferer and a receiver
V_i	Additive Gaussian noise vector
W	Signal bandwidth
W_i	Beamforming vector
X_i	Node i
Z'	The random variable denoting the effective transmission distance of a flow

Chapter 1

Introduction

Motivated by the rapid evolution of wireless communication technologies and the ever-growing demands of multimedia services, a diversity of wireless networks have been emerging in the last decade, e.g., cellular networks, wireless metropolitan area networks (WMAN), IEEE 802.11 wireless local area network (WLANs), Bluetooth, Ultra-wideband (UWB), and millimeter wave (mmWave) based wireless personal area networks (WPANs) and wireless body area networks (WBANs), etc. While the network architecture, service scenarios, and underlying communication technologies make these networks fundamentally different from each other, these networks will co-exist and operate synergistically together in the next generation wireless networks to enable ubiquitous broadband wireless access anywhere anytime.

Wireless medium access control (MAC) defines a set of rules for multiple users to effectively and fairly share the radio resources [2]. As wireless communication is broadcast in nature, interference resulting from uncoordinated transmissions limits the link capacity or even cause link breakage, which degrades the performance of the network and individual users. Therefore, MAC design in wireless networks plays an important role in efficient resource utilization and quality of service (QoS) provisioning. Although wireless MAC has been heavily pursued in the literature, there still exists considerable

room for improving MAC protocol performance to meet the ever-increasing demands of broadband multimedia applications, as the emerging communication technologies bring both great opportunities and challenges in the MAC design.

1.1 Research Challenges and Motivations

It is known that a wireless channel exhibits characteristics different from wireline because of the hostile wireless fading environment. Due to the radio propagation property, e.g., multipath fading, signal reflection and wave dispersion, the achievable channel capacity is generally very limited and varies over time. With the recent advances in wireless communications, next generation wireless networks will employ various physical layer techniques, e.g., multiple input multiple output (MIMO) beamforming, directional antenna, etc., to improve the link capacity and reliability. On the other hand, the enhanced physical link capability opens a door for high rate multimedia services, which are usually infeasible in traditional low rate wireless networks. To explore efficient MAC design in next generation broadband wireless networks in support of multimedia services, we first present the challenging research issues in the MAC design as follows.

1. Capacity and fairness: The proliferation of consumer electronics and popularity of multimedia services lead to the ever-growing density of wireless networks. In addition, wireless users may create and exchange large volumes of rich multimedia information locally and remotely. Therefore, it is of critical importance to fully utilize the wireless resources to push up the limit of the number of users that can share the available wireless resources. Moreover, it is recognized that the conventional network resource allocation strategy that maximizes the network capacity is generally unfair because resources will always be allocated to the user that can best exploit it [3]. As fairness is another important performance metric in MAC design, a desirable MAC should well balance the capacity and fairness

performance, e.g., maximize the network capacity while maintain a good fairness performance as well.

2. QoS provisioning: Next generation wireless networks are envisioned to support a variety of multimedia applications. Various multimedia applications have different QoS requirements and evaluation criteria [4, 5]. For example, interactive media such as voice calls and video conferencing are very delay-sensitive while video streaming and bulky data transfer are more bandwidth-intensive but can tolerate a certain level of startup delay. Therefore, QoS aware MAC design along with admission control are required to assure the QoS performance of multimedia applications.

3. Infrastructure support: Some wireless networks usually operate in an infrastructure mode, such as a WLAN with an access point (AP) and a WPAN with a piconet coordinator (PNC). However, infrastructure support may not always be available due to user failure or mobility, and the network scaling. In this case, a simple but robust MAC with distributed function is more desirable. It is known that the widely deployed IEEE 802.11 Distributed Coordination Function (DCF) does not scale well and suffer from hidden/exposed terminal problems in a multi-hop network. How to improve the efficiency of distributed MAC in a densely deployed multi-hop wireless network needs further investigation.

4. Peer to peer communications: To improve the transmission efficiency, users in a WPAN can communicate with each other via direct link peer to peer transmissions. In this case, the central controller is no longer the performance bottleneck as the AP in an infrastructure WLAN. However, it is difficult to collect the channel condition information between peer to peer links for efficient MAC layer resource allocation. Thus, scheduling of peer to peer transmissions is much more challenging than that of uplink or downlink transmissions.

5. Multiple-antenna system: The next generation IEEE 802.11n WLAN uses MIMO technology to improve communication performance. The emerging mmWave WPAN will also employ directional antennas with a higher directivity gain to combat high path loss in the 60 GHz mmWave channel. On one hand, the use of multiple antennas increases the antenna operation complexity. On the other hand, multiple antennas provides great potential for efficient space utilization of wireless channels. Thus, we should investigate the impacts of multiple antenna techniques on the MAC protocol design and explore low complexity scheduling algorithms for efficient spatial division multiple access, considering different network scenarios and communication features of the wireless access networks.

As a summary, efficient MAC design in next generation broadband access networks is still fraught with many fundamental challenges because of the ever-increasing user density, diverse QoS requirements, and various network communication scenarios, etc. The main objective of this dissertation is to address the above research issues in the design and analysis of MAC protocols in different wireless networks, including single- and multi-hop networks, with and without infrastructure support, employing a single and multiple antenna system. Towards this goal, we will 1) analytically study the existing MAC protocols and identify their performance bottlenecks and constraints; 2) explore the salient features of the emerging communication technologies, i.e., multiple antenna system, UWB, and mmWave, etc., in the MAC protocol design to mitigate the identified problems; and 3) maximize the network capacity by analyzing the MAC protocols and fine tuning protocol parameters.

1.2 Research Contributions

The main contributions of this dissertation are listed as follows.

- Development of a generic analytical framework for the performance study of IEEE

802.11 DCF MAC in support of non-persistent asymmetric multimedia traffic flows [6]. The obtained voice and video capacity can be applied for effective call admission control to maintain QoS performance of voice and video connections in a WLAN. Different enhanced MAC mechanisms, e.g., aggregation and bidirectional transmission, are also compared [7].

- Proposal of a distributed MIMO-aware multi-user MAC in a WLAN with MIMO capability [8]. By exploiting multi-user MIMO beamforming in the downlink, the proposed MAC effectively mitigates the AP bottleneck effect in an infrastructure based WLAN.

- Analytical study of IEEE 802.15.3 MAC [9]; the intrinsic relationship among protocol parameters is disclosed, which provides important guideline for parameter setting for a hybrid MAC.

- Proposal of a novel ER based MAC for a single hop UWB/mmWave WPANs with omni- and directional antennas [10]. By efficiently exploiting the spatial reuse opportunities in a UWB/mmWave channel, the proposed MAC achieves high network capacity and maintain good fairness among multiple users as well.

- Development of an analytical framework to study the spatial multiplexing capacity of a UWB/mmWave WPAN with random topology, considering the physical layer rate adaptation and different types of antennas [11, 12]. The analytical framework can be applied to other distributed MAC protocols in general. The analytical results reveal the main factors affecting the network (transport) capacity, and the best protocol parameters, e.g., exclusive region size, can be determined toward the maximum network capacity.

- Proposal of a distributed asynchronous ER based MAC for a multi-hop UWB network [13]. Appropriate ER size is analytically derived to maximize the network

capacity. The proposed MAC significantly outperforms IEEE 802.11 DCF in terms of network throughput, transport throughput, and fairness. In addition, the distributed and asynchronous nature makes it robust and scalable in a densely deployed multi-hop network.

1.3 Outline of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 reviews the background of wireless MAC protocols. Chapter 3 describes the system model under consideration. In Chapter 4, we first study MAC protocol design and analysis in WLANs. We analyze the MAC performance of legacy IEEE 802.11 DCF and some enhanced mechanisms. A MIMO-aware multi-user MAC in WLANs with MIMO capability is also proposed and analyzed. In Chapter 5, we investigate the UWB and mmWave characteristics and propose a novel ER based MAC in centralized WPANs. Theoretical network capacity of a WPAN is analyzed and the best ER size is derived. In Chapter 6, we design a distributed asynchronous ER based MAC for a densely deployed multi-hop UWB network. Finally, Chapter 7 gives concluding remarks and outlines future work.

Chapter 2

Overview of Wireless MAC Protocols

2.1 Wireless MAC Classification

Generally, wireless MAC protocols can be classified into three categories [2]: random access, guaranteed access and hybrid access protocols.

2.1.1 Contention-based Random Access

Random access protocols, such as pure Aloha, slotted Aloha, carrier sense multiple access with collision avoidance (CSMA/CA), and non/p/1-persistent CSMA, *etc.*, are contention-based protocols that can operate in either infrastructure-based wireless networks or infrastructureless (*ad hoc*) networks. As collision detection is not available in wireless networks, usually a random backoff procedure, e.g., binary exponential backoff, is applied to reduce potential collisions. The main advantage of the random access protocol is that it operates in a distributed manner and thus does not require a centralized controller. In addition, synchronization among users is not a necessity and thus it is relatively simple for implementation. Asynchronous data transmissions has been a

key factor to the overwhelming success of the IEEE 802.11 WLANs. However, as collision is inevitable and the collision resolution overhead degrades the network resource utilization, random access provides no satisfactory QoS guarantee.

2.1.2 Contention-free MAC

Unlike random access protocols, guaranteed access are contention-free protocols in which wireless users access the medium in an orderly manner, e.g., in a round-robin fashion. Contention-free access can also be divided into two sub-categories: Channel partitioning access and turn-taking access. In the channel partitioning access, channel access can be partitioned in the time domain, e.g., time division multiple access (TDMA), frequency domain, e.g., frequency division multiple access (FDMA), and code domain, e.g., code division multiple access (CDMA), or any combinations of them. For example, wideband-CDMA combines the frequency domain and code domain channel partitioning for multiple access. Turn-taking protocols includes polling based and token passing based protocols. Polling based scheme is usually used in a centralized network. The controller polls other mobile users according to a pre-defined scheduling scheme and the user can transmit without contention only when it is polled by the controller [14]. In a token based protocol, e.g., wireless token ring protocol [15, 16], the ring owner determines the transmission order and the maximum transmission time of each user. A user can transmit when he holds the token and the token has to be passed to the next user when the maximum time is reached. Therefore, contention-free MAC can provide a certain level of QoS in terms of bounded access delay and reserved bandwidth.

2.1.3 Hybrid MAC

Hybrid MAC protocol operates in a request-grant fashion. In a hybrid MAC, each user sends a request to the network controller indicating how much time or bandwidth is

required for data transmission using a random access protocol. Based on the received requests, the controller allocates time slots and sends grants to the requesting users indicating the start time and the duration of data transmissions. Users can transmit data in the reserved time duration without contentions. Thus, it is also called contention based reservation access. Hybrid protocols combine the best qualities of the random access and guaranteed access protocols to achieve flexibility, efficiency and QoS provisioning [2].

2.2 MAC Standards in Wireless Local/Personal Area Networks

2.2.1 IEEE 802.11 MAC

The IEEE 802.11 standard defines two modes in the MAC protocol: the mandatory DCF mode and the optional Point Coordination Function (PCF) mode. Although the PCF mode is designed for real-time traffic [17, 18], it is not widely deployed due to its inefficient polling schemes, limited QoS provisioning, and implementation complexity[1].

The IEEE 802.11 DCF uses the CSMA/CA mechanism [14]. A user monitors the medium before attempting a transmission. If the medium is sensed busy, the user defers transmission until the medium is sensed idle for a period of time equal to a DCF InterFrame Space (DIFS). After a DIFS medium idle time, the user enters the backoff phase and sets a random backoff counter randomly chosen from $[0, CW)$, where CW is the contention window (CW) size. The backoff counter decreases by one for every time slot if the medium is idle; otherwise, the counter freezes, and the decrement resumes after the medium is sensed idle again for a DIFS. When the backoff counter reaches

[1]Since both DCF and PCF have limited support for real-time applications, the IEEE 802.11e has been proposed to enhance the current 802.11 MAC to support applications with stringent QoS requirements [19].

zero, the user transmits the frame. If another user transmits a frame at the same time, a collision occurs and both transmissions fail. After an unsuccessful transmission, CW is doubled until it reaches the maximum value ($CW_{\max}$), and the sender reschedules the transmission by randomly choosing a backoff counter in $[0, CW)$. The frame is dropped when the retransmission limit is reached. After a successful transmission, CW is reset to its minimum value ($CW_{\min}$). Upon receiving a frame successfully, the receiver transmits an acknowledgment (ACK) following a Short InterFrame Space (SIFS). Two medium access techniques are specified in DCF: the basic access mechanism and the request to send/clear to send (RTS/CTS) mechanism. Frames are transmitted using the RTS/CTS mechanism if their payload exceeds a given threshold; otherwise, the basic access is used.

With the IEEE 802.11 DCF-based MAC, all users have the same priority to access the channel. This is unfavorable to the AP which usually has much higher traffic loads. In addition, the CSMA/CA mechanism was originally designed for asynchronous data transmission, without considering delay-sensitive multimedia traffic. Before being successfully transmitted, each frame has to wait a random time period which depends on the network load and collisions it experienced. A high collision probability reduces the frame service rate and accentuates the queue length and delay. On the other hand, in order to provide satisfactory QoS for multimedia services over contention based IEEE 802.11 MAC, it is essential to quantify the network capacity for effective admission control so that the network contentions can be controlled at a certain level to assure the delay performance of WLAN users.

2.2.2 IEEE 802.15.3 MAC

IEEE 802.15.3 standard uses a hybrid MAC protocol, with random access periods for network initiation/association and resource allocation, and contention-free periods for data transmissions. Multiple users can autonomously form a piconet in which one

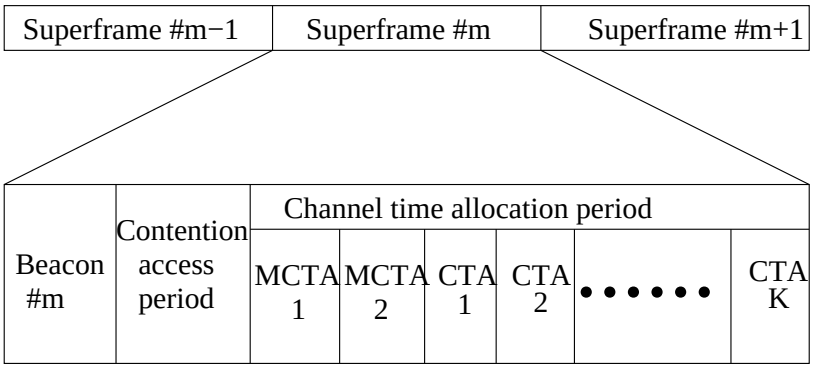


Figure 2.1: 802.15.3 Superframe

of them is selected as the PNC. The PNC can collect global user information in the piconet and allocate radio resources or schedule channel times to all user in the piconet according to their requirements. Based on the scheduling, all devices can communicate in a peer-to-peer fashion. Such a semi-ad hoc setting can provide better QoS than a pure ad hoc network.

Timing in IEEE 802.15.3 is based on the superframe structure as illustrated in Fig. 2.1. IEEE 802.15.3 defines three methods for data transmissions between wireless users [20]: a) transmitting asynchronous data or communication commands in the contention access period (CAP); b) allocating channel time for isochronous streams in the channel time allocations period (CTAP); and c) allocating asynchronous channel time in the CTAP. Although both commands and asynchronous data can be transmitted in the CAP, it is recommended that only commands be transmitted to minimize the length of contention period. This is desirable for reducing the protocol overheads and inefficiency resulting from the collision avoidance mechanism used for random access. In addition, for multimedia applications, users may need channel times on a regular basis, and they send channel time requests during the CAP to reserve isochronous

channel time in the following CTAP. Based on the successfully received requests from all users, the PNC will schedule and allocate channel time in the CTAP in a TDMA manner.

IEEE 802.15.3 MAC requires a centralized network architecture and suffers from single-point-of-failure problem. Thus, IEEE 802.15.3 is more suitable for a network where a central controller is usually available and network topology is relatively stable. In public hotspots, a large number of wireless users may join and leave the network at any time and the network topology changes dynamically. If the current PNC disappears (e.g., powers off or moves away), it may take several seconds before the remaining users reorganize and re-elect a new PNC [21]. In addition, when multiple piconets are used in public hotspots to extend the communication coverage, it is very challenging and costly to manage the inter-piconet interference which may degrade the network performance significantly.

2.2.3 WiMedia MAC

To address the aforementioned problems in IEEE 802.15.3 MAC, the WiMedia alliance has specified a distributed MAC based on Multi-band Orthogonal Frequency Division Multiplexing (MB-OFDM) UWB [22]. Similar to IEEE 802.15.3, timing in WiMedia MAC is based on the slotted superframe structure, as shown in Fig.2.2. Each superframe starts with a beacon period (BP). A user first senses the UWB channels for several superframes and selects a channel for communication if a beacon is received; otherwise, the user will select a channel and initiate a BP by itself. All users communicating in the same channel will collect the beacons from their neighbors and pick up unoccupied beacon slots to transmit their own beacon frames. Data frames are transmitted during the data transmission period (DTP), which consists of multiple medium access slots (MASs). During the DTP, users can access the channel in an asynchronous manner via the prioritized channel access (PCA) protocol, which is sim-

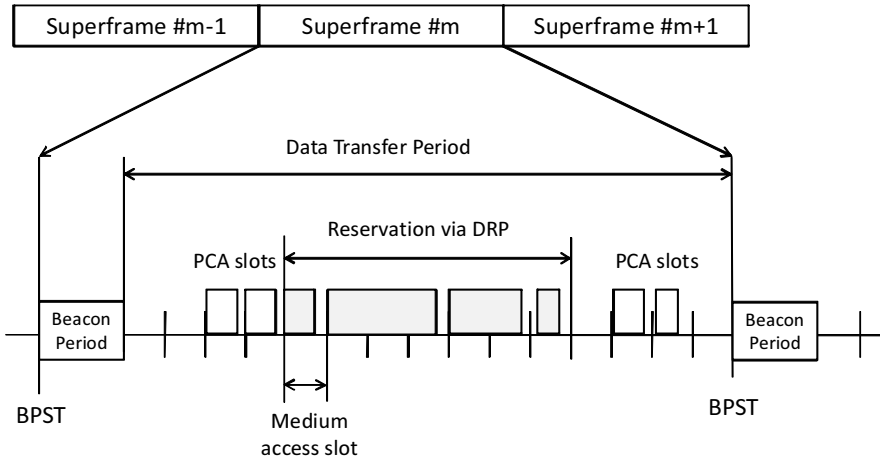


Figure 2.2: WiMeda MAC Superframe Structure

ilar to the enhanced distributed channel access (EDCA) specified in IEEE 802.11e. The basic difference between PCA and EDCA lies in the physical layer. Due to the low power level of UWB signals, PCA uses preamble sensing instead of energy detection based carrier sensing in EDCA. Users carrying isochronous traffic can also reserve multiple MASs for contention-free channel access via a distributed reservation protocol (DRP). A device first sends a reservation request to the receiver either in the beacon or using DRP or PCA. The receiver analyzes the channel time utilization of its neighbors and responds to the sender. If the requested MASs are not available, the receiver will provide additional information (e.g., available MASs in its beacon group) to the sender. Otherwise, a successful reservation is announced in the beacons so that other users within the transmission range become aware of the reservation and defer their channel access during that period.

WiMedia MAC is still a TDMA based MAC and requires channel time synchronization among users, i.e., all users communicating in the same channel need to synchronize the BP starting time with each other. However, synchronization is difficult and costly in a multi-hop network. In a densely deployed multi-hop network where multiple bea-

con groups may overlap with each other, merging different BPs into one single common BP is not a trivial task. Another problem is when a burst of users join in the network during one superframe, it is very likely that two or more users may select the same beacon slot which causes beacon collisions. A user can only determine a beacon collision or transmission error if its own address is not included in its neighbors' beacons for multiple continuous superframes. In other words, it may take hundreds of milliseconds for a user to detect a beacon collision.

2.3 Summary

Due to the low resource utilization and lack of QoS guarantee in the random access protocols and the inflexibility of guaranteed access protocols, we anticipate a hybrid MAC will be deployed in next generation wireless networks with infrastructure support for providing satisfactory QoS performance for delay-sensitive multimedia applications; while a robust asynchronous random access protocol is more desirable for distributed wireless networks without infrastructure support.

Chapter 3

System Model

Generally, wireless networks can be classified into different categories. According to the communication coverage, wireless networks can be grouped into wireless body, personal, local, metropolitan, and wide area networks. Based on the network architecture, wireless networks can be divided into infrastructure-based and infrastructureless networks. Based on the network scale, wireless networks can also be categorized into single hop and multi-hop networks. Usually, an infrastructure-based network is a single hop network in which a base station or AP serves as the portal to a network infrastructure so that wireless users in its coverage can access a backbone network such as the Internet for global communication. An infrastructureless network can be either a single hop or a multi-hop network which autonomously operates in an ad hoc mode without a central controller. Since current backbone network is rapidly upgraded to terabit speed, the last-mile wireless access networks, e.g., WLANs and WPANs, are usually considered as the bottleneck for achieving the full potential of high speed Internet access. Therefore, in this study, we focus on high rate wireless personal and local area networks in support of various multimedia applications, e.g., video, voice, and data. We study efficient MAC protocol design for both single-hop and multi-hop networks with and without network infrastructure.

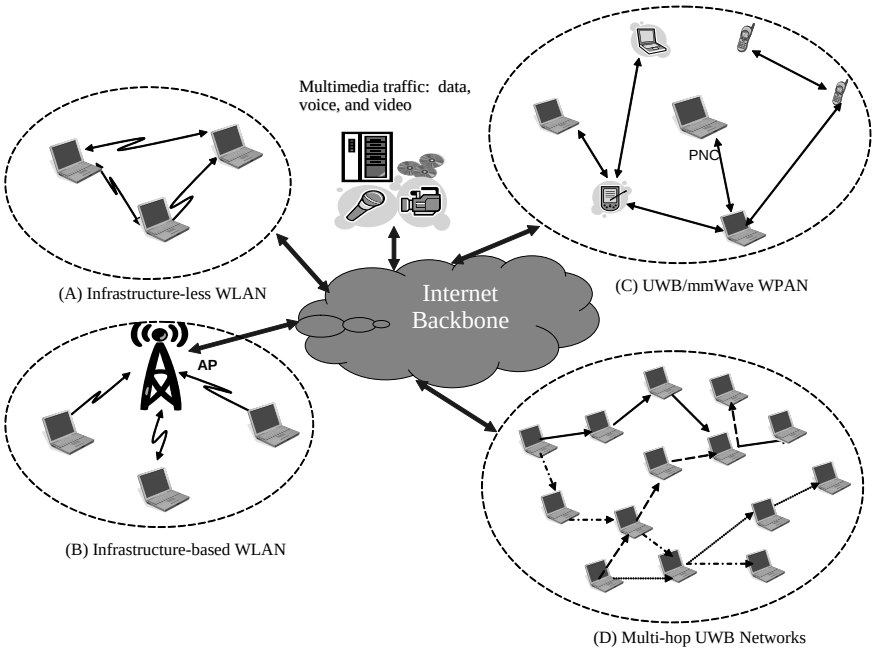


Figure 3.1: System Model

3.1 Network Model

3.1.1 Wireless Local Area Network

As one of the most successful wireless networks deployed all over the world, a WLAN offers a quick and effective wireless extension to the wired Ethernet. A typical WLAN usually covers a small geographical area and is a single hop network. For example, in an open indoor office environment, an IEEE 802.11b device using 100 mW transmission power and 2.2 dBi gain diversity dipole antenna can transmit at 11 Mbps within 48 m; an IEEE 802.11g device using 30 mW transmission power and the same type of antenna can transmit at 54 Mbps within 27 m.

We consider a single hop fully connected WLAN with N users, which can operate in both ad hoc mode and infrastructure mode, as shown in Fig. 3.1 (A) and (B),

respectively. In an ad hoc WLAN, users communicate with each other in a peer-to-peer manner; while in an infrastructure-based WLAN, all the traffic to and from the WLAN go through the AP. We use the mandatory IEEE 802.11 DCF MAC that all N users, including the AP if an infrastructure WLAN is considered, contend for channel access. Every user has a half-duplex transceiver thus can not receive and transmit simultaneously. An omni-directional antenna is used for both transmitting and receiving. All users can sense the status of the shared wireless channel and are well synchronized with each other. Unless otherwise specified, we assume an ideal wireless channel such that all transmitted frames can be received error-free if there is no collision. The performance of a single hop WLAN will be studied in Chapter 4.

3.1.2 UWB/mmWave based Wireless Personal Area Network

UWB communication technology can achieve very high data rate at short distance and is considered one of the best candidates for high rate WPANs. The first approved commercial UWB system operates in $3.1-10.6$ GHz frequency band. Several physical layer have been proposed for the UWB system: continuous wave UWB (C-UWB), direct sequence UWB (DS-UWB), and MB-OFDM UWB. C-UWB uses bursts of pulses and variable spreading codes to trade data rate for communication range and is specified in IEEE 802.15.4a for low rate WPANs [23]. DS-UWB and MB-OFDM are two physical specifications for high rate WPANs. DS-UWB is based on direct sequence spread spectrum (DSSS) technology and MB-OFDM uses a combination of frequency hopping and OFDM technologies. To allow a UWB system to co-exist with other systems, e.g., WiFi, cellular, the FCC power spectral density emission limit for UWB devices is -41.3 dBm/MHz. Because of the stringent power constraint and the wide bandwidth in UWB communications, normally the UWB transmission power can not be adjusted, and spreading technologies in both the time domain and the frequency domain are used to vary the data rates [24]. The inherent characteristics of spreading technology make

UWB systems immune to interference. Thus, different from narrowband systems (e.g., WiFi) where simultaneous transmissions from nearby neighbors causes severe interference to each other, multiple concurrent transmissions are possible in UWB systems if and only if the multi-user interference is properly managed. In addition, UWB can provide high precision ranging which facilitate location-aware applications and protocol design in UWB networks [25, 26].

Another UWB system operates in $57-64$ GHz frequency band, i.e., mmWave frequency band. In this dissertation, we denote UWB as UWB operating in $3.1-10.6$ GHz frequency band, and mmWave as UWB operating in $57-64$ GHz frequency band. Because of the unique characteristics of 60 GHz channel, i.e., high path loss due to oxygen absorption and atmospheric attenuation, and the stringent power limitation for wireless devices, it is highly desirable to use directional antenna to achieve high directivity and diversity gains. On the other hand, the size of the antenna used for mmWave band could be very small, and thus it is feasible to deploy a directional antenna with multiple antenna elements in a single device. For large bandwidth mmWave communications supporting extremely high data rates, the negative impacts of intersymbol interference (ISI) due to multipath propagation become significant. Orthogonal frequency-division multiplexing (OFDM) signals are intrinsically multipath robust due to the low symbol rate in each of the subcarriers, so OFDM will be a good candidate for mmWave communications[1]. For RF oscillators at mmWave spectrum, it is very difficult if not impossible to maintain a low level phase noise, which affects the signal in the frequency conversion operations, and results in higher bit error rate (BER) for effective communications. Different multiple access techniques, including OFDM/TDMA, direct sequence (DS)-CDMA, Multi Carrier (MC)-CDMA, and MC-DS-CDMA, have different sensitivities to phase noise. Among these techniques, MC-DS-CDMA is the most robust against phase noise and multiple access interference (MAI) in 60 GHz wireless channel [27].

[1]Although we use OFDM in our system model, our work is independent of any particular modulation schemes.

In this dissertation, we consider a dense UWB-based single hop WPAN with the piconet as the basic network element, as shown in Fig. 3.1 (C)[2]. Wireless users are randomly distributed over a square room. Based on IEEE 802.15.3 MAC specification for high rate WPANs introduced in Chapter 2.2.2, the PNC provides the basic timing and schedules peer to peer communications between wireless users. Spreading techniques are applied for UWB in both time and frequency bands, e.g., DSSS for UWB and MC-DS-CDMA for mmWave. Denote G_0 as the cross correlation between two (pseudo-random) spreading codes used by two different flows. In the special case of $G_0 = 1$, two flows use a common speading code for transmissions. All senders use the maximum transmission power and the data rate is adapted according to the received signal-to-interference-plus-noise ratio (SINR). Taking advantage of the ranging service in UWB, users can determine their locations by measuring the beacon signals and report the location information to the PNC. Thus, the PNC has the topology information of the network. We use omni-directional antenna for UWB and directional antenna for mmWave networks. For directional antenna, we first employ an ideal flat-top model [28, 29] to study the capacity of a UWB/mmWave WPAN, assuming the antenna gain is constant within the beamwidth θ, $G_m = 2\pi/\theta$, and zero outside the beamwidth, $G_s = 0$. We also employ a simplified cone plus circle model in a two-dimensional plane to consider the sidelobe effect [30], which defines the antenna gains of the mainlobe and sidelobe as $G_m = \varsigma\frac{2\pi}{\theta}$ and $G_s = (1-\varsigma)\frac{2\pi}{2\pi-\theta}$, respectively, where ς is the antenna radiation efficiency.

3.1.3 Multi-hop Wireless Network

Due to the limited transmission range of UWB communications, multi-hop relay is necessary to meet the escalating demands for high rate wireless connections anywhere and

[2]The high dense network scenarios can be in future stock exchange rooms, conference rooms, shipping malls, expo rooms, *etc.*

at any time. We consider a dense multi-hop UWB network with no central controller for time synchronization and transmission scheduling. It is found in [31, 32] that the deployment of directional antenna is not suitable for a fully distributed system because the increased hidden terminal/beam, deafness, and 'ACK suicide' problems, etc., which could significantly degrade the network performance. Therefore, we use omni-directional antenna for both transmitting and receiving in a distributed multi-hop UWB network[3]. Similar to the UWB WPAN model in Section 3.1.2, all senders use the maximum transmission power and the data rate is adapted according to the received SINR. We assume all users share a pool of spreading codes, one of which is chosen for control message exchange and others are used for data transmissions. Different data flows can use different spreading codes to reduce the mutual interference level among concurrent transmissions. Asynchronous distributed MAC design for a multi-hop UWB network will be discussed in Chapter 6.

3.2 Traffic Model

Although original WLAN applications are mainly data centric, there is a growing demand for multimedia applications over WLAN, e.g., voice over IP (VoIP) and video conferencing. Recent advances in UWB communications further open a door for emerging killer multimedia applications such as Internet Protocol TV (IPTV). Compared with best effort data traffic, multimedia traffic usually have stringent delay and bandwidth requirements. To provide better services to wireless users, we need to theoretically study the network capacity in support of these multimedia applications, where the network capacity is defined as the maximum number of multimedia flows that can be supported in a WLAN/WPAN with satisfactory user-perceived quality during any one use. The theoretical study of network capacity is of critical importance for effective

[3]For UWB operating in mmWave band, it is also possible if each node is equipped with a number of beams and switches on/off all beams to work in an omni-directional mode.

Table 3.1: Frequently Used Voice Codecs

Voice Codec		G.711	G.723a	G.729	iLBC
Codec Bit Rate		(64Kbps)	(5.3/6.3Kbps)	(8Kbps)	(15.2/13.3Kbps)
Sample Period	Arrival Rate (frames/sec)	Payload (Byte)	Payload (Byte)	Payload (Byte)	Payload (Byte)
10ms	100	80		10	
20ms	50	160		20	38
30ms	33.33	240	20/24	30	50
40ms	25	320		40	
50ms	20	400		50	
60ms	16.67	480	40/48	60	

admission control and scheduling to guarantee the QoS performance of multimedia applications.

As an IEEE 802.11b/n based WLAN can provide from tens to hundreds of Mbps data rate, we consider both voice and low/medium rate video applications over WLAN. UWB and mmWave based WPANs can achieve up to several Gbps data rate and is considered an ideal candidate for last meter high definition (HD) IPTV distribution [33]. Thus, we deploy high volume data and video traffic in a UWB/mmWave based WPAN.

- Voice: VoIP application is one of the fastest growing Internet applications. In a VoIP system, analogue voice signals are first digitized, compressed, and encoded into voice streams, and then packetized into constant-bit-rate (CBR) flows. We use CBR voice model because of the following three reasons: 1) many voice codecs do not use silence suppression; 2) if silence suppression is used and voice traffic exhibits on-off characteristics, a tighter bound derived is robust in the worst case when all voice flows are in the "on" state; and 3) even if silence suppression is used,

Table 3.2: Levels in H.264/MPEG-4 AVC [1]

Level Number	video bit rate (bps)	resolution & frame rate(fps)	average frame payload (bytes)
1b	128k	128x96 & 15	1067
1.2	384k	320x240&20	2400
1.3	768k	352x288&30	3200
2	2M	352x288&30	8333
		320x240&36	6944

some packets are still transmitted intermittently during "off" period to maintain a better voice quality [34]. The main attributes of some commonly used voice codecs with different packetization intervals are tabulated in Table 3.1. Different codecs use different compression algorithms resulting in different bit rates. For example, G.711 codec provides high quality voice with low compression ratio and high bit rate of 64 Kbps, while iLBC is a very bandwidth efficient codec with a bit rate of 13.3Kbps and has been used in Internet soft-phone applications, e.g., skype.

A two-way conversation is very sensitive to packet delay and jitter, but it can tolerate certain degree of packet losses. A playout buffer at the receiver can efficiently remove the delay jitter and thus we only consider the delay and packet loss in our study. According to International Telecommunication Union (ITU) standards, the one-way end-to-end delay of voice traffic should be no greater than 150 ms for good voice quality and up to 400 ms for acceptable voice quality, with an echo canceller [35]. Also, usually the packet loss rate should be no more than 1% to maintain satisfactory voice quality.

- Video: With the advances of wireless communication technologies, video streaming with stringent QoS requirements becomes possible to deliver in high performance WLANs and UWB networks. For example, video telephony and video conferencing are in high demand in WLANs for various applications such as telecommuting, telemedicine, e-training and e-learning, etc. IPTV with high definition content is an emerging killer application in multi-Gigabit UWB WPANs. Video traffic is usually modelled as a variable bit rate (VBR) flow with different compression ratios and various payload formats in the codec. In this study, we consider H.264 codec, which supports very efficient video compression and is applied in a broad range of video applications from low rate Internet video streaming to high definition video (HDV). H.264 defines 16 different levels, tied mainly to the picture size and frame rate [1]. Some examples for various resolution, frame rate, and maximum compressed video rate in five levels are listed in Table 3.2 for video capacity evaluation. The level 2 supports up to 2 Mbps video rate, with the frame rate of 30 frames per second (fps) at the frame resolution of 320×240 pixels, or with a higher frame rate of 36 fps at a lower resolution of 352×288 pixels. Higher resolution provides better image quality and higher frame rate results in a smoother motion video. The emitted frames are then encapsulated into RTP packets for transmission, depending on the maximum transmission unit (MTU) of underlying layers.

  The general QoS metrics of video applications include throughput, delay, jitter, and packet loss. The throughput demands vary among different video flows. We consider low/medium rate video telephony applications in WLANs and high rate video distribution in UWB WPANs. Similar to voice traffic, we do not consider jitter because a playout buffer at the receiver can efficiently remove the delay variations. For interactive applications, e.g., video telephony, the normal tolerable delay should be less than 100 ms and packet loss rate should be below

1%.

- Data: There are many data applications including email, web browsing, file transfer, etc. We also adopt a simple saturated data model in the performance study of wireless networks, i.e., a user always has a frame in the queue ready for transmission. The saturated data model is applicable for large volume bulk data transfer applications.

  Data traffic is usually delay-insensitive, but require no transmission error. Transmission errors can be improved by link layer automatic repeat request (ARQ) and transport layer reliable transmission control protocol, etc. In this dissertation, we evaluate the QoS metrics for data traffic in terms of throughput and fairness. We aim to design a MAC that allow multiple data flows to fairly and efficiently share the wireless resources, in both single hop and multi-hop networks.

Chapter 4

MAC Protocol in Wireless Local Area Networks

As IEEE 802.11 WLANs are one of the most successful wireless networks which have been widely deployed all over the world, we first study the performance of a WLAN using IEEE 802.11 MAC. By developing an analytical framework to study the network capacity in terms of the maximum number of multimedia flows that can be supported with satisfactory user-perceived quality, we point out the drawbacks and performance constraints of the existing MAC [6]. Some enhanced MAC mechanisms to mitigate the performance bottleneck are investigated [7]. In response to the escalating demands for high throughput WLANs with the rise of broadband multimedia applications, the next generation WLAN will employ MIMO technology in the physical layer to improve the capacity and reliability of the wireless channel. A distributed MIMO-aware multi-user MAC protocol is also proposed to exploit the multi-user degree of freedom in a MIMO system in future WLANs with MIMO capability [8].

4.1 Related Work

Existing performance studies of IEEE 802.11 WLAN can be broadly classified into two categories, saturation and non-saturation cases. The early IEEE 802.11 models in the literature mainly focus on saturation case that a user always has data for transmission. Bianchi [36] develops a bi-dimensional discrete-time Markov chain model to calculate the network throughput as a function of the number of saturated users. Since then, a large number of follow-up papers have appeared. The Markov model is improved in [37] with the consideration of the frame retry limits specified in the standard. The delay performance under different traffic loads is analyzed in [38]. By integrating frame loss in wireless fading channels, the mean throughput and delay are studied in [39]. Furthermore, Bianchi's model is extended to study QoS enabled EDCA in [40, 41]. These works provide important insight into the detail protocol operations. However, saturation assumption simplifies the performance analysis by ignoring the traffic arrival characteristics. In reality, some users, especially those with real-time multimedia traffic, are unsaturated. It is also found in [42] that the maximum throughput of WLAN can be achieved only in the unsaturated user case when there are more than one users sharing the wireless resource. To provide more thorough studies on the performance of IEEE 802.11 MAC, some recent papers relax the saturation assumption to investigate network throughput and delay with unsaturated users. An extended Markov chain based on Bianchi's model is proposed in [43], assuming traffic arrives follow a Poisson process. The channel access delay is studied in [44] where an on-off model is applied for traffic arrivals. A more general model is presented in [45, 46] which accounts for arbitrary arrival patterns. The queue of an unsaturated user is modeled as a discrete time G/G/1 queue and the second-order queueing analysis techniques are applied for service time distribution. Complex mathematical computations are involved in G/G/1 queueing formulations. In addition, to the best of our knowledge, most existing analytical models simply assume a homogeneous network in an independent BSS where all

users have the same traffic arrival patterns. However, different users may carry various applications and have different traffic loads. Moreover, the majority of WLANs are operating in an infrastructure mode, where the AP usually has much higher traffic load and is the bottleneck. In this study, we will develop a generic analytical framework for IEEE 802.11 MAC with saturated or unsaturated users carrying homogeneous or heterogeneous traffic loads. We use the framework to analyze the AP-bottleneck effect in infrastructure-based WLANs, which provides important guidelines for network planning, protocol design and optimization. The generic framework can also be applied to study the capacity region and optimal parameter setting in multi-class WLANs [47, 48].

Some enhanced MAC mechanisms, *i.e.*, frame concatenation (or aggregation) and piggyback are proposed for high throughput WLANs. The two mechanisms are analyzed in [49] under the best-case and saturation scenarios. The saturation throughput of an IEEE 802.11 WLAN using bidirectional frame aggregation is analyzed in [50]. Studies in [49, 50] show that these two mechanisms can greatly improve the network throughput in the saturation case. We further analyze and compare the performance of these enhanced MAC mechanisms in support of non-persistent traffic, e.g., voice and video applications.

The next generation WLANs will employ MIMO technology to improve link capacity and reliability. MAC design for a MIMO-enabled WLAN has drawn great attention recently. In [51], a unified MAC framework for ad hoc networks with smart antennas is proposed, assuming ideal interference cancellation in a closed-loop MIMO system. MAC design and routing issues in mobile ad hoc networks are investigated in [52], focusing on exploiting spatial diversity. A MIMO-DCF protocol is presented in [53], using modified RTS/CTS frames to exchange antenna selection information and exploiting diversity and multiplexing gains. A distributed MIMO-aware MAC is proposed in [54], which is specifically designed for a three element antenna array based MIMO system that allows two simultaneous transmissions in a single collision domain. Due to the size

and cost constraints of antennas, it is likely that AP will be equipped with multiple antenna elements while mobile users can only employ one or a limited number of antennas. Based on this network scenario, a space-division time-division multiple access scheme is proposed in [32]. It is shown that significant throughput enhancements can be achieved at the expense of redesigning the asynchronous IEEE 802.11 MAC to be a synchronous MAC to avoid unwanted interferences. On the other hand, it is well known that MIMO capacity increases linearly with the number of employed antennas at the transmitter and receiver. Some recent results indicate that similar capacity scaling also applies when an AP with multiple antennas communicates with multiple users [55]. Considering a homogeneous network where each users is equipped with M antennas, it is shown in [56] that multi-user transmission has higher link utilization than a single user case due to a large degree of multi-user diversity, using contention-free transmissions. In a centralized network where the channel conditions are usually available at the base station via a feedback channel, multi-user beamforming can provide a substantial gain in downlink throughput [57]. However, channel information is generally not available in a distributed random access network and thus it is difficult to exploit the multi-user diversity gain. To the best of our knowledge, how to efficiently exploit multi-user degree of freedom of a MIMO system to minimize the AP-bottleneck effect in an infrastructure based WLAN remains an open issue. In addition, previous studies either focus on the design of MIMO techniques or the MAC performance analysis. There is little analytical work in the literature which addresses the performance of asynchronous random access MAC in a MIMO system.

4.2 Analytical Framework of IEEE 802.11 DCF

As discussed in Chapter 3.1.1, we consider an IEEE 802.11 DCF-based WLAN with N users. Denote the mean traffic arrival rate and frame service rate of user i as λ_i and μ_i frames per slot, respectively, where $i = 0, 1, ...N-1$. The queue utilization ratio, or

the traffic intensity of user i is $\rho_i = \lambda_i/\mu_i$. Define the conditional collision probability p_i as the probability of a collision seen by a frame being transmitted by user i. The assumption that p_i is constant and independent from the number of retransmissions the frame has experienced [36] is adopted in the analysis. As the probability of three or more simultaneous transmissions is very small, in what follows, we assume that collisions are due to two users transmitting simultaneously.

Define $p_i[T]$ the probability that user i transmits a frame in a randomly chosen slot. Conditional on the queue state, the transmission probability of user i is derived as

$$p_i[T] = p_i[T|QE]p_i[QE] + p_i[T|QNE]p_i[QNE], \tag{4.1}$$

where $p_i[QE]$ and $p_i[QNE]$ are the probabilities of an empty queue and a nonempty queue of user i, respectively. The queue of a user is considered empty when the user is idle, *i.e.,* no frame is in service or waiting for service. A user is idle with probability $1 - \rho_i$ for $\rho_i < 1$. Note that a saturated user always has data for transmission and $\rho_i = 1$. For a unsaturated user, $\rho_i = \lambda_i/\mu_i < 1$. Therefore,

$$p_i[QE] = 1 - \rho_i, \quad \text{and} \quad p_i[QNE] = \rho_i. \tag{4.2}$$

Since a user never transmits with an empty queue, $p_i[T|QE] = 0$. Let $\tau_i = p_i[T|QNE]$ to simplify the notation, the transmission probability of user i is given by

$$p_i[T] = 0 * (1 - \rho_i) + \tau_i * \rho_i = \rho_i\tau_i. \tag{4.3}$$

If user i transmits in a given slot, a collision occurs if at least one of the remaining users also transmits in the same slot. We have

$$p_i = 1 - \prod_{j=0, j\neq i}^{N-1} (1 - p_j[T]) = 1 - \prod_{j=0, j\neq i}^{N-1} (1 - \lambda_j\tau_j/\mu_j), \tag{4.4}$$

where $i = 0, 1, ..., N-1$, and $\rho_i = \lambda_i/\mu_i \leq 1$.

The exponential backoff procedure of CSMA/CA can be modeled as a truncated geometrical random variable, and the average backoff time of user i is derived as

$$\begin{aligned} E[W_i] &= (1-p_i)\frac{CW_0}{2} + p_i(1-p_i)\frac{CW_0+CW_1}{2} + \cdots + p_i^m \frac{\sum_{j=0}^{m} CW_j}{2} \\ &= \sum_{k=0}^{m-1} p_i^k (1-p_i) \sum_{j=0}^{k} \frac{CW_j}{2} + p_i^m \sum_{j=0}^{m} \frac{CW_j}{2}, \end{aligned} \tag{4.5}$$

where CW_j is the contention window in the j-th backoff stage and m is the retransmission limit. During the period of $E[W_i]$, user i makes A_i transmission attempts, which can also be modeled as a truncated geometrical random variable with mean

$$\begin{aligned} E[A_i] &= (1-p_i)\cdot 1 + p_i(1-p_i)\cdot 2 + \cdots + p_i^m \cdot (m+1) \\ &= \sum_{k=0}^{m-1} p_i^k (1-p_i)(k+1) + p_i^m (m+1) = \frac{1-p_i^{m+1}}{1-p_i}. \end{aligned} \tag{4.6}$$

As a user initiates a transmission at the beginning of each slot, the transmission probability τ_i is derived as

$$\tau_i = \frac{E[A_i]}{E[W_i] + E[A_i]}. \tag{4.7}$$

Note that (4.4) and (4.7) yield the same results as Bianchi's Markov model for the saturation case with $\rho_i = 1$.

During $1/\mu_i$, from the time instant that a frame is ready for transmission to the time instant that the frame is successfully transmitted, the following events may occur: 1) a successful transmission by the tagged user i; 2) successful transmissions by the remaining $N-1$ users; 3) collisions; 4) channel idle when user i is in its backoff stage(s).

Denote T_{s_i} as a frame transmission time of user i, and T_{c_i} as the collision time user i experiences each time a collision occurs. In the basic access mode, a frame transmission time includes the transmission of data plus headers encapsulated in each layer, a SIFS, an ACK frame, and a DIFS. A collision duration consists of the time for data frame transmission, the time waiting for ACK timeout, and a DIFS.

$$T_{s_i} = T_{data} + SIFS + T_{ACK} + DIFS \tag{4.8}$$

$$T_{c_i} = T_{data} + ACK_{timeout} + DIFS \tag{4.9}$$

As SIFS, DIFS and ACK frame length are deterministic according to the standard, T_{s_i} and T_{c_i} only depend on T_{data}, which is proportional to the data frame length. In IEEE 802.11, RTS and CTS frames are used to reserve channel access for data transmission when data frame exceed a given threshold. In the RTS/CTS mode, the frame transmission and collision times are

$$T_{s_i} = T_{RTS} + SIFS + T_{CTS} + SIFS + T_{data} + SIFS + T_{ACK} + DIFS, \tag{4.10}$$

$$T_{c_i} = T_{RTS} + CTS_{timeout} + DIFS. \tag{4.11}$$

(4.10) and (4.11) show that the RTS/CTS frames introduce extra overhead for RTS/CTS exchange, i.e., $T_{RTS} + SIFS + T_{CTS} + SIFS$, which lowers the network throughput. On the other hand, T_{c_i} in the RTS/CTS mode is deterministic and independent of the data frame length. In other words, the time a user waits during collisions is reduced when the payload size of the data frame is much larger than that of a RTS frame, which improves the throughput to some extent.

We study a system operating in the stable state, *i.e.,* all incoming data are eventually served within a finite delay. In IEEE 802.11 standard, the retransmission limit is $m = 4$ for long data frames and $m = 7$ for short data frames. The frame drop probability, $p_{drop} = p_i^{m+1}$, is negligible with a small p_i and a large m. Therefore, we assume all frames are eventually received successfully in a stable system. During $1/\mu_i$, on average the remaining users successfully transmit $\frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j$ frames, which contribute to $\frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j T_{s_j}$ time slots. Before the users successfully transmit the frames, the total amount of collision time each user experiences is $\frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j\overline{T_{c_j}} + \overline{T_{c_i}}$, where $\overline{T_{c_i}}$ is the average collision time of a frame transmitted by user i. Denote T_{c_i} the collision

time user i experiences each time a collision occurs; $\overline{T_{c_i}}$ can be derived as

$$\begin{aligned}\overline{T_{c_i}} &= \sum_{k=1}^{m} p_i^k(1-p_i)\cdot kT_{c_i} \\ &= \frac{p_i(1-(m+1)p_i^m+mp_i^{m+1})}{1-p_i}T_{c_i} \\ &\approx \frac{p_i}{1-p_i}T_{c_i}. \end{aligned} \tag{4.12}$$

Given the frame length of user i, T_{s_i} and T_{c_i} can be easily obtained. Since a collision is assumed to occur due to two simultaneous transmissions, the duration of the channel to be busy because of collisions equals half of the total amount of collision time experienced by all users, which is $\frac{1}{2}(\frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j\overline{T_{c_j}}+\overline{T_{c_i}})$. Finally, user i spends $E[W_i]$ in the backoff stage before it successfully transmits the current frame. Therefore, we have

$$\frac{1}{\mu_i} = T_{s_i} + \frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j T_{s_j} + \frac{1}{2}(\frac{1}{\mu_i}\sum_{j=0,j\neq i}^{N-1}\lambda_j\overline{T_{c_j}}+\overline{T_{c_i}}) + E[W_i], \tag{4.13}$$

where $i = 0, 1, ..., N-1$.

Given the arrival rates $\vec{\lambda} = [\lambda_0, \lambda_1, ..., \lambda_{N-1}]$, the equation sets (4.4) and (4.13) can be solved numerically to obtain $\vec{p} = [p_0, p_1, ..., p_{N-1}]$, $\vec{\mu} = [\mu_0, \mu_1, ..., \mu_{N-1}]$, and $\vec{\rho} = [\rho_0, \rho_1, ..., \rho_{N-1}]$.

4.3 Voice Capacity Analysis of an Infrastructure-based WLAN with Unbalanced Traffic

In this section, we study the network performance of an infrastructure-based WLAN, considering the practical issue induced by unbalanced traffic. We first apply the developed analytical framework to study the capacity of a WLAN with non-persistence unbalanced traffic flows. Given the parameters of IEEE 802.11 MAC and voice codecs, we then quantify the voice capacity of an infrastructure-based WLAN. Our voice capacity analysis is compared with that in [58, 59]. In [58], it is assumed there is no collision

during transmissions and all mobile users take advantage of the backoff time of the AP to fulfill their own backoff requirements. The work in [59] takes contention into consideration yet assumes that there are always two and only two active users competing for the wireless channel. The voice capacity obtained in [58, 59] may be over-optimistic due to these simplified assumptions. A loose estimation of voice capacity is harmful for admission control, since, once traffic load exceeds network capacity, the quality of all on-going voice traffic will be jeopardized [60]. We show that our analysis gives a tighter upper bound of voice capacity by considering detail MAC behaviors in an IEEE 802.11 WLAN.

4.3.1 Analytical Model

We consider an infrastructure based WLAN consisting of one AP and $N-1$ users. The AP and correspondent users are connected to the Internet backbone. Multimedia connections are established between WLAN users and correspondent users, through the AP. We assume every user has the same traffic load with an arrival rate of λ_1 frames per slot and the AP carries half of the traffic in the WLAN, $\lambda_0 = (N-1)\lambda_1$ frames per slot. This scenario is valid for an infrastructure WLAN with two-way VoIP connections. Denote the frame service time of the AP and a user as μ_0 and μ_1, respectively. Accordingly, the queue utilization ratios at the AP and a user are

$$\begin{cases} \rho_0 &= \lambda_0/\mu_0 = (N-1)\lambda_1/\mu_0 \\ \rho_1 &= \lambda_1/\mu_1 \end{cases} \tag{4.14}$$

Similarly, we define the conditional collision probability p_i as the probability of a collision seen by a frame being transmitted by the tagged node, which is either the AP ($i = 0$) or any user ($i = 1$). If the AP transmits in a given slot, a collision occurs if at least one of the users also transmit in the same slot. If the tagged node is a user, a collision occurs if either the AP or one of the remaining $N-2$ users transmit in the

same slot. According to (4.4), the conditional collision probabilities p_0 and p_1 are

$$\begin{cases} p_0 &= 1-(1-\rho_1\tau_1)^{N-1} \\ p_1 &= 1-(1-\rho_1\tau_1)^{N-2}(1-\rho_0\tau_0). \end{cases} \tag{4.15}$$

From (4.7), (4.5) and (4.6), the transmission probabilities of the AP and a user at any given slot, τ_0 and τ_1, are functions of p_0 and p_1, respectively,

$$\begin{cases} \tau_0 &= E[A_0]/(E[A_0]+E[W_0]) \\ \tau_1 &= E[A_1]/(E[A_1]+E[W_1]), \end{cases} \tag{4.16}$$

where

$$\begin{cases} E[A_0] &= (1-p_0^{m+1})/(1-p_0) \\ E[A_1] &= (1-p_1^{m+1})/(1-p_1) \\ E[W_0] &= \sum_{k=0}^{m-1} p_0^k(1-p_0)\sum_{j=0}^{k}\frac{CW_j}{2} + p_0^m\sum_{j=0}^{m}\frac{CW_j}{2} \\ E[W_1] &= \sum_{k=0}^{m-1} p_1^k(1-p_1)\sum_{j=0}^{k}\frac{CW_j}{2} + p_1^m\sum_{j=0}^{m}\frac{CW_j}{2}. \end{cases}$$

The average collision time of a frame transmitted by the AP and by a user can be obtained from (4.12):

$$\begin{cases} \overline{T_{c_0}} &= \frac{p_0}{1-p_0}[1-(m+1)p_0^m + mp_0^{m+1}]T_c \approx \frac{p_0}{1-p_0}T_c \\ \overline{T_{c_1}} &= \frac{p_1}{1-p_1}[1-(m+1)p_1^m + mp_1^{m+1}]T_c \approx \frac{p_1}{1-p_1}T_c. \end{cases} \tag{4.17}$$

From the time an AP attempts to transmit a frame till the frame is transmitted successfully, the time interval $1/\mu_0$ consists of four parts: (a) on average the remaining $N-1$ users successfully transmit $(N-1)\lambda_1/\mu_0$ frames, which contribute $(N-1)\lambda_1 T_s/\mu_0$; (b) the AP spends T_s in transmitting the current frame; (c) before the users successfully transmit these frames, the total time that the channel is sensed busy due to failed transmissions is $[(N-1)\lambda_1\overline{T_{c_1}}/(2\mu_0) + \overline{T_{c_0}}/2]$; and (d) $E[W_0]$ is the average backoff time the AP experiences before it successfully transmits the current frame.

Similarly, the time interval $1/\mu_1$ also consists of four parts: (a) the remaining $N-2$ users and the AP contribute $(N-2)\frac{\lambda_1}{\mu_1}T_s$ and $\frac{(N-1)\lambda_1}{\mu_1}T_s$ in successful transmissions, respectively; (b) the tagged user spends T_s in transmitting the current frame; (c) the average collision time is $\frac{1}{2}[(N-2)\frac{\lambda_1}{\mu_1}\overline{T_{c_1}}+\overline{T_{c_1}}+\frac{(N-1)\lambda_1}{\mu_1}\overline{T_{c_0}}]$; and (d) the average backoff time of the tagged user is $E[W_1]$. Therefore, the average service time for the AP and the users are given by

$$\begin{cases} \frac{1}{\mu_0} = \left((N-1)\frac{\lambda_1}{\mu_0}+1\right)T_s + E[W_0] + \frac{1}{2}\left((N-1)\frac{\lambda_1}{\mu_0}\overline{Tc_1}+\overline{Tc_0}\right) \\ \frac{1}{\mu_1} = \left((N-2)\frac{\lambda_1}{\mu_1}+1+\frac{(N-1)\lambda_1}{\mu_1}\right)T_s + E[W_1] + \frac{1}{2}\left(\left((N-2)\frac{\lambda_1}{\mu_1}+1\right)\overline{Tc_1}+\frac{(N-1)\lambda_1}{\mu_1}\overline{Tc_0}\right). \end{cases} \tag{4.18}$$

Equations (4.15), (4.18), along with (4.16)-(4.17), can be solved numerically to obtain p_0, p_1, μ_0, μ_1, ρ_0 and ρ_1.

A user is considered stable only if its traffic intensity or queue utilization ratio $\rho_i < 1$, *i.e.,* the traffic arrival rate is less than the frame service rate. In an infrastructure-based WLAN, the AP is the bottleneck since the traffic to all users has to go through the AP. Therefore, the maximum number of non-persistent connections that can be accommodated in a WLAN can be obtained under the constraint that the AP is stable, *i.e.,* the queue utilization ratio of the AP, $\rho_0 < 1$. The number of active users, i.e., a user that currently has a frame in service, in the WLAN can be obtained as

$$\sum_{i=0}^{N-1}\rho_i = \rho_0 + (N-1)\rho_1. \tag{4.19}$$

4.3.2 Voice Capacity Evaluation

We investigate the maximum number of VoIP connections that can be supported in a single-AP WLAN and compare our analysis with that in [58, 59]. The main parameters of the IEEE 802.11a/b and the upper-layer-header overheads of voice frames are listed in Table 4.1. The highest data rate is 11 Mbps in an IEEE 802.11b WLAN. The values of a slot duration, DIFS, and SIFS are 20 μs, 50 μs and 10 μs, respectively. Each

Table 4.1: Parameters of Voice over 802.11

		802.11b	802.11a
Highest Channel Rate		11Mbps	54Mbps
Slot Time		20 µs	9 µs
SIFS		10 µs	16 µs
DIFS		50 µs	34 µs
CW_{min}		32	16
CW_{max}		1024	1024
Retry Limit		7	7
T_{voice}	PLCP & Preamble	192 µs	24 µs
	MAC Header + FCS	24.7 µs	5 µs
	RTP/UDP/IP Header	29.1 µs	6 µs
	Voice Payload	(payload *8/11) µs	(payload *8/54) µs
T_{ACK}	PLCP & Preamble	192 µs	24 µs
	ACK Frame	10.2 µs	2.1 µs

ACK frame has 14 bytes, and it takes $14 * 8/11 = 10.2$ μs for transmission. Each data frame has a 34 bytes MAC layer overhead and a 40 bytes RTP/UDP/IP-header overhead, which take $34 * 8/11 = 24.7$ μs and $40 * 8/11 = 29.1$ μs to transmit. In addition, it takes 192 μs to transmit the physical layer overheads consisting of 48 μs Physical Layer Convergence Protocol (PLCP) header and 144 μs preamble. In the IEEE 802.11a standard, the maximum data rate is 54 Mbps, approximately five times of that of 802.11b. It takes 2.1 μs, 5 μs, and 6 μs to transmit the ACK frame, MAC layer overhead, and the RTP/UDP/IP headers, respectively. The values of a slot time, DIFS, and SIFS are 9 μs, 34 μs and 16 μs, respectively; and it takes 24 μs to transmit the physical layer overhead, which is eight times smaller than that in 802.11b.

We use Maple 9.5 [61] to calculate the analytical results. Fig. 4.1 shows the conditional collision probabilities of the AP and mobile nodes (MN) with G.711 and G.729 codecs and a 10 ms packetization interval, in an IEEE 802.11b WLAN. The collisions increase with the number of voice connections. Due to the larger payload, the collision probability of G.711 is higher than that of G.729. Since the traffic load of the AP is $N - 1$ times the load of a user, collisions are more likely to occur from the viewpoint of a user than that from the AP.

Real-time applications are very sensitive to delay and jitter. With constant arrival

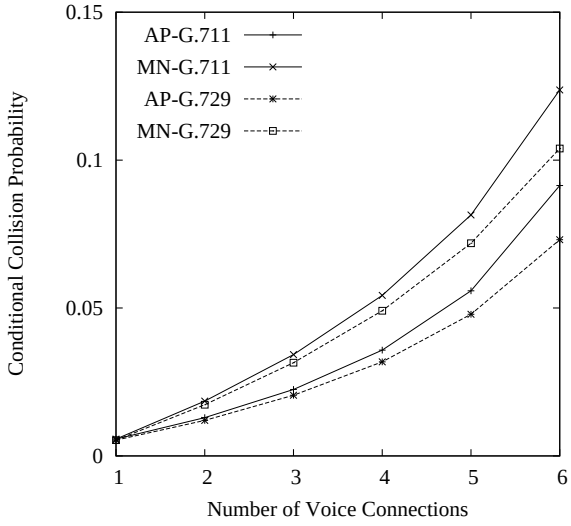


Figure 4.1: Comparison of the Conditional Collision Probabilities of the AP and MNs (802.11b)

rate, delay guarantee of real-time applications is possible only when the traffic arrival rate is less than the service rate ($\rho_i < 1$). A user is considered unstable if its queue utilization ratio $\rho_i \geq 1$. For an unstable user, the queue will be built up, and thus the real-time applications will be damaged because of the ever increasing queuing delay and packet losses due to buffer overflow.

We use G.729 with a 10 ms packetization interval for illustration. Due to the characteristics of the voice traffic, the traffic arrival rate of a user is constant. With the increase of the number of users, the traffic arrival rate of the AP increases linearly while the frame service rate exhibits a non-linear decreasing trend, as shown in Fig. 4.2. Although the frame service rate of a user degrades more rapidly than that of the AP due to the higher collision probability, the AP enters the unstable state before users because of its much higher traffic load. It can be seen from Fig. 4.2, when the seventh G.729 voice connection joins in, the queue of the AP is no longer stable. Therefore, with G.729 and a 10 ms packetization interval, at most six bi-directional VoIP connections can be

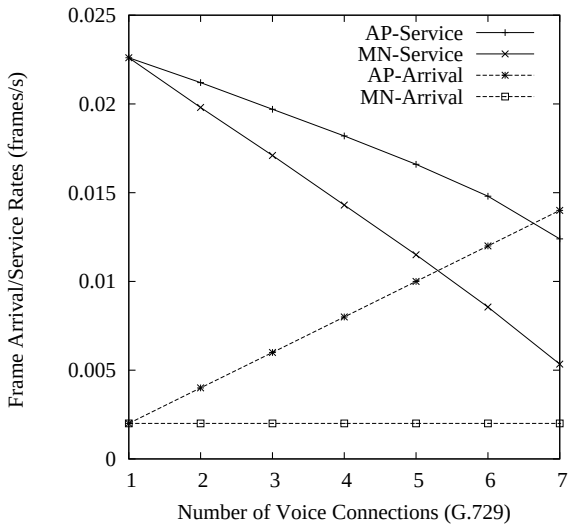


Figure 4.2: Traffic Arrival Rate and Frame Service Rate (802.11b)

supported in an IEEE 802.11b WLAN. One more VoIP connection will jeopardize the performance of all voice connections. Therefore, an accurate upper bound is critical for connection admission control to maintain QoS performance of all VoIP connections.

In Fig. 4.3, it can be seen that the queue utilization ratio of the AP (ρ_0) is always much higher than that of a user (ρ_1) due to the higher traffic load. The maximum number of voice connections with $\rho_0 < 1$ can also be observed. With the G.729 codec, six voice connections with a 10 ms packetization interval, thirteen connections with a 20 ms interval, and nineteen with a 30 ms interval can be supported in an 802.11b WLAN. It is also observed that using G.729 or G.723 makes little difference on the maximum number of voice connections being supported in the WLAN. With G.729 or G.723, up to nineteen simultaneous voice connections with a 30 ms packetization interval can be supported. The payload of G.711 is eight times that of G.729, but only two fewer connections can be accommodated. Compared to the huge overheads specified in the physical and MAC layers, the payload difference between different codecs is relatively small. The maximum number of connections with iLBC is similar to that with

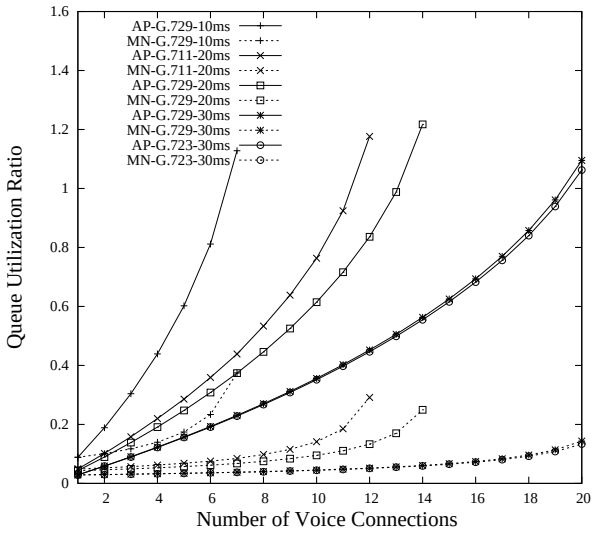


Figure 4.3: Queue Utilization Ratio of the AP and MNs (802.11b)

G.723 and G.729. For VoWLAN, G.729 and iLBC are preferred over G.723 because less compression is required. Another observation is that more VoIP connections can be accommodated when the packetization interval is enlarged. However, larger packetization interval will result in a longer delay. There is a tradeoff between the delay constraint and the voice capacity.

Table 4.2 tabulates the maximum number of VoIP connections for various codecs in an 802.11b WLAN. It shows that only a very limited number of voice connections can be

Table 4.2: The Maximum Number of VoIP Connections (802.11*b*)

Audio (ms)	G.711			G.729			G.723			iLBC
	Proposed Analysis	[59]	[58]	Proposed Analysis	[59]	[58]	Proposed Analysis	[59]	[58]	Proposed Analysis
10	6	6	6	6	7	7				
20	11	12	12	13	14	14				12
30	15	17	18	19	21	22	19	21	22	18
40	19	21	22	25	28	28				
50	22	25	26	31	34	35				
60	25	28	29	37	41	42	37	42	42	

Table 4.3: Comparison of the Maximum Number of VoIP Connections (802.11a)

	G.711		G.729		G.723		iLBC
Audio (ms)	Proposed Analysis	[59]	Proposed Analysis	[59]	Proposed Analysis	[59]	Proposed Analysis
10	25	30	27	32			
20	47	56	53	64			53
30	66	79	79	95	80	96	78
40	82	98	105	126			
50	97	116	130	156			
60	110	131	155	185	158	187	

supported in a WLAN, even with efficient codecs such as G.723 and iLBC. Compared to the results in [58, 59], the obtained analytical upper bounds are much tighter. When the packetization interval is enlarged to accommodate more voice connections, analytical results given in [58, 59] become too optimistic. This is because in [58] it is assumed that any transmitted frame is received successfully without any collision. This assumption may not hold, especially when the number of voice connections is close to capacity. A simple approximation is made in [59] that there are always two active users (one is the AP and the other is a user) in the network and the collision probability keeps as low as 0.03, independent of the number of voice connections. However, in the unsaturated-user scenario, the number of active users is not a constant but increases with the number and the traffic intensity of users in the network. The AP has a frame in service with probability ρ_0 while each user has a frame in service with probability ρ_1. On average, there are $\rho_0 + (N-1)\rho_1$ users that have a frame in service. As shown in Fig. 4.4, the average number of active users in the WLAN varies from 0.02, when there is only one voice connection, to above 3, when the AP is nearly saturated.

The data rate of an 802.11a WLAN is roughly five times of that of an 802.11b WLAN. However, the voice capacity of 802.11a is less than five times of that of 802.11b due to the different parameter values specified in the standard, including the minimum contention window, duration of a slot, DIFS, SIFS, *etc.* For example, a smaller mini-

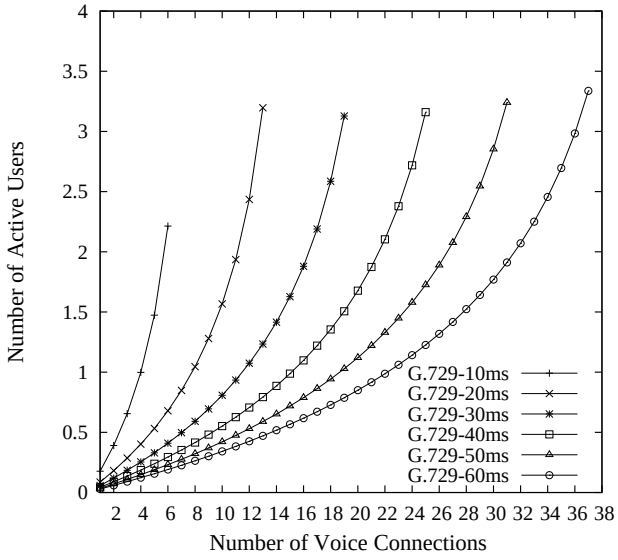


Figure 4.4: Number of Active Users

mum contention window may result in more collisions and a larger SIFS causes longer service time, so both of them may reduce the voice capacity. On the other hand, a smaller physical-layer overhead and shorter slot duration result in higher voice capacity. The effect of different codecs and packetization intervals on the voice capacity of an 802.11a WLAN is given in Table 4.3.

We further validate the analytical results by extensive simulations using the Network Simulator (NS2-2.27) [62]. We use the same parameter values of the IEEE 802.11b as those listed in Table 4.1. The IEEE 802.11 code in NS2 is rigorously checked and some modifications are made according to the standard: the ACK transmission rate is set to 11 Mbps and the preamble transmission rate is kept at 1 Mbps. The network topology is shown in Fig. 3.1 (B). The links connecting the AP and the correspondent users have a data rate of 100 Mbps with a 20 ms propagation delay. The end-to-end packet delay bound is set to 150 ms to maintain good voice quality [35]. Any packets arrive after 150 ms will be discarded from the receiver's playout buffer. In order to show the queue

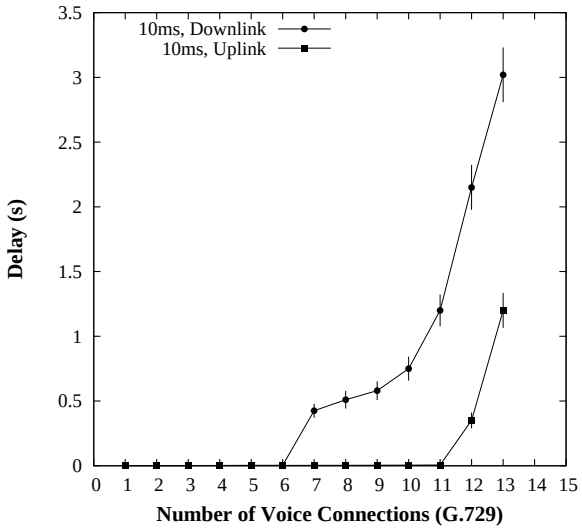


Figure 4.5: Delay Comparison between Uplink and Downlink Voice Flows

accumulating effect in the AP, the buffer size of the AP is set to 300 packets. Initially, a voice connection is established every 10 ms to gradually approach the network capacity, with the starting time randomly chosen over [0, 10] ms. To eliminate the warming-up effects, the simulation data are collected from 10 s to 100 s. We repeat each simulation ten times and calculate the average values.

Fig. 4.5 shows the mean and variation of the delay of G.729 voice flows with a 10 ms packetization interval. The delay of the uplink (from a user to the AP) and downlink (from the AP to a user) voice flows is very low when there are fewer than six connections in the WLAN. When the seventh user joins the system, the delay of the downlink flow increases rapidly while the delay of the uplink is as low as 2 ms. It implies that the AP is saturated when the queue utilization ratio $\rho_0 \geq 1$. Meanwhile, the queue utilization ratio of the users ρ_1 is much less than 1. When more users join, which result in more collisions in the network and decreases the frame service rate, the delay of the uplink flow also increases to more than 300 ms, implying that the users become saturated when there are more than twelve voice connections.

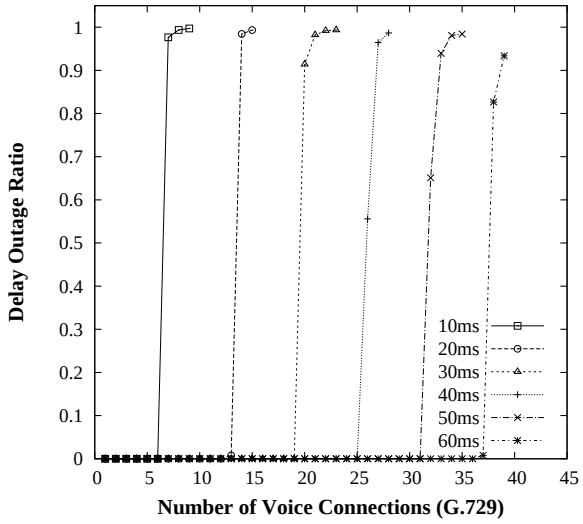


Figure 4.6: Delay Outage Ratio of Voice Traffic

Since the downlink transmissions always suffer longer queuing delays at the AP than the uplink transmissions at the users, we are more interested in the delay of downlink flows due to this bottleneck effect. Fig. 4.6 shows the delay outage ratio (the ratio of the packets with end-to-end delay exceeding 150 ms over the packets being transmitted) of downlink flows, with G.729 and different packetization intervals. Due to the non-bursty characteristics of voice traffic, packet delay is quite low and no packet is discarded from the playout buffer when all users are not saturated. However, the outage ratio of downlink flows becomes significant when the AP is saturated, due to the ever increasing queuing delay at the AP.

In the simulation, the maximum number of voice connections is obtained in the way that one more connection will result in the delay outage ratio larger than 1%. As shown in Fig. 4.7, the simulation results conform with our analysis results quite well, and the obtained upper bounds are more accurate than the result in [58, 59], since we consider the different collision probabilities and queue states of the AP and users. Therefore, our analytical results can be used as a guideline for admission control. Also from the

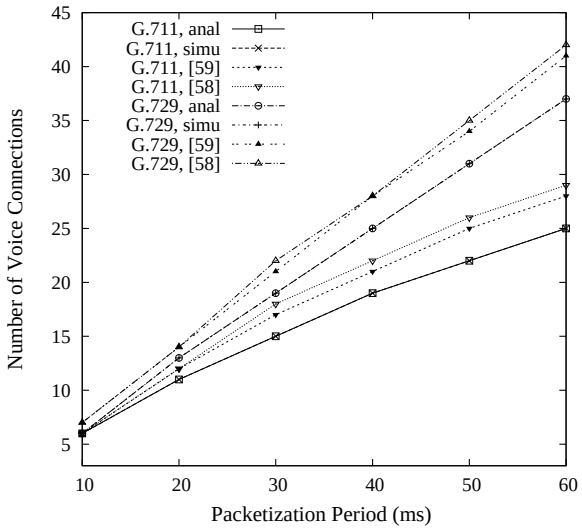
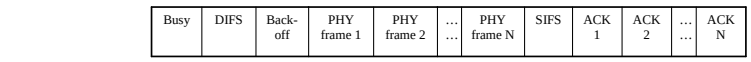


Figure 4.7: Maximum Number of Voice Connections

simulation results, all frames are transmitted within five retransmissions and none is dropped by the MAC due to excessive number of retransmissions, which validates our assumption in Section 4.2.

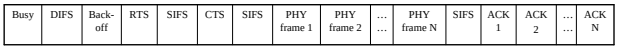
4.4 Analysis of Enhanced MAC Mechanisms

In response to the demand for high throughput WLANs to support multimedia applications such as voice, video conferencing and mobile television, the IEEE 802.11n task group has been established to standardize the next generation WLAN to provide over 100 Mbps throughput at the MAC data service access point (SAP). Different from IEEE 802.11b/a/g, which aims to improve raw data rates with different PHY specifications, IEEE 802.11n aims to achieve higher MAC layer throughput with both PHY and MAC enhancements. In this section, we first present some enhanced MAC mechanisms. We then analytically study the capacity improvements and quantify voice and video capacities with these enhanced MAC mechanisms.

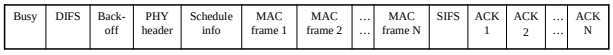


(a) PHY level, basic access, delayed ACK

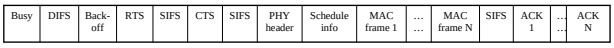
(b) PHY level, basic access, immediate ACK



(c) PHY level, RTS/CTS, delayed ACK



(d) MAC level, basic access, delayed ACK



(e) MAC level, RTS/CTS, delayed ACK

Figure 4.8: Frame Aggregation Mechanisms

4.4.1 MAC Enhancements

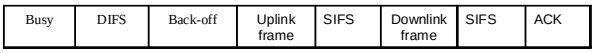
Several new MAC features have been proposed to improve throughput efficiency. One approach is frame aggregation, the idea of which is to aggregate multiple MAC/PHY frames into a single frame (or a train of frames) for transmission [63]. Generally, aggregation mechanisms can be classified into many different aspects: uplink vs. downlink, PHY-level vs. MAC-level, immediate ACK vs. delayed ACK, single-destination vs. multi-destination, etc.

Some frame aggregation mechanisms are illustrated in Fig. 4.8. In Fig. 4.8(a), a train of n PHY frames are sent one by one with no inter-frame space (IFS). These frames can be transmitted to one or multiple destinations, and each destination acknowledges the received frame in the same order after a SIFS. In Fig. 4.8(b), each destination sends an ACK immediately after a SIFS when it successfully receives a frame.

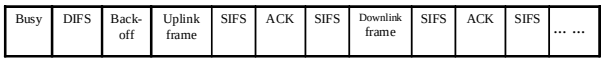
Maximizing throughput may require a large aggregation frame with length longer

than that specified in the current standard (4095 bytes) [63]. On the other hand, it is suggested that the total length of the aggregation frame should be smaller than a threshold since some huge frames may cause unfairness among different users. In addition, long data frames will result in a large collision time and thus reduce the transmission efficiency when collision probability is high. In legacy 802.11, the optional RTS/CTS mode is used to improve the transmission efficiency when the frame size is larger than a threshold ($0 - 2347$ bytes). However, RTS/CTS in legacy 802.11 is employed by a pair of sender and receiver for unicast transmission and is not suitable for the downlink aggregation mechanism which may involve multiple destinations. Therefore, we propose a modified RTS/CTS function that can be used with downlink aggregation to reduce collisions resulting from large data frames, as shown in Fig. 4.8(c). Generally, an RTS frame can be sent in a multicast fashion and all involved destinations need to send back CTS frames if they are available to receive data. In a single-hop WLAN with no hidden terminals, a modified RTS frame with aggregation information, *e.g.*, a list of destinations and transmission sequence, can be sent out and the destination users send back CTS frames in the same sequence of the destination list. Upon receiving the RTS frame, all users check the destination list. Users not on the list set the NAV and will not access the channel during the period indicated by NAV. To further reduce the CTS overhead, another option is that only a couple of destinations, which are chosen randomly by the transmitter, send back CTS frames in a predetermined sequence.

The above three mechanisms are PHY level aggregations. The PHY overhead can be further reduced through MAC level aggregations, which are shown in Fig. 4.8(d) and (e) for basic access mode and RTS/CTS mode, respectively. With these two mechanisms, n MAC frames for different destinations can be aggregated into one PHY frame. After the (shared) PHY preamble and header, destination users receive the scheduling information, based on which they can determine the time to receive the MAC frames if there is any. Using downlink multi-destination aggregation, the AP only



(a) Traditional piggyback



(b) An extension scheme for the downlink transmission

Figure 4.9: Bidirectional Transmission

needs to contend once to transmit an aggregated frame to multiple users, in contrast to multiple contentions and transmissions without frame aggregation.

Another approach to improve throughput efficiency is by allowing data transmission in both directions. That is, a receiver can piggyback aggregation frames to the transmitter without initiating a new transmission, as shown in Fig. 4.9(a). In an infrastructure WLAN, the AP can piggyback a frame to a user in the downlink after the user successfully transmits a frame to the AP in the uplink, and vice versa. Thus, the number of contentions in the WLAN can be significantly reduced and network throughput and capacity will be effectively improved. The bidirectional transmission mechanism is efficient when the traffic flows between the transmitter and receiver are symmetric, but it may not be useful for some other applications, *e.g.*, half-duplex voice services with silence suppression. In other words, bidirectional transmission cannot improve the transmission efficiency if there is no frame in the reverse direction for piggyback. Thus, we extend the bidirectional transmission scheme in a more general scenario to smooth the AP-bottleneck effect. As shown in Fig. 4.9(b), the heavily loaded AP can transmit one frame to any destination users without contentions upon receiving an uplink frame.

4.4.2 Analytical Model

A. Analysis of Frame Aggregation Mechanisms

We consider the same WLAN in Section 4.3.1 that the traffic load of the AP ($i = 0$) is $N-1$ times that of a user ($i = 1, ..., N$). With the legacy DCF MAC, all users have the same priority for channel access, which is unfavorable to the heavily loaded AP, which becomes unstable when $\rho_0 = (N-1)\lambda_1/\mu_0 \geq 1$. By efficient frame aggregations, the frame arrival rate of the AP, λ_0, can be effectively reduced. Without loss of generality, we consider that the AP aggregates $N-1$ frames for multiple users, using the aggregation schemes shown in Fig. 4.8. Therefore, λ_0 is reduced to λ_1 yet the transmission time of an aggregated frame is much larger than that of a single frame. Denote the frame transmission time in the downlink and uplink as T_D and T_U, respectively. The uplink transmission time is

$$T_U = DIFS + P_{OH} + T_{MF} + T_{ACK} + SIFS, \tag{4.20}$$

where P_{OH} is the transmission time of PHY overheads including the preamble and physical layer header, and T_{MF} is the transmission time of the MAC payload. For downlink aggregation schemes (a) and (b), the downlink transmission time T_D are

$$T_D = DIFS + (N-1)(P_{OH} + T_{MF} + T_{ACK}) + SIFS, \tag{4.21}$$

and

$$T_D = DIFS + (N-1)(P_{OH} + T_{MF} + T_{ACK}) + 2(N-1)SIFS, \tag{4.22}$$

respectively. During $1/\mu_0$, the AP spends T_D seconds transmitting an aggregation frame and users transmit $(N-1)\lambda_1/\mu_0$ frames which contribute $(N-1)\lambda_1 T_U/\mu_0$ seconds. Due to the different lengths of downlink and uplink frames, the collision times of downlink frames T_{CD} and uplink frames T_{CU} are different. The average collision time a downlink frame experiences is

$$\overline{T_{CD}} = \frac{p_0}{1-p_0}[1-(m+1)p_0^m + mp_0^{m+1}]T_D \approx \frac{p_0}{1-p_0}T_D. \tag{4.23}$$

An uplink frame may collide with another uplink frame with probability $(N-2)\rho_1/[(N-2)\rho_1+\rho_0]$ and a downlink frame with probability $\rho_0/[(N-2)\rho_1+\rho_0]$. The average collision time an uplink frame experiences is

$$\begin{aligned}\overline{T_{CU}} &= \frac{p_1}{1-p_1}[1-(m+1)p_1^m+mp_1^{m+1}](\frac{(N-2)\rho_1}{(N-2)\rho_1+\rho_0}T_U+\frac{\rho_0}{(N-2)\rho_1+\rho_0}T_D)\\ &\approx \frac{p_1}{1-p_1}(\frac{(N-2)\rho_1}{(N-2)\rho_1+\rho_0}T_U+\frac{\rho_0}{(N-2)\rho_1+\rho_0}T_D). \qquad (4.24)\end{aligned}$$

Thus, the equations of μ_i of the AP ($i=0$) and a user ($i=1$) are given by

$$\frac{1}{\mu_0}=(N-1)\frac{\lambda_1}{\mu_0}T_U+T_D+\frac{1}{2}((N-1)\frac{\lambda_1}{\mu_0}\overline{T_{CU}}+\overline{T_{CD}})+E[W_0], \qquad (4.25)$$

$$\frac{1}{\mu_1}=(\frac{(N-2)\lambda_1}{\mu_1}+1)T_U+\frac{\lambda_1}{\mu_1}T_D+\frac{1}{2}((\frac{(N-2)\lambda_1}{\mu_1}+1)\overline{T_{CU}}+\frac{\lambda_1}{\mu_1}\overline{T_{CD}})+E[W_1]. \qquad (4.26)$$

With CSMA/CA, users have to wait T_C each time a collision occurs. In the basic access mode, a long frame results in a large T_C, which degrades transmission efficiency significantly when the collision probability is high. The RTS/CTS mode is an option used by the legacy MAC to reduce data frame collisions when the frame payload exceeds a threshold. As shown in Fig. 4.8(c), in the RTS/CTS mode,

$$T_D=DIFS+RTS+CTS+(N-1)(P_{OH}+T_{MF}+ACK)+3SIFS. \qquad (4.27)$$

A CTS timeout implies a collision and the transmitter will re-initiate a transmission following the CSMA/CA mechanism. Since RTS and CTS frames are very small, the duration of a collision depends on the uplink data frame. When the frame length does not exceed the RTS threshold, the uplink frames are transmitted in the basic access mode, so the collision time of both uplink and downlink frames is

$$T_{CU}=T_{CD}=T_U; \qquad (4.28)$$

otherwise, the uplink frames are transmitted in the RTS/CTS mode, and

$$T_{CU}=T_{CD}=RTS+CTS_{timeout}+DIFS. \qquad (4.29)$$

The PHY overheads can be further reduced with a MAC-level aggregation scheme. As shown in Fig. 4.8(d),

$$T_D = DIFS + P_{OH} + T_{sch} + (T_{MF} + T_{ACK})(N-1) + SIFS, \tag{4.30}$$

where T_{sch} is the transmission time of the scheduling information that is transmitted before the MAC frames. After synchronization, users can determine when to receive their MAC frames by checking the scheduling information. Hence, strict synchronization is critical for the MAC aggregation scheme. Similar to Fig. 4.8(c), RTS/CTS can also be used in MAC aggregation for transmission efficiency, and the collision time can be reduced to either T_U or $RTS + CTS_{timeout} + DIFS$. Substitute the aggregation transmission time T_D and collision time T_{CU} and T_{CD} into (4.25) and (4.26), we can obtain p_i, τ_i, μ_i, and ρ_i of user i accordingly.

B. Analysis of Bidirectional Transmission Scheme

Assume symmetric traffic are transmitted between the AP and users. After a user successfully transmits a frame to the AP, the AP can piggyback data frames to the transmitter without initiating a new transmission, and vice versa. Therefore, half of the frames are piggybacked without contentions, while all frames have to be transmitted via contention using the legacy DCF MAC. For the downlink piggyback scheme, *i.e.*, the AP that carries half of the frames in the WLAN do not contend with the users. Thus, only the $N-1$ users contend for uplink transmissions. We have

$$p_1 = 1 - (1 - \rho_1\tau_1)^{N-2}, \tag{4.31}$$

$$\frac{1}{\mu_1} = ((N-2)\frac{\lambda_1}{\mu_1} + 1)T_S + \frac{1}{2}((N-2)\frac{\lambda_1}{\mu_1} + 1)\frac{p_1 T_C}{1-p_1} + E[W_1], \tag{4.32}$$

where τ_1 is a function of p_1 derived from (4.7). When the traditional piggyback is used, as shown in Fig. 4.9(a),

$$\begin{aligned} T_S &= DIFS + 2(P_{OH} + T_{MF}) + 2SIFS + T_{ACK} \\ T_C &= DIFS + 2(P_{OH} + T_{MF}) + SIFS. \end{aligned} \tag{4.33}$$

Table 4.4: Parameters of Voice/Video over 802.11(n)

PHY layer data rate	216 Mbps	T_{sch}	300 Bytes
PHY preamble & header (P_{OH})	24 μs	retry limit m	7
DIFS	34 μs	MAC header & FCS	34 Bytes
SIFS	16 μs	IP/UDP/RTP header	40 Bytes
a slot time	9 μs	T_{ACK},CTS	24.5μs
CWmin	16	RTS	24.7μs
CWmax	512		

In the extended scheme, as shown in Fig. 4.9(b),

$$\begin{aligned} T_S &= DIFS + 2(P_{OH} + T_{MF}) + 3SIFS + 2T_{ACK} \\ T_C &= DIFS + P_{OH} + T_{MF} + SIFS. \end{aligned} \tag{4.34}$$

Substitute T_S and T_C in (4.31) and (4.32), p_1 and μ_1 can be obtained.

4.4.3 Voice and Video Capacity Evaluation

To substantiate the analysis of enhanced MAC mechanisms, we calculate the capacity of an IEEE 802.11n WLAN supporting multimedia services, such as low rate voice and broadband video applications. We validate the analysis through extensive simulations with an event-driven simulator written in C language. The system parameters are listed in TABLE 4.4. The network topology is the same as in Section 4.3.2. Similarly, to show the queue accumulating effect in the AP, we set the buffer size of the AP to 300 packets. In the initial stage, a voice or a video connection is established during every codec sample period to gradually approach the network capacity, with the starting time randomly chosen over the sample period. To eliminate the warming-up effects, the simulation data are collected from 10 s to 200 s.

A. Voice Capacity Evaluation

The service rates of the AP and WLAN users carrying G.729 voice connections

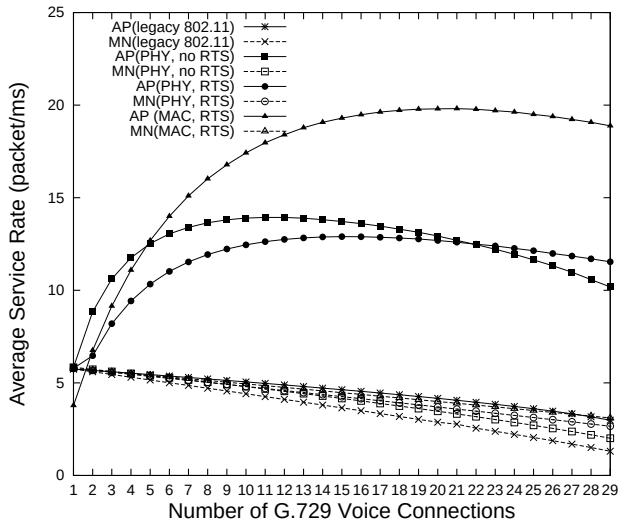


Figure 4.10: Service Rates of Different MAC Aggregation Schemes

with a 10 ms interval under different MAC mechanisms are compared in Fig. 4.10. With the legacy DCF MAC, every MAC frame needs to initiate a transmission and the service rates of the AP and users are quite low. With downlink aggregation, the AP only needs to contend once to transmit a train of frames for a single or multiple destinations. Although the AP requires a long time period to transmit the aggregated frame (which contains $N-1$ separate downlink frames), the average service time for a single downlink frame is significantly improved with the average service rate of $(N-1)\mu_1$ (since a train of $N-1$ frames can be served during the interval $1/\mu_1$). When the number of voice connections is small, it is more efficient to transmit multiple small G.729 packets in the basic access mode to avoid RTS/CTS overheads. When the number of connections increases, RTS/CTS outperforms basic access by reducing the collision time among large data frames. We also observe that MAC layer aggregations achieve much higher service rates of the AP and users than PHY level aggregation by further reducing the PHY overheads, although the scheduling information added may degrade the service rate slightly when the number of connections is quite small.

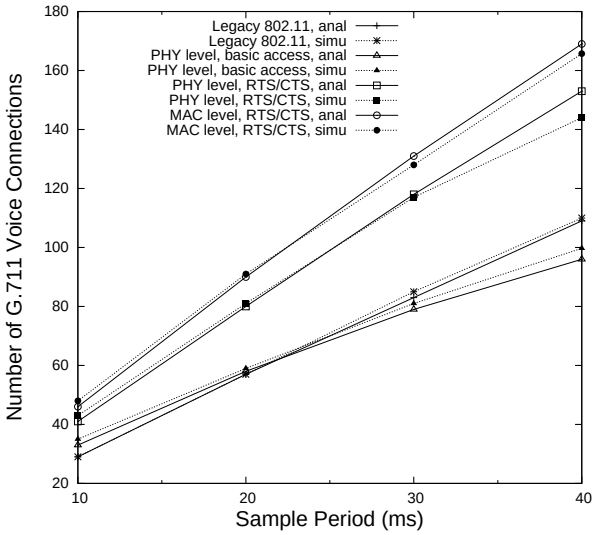


Figure 4.11: Capacity Comparisons of Various Aggregation Schemes

The capacity of various aggregation schemes are compared with that of legacy 802.11 DCF MAC in Fig. 4.11. It can be seen that physical aggregation in the basic access mode may not always improve the voice capacity. With a large sample period, more voice connections can be supported and more collisions may occur. Under high collisions, long frames may degrade the transmission efficiency significantly and thus reduce the voice capacity, even if the arrival rate of the AP is as low as that of an MN. When RTS/CTS is used, voice capacity can be improved by $30 \sim 40\%$ compared to that with the legacy 802.11 DCF. Voice capacity can be further improved by around 10% with MAC-level aggregation which reduces $N - 2$ physical overheads P_{OH} compared with PHY-level aggregation.

Voice capacity can be improved by $35 - 45\%$ with bidirectional transmission, as shown in Fig. 4.12. The analysis is for CBR traffic, in which case the AP always has a frame to transmit in the reverse direction during a sample period. Since voice application is almost half-duplex, the AP may not always be able to piggyback. With the consideration of the packets transmitted during the off period, bidirectional transmis-

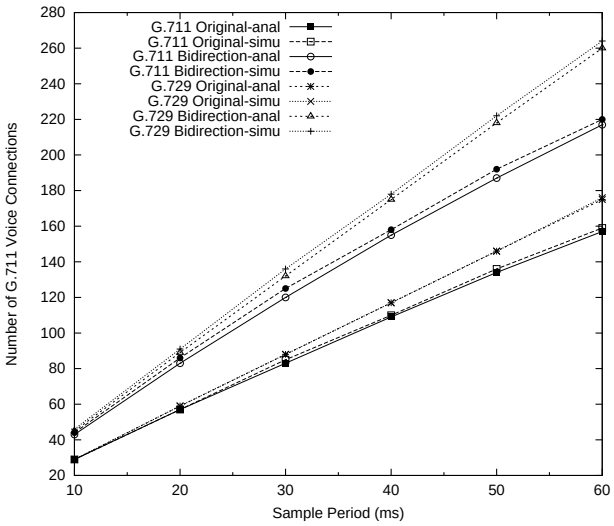


Figure 4.12: Bidirectional Transmission

sion mechanism is still useful for voice applications.

B. Video Capacity Evaluation

We further quantify the video capacity of the legacy DCF MAC, using the main parameters of H.264 video codec tabulated in TABLE 3.2. As shown in Fig. 4.13, the number of supported L1b video flows is quite large due to the low data rate (128 kbps) and encoding frame rate (15 fps). Low frame rate results in a longer frame interval, and thus more multiplexing gain can be achieved. The video capacity is non-decreasing with the increased MTU size. An L1b video flow requires a payload of 1067 bytes and video frames can be encapsulated in one RTP packet for transmission when $MTU \geq$ 1500 bytes, while a video frame may be fragmented and encapsulated over multiple RTP packets for transmission when a smaller MTU is used, *e.g.*, 3 RTP packets with an average payload of 355 bytes are output for transmission when $MTU = 500$ bytes and 2 packets with an average payload of 533 bytes when $MTU = 1000$ bytes. Since every packet needs to contend for transmission in the legacy DCF MAC, the MAC layer traffic arrival rate for $MTU = 1500, 1000, 500$ bytes are $15 \times 9 \times 10^{-6} = 0.135 \times 10^{-3}$,

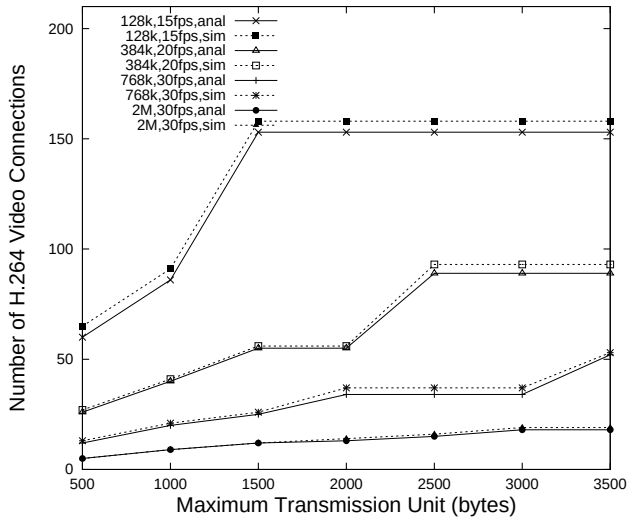


Figure 4.13: Video Capacity of the Legacy 802.11 DCF MAC

$15 \times 2 \times 9 \times 10^{-6} = 0.27 \times 10^{-3}$, and $15 \times 3 \times 9 \times 10^{-6} = 0.405 \times 10^{-3}$ MAC frames per slot, respectively. The number of output RTP packets is non-deceasing with the reducing MTU size. Thus, the MAC layer traffic arrival rate λ_i ($i = 0$ for AP and $i = 1$ for an MN) increases accordingly, which results in a lower capacity based on the AP-bottleneck constraint $(N-1)\lambda_0/\mu_0 < 1$.

Although a large number of low rate video flows can be supported with the enhanced data rate in the next generation WLAN, the video capacity for high rate video flows is still very limited. For L2 video with 2Mbps data rate, 30fps frame rate, and 1500 bytes MTU, a video frame is fragmented into 6 RTP packets and the MAC layer traffic arrival rate is as high as 1.62×10^{-3} MAC frames per slot. The maximum number of L2 video connections that can be supported with the legacy MAC is only 12. When $MTU = 1000$ bytes and 500 bytes, the traffic arrival rate increases to 2.43×10^{-3} and 4.59×10^{-3} MAC frames per slot, and the video capacity decreases to 9 and 5, respectively.

Video capacity under various MAC mechanisms are compared in Fig. 4.14. We

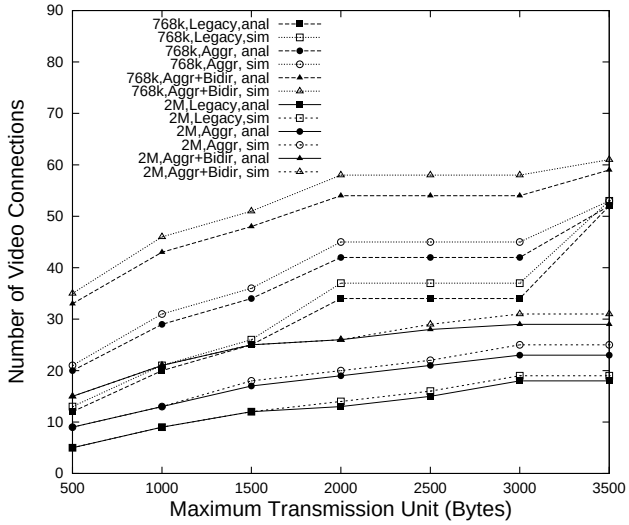


Figure 4.14: Video Capacity Comparison under Various MAC Mechanisms

employ a frame aggregation scheme as shown in Fig. 4.8(b). That is, a train of MAC frames that belong to one single video frame can be aggregated for one transmission. With this aggregation scheme, the MAC arrival rate is only determined by the frame rate, no matter how many fragments (or RTP packets) are output from one single video frame. However, the transmission time increases with the number of output fragments. We take L2 video with 2Mbps data rate and 30fps frame rate for illustration. When $MTU = 1500$ bytes, one video frame is fragmented over 6 RTP packets and the corresponding transmission time is $Ts = DIFS + 6(P_{OH} + T_{MF} + SIFS + T_{ACK} + SIFS) - SIFS$. It is observed that video capacity can be improved by up to 66% for L1.3 video flow (768 kbps) and 80% for L2 video (2 Mbps) with MAC level aggregation scheme in Fig. 4.8(b). More improvements can be achieved by further reducing the PHY, SIFS and ACK overheads using delayed ACK in Fig. 4.8(a), group ACK, or MAC level aggregation scheme in Fig. 4.8(d). When $MTU = 3500$ bytes, no fragmentation is required and thus the legacy MAC and the aggregation scheme achieve the same capacity performance.

Downlink aggregation scheme can effectively smooth the AP bottleneck effect and efficiently improve the voice capacity, as shown in Section 4.4.3 A. However, it may not be appropriate for high rate video connections with much higher payload because an overly long transmission time of the aggregated downlink flow may cause serious starvation of the uplink flows. On the other hand, bidirectional transmissions can be used to eliminate the AP-bottleneck effect and improve the capacity for both video and voice services. Therefore, we combine the bidirectional transmission and aggregation schemes for video services. That is, the heavily loaded AP do not contend with WLAN users, but can transmit an aggregated train of MAC frames belonging to one single video frame upon receiving an aggregated train of uplink frames. As shown in Fig. 4.14, the combined scheme can improve the video capacity of legacy 802.11 MAC by $2-3$ times.

4.5 Distributed MIMO-aware Multi-user MAC

Future WLAN will employ promising MIMO technology in the physical layer. Thus, it is important to study the impacts of MIMO technology on medium access. In this section, we explore the MIMO technology in the MAC design for multi-user downlink transmissions to mitigate the AP-bottleneck in an infrastructure based WLAN. We then extend the developed framework to study the performance of the proposed multi-user MIMO MAC.

4.5.1 Multi-user MIMO Beamforming

MIMO communication techniques have shown great potential to provide enhanced link capacity combined with increased diversity and interference suppression. Single-user MIMO techniques were first studied in the literature. Recently, multi-user MIMO has been an important research area as it is indicated multi-user MIMO can also achieve similar capacity scaling as single user MIMO system. An important practical issue

is that multiple antenna elements are hardly deployed over customer's mobile devices due to technique limitations as well as cost consideration. In a realistic WLAN, it is more likely that only AP be equipped with multiple antennas. Thus motivated, we study how to explore a multi-user MIMO system to mitigate the AP-bottleneck effect in traditional infrastructure based WLAN, where the AP with multiple antennas communicates with multiple users equipped with a single antenna simultaneously in the downlink.

We consider the same single-hop WLAN scenario as in Section 4.3. The AP employs M antennas and user i is equipped with M_i antennas, where $1 \leq M_i \leq M$. By using MIMO beamforming technology to effectively suppress co-channel interference (CCI) at the end users, the AP is able to communicate with K ($\leq M$) users in the same frequency channel simultaneously. Here, we propose a distributed MU MAC protocol based on a leakage-based precoding scheme for multi-user beamforming. A functional block diagram of the multi-user beamforming is shown in Fig. 4.15 [64, 65]. The data intended for user i, s_i, is multiplied by a beamforming vector $\mathbf{W}_i$ prior to transmission, and the overall data, $\sum_{k=1}^{K} \mathbf{W}_k s_k$, is broadcast over the wireless channel. At the receiver's end, the received data vector of size $M_i \times 1$ of user i is given by

$$\mathbf{y}_i = \mathbf{H}_i \sum_{k=1}^{K} \mathbf{W}_k s_k + \mathbf{V}_i = \mathbf{H}_i \mathbf{W}_i s_i + \mathbf{H}_i \sum_{k=1, k\neq i}^{K} \mathbf{W}_k s_k + \mathbf{V}_i \tag{4.35}$$

where $\mathbf{V}_i$ is an additive Gaussian noise vector and each element in $\mathbf{V}_i$ has zero mean and variance N_0, $\mathbf{H}_i$ is the MIMO channel matrix denoted by $\mathbf{H}_i = [h_i^{p,q}]_{M_i \times M}$ with (p,q)th entry $h_i^{p,q}$ representing the channel gain from the q-th antenna at the AP to the p-th antenna at user i. The channel gain $h_i^{p,q}$ is assumed to be independent complex Gaussian with zero mean and unit variance, and time-invariant over each transmission frame time. Based on the concept of power leakage introduced in [57], which is the total interference power leaked from user i to all other co-channel users, the signal-to-

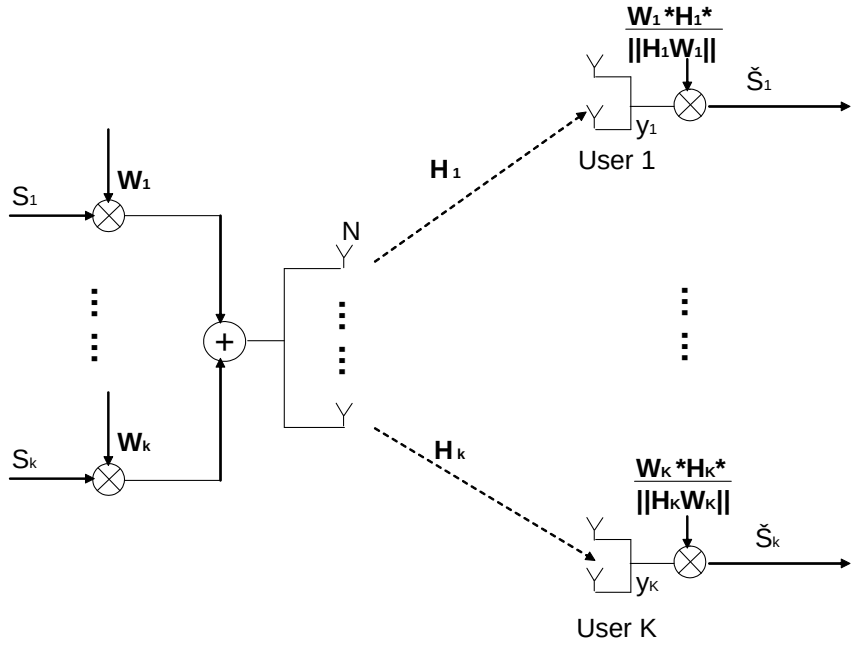


Figure 4.15: Block Diagram of Multi-user Beamforming System

leakage-noise ratio (SLNR) of user i is defined as

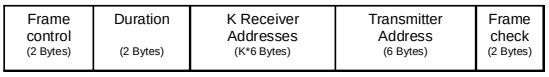
$$SLNR_i = \frac{||\mathbf{H}_i\mathbf{W}_i||^2}{M_iN_0 + \sum_{k=1,k\neq i}^{K}||\mathbf{H}_k\mathbf{W}_i||^2}. \tag{4.36}$$

The beamforming vectors, $\mathbf{W}_i$, $i = \{1, ..., K\}$, is determined such that (4.36) is maximized over $\mathbf{W}_i$ and subject to $\sum \mathbf{W}_i = P_T$, where P_T is the total transmission power. A closed form solution of the optimization problem is given by [64]

$$\mathbf{W}_i \propto \text{max. eigenvector}\left((M_iN_0\mathbf{I} + \tilde{\mathbf{H}}_i^*\tilde{\mathbf{H}}_i)^{-1}\right). \tag{4.37}$$

4.5.2 Multi-user MIMO MAC

Interoperation with the widely deployed legacy WLANs is critical for MAC protocol design in the next generation WLANs with MIMO capability. Therefore, the proposed



(a) Multi-user RTS (MU-RTS) frame

(b) Multi-user CTS (MU-CTS) frame

Figure 4.16: Control Frame Formats

MU MAC protocol integrates MIMO technology into IEEE 802.11 MAC without changing the asynchronous, distributed medium access mechanism of the legacy MAC.

A. Extended Control Frames

Prior to data transmission, users exchange RTS/CTS control frames, to reserve the channel for the subsequent data frames, share with each other the hardware capabilities, and estimate the associated MIMO channel matrix between the transmitters' and the receivers' antennas if possible. The control frames are generally transmitted at the basic rate, which is much lower than the data rate, to assure that all users in the WLAN are capable of decoding them. The formats of the extended control frames are shown in Fig. 4.16. The multi-user RTS (MU-RTS) frame is extended from the original RTS frame structure defined in the standard. To support multi-user transmissions, the extended MU-RTS frame includes multiple receiver address (RA) fields. The receiver estimates the MIMO channel matrix upon receiving the MU-RTS frame and attaches the measured channel information in the extended multi-user CTS (MU-CTS) frame.

B. Uplink Transmissions

Because of the difficulties for effectively managing the multi-user interference and acquiring synchronization among multiple uplink transmissions at the AP receiver, in the

proposed MU-MAC, all the users contend for channel access with IEEE 802.11 DCF. That is, only one user accesses the channel at a time in the uplink. In a MIMO/SIMO system, users still can exploit multi-antenna degree of freedom to improve the transmission rate and reliability. When a spatial diversity scheme is applied, the transmission rate of user i is given by [66]

$$R_{i,1}^{U} = \eta W \log_2(1+\gamma_i) \tag{4.38}$$

where η is a system coefficient related to the efficiency of the transceiver design, W is the signal bandwidth, $\gamma_i = \max\{\sum_{j=1}^{M}\gamma_{j,1}, \sum_{j=1}^{M}\gamma_{j,2}, ..., \sum_{j=1}^{M}\gamma_{j,M_i}\}$, and $\gamma_{j,k} = |h_i^{j,k}|^2 P_T/N_0$. When a spatial multiplexing scheme is used, the transmission rate of user i is given by [67]

$$R_{i,2}^{U} = \max \sum_{j=1}^{\min(M,M_i)} \eta W \log_2(1+\frac{P_j^* \lambda_j^2}{N_0}) \tag{4.39}$$

where λ_j is the singular value of $\mathbf{H}_i$ of user i, $P_j^* = \max\{\mu - \frac{N_0}{\lambda_j^2}, 0\}$ is the waterfilling power allocation, and μ is the Lagrange multiplier selected under the constraints of $\sum_j P_j^* = P_T$.

C. Multi-user Downlink Transmissions

In the downlink transmission, we exploit the multi-user degree of freedom for the AP to simultaneously serve multiple users. The multi-user downlink transmission diagram is shown in Fig. 4.17 and the operational procedure is as follows.

Step 1. The AP sends out an MU-RTS frame to a group of users (i.e., K users) at the basic rate.

Step 2. User j on the RA list measures the MIMO channel matrix $\mathbf{H}_j$ upon receiving the MU-RTS; and responds with an MU-CTS containing $\mathbf{H}_j$ sequentially in the same order as that in the MU-RTS if it satisfies the multi-user transmission condition. A SIFS is applied between any two consecutive MU-CTS frames. That is,

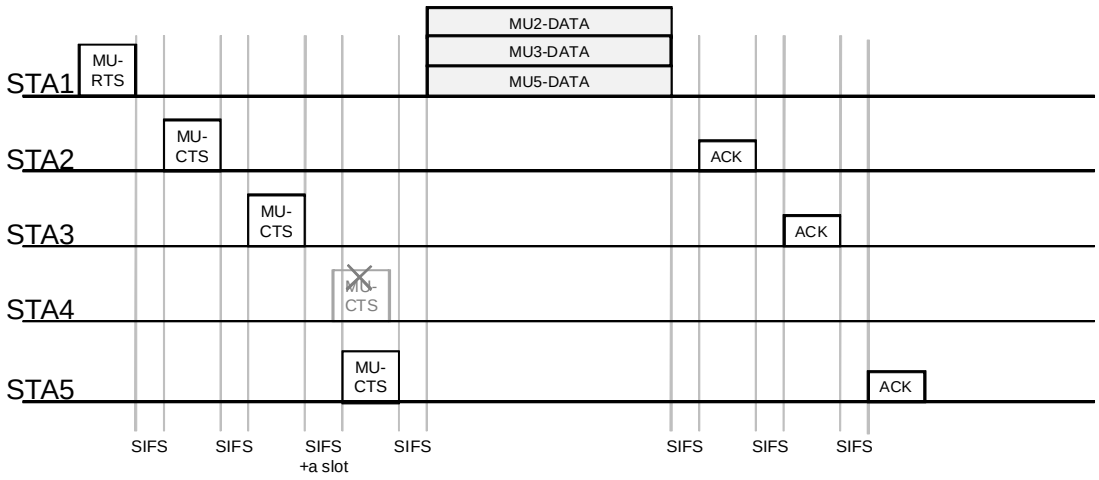


Figure 4.17: Multi-user Transmissions

based on $\mathbf{H}_j$ and the overheard $\mathbf{H}_i$ from the previous users, user j can decide if it can be admitted or not by measuring the SINR values of all admitted users plus itself,

$$SINR_i = \frac{||\mathbf{H}_i\mathbf{W}_i||^2}{M_iN_0 + \sum_{k=1,k\neq i}^{j} ||\mathbf{H}_i\mathbf{W}_k||^2} \geq \beta, i = 1, ..., j \tag{4.40}$$

where β is a threshold to ensure successful decoding. If the SINR of any user is below β, user j will not respond with an MU-CTS frame.The next user calculates the SINR at the same time as user j and sends an MU-CTS frame after $SIFS + a\ slot\ time$ if user j does not transmit and it satisfies the multi-user transmission condition.

Step 3. Based on the $\mathbf{H}_i$ in MU-CTS frames, the AP can decide the beamforming weights $\mathbf{W}_i$ for concurrent downlink transmissions, according to (4.37). A transmission opportunity (TXOP) and aggregation schemes can also be applied for downlink transmissions. The users not on the RA list set the network allocation vector (NAV) and defer their transmissions.

Step 4. After the longest downlink transmission completes and the channel is sensed idle for $SIFS$, users send back a delayed group ACK frame sequentially in the same order of MU-CTS frame transmissions.

Step 5. It is also possible that the AP may not be able to predict the NAV duration accurately due to the complicated scenario of multi-user transmissions, and the AP needs to set a long NAV to protect all ongoing transmissions. In this case, the AP transmits a CF-End frame after all downlink transmissions finish, explicitly indicating the completion of its TXOP.

D. Multi-user Selection

Generally, when a network operates in the stable state, all traffic can be served during a certain time interval. However, when a network is heavily loaded with bursty traffic, traffic from some users may suffer from overly long delays. In this case, it is critical for the MU-MAC to determine a proper set of users for multiple transmissions. If the channel conditions of all users are available, the AP can use exhaustive search or greedy user selection algorithms for downlink multi-user scheduling to enhance the downlink throughput. However, to accurately obtain timely channel state information (CSI) at the transmitter is very challenging and costly, especially in a distributed network where users randomly access the wireless channel. In the proposed MU-MAC protocol, the CSI is measured by the AP and other users through the RTS/CTS control message exchanges, as shown in Fig. 4.17. In other words, the AP can not determine the multi-user sets according to $\mathbf{H}_i$ before it successfully accesses the channel and exchanges control messages with other users. On the other hand, with the enhanced performance in terms of throughput, transmission reliability, and communication coverage, broadband multimedia services can be supported in WLANs. To provide QoS for multimedia services, we implement a simple priority queue (PQ) based scheduling scheme for multi-user downlink transmissions.

Data traffic can be generally classified into realtime (RT) and non-realtime (NRT) categories. Denote the waiting time of the head-of-the-line frame in the queue and the number of frames to user i at time t as $\omega_i(t)$ and $q_i(t)$, respectively. The AP first

checks the queue of the RT traffic, Q_{RT}. If at least K users have traffic in Q_{RT}, the AP selects the set of K users with the largest $\sum_i \omega_i(t)$. Otherwise, the AP needs to select the remaining users with traffic in the NRT queue, Q_{NRT}. To further improve the efficiency of multiple downlink transmissions, the users with similar transmission duration in Q_{NRT} will be selected. Notice that a strict priority scheduling policy is implemented to provide QoS guarantee for delay-sensitive RT traffic, i.e., the AP will choose users from the NRT queue only when the RT queue is empty. For scheduling relatively delay-tolerable NRT traffic, the AP selects the set of users with the largest queue lengths, i.e., $\sum_i q_i(t)$, to guarantee the system stability and improve the downlink NRT throughput.

4.5.3 Performance Analysis

We extend the analytical framework to study the performance of the proposed MU-MAC protocol supporting non-persistent traffic flows. During the service time interval $1/\mu_0$, AP can successfully access the channel and serve multiple users in the downlink simultaneously. The number of users in the multi-user downlink transmission is determined by the instantaneous channel conditions and the threshold β for successful decoding, as in (4.40). Denote $\bar{K}$ as the average number of users in the downlink transmissions, the average downlink service rate of the AP is $\bar{K}/(1/\mu_0) = \bar{K}\mu_0$. Thus, we have

$$\rho_i = \begin{cases} \lambda_0/\bar{K}\mu_0, & i = 0 \quad \text{(AP)} \\ \lambda_i/\mu_i, & 1 \le i \le N \quad \text{(users)}. \end{cases} \tag{4.41}$$

We then extend (4.13) to obtain the average service rate of the AP (μ_0) and any user (μ_i) as

$$\frac{1}{\mu_0} = T_{s_0} + \frac{1}{\mu_0} \sum_{j=1, j\neq i}^{N} \lambda_j T_{s_j} + \frac{1}{2}\left(\frac{1}{\mu_0} \sum_{j=1, j\neq i}^{N} \lambda_j \overline{T_{c_j}} + \overline{T_{c_0}}\right) + E[W_0] \tag{4.42}$$

and

$$\frac{1}{\mu_i} = T_{s_i} + \frac{1}{\mu_i}(\sum_{j=1,j\neq i}^{N} \lambda_j T_{s_j} + \frac{\lambda_0}{\bar{K}} T_{s_0}) + \frac{1}{2}(\frac{1}{\mu_i}\sum_{j=1,j\neq i}^{N} \lambda_j \overline{T_{c_j}} + \frac{\lambda_0}{\mu_i \bar{K}}\overline{T_{c_0}} + \overline{T_{c_i}}) + E[W_i] \quad (4.43)$$

Since we use RTS/CTS frames to measure channel conditions and exploit MIMO capabilities, we have

$$\begin{aligned} T_{s_0} &= DIFS + T_{RTS} + SIFS + KT_{CTS} + KSIFS + T_{Data_0} + KSIFS + KT_{ACK}, \\ T_{s_i} &= DIFS + T_{RTS} + SIFS + T_{CTS} + SIFS + T_{Data_i} + SIFS + T_{ACK}, \end{aligned} \quad (4.44)$$

and

$$T_c = DIFS + T_{RTS} + CTS_{timeout}. \quad (4.45)$$

The data transmission time is given by $T_{Data_i} = P_{OH} + L_i/R_i$, where P_{OH} is the transmission time of the physical overheads, i.e., physical header and preamble, L_i is the frame payload length, and R_i is the transmission rate of user i. The uplink transmission rate, R_i^U, is determined by the MIMO spatial diversity/multiplexing scheme, as derived in (4.38) and (4.39). For multi-user downlink transmissions, we approximately use the average transmission rate of the AP to user i, which is given by

$$R_i^D = E[\eta W \log_2(1 + SINR_i)] \quad (4.46)$$

where $SINR_i$ is given in (4.40). Notice that, during the service time interval $1/\mu_i$, on average the successful transmissions of the AP equals λ_0/μ_i under the stable state [6], but contributes only $\frac{\lambda_0}{\bar{K}\mu_i} T_{s_0}$ in (4.43) since the AP serves $\bar{K}$ users in the downlink simultaneously. In a stable network, all arrival traffic will eventually be served, thus the average network throughput under the stable state is

$$E[S] = \lambda_0 L_0 + \sum_{i=1}^{N} \lambda_i L_i \quad (4.47)$$

where N is the maximum number of the user supported in a WLAN, which is bounded by the stable state condition. That is, when one more user joins the network, the system will become unstable for either $\lambda_0/\bar{K}\mu_0 \geq 1$ or $\lambda_i/\mu_i \geq 1$ ($1 \leq i \leq N$).

Table 4.5: PHY Modes in IEEE 802.11a

Mode	Modulation	Code Rate	Data Rate	Bytes/Symbol
1	BPSK	1/2	6 Mbps	3
2	BPSK	3/4	9 Mbps	4.5
3	QPSK	1/2	12 Mbps	6
4	QPSK	3/4	18 Mbps	9
5	16-QAM	1/2	24 Mbps	12
6	16-QAM	1/2	36 Mbps	18
7	64-QAM	2/3	48 Mbps	24
8	64-QAM	3/4	54 Mbps	27

4.5.4 Numerical Results

We use Matlab 7.0 and an event-driven simulator written in C to obtain the simulation results. The physical modes is based on IEEE 802.11a, as listed in Table 4.5. Hard-decision Veterbi decoding is used at the receiver. We use the same physical and MAC parameters as listed in Table 4.4. Users have the same CBR traffic with the frame payload $L_i = 1024$ bytes and average arrival rate λ_i, and the AP carries half of the traffic in the network, $\lambda_0 = N\lambda_i$. The AP employs $M = 6$ antenna elements while each user only has one antenna. The signal bandwidth is 20MHz. We first evaluate the average bit error rate (BER) performance of all physical modes with the leakage-based precoding technique. Then we obtain the average number of users that can be served simultaneously in the downlink with satisfactory BER performance, e.g., BER is not larger than 10^{-4}, and investigate the performance of the proposed MAC with multi-user beamforming in a stable network.

Fig. 4.18 shows the relationship between the number of users that are simultaneously served by the AP in the downlink and the achieved BER of each user. The normalized SNR is defined as the ratio of the total transmission power over the Gaussian noise,

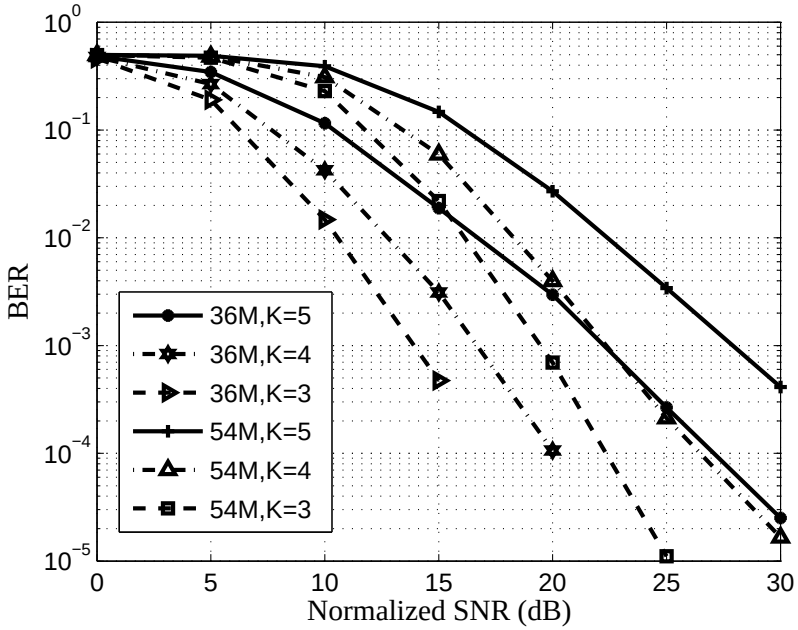


Figure 4.18: BER vs SNR

P_T/N_0. The BER performance increases with the SNR for a given number of users, but degrades when more users are supported for a given PHY mode. This is because the co-channel interference among multiple users are not fully suppressed with the leakage-based precoding scheme[1]. For mode 8 and $SNR = 30$ dB, the BER is on the order of 10^{-5} when 4 users are served simultaneously, and the BER degrades to be on the order of 10^{-4} for 5 users due to higher co-channel interference. It is also observed that 5 users can be supported with BER less than 10^{-4} when a lower mode, mode 6, is considered. The average number of users that can be supported simultaneously with the downlink beamforming is given in Table 4.6.

We compare the traffic intensities of the AP and mobile nodes (MN) in Fig. 4.19. The traffic arrival rate of a single user is 50 packets per second. For $SNR = 30$ dB, the

[1]To obtain the optimal W_i to maximize the received SINR at each user is very challenging, and thus the leakage-based precoding uses the alternative SLNR criterion instead [65].

Table 4.6: Maximum Number of Users and Average Data Rate

	5dB	10dB	15dB	20dB	25dB	30dB
6M	3	4	4	5	5	5
9M	2	3	4	5	5	5
12M	2	3	4	5	5	5
18M	-	2	3	4	5	5
24M	-	2	3	4	5	5
36M	-	-	2	3	4	5
48M	-	-	-	2	4	4
54M	-	-	-	2	4	4

AP can serve up to 4 users simultaneously, each of which has a data rate of 54 Mbps. The uplink transmission rate under $SNR = 30$ dB can also reach 54 Mbps with either diversity (4.38) or multiplexing (4.39) scheme. Since the AP with a higher traffic load has the same priority for channel access as other mobile users, the traffic intensity of the AP is generally larger than that of a mobile user. Thus, the maximum number of users, N, that can be supported in a stable WLAN is obtained such that the traffic intensity of the AP is less than 1, $\rho_0 < 1$. It is shown that for SU MAC, $N = 21$ and the maximum allowed traffic intensity of the AP is 6 times that of a user, $\rho_0/\rho_1 = 6$. For MU MAC supporting $K = 2$ users in the downlink transmissions, $M = 26$ and $\rho_0/\rho_1 = 4.5$. When $K = 4$, M increases to 29 and ρ_0/ρ_1 further decreases to 3.9. By employing multi-user beamforming at the AP, the proposed MU MAC protocol can significantly improve the network performance by minimizing the AP-bottleneck effect in legacy WLANs.

We further investigate the MU MAC performance under various traffic loads in Figs. 4.20 and 4.21. The maximum number of users supported in a stable WLAN decreases when the traffic arrival rate of each user increases. For random channel access,

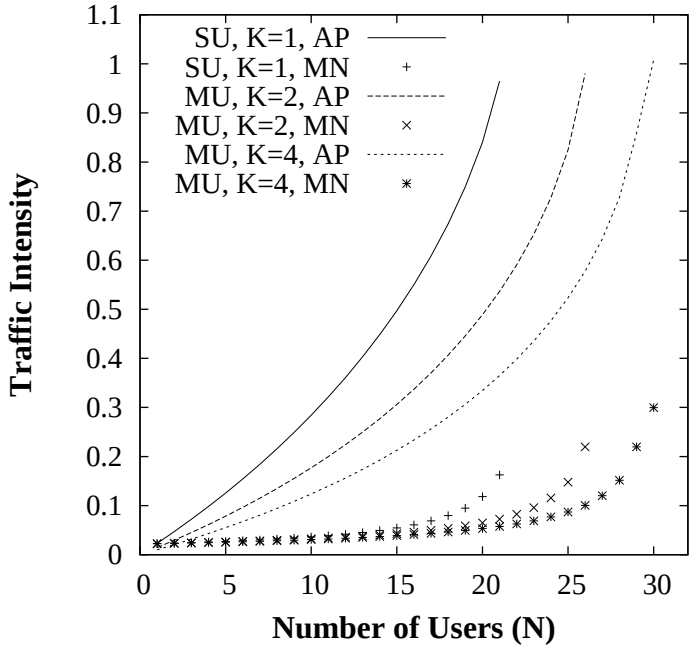


Figure 4.19: Traffic Intensity Comparison

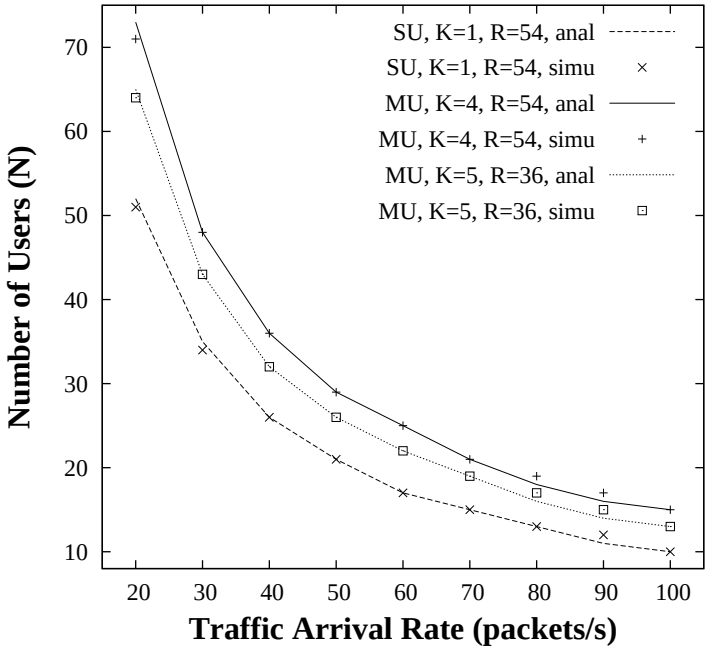


Figure 4.20: Maximum Number of Users under Various Traffic Arrivals

all users with a non-empty queue contend for channel access. A higher traffic load results in more collisions in the network, which can significantly degrade the network performance. It is shown in Fig. 4.20 that M decreases when the traffic arrival rate of a user increases. M decreases from 52 to 10 when the traffic arrival rate of a user increases from 20 to 100 packets per second, in the SU MAC supporting 54 Mbps for a single user transmission. According to Table 4.6, for $SNR = 30$ dB, the AP can support up to $K = 4$ users with a rate of 54 Mbps and up to $K = 5$ users with a rate of 36 Mbps each. For MU MAC, M decreases from 73 to 15 for $K = 4$, and 65 to 13 for $K = 5$. A larger K may not necessarily improve the network performance due to the larger MAC layer overheads, including a larger MU-RTS frame, more MU-CTS frames and the involved IFSs, and the degraded data rate resulting from the co-channel interference. The maximum network throughput under the stable state is shown in Fig. 4.21. When the traffic load of each user increases, the number of users, N, that can be supported in the network decreases. Hence, the maximum network throughput under the stable state does not change much with the traffic loads. By employing multi-user beamforming at the AP, the proposed MU MAC can outperform SU MAC by 40 percent. The proposed MU MAC with $K = 4$ achieves the highest network throughput, as compared with SU MAC and MU MAC with $K = 5$.

4.6 Summary

In this chapter, we analytically study the performance of IEEE 802.11 WLAN in support of non-persistent multimedia traffic flows. The main accomplishments of this chapter are summarized as follows:

- We have developed a generic analytical framework to study the performance of an IEEE 802.11 WLAN, considering the practical issues induced by asymmetric non-persistent multimedia traffic flows.

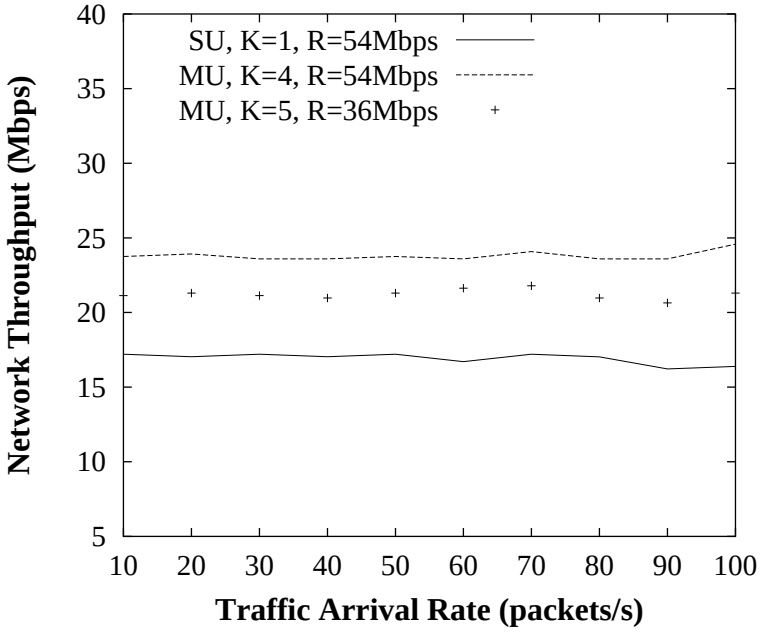


Figure 4.21: Network Throughput

- We have applied the analytical framework to investigate the voice capacity of an infrastructure based WLAN. The analysis reveals how the protocol overheads from different layers, the selection of codecs, and the heterogeneous load distribution in the network, affect the capacity of a WLAN.

- We have further analytically studied some enhanced MAC mechanisms, i.e., frame aggregation and bidirectional transmission, which have been proposed for next generation high throughput WLANs. With various MAC mechanisms, voice and video capacities have been quantified and compared as well.

- We have explored multi-user MIMO technique in MAC design for future WLANs with MIMO capability and proposed a distributed multi-user MIMO MAC. The proposed multi-user MAC significantly outperforms the legacy single-user MAC by effectively mitigating the AP-bottleneck effect.

Chapter 5

MAC Protocol in UWB/mmWave Wireless Personal Area Networks

In this chapter, we first analytically study the performance of existing IEEE 802.15.3 MAC [68], and discuss the potential MAC enhancements in UWB/mmWave WPANs, considering the salient features of UWB and mmWave communication technologies. Then, We propose exclusive regions (ER) based MAC layer resource allocation to explore the spatial multiplexing gain of UWB/mmWave WPANs by allowing appropriate concurrent transmissions [10]. We further theoretically derive the spatial multiplexing capacity and obtain the best protocol parameters, e.g., the ER size, in order to maximize the network capacity [12].

5.1 Related Work

5.1.1 Analysis of IEEE 802.15.3 MAC

IEEE 802.15.3 is a hybrid MAC designed for high rate WPANs. A simulation study of IEEE 802.15.3 MAC is presented in [69] to investigate the performance of real-time and best-effort traffic with various superframe lengths and different ACK policies.

In [70], the performance of intra-piconet communications is enhanced by taking advantage of the multi-rate support in the physical layer. In [71], the relationship between the duration of contention period and the number of requesting users is studied. The analysis in [71] is based on the result from [72] which studies the capacity of the p-persistent CSMA and may not be suitable for the traditional CSMA/CA adopted in IEEE 802.15.3. In [71, 72], the channel is observed at the end of each successful transmission, and the time interval between two successful transmissions is considered regenerative. However, due to the very small size of the minimum contention window and maximum retry limit specified in IEEE 802.15.3, which are 7 and 3, respectively, the frames are not always successfully transmitted and may be dropped after 3 retransmissions. The frame drop probability increases with the number of requesting users and may not be negligible. Thus, the general assumption of no frame drop in IEEE 802.11 and the previous model in [71] does not hold in IEEE 802.15.3. Moreover, the regenerative property of each successful transmission holds when all users continuously have packets for transmission during the contention period, but it does not hold in the case that each user only transmits at most one request per superframe, which is a special case in non-saturation scenarios. None of the existing models can be directly used for investigating this special case. To the best of our knowledge, existing works on IEEE 802.15.3 either study the throughput and delay performance in the CAP and CTAP separately or focus on scheduling in the CTAP. There is little analytical study on the overall performance of a hybrid MAC with respect to the tunable parameters [68], including the initial contention window, the retry limit, the length of contention period CAP and contention free period CTAP, *etc.* As discussed in Section 2.1.3, a hybrid MAC usually adopts a request-grant mechanism to provide a certain level of QoS for realtime applications. That is, users send requests in the CAP and transmit data in the CTAP if their requests are successfully granted. Given a superframe, more users can successfully send requests in a longer CAP, but fewer slots can be allocated for

data transmissions in the CTAP, and vice versa. Therefore, appropriate lengths for contention and contention free periods are critical to IEEE 802.15.3 MAC.

5.1.2 UWB/mmWave MAC

Besides the MAC protocols specified in the current high rate WPAN standards, i.e., IEEE 802.15.3 and WiMedia MAC, there has been significant research work on UWB MAC and related MAC layer resource allocation issues. MAC layer generally uses temporally exclusive mechanisms in the time, frequency, or space domains to eliminate or reduce interference from simultaneous transmissions. Technically, interference and multiple access can be effectively managed through power control, rate control, or mutual exclusion. Recognizing the UWB characteristics, e.g., low power emission and interference levels, MAC protocols based on the concept of maximum sustainable interference (MSI) have been studied in the literature. In [73], a proactive and adaptive MAC is presented for joint power and transmission rates assignment. However, it is shown in [74] that complex power control scheme can only provide marginal throughput gain in UWB systems due to the stringent transmission power limit. In [75], a MSI-based MAC layer resource allocation scheme is proposed to resolve the near-far problem and alleviate the negative effect of long acquisition time of UWB transmissions. However, frequency control message exchanges are required, which introduces heavy overheads and degrades the network performance. In [76], measurement based interference management in wireless mesh networks is studied. A rate control based interference mitigation scheme is proposed in [77] to mitigate the impacts of interfering pulses to the receiver in a pulsed time-hopping (TH) UWB system. If pulses from a strong interferer are larger than the erasure threshold, the scheme will replace them by erasures (i.e., skipping in the decoding process). Simulations in a symmetric topology show that a source can always send and continuously adapts its rate without mutual exclusion when the physical layer interference mitigation scheme is properly applied.

Another general approach is to use an exclusive mechanism in the space domain. In [74], a general model for joint scheduling, power allocation and routing optimization problem is studied for UWB networks. The most important finding is that the optimal MAC protocol in an UWB network should be a combination of rate adaptation and mutual exclusion. That is, to improve the resource utilization in UWB networks, it is optimal to allow concurrent transmissions, as long as all interferers are outside a well-defined ER around the destinations. Based on the mutual exclusion concept, exclusive region based resource allocation scheme appears. In [78], two simple heuristic scheduling algorithms with polynomial time complexity are presented. In [79], a multi-class scheduling problem for UWB networks supporting heterogeneous traffic is formulated as a utility maximization problem under the fairness constraint. They show that ER based scheduling scheme can greatly improve the network performance with a carefully selected ER. However, how to determine the best ER size in a random network to maximize the capacity of UWB WPANs remains an open research issue.

On the other hand, considering the salient features of mmWave communications with directional antenna, directional MAC design in mmWave WPANs has been emerging recently. An mmWave WPAN architecture is proposed in [80], where an intermediate node is selected as the relay when the LOS link between source and destination is blocked by moving obstacles. The proposed architecture is shown to be effective to keep the network connectivity when serious link outage happens due to obstacles. An enhanced MAC with spatial reuse TDMA is presented in [81]. Simulations show that the enhanced MAC improves the throughput and delay performance by allowing spatial reuse in mmWave channel. To the best of our knowledge, there is little theoretical work on the resource management issues in mmWave WPANs. How to design ER based MAC and determine appropriate spatial reuse parameters in the mmWave WPAN is beckoning for further investigation.

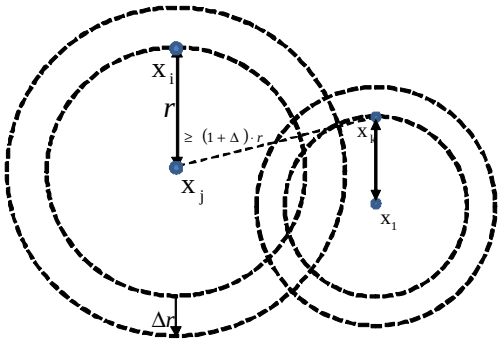


Figure 5.1: Protocol Model

5.1.3 Capacity Analysis

Optimizing resource utilization of wireless networks is crucial because wireless resource in general is at a premium. To meet the ever-growing demand of anytime, anywhere, higher data-rate wireless connectivity, we are striving to explore the wireless capacity in a *networked* environment. In the pioneer work of Gupta and Kumar [82], the asymptotic bounds of network transport capacity have been derived, given the node density in an arbitrary or random wireless network. To quantify the network capacity, two transmission models are proposed, namely, the protocol model and physical model, which define an exclusive region or guard zone between the transmitter and interferers. They extended their work to three dimensional space in [83], considering all nodes are located in a sphere. Since then, a large number of follow-up papers has appeared. The impact of user mobility on network capacity has been studied in [84, 85]. It was found that node mobility can be exploited to increase the network capacity [84], but excessive mobility contrarily limits the capacity [85]. The capacity of different networks with different traffic patterns, *e.g.*, relay traffic, convergent traffic (in sensor networks), and broadcast traffic, have been investigated in [86–89]. The upper and lower bounds of throughput capacity of UWB networks have been derived in [90, 91], considering

power constraints of UWB communications and link layer packet loss. It was found in [90] that different properties of the physical layer may dramatically alter the network capacity. In contrast with the result in [82] that the capacity per node is a decreasing function of node density, the capacity bounds derived in [90] increase with the node density, assuming that the interference is negligible for UWB networks with low transmission power. To the best of our knowledge, most of the previous work studied the capacity region, or the upper/lower bounds on network capacity are based on the same transmission models proposed in [82]: the protocol model and the physical model. In the protocol model, the transmission between two nodes X_i and X_j is successful if

$$|X_k - X_j| \geq (1 + \Delta)|X_i - X_j|$$

for every other node X_k simultaneously transmitting over the same subchannel. In other words, to ensure the transmission from X_i to X_j to be successful, there is a guard zone centered at X_j with radius proportional to $|X_i - X_j|$, as shown in Fig. 5.1 [82]. No other node inside the guard zone should transmit concurrently. In the physical model, a transmission from a node X_i is successfully received by a node X_j if

$$\frac{\frac{P_i}{|X_i - X_j|^\alpha}}{P_n + \sum_{k \neq i} \frac{P_k}{|X_k - X_j|^\alpha}} \geq \beta \ ,$$

where P_i and P_k are the transmission power levels of node X_i and X_k, respectively, P_n is the ambient noise power level, and α is the path loss exponent. This model indicates that the transmission from X_i to X_j can be successful only if the SINR at the receiver exceeds a certain threshold. If not, the transmission will fail and the flow throughput will be zero. In these two models, the transmission rate is a binary function, which is simple and thus analytically attractive for capacity analysis. However, with the recent advances in wireless communication technologies, the two models are no longer applicable in some emerging wireless networks. For instance, in a UWB system, the sender can adjust the data rate according to the received SINR. Such rate adaptation

scheme is also used in other wireless communication systems including IEEE 802.11 WLAN, IEEE 802.15.3c, and IEEE 802.16 WiMAX. Therefore, even if there is an interferer which is located inside the guard zone specified in the protocol model in [82], the flow throughput may not immediately drop to zero. Similarly, there is no single threshold β (in the physical model in [82]) that decides the transmission rate, which is adaptive according to SINR. To model realistic UWB and mmWave networks, it is necessary to employ a more general communication model that captures the rate adaptation in the physical layer. In addition, the asymptotic capacity bounds derived in the previous work, especially those for arbitrary networks, may be too loose to be useful in realistic networks where the network topology is random and may change from time to time due to user mobility. Therefore, we are more interested in deriving the expected network capacity or network transport capacity of a random network, and maximizing them for MAC layer resource allocation by fine tuning the protocol parameters.

5.2 Performance Analysis of IEEE 802.15.3 MAC

We consider a WPAN consisting of N users, as shown in Fig. 3.1(C). The PNC serves as an AP or gateway when a piconet user establish a two-way multimedia connection with a correspondent user outside the WPAN. Let the achievable data rate with transmission distance $d_{\max}$ be R Mbps, and the distance between the i-th user to the PNC be d_i, where $0 \leq d_i \leq d_{\max}$ for $i = 1, 2, ...N$. The achievable data rate of the i-th user is $R_i = (d_i/d_{\max})^{-\alpha} R$. To support a number of multimedia connections in a WPAN, we first quantify the appropriate CAP duration to ensure that the requests for channel times can be successfully sent to PNC. Second, we calculate the number of multimedia connections that can be accommodated in the CTAP period. We then substantiate the analysis by calculating how many voice and video calls can be supported in a high rate WPAN [9].

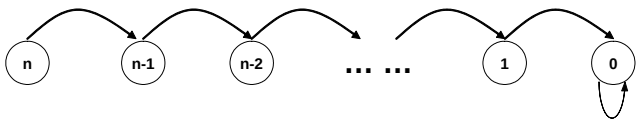


Figure 5.2: Process of Frame Service

5.2.1 Contention Access Period

For constant rate multimedia traffic, *e.g.*, voice applications, the user may only need to send one request per call and reserve channel time for contention free data transmission in the CTAPs. Therefore, we denote the number of contending users at the beginning of each CAP as $n \leq N$. As shown in Fig. 5.2, the process of frame service continues until all channel requests are served. Note that a served frame can be either successfully transmitted or dropped due to excessive retransmissions. Define $p|k$ the conditional collision probability and $\tau|k$ the transmission probability of a user at any slot when there are k contending users in the piconet. Conditioning on the number of contending user k, we have

$$p|k = 1 - (1 - \tau|k)^{k-1}. \tag{5.1}$$

The transmission probability $\tau|k$ can be derived from (4.7) and is a function of $p|k$. Conditioning on k contending users in the piconet, a frame is dropped with probability $(p|k)^{m+1}$ when the retry limit m is reached, and successfully transmitted with probability $1 - (p|k)^{m+1}$. Therefore, during the contention period, the total number of successfully transmitted requests can be approximated as

$$n_s \approx \sum_{k=1}^{n} (1 - (p|k)^{m+1}), \tag{5.2}$$

and the number of failed requests is $n_f = n - n_s$. To avoid collisions, the failed requests will not contend with others in the same CAP, but will be re-initiated in the next contention period.

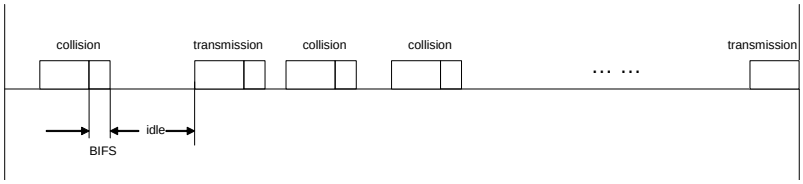


Figure 5.3: Channel State During the Contention Access Period

For a given number of users contending at the beginning of each CAP, the duration of the contention periods in a superframe can be considered as a renewal process. During each CAP, the number of contending users monotonically decreases with time. To analyze the duration of the CAP, we need to track n at every time slot, which can be very complicated. To simplify the analysis, we consider the average number of contending users over the CAP, which is $l = n/2$. We can then obtain the average conditional collision probability p and average transmission probability τ of a user. We investigate the network performance from the viewpoint of the channel status, which can be either idle or busy. The channel is idle only when there is no transmission, and sensed busy because of collisions or successful transmissions, as shown in Fig. 5.3. Thus, the length of CAP (T_{CAP}) is the summation of the busy periods and idle periods. Here we consider an ideal duration of CAP that ends when the last request is served.

$$T_{CAP} = E[N_{busy}] \cdot T_{busy} + E[N_{idle}] \cdot T_{slot}, \tag{5.3}$$

where $E[N_{busy}]$ is the average number of busy slots, $E[N_{idle}]$ is the average number of idle slots, and T_{busy} and T_{slot} are the duration of a busy slot and an idle slot, respectively. According to IEEE 802.15.3 [20], $T_{slot} = 17.3\ \mu s$, and $T_{busy} = T_{req} + SIFS + T_{ACK} + BIFS$, where T_{req} is the transmission time of a request. T_{busy} is constant with a fixed request frame length. To obtain the duration of the CAP, we need to compute $E[N_{busy}]$ and $E[N_{idle}]$. Since the channel is idle only when all requesting users are in the backoff stage, $E[N_{idle}]$ is the average backoff time the last requesting user experiences, which

can be approximated as

$$E[N_{idle}] = \sum_{i=0}^{m-1} p^i(1-p) \sum_{j=0}^{i} \frac{2^j W - 1}{2} + p^m \sum_{j=0}^{m} \frac{2^j W - 1}{2}. \tag{5.4}$$

We calculate $E[N_{busy}]$ from its probability distribution function

$$E[N_{busy}] = \sum_{i=0}^{N_B} i \cdot P[N_{busy} = i], \tag{5.5}$$

where N_B is the maximum number of busy slots. For a retry limit m, n frames can be transmitted at most $(m+1)n$ times. During the CAP, n request frames can be served, but only n_s are successfully transmitted. Therefore, among $(m+1)n$ busy slots, n_s are successful transmissions and $(m+1)n - n_s$ are collisions. With most collisions caused by two simultaneous transmissions, the maximum number of busy slots is $N_B = [(m+1)n - n_s]/2 + n_s = [(m+1)n + n_s]/2$.

Given the channel is busy, implying at least one user is transmitting in the given slot, two events may occur: 1) a collision resulting from multiple simultaneous transmissions; 2) a successful transmission when only one user transmits in that slot. Thus, the probability of a collision and a successful transmission can be derived as

$$p_c = \frac{1 - (1-\tau)^l - l\,\tau(1-\tau)^l}{1 - (1-\tau)^l}, \tag{5.6}$$

$$p_s = \frac{l\,\tau(1-\tau)^l}{1 - (1-\tau)^l}. \tag{5.7}$$

The number of busy slots can be considered as a binomial random variable. In other words, in i busy slots, n_s slots are due to successful transmissions and the remaining $i - n_s$ slots are due to collisions.

$$P[N_{busy} = i] = \begin{pmatrix} i \\ n_s \end{pmatrix} (p_s)^{n_s} (p_c)^{i-n_s} \tag{5.8}$$

Substitute (5.4)-(5.8) into (5.3), we can obtain the CAP duration in which n requests can be served but only n_s requests are transmitted successfully.

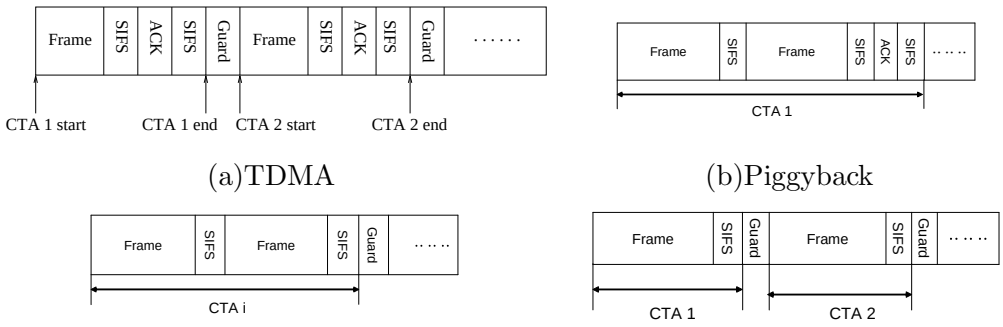


Figure 5.4: Channel Time Allocation Policies

5.2.2 Channel Time Allocation Period

After receiving requests, the PNC can allocate network resources to the requesting flows and, in the mean time, reserve resource for the isochronous flows that have requested in the previous suprframes.

TDMA — In IEEE 802.15.3 adopting standard TDMA, guard times (T_g) are required to keep transmissions in adjacent CTAs from colliding. Assume all multimedia connections have a constant packet inter-arrival time. If T_{sf} is less than the packet inter-arrival time, all N connections can be supported in the UWB network. Then the duration of the superframe, T_{sf}, is

$$T_{sf} = T_{beacon} + T_{CAP} + (2N+1)T_g + 2NT_{CTA}. \tag{5.9}$$

where T_{CTA} is the time unit for a channel time allocation. In the CTAP, a SIFS time is required to ensure sufficient turnaround time between transmissions,

$$T_{CTA} = k(T_{frame} + SIFS + T_{ACK} + SIFS),$$

where $k = 1$ without link layer fragmentation and $k > 1$ in other cases. Denote L the payload of a packet. The frame payload, including L and RTP/UDP/IP headers (RUI_h), will be transmitted at a data rate of R_i. The PHY header (PHY_h), MAC header (MAC_h), Header Check Sequence (HCS) and Frame Check Sequence (FCS)

will be transmitted at a lower basic rate of R_0. The preamble is used for clock/carrier acquisition and receiver training. The length of the preamble should be long enough for synchronization between the users in the network, *i.e.,* the preamble time T_a of the i-th user is determined by d_i. Therefore, T_{frame} and T_{ACK} are obtained as

$$T_{frame} = T_a + (p_l + RUI_h)/R_i + (PHY_h + MAC_h + HCS + FCS)/R_0, \quad (5.10)$$

$$T_{ACK} = T_a + (PHY_h + MAC_h + HCS)/R_0. \quad (5.11)$$

Since $d_i \leq d_{\max}$ for all connections, the lower bound of the number of multimedia connections being supported, N_{min}, can be obtained as

$$N_{min} = \frac{T_{sf} - T_{beacon} - T_{CAP} - T_g}{2T_g + 4k\ SIFS + 2k\ T_{frame} + 2k\ T_{ACK}} \quad (5.12)$$

Piggyback — As shown in Fig. 5.4(b), the PNC only needs to allocate one CTA for a two-way multimedia connection by allowing the PNC and the user to exchange their frames. Therefore, N CTAs are required for N multimedia connections while $2N$ CTAs are needed when traditional TDMA is used. The duration of the superframe is

$$T_{sf} = T_{beacon} + T_{CAP} + (N+1)T_g + NT_{CTA}, \quad (5.13)$$

where $T_{CTA} = 2kT_{frame} + (2k+1)SIFS + T_{ACK}$. And N_{min} is obtained as

$$N_{min} = \frac{T_{sf} - T_{beacon} - T_{CAP} - T_g}{T_g + (2k+1)SIFS + 2k\ T_{frame} + T_{ACK}} \quad (5.14)$$

No-ACK — With error-resilient source coding schemes, many multimedia applications can tolerate a certain degree of packet loss, and thus *No-ACK* policy can be used to reduce the ACK overhead. In this case, T_{CTA} is reduced to $T_{CTA} = kT_{frame} + k\ SIFS$, as shown in Fig. 5.4(c). Accordingly, N_{min} is obtained as

$$N_{min} = \frac{T_{sf} - T_{beacon} - T_{CAP} - T_g}{2T_g + 2k\ SIFS + 2k\ T_{frame}} \quad (5.15)$$

The overhead of guard times can be further reduced by allowing the sender and receiver to exchange data in one CTA, With the combined CTA, the duration of the superframe

is

$$T_{sf} = T_{beacon} + T_{CAP} + (N+1)T_g + NT_{CTA}, \tag{5.16}$$

where $T_{CTA} = 2kT_{frame} + 2k\ SIFS$, as shown in Fig. 5.4(d). Compared with separate CTA, the number of guard time is reduced from $2N_f + 1$ to $N_f + 1$. Therefore, N_{min} is obtained as

$$N_{min} = \frac{T_{sf} - T_{beacon} - T_{CAP} - T_g}{T_g + 2k\ SIFS + 2kT_{frame}} \tag{5.17}$$

No-ACK can effectively reduce the ACK overhead at the cost of reliability because it disables the link layer recovery. Therefore, *No-ACK* is not desired when channel error rate is high. On the other hand, *piggyback* is highly recommended when the users in the WPAN carry symmetric traffic, such as two-way voice communications. With dynamically asymmetric traffic, it is difficult for the PNC to efficiently allocate combined CTAs for a pair of sender and receiver.

5.2.3 Numerical Results

To substantiate the analysis, we first calculate the voice capacity of a DS-UWB WPAN as an example, since voice capacity of other wireless networks (*e.g.*, cellular, WLAN) has been heavily studied and it can be used for direct comparison. We then calculate the maximum number of high data rate multimedia connections being supported in a WPAN with different data rates. To maximize the capacity, we obtain the required minimum duration of CAP. Extensive simulations are conducted to validate the analysis using NS-2 [62].

Let a UWB PNC support a large number of VoIP calls with a comparatively large range, 10 m. The system parameters for DS-UWB are listed in Table 5.1. DS-UWB supports two independent bands of operation. The lower band occupies the spectrum from 3.1 GHz to 4.85 GHz and the upper band occupies the spectrum from 6.2 GHz to 9.7 GHz. R_0 is 28 Mbps for the lower band, and 55 Mbps for the upper band. We

Table 5.1: System Parameters (DS-UWB)

Data rate R	110 Mbps
Basic rate R_0	28 Mbps
Preamble T_a	5 or 15 or 30 μs [92]
PHY header	2 Bytes [92]
MAC header	10 Bytes [20]
HCS	2 Bytes [92]
FCS	4 Bytes [92]
SIFS	10 μs [20]
IP/UDP/RTP header	40 Bytes

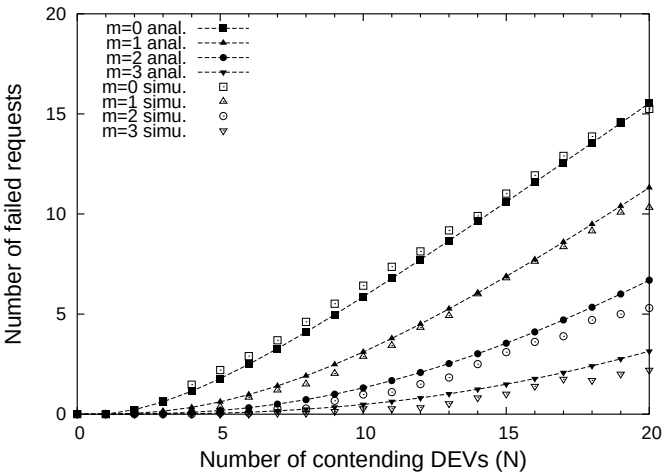


Figure 5.5: Number of Failed Requests

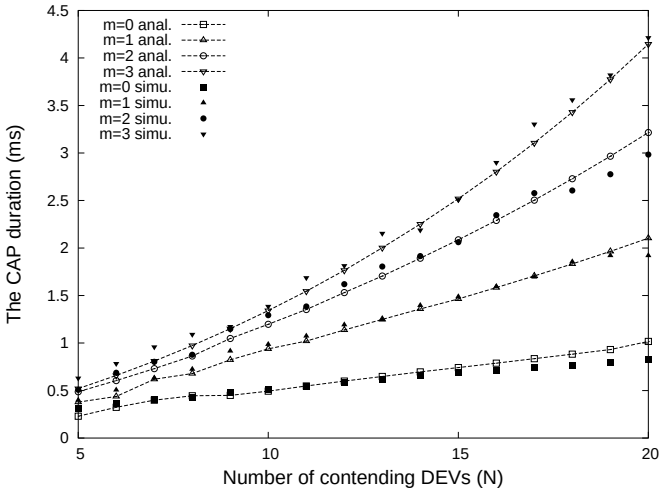


Figure 5.6: The CAP Duration (ms)

choose the basic rate $R_0 = 28$ Mbps and data rate $R = 110$ Mbps to calculate the voice capacity in this paper. There are three preambles defined in [92]: 1) short preamble: 5 μs in length that requires a high SINR with low channel dispersion – it is most suitable for short range links (< 3 meters); 2) nominal preamble, 15 μs in length that requires a nominal SINR with a nominal channel – it is the default preamble choice; and 3) long preamble, 30 μs in length that is used for a poor SINR and/or highly dispersive channel. Long preamble is intended for extended range applications. The Beacon of each superframe contains variable number of information elements (IEs), 4 bytes FCS, and 21 bytes synchronization parameters [20]. The beacon length is approximately 200 bytes plus $34n$ bytes for CTA IEs and CTA status IEs. Guard times are calculated based on the worst-case drift in a superframe and the maximum allowed number of lost beacons g [20], $T_g = g * (\text{Clock accuracy}/10^6) * T_{sf}$, where $g = 10$ and the clock accuracy is ± 25 ppm. The voice codecs used for performance evaluation are tabulated in Table 3.1.

We first determine the proper duration of CAP. At the beginning of a superframe, each contending user sends a request with the starting time randomly chosen over the

minimum contention window, *i.e.*, $[0, 7]$. The required CAP duration is obtained that n requests are served, either successfully received or dropped. The channel request frames are transmitted at the basic rate of 28 Mbps and the voice/video frames are transmitted at a data rate of 110 Mbps. Fig. 5.5 shows the number of failed requests for a given number of contending users and retry limit m. It is observed that the increase rate of the failed transmissions goes higher with a larger n and smaller m. In IEEE 802.11, the minimum contention window CW_{min} and m are relatively large ($CW_{min} = 32$ and $m = 7$), and most frames will be eventually transmitted successfully. But in IEEE 802.15.3, due to the smaller minimum contention window ($CW_{min} = 8$) and retry limit ($m \leq 3$), the collision probability could be very high when n is large. The request frames are more likely to be dropped and thus fewer requests can be successfully transmitted. If a request is dropped in the CAP, the user needs to re-initiate the request in the following superframe, which degrades the network performance due to the waste of network resources caused by retransmissions, and also affects user-perceived QoS with a longer delay. The duration of the CAP for a given n and m is shown in Fig. 5.6. Note that not all served requests can be successfully transmitted. For $m \leq 1$, most request frames are dropped due to excessive transmissions for a large n, as shown in Fig. 5.5, and it is no use to enlarge CAP in this case. It is recommended that a large m be used to improve the number of successful requests when n is large. Although it is not desirable for the PNC to change the CAP duration in every superframe, it is necessary for the PNC to determine the CAP duration from time to time, based on the varying number of contending users in the network. For real time multimedia connections that require channel time in a regular basis, we expect that each connection can last for a certain period, and there are only a limited number of new connections that contend in the CAP. In the simulation, we set the retry limit to the maximum value defined in IEEE 802.15.3, i.e., $m = 3$. We choose $T_{CAP} = 2$ ms when $T_{sf} = 10$ ms, which is suitable to serve 12 contending users. The T_{CAP} increases by 0.5 ms when

Table 5.2: Number of VoIP Connections Supported in a UWB WPAN with Different Preambles and Voice Codecs

Audio	G.711			G.729			G.723			iLBC		
ms	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs
10	71	52	37	79	56	39						
20	137	104	76	163	118	84				157	115	82
30	187	146	110	238	175	126	238/237	176/175	126/126	229	171	124
40	227	181	139	304	228	165						
50	259	211	165	364	276	203						
60	286	237	188	418	321	238	433/431	330/329	243/242			

Table 5.3: Number of Video Connections Supported in a UWB Network with Different Data Rates

Preamble	Data Rate				
(μs)	1 Mbps	2 Mbps	3 Mbps	4 Mbps	5 Mbps
5	32	16	11	8	6
15	28	14	9	7	5
30	24	12	8	6	4

the superframe length increases by 10 ms to accommodate more connections. When $T_{sf} = 60$ ms, $T_{CAP} = 4.5$ ms, which is suitable for serving 20 contending users.

The number of voice calls supported in a UWB WPAN for different voice codecs is listed in Table 5.2. The voice capacity of G.729 is slightly higher than that of G.711 (less than 30%), although the payload of G.729 is much smaller than that of G.711. High compression codec cannot significantly improve the voice capacity of a high rate UWB network. Thus, for low rate voice applications over high rate UWB networks, it is recommended to use low compression codecs with low complexity and high voice quality. The voice capacity of UWB networks will surpass that of all existing wireless technologies by a large margin. UWB has great potential to support other high data rate multimedia services that are otherwise impossible with existing narrow band communications technologies. On the other hand, without the protocol overheads, a

Table 5.4: Number of Connections Supported in a UWB Network Using Different Policies

MAC Policies	Data Rate						
	8 kbps	64 kbps	1 Mbps	2 Mbps	3 Mbps	4 Mbps	5 Mbps
TDMA	39	37	20	10	7	5	4
Piggyback	51	48	23	12	8	6	5
No-ACK with Separate CTA	70	64	27	13	9	7	5
No-ACK with Combined CTA	72	66	27	13	9	7	5

110 Mbps channel should be able to support $110000/64 \approx 1718$ G.711 voice calls with 64 Kbps codec rate. The actual capacity of a UWB network is far below that upper bound, which reflects the inefficiency of the protocol. Some multimedia applications require high data rate. We list some examples of high data rate video connections and calculate their capacity in Table 5.3. The superframe length is 30 ms, and the CAP duration is 3 ms. With link layer fragmentation, one video packet corresponds to several link layer frames transmitted in a CTA. The average payload of a frame is 1250 bytes. Therefore, for a 2 Mbps video flow, the number of link layer frames transmitted in a 30 ms interval is $2 \times 10^6 \times 30 \times 10^{-3}/(1250 \times 8) = 6$. It is observed that the number of high rate video connections being supported is very limited due to the high bandwidth requirement and protocol overheads. Only 32 one Mbps two-way video connections can be supported, and the video capacity decreases proportionally when the rates of video connection increases. Compare to voice connections, where the payload size is quite small and no frame fragmentation is needed, the fragmentation overheads of video traffic (*e.g.*, SIFSs between frames, header overheads in each frame, *etc.*) are quite large. The number of supported connections is even less if high definition television (HDTV) is considered, which requires high bandwidth from tens of Mbps (using efficient compression algorithms) to up to 1.5 Gbps (without compression). Thus, there remain considerable issues that need further improvement.

The numbers of supported voice or video connections under different policies are compared in Table 5.4. The preamble is 30 μs and superframe length is 10 ms. It is observed that with *piggyback* and *No-ACK* policies, the number of high data rate video connections being supported can only be improved by around $15-35\%$, while the voice capacity can be improved by $60-85\%$. This is because the overheads of SIFS and ACK become relatively small compared with the large video payload. Therefore, piggyback is more efficient for small payload and highly recommended for two way symmetric, low data rate voice applications. With a limited number of high data rate connections, the guard times are negligible and there is no capacity difference under separate and combined CTA policies.

5.2.4 Discussion

IEEE 802.15.3 is a hybrid MAC combining CSMA/CA based random access and TDMA based guaranteed access to provide a certain level of QoS for multimedia application. Due to the large protocol overheads resulting from contentions in CSMA/CA-based CAP, it is highly recommended data be transmitted in a contention-free manner and only commands, *e.g.*, channel time request, be transmitted in the CAP. Moreover, channel time requests should be performed on a per-flow basis rather than on a per-packet basis. The numerical results indicate although a UWB WPAN can provide very high voice capacity for low rate voice calls, the number of high rate video flows supported in a UWB WPAN is still limited due to the protocol inefficiency and overheads. Thus, it is essential to explore enhanced MAC design in IEEE 802.15.3 based WPANs to support high volume multimedia applications.

Different from narrow band communications where simultaneous transmissions in nearby neighbors collides with each other, wideband communication system can support multiple concurrent transmissions if the multi-user interference is properly managed. On the other hand, taking advantage of precise localization capability of UWB system,

location-aware MAC is possible. Therefore, to improve the network capacity for high volume multimedia applications, we can utilize the network topology information for efficient interference management and explore the spatial reuse opportunity in UWB channel for appropriate concurrent transmissions. The use of directional antenna in mmWave networks can further improve the network capacity with a higher directivity gain and more efficient spectrum reuse. In the following sections, we will present the detail enhanced MAC design and theoretical analysis of the network capacity of UWB/mmWave WPANs.

5.3 ER-based MAC Design

5.3.1 Exclusive Regions with Different Types of Antenna

We consider a IEEE 802.15.3 based UWB/mmWave WPAN consisting of N_f flows, $\{f_i, |i \in 1, 2, ..., N_f\}$, with the piconet as the basic network element, as shown in Fig. 3.1(C). According to the Shannon limit of an additive white Gaussian noise (AWGN) channel, the achievable data rate of f_i is

$$R_i = W \log_2(1 + \frac{P_R(i)}{N_0 W + \sum_{j \neq i, j \in \gamma} I_{j,i}}) \tag{5.18}$$

where $P_R(i)$ is the receiving power of flow f_i, N_0 is the one-sided power spectral density of additive white Gaussian noise, W is the signal bandwidth, γ is the set of flows concurrently transmitting with flow f_i, and $I_{j,i}$ is the interference power of the sender of flow f_j to the receiver of flow f_i. Due to the low power level of UWB systems, all flows use the maximum transmission power P_T. Denote the distance between the sender and receiver of the i-th flow as d_i, and the distance between the sender of flow j and the receiver of flow i as $d_{j,i}$. The received signal power of flow f_i is given by

$$P_R(i) = k_1 P_T G_T(i) G_R(i) d_i^{-\alpha} S_i, \tag{5.19}$$

where $k_1 \propto (\lambda/4\pi)^2$ is a constant coefficient dependent of the wavelength λ, α is the path loss exponent dependent of the propagation environment and usually takes the value between 2 to 6 [93], and S_i captures the fast fading characteristics and equals the instantaneous channel gain normalized by its time-average value. Denote G_0 as the cross correlation between coefficient between two concurrent flows. The received interference from flow j is

$$I_{j,i} = k_1 P_T G_0 G_T(j) G_R(i) d_{j,i}^{-\alpha} S_{j,i} \tag{5.20}$$

Let k flows transmit in k time slots in a serial TDMA fashion, each flow achieves a higher instant throughout in its assigned slot without interference and the achievable throughput over k slots is

$$R_i^T = \frac{1}{k} W \log_2(1 + \frac{k_1 P_T G_T(i) G_R(i) d_i^{-\alpha} S_i}{N_0 W}). \tag{5.21}$$

If all flows can transmit concurrently in each of the k slot, the instant throughput of each flow is lower due to mutual interference but each flow can transmit during more slots, and the achievable throughput of f_i is

$$R_i^C = W \log_2(1 + \frac{k_1 P_T G_T(i) G_R(i) d_i^{-\alpha} S_i}{N_0 W + \sum_{j \neq i}(I_{j,i})}). \tag{5.22}$$

To investigate when concurrent transmission is preferable, we first ignore the short term fast fading effect and consider long term path loss fading to compare the achievable throughput of the serial TDMA transmission scheme and concurrent transmission scheme. To compare R_i^T and R_i^C, we consider two cases separately. First, if $SINR << 1$, the achieved data rate can be approximated as

$$k_1 W \log_2(\text{SINR} + 1) \approx k_1 W \times \text{SINR} \log_2 e. \tag{5.23}$$

With this approximation, from (5.21) and (5.22), a sufficient condition to ensure that $R_i^C \geq R_i^T$ is $I_{j,i} \leq N_0 W, \quad \forall j \neq i$, *i.e.,* the average interference power from any other

flow should be less than the background noise[1]. Thus, if we allow flows with mutual interference less than that of the background noise to transmit simultaneously, the throughput of each flow can be higher than that of serial TDMA transmissions.

Second, if SINR is close to or larger than one, the approximation in (5.23) does not hold when SINR ≥ 1. Nevertheless, the previous derived sufficient condition can still ensure that $R_i^C \geq R_i^T$. Since $\log_2(x/k+1) \geq (1/k)\log_2(x+1), \ \ \forall x \geq 1, \ k \geq 1$, and if $I_{j,i} \leq N_0 W$, $R_i^C/R_i^T \geq 1/k \log_2(\text{SNR}+1)/\log_2(\text{SNR}/k+1) \geq 1$. Thus, the derived sufficient condition is always applicable.

To ensure that the interference power from a concurrent flow is less than the noise level, we should not allow any interferer inside an ER around the receiver. In other words, an interferer, the sender of f_j, should be at least $r(i)$ away from the receiver of the i-th flow, where $r(i)$ is given as

$$r(i) = (\frac{k_1 G_0 G_T(j) G_R(i) P_T(j)}{N_0 W})^{1/\alpha}. \tag{5.24}$$

In the following, we consider three cases and derive ERs according to the types of transmitting and receiving antennae, *i.e.*, omni- or directional.

Case 1: Omni-antenna to Omni-antenna

In this case, both the transmitters and receivers use omni-antennae, $G_T(i) = G_R(i) = 1, \forall i \in 1, 2, ..., N_f$. To ensure that the interference from any interferer to be lower than the noise level, all interfering sources should be at least r_0 away from the receiver of the i-th flow ($d_{j,i} \geq r_0$), where r_0 is given by

$$r_0 = (\frac{k_1 G_0 P_T}{N_0 W})^{1/\alpha}. \tag{5.25}$$

Therefore, the ER is a circle centered at the receiver, with radius r_0, as shown in Fig. 5.7 (a).

[1]The necessary and sufficient condition to ensure that $R_i^C \geq R_i^T$ is $\sum_{j \neq i} I_{j,i} \leq (k-1) W N_0$, where flow j is scheduled to transmit concurrently with flow i. The sufficient condition given in the main text is more conservative, but it allows to design much simpler and practically more feasible scheduling algorithms.

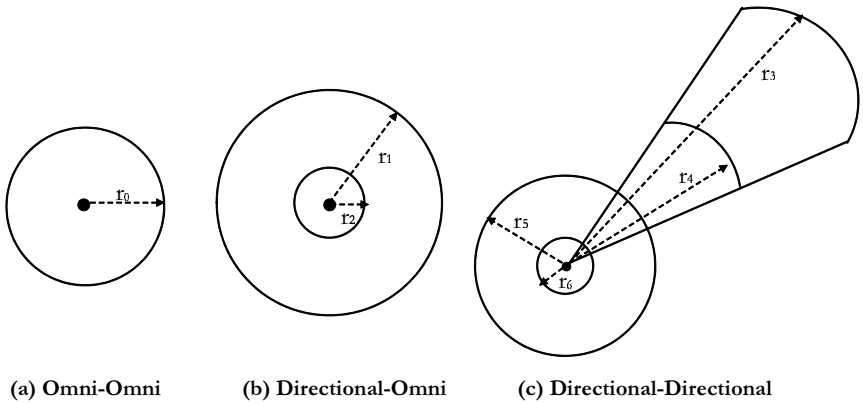


Figure 5.7: Exclusive Regions for Omni-directional and Directional Antennae.

Case 2: Directional-antenna to Omni-antenna

In the second case, the transmitter antennae are directional and the receiver antenna are omni-directional ($G_R(i) = 1$). The directional antenna pattern consists of a mainlobe of gain G_{T_M} with beamwidth θ and a sidelobe of gain G_{T_S} with beamwidth $2\pi - \theta$.

As shown in Fig. 5.7(b), if a receiver is inside the radiation angle of an interferer, the average interference is $I_{j,i} = k_1 G_0 G_{T_M} P d_{j,i}^{-\alpha}$. Thus, an interferer should be outside the circle centered at the receiver with radius r_1:

$$r_1 = (\frac{k_1 G_0 G_{T_M} P_T}{N_0 W})^{1/\alpha}. \tag{5.26}$$

If a receiver is outside the radiation angle of an interferer, we have $I_{j,i} = k_1 G_0 G_{T_S} P d_{j,i}^{-\alpha}$, and the ER is a circle with radius r_2:

$$r_2 = (\frac{k_1 G_0 G_{T_S} P_T}{N_0 W})^{1/\alpha}. \tag{5.27}$$

If an ideal directional antenna is considered, $G_{T_S} = 0$, and the ERs are reduced to one circle with radius r_1.

Case 3: Directional-antenna to Directional-antenna

When both the transmitter and receiver antennae are directional, the ER contains four zones. If an interferer is located within the beamwidth of the receiver, and the receiver is also within the beamwidth of the interferer, the interferer should be at least r_8 away from the receiver:

$$r_3 = (\frac{k_1 G_0 G_{T_M} G_{R_M} P_T}{N_0 W})^{1/\alpha}. \tag{5.28}$$

Therefore, the first ER zone is a cone with angle θ and radius r_8.

If an interferer is within the radiation angle of the receiver, but the receiver is outside the radiation angle of the interferer, the second ER zone is a cone with angle θ and radius r_6:

$$r_4 = (\frac{k_1 G_0 G_{T_S} G_{R_M} P_T}{N_0 W})^{1/\alpha}. \tag{5.29}$$

If an interferer is outside the radiation angle of the receiver with its radiation beamwidth toward the receiver, the third ER zone is a sector with angle $2\pi - \theta$ and radius r_7:

$$r_5 = (\frac{k_1 G_0 G_{T_M} G_{R_S} P_T}{N_0 W})^{1/\alpha}. \tag{5.30}$$

If both the interferer and the receiver are outside of each other's radiation beamwidth, the last ER zone is a sector with angle $2\pi - \theta$ and radius r_5:

$$r_6 = (\frac{k_1 G_0 G_{T_S} G_{R_S} P_T}{N_0 W})^{1/\alpha}. \tag{5.31}$$

The four ER zones for this case are shown in Fig. 5.7 (d). Similarly, if ideal directional antenna is assumed, four ER zones are reduced to one cone with radius $r8$.

Different from the guard zone specified in the protocol model in [82], we derive a general exclusive region surround the receiver, which is independent of the transceiver distance. The total interference level at the tagged receiver can be bounded for a given ER. In this section, we give a sufficient condition to show that concurrent transmissions are preferable. A detailed concurrent scheduling scheme and parameter setting will be presented in Sections 5.3.2 and 5.4, respectively.

5.3.2 Exclusive Region based Concurrent Transmissions

In UWB/mmWave WPANs, concurrent transmissions are possible and favorable compared with serial TDMA transmissions if all interfering users are sufficiently far apart, *i.e.*, outside a well defined ER of the other receivers. In other words, network throughput can be greatly improved by exploiting the spatial reuse opportunity of the wireless channel for concurrent transmissions. With a random network topology, the optimal scheduling problem for peer-to-peer concurrent transmissions is known to be *NP-hard* [78, 94]. Unlike the traditional scheduling problems, each flow's throughput per time slot in UWB/mmWave WPANs is unknown before the scheduling decision, and it depends on network topology, user deployment, transmission power, cross-correlations of interfering signals, and the scheduling decision itself.

In the following, we propose REX [10, 95], a randomized ER based scheduling scheme in a centralized mmWave WPAN, with computational complexity $O(N_f^2 \log N_f)$ to allocate a time slot. The PNC of a WPAN has the global information of the WPAN, *e.g.*, the number of active flows and their demands for network resources, and the location of all users, *etc.*, based on which the PNC schedules peer-to-peer transmissions for active flows. Denote the set of all active flows as $S\{N_f\}$ of N_f elements. A subset of flows $\gamma_i \subset S\{N_f\}$ contains the flows scheduled in slot i that satisfy the conditions favoring concurrent transmissions, as derived in (5.25)-(5.31). Denote FS the set of scheduled flows in $S\{N_f\}$ and $T_a(j)$ a scheduling weight associate with flow j. $T_a(j)$ can be defined as a function of the allocated network resource (e.g., the achieved throughput), the resource demands, traffic class, and buffer occupation, etc. In this work, we define $T_a(j)$ the number of slots allocated to flow j. Initially, $FS = \gamma_i = NULL$ and $T_a = 0$ for all flows in any slot. The proposed REX operates as follows.

- Step 1: Randomly choose one flow with the minimum T_a and schedule it in slot i (initially, $i = 1$ for the first slot). Add this flow to the subsets γ_i. If the flow is not included in FS, add it to FS;

- Step 2: Check all the remaining active flows in the set $S\{N_f\} - \gamma_i$ for concurrent transmission conditions, starting from the flow with the smallest $T_a(j)$. If any flow satisfies the concurrent transmission condition, *i.e.*, the new flow and the flows in set γ_i are mutually outside each other's exclusive regions, add it to γ_i and increase $T_a(j)$ of the flow by one. If this flow is not included in FS, add it to FS;

- Step 3: Increase the slot number i by one and sort flows according to $T_a(j)$ in ascending order;

- Step 4: Repeat Steps 1-3 until all flows are scheduled, $FS = S\{N_f\}$.

The procedure can also be repeated until the requirements of all active flows are fully satisfied. It is worth noting that although sorting flows according to their $T_a(j)$ in step 2 will increase the computational complexity by $O(N_f \log N_f)$, it is essential for maintaining fairness among flows. If we search flows in a deterministic sequence for slot allocation, those flows with smaller sequence number are more likely to be scheduled in γ_i. This will cause serious unfairness problem, as shown in the simulation results in a later section. With the searching sequence used in Step 2, the maximum access delay of all flows can be bounded.

The results of whether two flows are mutually exclusive can be saved in a look-up table to reduce the execution time of REX. In WPANs, the mobility is typically low, e.g., less than 1 m/s, and the superframe duration is usually on the order of tens of ms[2]. Thus, the node movement is normally much less than 0.1 m during the superframe duration. Such small change in location will not significantly affect the received power and interference power level, and it is acceptable to ignore mobility for scheduling decision. Due to low mobility in WPANs, the frequency of updating the table of mutual exclusive condition is also low. In the case a transmitter or a receiver detects

[2]The maximum length of a superframe specified in IEEE 802.15.3 is 65535 μs.

the topology change by measuring the link quality between them, the neighbor discovery (ND) process embedded in the association process will be automatically triggered and the network topology will be updated accordingly. In this work, we focus on the enhanced MAC layer resource allocation and do not consider the impact of the ND process on network capacity.

5.4 Capacity Analysis and Parameter Setting

In this section, we develop an analytical model to study the network capacity of a WPAN using ER based concurrent transmissions [11]. Our objective is to control the interference level and set the best ER r to maximize the spatial multiplexing capacity of a WPAN with random topology. In specific, given a two-dimensional or three-dimensional space with a number of users randomly distributed over the space, we analyze the expected per-flow throughput, total network throughput, and network transport throughput. The analysis can be used to select the best ER size to maximize the network capacity.

5.4.1 Average Number of Concurrent Transmissions

Given the number of active users in an area, what is the number of flows that can transmit simultaneously under the constraint of the ER condition? Since network topology and user deployment drastically affect the network performance, we focus on the expected number of concurrent transmissions, which is general and independent of network topology and user deployment.

Consider an $L \times L$ square room containing N_f active flows, with N_f transmitters and N_f receivers randomly deployed. Define $P(k, n)$ as the probability that only k flows satisfy the ER condition and can be scheduled for concurrent transmissions, after checking the first $n \leq N_f$ flows one by one. Without loss of generality, we check flows

in ascending order $1, 2, ..., N_f$. The first flow f_1 will be scheduled for transmission in the set γ, and we have $P(1,1) = 1$. Flow f_2 will be added to γ if it does not conflict with flow f_1. Define Q as the probability of a transmitter lying outside an ER of a receiver. The probability that a flow does not conflict with another flow is Q^2, because both transmitters should be outside the ERs of the other receivers. Accordingly, the probability that two flows do not satisfy the ER condition is $1 - Q^2$. Therefore, in the two-flow case, we have $P(2,2) = Q^2$ and $P(1,2) = 1 - Q^2$. After we check the first n flows, there are k flows in γ if a) there are $k-1$ flows in γ when we check the first $n-1$ flows, and the n-th flow does not conflict with the other $k-1$ flows in γ; or b) there are k flows in the set when we check the first $n-1$ flows, and the n-th flow conflicts with one of the k flows in γ. The probability that a flow does not conflict with any of the other $k-1$ flows is $Q^{2(k-1)}$.

$$P(k,n) = P(k-1,n-1)Q^{2(k-1)} + P(k,n-1)(1-Q^{2k}) \quad \text{for } k < n. \tag{5.32}$$

If, among the n flows, only the first flow can be added in γ, implying that the following $n-1$ flows do not satisfy the ER condition, we have

$$P(1,n) = (1-Q^2)^{n-1} \quad \text{for } k = 1. \tag{5.33}$$

Another extreme case is that all n flows can be scheduled concurrently, which means that none of the flows conflicts with the remaining $n-1$ flows,

$$P(n,n) = (Q^{n-1})^n \quad \text{for } k = 1. \tag{5.34}$$

Given the initial values of $P(1,1)$, $P(1,2)$ and $P(2,2)$, we can iteratively obtain $P(k,N_f)$ as a function of Q for $\forall k, 1 \le k \le N_f$. The expected number of concurrent transmissions is

$$E[CT] = \sum_{k=1}^{N_f} kP(k,N_f). \tag{5.35}$$

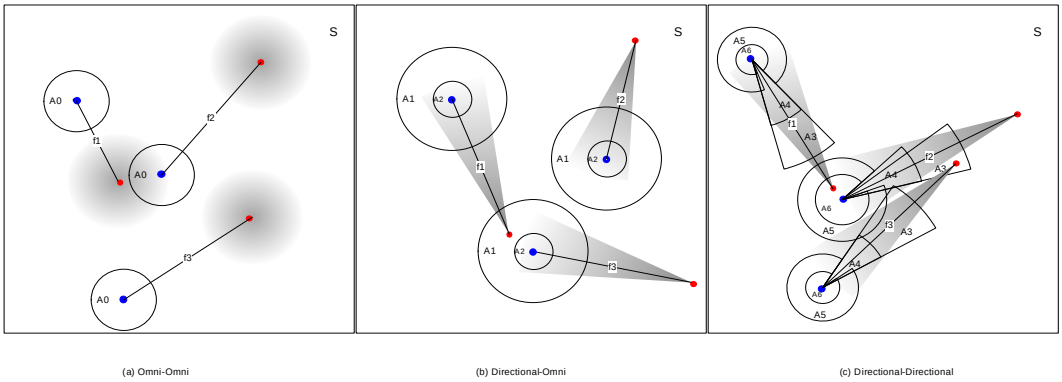


Figure 5.8: Exclusive Region for Concurrent Transmissions

Derivation of Q: Approach I

To obtain $E[CT]$, we need to know Q. Let the size of the ER of a receiver be A, and total area $S = L^2$. As shown in Fig. 5.8, with each user randomly deployed in the room, an interferer of one flow is outside the ER of the receiver of another flow with probability $Q = 1 - A/S$ Since the ER region and Q are related to the types of antennae used, in the following, we derive Q by considering the three cases shown in Fig. 5.7.

Case 1: Omni-antenna to Omni-antenna

In case 1, the ER is a circle with radius r_0 and $A_0 = \pi r_0^2$, as shown in Fig. 5.8(a). The probability that an interferer is outside the ER of a receiver is given by

$$Q_1 = 1 - \frac{A_0}{S} = 1 - \frac{\pi r_0^2}{S}, \text{ for } r_0 << L. \tag{5.36}$$

Case 2: Directional-antenna to Omni-antenna

Due to the omni- receivers and directional transmitters, the ER in case 2 contains two zones, a circle with radius r_1 and another circle with radius r_2, as shown in Fig. 5.8(b). Accordingly, the areas of the two zones are $A_1 = \pi r_1^2$ and $A_2 = \pi r_2^2$. If a receiver is within the radiation angle of an interferer with probability $\theta/2\pi$, the interferer is outside the first ER zone (A_1) with probability $1 - A_1/S$. Similarly, if a receiver is outside the radiation angle of an interferer with probability $1 - \theta/2\pi$, the

interferer is outside the second ER zone (A_2) with probability $1 - A_2/S$. Therefore, the probability that an interferer is outside the ER of a receiver is given by

$$Q_2 = 1 - \frac{\pi r_2^2}{S} + \frac{r_2^2\theta}{2S} - \frac{r_1^2\theta}{2S}, \text{ for } r_1, r_2 << L. \tag{5.37}$$

Case 3: Directional-antenna to Directional-antenna

The ER in this case contains four zones, as shown in Fig. 5.8(c). The first zone A_6 is a sector with radius r_6 and angle $2\pi - \theta$ and the second zone A_4 is a sector with radius r_4 and angle θ. The areas of these two exclusive zones are $A_6 = \pi r_6^2(1 - \theta/2\pi)$ and $A_4 = \pi r_4^2(\theta/2\pi)$. Note that the two areas $A4$ and $A6$ are exclusive to each other, i.e., $A4 \cap A6 = Null$. If a receiver is outside the radiation angle of another sender, the sender can concurrent transmit if it is outside the ER zones of the receiver (A_4 and A_6). A receiver is outside the radiation angle of another sender with probability $1 - \theta/2\pi$, and the sender is outside the ER zones of the receiver with probability $1 - (A_4 + A_6)/S$. Similarly, if a receiver is within the radiation angle of an interferer with probability $\theta/2\pi$, the interferer is outside the ER zones of the receiver (A_3 and A_5) with probability $1 - (A_3 + A_5)/S$, where $A_5 = \pi r_5^2(1 - \theta/2\pi)$ and $A_3 = \pi r_3^2(\theta/2\pi)$. Therefore, we have

$$Q_3 = (1 - \frac{A_5 + A_3}{S})\frac{\theta}{2\pi} + (1 - \frac{A_4 + A_6}{S})(1 - \frac{\theta}{2\pi}), \tag{5.38}$$

for $r_3, r_4, r_5, r_6 << L$.

Edge Effect — If the areas of ER zones are relatively large compared to the room area or a user is located near the edge of the room, it is likely that some parts of the ER zones will be outside the room, and this is referred to as "edge effect". Using case 1 as an example, if the receiver is at the corner of the square room, we have $A' = \pi r_0^2/4 < A$. The actual probability that a random interferer is within the ER of a receiver should be $A'/S \leq A/S$. Thus, the developed analytical model without considering the edge effect may result in a conservative estimation of $E[CT]$.

To mitigate the edge effect, we set a constraint for the area of ERs such that $A_i \leq S$ and $r_i \leq \sqrt{2}L$, $\forall i, i = 0, ..., 6$, in all cases.

Derivation of Q: Approach II

Another approach to derive Q is based on the distance distribution among any two users. Given the distance between two users at random from the interior of a unit square [96], the pdf of the distance between two randomly deployed users in an $l \times l$ square room, $f(x)$, is derived as

$$f(x) = \begin{cases} 2\frac{x}{l^2}(\frac{x^2}{l^2} - 4\frac{x}{l} + \pi) & \text{for} \quad 0 \leq x \leq l \\ 2\frac{x}{l^2}(4\sqrt{\frac{x^2}{l^2} - 1} - (\frac{x^2}{l^2} + 2 - \pi) - 4\tan^{-1}\sqrt{\frac{x^2}{l^2} - 1}) & \text{for } l < x \leq \sqrt{2}l). \end{cases} \tag{5.39}$$

Case 1: Omni-antenna to Omni-antenna In case 1, an interferer of flow f_j is outside of the ER of the receiver of flow f_i if the distance $d_{j,i}$ is larger than r,

$$Q_1 = P(d_{j,i} \geq r) = \int_r^{\sqrt{2}l} f(x)dx. \tag{5.40}$$

Case 2: Directional-antenna to Omni-antenna In case 2, an interferer is inside an ER of the receiver of flow f_i if the receiver is within the radiation angle θ of the interferer and the distance $d_{j,i}$ is less than the exclusive distance r_1; or if the receiver is not within the radiation angle and the distance is less than r_2. Thus,

$$Q_2 = 1 - \frac{\theta}{2\pi}\int_0^{r_1} f(x)dx - (1 - \frac{\theta}{2\pi})\int_0^{r_2} f(x)dx. \tag{5.41}$$

Case 3: Directional-antenna to Directional-antenna

Similarly, in case 3, an interferer is inside an ER of a receiver if 1) both the interferer and the receiver are within the radiation angle θ of each other and the distance $d_{j,i} < r_3$; or 2) if an interferer is within the beamwidth of the receiver but the receiver is outside of the beamwidth of interferer, and the distance $d_{j,i} < r_4$; or 3) an interferer is outside the radiation angle θ of the receiver but the receiver is within the angle of interferer,

and their distance $d_{j,i} < r6.$ or 4) both the interferer and the receiver are outside each other's radiation angle and their distance $d_{j,i} < r_6$.

$$Q_3 = 1-(\frac{\theta}{2\pi})^2 \int_0^{r_3} f(x)dx - \frac{\theta}{2\pi}(1-\frac{\theta}{2\pi})(\int_0^{r_4} f(x)dx + \int_0^{r_5} f(x)dx) - (1-\frac{\theta}{2\pi})^2 \int_0^{r_6} f(x)dx. \tag{5.42}$$

5.4.2 Network Capacity and Transport Capacity

Given the transmission distance z' and the channel gain s between the transmitter and the receiver, the received signal power, P_R, is given by (5.19)

$$P_R = k_1 G_T G_R z'^{-\alpha} s P_T. \tag{5.43}$$

Accordingly, the achievable data rate of the flow is calculated as

$$R = \eta W \log_2(1 + \frac{k_1 G_T G_R z'^{-\alpha} s P_T}{N_0 W + \sum I_i}), \tag{5.44}$$

where η is a constant coefficient related to the efficiency of the transceiver design. As UWB fast fading characteristics follow the Nakagami distribution [97][3], and the pdf of the Nakagami channel gain is given by

$$f_S(s) = (\frac{m}{\omega})^m \frac{s^{m-1}}{\Gamma(m)} e^{\frac{-ms}{\omega}}, \tag{5.45}$$

where m is the Nakagami parameter denoting the channel fading conditions and ω is the average received power. Notice that (5.44) does not apply when $z' \to 0$, and the achievable data rate is actually bounded in realistic communication systems. We assume that the maximum data rate is achieved when the transmission distance is not larger than the reference distance $d_{\min}$, $z' \leq d_{\min}$.

[3]Although we adopt a UWB channel model in this work, other channel models can also be applied in the analytical framework in general.

Let the random variable Z' denote the effective transmission distance, which can be represented as a truncated random variable over $[d_{\min}, d_{\max}]$ with probability density function of $f_{Z'}(z')$. The derivation of $f_{Z'}(z')$ in 2-D and 3-D space is given in Appendix 7.2.4.

Denote the interference distance and the gain of the channel between the ith interferer to a receiver as random variables V_i over $[r, d_{max}]$ and S_i, respectively. The pdf of V_i is given by

$$f_{V_i}(v) = \frac{f_Z(v)}{\int_r^{d_{max}} f_Z(z)dz}. \tag{5.46}$$

The channel between the interferer and the receiver also follows Nakagami fading model and the pdf of R.V. S_i is the same as in (5.45).

Given the interference distance v and the channel fading factor s_i, the interference from a single interferer is

$$I_1 = k_1 C_0 G_T G_R v^{-\alpha} s_i P_T. \tag{5.47}$$

Given that there are k concurrently transmitting flows, each flow has $k-1$ interferers. Since the received signal strength and the interference strength are independent R.V.s, the received $SINR$ of a single flow can be obtained as

$$\begin{aligned} SINR|k &= k_1 G_T G_R z'^{-\alpha} s P_T / [N_0 W + \sum_{i=1}^{k-1} I_i] \\ &\approx k_1 G_T G_R P_T s z'^{-\alpha} / [N_0 W + (k-1) I_1]. \end{aligned} \tag{5.48}$$

Accordingly, the average transmission rate and the bit-meter product of a single flow under $k-1$ interferers are given by

$$\begin{aligned} & E[T_S|k] \\ = \;& \eta W E[\log_2(1 + SINR|k)] \\ \approx \;& \iiiint \eta W \log_2(1 + \frac{k_1 G_T G_R z'^{-\alpha} s P}{N_0 W + (k-1) K C_0 G_T G_R P_T v_i^{-\alpha} s_i}) \\ & \cdot f_{V_i}(v_i) f_{S_i}(s_i) f_S(s) f_{Z'}(z') \, \mathrm{d}v_i \, \mathrm{d}s_i \, \mathrm{d}s \, \mathrm{d}z', \end{aligned} \tag{5.49}$$

and

$$E[Tr_S|k] \approx \iiiint z'\eta W \log_2(1+\frac{k_1 G_T G_R z'^{-\alpha} s P_T}{N_0 W + (k-1)KC_0 G_T G_R P_T v_i^{-\alpha} s_i}) \cdot f_{V_i}(v_i) f_{S_i}(s_i) f_S(s) f_{Z'}(z')\ \mathrm{d}v_i\ \mathrm{d}s_i\ \mathrm{d}s\ \mathrm{d}z'. \quad (5.50)$$

We then remove the condition of $k-1$ interferers to obtain the average single flow throughput $E[T_S]$ and bit-meter product $E[Tr_S]$ when it is scheduled for transmission as

$$E[T_S] = \sum_{k=1}^{N_f} E[T_S|k]p(k, N_f), \quad (5.51)$$

and

$$E[Tr_S] = \sum_{k=1}^{N_f} E[Tr_S|k]p(k, N_f), \quad (5.52)$$

where $p(k, N_f)$ is obtained in Section 5.4.1.

The capacity and transport capacity of the network are the sum of the throughput and bit-meter products of all flows concurrently transmitting, which are obtained as

$$E[T] = \sum_{k=1}^{N_f} kE[T_S|k]p(k, N_f), \quad (5.53)$$

and

$$E[Tr] = \sum_{k=1}^{N_f} kE[Tr_S|k]p(k, N_f). \quad (5.54)$$

Given the system parameters such as α, G_0, P, *etc.*, $E[T]$ and $E[Tr]$ can be obtained as a function of r. Consequently, we can obtain the best value of r to maximize network capacity and transport capacity.

5.5 Numerical Results and Discussions

We use Maple 10 to calculate the analytical results and validate them through extensive simulations with C. We set up the experimental network in a 10 m $\times$ 10 m 2 dimensional

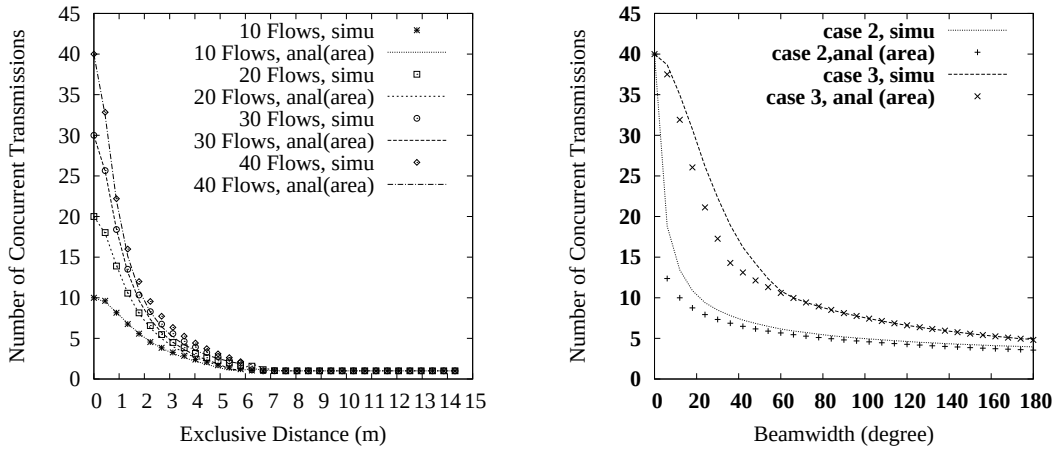
and a 10 m × 10 m × 2 m three dimensional square room, respectively. A number of flows with distinctive senders and receivers are randomly deployed in the space. All flows keep active for a long period of time that the transmitters always have data for transmission. We consider the use of omni-directional antenna in a UWB network, and directional antenna in a mmWave network. k_1 is set to 1 for calculation simplicity. The parameters used in the analysis and simulations are listed in Table 5.5. All flows use the maximum transmission power. The background noise level is $N_0 = 114$ dBm/MHz over a 500 MHz signal bandwidth. The path loss exponent is 4 and the cross correlation among concurrent flows $G_0 = 1$ if not otherwise specified. The reference distance is set as $d_{min} = 1$ m and the path loss at d_{min} is 43.9 dB. We set the transceiver efficiency $\eta = 0.21$ so the expected achievable data rate at 1 m is $R = \eta W \log_2(1 + SINR) = 1$ Gbps. In the simulations, the scheduler assigns time slots to flows that can be transmitted concurrently using the random selection algorithm introduced in Section 5.3.2. We also use the UWB channel proposed in [97] to study the impact of channel fading on the network performance. Independent Nakagami fading channels with $\omega = 1$ and $m = 1 \sim 6$ are applied between any two users. Since the number of concurrent transmissions and network capacity largely depend on the user deployment, we repeat the simulation 500 times with different seeds and calculate the average values.

5.5.1 Concurrent Transmissions vs Serial Transmissions

We first use the simplified cone plus circle model in a two-dimensional plane to compare the performance of the concurrent transmissions using the ER distance derived in (5.25)-(5.31) with that of serial TDMA transmissions. Based on the sufficient condition, the ER distance varies with the antenna beamwidth θ. A small θ results in a higher antenna gain, and thus a larger r. As shown in Figs. 5.9 (a) and (b), since the ER area increases with the ER distance r or the beamwidth θ, fewer flows can be scheduled for concurrent transmissions. In cases 2 and 3, due to the use of directional

Table 5.5: Simulation Parameters

Signal Bandwidth (W)	500 MHz
Transmission Power (P_T)	-41.3 dBm/MHz
Noise Power Density (N_0)	114 dBm/MHz
Path Loss Exponent (α)	$2.5 - 6$
Reference Distance (d_{min})	1 m
Path loss at d_{min} (PL_0)	43.9 dB
Cross Correlation (G_0)	0.1-1
Nakagami Factor (m)	$1 \sim 6$
Transceiver Efficiency (η)	0.21



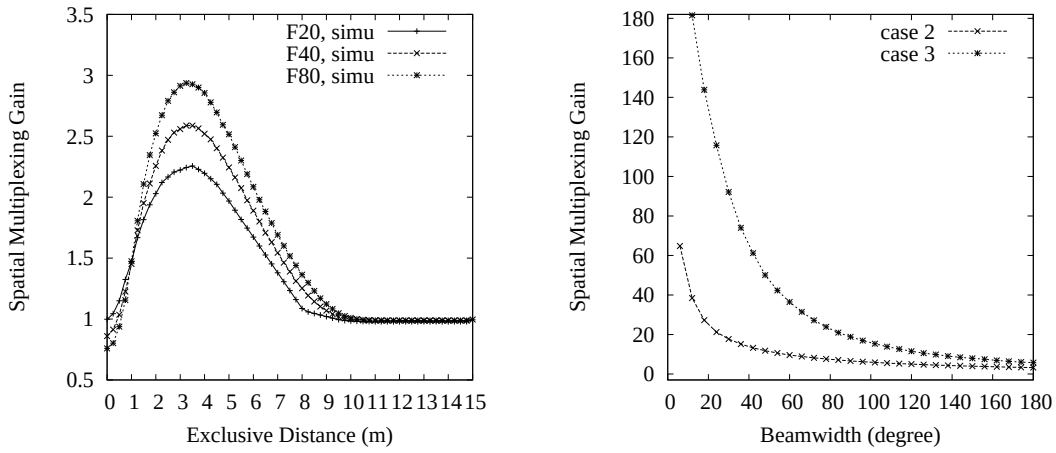
(a) case 1 (b) cases 2 - 3

Figure 5.9: Number of Concurrent Transmissions ($\alpha = 4$) ($\varsigma = 1$)

antenna, a receiver is less likely to be in the radiation angle of an interferer, and vice versa. Thus, more flows can be scheduled for concurrent transmissions. We use the area based approach I to calculate $E[CT]$. Our analysis is slightly conservative with a small beamwidth due to the edge effect. Overall, simulation results validate the accuracy of our analysis.

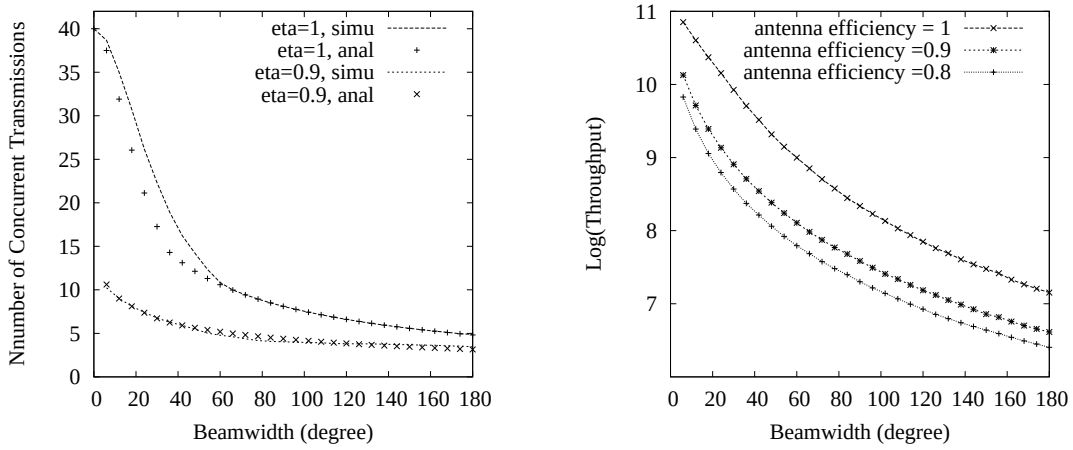
The spatial multiplexing gain, defined as the ratio of network throughput with concurrent transmissions to that with serial TDMA transmissions, is shown in Fig. 5.10. For case 1 (omni-to-omni), the sufficient condition to favor concurrent transmissions is to set the ERs as circles with radius $r_0 = 5.24$ m centered at each receiver, as derived in (5.25). When the radius is set close to 0, almost all flows can transmit concurrently. However, concurrent transmissions may achieve lower throughput than TDMA due to severe mutual interference. When the ER size is too small, the interference among concurrent flows is so high that the concurrent transmissions will not always outperform TDMA (spatial multiplexing gain is less than 1), especially in a dense network. When the ER size increases, fewer flows can be transmitted concurrently and thus more time slots are required to schedule all flows. When the ER size is large enough to forbid any concurrent transmission, the performance of the proposed scheme is the same as that of TDMA and the spatial multiplexing gain equals 1. Although network throughput may vary with different user deployments, we find that the spatial multiplexing gain achieved with r_0 is much higher than 1, as shown in Fig. 5.10 (a). Therefore, by scheduling concurrent transmissions with appropriate ERs, significant improvement on network throughput can be achieved. Fig. 5.10(b) shows the spatial multiplexing gains obtained in cases 2-3. Using a directional antenna with a small beamwidth, most flows can be transmitted concurrently and the spatial multiplexing gain is very high in both case 2 and case 3. In all cases, concurrent transmissions always outperforms serial TDMA transmissions under the derived sufficient condition of ERs.

In reality, the antenna gain is a product of the antenna directivity gain and the



(a) case 1 (b) cases 2 - 3

Figure 5.10: Spatial Multiplexing Gain ($G_0 = 1$)



(a) Number of Concurrent Transmissions (b) Network Spatial Capacity

Figure 5.11: Impacts of Sidelobe Effects (Case 3)

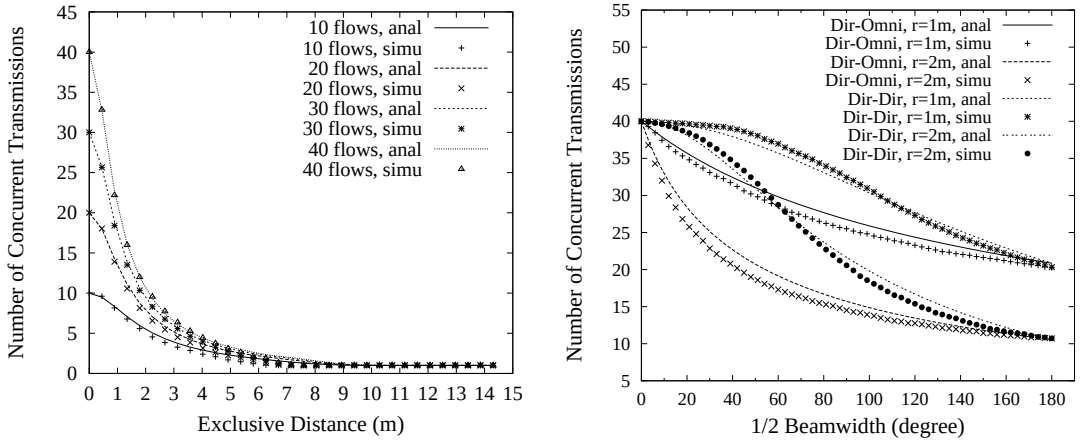
radiation efficiency $\varsigma < 1$, and the sidelobe is not negligible. We then investigate the expected number of concurrent transmissions in case 3 under various antenna efficiencies, and the numerical results are shown in Fig. 5.11(a). The antenna gain within the beamwidth deceases by a factor of ς, which results in a smaller ER sector with radius r_3. However, other ER zones (sectors with radius r_4, r_5, and r_6) are non-zero when the sidelobes are considered. Take $\theta = 40$ degree for example. The antenna gain is 9 when $\eta = 1$, and decreases to 8.1 when $\eta = 0.9$. Without considering the sidelobe, the ER for $\eta = 1$ is a sector of a circle with radius $r_3 = 13.4$ m. When $\eta = 0.9$, the ER contains four zones, as shown in Fig. 5.7(c) with $r_6 = 1.5$ m, $r_4 = r_5 = 4.5$ m and $r_3 = 12.7$ m. Here, $r_4 = r_5$ because the antennae gains used for the transmission and receiving are the same. Obviously, the areas of the four ER zones for $\eta = 0.9$ are larger than those for $\eta = 1$. According to (5.36)-(5.38), the probability, Q, that a transmitter is outside the ER of a receiver decreases when the ER areas increase, and thus fewer concurrent transmissions can be scheduled. As shown in Fig. 5.11(a), when ς is reduced from 1 to 0.9, the expected number of concurrent transmissions decreases significantly due to the increased ER areas. The base 10 logarithmic throughput in case 3 is shown in Fig. 5.11(b). It can be seen that the spatial multiplexing throughput also decreases with ς.

5.5.2 Network Spatial Multiplexing Capacity

In this subsection, we apply an ideal flat-top antenna model to study parameter setting for concurrent transmissions in UWB/mmWave WPANs.

Number of Concurrent Transmissions

We use the distance based approach II in (5.40-5.42) to calculate the expected numbers of concurrent transmissions under various ER distance r in Fig. 5.12. Note that in Fig. 5.9, r is a function of θ, while in Fig. 5.12, r is a tunable parameter that is inde-



(a) Case 1 (b) Cases 2-3

Figure 5.12: Expected Number of Concurrent Transmissions vs ER Distance ($\alpha = 4$)

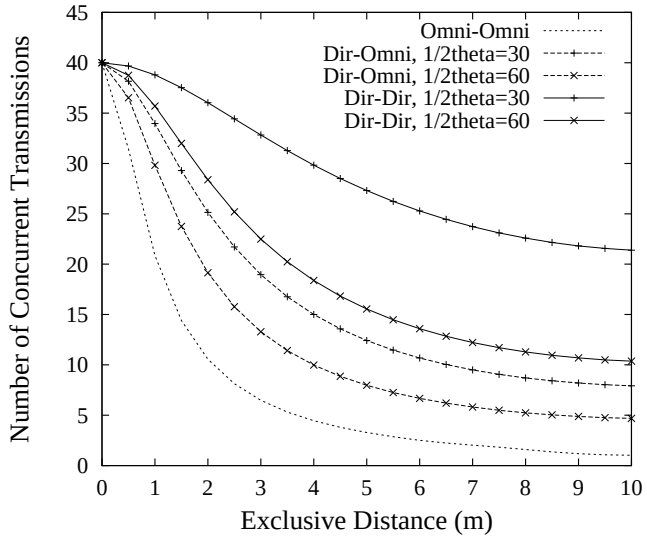


Figure 5.13: Comparison of Expected Number of Concurrent Transmissions in Cases 1-3

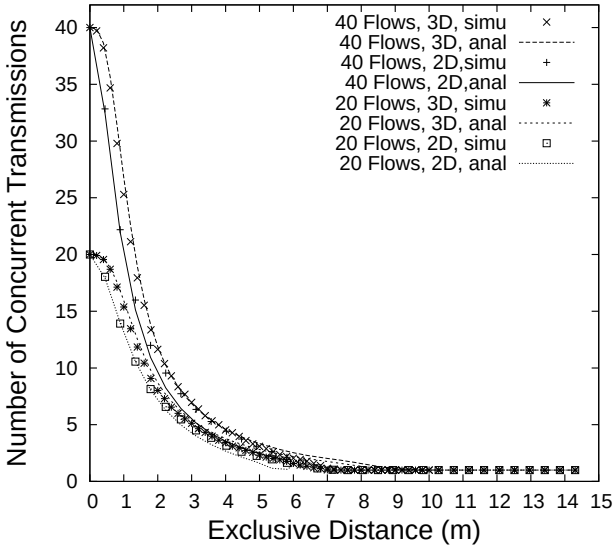


Figure 5.14: Expected Number of Concurrent Transmissions in 2D and 3D Space

pendent of θ. In all cases, the expected number of concurrent transmissions decreases when the exclusive distance r increases. Both area and distance based methods are validated by the simulation results. The number of $E[CT]$ under various r in all cases are compared in Fig. 5.13. It is shown that $E[CT]$ decreases much more drastically in case 1 than that in cases 2 and 3.

The expected number of concurrent transmissions in a 3-D space ($10\,m \times 10\,m \times 2\,m$) is shown in Fig. 5.14. The obtained $E[CT]$ in a 3-D space is slightly larger than that in a 2-D space by considering the third dimension of the space. The relationship of $E[CT]$ and the network density is shown in Fig. 5.15. It is observed that $E[CT]$ increases rapidly with the increase of the total number of flows when r is small, and the increase ratio decreases when r becomes large. For example, when r is large enough to prevent most concurrent transmissions, $E[CT]$ does not vary much with the network density.

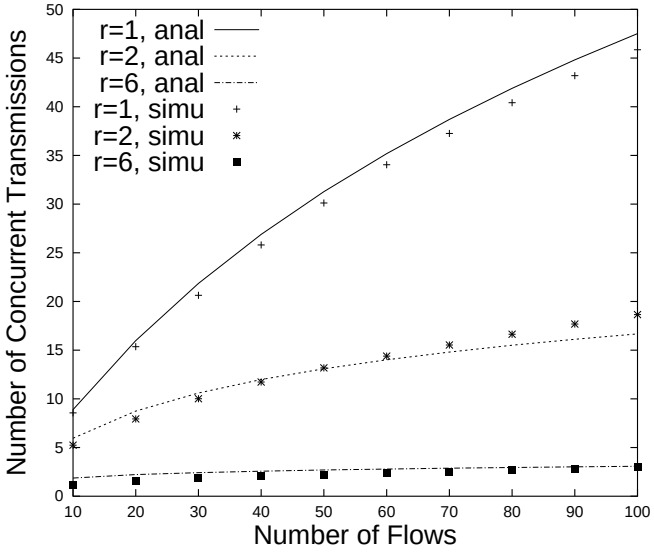
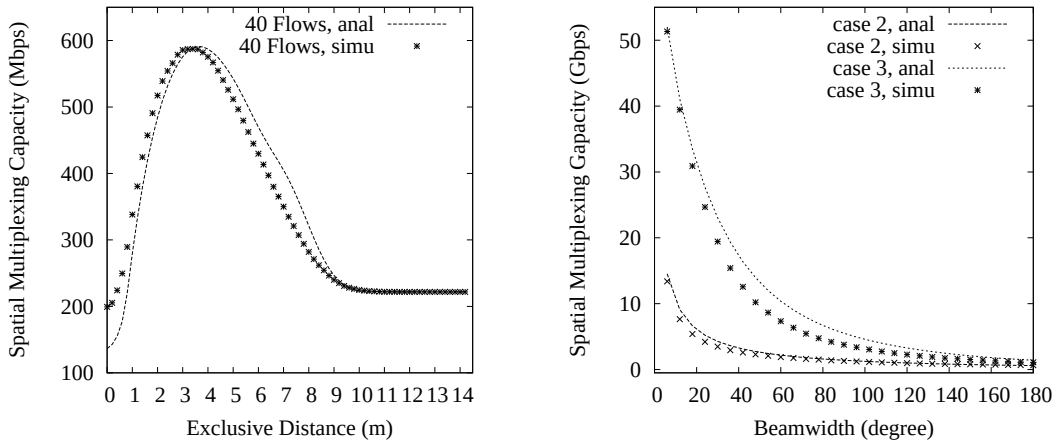


Figure 5.15: Expected Number of Concurrent Transmissions vs Number of Flows (3D)

Network capacity and transport capacity

We first evaluate the network performance without the channel fast fading, i.e., considering constant channel gain $S = 1$. The network spatial multiplexing capacity in cases 1-3 is shown in Fig. 5.16. For a small exclusive distance r, the spatial capacity is quite low, although a large number of flows can be scheduled for concurrent transmissions. This is because, the more the concurrent transmissions, the larger the interference from other flows. We also observe that in Fig. 5.16(a), the maximum capacity is achieved when r is 3.5 m. We further enlarge r, the network capacity decreases and becomes constant when $r > 10$ m. In Fig. 5.16(b), the network capacity in cases 2 and 3 is much higher than that in cases 1 due to greater antenna gain and more aggressive spatial reuse, especially when θ is small. As shown in Fig. 5.17, the network capacity peaks under certain exclusive distances in all cases.

We also study the relationship of the exclusive region radius r and the data rate of a single flow during transmission. Basically, the average data rate of a flow is



(a) Case 1 (b) Cases 2-3 ($r = 1$ m)

Figure 5.16: Network Spatial Multiplexing Capacity

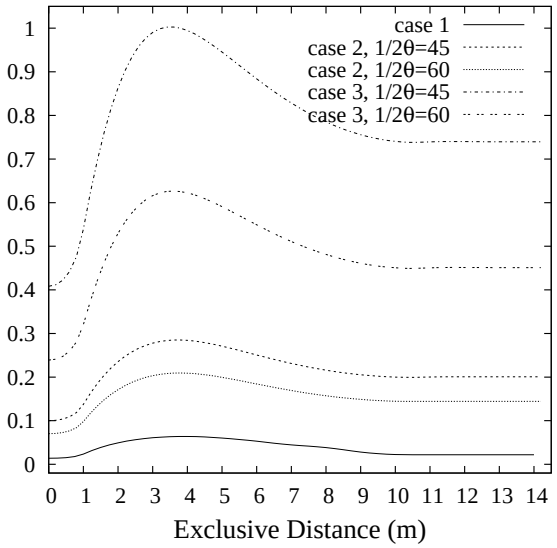


Figure 5.17: Normalized Network Spatial Multiplexing Capacity vs Exclusive Distance

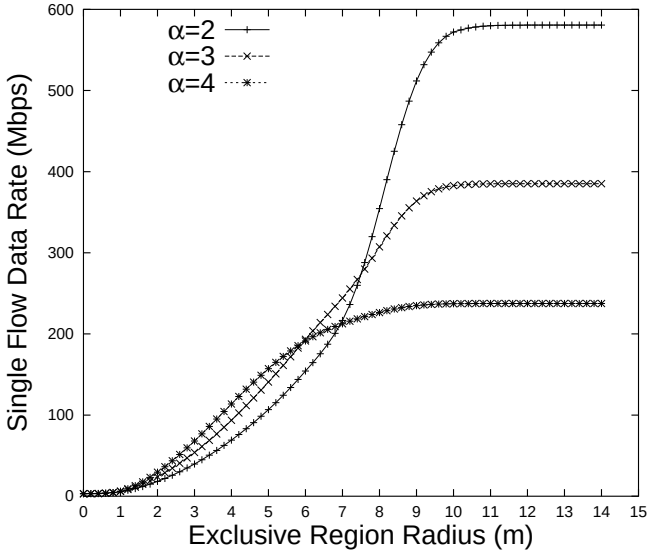


Figure 5.18: Single Flow Data Rate During Transmission (3D)

proportional to its average received $SINR$, which increases when the exclusive region radius r becomes larger and the number of concurrent transmissions decreases. When r is small, more flows can transmit concurrently, which results in higher interference level that degrades the data rate of each flow significantly. When r is sufficiently large to forbid any concurrent transmission, the maximum data rate of a single flow can be achieved with no interference, as shown in Fig. 5.18. Notice that both signal and interference power levels are dependent on the path loss exponent α. When the exclusive region radius r is small, interference from concurrent transmission could be serious. A larger value of α may result in a drastic decrease of interference level and a higher $SINR$ can be achieved. When r is large and there is no serious interference, the signal power decreases as α increases, so a lower data rate is achieved when α is larger. As shown in Fig. 5.18, when $r = 3$ m, flow data rate for $\alpha = 4$ is slightly larger than that for $\alpha = 3$ and $\alpha = 2$, in which cases, interference is the dominant factor of the single flow data rate. When $r \geq 9$ m, the data rate for $\alpha = 4$ is much less than that for $\alpha = 3$ and $\alpha = 2$, because in these cases the signal strength becomes the dominant

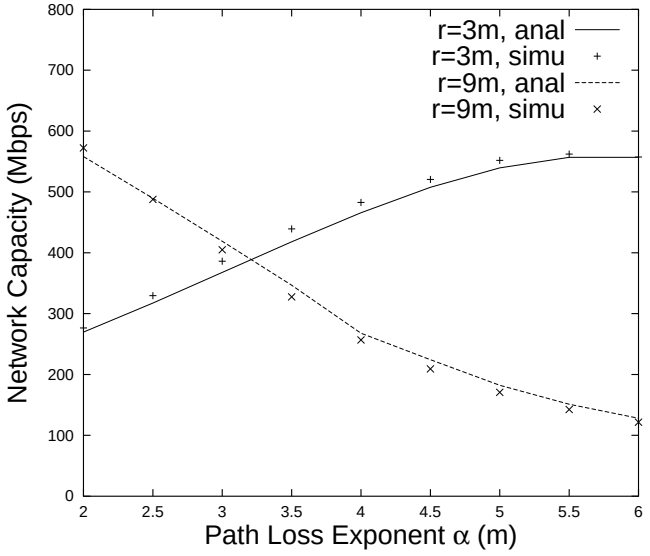


Figure 5.19: Network Capacity vs Path Loss Exponent(3D)

factor.

The network spatial multiplexing capacity with omni antenna in a 3D space under various path loss exponents is shown in Fig. 5.19. It is observed that the network capacity increases with α when $r = 3$ m, but decreases when $r = 9$ m. Although the single flow data rate of $r = 3$ m is always much smaller than that of $r = 9$ m, the total throughput when $r = 3$ m may be larger when the path loss is severe enough to significantly reduce the interference level. For $r = 9$ m, there is little spatial multiplexing gain since only one flow transmits at a time; the network throughput equals to that of a single flow throughput which decreases with α due to signal dispersion over distance.

We further investigate the impacts of spread spectrum techniques on spatial multiplexing gain. As shown in Fig. 5.20, the network spatial multiplexing capacity in a 3D space is a concave curve with different exclusive distances as in a 2D space. The network capacity is bounded by high interference level when r is small; and the capacity becomes constant and equals the average throughput of a single flow when r is large enough to forbid any concurrent transmissions. When a certain spreading technique

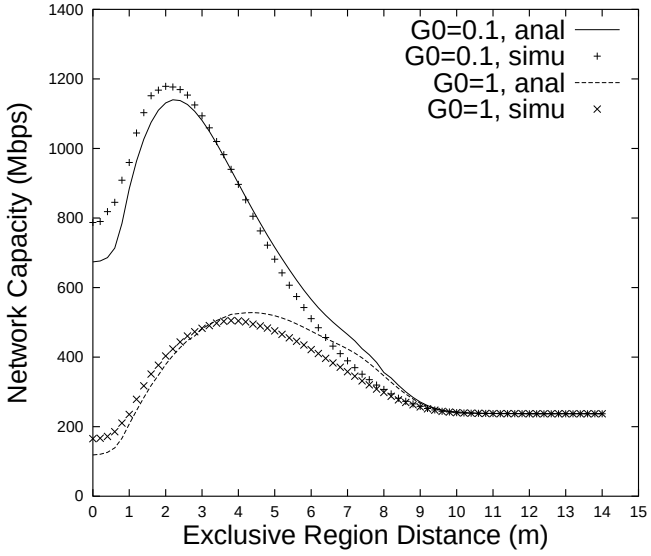


Figure 5.20: Network Capacity vs ER Radius(3D)

(FDMA or CDMA) is used to reduce the cross correlation among concurrent flows, a smaller G_0 can be achieved, which can significantly reduce the interference level among concurrent flows and thus improve the network capacity. For $G_0 = 0.1$, the maximum network capacity is higher than that of $G_0 = 1$. However, in wireless networks, it is difficult to maintain the orthogonality among different spreading codes. In addition, efficiently distributing spreading codes among multiple flows is not a trivial task. Simulation results validate the accuracy of our analysis.

The transport capacity of the network in a 3D space is shown in Fig. 5.21. There are 40 flows in the network. Similar to the network capacity, the transport capacity of the network is also a concave curve under different exclusive region radii. The relationship between the transport capacity and network density under different exclusive regions is investigated in Fig. 5.22. When the network is sparse, the transport capacity is relatively low because spatial reuse is not fully deployed. We increase the number of flows in the network, and the transport capacity improves with more flows transmitting concurrently. Given the exclusive region size r, the expected number of concurrent

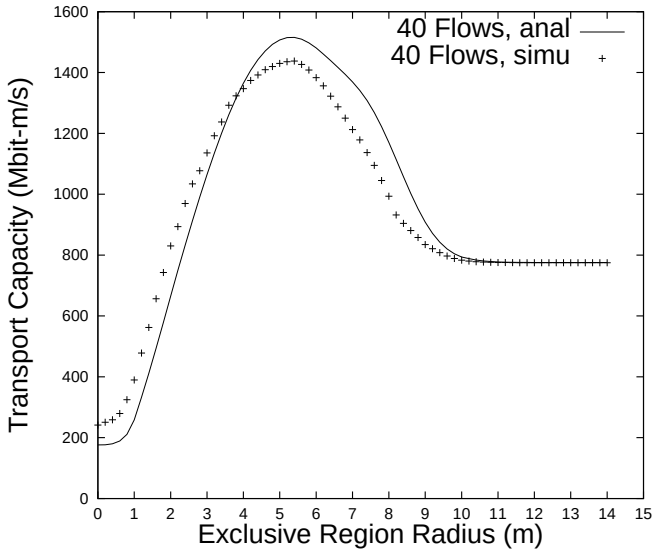


Figure 5.21: Transport Capacity vs ER Radius (3D)

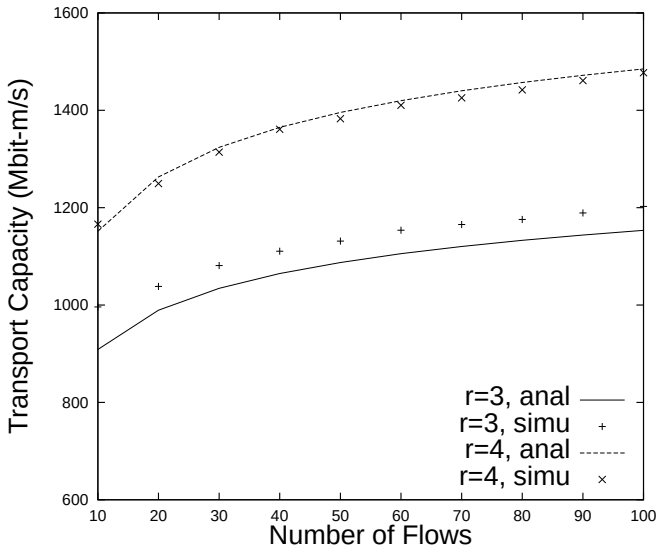


Figure 5.22: Transport Capacity vs Number of Flows (3D)

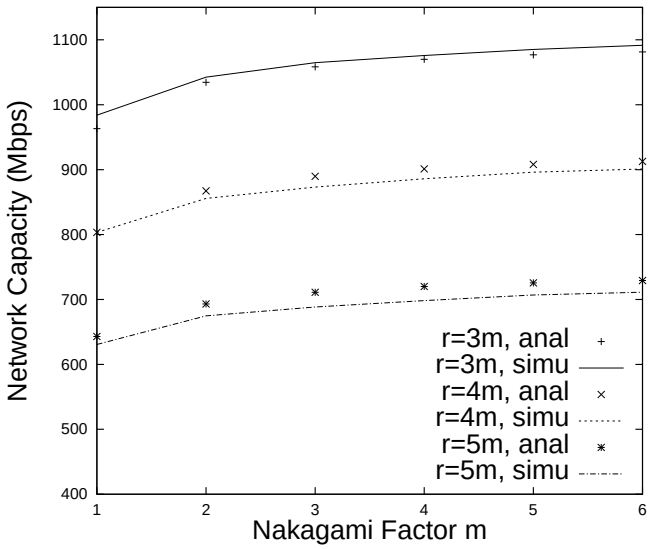


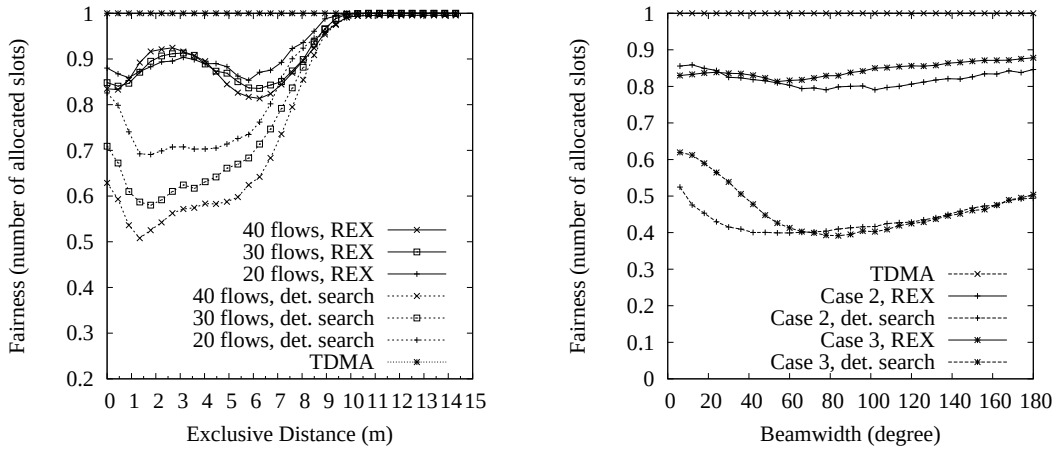
Figure 5.23: Impacts of Nakagami Channel Fading (3D, $G_0 = 0.1$)

transmissions will not increase rapidly with the total number of flows in a dense network, as shown in Fig. 5.15, and the increment of the transport capacity slows down accordingly.

We then study the impacts of the Nakagami channel fading on network capacity. The network capacity under various Nakagami fading parameter m ($m = 1 \sim 6$) is shown in Fig. 5.23. The parameter m generally reflects the severity of the channel fading conditions. The larger the m is, the more likely there is a line-of-sight path, and thus the better channel condition we have. It is observed that the network capacity increases as m increases.

Fairness

We further investigate the fairness performance of the proposed REX. Fairness is measured by Jain's fairness index [98], $(\sum x_i)^2/(N\sum x_i^2)$, in terms of the total number of time slots allocated to each flow. The fairness index in cases 1 to 3 are shown



(a) case 1 (b) cases 2 - 3

Figure 5.24: Fairness Comparison w.r.t Number of Allocated Slots

in Figs. 5.24. TDMA can always achieve perfect fairness since it allocates time slots evenly to all flows. With the deterministic searching scheme, all 40 flows are examined in ascending order of the flow number, *i.e.*, from flow 1 to flow 40, thus the flows with smaller sequence numbers are more likely to be scheduled. As shown in Fig. 5.24(a), with the deterministic searching sequence, fairness degrades drastically when the number of flows increases, while our proposed scheduling scheme using sorted searching sequence can achieve better fairness, regardless of the number of flows. In Fig. 5.24(b), we compare Jain's fairness index in cases 2 to 3. The scheduler with deterministic searching sequence provides poor fairness performance because those flows with smaller sequence number are more likely to be scheduled. The proposed REX scheme can achieve better fairness since it always gives higher priority to the flows with fewer number of assigned slots. On the other hand, for fair resource allocation, it is desirable to maximize the minimal flow throughput among all competing flows. We compare the per-flow throughput of different transmission schemes in Fig. 5.25. The simulation results show that, although Jain's fairness index of the proposed scheme is

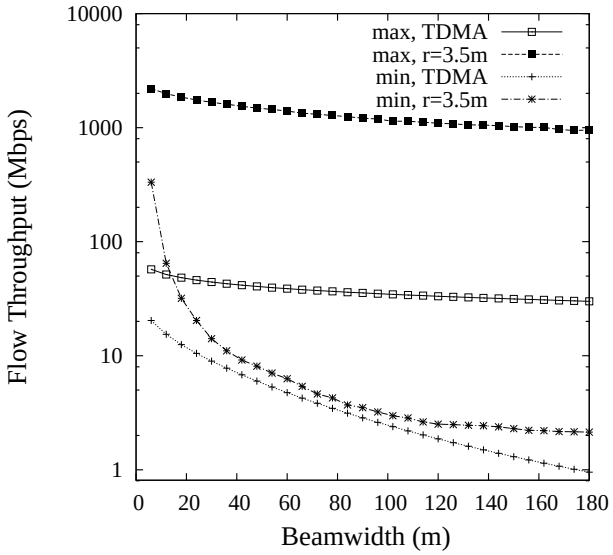


Figure 5.25: Singe Flow Throughput (Case 3)

lower than that of the TDMA scheme, the minimal per-flow throughput is improved with the proposed scheme. Thus, the proposed scheme achieves better Min-Max fairness and it is more desirable for user experience.

5.6 Summary

In this chapter, we analytically study the performance of IEEE 802.15.3 MAC and investigate the enhanced MAC layer resource allocation in UWB/mmWave WPANs. The main accomplishments of this chapter are summarized as follows:

- We have developed an analytical model to study IEEE the performance 802.15.3 MAC. Our analysis disclose the interaction between the tunable parameters, including the initial contention window, the retry limit, the length of contention period CAP and contention free period CTAP, *etc.*, and thus provides important guideline for parameter setting.

- We have proposed a novel ER based MAC for UWB/mmWave WPANs to efficiently explore the spatial capability of UWB/mmWave channel, considering the salient features of UWB and mmWave communications. The proposed MAC not only achieves high network capacity, but also provides good fairness among multiple users for channel access.

- We have further analyzed the spatial multiplexing capacity of a UWB/mmWave WPAN using the proposed ER based MAC. Given a 2D or 3D space, we have derived the expected number of concurrent transmissions, per-flow throughput, and network (transport) throughput of a WPAN. The analytical results reveal what are the main factors affecting the network (transport) capacity, and how to determine the best protocol parameters, e.g., exclusive region size, to maximize the network capacity.

Chapter 6

MAC Protocol in Distributed Multi-hop UWB Networks

6.1 Related Work

6.1.1 Distributed MAC

MAC protocols in general can be classified into two categories: centralized and distributed MAC. Extensive research has been conducted on developing efficient centralized MAC protocols for wireless networks, with dedicated centralized coordinators or randomly chosen coordinators [99]. A centralized MAC protocol usually provides more reliable and predictable services than distributed MAC at the expense of control overheads [79, 82, 100–102]. However, a centralized solution may not be desirable for large-scale, multi-hop wireless networks for the following reasons: a) centralized schemes normally have significant communication and computational overheads. With high data rate (up to Gbps) UWB communication technologies, the transmission time is usually on the order of μs, so any packet-level scheduler with complexity more than $O(1)$ becomes less desirable. On the other hand, the traffic of many applications, *e.g.*, data and video traffic, are bursty in nature, thus it is difficult to reserve an appropri-

ate amount of resources for these traffic flows; b) tight synchronization among users is costly, especially in a large scale multi-hop UWB networks; c) centralized architecture is not scalable, and it may suffer the single-point-of-failure problem; d) when a hierarchical structure is used to divide the entire network into multiple small piconets, efficient coordination among piconets is not an easy task and interference resulting from inter-piconets communication may severely degrade the network performance.

Recently, the WiMedia Alliance has launched PHY and MAC layer specifications based on MB-OFDM UWB technology [24]. The Wimedia MAC specification uses a combination of CSMA and TDMA mechanisms to provide a certain level of quality of service for isochronous traffic in a distributed manner. In [103], a contention based distributed algorithm, RCAMA, is proposed, using a physical interference model. However, both RCAMA and Wimedia MAC are based on a time slotted frame structure with a implicit assumption of negligible overhead in acquiring tight MAC layer synchronization among users. The maximum throughput region attained by a distributed scheduling strategy under arbitrary network topology and interference models was given in [104]. A distributed maximal matching scheduling strategy was presented in [105] to guarantee a certain fraction of the optimal throughput region. In [106], a distributed greedy scheduling scheme based on a more general interference model was proposed, and a lower bound on the capacity region was also investigated. However, to guarantee a certain throughput region, the schemes in [104–106] may require many rounds of computation and control message exchanges, and thus are not scalable because the overheads increase with the network scale. These works mainly focused on proposing new protocols or algorithms with more complicated control messages and computations to improve the resource utilization. In addition, the rate adaptive characteristics of the UWB communication technologies were not considered. A joint PHY/MAC architecture for impulse-based time-hopping UWB was proposed in [107], considering power control, rate adaptation, and mutual exclusive region. The approach mainly focused

on effective physical layer modulation schemes to cancel the interfering energy, and effective MAC protocol design still remains an open issue.

6.1.2 Asynchronous MAC

IEEE 802.11 DCF MAC is one of the most successful MAC protocols that has been widely deployed over the world due to its asynchronous data transmissions. Several recent works have been proposed to adaptively adjust the carrier sensing range of the transmitters to improve the spatial reuse performance of IEEE 802.11 MAC [108–112]. In [109], the relationship between transmission power and the carrier sensing threshold was studied, assuming a perfect MAC protocol that all communication channels are fully utilized. The impact of the carrier sensing threshold on the network capacity was investigated in [110]. The optimal carrier sensing threshold that maximizes spatial reuse for several regular topologies was obtained in [111]. It was found in [112] that the optimal carrier sensing range of IEEE 802.11 MAC should consider the tradeoff between the spatial reuse and the packet collision probability, and that an optimal carrier sensing range can be obtained based on a reward formulation. However, all the previous works use a simple collision model[1] in a WLAN environment, where the transmission and interference levels are much higher than those in UWB networks. Due to the stringent power emission regulation and the wide bandwidth of UWB communications, spreading technologies are usually employed to allow multiple concurrent transmissions [24, 92]. Thus, the simple collision model used in WLANs does not hold in UWB systems. In addition, adjusting the carrier sensing range of the transmitters can only reduce possible collisions among those transmitters within their carrier sensing ranges, but cannot guarantee successful receptions at the receivers. Therefore, instead of adjusting sensing ranges around the transmitters, we define exclusive regions around the receivers to assure that the ongoing transmission to the tagged receiver will not be interrupted

[1] A collision occurs if two or more stations within their transmission ranges transmit simultaneously.

by other interferers. To the best of our knowledge, little work has been done for asynchronous distributed MAC design and optimization for multi-hop UWB wireless networks, considering the characteristics of UWB communication technologies. Thus motivated, we utilize the salient features of UWB communications and propose the Distributed EXclusive (DEX) region based MAC protocol to efficiently exploit the spatial capacity of multi-hop UWB networks.

6.2 Distributed ER-based MAC Protocol

In this section, we extend our ER based MAC design for centralized WPAN to a distributed multi-hop wireless network [13, 113]. We consider a multi-hop UWB network as shown in Fig. 3.1(D). Spreading techniques are employed for multiple access and all users share a pool of spreading codes, which are numbered as $1, 2, ..., \kappa$. One common spreading code is chosen for control message exchange, *e.g.*, for RTS and CTS frames. Each user maintains a code table to record all the spreading codes used by the ongoing neighboring transmissions. The procedures to choose codes and initiate transmissions at the sender and receiver sides are given in Algorithm 1 and Algorithm 2, respectively.

Initially, the code table and the NAV maintained by each node are empty, the initial contention window size equals the minimum window size, i.e., $CW = CW_{\min}$, and the node attempting retransmission ($retry$) sets its counter to 0.

If node A receives data from the upper layer for transmission to node B, A will use a hash function to obtain a spreading code: $X = Hash(A + B)$ for the transmission, where A and B used in the hash function are related to their MAC addresses. A starts channel sensing when its NAV reaches zero. If the channel is sensed idle for a backoff interframe space ($BIFS$), A transmits an RTS frame to B, including the chosen code X and the transmission time $T_2 = RTS + SIFS + CTS + SIFS + DATA + ACK$. Otherwise, A enters a backoff procedure and sets a backoff counter (BC) uniformly distributed over [0, CW) for the first transmission attempt and A freezes its BC until

the channel is sensed idle for a $BIFS$.

If the channel is sensed busy but A has not successfully received an RTS or CTS, A needs to continue channel sensing till the channel is idle for $BIFS$. If A overhears an RTS or CTS frame from another transmission f_i, A checks the ER condition: 1) if either the transmitter or the receiver of f_i is in A's ER region, A should postpone its own transmission until the ongoing transmission f_i completes, and A updates its NAV according to T_2; 2) if A is outside the ER of f_i, A only needs to wait until RTS times out and updates its NAV according to $T_1 = RTS + SIFS + CTS$. An example of NAV setting is shown in Fig. 6.1, where A and B exchange RTS and CTS messages, C and D are neighbors within the ER of flow AB, respectively, and E is another neighbor outside the ER of flow AB. Since A is outside the ER of f_i and concurrent transmission is allowed, A adds the spreading code used by f_i in its code table and assures that its own code X does not conflict with any record in its code table. If code collision occurs, A can hash again till there is no code collision. Each record in the code table is associated with a time to live (TTL) parameter and will be removed from the table if TTL expires.

If A successfully receives a CTS from B after an interval SIFS, implying that B is available for the transmission using the spreading code X, A starts to transmit data to B at a rate of $R(i)$ after a SIFS. For implementation simplicity, the rate $R(i)$ is not determined based on the measurement of the instantaneous interference and noise level of the tagged transmission, but on the worst case scenario that assumes the maximum number of dominant interferers. Therefore, DEX is robust against interference from neighborhood asynchronous transmissions. The detailed derivation of $R(i)$ is presented in Section 6.3. If no CTS is received successfully, implying that B is not available at this moment to receive data using code X, A will enter the backoff stage and attempt retransmission thereafter, until the retransmission limit m is reached. The backoff procedure in DEX is the same as that in IEEE 802.11. Each time A retransmits RTS,

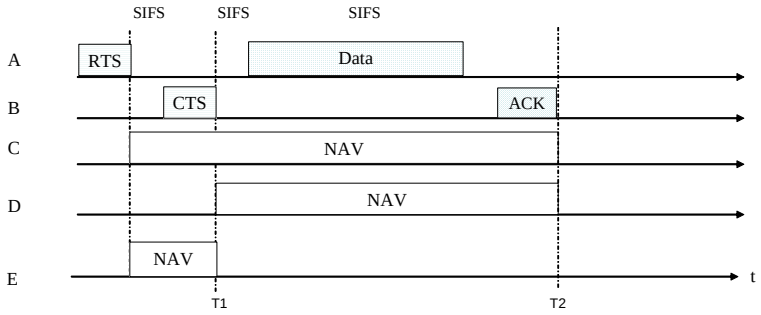


Figure 6.1: Network Allocation Vector Update

it will also choose a different code X by repeating hash functions because the code it chose previously may not be acceptable for use at B, (although the probability of code collision at B is very low).

To further improve the protocol efficiency, a TXOP is employed, *i.e.*, a time duration T is reserved in each RTS/CTS that a transmitter can transmit a burst of data frames during T. The longer the T, the better resource utilization will be, because less overhead is involved in each transmission. But a longer T leads to a larger access delay. Therefore, T should be chosen appropriately so that the access delay is tolerable for other flows in the ER region. On the other hand, a smaller ER region allows for more concurrent transmissions, which reduces the access delay of each flow. Thus, it is possible to choose a larger T for DEX and still well maintain the desired delay and fairness performance.

At the receiver side, B is ready for channel sensing or receiving only if its $NAV = 0$. Whenever B overhears an RTS or CTS frame from its neighboring node, B will update its NAV and code table in the same way as sender A does. Upon successfully receiving an RTS from A, B sends back a CTS if X does not conflict with any record in B's code table and the channel is idle for a SIFS period. Otherwise, B keeps silent and A may retransmit an RTS and choose another code after the RTS timeout.

Algorithm 1 Sender

```
1: if (A has data for B) then
2:    X = Hash(A + B);
3:    if X conflicts with any record in A's code table, repeat X = Hash(A+X);
4:    wait till NAV = 0
5:    channel sensing;
6:    if (Channel = idle for BIFS) then
7:       go to Line 29
8:    else
9:       exponential random backoff;
10:   end if
11:   while (BC > 0) do
12:      channel sensing;
13:      if (Channel = idle ) then
14:         decrease BC by 1 for each idle slot;
15:      end if
16:      if (Channel = busy that A overhears an RTS/CTS) then
17:         if (the overheard sender is in A's ER region) then
18:            update NAV according to T2;
19:         else
20:            update NAV according to T1;
21:         end if
22:         update A's code table;
23:         if X conflicts with any record in A's code table, repeat X=Hash(A+X) ;
24:         wait till NAV = 0;
25:      else
26:         freeze BC; wait till channel is idle;
27:      end if
28:   end while
29:   transmit RTS to B;
30:   if (receive CTS from B before timeout) then
31:      transmit DATA at rate R after SIFS;
32:   else
33:      increase retry by 1;
34:      if ( retry > m) then
35:         drop the current frame;
36:      else
37:         if X conflicts with any record in A's code table, repeat X=Hash(A+X) ;
38:         exponential backoff, and go to Line 4;
39:      end if
40:   end if
41: end if
```

Algorithm 2 Receiver

1: **if** ($NAV! = 0$) **then**
2: wait;
3: **else**
4: ready for receiving;
5: **end if**
6: **if** (*overhear RTS/CTS*) **then**
7: **if** (*the overheard transmitter/receiver is not in B's ER*) **then**
8: update B's code table;
9: update NAV according to T_1;
10: go to Line 1;
11: **else**
12: update NAV according to T_2;
13: go to Line 1;
14: **end if**
15: **end if**
16: **if** receive RTS from A **then**
17: **if** (*X does not conflict with any record in B's code table* and *Channel = idle for SIFS*) **then**
18: transmit CTS; receive data; send ACK; go to Line 4;
19: **else**
20: silent; go to Line 4;
21: **end if**
22: **end if**

6.3 Performance Analysis and Parameter Setting

In this subsection, we first prove that the DEX protocol is correct, *i.e.*, for each pair of users that successfully exchanges RTS/CTS messages, their data transmissions can be collision-free. Here, collision-free means that the interference from other concurrent transmissions is less than the tolerable threshold. We also study the performance bounds of a dense multi-hop UWB network and propose a method to choose ER size appropriately towards the maximum network throughput in a randomly deployed network.

To evaluate the network performance, we use the log distance path loss model for signal loss in an indoor radio propagation channel, which is given by

$$PL(d) = PL_0 + 10\alpha \log_{10}(d/d_{ref}), \tag{6.1}$$

where $PL(d)$ (dB) is the total path loss at distance d, PL_0 is the path loss at the reference distance $d_{ref} = 1$ m, α is the path loss exponent. Under this channel model, the received SINR of flow i is given by

$$SINR(i) \quad = \quad \frac{P(s_i)d_i^{-\alpha}}{N_0 W + \sum_{j \neq i} P(s_j) d_{j,i}^{-\alpha} G_0}, \tag{6.2}$$

where $P(s_i)$ is the transmission power of the sender of flow i, $d_{j,i}$ is the distance between the sender of flow j and the receiver of flow i.

6.3.1 Protocol Correctness

For a network with random topology and user deployment, we first consider the maximum amount of interference generated by concurrent transmissions from other users to the tagged receiver r_i.

The highest interference level is related to the "circle packing" problem. Each sender is located in the mid-point of an ER which is a circle. The maximum interference occurs

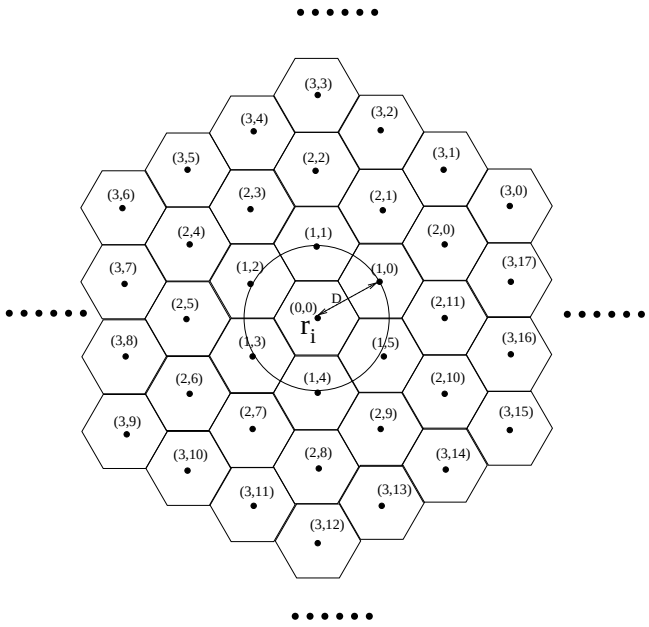


Figure 6.2: Worst-case Scenario with the Maximum Interference

when the non-overlapping circles are packed in the plane with highest density, which has been proved to be the hexagonal packing. Thus, the maximum interference to the receiver r_i occurs when all interferers are located at the center of each hexagonal cell around r_i, as shown in Fig. 6.2, i.e., there are 6 interferers located in the first tier of cells of r_i, and $6k$ interferers located in the k-th tier of cells. The distance from the k-th tier interferers to r_i is no less than $\sqrt{3}kD/2$. Thus, the total interference to r_i, I_r, is bounded by

$$\begin{aligned} I_{r,D} &< \sum_{k=1}^{\infty} 6PG_0(\sqrt{3}D/2)^{-\alpha}(k)^{1-\alpha} \\ &= 6PG_0(\sqrt{3}D/2)^{-\alpha}\zeta(\alpha-1). \end{aligned} \tag{6.3}$$

The above Riemann Zeta-function, $\zeta(\alpha-1)$, converges *iff* $\alpha > 2$. Therefore, if the path loss exponent α is a constant not larger than 2, an infinite coverage area cannot allow an infinite number of concurrent transmissions, and vice versa. Fortunately, empirical evidences from experimental field studies suggest that while path loss exponent near the

transmitter is likely to be 2, at large distance, it is larger than 2, and the received power level even decays exponentially with distance if the distance is quite large. Therefore, the interference to a tagged user can be well bounded. With a random network setting, it is practical to assume that the maximum interference to r_i, $I_{r,D}$, is $6PG_0D^{-\alpha}$, because (a) the number of first tier interferers is less than six almost surely, (b) the number of interferers is finite, and (c) the path loss exponent is large for high dense wireless networks, so the value of $\zeta(\alpha - 1)$ is close to one.

Now, we can prove that the sender and receiver which successfully exchange RTS/CTS can successfully transmit without being interrupted by other users at rate $R(i) = \eta W \log_2(P(r_i)/[N_0W + I_{r,D}] + 1)$. First, since the RTS/CTS of the pair has been successfully exchanged, all other users within their ER will not interrupt the tagged transmission. Second, the actual SINR should be larger than $I_{r,D}$. Thus, the transmission can be successful because the data rate chosen by the pair is more conservative than the actual achievable one.

6.3.2 Hidden Terminal and Exposed Terminal

We examine the hidden terminal and exposed terminal problems in multi-hop wireless networks. The hidden terminal problem exists for RTS transmissions. Since DEX allows concurrent transmissions and each pair of users can transmit data/ACK for a comparatively long time T consecutively, the number of RTS messages exchanged is reduced, so the collisions due to hidden terminal are reduced. In addition, if we can set the carrier sensing range to be the sum of the transmission range and the interference range of RTS, we can eliminate hidden terminals.

Using 802.11 DCF, there are proposals to mitigate the hidden terminal problem, which usually leads to more severe exposed terminal problem. The nice feature of the proposed DEX protocol is that the reserved space by RTS/CTS is determined by the ER region, instead of the carrier sensing range, so it does not suffer from the exposed

terminal problem as much as the IEEE 802.11 DCF protocol. Since the ER is much smaller than their carrier sensing regions, more flows can transmit concurrently, and the spatial multiplexing gain of DEX is higher than that of 802.11 DCF.

Another issue may result in performance degradation is that the RTS message in 802.11 DCF will notify all other users within the carrier sensing range of the sender to refrain from contention till the end of the transmission. If the receiver cannot send CTS because it is inside the carrier sensing range of some ongoing transmissions, the reservation of the RTS is totally wasted. In addition, if both the sender and the receiver of a flow need to compete with a large number of other users, the flow may starve due to the low possibilities that both the sender and the receiver can successfully access the channel. With DEX, since the ER is smaller than the carrier sensing region, the chance that the receiver cannot reply to the RTS is much lower, and the starvation problem can be alleviated.

6.3.3 Network Performance Bounds

We investigate the performance of the DEX protocol and derive the performance bounds. A node's ER is a circle centered at the node with radius D. D is a key parameter, which affects the number of concurrent transmissions in an area and the interference level to a tagged user. Considering a dense network, we are interested in obtaining the best value of D which can maximize the expected network transport throughput. To derive the throughput, we need to know the average number of concurrent transmissions (CTs), which is very difficult to obtain because it is sensitive to the network topology and the sequence of users initiating transmissions. Thus, we first obtain the theoretical upper bound and lower bound of the number of CTs in a dense network.

Lemma 1 *In an area of $L \times L$, for a given ER with radius D, the upper bound of the number of CTs is $2L^2/(D^2\sqrt{3})$.*

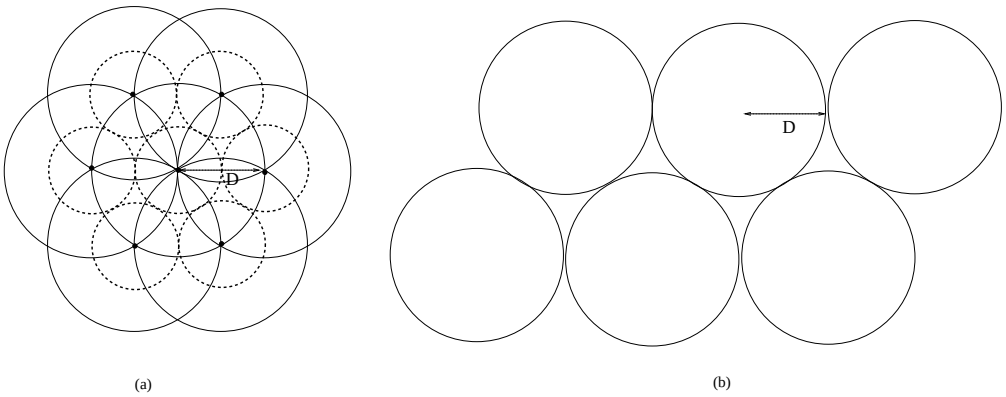


Figure 6.3: (a) Circle Packing Problem, (b) Circle Covering Problem

Proof 1 *As shown in Fig. 6.3, in the extreme case that, for each flow, the ER region (the solid circles) of the sender and that of the receiver fully overlap, the maximum number of CTs is equivalent to the maximum number of circles with radius $D/2$ (the dashed circles) that can be packed in the area. This is the classical circle packing problem. Toth proved that the hexagonal lattice is indeed the densest of all possible plane packings [114]. Accordingly, the maximum number of CTs is* $2L^2/(D^2\sqrt{3})$.

Remark: In a sparse network, we can improve the network throughput by increasing the user density to enlarge the number of CTs. However, according to Lemma 1, once the user density is large enough to saturate the network, further increasing the user density cannot improve the network throughput, but only increase the competition levels of all users within the associated ER and results in severe collisions.

The lower bound of the number of CTs in a sparse network can be as low as zero and it is not of our interest. In the following, we consider the lower bound of a *saturated network*. Wireless resources have three dimensions: time, frequency, and space. In a *saturated network*, we assume that if there is any unoccupied time/frequency/space to allow new collision-free transmissions, some user will initiate a transmission. As one of the main concerns for MAC protocols is to control congestion in the link layer, performance study for saturated networks can provide important insights and guidelines.

Lemma 2 *Given the ER with radius D in a saturated network of size $L \times L$, the lower bound of the number of CTs is $L^2/(\sqrt{27}D^2)$.*

Proof 2 *We consider the extreme scenario that the ERs of all senders and receivers do not overlap with each other. For a given ER with radius D and a saturated network of size $L \times L$, the lower bound of the number of CTs is equivalent to half of the minimal number of circles with radius D that can cover the area. This is the classical circle covering problem. It has been derived that a lower bound for a covering using equivalent circles is $2\pi/\sqrt{27}$ [115]. Thus, the minimal number of circles covering an area equals $(2\pi/\sqrt{27})(L^2/\pi D^2) = 2L^2/(\sqrt{27}D^2)$. Given that each flow has two non-overlapping circles, the minimal number of CTs in a saturated network is $L^2/(\sqrt{27}D^2)$.*

6.3.4 Exclusive Region Size

From Lemmas 1 and 2, for a saturated network, the upper bound of the number of CTs is six times that of the lower bound, and both are proportional to D^{-2}. The distribution of the number of CTs in a random network is very difficult if not impossible to obtain. Nevertheless, the expected number of CTs should be proportional to D^{-2}: $\mathrm{E[CT]} = \mathrm{k_1/D^2}$, where k_1 is an unknown coefficient.

As the REX sender will use $6PG_0D^{-\alpha}$ as the interference level to set the transmission rate, we then estimate the expected network transport throughput, as given by

$$\frac{k_1}{D^2}\eta W E[d] \log_2(1 + \frac{PE[d]^{-\alpha}}{N_0W + 6PG_0D^{-\alpha}}), \tag{6.4}$$

where $E[d]$ is the expected transmission distance. As the expected transport throughput is a non-linear function of d, using $E[d]$ to get the expected transport throughput is an approximation. Simulation results show that the above approximation is acceptable. Taking the derivative of (6.4), we can obtain the best D value which maximizes the expected network transport throughput.

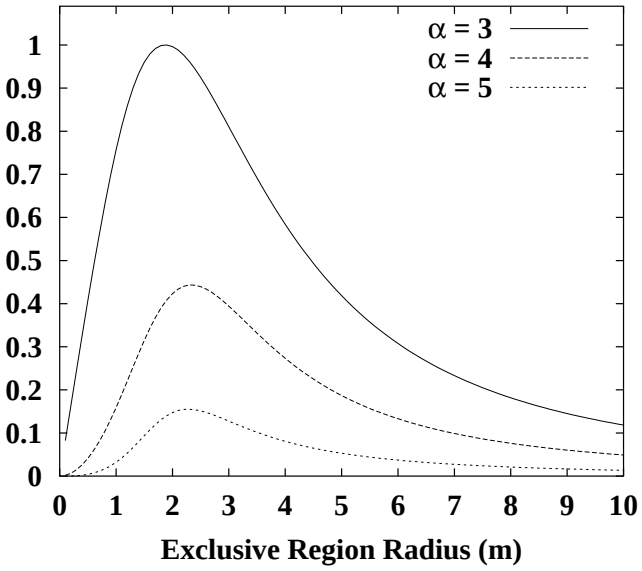


Figure 6.4: Normalized Expected Network Transport Throughput

Fig. 6.4 shows the normalized expected network transport throughput as a function of the exclusive region of radius D, with $G_0 = 0.01$, using the parameters listed in TABLE 6.1. It is shown that the expected network transport throughput is a concave function of D, while fixing other parameters, including P, W, N_0, *etc.* The best value of D can be determined when the maximum expected throughput is achieved. The analytical results of best D under different parameter values of α and G_0 are listed in TABLE 6.2. It is observed that the best D becomes larger when G_0 increases, but changes less with α. This is because the path loss exponent α affects both the received signal strength and the interference level and the corresponding SINR does not change much, while the cross correlation, G_0, determines the interference only (in the denominator of SINR). A greater G_0 results in higher interference among concurrent transmissions, and thus a larger D is required to bound the total interference level to achieve high network throughput.

In practical, the value of α may not be accurately measured or estimated, so the

Table 6.1: System Parameters

W	500 MHz	BIFS	20 μs
$P(s_i)$	-41.3 dBm/MHz	SIFS	10 μs
N_0	-114 dBm/MHz	a slot time	20 μs
α	2.5-6	$CW_{\min}$	31
G_0	0.01-1	$CW_{\max}$	1023
η	0.21	maximum retry limit	7
d_{ref}	1m	RTS/CTS	20 μs
PL_0	43.9 dB	Transmission range	10 m

Table 6.2: Best Exclusive Region Size (Analysis)

α	$G_0 = 0.01$	$G_0 = 0.1$	$G_0 = 1$
3	1.87 m	4.03 m	8.69 m
4	2.34 m	4.15 m	7.39 m
5	2.28 m	3.61 m	5.72 m
6	2.11 m	3.10 m	4.55 m

value of D may not be the best. However, as shown in Fig. 6.4, the best values of D for $\alpha = 3, 4, 5$ only have small difference. Thus, even we under- or over- estimate α, the value of D chosen by the DEX protocol can still be close to the best value.

6.4 Simulation Results

In this section, we evaluate the performance of the proposed DEX protocol in terms of transport throughput, fairness, and access delay, and compare it with that of the IEEE 802.11 DCF via simulations. We choose the IEEE 802.11 DCF protocol as benchmark since it is the most popular asynchronous MAC protocol widely adopted.

6.4.1 Simulation Settings

The simulated network is set up in a 20 m $\times$ 20 m square room, which contains up to 100 active flows, with distinct senders and receivers uniformly distributed in the room. The simulation parameters are listed in TABLE 6.1. The senders use the maximum transmission power and the transmission range is 10 m. The background noise power density is -114 dBm/MHz [92]. All N_f active flows contend for channel access in an asynchronous manner, with its initial arrival time uniformly chosen over $[0, 32)$ time slots. The reference distance is set as $d_{ref} = 1$ m, and the path loss at d_{ref} is 43.9 dB. Thus, the maximum data rate at 1 m, i.e., $R = \eta W log_2(1 + SINR) = 1$ Gbps. The achievable data rate decreases with distance d, *e.g.*, given the path loss exponent $\alpha = 4$, the received $SINR$ degrades from 28.8 dB at $d_{ref} = 1$ m to 16.76 dB at $d = 2$ m and the achievable data rate at 2 m decreases to $R = 585$ Mbps. To eliminate the warming-up effects, the simulation data are collected from 10 s to 60 s. We repeat each simulation 10 times with different random seeds and calculate the average values.

6.4.2 Transport Throughput

The transport throughput of a dense network with 40 flows using the DEX protocol is shown in Fig. 6.5. The data transmission time is $T = 10$ ms. When the ER radius D is very small, more flows are likely to be outside of each other's ER to transmit concurrently; however, a smaller D results in a higher interference level that decreases the data transmission rate.It is observed in Fig. 6.5 that the total transport throughput of the network is maximized if the value of D is close to the best value obtained from the analysis. When the cross correlation G_0 is larger, the interference level among concurrent flows becomes more serious so we should enlarge the value of D accordingly. Simulation results validate the accuracy of our analysis, which demonstrate the significant spatial multiplexing gain achieved by the proposed DEX protocol.

Another observation from Fig. 6.5 is that, if the value of D is slightly different from

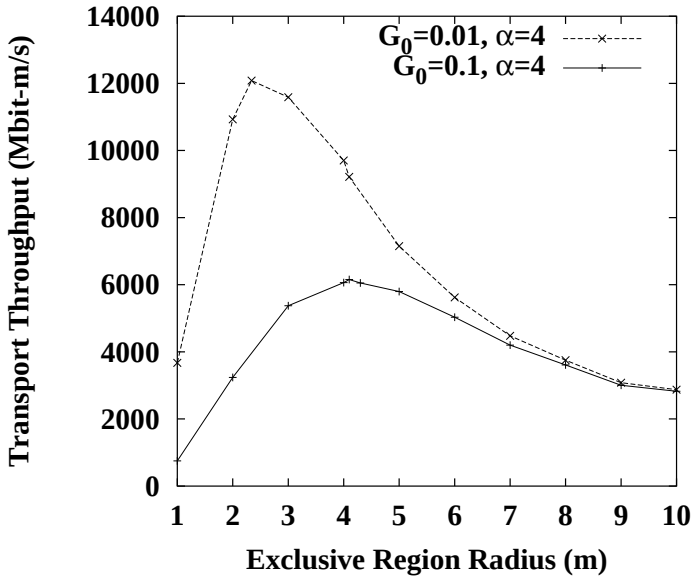


Figure 6.5: Transport Throughput vs ER Radius

the best setting, the throughput is slightly below the highest one. Combining this observation with the results shown in Fig. 6.4, we can claim that, even if the value of α is not accurately obtained, the protocol performance of DEX will not degrade significantly.

We then investigate the network transport throughput under various network densities, and compare the performance of the DEX protocol (with ER radius $D = 4.15$ m, $G_0 = 0.1$, $T = 10$ ms and $\alpha = 4$) with that of IEEE 802.11 DCF (with carrier sensing range of 10 m) in Fig. 6.6. When there are only 10 flows in a 20 m $\times$ 20 m square room, the network is relatively sparse, and the transport throughput of the DEX protocol is 1.45 times that of IEEE 802.11 DCF. When the number of flows exceeds 30, the network throughput of IEEE 802.11 DCF decreases due to serious collisions among the competing flows; while with the proposed DEX, more concurrent transmissions are allowed when the number of active flows increases. The achieved transport throughput of DEX is around 2 times that of IEEE 802.11 DCF with 30 active flows, and 2.7 times

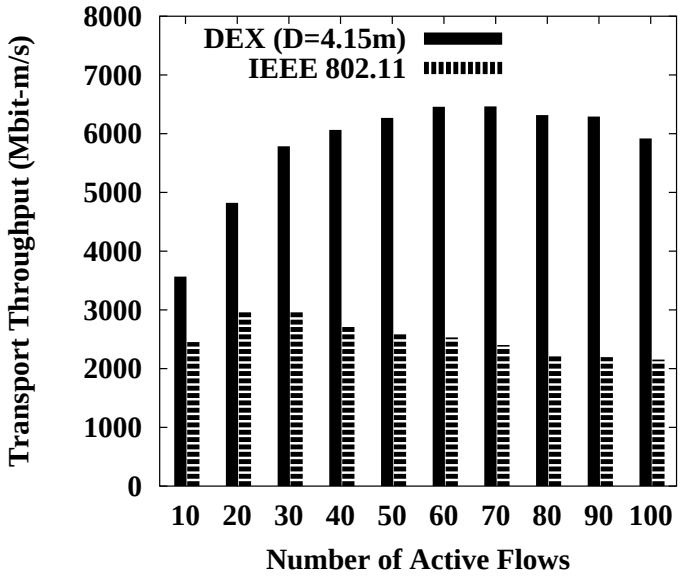


Figure 6.6: Transport Throughput Comparison of Different Protocols

with 70 active flows. It is also shown in Fig. 6.6 that the network transport throughput of DEX increases w.r.t. the network density until the number of active flows in the network exceeds 70, when serious collisions degrade the network performance. In all cases, the proposed DEX significantly outperforms IEEE 802.11 DCF by aggressively exploiting spatial reuse opportunities, and it is more suitable for a dense multi-hop UWB network.

We further study the impact of data transmission time T on both protocols. In the DEX protocol, we use the best ER radius $D = 4.15$ m for $G_0 = 0.1$ and $\alpha = 4$. It is observed in Fig. 6.7 that the transport throughput of 40 active flows increases with the data transmission time T in both DEX and IEEE 802.11.With a larger T, the protocol overheads, including RTS/CTS, backoff time, interframe space, *etc.*, become relatively smaller, and more flows can transmit concurrently to achieve a higher spatial multiplexing gain with DEX. As shown in Fig. 6.7, the ratio of the achieved transport throughput using DEX at the best D to that of IEEE 802.11 increases from 1.6 for

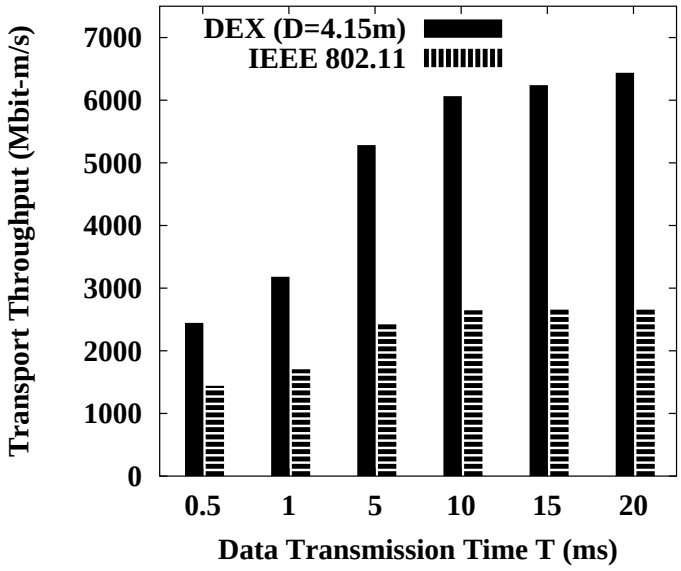


Figure 6.7: Transport Throughput vs Data Transmission Time T

T = 0.5 ms to 2.29 for T = 10 ms. The proposed DEX always outperforms IEEE 802.11 w.r.t. various T values. It is worth noting that a large value of T is preferable for network throughput, but it will result in unfairness problem and longer access delays for other flows in the same contention region. In the following subsections, we investigate the fairness and access delays.

6.4.3 Fairness

Fairness is evaluated using Jain's fairness index [98], in terms of the network transport throughput. We first compare the fairness performance of DEX under various D values with that of IEEE 802.11 DCF in Fig. 6.8. It is well known that the 802.11 DCF based MAC exhibits serious unfairness among competing flows in a multi-hop environment. Some "lucky" flows are more likely to access the channel, while other "unlucky" ones may suffer from complete throughput starvation. Starvation phenomenon becomes

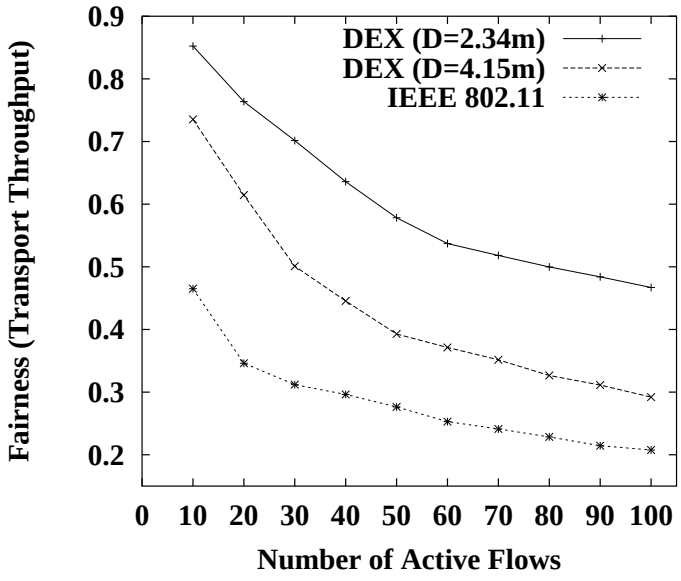


Figure 6.8: Fairness (Transport Throughput) Comparison of Different Protocols

more severe in a denser network. As shown in Fig. 6.8, the fairness performance of both protocols decreases when the number of active flows increases. However, the proposed DEX with smaller ER radius D achieves better fairness than the 802.11 DCF. This is because smaller ER region can effectively reduce the flow starvation by reducing the number of competing flows, and accordingly improve the fairness performance.

The impact of T on fairness is shown in Fig. 6.9. We consider 40 active flows in the room. With a larger value of T, all neighbors within the ER of the tagged sender and receiver have to postpone their transmissions for a longer duration, and thus they are more likely to starve, especially when the ER radius is large and there are many competing flows in the neighborhood. As shown in Fig. 6.9, the fairness performance degrades significantly when the ER radius D and data transmission time T increase.

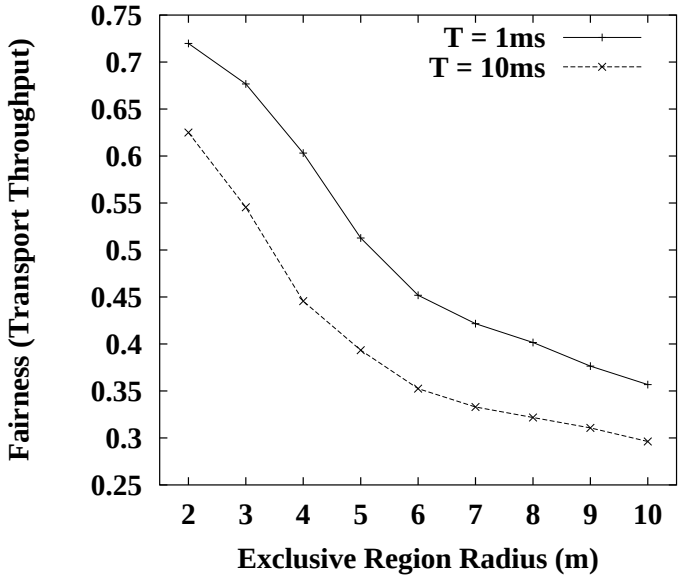


Figure 6.9: Fairness (Transport Throughput) under Various ER Radius

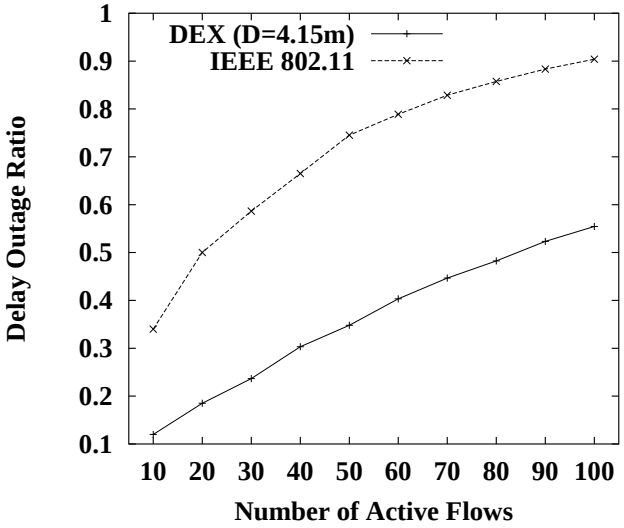


Figure 6.10: Delay Outage Ratio Comparison of Different Protocols

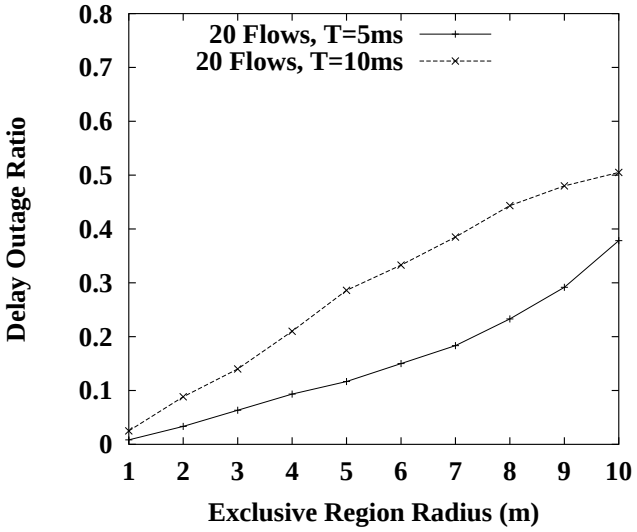


Figure 6.11: Delay Outage Ratio vs ER Radius

6.4.4 Delay Outage Ratio

Access delay is another important performance metric for evaluating a MAC protocol. We define the delay outage ratio as the ratio of the number of attempts with access delay exceeding the delay threshold to the total number of attempts. We set the delay threshold to 150 ms, $G_0 = 0.1$ and $\alpha = 4$. As shown in Fig. 6.10, the delay outage ratio increases with the number of active flows because more collisions among competing flows result in more backoff and thus longer access delays. However, the proposed DEX allows for more concurrent transmissions, so the average access delay of each flow is reduced when the ER radius is smaller.

For choosing the value of T, there is a tradeoff between fairness/delay and transport throughput. We can choose the maximum value of T, under the constraints that the corresponding fairness index is above certain threshold and the delay outage ratio is below certain threshold. Since the ER of DEX is much smaller than the carrier sensing region of 802.11 DCF, DEX can use a larger value of T for higher throughput and still

maintain a desirable delay and fairness performance.

6.5 Summary

In this chapter, we study distributed asynchronous MAC design to effectively utilize the wireless spatial capacity of a multi-hop UWB network. The main accomplishments of this chapter are summarized as follows:

- We have proposed a distributed asynchronous MAC for a multi-hop UWB network, based on the ER concept applied in centralized MAC design in WPANs. Considering UWB characteristics, the proposed DEX can achieve significantly higher network throughput and better fairness performance, while fairness and throughput are usually treated as a tradeoff in other MAC protocols.

- We have systematically analyzed the performance of the proposed MAC. The analytical framework can be applied to other distributed MAC protocols in general.

- We have further investigated the parameter setting to maximize the expected network transport throughput of a randomly and densely deployed multi-hop UWB network. Extensive simulations are performed to demonstrate the effectiveness and efficiency of the proposed MAC.

Chapter 7

Conclusions and Future Work

This dissertation aims at efficient MAC protocol design and analysis in broadband wireless networks, including WLANs, UWB/mmWave based WPANs, and multi-hop UWB networks. In this chapter, we conclude this dissertation by summarizing our contributions and proposing future work.

7.1 Major Research Contributions

7.1.1 Design and Analysis of MAC Protocols in IEEE 802.11 based WLANs

IEEE 802.11 has witnessed the overwhelming success over the last decades. However, IEEE 802.11 WLAN is originally designed for asynchronous data transmission, the network capacity of a WLAN in support of non-persistent realtime multimedia applications is still unknown. In addition, to improve the throughput in the MAC layer, IEEE 802.11n specifies several MAC layer and PHY layer enhancements. The impacts of these enhancements on the MAC layer performance needs to be further investigated. In this dissertation, we have analytically studied IEEE 802.11 MAC and identified the performance bottleneck in an infrastructure WLAN. We have also shown how to mit-

igate the bottleneck effect by exploiting enhanced MAC design and multi-user MIMO beamforming technology in the physical layer. The main contributions are:

- We have developed an analytical framework for capacity analysis of a WLAN by studying the queue utilization of the users. The maximum capacity of the network is obtained such that one more traffic flow will result in at least one of the users' queues become unstable. The generic framework can be applied for WLANs with and without infrastructure support, carrying saturated or non-saturated traffic flows. In addition, the computational complexity of the framework does not scale up with the backoff stages as in other Markov based approaches. We have further applied the framework to obtain the voice and video capacity of a WLAN, i.e., the maximum number of voice and video flows that can be supported in the WLAN with satisfactory user-perceived quality. We found that the queue utilization of the AP is much higher than other users due to heavier traffic loads in an infrastructure mode, which limits the network capacity. Several MAC enhancements, e.g., frame aggregation and bidirectional transmissions, have also also been studied and compared with the legacy MAC.

- IEEE 802.11n WLANs will employ MIMO technology to improve link capacity and reliability. However, it is not practical to implement multiple antennas in a mobile user's device in the near future due to the cost and technical issues. It is more likely that the AP will be equipped with multiple antennas while mobile user only has one omni antenna. We propose a MIMO-aware multi-user MAC to exploit the multi-user MIMO beamforming at the AP. By allowing the AP to communicate with multiple users in the downlink simultaneously, the AP-bottleneck effect in conventional WLAN can be greatly mitigated and the network capacity can be significantly improved.

7.1.2 Efficient Resource Allocation and Capacity Analysis in Centralized UWB/mmWave WPANs

The inherent ultra wide spectrum of UWB and mmWave communications favors concurrent transmissions in the proximity. Oxygen absorption and atmospheric attenuation make it essential to use directional antenna in 60 GHz mmWave systems, which can further improve the spatial utilization of mmWave based WPANs. As the instantaneous interference between peers are usually not available at the PNC, we propose a simple yet efficient ER based resource management schemes to exploit the spatial capacity of UWB/mmWave WPANs, taking advantages of the salient features of UWB and mmWave communication technologies, e.g., wide spectrum, low interference, and precise localization service, etc. The main contributions are:

- We have analyzed the performance of IEEE 802.15.3 MAC and revealed the intrinsic relationship among protocol parameters. That is, given a superframe length, more users can successfully send requests in a longer CAP, but fewer slots can be allocated for data transmissions in the CTAP, and vice versa. The analysis provides important guideline for critical parameter setting to improve the IEEE 802.15.3 MAC performance. We have also quantified the video and voice capacity of a UWB based WPAN using IEEE 802.15.3 MAC. Analytical results show that without considering the particular features of UWB communications, current MAC design is inherent inefficient for supporting emerging broadband killer applications.

- We have justified a sufficient condition to assure that ER based concurrent transmissions outperform serial TDMA transmissions. Then we have derived different ERs based on various types of antenna applied by transmitters and receivers. A simple scheduling scheme has been proposed to improve the spatial utilization of UWB and mmWave based WPANs and provide good fairness performance as

well.

- With the ever-increasing density of wireless devices, researchers are striving to push up the limit of the number of users that can share the available resources. Given a 2D or 3D space with the number of active users randomly distributed, what are the expected per-flow throughput, network capacity, and network transport capacity, and how to maximize the space resource utilization are still open issues. Instead of deriving the asymptotic capacity bounds of arbitrary networks which may be too loose to be useful in realistic networks, we have analytically studied the expected spatial multiplexing capacity of a UWB/mmWave based WPAN. With the increasing number of concurrent transmissions, the interference level becomes higher that degrades the per-flow throughput. Therefore, there is a tradeoff between the per-flow throughput and spatial utilization. Our analysis show that the expected network spatial multiplexing capacity can be maximized under the best ER size. The analytical framework is applicable to other networks in general.

7.1.3 Asynchronous Distributed MAC Design in Multi-hop UWB Networks

UWB is a promising enabling technology to provide very high data rate at a short distance. To meet the escalating demands for high rate wireless connections anywhere and at any time, multi-hop relay is necessary for extending the communication coverage. Since centralized MAC usually involves a large number of communication and computation overheads, it does not scale well with the network size. Moreover, synchronization in a large scale multi-hop UWB networks is very costly if not impossible. Therefore, a simple, scalable, distributed, and asynchronous MAC is more desirable in a multi-hop environment. The main contributions are:

- We have proposed an asynchronous distributed ER based MAC for a densely deployed multi-hop UWB network. The proposed MAC capitalizes on the achievable spatial multiplexing gain of UWB networks by reserving ERs surrounding the sender and receiver for data and ACK transmissions, so that users can efficiently and fairly share network resources in a distributed and asynchronous manner. Further, we have shown that by reserving two small ERs around the transceivers, the proposed MAC does not suffer from the hidden and exposed terminal problems as much as the IEEE 802.11 MAC. Therefore, the proposed MAC can achieve a higher spatial capacity and provides better fairness performance in the mean time.

- Analyzing the capacity of a multi-hop distributed network is very challenging, since it is highly sensitive to the network topology and the sequence of users initiating transmissions. To evaluate the network performance, we first obtain the theoretical upper bound and lower bound of the number of concurrent transmissions in a dense network. We have found that we can improve the network capacity by increasing the number of flows in a sparse network; however, analytical results show that once the user density is large enough to saturate the network, further increasing the user density cannot improve the network capacity, but only increase the competition levels in the network. Instead of depending on complicated control messages, we have studied how to determine protocol parameters to improve the protocol performance in a dense network environment.

7.2 Future Work

This dissertation focuses on MAC design and analysis in various broadband wireless networks, e.g., WLANs and UWB/mmWave networks. There are many relevant research issues that are worth further investigation.

7.2.1 Interference-aware MAC Enhancement

Our current asynchronous UWB MAC design uses a conservative estimation, i.e., an upper bound of the total interference, to decide the transmission rate in order to assure the successful transmissions in a distributed random access network. As wireless users are able to measure and estimate the interference level over the fading channel distributedly, it is possible to improve the transmission rate according to the statistical interference information while still maintain the robustness of the protocol. In addition, most of our work uses a simple path loss model for wireless channels without considering other channel impairments due to fast fading and shadowing in WLAN and mmWave channels. How to accurately estimate the accumulative interference levels and appropriately exploit the diversity gain of wireless channels in a multi-hop network to further improve the network performance and maintain fair channel access is an interesting research issue.

7.2.2 Distributed QoS Control in Multihop UWB/mmWave Networks

The next generation wireless network is envisioned to support heterogeneous multimedia applications with various QoS requirements and evaluation criteria. Although we study the network capacity of WLANs and WPANs in support of voice and video flows, great efforts are still needed for QoS provisioning in emerging broadband multimedia networks. Admission control and MAC layer scheduling are two general QoS control approaches that are relatively easy implement in a single hop or centralized wireless network for QoS provisioning. However, distributed QoS control in a multihop UWB/mmWave network poses many challenges as wireless users only have very limited and inconsistent network information from their neighbors. How to extend our asynchronous MAC design in a multi-hop UWB network for distributed QoS control

requires further research.

7.2.3 Cross-layer Design of Scalable Routing and MAC Protocols

This dissertation primarily focuses on MAC layer issues, and assume a routing protocol is in place to choose the appropriate path from the source to the destination. For future research, a cross-layer design of scalable routing and MAC protocols should be considered. Geographical routing is an efficient approach for wireless ad hoc networks, thanks to the low complexity of the localized routing algorithm, scalability with the network size, and fast route convergence when network topology changes. Taking advantage of the localization service of the UWB/mmWave systems, users are able to decide their locations which makes geographical routing in multi-hop UWB and mmWave networks possible. As the available radio resources in wireless networks are highly dynamic, opportunistic channel access in the MAC layer is also desirable. Therefore, further efforts are needed to exploit the cooperations among physical, MAC and network layers for proper routing decisions.

7.2.4 Energy Efficient Protocol Design in Green Communication Networks

Driven by the steady increase of relevant global CO_2 emission and energy cost, green communication approaches have attracted great research attentions for reducing power consumptions of electrical devices and promoting environment conditions. For example, wireless users may use alternative sustainable energy source such as solar or wind power to replace the traditional AC power. To improve the energy and spectrum efficiency in hybrid solar or wind powered WLANs and WPANs, physical layer power control, link layer resource allocation and opportunistic spectrum access should be

jointly considered, which is a critical yet challenging research issue that deserves further study.

Appendix A

Probability Density Function of Distance in 3D Space

To measure the signal and interference levels for network capacity study, the probability density functions (pdf) of the distance among devices is required. In a three-dimensional space, the distance distribution between two devices can be derived given the distributions of their coordinators in the three dimensions. The following is an example of how to obtain the pdf function of the distance for devices randomly deployed in an $l \times l \times h$ space, i.e., the three coordinators of each device are random variables with uniform distribution.

The probability distribution of the distances between two devices randomly (uniformly) located on a one-dimensional line or in a two-dimensional plane are known [96]. Based on the pdf of the distance on a unit line and in a unit square, we derive the cumulative distribution function (CDF) of the distance in 3D space and obtain its pdf as follows.

Let X be the distance between two devices in an $l \times l$ square with uniform distri-

bution. The pdf of X, $f_X(x)$, is a piecewise function given by,

$$f_X(x) = \begin{cases} 2\frac{x}{l^2}(\frac{x^2}{l^2} - 4\frac{x}{l} + \pi) & \text{for} \quad 0 \le x \le l \\ 2\frac{x}{l^2}(4\sqrt{\frac{x^2}{l^2} - 1} - (\frac{x^2}{l^2} + 2 - \pi) - 4\tan^{-1}\sqrt{\frac{x^2}{l^2} - 1}\ ) & \text{for } l < x \le \sqrt{2}l. \end{cases} \tag{A-1}$$

Let Y be the distance between two devices randomly located on a line of length h. The pdf of Y, $f_Y(y)$, is given by

$$f_Y(y) = \frac{2}{h}(1 - \frac{y}{h}) \quad \text{for } 0 \le y \le h. \tag{A-2}$$

Let Z be the distance between two devices randomly located in an $l \times l \times h$ space. The CDF of Z is derived as

$$\begin{aligned} & \Pr(Z \le z) \\ = & \int_0^h \Pr(X \le x | Y = y) f_Y(y) dy \\ = & \begin{cases} \int_0^h F_X(\sqrt{z^2 - y^2}) f_Y(y) dy & \text{for } z > h \\ \int_0^z F_X(\sqrt{z^2 - y^2}) f_Y(y) dy & \text{for } z \le h, \end{cases} \end{aligned} \tag{A-3}$$

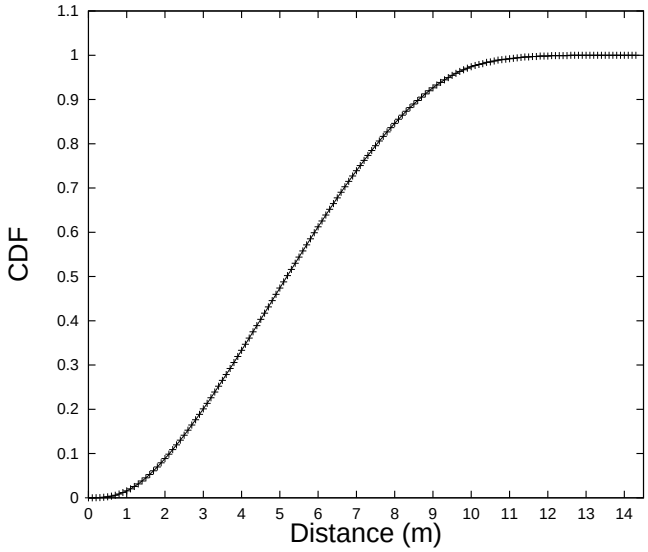
where $F_X(\sqrt{z^2 - y^2})$ can be obtained from (A-1),

$$F_X(\sqrt{z^2 - y^2}) = \int_0^{\sqrt{z^2 - y^2}} f_X(x) dx.$$

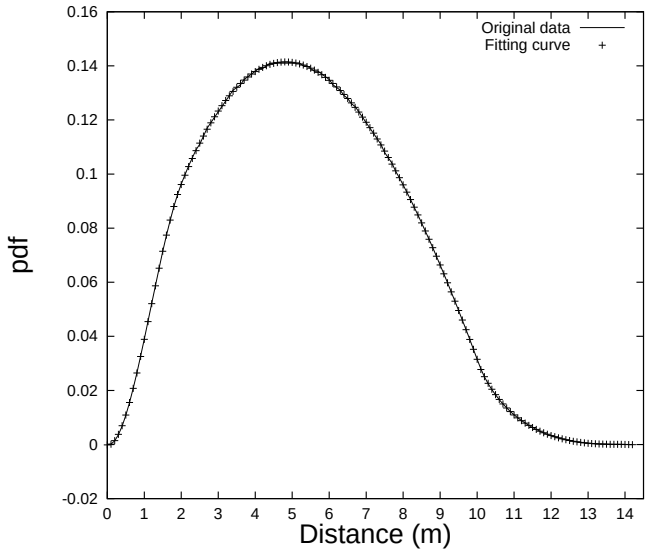
According to (A-1) and (A-3), $f_Z(z)$ is also a piecewise function. To simplify the presentation and analysis, we can use polynomial functions to fit the $f_Z(z)$ in each range:

$$f_Z(z) = \begin{cases} \sum_{i=1}^{j} a_{1i} z^i & \text{for } 0 \le z \le h \\ \sum_{i=1}^{j} a_{2i} z^i & \text{for } h < z \le l \\ \sum_{i=1}^{j} a_{3i} z^i & \text{for } l < z \le \sqrt{2l^2 + h^2}, \end{cases} \tag{A-4}$$

where j is the degree of the polynomials. For instance, using the polynomial fitting function in Matlab, we obtain the numerical results of the CDF and pdf of the distance



(a) CDF



(b) pdf

Figure A-1: Distribution and Probability Functions ($l = 10$, $h = 2$)

between two devices randomly located in a room of $l = 10$ m and $h = 2$ m. The probability functions are shown in Fig. A-1. The coefficient vectors of the three polynomials are

$$\begin{cases} \vec{a_1} = [-0.016556, 0.058759, -0.00231682, -0.00012774], \\ \vec{a_2} = [0.00019773, -0.007981, 0.063154, 0.0003309], \\ \vec{a_3} = [-0.00053767, 0.022103, -0.30305, 1.3859]. \end{cases}$$

The errors introduced by the fitting functions of the third degree polynomials are less than 0.6%, as shown in Fig. A-1.

Bibliography

[1] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Proc. SPIE on Applications of Digital Image Processing XXVII*, Aug. 2004.

[2] A. Chandra, V. Gummalla, and J. O. Limb, "Wireless medium access control protocols," *IEEE Commun. Surv.*, vol. 3, no. 2, pp. 2–15, 2000.

[3] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[4] H. Luan, L. X. Cai, and X. Shen, "Impact of network dynamics on user's video quality: analytical framework and QoS provision," *IEEE Trans. Multimedia*, to appear.

[5] F. Hou, L. X. Cai, P. H. Ho, X. Shen, and J. Zhang, "A cooperative multicast scheduling scheme for multimedia services in ieee 802.16 networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1508–1519, Mar. 2009.

[6] L. X. Cai, X. Shen, J. W. Mark, L. Cai, and Y. Xiao, "Voice capacity analysis of WLAN with unbalanced traffic," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 752–761, May 2006.

[7] L. X. Cai, X. Ling, X. Shen, J. W. Mark, and L. Cai, "Supporting voice and video applications over IEEE 802.11n WLANs," *Wireless Networks*, vol. 15, no. 4, pp. 443–454, Nov. 2009.

[8] L. X. Cai, H. Shan, W. Zhuang, X. Shen, J. W. Mark, and Z. X. Wang, "A distributed multi-user MIMO MAC protocol for wireless local area networks," in *Proc. IEEE Globecom'08*, Nov. 2008, pp. 1–5.

[9] L. X. Cai, X. Shen, J. W. Mark, and L. Cai, "Capacity analysis and MAC enhancement for UWB broadband wireless access networks," *Computer Networks(Elsevier)*, vol. 51, no. 11, pp. 3265–3277, Aug. 2007.

[10] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "REX: a randomized exclusive region based scheduling scheme for mmWave WPANs with directional antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, Jan. 2010.

[11] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "Spatial multiplexing capacity analysis of mmWave WPANs with directional antennae," in *Proc. IEEE Globecom'07*, Nov. 2007, pp. 4744 – 4748.

[12] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "Capacity analysis of UWB networks in three-dimensional space," *IEEE/KICS J. Commu. and Networks (JCN)*, vol. 11, no. 3, pp. 287–296, 2009.

[13] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "MAC protocol design and optimization for multi-hop ultra-wideband networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4056–4065, Aug. 2009.

[14] IEEE 802.11 WG, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification," Aug. 1999.

[15] M. Ergen, D. Lee, R. Sengupta, and P. Varaiya, "Wtrp - wireless token ring protocol," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1863–1881, Nov. 2004.

[16] Y. Bi, K. H. Liu, L. X. Cai, X. Shen, and H. Zhao, "A multi-channel token ring protocol for QoS provisioning in inter-vehicle communications," *IEEE Trans. Wireless Commun.*, to appear.

[17] B. P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai, "Investigation of the IEEE 802.11 medium access control (MAC) sublayer functions," in *Proc. IEEE Infocom'97*, vol. 13, Apr. 1997, pp. 126–133.

[18] M. Veeraraghavan, N. Cocker, and T. Moors, "Support of voice services in IEEE 802.11 wireless LANs," in *Proc. IEEE Infocom'01*, vol. 1, Apr. 2001, pp. 448–497.

[19] Y. Xiao, "IEEE 802.11e: QoS provisioning at the MAC layer," *Wireless Commun.*, vol. 11, no. 3, pp. 72–79, Jun. 2004.

[20] IEEE 802.15.3 WG, "IEEE std 802.15.3 - 2003: Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs)," Sept. 2003.

[21] J. D. P. Pavon, S. N. Shankar, V. Gaddam, K. Challapali, and C. T. Chou, "The MBOA-WiMedia specification for ultra wideband distributed networks," *Wireless Commun. Mag.*, vol. 44, no. 6, pp. 128–134, Jun. 2006.

[22] S. ECMA-368, "High Rate Ultra Wideband PHY and MAC Standard, 2nd Ed." Dec. 2007. [Online]. Available: http://www.ecma-international.org/publications/standards/Ecma-368.htm

[23] IEEE 802.15.4 TG, "IEEE P802.15.4a/D5 (Amendment of IEEE Std 802.15.4), Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LRWPANs)," Sept. 2006.

[24] Standard ECMA-368, "High rate ultra wideband PHY and MAC standard, 2nd ed." Dec. 2007. [Online]. Available: http://www.ecma-international.org/publications/standards/Ecma-368.htm

[25] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

[26] L. X. Cai, X.Shen, and J. W. Mark, "Efficient MAC protocol for ultra-wideband networks," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 179–185, Jun. 2009.

[27] C. Garnier, L. Clavier, Y. Delignon, M. Loosvelt, and D. Boulinguez, "Multiple access for 60 GHz mobile ad hoc network," in *Proc. IEEE VTC'02*, vol. 3, May 2002, pp. 1517 – 1521.

[28] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy-limited wireless networking with directional antennas: the case of session-based multicasting," in *Proc. IEEE Infocom'02*, vol. 1, Jun. 2002, pp. 190–9.

[29] I. Kang and R. Poovendran, "Power-efficient broadcast routing in adhoc networks using directional antennas: Technology dependence and convergence issues," UWEETR-2003-0015, Tech. Rep., Jul. 2003.

[30] R. Ramanathan, "On the performance of ad hoc networks with beamforming antennas," in *Proc. MobiHoc'01*, Oct. 2001, pp. 95–105.

[31] B. Jose, H. Yin, P. Mehrotra, and E. Casas, "MAC layer issues and challenges of using smart antennas with 802.11," in *IEEE VTC'03*, vol. 5, Oct. 2003, pp. 3169–3173.

[32] G. Bianchi, D. Messina, L. Scalia, and I. Tinnirello, "A space-division time-division multiple access scheme for high throughput provisioning in WLANs," in *IEEE ICC'05*, vol. 4, May 2005, pp. 2728– 2733.

[33] L. X. Cai, L. Cai, X.Shen, and J. W. Mark, "Resource management and QoS provisioning for IPTV over mmwave-based WPANs with directional antenna," *ACM Mobile Networks and Applications*, vol. 14, no. 2, pp. 210–219, 2009.

[34] International Telecommunication Union, "G.729 annex b: A silence compression scheme for g.729 optimized for terminals conforming to recommendation v.70," in *Telecommunication Standardization section of ITU*, Nov. 1996.

[35] International Telecommunicaion Union, "General characteristics of international telephone connections and international telephone circuits one-way trans mission time," Feb. 1996.

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

[37] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement," in *Proc. IEEE Infocom'02*, vol. 2, Jun. 2002, pp. 599–607.

[38] P. Chatzimisios, V. Vitsas, and A. C. Boucouvalas, "Packet delay analysis of IEEE 802.11 MAC protocol," in *Electronics Letters*, vol. 39, Sept. 2003, pp. 1358– 1359.

[39] Z. Hadzi-Velkov and B. Spasenovski, "Saturation throughput - delay analysis of IEEE 802.11 dcf in fading channel," in *Proc. IEEE ICC'03*, vol. 1, May 2003, pp. 121–126.

[40] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of ieee 802.11e enhanced distributed coordination function," *IEEE J. Sel. Area Comm.*, vol. 22, pp. 917– 928, Jun. 2004.

[41] Z. N. Kong, D. Tsang, B. Bensaou, and D. Gao, "Performance analysis of ieee 802.11e contention-based channel access," *IEEE J. Sel. Area Comm.*, vol. 22, pp. 2095–2106, 2004.

[42] H. Zhai, X. Chen, S. Member, Y. Fang, and S. Member, "How well can the IEEE 802.11 wireless LAN support quality of service?" *IEEE Trans. Wireless Comm.*, vol. 4, pp. 3084–3094, Nov. 2005.

[43] F. A. Shabdiz and S. Subramaniam, "A finite load analytical model for the IEEE 802.11 distributed coordination function MAC," Mar. 2003.

[44] C. H. Foh and M. Zukerman, "Performance analysis of the IEEE 802.11 MAC protocol," in *Proc. European Wireless*, Feb. 2002, pp. 184–190.

[45] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE Infocom'04*, vol. 2, Mar. 2004, pp. 1404–1413.

[46] O. Tickoo and B. Sikdar, "A queueing model for finite load IEEE 802.11 random access MAC," in *Proc. IEEE ICC'04*, vol. 1, Jun. 2004, pp. 175–179.

[47] Y. Cheng, X. Ling, W. Song, L. X. Cai, W. Zhuang, and X. Shen, "A cross-layer analytical approach for WLAN voice capacity region," *IEEE J. Select. Areas Commun.*, vol. 25, no. 4, pp. 678–688, May 2007.

[48] Y. Cheng, X. Ling, L. X. Cai, W. Song, W. Zhuang, X. Shen, and A. Leon-Garcia, "Statistical multiplexing, admission region, and contention window optimization in multiclass wireless LANs," *ACM Wireless Networks*, vol. 15, no. 1, pp. 73–86, Jan. 2009.

[49] Y. Xiao, "IEEE 802.11 performance enhancement via concatenation and piggyback mechanisms," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2182– 2192, Sept. 2005.

[50] C. Liu and A. P. Stephens, "An analytic model for infrastructure WLAN capacity with bidirectional frame aggregation," in *Proc. IEEE WCNC'05*, Mar. 2005, pp. 113–119.

[51] K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," *IEEE/ACM Trans. Networking*, vol. 15, no. 3, pp. 546–559, Jun. 2007.

[52] M. Hu and J. Zhang, "MIMO ad hoc networks: Medium access control, saturation throughput, and optimal hop distance," *IEEE/KICS J. Commu. and Networks (JCN)*, vol. 6, no. 4, pp. 317–330, Dec. 2004.

[53] J. Mirkovic, G. Orfanos, H.-J. Reumerman, and D. Denteneer, "A MAC protocol for MIMO based IEEE 802.11 wireless local area networks," in *Proc. IEEE WCNC'07*, Mar. 2007, pp. 2131–2136.

[54] D. J. Dechene, K. A. Meerja, A. Shami, and S. Primak, "A novel MIMO-aware distributed media access control scheme for IEEE 802.11 wireless local area networks," in *Proc. IEEE LCN'07*, Oct. 2007, pp. 125–132.

[55] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[56] Y.-J. Choi, N.-H. Lee, and S. Bahk, "IEEE 802.11 performance enhancement by MIMO spatial multiplexing," in *Proc. IEEE PIMRC'05*, vol. 1, Sept. 2005, pp. 87–91.

[57] A. Tarighat, M. Sadek, and A. Sayed, "A multi-user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios," in *Proc. IEEE ICASSP'05*, vol. 3, Mar. 2005, pp. 1129–1132.

[58] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in *Proc. IEEE ICC'04*, Jun. 2004, pp. 196–201.

[59] S. Garg and M. Kappes, "Can i add a VoIP call?" in *Proc. IEEE ICC'03*, May 2003, pp. 779–783.

[60] L. X. Cai, Y. Xiao, X. Shen, L. Cai, and J. Mark, "VoIP over WLAN: Voice capacity, admission control, QoS, and MAC," *Int. J. Communication Systems (Wiley)*, vol. 19, no. 4, pp. 491–508, May 2006.

[61] A. Heck, *Introduction to Maple (3rd ed.)*. New York: Springer-Verlag, 2003.

[62] NS-2, "Network Simulator (2.27)." [Online]. Available: Available at http://www.isi.edu/nsnam/ns/

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

[64] M. Sadek, A. Tarighat, and A. H. Sayed, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1498–1510, Apr. 2007.

[65] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[66] S. Catreux, L. J. Greenstein, , and V. Erceg, "Some results and insights on the performance gains of mimo systems," *IEEE J. Sel. Area Comm.*, vol. 21, no. 5, pp. 839–847, 2003 Jun.

[67] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[68] L. X. Cai, K.-H. Liu, H. Rutagemwa, X. Shen, and J. W. Mark, "Performance analysis of hybrid medium access protocol in an IEEE 802.15.3 WPAN," in *Proc. IEEE CCNC'07*, Jan. 2007, pp. 445–449.

[69] K.-W. Chin and D. Lowe, "Simulation study of the IEEE 802.15.3 MAC," in *Proc. ATNAC'04*, Dec. 2004.

[70] Z. Yin and C. M. Leung, "Connection data rate optimization of ieee 802.15.3 scatternets with multi-rate carriers," *International Journal of Sensor Networks*, vol. 3, no. 2, pp. 95–106, Feb. 2008.

[71] Y.-H. Tseng, H.-K. Wu, and G.-H. Chen, "Maximum traffic scheduling and capacity analysis for IEEE 802.15.3 high data rate MAC protocol," in *Proc. VTC'03*, vol. 3, Oct. 2003, pp. 1678–1682.

[72] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Networking*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[73] R. Jurdak, P. Baldi, and C. V. Lopes, "U-mac: a proactive and adaptive uwb medium access control protocol," *Wirel. Commun. Mob. Comput.*, vol. 5, no. 5, pp. 551–566, Aug. 2005.

[74] B. Radunovic and J. L. Boudec, "Optimal power control, scheduling, and routing in UWB networks," *IEEE J. Sele. Area Comm.*, vol. 22, no. 7, pp. 1252–1270, Sept. 2004.

[75] H. Jiang, K. Liu, W. Zhuang, and X. Shen, "An effective resource management scheme for UWB networks with simultaneous transmissions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3005–3015, Aug. 2007.

[76] V. Rajakumar, M. N. Smadi, S. C. Ghosh, T. D. Todd, and S. Hranilovic, "Interference management in WLAN mesh networks using free-space optical links," *Journal of Lightwave Technology*, vol. 26, no. 13, pp. 1735–1743, Jul. 2008.

[77] R. Merz, J. Widmer, J. L. Boudec, and B. Radunovic., "A joint phy/mac architecture for low-radiated power th-uwb wireless ad-hoc networks," *Wirel. Commun. Mob. Comput.*, vol. 5, no. 5, pp. 567–580, Aug. 2005.

[78] K. H. Liu, L. Cai, , and X. Shen, "Exclusive-region based scheduling algorithms for UWB WPAN," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 933–942, Jan. 2008.

[79] K. H. Liu, L. Cai, and X. Shen, "Multi-class QoS scheduling in UWB networks using discrete stochastic optimization," *IEEE Trans. Veh. Technol.*, to appear.

[80] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. J. W. Rodwell, "Millimeter wave wpan: Cross-layer modeling and multihop architecture," in *Proc. IEEE Infocom'07*, May 2007, pp. 2336–2240.

[81] X. An, S. Zhang, and R. Hekmat, "Enhanced MAC layer protocol for millimeter wave based WPAN," in *Proc. IEEE PIMRC'07*, Sept. 2008, pp. 1–5.

[82] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Theory*, vol. IT-46, pp. 388–404, Mar. 2000.

[83] P. Gupta and P. Kumar, "Internets in the sky: The capacity of three dimensional wireless networks," *Communications in Information and Systems*, vol. 1, pp. 39–49, 2001.

[84] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proc. IEEE Infocom'01*, Apr. 2001, pp. 1360–1369.

[85] S. A. Jafar, "Too much mobility limits the capacity of wireless ad hoc networks," *IEEE Trans. Info. Theory*, vol. 51, pp. 3954–3965, Nov. 2005.

[86] M. Gastpar and M. Vetterli, "The capacity of wireless networks: The relay case," in *Proc. IEEE Infocom'02*, vol. 3, Jun. 2002, pp. 1577–1586.

[87] H. Gamal, "On the scaling laws of dense wireless sensor networks: the data gathering channel," *IEEE Trans. Info. Theory*, vol. 51, pp. 1229–1234, 2003.

[88] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proc. MobiCom'06*, Sept. 2006, pp. 239–250.

[89] R. Zheng, "Information dissemination in power-constrained wireless networks," in *Proc. IEEE Infocom'06*, Apr. 2006, pp. 1–10.

[90] R. Negi and A. Rajeswran, "Capacity of power constrained ad-hoc networks," in *Proc. IEEE Infocom'04*, vol. 1, Mar. 2004, pp. 443–453.

[91] V. P. Mhatre and C. P. Rosenberg, "The capacity of random ad hoc networks under a realistic link layer model," *ArXiv Computer Science e-prints*, Apr. 2005.

[92] IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a, "DS-UWB physical layer proposal," IEEE P802.15-04/0137r4, Jan. 2005.

[93] J. S. Davis, "Indoor wireless RF channels." [Online]. Available: http://wireless.per.nl/reference/chaptr03/indoor.htm

[94] S. Ramanathan, "A unified framework and algorithms for (T/F/C)DMA channel assignment in wireless networks," in *Proc. IEEE Infocom'97*, Apr. 1997, pp. 900–907.

[95] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "Efficient resource management for mmwave WPANs," in *Proc. IEEE WCNC'07*, Mar. 2007, pp. 3816–3821.

[96] Math world, "Square line picking," Jan. 2005. [Online]. Available: http://mathworld.wolfram.com/SquareLinePicking.html

[97] A. F. Molisch, "Ultrawideband propagation channels-theory, measurement and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1528–1545, Sept. 2005.

[98] R. Jain, A. Durresi, and G. Babic, "Throughput fairness index: an explanation," in *ATM Forum Document Number: ATM Forum/990045*, Feb. 1999.

[99] X. Shen, W. Zhuang, H. Jiang, and J. Cai, "Medium access control in ultrawideband wireless networks," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1663–1677, Sept. 2005.

[100] L. X. Cai, L. Cai, X. Shen, and J. Mark, "Capacity of UWB networks supporting multimedia services," in *Proc. IEEE Qshine'06*, Aug. 2006.

[101] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks–part I: The fluid model," *IEEE/ACM Trans. Networking*, vol. 14, no. SI, pp. 2568–2592, 2006.

[102] T. Moscibroda, R. Rejaie, and R. Wattenhofer, "How optimal are wireless scheduling protocols?" in *Proc. IEEE Infocom'07*, May 2007, pp. 1433–1441.

[103] Y. Yi, G. de Veciana, and S. Shakkottai, "On optimal MAC scheduling with physical interference," in *Proc. IEEE Infocom'07*, May 2007, pp. 294–302.

[104] K. K. P. Chaporkar and S. Sarkar, "Throughput guarantees through maximal scheduling in wireless networks," in *Proc. Allerton Conference on Communication, Control and Computing*, Sept. 2005, pp. 28–30.

[105] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proc. IEEE Infocom'05*, vol. 3, Mar. 2005, pp. 1804–1814. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1498460

[106] X. Wu, R. Srikant, and R. James, "Scheduling efficiency of distributed greedy scheduling algorithms in wireless networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 595–605, Jun. 2007.

[107] R. Merz, J. Widmer, J.-Y. L. Boudec, and B. Radunovic, "A joint PHY/MAC architecture for low-radiated power TH-UWB wireless ad-hoc networks: Research articles," *Wirel. Commun. Mob. Comput.*, vol. 5, no. 5, pp. 567–580, Aug. 2005.

[108] B. Hajek, A. Krishna, and R. Lamaire, "On the capture probability for a large number of stations," *IEEE Trans. Commun.*, vol. 45, pp. 254–260, Feb. 1997.

[109] T. Kim, H.Lim, and J. C. Hou, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. MobiCom'06*, Sept. 2006, pp. 366–377.

[110] X. Yang and N. H. Vaidya, "On physical carrier sensing in wireless ad hoc networks," in *Proc. IEEE Infocom'05*, vol. 4, Mar. 2005, pp. 2525–2535.

[111] J. Zhu, S. Roy, X. Guo, and W. S. Conner, "Maximizing aggregate throughput in 802.11 mesh networks with physical carrier sensing and two-radio multi-channel clustering," in *Proc. NSF-FPI workshop*, 2004.

[112] D. Jing, B. Liang, and P. K. Varshney, "Tuning the carrier sensing range of IEEE 802.11 MAC," in *Proc. IEEE Globecom'04*, vol. 5, Nov. 2004, pp. 2987–2991.

[113] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "Optimizing distributed MAC protocol for ultra-wide band wireless networks," in *Proc. IEEE Infocom'08*, Apr. 2008, pp. 1912–1920.

[114] H. SteinhausM, *Mathematical Snapshots (3rd ed.)*. New York: Dover, 1999.

[115] R. Williams, "Circle coverings," *Geometrical Foundation of Natural Structure: A Source Book of Design*, pp. 51–52, 1979.