# Design, Modeling, and Analysis for MAC Protocols in Ultra-wideband Networks

by

Kuang-Hao Liu

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, 2008

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

Ultra-wideband (UWB) is an appealing transmission technology for short-range, bandwidth demanded wireless communications. With the data rate of several hundred megabits per second, UWB demonstrates great potential in supporting multimedia streams such as high-definition television (HDTV), voice over Internet Protocol (VoIP), and console gaming in office or home networks, known as the wireless personal area network (WPAN). While vast research effort has been made on the physical layer issues of UWB, the corresponding medium access control (MAC) protocols that exploit UWB technology have not been well developed.

Given an extremely wide bandwidth of UWB, a fundamental problem on how to manage multiple users to efficiently utilize the bandwidth is a MAC design issue. Without explicitly considering the physical properties of UWB, existing MAC protocols are not optimized for UWB-based networks. In addition, the limited processing capability of UWB devices poses challenges to the design of low-complexity MAC protocols. In this thesis, we comprehensively investigate the MAC protocols for UWB networks. The objective is to link the physical characteristics of UWB with the MAC protocols to fully exploit its advantage. We consider two themes: centralized and distributed UWB networks.

For centralized networks, the most critical issue surrounding the MAC protocol is the resource allocation with fairness and quality of service (QoS) provisioning. We address this issue by breaking down into two scenarios: homogeneous and heterogeneous network configurations. In the homogeneous case, users have the same bandwidth requirement, and the objective of resource allocation is to maximize the network throughput. In the heterogeneous case, users have different bandwidth requirements, and the objective of resource allocation is to provide differentiated services. For both design objectives, the optimal scheduling problem is NP-hard. Our contributions lie

in the development of low-complexity scheduling algorithms that fully exploit the characteristics of UWB.

For distributed networks, the MAC becomes node-based problems, rather than link-based problems as in centralized networks. Each node either contends for channel access or reserves transmission opportunity through negotiation. We investigate two representative protocols that have been adopted in the WiMedia specification for future UWB-based WPANs. One is a contention-based protocol called prioritized channel access (PCA), which employs the same mechanisms as the enhanced distributed channel access (EDCA) in IEEE 802.11e for providing differentiated services. The other is a reservation-based protocol called distributed reservation protocol (DRP), which allows time slots to be reserved in a distributed manner. Our goal is to identify the capabilities of these two protocols in supporting multimedia applications for UWB networks. To achieve this, we develop analytical models and conduct detailed analysis for respective protocols. The proposed analytical models have several merits. They are accurate and provide close-form expressions with low computational effort. Through a cross-layer approach, our analytical models can capture the near-realistic protocol behaviors, thus useful insights into the protocol can be obtained to improve or fine-tune the protocol operations. The proposed models can also be readily extended to incorporate more sophisticated considerations, which should benefit future UWB network design.

**Acknowledgements**

My PhD study at the University of Waterloo has been a journey of discovery and professional growth. Having arrived where I am today, I owe thanks to many people.

I am sincerely grateful to my supervisor, Professor Xuemin (Sherman) Shen, for his continuous guidance, encouragement, patience, and financial support during my PhD study. His invaluable advice on how to select research topics, define the scope of problems, further work on them, and scientifically present the results has shined me through and will definitely help me in my future career.

I am very grateful to my thesis committee members: Professor Guoliang Xue, Professor Yulia Gel, Professor Sagar Naik and Professor Dongmei Zhao. Their precious time in reviewing my thesis and providing insightful counsels on both research and writing are highly appreciated.

I would also like to thank Professor Jon W. Mark for his valuable suggestions on my research. I have been fortunate to have chances of working with him.

My deep appreciation goes to Drs. Jun Cai, Lin Cai, Hai Jiang, Xinhua Lin, and Humphrey Rutagemwa for their fruitful discussions and professional collaborations. I would also like to thank all my colleagues in the Broadband Communications Research Group (BBCR) who have supported me at various stages of my graduate career.

The administrative staff: Ms. Wendy Boles, Ms. Lisa Hendel and Ms. Karen Schooley have offered a great deal of assistance on organizing my defense.

Special thanks to my mentors in Taiwan: Professor Jennhua Chang, Professor Wu-Chung Su, and Ms. Lynn Chu for being so supportive all these years.

Last but not least, I would like to express my heartfelt gratitude to my parents, my sister, and Denise Wang. Throughout the year of my study, they have provided me a stable background and no doubt that their unflagging encouragement and priceless love have made it possible for me to carry out my research.

## Dedication

*Dedicated to the memories of my grandparents*

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Recent advances in semiconductor industry have bolstered the development of portable devices and their applications on information exchange and data delivery. Users equipped with such devices can be connected through various wireless networks, such as cellular networks, wireless metropolitan area networks (WMANs), wireless local area networks (WLANs), and wireless personal area networks (WPANs). While these wireless networks are fundamentally different in their service scenarios and underlying technologies, a common trend is that they all have been developed for providing higher bandwidth to satisfy the growing demand of bandwidth-required multimedia content distribution on portable devices.

The ultra-wideband (UWB) is a promising technology for ubiquitous connectivity in home/personal space. Devices equipped with UWB transceivers can carry a wide set of multimedia applications such as high-definition (HD) television, photos, videos, console gaming, voice over Internet Protocol (VoIP), music sharing, *etc.*, with high display quality and sustainable battery life. While vast research effort has been dedicated to physical layer issues of UWB, medium access control (MAC) and higher layer protocols that exploit UWB technology have not been well addressed.

The importance of the MAC protocol can be understood from its functionalities. A MAC protocol is the process of coordinating the competing users in accessing the wireless medium and sharing the radio resource. In wireless communications, the uncoordinated transmission is regarded as interference that may cause link broken or packet lose. The MAC protocol ensures that wireless medium is efficiently used and fairly shared by users. Therefore, the design of MAC protocols has been deemed the key factor in determining the success of providing wireless services to users.

Generally MAC protocols can be classified into centralized and distributed ones. In centralized MAC protocols, the MAC functionality is performed at a central controller. In distributed MAC protocols, each node performs the MAC functionally according to the locally collected information. Despite that numerous MAC protocols have been proposed in the literature, specific considerations should be made for UWB networks in order to leverage the advantage of UWB in providing multimedia services. Lacking of explicit considerations of the physical properties of UWB, existing solutions remain considerable space for improvement. Therefore, the first objective of this thesis is devoted to *linking the physical characteristics of UWB communications with the MAC protocol design.*

The increasing demand of high-resolution digital content has driven the bandwidth requirements at a dramatic pace. For emerging multimedia applications, conventional wireless systems that are mainly designed to provide audio and lately data services are short of supporting high bandwidth-demanded multimedia applications. While the intrinsic wide bandwidth of UWB makes it a promising candidate for various multimedia streams in digital home and office, it is imperative to take into account the characteristics of multimedia traffic in developing appropriate MAC protocols. The second objective of this thesis is to *assess the performance of MAC protocols in supporting multimedia applications over UWB networks.* In the next section, we

identify the main challenges to MAC protocol design in UWB networks for achieving the aforementioned two goals.

## 1.1 Challenges to UWB MAC Protocol Design

With the capability of supporting very high data rates within short range, UWB is the best candidate for the high rate WPAN. We introduce the WPAN by highlighting its features as follows:

1. *Infrastructure-free*: Unlike cellular networks or WLANs, there is no fixed infrastructure in WPANs. Several devices can autonomously form a WPAN, which can be flexibly reestablished, merged, or split in response to the environment change.

2. *Dense-deployed devices with limited computational capability*: Communications in WPANs are confined to a small area such as home, office, around persons or objects. Thus devices may be densely deployed in the network. On the other hand, UWB aims at low-cost, low-power devices. UWB devices can only afford limited computation power.

3. *Peer-to-peer*: Devices in WPAN can communicate with each other via peer-to-peer connections. Consequently, the central controller is no longer the traffic bottleneck as that in cellular networks and WLANs, where all traffic is routed through the central entity. On the other hand, this increases the difficulty of acquiring the link quality information, which is the necessary input for making scheduling decisions.

4. *Heterogeneous and high bandwidth demand*: A WPAN powered by UWB can support multiple high-definition streams with bandwidth requirements on the

order of hundred megabits per second Mbps. Meanwhile, users can also launch real-time audio applications and bulky data transfer. Therefore, a UWB-based WPAN presents the requirement of supporting heterogeneous traffic with high bandwidth requirements.

Based on the above observations, in what follows, we identify the main challenges to MAC protocol design for UWB-based WPAN. We discuss two distinct types of UWB networks: *centrally controlled* and *distributively controlled*, since they possess different design requirements, and are being developed by different standardization bodies.

### 1.1.1 Centralized WPAN

In centralized WPAN under the consideration of IEEE 802.15.3 [1], the central controller, called the piconet controller (PNC), is responsible for allocating the radio resource to its members. Resource allocation in wireless networks generally consists of three themes: throughput maximization, fairness, quality-of-service (QoS) provision. For throughput maximization, information-theoretical analysis [2] has shown that the optimal scheduling for maximizing the network throughput in UWB networks is to properly allow the concurrent transmission. To determine the optimal concurrent transmission set, the PNC needs the feedback information about the instantaneous channel gain from each link. For WPANs with peer-to-peer connections, this is practically not feasible. Furthermore, the optimal concurrent transmission set problem can be shown to be NP-hard. Therefore, how to balance the achieved network throughput and the induced computational effort is critical for UWB devices equipped with limited computation capability.

Another import issue in resource allocation is fairness. A resource allocation

strategy that maximizes the network throughput is generally unfair, because radio resource is always allocated to the user that can best exploit it, *i.e.*, the one with the strongest channel [3]. On the other hand, if fairness is defined as the equal share of resource, the network throughput cannot be maximized. There has been a significant research effort on investigating the tradeoff between the efficiency and fairness of resource allocation.

As a summary, providing QoS in UWB networks confronts different challenges from classical wireless systems. Since UWB is envisioned to support various traffic types with different bandwidth requirements, providing differentiated service becomes essential to ensure QoS. Furthermore, fairness needs to be redefined to reflect heterogeneous bandwidth requirements.

### 1.1.2 Distributed WPAN

The distributed WPANs possess some different design challenges from centralized ones. Without the aid of the central controller, maximizing the network throughput in distributed networks is more difficult. Although several distributed MAC protocols have been developed to exploit the capability of UWB of allowing concurrent transmission [4–6], these protocols commonly introduce considerable signaling overhead. Thus they are not suitable for UWB-based WPAN that requires fast deployment and low operation complexity.

As a matter of practical concern, the WiMedia Alliance has included two simple mechanisms in their specification for future UWB-based WPANs. One is a contention-based protocol employing the same mechanisms as the enhanced distributed channel access (EDCA) in IEEE 802.11e. The EDCA protocol has been successfully deployed in WLANs, due to its effectiveness in supporting heterogeneous traffic. Although significant research has been devoted to studying the performance of EDCA, its ca-

pability in serving multimedia applications remains unexplored. The main difficulty in analyzing the EDCA protocol arises from the complex contention behaviors among traffic streams belonging to different classes. Meanwhile, the inherent characteristics of multimedia traffic (bursty and correlated inter-packet arrival times) render the delay performance difficult to analyze. For EDCA protocol to continue providing effective service differentiation, one of the key issues is to thoroughly study and improve its performance in distributed UWB networks.

Another distributed MAC protocol defined in WiMedia specification is called the distributed reservation protocol (DRP). The DRP allows users to reserve time slots, or equivalently, bandwidth, in a distributed manner. Guaranteed bandwidth is important to real-time applications to ensure QoS. While the PCA protocol assigns real-time traffic with the highest priority, bandwidth guarantee is still difficult in a contention-based protocol. On the other hand, reservation-based protocols such as time division multiple access (TDMA) can more effectively bound the delay by allocating each user a fixed number of time slots and thus guaranteed bandwidth. However, when the reservation procedure is fully distributed as in DRP, the available time slots for reservation may non-uniformly spaced, leading to arbitrary reservation pattern. It is well-known the reservation pattern has a significant impact to the delay performance in reservation-based protocols. Even with granted time slots, the shadowing effect may also drastically degrade the quality of real-time applications. In an indoor environment, the shadowing effect often takes place when a person walks through the line-of-sight (LOS), which contains the most significant transmission power of an ongoing link. Despite the shadowing is a short-term phenomenon, its effect may expand to a large amount of data, especially when the transmission rate is fast. In the literature, the centralized TDMA protocol and its variants have been extensively studied [7–9]. The aforementioned problems in the distributed slot

reservation protocol have not been addressed. In view of the ugent need of DRP to support multimedia applications in UWB networks, it is important to understand the impact of reservation pattern and the shadowing effect to the performance of DRP over UWB channels.

## 1.2 Problem Description

The objective of this research is to comprehensively investigate the MAC protocols in UWB networks. Both centralized and distributed UWB networks are considered, as described as follows.

**Design and Evaluation for Centralized UWB Networks**

For centralized UWB networks, the main challenge is on how to efficiently allocate the network resource, taking into account the characteristics of UWB. Here the term "resource" explicitly means the "time slots". Considering the capability of UWB of allowing concurrent transmission, the resource allocation problem in UWB networks can be regarded as two subproblems. Firstly, how to schedule the concurrent transmission links within one slot. Secondly, how to allocate time slots to all links. Because of the increased interference introduced by concurrent transmission, arbitrarily allowing concurrent transmission may lead to lower aggregate throughput than TDMA, *i.e.*, a scheduling without concurrent transmission. Therefore, we consider TDMA as the benchmark and try to find the condition when concurrent transmission can produce higher aggregate throughput than TDMA.

Based on the derived condition, we can determine the concurrent transmission links within one slot. The next question is how the resource, namely the total time slots, should be allocated. We consider two scenarios: homogeneous and heterogeneous networks. The homogeneous network corresponds to the case that users

have the same bandwidth requirement, and the objective of resource allocation is to maximize the network throughput. In heterogeneous case, users lunch different applications and thus have different bandwidth requirements. We adopt the notion of "utility" to reflect the distinct bandwidth requirements of individual traffic class, leading to a utility maximization problem.

**Modeling ana Analysis for Distributed UWB Networks**

In distributed UWB networks, the resource allocation is performed locally at each device. We consider a beacon-enabled network such that the time can be aligned within at least two-hop radio range. Furthermore, time is divided into the superframe consisting of multiple time slots. Users acquire time slots for packet transmission via contention-based or reservation-based approach. Both approaches are considered in this thesis, with the focus on their capability in supporting multimedia applications.

For contention-based approach, we consider priority channel access (PCA) protocol specified in the WiMedia standard. The main reason of choosing PCA protocol lies in its support of heterogeneous traffic for future UWB-based WPAN. We investigate the capability of PCA protocol in providing QoS to multimedia traffic. This requires us to look into the characteristics of multimedia traffic that are significantly different from typical data traffic. Specifically, the objective of this research is to develop an appropriate analytical model, whereby we can observe the impact of various system parameters, such as burstiness/correlation in interarrival times, contention window size, and interspace between contention zones of different traffic classes, to the protocol performance. Based on these observations, we can improve the protocol to best support multimedia traffic in UWB networks.

For reservation-based approach, we consider the DRP, where the time slot can be reserved through a negotiation procedure between the sender and the receiver. We identify three issues relevant to the QoS provisioning for delay-sensitive multi-

media traffic. Firstly, the available slots for reservation may arbitrarily located in the superframe, leading to the "arbitrary reservation pattern". This is fundamentally different from the ordinary TDMA protocol, where time slots are either contiguous or uniformly spaced. We investigate the impact of reservation pattern to the protocol performance in terms of throughput and delay. Secondly, the reserved time slots may not be used if the sender buffer becomes empty from time to time. Alternatively, the unused time slots that are indicated as "soft reservation" can be accessed by other users if the slots are not used. From network-centric viewpoint the soft reservation is advantageous to reducing channel idle time and thus improves the channel utilization. From user-centric perspective, we challenge whether the use of soft reservation is beneficial to the delay performance. Apart from the protocol operation, we are also interested in the impact from the propagation environment. UWB applications will mostly take place in the indoor environment, *e.g.*, living room or office building, where people might frequently walk through the LOS of an ongoing link and introduce the short-term shadowing. Our research goal is to identify the impact of the short-term shadowing to the DRP performance. Thus, we develop an analytical model that captures the joint behavior of the queue length variation and the time-varying UWB channel under the given reservation pattern and method.

## 1.3 Organization of This Thesis

This thesis is divided into two main parts. In the first part, including Chapter 2, 3 and 4, we consider resource allocation problems in centrally controlled UWB networks.

In Chapter 2, we review the fundamental properties of UWB and the related work on UWB MAC design. The two standardization bodies of UWB and their specifications are also briefly reviewed.

Chapter 3 considers the link scheduling problem for throughput maximization in centralized UWB networks. We attempt to link the physical properties of UWB to the design of scheduling algorithm. To achieve this, we derive the sufficient condition when the concurrent UWB transmissions yields higher throughput than TDMA scheduling. Based on this result, we propose low-complexity scheduling algorithms for UWB networks.

In Chapter 4, we further consider the network consisting of heterogeneous traffic with different bandwidth requirements. The assumption of perfect distance information is also relaxed that leads to a stochastic optimization problem. Based on the global search algorithm (GSA) [10], we propose an improved heuristic algorithm tailored for UWB networks.

In the second part, containing Chapter 5 and Chapter 6, we conduct performance modeling and analysis for distributive MAC protocols, aiming at their capabilities of supporting multimedia traffic in UWB networks.

Chapter 5 considers the contention-based PCA protocol specified in the WiMedia standard. We develop an analytical model for studying the performance of PCA, considering the bursty/correlated interarrival times of multimedia traffic. We perform extensive simulations to verify the accuracy of the proposed analytical model and provide insights into the PCA performance.

In Chapter 6, we consider the reservation-based DRP. We develop an analytical model, which captures the joint behavior of queue variation, non-uniformly spaced reserved slots, and the time-varying UWB channel due to a single moving obstacle. Different reservation methods are also considered. Based on the proposed analytical model, we conduct a thorough investigation on the interaction of DRP performance with numerous system parameters.

### 1.3.1 Bibliographic Notes

Most of the research results reported in this thesis have disseminated through the following publications: Chapter 3 and Appendix A appeared in [112, 116]; Chapter 4 appeared in [111, 115]; Chapter 5 and Appendix B appeared in [110]; Chapter 6 and Appendix D appeared in [113, 114].

# Chapter 2

# Related Work

The distinct traits of UWB from traditional narrowband communication systems[1] stimulates the need of revisiting various MAC issues such as resource allocation and scheduling, as will be investigated in this thesis. To explore the design space of UWB networks, this chapter overviews some fundamental properties of UWB and the related work. In Section 2.1, we briefly introduce UWB and highlight the main features, followed by the current standards for UWB networks in Section 2.2. Related literature on UWB is reviewed in Section 2.3.

## 2.1 Ultra Wideband

UWB technology encompasses a broad range of signal forms and design approaches. Any wireless transmission scheme is qualified as UWB as long as it achieves the transmission bandwidth more than 500 MHz or a fractional bandwidth of at least 25%,

[1] A radio system whose fractional bandwidth is of the order of 1% or less is considererd as narrowband. The fractional bandwidth is obtained by dividing the transmission bandwidth by the center frequency.

as defined by the Federal Communication Commission (FCC) [11]. The deployment of UWB systems is approved on the unlicensed 3.1–10.6 GHz band subject to a strict power spectral density (PSD) ($-41.25$ dBm/MHz) [11] to prevent interfering with other radio systems. The application scenarios for UWB communications fall in either high data rates (HDRs) over short range, under the consideration of the IEEE 802.15.3a Task Group, or low data rates (LDRs) over medium-to-long ranges, considered by the IEEE 802.15.4a Task Group. This thesis focuses on the HDR scenario of UWB.

The very first UWB system is known as "impulse radio" (IR), where each signal is transmitted using an extremely short pulse of duration $T_c$ (typically, $T_c = 0.1 \sim 1.5$ nanoseconds). To increase the reliability, the same symbol is repeated a certain number of times $N_c$. The bit interval $T_b$ is divided into multiple frames, each with duration $T_f = T_b/N_c$, where $T_f$ is the pulse repetition time typically being a hundred to a thousand times $T_c$. Therefore a UWB signal has a very low duty cycle inherently providing the capability of multiple accessing. When multiple users access the channel concurrently, a large number of pulses might arrive at the receiver concurrently if pulses are uniformly spaced, resulting in catastrophic collisions. To reduce the possibility of catastrophic collisions, each user can use a different pseudo-random sequence to determine in which frame should a pulse be transmitted. Such a pseudo-random sequence is called a time-hopping (TH) sequence, and it can be deemed a means to spread the signal over a wide spectrum [12].

The IR-UWB can be jointly used with numerous modulation methods. Alternatively, UWB pulses can be transmitted without modulation onto a sine wave carrier frequency, in contrast with conventional radio frequency technology. The carrier-less modulation leads to low-power and low-cost transceiver circuitry. The IR or TH-UWB has been widely used in radar and military applications due to its high spatial

resolution, immunity to passive interference, low probability of detection, etc [13]. However, it is not suitable for high-rate wireless networks, since the low-duty cycle property of IR-UWB limits the maximum achievable bit rate up to several tens of megabits per second. Additionally, the IR-UWB employs carrier-less modulation challenges the carrier-sensing mechanism that is widely used in distributed wireless networks to detect the channel activity. Therefore, modern UWB systems employ other techniques such as impulsive direct-sequence spread-spectrum (DSSS) [14] or nonimpulsive orthogonal frequency division multiplexing (OFDM) [15] to achieve high data rate and spread the signal over the wide spectrum.

The UWB technology promises short-range wireless applications with high bandwidth demand. The general features of UWB systems can be summarized as follows.

- The extremely wide bandwidth of UWB inherently allows multiple concurrent transmissions within close proximity that facilitates indoor wireless applications with dense user population [16]. In addition, UWB signals are less sensitive to multipath fading [17] and can provide accurate distance information [18].

- The power consumption of UWB transmission is very low (about 0.05 mW)[2], which promises its application on portable devices with limited power supply.

## 2.2 MAC Standards for UWB Networks

As UWB is envisioned to be the transmission technology for HDR WPANs, we briefly introduce the MAC protocols defined in current WPAN standards. The main functionality of MAC is to coordinate channel use among multiple users. The MAC protocol plays an important role in enabling various networking functionality and

[2]Existing WLAN technologies require 50 mW in average power consumption.

providing QoS to user applications.

### 2.2.1 IEEE 802.15.3

To allow fast network formulation among users with QoS assurance, the IEEE 802.15.3 working group has defined the MAC mechanisms for WPAN, by which a number of devices can exchange high-volume data with each other and free from needing any pre-established infrastructure [1]. Among a set of devices in the vicinity, one of them is elected as the piconet coordinator (PNC), responsible for the channel access control and radio resource allocation.

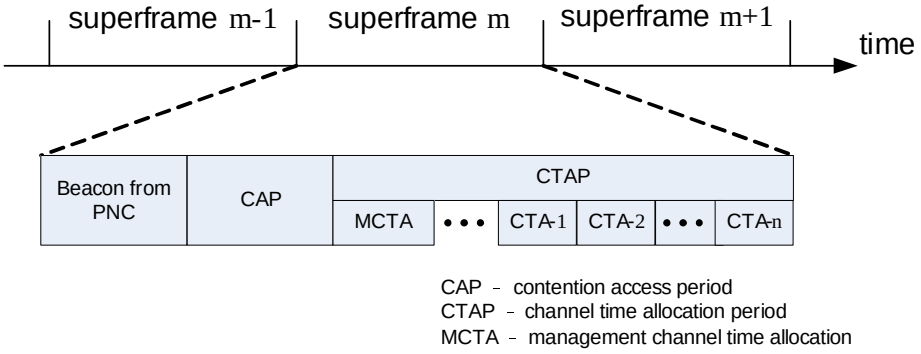



Figure 2.1: Superframe structure defined in IEEE 802.15.3 MAC protocol.

In IEEE 802.15.3 standard, the channel time is divided into superframes, with the structure shown in Figure 2.1. Each superframe starts with a beacon period (BP), during which the PNC sends the beacon containing network synchronization and control message. Devices then can access the channel using either the contention or contention-free mechanisms. In contention access period (CAP), devices send their resource requests to the PNC using carrier sensing multiple access with collision avoidance (CSMA/CA).

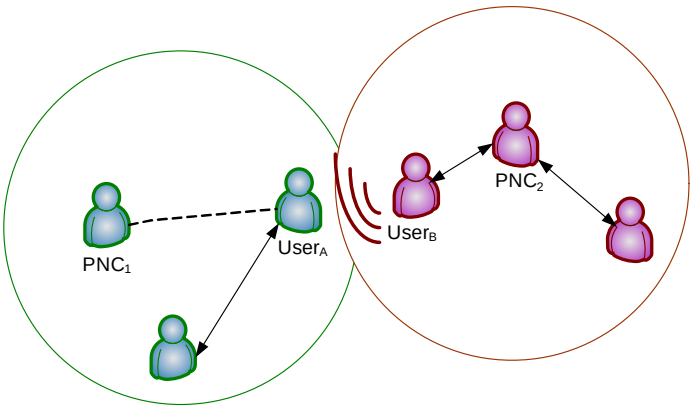



Figure 2.2: Illustration of inter-piconet interference in the IEEE 802.15.3 MAC protocol. Due to lack of coordination, $\text{User}_\text{B}$ might interfere with $\text{User}_\text{A}$.

Besides BP and CAP periods, the remaining time in a superframe allocated for data transmission is called channel time allocation period (CTAP), which is further divided into equal-length time slots or called channel time allocations (CTAs). The PNC of a WPAN allocates CTAs to both asynchronous and isochronous data (periodic traffic, *e.g.*, video and audio streams with time bound) using the TDMA discipline. CTAs that are used for PNC to communicate with devices for command exchange are called management channel time allocations (MCTAs). We summarize the IEEE 802.15.3 MAC by discussing its pros and cons.

- Unlike conventional centralized solutions such as the base station in cellular networks or the access point in WLANs, the PNC is not restricted to be a fixed facility. This allows fast network deployment.

- Users in WPAN can communicate in a peer-to-peer fashion that relieves the load of PNC, which needs to relay the user traffic in the legacy Bluetooth based WPAN, due to its master-slave configuration.

- The hybrid MAC protocol, combining contention and contention-free, is more efficient than the use of single one.

- In case the PNC is shut down or leaves abruptly, it may take several seconds for the rest of the devices to reorganize and elect a new PNC. Thus the network is prone to single point of failure.

- The centralized WPAN suffers from inter-piconet interference as illustrated in Figure 2.2. Without inter-piconet signaling, $\text{User}_\text{A}$ in piconet 1 is interfered with the $\text{User}_\text{B}$ of piconet 2. In this scenario, $\text{PNC}_1$ and $\text{PNC}_2$ are hidden from each other.

### 2.2.2 WiMedia MAC

The IEEE 802.15.3 was defined based on a narrow-band physical layer without explicit consideration of UWB. There has been a standard debate between the choice of direct-sequence based and OFDM based UWB for future WPAN. In 2005, the WiMedia Alliance launched its PHY and MAC specifications based on multiband-OFDM (MB-OFDM) technology [19]. The most distinct difference of WiMedia MAC from IEEE 802.15.3 is that there is no need of a central controller, as the MAC functionality is fully distributed.

The WiMedia MAC also employs a superframe structure consisting of two parts: the BP and the data transfer period (DTP). The beacon frame transmitted in the BP contains important timing and control information to maintain the fully distributed channel access. Each device needs to send its beacon frame during the BP. To deal with the possible collisions of beacon frames, a beacon collision resolution protocol (BCRP) is suggested in the standard that guarantees all devices within two-hop radio range use a different beacon slot. The set of devices that have the same beacon

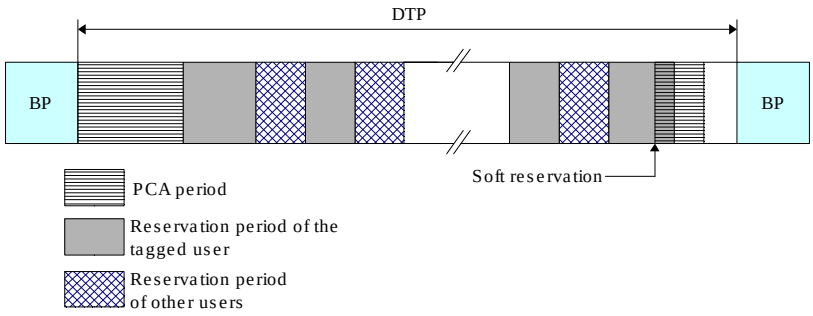



Figure 2.3: Superframe structure defined in WiMedia MAC protocol.

period start time (BPST) forms a logical group, or called beacon group. The union of a device's beacon group and the beacon groups of all devices in the tagged device's beacon group becomes an extended beacon group.

Besides BP, the remaining time in the superframe is used for data transmissions. During DTP, devices can access the channel using two different methods: a contention based prioritized channel access (PCA) protocol, and a reservation based distributed reservation protocol (DRP). The PCA provides differentiated channel access via the same differentiation mechanisms as the enhanced distributed channel access (EDCA) in IEEE 802.11e. The DRP can be regarded as a distributed TDMA protocol. Through DRP, guaranteed bandwidth can be reserved locally, which is important to streaming applications. The operations of these two protocols will be detailed in Chapter 5 and Chapter 6, respectively. The WiMedia MAC is summarized as follows:

- It eliminates the need of a central entity.
- It eliminates the hidden terminal problem in typical ad-hoc networks by beacon exchange.
- The use of hybrid MAC protocols improves channel utilization.

- Devices in an extended beacon group can not transmit in the same beacon slot. As the node density increases, the excessive contentions may degrade the protocol efficiency dramatically.

## 2.3 Related Work

There has been a significant research on UWB MAC and related issues. With the same goal of fully exploiting the advantage of UWB, several MAC protocols based on MSI have been proposed. Another line of research focuses on the joint consideration across protocol stacks. Since the high data rate transmission provided by the UWB physical layer introduces significant overhead to the upper layers, how to reducing this side effect has also drawn a lot of attentions.

### MSI-based MAC Protocols

One of the most notable characteristics of UWB is that, its extremely wide transmission bandwidth inherently allows concurrent transmission. To exploit this feature, several MAC protocols based on the notion of maximum sustainable interference (MSI) have been developed for UWB networks. The MSI indicates the maximum tolerable interference of a node to maintain the required SINR. For centralized networks, a node wishing to join sends a request to the central entity. According to the interference level introduced by this new node, the central entity admits the new node if its transmission is less than the MSI of existing nodes. The employment of MSI-based schemes can be tracked back to CDMA-based ad hoc networks. This approach is recently revisited in [4], and adopted by several companion papers for designing the MAC protocol based on TH-UWB [5, 6, 20]. The admission control can also be performed locally, but frequent exchanges of control messages among nodes

are necessary that introduce significant protocol overhead. In addition, MSI-based protocols suffer the near-far problem, because of the analogue of TH-UWB to CDMA systems. In [6], a MSI-based MAC protocol is proposed to resolve severe near-far problem. Each user selects the time slot such that its transmission introduces the least interference from/to other existing users at that slot. Hence, the nodes whose transmission have a larger impact to each other can be scheduled exclusively.

Traditional MAC protocols that enforce temporally mutual exclusion between concurrent transmissions are over-cautious for UWB networks. While MSI-based protocols are more suitable and effective for UWB networks, some issues need to be further addressed, *e.g.*, enormous signaling overhead, its application to multi-hop communications, and so on.

### Cross-Layer Design

In view of the essentiality of UWB as a physical layer technology, a cross-layer approach is expected to provide more comprehensive guideline for UWB network design. In [21], a joint power control, scheduling, and routing problem for UWB ad hoc networks is investigated. The cross-layer optimization is formulated as a mixed-integer nonlinear programming (MINP) problem and solved by a heuristic approach for a small-sized network. The most important finding is that, with the linear rate function provided by extremely wide bandwidth of UWB, the proportional rate allocation (by maximizing the sum of logarithmic rates) can be achieved using a simple on-off power allocation, *i.e.*, each link transmits with either zero or full power. This result is independent of the choice of routing and MAC protocol. In addition, by implementing an exclusive region around each receiver, the optimal flow rate can be attained and the radius of exclusive region is relevant to the achievable flow rate. However, how to determine a proper exclusive region size is untouched. The same optimization

problem is revisited in [22] with more focus on algorithm development for solving the nonlinear optimization problem.

Different from conventional wireless networks with narrow bandwidth and large transmitting power, UWB radios possess extremely wide bandwidth and low transmitting power. In the limit of infinite bandwidth, it is preferable to transmit over the entire bandwidth, implying the optimality of CDMA for multiple access [2]. For UWB networks with finite but large bandwidth, concurrent transmissions are still favorable to maximizing network throughput. However, certain exclusive mechanisms are required, *e.g.*, assigning links with different subbands [23]. Although the result is obtained based on the ideal assumption that the entire bandwidth can be partitioned arbitrarily small, it indeed indicates that the fundamentally different characteristics of UWB could lead to a new dimension of UWB MAC design.

For single-hop networks, the maximum network throughput is usually achieved by coupling power allocation with link scheduling. In [24], it has shown that for UWB networks, the throughput improvement resulted from power allocation is relatively small as compared with that from proper scheduling. Thus sophisticated power allocation may not be necessary if the primary goal is to maximize the network capacity.

### Overhead Reduction

UWB is capable of providing several hundreds Mbps data rate that can benefit many multimedia applications with high bandwidth demands. However, the very high rate of UWB transmissions also introduces some downsides to the MAC protocol.

The first issue is the overhead caused across protocol stacks. For example, the physical and MAC headers that contain important control information are usually protected with higher redundancy and thus transmitted using a much lower rate than that for payload. This leads to low channel utilization as a relatively large proportion

of time is consumed on overhead. Additionally, the high rate transmission of UWB challenges the synchronization speed. To achieve bit synchronization, the sender's clock and the receiver's one must be aligned before starting the communication. For UWB signals with ultra short duration, the time required to establish synchronization can be prohibitive. In the presence of multipath fading, rapid synchronization is more challenging.

Several research works have been conducted in attempt to reduce the overhead and its effect to UWB networks. In [25], a number of techniques are proposed to address the problem of long timing acquisition at the MAC layer. The basic idea is to pack multiple upper-layer packets in one MAC frame and transmit the MAC frame as one unit. By aggregating multiple packets into one MAC frame, the synchronization overhead can be greatly reduced. The drawback of this approach is, a higher packet error probability may be expected when transmitting a large frame over the error-prone wireless channel. Another solution proposed in [26] is to persistently maintain the physical link by sending low-rate control packets if there is no data to send. Thereby the sender and receiver do not need to re-synchronize once they are engaged. This approach inevitably increases the transmission time leading to increased power consumption of an ongoing link and interference to other links in the mean time. In [6], a distributed MAC protocol is proposed for UWB networks, where consecutive slot reservation is favored when a link is admitted into the network. By scheduling a link in consecutive slots, the acquisition overhead except the first slot is eliminated since the receiver and sender have been synchronized. Such a strategy does not introduce extra signaling overhead nor incurs higher probability of transmission errors and thus is more suitable to high-rate UWB networks with limited power supply.

Another source of overhead is the interframe space between two successive frames. The interframe space is used for the receiver to tell the end of one frame transmissions

and potentially the start of the next one. During this interval, the channel is idle and thus is regarded as an overhead. As the transmission speed increases, the overhead of interframe space becomes more notable. However, the length of interframe space can not be arbitrarily shortened. Alternatively, the overhead of interframe space can be reduced by using one acknowledgement (ACK) frame to acknowledge multiple data frames. In general, a data frame is followed by the ACK frame to signal the sender if retransmissions are required. By doing so, the number of interframe spaces can be reduced so as the ACK frames. Such an ACK mechanism is called Delayed ACK (Dly-ACK) or burst ACK (B-ACK) and has been adopted in IEEE 802.15.3 [1] for WPAN and IEEE 802.11e [27] for WLAN. The number of frames that can be transmitted consecutively before the ACK is called the burst size. Setting a proper burst size is critical to the protocol performance. A large burst size is beneficial to improve channel utilization, since the channel is used more often for delivering data. However, increasing the burst size also results in longer inter-packet delay that may not be acceptable to time-sensitive applications [28]. Some recent work has explored the optimal configuration of Dly-ACK, under different objectives, *e.g.*, optimal burst size in minimizing packet delay with respect to the traffic load [29], and optimal payload size for throughput maximization [30].

## 2.4 Summary

In this chapter, we first introduced the essential properties of UWB and the current MAC standards for UWB networks. We then reviewed the previous work relevant to UWB MAC design, including three major issues: MAC protocols for concurrent transmission, cross-layer considerations aiming at throughput maximization, and techniques to reduce the impact of protocol overhead in UWB networks. Lacking

of explicit considerations of the physical properties of UWB, there remains a large space for existing solutions to be further improved. In the following chapters, we will explore the MAC protocol design for UWB networks. Our focus is on linking the fundamental properties of UWB to MAC protocols.

# Chapter 3

# Throughput Improvement via Concurrent Link Scheduling

## 3.1 Introduction

Given the extremely wide UWB bandwidth, a fundamental question is how this fat wireless pipe can be optimally and efficiently utilized. Achieving optimal channel utilization in wireless networks generally is achieved by solving a cross-layer problem, involving power allocation, scheduling, and routing. Although such a cross-layer design can achieve the optimality, the formulated optimization problem is commonly complex and requires high computational effort to solve. The problem is even more difficult when the user QoS requirements are concerned. As mentioned in Chapter 1, complexity issue is crucial to UWB network design.

Previous work has suggested that, in the wide-band region, the optimal channel utilization can be achieved by allowing concurrent transmission [2] using a simple

on-off power allocation[3]. For UWB networks with large but finite bandwidth, concurrent transmission can still be beneficial to improve the network throughput, where the major improvement is resulted from proper scheduling for concurrent transmission rather than from power allocation [21, 24]. In fact, the stringent emission masks imposed on UWB according to the FCC regulations have limited the maximum transmission power of UWB to a very low level. Thus we eliminate sophisticated power allocation and concentrate on the design of achieving optimal channel utilization in UWB networks, which can be formulated as a link scheduling problem.

The impact of concurrent transmission is twofold. Due to the increased multiple user interference (MUI) from concurrent transmission, the received signal quality of individual link is degraded and thus the link throughput. On the other hand, by properly allowing concurrent transmission, each link can obtain more transmission opportunity than using a temporal exclusive mechanism such as TDMA. Consequently, the total throughput of the network can be improved, or equivalently, better channel utilization.

This chapter is dedicated to the design of scheduling algorithms for concurrent UWB links with the goal of maximizing the network throughput. The optimal scheduling for the concurrent transmission can be shown to be NP–hard. Thus we aim at low complexity algorithms for practical concern. Because of the increased interference introduced by concurrent transmission, arbitrarily allowing concurrent transmission may not be better off than TDMA, *i.e.*, a scheduling without concurrent transmission. Therefore, we consider TDMA as the benchmark and try to find the condition when concurrent scheduling can produce higher aggregate throughput than TDMA. Based on this condition, we develop low-complexity scheduling algorithms for improving the networking throughput.

[3]The definition of on-off power allocation can be referred to Chapter 2.

The remainder of this chapter is organized as follows. Section 3.2 presents the system model. In Section 3.3 we analyze the condition when concurrent transmission achieves higher network throughput than TDMA. Based on this result, two low-complexity scheduling algorithms that improve the network throughput are presented in Section 3.4. The performances of the proposed scheduling algorithms are evaluated by simulations in Section 3.5. Section 3.6 summarizes this chapter. Some terminologies will be frequently used in this chapter with the following definitions:

- Channel utilization: It is the achieved throughput normalized by the channel capacity. A higher channel utilization immediately implies a higher network throughput.
- MUI factor: It is the reciprocal of the processing gain. In a spread spectrum system, the processing gain is the ratio of the spread bandwidth to the unspread bandwidth. We use the inverse of the processing gain to characterize the robustness of the UWB signal against MUI.

## 3.2 System Model

The considered network is a beacon-enabled UWB network, where one of the devices is selected as the piconet controller (PNC) responsible for broadcasting beacons and scheduling the channel access for its members. The beacon frame contains synchronization information, control messages, and scheduling decisions. We consider a superframe structure as the one used in the IEEE 802.15.3 (see Figure 2.1). Devices can reserve time slots by sending their requests using CAP. Upon receiving the request for resource, the PNC makes the scheduling decision and encodes the result in the beacon. The scheduling algorithm is performed only when there are changes taken place in the network topology or the bandwidth demand of individual user.

The transmitted signal passes through a channel which introduces additive white Gaussian noise (AWGN), power attenuation due to propagation path loss, and interference. The AWGN channel is assumed on the basis that UWB signals experience less fast multipath fading due to its fine multipath separation and the multipath combining at the receiver end [17], and that slow fading has a larger time-scale compared to a packet transmission time [21]. Unlike typical centralized wireless networks, the PNC in a peer-to-peer network is not capable of learning the instantaneous channel gain of each link. Alternatively, the channel quality can be measured using the link distance information provided by the accurate ranging capability of UWB. Denote the distance between the receiver and the transmitter of link $i$ by $d_i$, the received power $p_r(i)$ can be estimated as $p_r(i) = p_t d_i^{-\gamma}$, where $\gamma$ is the path-loss exponent. Shannon's upper bound gives the achievable throughput $R_i$ as $W \log_2(1 + SINR)$, where $SINR = \frac{p_r(i)}{(\eta+I_0)W}$, $W$ denotes the transmission bandwidth, $\eta$ is the one-sided power spectral density (PSD) of thermal noise, and $I_0$ is the PSD of interference[4], respectively. In the wideband regime, *i.e.*, $W \to \infty$ [32],

$$R_i \approx \frac{p_r(i)}{\eta + I_0} \log_2 e. \tag{3.1}$$

## 3.3 Scheduling for Concurrent Transmission

As discussed in the beginning of this chapter, multiple UWB links can transmit concurrently to improve the channel utilization. The effect of concurrent transmission is twofold. For individual link, the concurrent transmission degrades the received signal quality and thus the instantaneous data rate. On the other hand, by allowing concurrent transmission, each user can be assigned with more transmission opportunities such that their total throughput is possibly increased. Such a tradeoff can be

[4]The Gaussian approximation holds in the presence of a large number of interferers [31]

characterized by the concept of *exclusive region*, defined as the spatial region around each receiver in which no concurrent transmission is allowed. A larger exclusive region helps to limit the aggregate interference by inhibiting the nearby dominant interferers, with the cost of less spatial reuse. For the extreme case that the exclusive region is infinitely large or small, the scheduling discipline reduces to TDMA (*i.e.*, no concurrent transmission), or all-at-once (*i.e.*, all links transmit concurrently), respectively. Considering TDMA as the benchmark, the following lemma states the sufficient condition that the concurrent UWB transmission is preferable to TDMA in improving the network throughput.

**Lemma 3.1.** *Without loss of generality, consider scheduling N links during a scheduling cycle of N time slots. The link throughput, resulted from concurrent transmission with the exclusive region of radius $\delta$, is larger than that from TDMA when*

$$I_{j,i} \leq \eta, \quad j \neq i \tag{3.2}$$

*where $I_{j,i}$ represents the interference from link $j$'s source to link $i$'s destination.*

*Proof.* Denote $R_i^{\mathsf{T}}$ and $R_i^{\mathsf{C}}$ the throughput of link $i$ under TDMA and concurrent transmissions, respectively, during a scheduling cycle. With TDMA scheduling, the achievable data rate of link $i$ is

$$R_i^{\mathsf{T}} = k' p_r(i)/\eta = k' p_t d_i^{-\gamma}/\eta, \tag{3.3}$$

where $k' = \log_2 e$. For concurrent transmissions, the achievable data rate of link $i$ is

$$R_i^{\mathsf{C}} = \frac{N k' p_r(i)}{\eta + \sum_{j \neq i} I_{j,i}} = \frac{N k' p_t d_i^{-\gamma}}{\eta + \sum_{j \neq i} I_{j,i}}, \tag{3.4}$$

where $I_{j,i}$ is the interference from link $j$'s source to link $i$'s destination with distance $d_{j,i}$. Let $\delta$ be the distance such that $I_{j,i}$ equals $N_0$. If all interferers are at least $\delta$

distance away from the receiver of link, *i.e.*, $d_{j,i} \geq \delta$, it implies $I_{j,i} \leq \eta$ for all $j \neq i$. The resultant link throughput with the concurrent transmission is

$$R_i^{\mathsf{C}} > \frac{Nk' p_t d_i^{-\gamma}}{\eta + (n-1)\eta} = \frac{k' p_t d_i^{-\gamma}}{\eta} = R_i^{\mathsf{T}}. \tag{3.5}$$

□

**Theorem 3.1.** *Denote by $\delta$ the radius of exclusive region, as defined in (3.2). Any two links $i$ and $j$ can transmit concurrently such that (3.5) is satisfied if the following condition holds:*

$$d_{i,j} > \delta \text{ and } d_{j,i} > \delta. \tag{3.6}$$

The minimal radius of the exclusive region $\delta$ that ensures $I_{j,i} \leq \eta$ depends on the processing gain of UWB signals and the background noise level, and it is independent of the link length. The condition (3.6) implies that scheduling concurrent UWB communications is preferable to TDMA transmissions so long as all interferers are outside the exclusive regions of other receivers.

To achieve the maximum throughput, the optimal radius of exclusive region needs to be decided. In general, there is no close-form solution for arbitrary network topology. In the following an asymptotic analysis is conducted to identify the performance bound. For the random distributed network, simulations are used to study the impact of exclusive region in Section 3.5.

**Asymptotic Analysis**

Consider a squared area of size $\mathcal{A}$ in which nodes are densely and uniformly distributed as shown in Figure 3.1. Let the radius of exclusive region be $\delta$, such that the shortest distance from the interferers to the tagged receiver located at the center of the square is $\delta$. For a tagged receiver, there are four interferers with distance $n\delta$ for $n = 1, 2, 3, ...$, four interferers with distance $\sqrt{2}n\delta$, and $8(n-1)$ interferers with

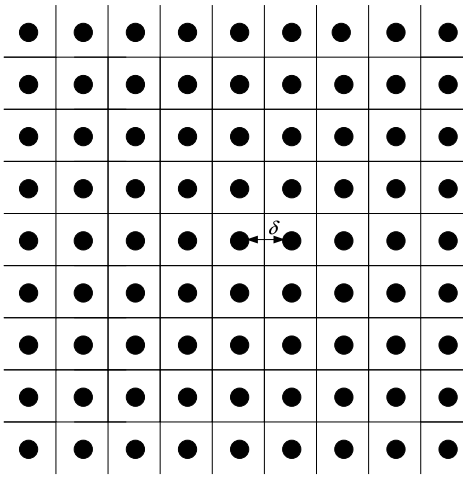



Figure 3.1: The Voronoi diagram of a dense UWB network.

distance in between $nr$ and $\sqrt{2}n\delta$. Denote $b$ the MUI factor, which represents the cross-correlation of the target signal and interfering signal. For the tagged receiver, the total interference from all of the other concurrent transmitters can be approximated as,

$$\begin{aligned} I &\approx bp_t \sum_{n=1}^{\infty} \{4[(n\delta)^{-\gamma} + (\sqrt{2}n\delta)^{-\gamma}] + 8(n-1)[(1+\sqrt{2})n\delta/2]^{-\gamma}\} \\ &= 4bp_t\delta^{-\gamma}[1 + 2^{-\gamma/2} - (1+\sqrt{2})^{-\gamma}2^{1+\gamma}]\zeta(\gamma) + 8bp_t(\delta/2)^{-\gamma}(1+\sqrt{2})^{-\gamma}\zeta(\gamma-1), \end{aligned} \tag{3.7}$$

where $\zeta$ is the Riemann Zeta-function defined as $\zeta(\gamma) = \sum_{n=1}^{\infty} n^{-\gamma}$. Note that the convergence of $\zeta(\gamma)$ for $\gamma > 1$ implies that (3.7) converges when $\gamma > 2$. Therefore, even with infinite numbers of interfering nodes, the total interference $I$ is bounded for $\alpha > 2$. For indoor environment, the pathloss exponent typically ranges from 2 to 7 [33]. To ease the presentation, (3.7) can be simplified as $I = b\delta^{-\gamma}C_\gamma$, where $C_\gamma$, as a function of $\gamma$, contains all the remaining terms. As a result, the achievable data

rate of a tagged user $i$ can be given by

$$R_i = k' \frac{p_r(i)}{\eta + bC_\gamma \delta^{-\gamma}}. \tag{3.8}$$

With the number of nodes approximated by $n \approx \frac{\mathcal{A}}{\delta^2}$, the aggregate throughput of concurrent transmissions is thus

$$\begin{aligned} \sum_{i=1}^{n} R_i &= \frac{k' \sum_{i=i}^{n} p_r(i)}{\eta + bC_\gamma \delta^{-\gamma}} \\ &= \frac{k' \mathcal{A} \bar{p}}{\eta \delta^2 + bC_\gamma \delta^{2-\gamma}}, \end{aligned} \tag{3.9}$$

where $\bar{p} = 1/n \cdot \sum_{i=1}^{n} p_r(i)$. The maximum throughput is achieved when $[\eta\delta^2 + bC_\gamma\delta^{2-\gamma}]$ is minimized. The optimal radius of exclusive region $\delta^*$ can thus be obtained as

$$\delta^* = [\frac{(\gamma - 2)bC_\gamma}{2\eta}]^{1/\gamma}. \tag{3.10}$$

It can be seen that $\delta^*$ is a function of the pathloss exponent, the background noise level, and the MUI factor. The maximum network throughput is given as

$$R_{\max} = \frac{2k' \mathcal{A} \bar{p}}{bC_\gamma (\delta^*)^{2-\gamma}}. \tag{3.11}$$

**Validation**

To verify the above analysis, we simulate a dense network with the same constellation as in Figure 3.1. The network size is 400m × 400m, where nodes are uniformly distributed following the grid topology. The tagged receiver is located at the center of the network, and the corresponding sender is 2 m away from the receiver. The simulation parameters are $p_t = 0.05$ mW, $N_0 = 2.5 \times 10^{-8}$ mW, and $b = 10^{-4}$. In simulations, the distance between the closest interferer $j$ and the target receiver $i$, denoted as $d$, is a multiple of $\delta$ as that defined in Lemma 3.1. Then the value $d$ that achieves the maximum aggregate throughput is regarded as the optimum size of

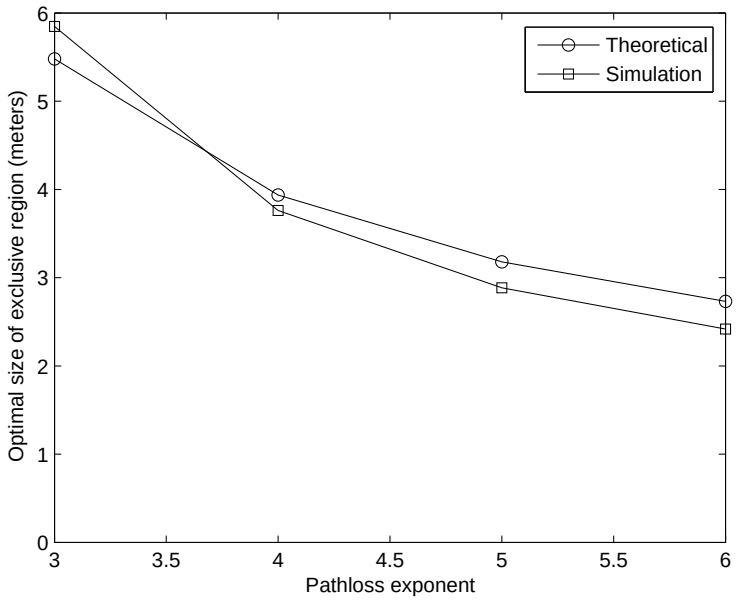



Figure 3.2: Optimal radius of exclusive region obtained from (3.10) and simulations.

exclusive region. Figure 3.2 shows the optimal radius of exclusive region with respect to different values of path-loss exponent obtained from (3.10) and from simulations. The path-loss exponent $\gamma$ is selected according to the measurement results reported in [33]. In office and residential environments, $\gamma$ ranges from 3 to 4 for soft non-line-of-sight (NLOS) and from 4 to 7 for hard-NLOS. It can be seen the theoretical values match the simulated ones very well.

Scheduling concurrent transmission can achieve both spatial reuse and time multiplexing. The following proposition gives the throughput gain of concurrent transmission.

**Proposition 3.1.** *Define the spatial reuse factor as the ratio between the per link throughput obtained from concurrent transmission with the optimal exclusive region $\delta^*$ and that without concurrent transmission. Scheduling concurrent transmissions*

*can achieve the spatial reuse factor up to* $1-\frac{2}{\gamma}$*, and the time multiplexing gain of* $\frac{\mathcal{A}}{(\delta^*)^2}$*. Hence, the throughput gain of concurrent transmissions using optimal exclusive region over TDMA is* $(1-\frac{2}{\gamma})\frac{\mathcal{A}}{(\delta^*)^2}$*.*

Proposition 3.1 can be readily obtained using the fact that for each link, the spatial reuse gain is $1-\frac{2}{\gamma}$, by substituting $\delta^*$ into (3.8). Meanwhile, the time multiplexing gain resulted from concurrent transmissions is $\frac{\mathcal{A}}{(\delta^*)^2}$, as this is the number of links allowed to transmit concurrently. During the time window of $n$ slots, where $n=\frac{\mathcal{A}}{(\delta^*)^2}$, each link has a throughput increase of $(1-\frac{2}{\gamma})\frac{A}{(\delta^*)^2}$, compared to TDMA that each link can use only one slot during the same time window. The analysis suggests that, properly scheduling concurrent transmission can improve both per link throughput and network throughput. In practice, the network topology may be arbitrary and finding the optimal schedule for concurrent transmission becomes computationally expensive. Based on Theorem 3.1, the rest of this chapter is devoted to designing of low-complexity scheduling algorithms for UWB networks.

## 3.4 Efficient Scheduling Algorithms

Designing appropriate scheduling algorithms for UWB networks should take into account two important requirements: (i) Low computational complexity, due to the limited processing power of PNC. (ii) The magnitude of MUI needs to be estimated to decide the concurrent transmission set. Concerning the complexity, the scheduling problem for finding the optimal concurrent transmission set for maximizing the throughput is *NP-hard* (see Appendix A). Obtaining the optimal solution within polynomial time is thus infeasible. The second requirement is also challenging because the PNC is not capable of acquiring instantaneous channel status of individual links, each of which is point-to-point connected. Alternately, the location/distance

information obtained from accurate ranging capability of UWB can be used to estimate the link condition. By jointly considering the above requirements, two heuristic scheduling algorithms are proposed in the following.

### 3.4.1 Proportional Allocation Algorithm

The goal of the scheduling algorithm is to decide a group of links that produce the highest aggregate throughput when they are scheduled at the same time slot. Suppose there are $N$ peer-to-peer links to be scheduled. Without loss of generality, the number of time slots $K$ is larger than or equal to $N$. Denote $S_i$ the $i$-th group of links which can transmit concurrently, and $ER_l$ the exclusive region of link $l$.

We elaborate the first algorithm, called the proportional allocation algorithm (PaA), as follows. Define $UA$ the set of links not belonging to any group in each iteration, and initially, $UA := \{1, \cdots, N\}$. The algorithm first randomly chooses a link from $UA$ (`Lines 2-4`). Then, the link which does not conflict with any links within $S_i$ is added such that all interferers are outside the exclusive region of each other. The above procedure is repeated until $UA$ is empty (`Lines 5-12`). Now all links have been grouped into $k$ sets. The algorithm then proportionally allocates $K \cdot |S_i| / \sum_{x=1}^{k} |S_x|$ slots to the $i$-th link group, for $1 \leq i \leq k$, where $|S_i|$ is the number of links in the $i$-th group (`Lines 17-19`). The computational complexity of PaA is $O(N^2)$.

It can be observed that in PaA, (i) each link belongs to at least one group, and is assigned at least one slot; (ii) in each group, all links do not conflict with each others' exclusive regions; (iii) the time slots allocated to each group are proportional to the number of links that can transmit concurrently in that group. The rationale behind PaA is that, the multi-user wireless network can be modeled as a conflict graph [34]. Then the maximum network throughput is achieved when the group of

**Algorithm 3.1** Proportional Allocation Algorithm

Initialization: $i := 1$; $S_i := \emptyset; UA := \{1, \ldots, N\}$
**repeat**
  **for** a flow $f$ randomly chosen from $UA$ **do**
    $S_i \leftarrow S_i \cup \{f\}$
    $UA \leftarrow UA \backslash \{f\}$
    **for** any flow $f'$ other than $f$ **do**
      **if** $f' \notin ER_l \quad \& \quad l \notin ER_{f'} \quad \forall\, l \in S_i$ **then**
        $S_i \leftarrow S_i \cup \{f'\}$
      **end if**
      **if** $f' \in UA$ **then**
        $UA \leftarrow UA \backslash \{f'\}$
      **end if**
    **end for**
  **end for**
  $i \leftarrow i + 1$
**until** $(UA := \emptyset) \vee (i > K)$
$k \leftarrow i$
**for** $i = 1$ to $k$ **do**
  allocate $K \cdot |S_i| / \sum_{j=1}^{k} |S_j|$ to $S_i$
**end for**
**Output:** $S_i$

links scheduled in each time slot is the maximum weighted independent set (MWIS) on the conflict graph, where the vertex weight is identical to the link throughput. In order to maximize the aggregate throughput, PaA allocates more slots to the link group with larger size, which is hypothetically close to the MWIS.

### 3.4.2 Repeating Allocation Algorithm

The second algorithm is called repeating allocation algorithm (RaA). Let $\phi_l$ be the number of slots being allocated to link $l$. A slot $i$ is assigned to a group of flows $S_i$ according to the following rule. The algorithm starts by randomly picking a link with

**Algorithm 3.2** Repeating Allocation Algorithm

```
1: Initialization: φ_l = 0; S_i := ∅
2: for i = 1 to K do
3:    f* ← arg min_l{φ_l}
4:    S_i ← S_i ∪ {l*}
5:    φ_{l*} ← φ_{l*} + 1
6:    for any flow other than l* do
7:       if l ∉ ER_k  &  k ∉ ER_l  ∀ k ∈ S_i then
8:          S_i ← S_i ∪ {l}
9:          φ_l ← φ_l + 1
10:      end if
11:   end for
12: end for
Output: S_i
```

minimal $\phi_l$, and adds it to $S_i$ (`Lines 2-4`). Then, other links which do not conflict with any link in $S_i$ (`Lines 5-10`) are added. This procedure repeats for all time slots. The RaA has a higher computational complexity of $O(KN^2 \log N)$, but with better fairness support in terms of the number of slots assigned to each link, since RaA essentially follows the max-min fairness discipline.

## 3.5 Numerical Results and Discussion

This section evaluates the performance of the proposed scheduling algorithms via simulations. A network with 40 peer-to-peer links uniformly distributed in a $10 \times 10$ $m^2$ area is simulated. The link distance is at least 1 m. The physical parameters used in the simulations are given in Table 3.1. All links use an equal transmission power $p_t$, and have the same background noise power $p_N$. We vary the value of MUI factor $b$ to reflect the impact of system parameters such as the characteristics of the pulse shape of transmitted signals, propagation conditions, cross-correlation of the target

signal and interfering signal to the scheduling performance. Each simulation scenario is repeated 10 times with different random seeds.

Table 3.1: Simulation parameters

| | |
|---|---|
| Bandwidth ($BW$) | 1 GHz |
| Center frequency ($f_c$) | 5.092 GHz |
| Transmitting power ($p_t$) | 0.0397 mW |
| Noise power ($p_N$) | $3.9811 \times 10^{-9}$ mW |
| Shadowing parameter ($\sigma_G$) | 4.3 |
| Path-loss exponent ($\gamma$) | 4 |
| Nakagami factor ($m$) | $1 \sim 6$ |
| MUI factor ($b$) | $5 \times 10^{-3} \sim 5 \times 10^{-2}$ |

### 3.5.1 Experimental UWB Channel Model in Simulations

In simulations, the network topology is fixed during each scheduling cycle. As indicated in [35], a single slope model for path-loss at distance $d$ is given by

$$PL[\text{dB}] = PL(d_0)[\text{dB}] + 10\gamma \log_{10}(\frac{d}{d_0}), \tag{3.12}$$

where $\gamma$ is the path-loss exponent which depends on the environment, $d_0 = 1$ m is the reference distance. $PL(d_0)$ is estimated using the Friis free-space equation given by

$$PL_{FS}(d_0)[\text{dB}] = 10 \cdot \log_{10}\left(\frac{G_{tx}G_{rx}\lambda^2}{(4\pi)^2 d_0^2 L}\right), \tag{3.13}$$

where $\lambda$ is the wavelength corresponding to the frequency $f_c$ approximated by the geometric mean of the lower and upper band edge frequencies, $L$ is the system loss factor, $G_{tx}$ and $G_{rx}$ denote the transmitter and receiver antenna gains, respectively. In simulations, the path-loss associated with each flow is generated in advance according to (3.12) and (3.13), where $L = G_{tx} = G_{rx} = 1$. Denote $G$ the total fading power across all paths due to the shadowing. Then, the dB value of $G$ follows a normal

distribution with mean value given from (3.12)

$$G \sim \mathcal{N}(-PL, \sigma_G^2). \tag{3.14}$$

The long-term SNR is

$$\mathrm{SNR}_{LT}[\mathrm{dB}] = P_t - G(d) - P_N, \tag{3.15}$$

where $P_N$ is the noise power. As to small-scale fading, it has been shown that the UWB fading amplitude can be well fitted by the Nakagami distribution [33]. The Nakagami fading gain, denoted by $g$, can be derived from a Ricean random variable using the conversion equations

$$K = \frac{\sqrt{m^2 - m}}{m - \sqrt{m^2 - m}},$$
$$m = \frac{(K+1)^2}{(2K+1)},$$

where $m > 1$ and $K > 0$ are the Nakagami and Rice factors, respectively. The corresponding Ricean random variable is obtained from a complex Gaussian random variable with mean $\sqrt{K/(K+1)}$ and variance $1/(2(K+1))$, where $K \geq 0$. Finally, the SNR of the received signal can be expressed as

$$\mathrm{SNR}\ [\mathrm{dB}] = 20 \cdot \log_{10} g + \mathrm{SNR}_{LT}. \tag{3.16}$$

Notice that the total received energy is spread over multiple paths with different weights following a lognormal distribution. The clustering phenomenon of UWB multipath channel is not included in the simulations since the mean energy given by (3.14) is used to represent the average received energy.

### 3.5.2 Scheduling Performance

This subsection compares the performance of the proposed scheduling algorithms, namely the PaA and RaA, and that of TDMA scheduling under different propagation

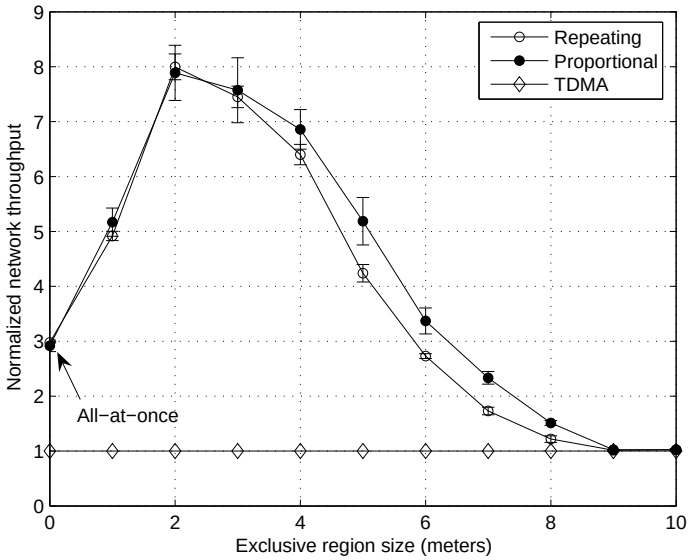



Figure 3.3: Comparisons of network throughput. The error bars represent 95% confidence interval.

environments characterized by Nakagami-$m$ factor, and interference level characterized by MUI factor $b$. Two performance measures are considered: network throughput and fairness index. Given a particular network size, the network throughput measures the aggregate data rate achieved by the scheduling algorithm. To demonstrate the performance gain, all results are normalized to the value obtained by TDMA scheduling. For fairness, the widely accepted Jain's fairness index[5] is used to evaluate the scheduling algorithms.

### Impact of Exclusive Region Size

The normalized network throughput versus the radius of exclusive region is shown in Figure 3.3, with Nakagami-$m = 4$ and MUI factor $b = 10^{-2}$. The 95% confidence interval is also plotted as an error bar for each proposed algorithm. An exclusive region with radius equal to zero implies all flows can transmit concurrently, and we call this the all-at-once scheme. As the exclusive region increases, less link can be allowed in concurrent transmissions. Eventually the proposed scheduling algorithms behave the same as TDMA when the exclusive size is sufficiently large. The simulation results show that both the proposed scheduling algorithms achieve similar network throughput. Under this setting, the radius of the optimal exclusive region is between $2 \sim 4$ m. It can be seen that by properly selecting the exclusive region, the proposed PaA and RaA can achieve network throughput 768% and 830% times higher than TDMA. The proposed algorithms also provide 280% gain compared to the all-at-once scheme.

### Fairness

Maintaining fair resource allocation among users is always an import subject. For the scheduling problem under investigation, the fairness can be evaluated in terms of the number of slots assigned to each link, and the throughput achieved by each link. Figure 3.4 compares the Jain's fairness index in terms of the number of slots assigned to each flow. The resultant 95% confidence interval corresponding to each point on the $x$-axis is also plotted. TDMA scheduling provides satisfactory fairness as it treats all links equally. Without considering different link qualities, this strategy, however, does not efficiently utilize the spectrum. As to the proposed scheduling algorithms,

[5]Denote $x_i$ as the share of the resource allocated to user $i$, the fairness index given in [36] can be calculated as $\frac{(\sum_i x_i)^2}{n \cdot \sum_i x_i^2}$, which ranges from $1/n$ (worst case) to 1 (best case).

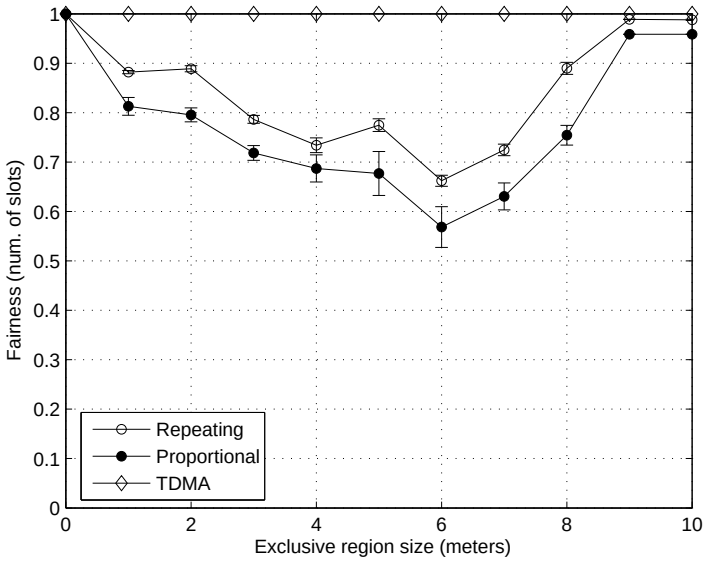



Figure 3.4: Comparisons of fairness w.r.t. number of slots per link. The error bars represent 95% confidence interval.

the general tradeoff between throughput maximization and fairness can be observed by comparing Figure 3.3 and Figure 3.4. The PaA shows inferior support in fair slot sharing than RaA because of the proportional rule for slot allocation.

From the viewpoint of users, it is more proper to consider fairness as the amount of throughput that one can obtain with respect to others. In Figure 3.5, all scheduling algorithms reveal poor fairness support in terms of throughput. Indeed, with the objective of increasing network throughput, the proposed algorithms favor the links with shorter communication distance and less peer interferers. For TDMA scheduling, the achievable throughput of each link is dominated by the link quality. When each user is assigned equal number of slots, throughput fairness can not be achieved. How to make better tradeoff between efficiency and fairness should deserve further study. An intuitive solution is to define a unified cost function which tradeoffs the two

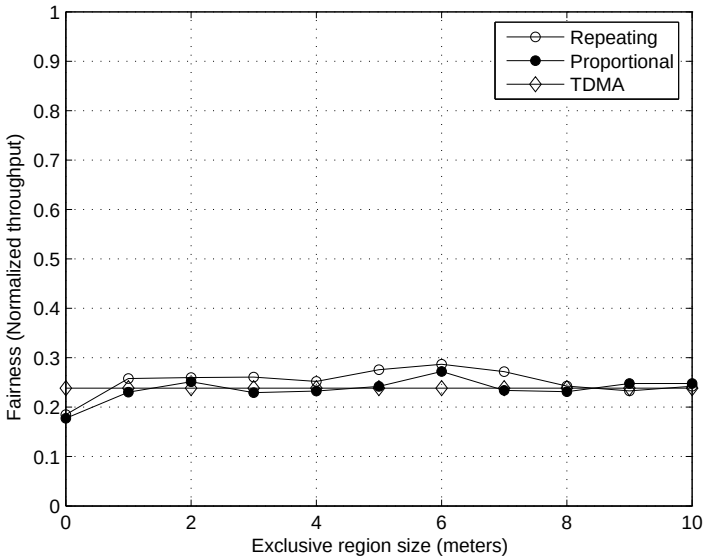



Figure 3.5: Comparisons of fairness w.r.t. normalized link throughput.

objectives.

Another important performance measure is the minimum throughput over all links provided by each algorithm. A greedy scheduling algorithm aiming to maximize the throughput may favor certain links and sacrifice others. In Figure 3.6, it can be seen that the proposed algorithms with an appropriate exclusive region can largely increase the minimal throughput among all competing flows, and thus more satisfactory services can be provided.

### Impact of MUI

An important factor on determining the throughput gain of concurrent transmissions lies in the effectiveness of compressing the MUI. The lower the MUI, the higher the achievable throughput of individual links in the concurrent transmission group. The effectiveness of MUI depends on the system parameters such as the monocycle shape

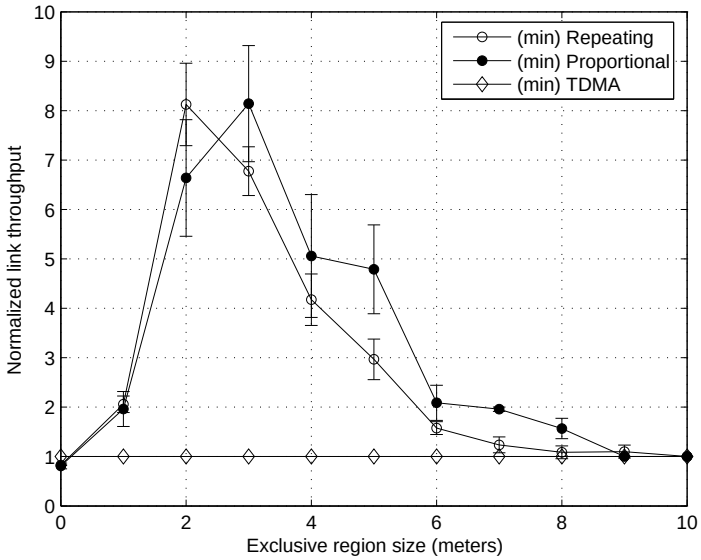



Figure 3.6: Comparisons of the minimum link throughput achieved by different algorithms. The error bars represent 95% confidence interval.

in DS-UWB systems. It has been reported that using a higher-order monocycle can increase the robustness against MUI given the fixed pulse width [37], while the complexity and susceptibility to timing jitter in the receiver are also increased. To study the impact of MUI, the effectiveness of MUI compression is represented by a MUI factor $b$. Generally perfect orthogonality (*i.e.*, $b = 0$) is impossible, even pseudo-random codes are used because of the multipath effect and when the spreading codes are generated locally by the communication peer. Figure 3.7 shows the impact of MUI with Nakagami-$m$ being 4. When $b$ is sufficiently small (*e.g.*, $b < 10^{-2}$) in Figure 3.7, the MUI is significantly reduced such that the all-at-once scheme (*i.e.*, infinitely small exclusive region) yields higher network throughput than that of TDMA. Nevertheless, a UWB system is limited to certain capability of MUI suppression. Therefore the simple all-at-once rule may be worse than TDMA, for instance, with $b = 5 \times 10^{-2}$.

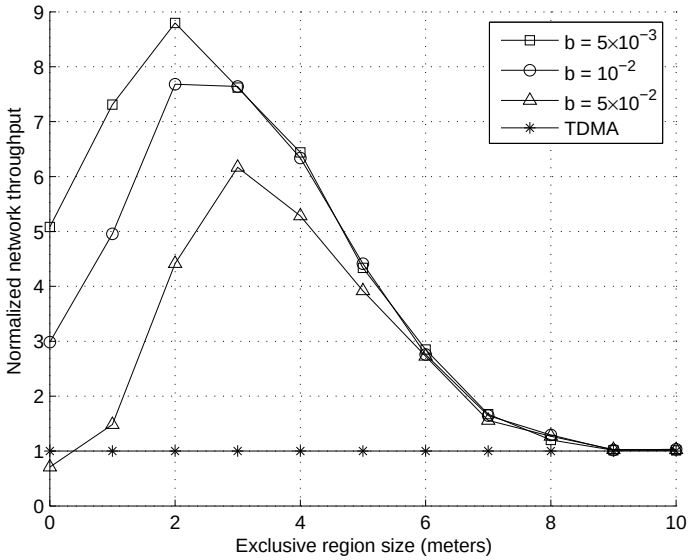



Figure 3.7: Normalized network throughput *vs.* radius of exclusive region in different MUI factor $b$. .

On the other hand, the proposed scheduling algorithms can guarantee higher network throughput than TDMA by deploying the proper exclusive region size. The optimal size of exclusive region leading to the maximum network throughput shown in Figure 3.7 is between $2 \sim 4$ meters. Finally, we investigate the relationship between exclusive region radius and the MUI factor. With less protection against MUI (*i.e.*, larger $b$), a larger exclusive region radius is required to achieve the maximum network throughput. Thus the optimal exclusive region radius can also be seen as a function of MUI factor, as shown in the asymptotic analysis (Section 3.4).

### Impact of Nakagami Fading

The impact of Nakagami fading parameter $m$ is shown in Figure 3.8. The value $m$ varies from 1 to 6, according to the measurement results suggested in [35]. For the

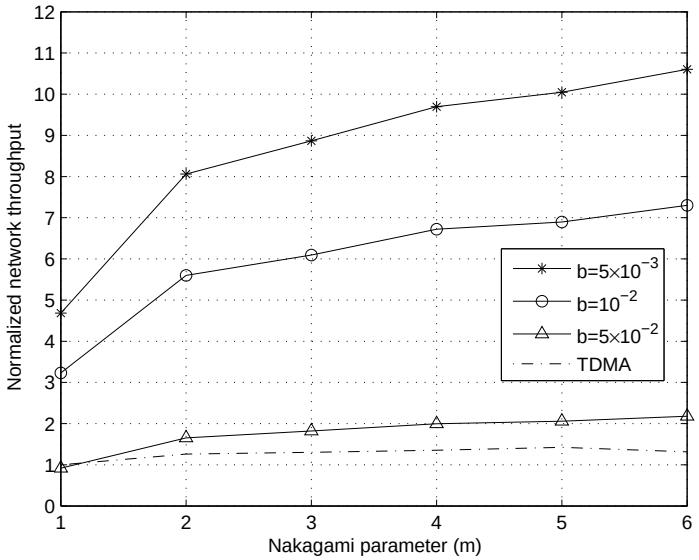



Figure 3.8: Nakagami parameter $m$ *vs.* normalized network throughput with different MUI factor $b$.

sake of brevity, only the network throughput achieved by the RaA in the exclusive region radius of 2 m is shown. Generally, the Nakagami parameter $m$ characterizes the severity of fading condition. A larger value of $m$ represents less fading and a stronger line-of-sign path. Therefore, the network throughput is an increasing function with respect to $m$.

### 3.5.3 Remarks on Implementation Issues

As mentioned in Section 3.4, algorithm efficiency is an important requirement toward practical implementation. In our simulations, each scheduling algorithm is coded in C language and executed on a Pentium-4 2.8 GHz CPU. The execution time for RaA is 0.58 ms and 0.11 ms for PaA on average when there are 40 link to be scheduled in 80

slots. The PaA has lower complexity and thus less execution time as expected. Both algorithms can be performed within a superframe time, which is 65 ms specified in the IEEE 802.15.3 standard. The computation overhead can be reduced by altering the scheduling cycle, for example, the PNC can perform the scheduling algorithm by several superframes for moderate traffic density.

## 3.6 Summary

Without explicit considerations of the physical properties of UWB, existing MAC protocols remain considerable space to improvement. This chapter attempts to bridge the gap between the physical characteristics of UWB communications and the MAC protocol design. The main accomplishments of this chapter are summarized as follows:

- We exploited the physical property of UWB and showed that the network throughput can be greatly improved by properly allowing concurrent transmission. The sufficient condition ensuring that concurrent transmission yields higher network throughput than TDMA is derived.

- We presented low-complexity scheduling algorithms aiming at improving network throughput for UWB networks.

- We comprehensively evaluated the proposed algorithms through simulations. Numerical results show that our solutions can increase network throughput by 800% in comparison to the standard solution using TDMA with decent fairness.

We have made some simplified assumptions such as UWB propagation model and quasi-static network topology for analytical tractability. We further assume that nodes are uniformly distributed in the network. The network with uneven node density, *e.g.*, a cluster network, should be more realistic and will be considered in the

future work. It is also noticed that the impact of inter-piconet interference is not considered. This problem has been discussed in Bluetooth-based networks, and recently being studied for UWB-based networks [38,39]. How to extend our work taking the possible inter-piconet interference into account should deserve further study. In addition, we focus on improving the network throughput in this chapter. On the other hand, users may launch different applications with heterogeneous bandwidth demands. For this scenario, the resource allocation strategy developed in this chapter may benefit certain users while starve others at the same time. In next chapter, we address the problem of providing QoS for UWB networks with heterogeneous bandwidth requirements.

Table 3.2: List of local notations for Chapter 3

| **Notation** | **Definition (unit)** |
|---|---|
| $p_t$ | transmitting power (mW) |
| $p_r$ | receiving power (mW) |
| $p_N$ | noise power (mW) |
| $\sigma_G$ | shadowing parameter |
| $\alpha$ | processing |
| $\gamma$ | Pathloss exponent |
| $m$ | Nakagami factor |
| $b$ | MUI factor |

# Chapter 4

# Multi-Class QoS Support for UWB Networks

## 4.1 Introduction

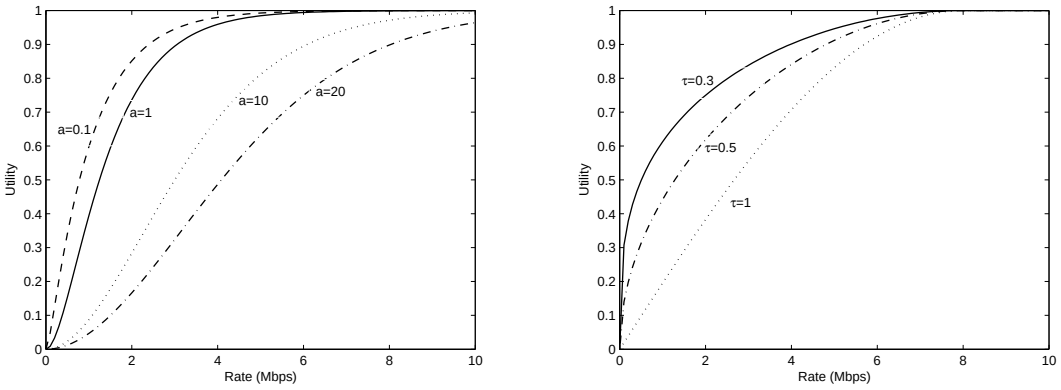
In Chapter 3, we have demonstrated that the network throughput can be greatly increased by strategically scheduling concurrent transmissions via the notion of "exclusive region". Besides throughput maximization, providing QoS for heterogeneous users is another important issue. Users may launch various applications such as real-time video/audio streaming and bulky file transfer. In this context, the scheduling algorithm that maximizes the network throughput cannot guarantee the diverse bandwidth requirement of users. Furthermore, equal share of resource is not necessary to be fair from the user viewpoint, as the same amount of bandwidth is consciously different to different applications. For instance, real-time video applications generally need a certain minimum bandwidth to initiate the application while a higher bandwidth can help to improve the display quality with certain extend. Voice traffic also

needs a strict minimum bandwidth to establish a connection, but providing additional bandwidth to the requirement does not improve the voice quality.

To satisfy different bandwidth demands of heterogeneous applications, it is important that the resource allocation mechanism takes traffic characteristics into account to achieve efficient resource utilization. For this sake, the notion of utility function is useful to quantitatively characterize the user satisfaction level in response to the amount of assigned bandwidth. In this chapter, we jointly consider the physical property of UWB communications and the user satisfaction levels to bandwidth, and formulate the resource allocation problem as a utility optimization problem. The objective is to maximize the total utility under the constraint of fairness among heterogeneous users. This chapter is organized as follows. In Section 4.2, we briefly introduce the utility functions for characterizing different traffic types. The we present the problem formulation in Section 4.3. The proposed algorithms for solving the utility maximization problem are presented in Section 4.4, followed by numerical results in Section 4.5. We summary this chapter in Section 4.6.

## 4.2 Utility Functions for Heterogeneous Traffic

Utility is defined as the satisfaction level of a user with respect to the amount of allocated bandwidth. Various utility functions have been proposed to characterize different traffic types in the Internet [40, 41] and cellular networks [42–45], respectively. To deal with heterogeneous traffic, traffic types are classified into three classes. Class-1 includes the applications that need a fixed amount of bandwidth continuously available for the lifetime of the connection (*e.g.*, CBR voice applications), which is




(a) (b)

Figure 4.1: Illustrative utility functions for Class 2 and Class 3 traffic. (a) Utility functions for Class 2 ($b = 1$). (b) Utility functions for Class 3 ($r_{\max} = 8$Mbps).

characterized by a step function:

$$U(R) = \begin{cases} 1 & \text{if } R \geq r_{\min}, \\ 0 & \text{if } R < r_{\min}, \end{cases} \tag{4.1}$$

where $r_{\min}$ is the minimum bandwidth requirement for a connection. Class-2 applications can adapt to the allocated bandwidth to certain extent (*e.g.*, compressed video streams) described by the following sigmoidal-like utility function:

$$U(R) = 1 - e^{-\frac{bR^2}{a+R}} \tag{4.2}$$

where the parameters $a$ and $b$ can be adjusted to determine the shape of $U(R)$, as shown in Figure 4.1(a). Class-3 applications are most flexible to the available bandwidth. Typical transmission control protocol (TCP) data applications belong to this class. Since there is no minimum rate requirement for such traffic class, the

utility can be modeled by the following function:

$$U(R) = \begin{cases} 1 & \text{if } R \geq r_{\max}, \\ \sin^{\tau}\left(\frac{\pi}{2} \cdot \frac{R}{r_{\max}}\right) & \text{if } R < r_{\max}, \end{cases} \tag{4.3}$$

where parameter $\tau$ controls the shape of $U(R)$, as shown in Fig. 4.1(b). Extensive discussions on different utility functions can be found in [40]. Generally, the choice of utility functions would affect the efficiency of resource allocation, and introduce different degrees of complexity to the utility maximization problem [41]. In this chapter, utility functions are continuous and non-decreasing functions with values within $[0, 1]$.

## 4.3 Optimal Scheduling Formulation

To increase the total throughput, it is desirable to have concurrent transmissions if condition (3.6) holds. Therefore, the objective of scheduling is to select the optimal link set for each time slot such that the total user utilities are maximized. Taking the MUI of concurrent transmissions into account, the network can be modeled as a conflict graph $G = (V, E)$, where the vertex set $V$ represents the set of flows requesting for transmission and $E$ is the set of edges. Two vertices in $G$ are connected if they are not allowed to transmit in the same time slot according to the exclusive-region condition (3.6). The connected vertices (*i.e.*, links) are called *neighbors*. Let $N(i)$ represent the neighbors of a vertex $i \in V$. Define the following binary variables: $x_i^s = 1$ if link $i$ is allocated in slot $s$; otherwise $x_i^s = 0$. $y_{ij}^s = 1$ if node $j$ is the recipient of node $i$ in slot $s$; otherwise $y_{ij}^s = 0$. The scheduling problem for a single

slot can be formulated as a nonlinear mixed-integer optimization problem[6]:

$$(\mathcal{P}) \quad \max \sum_{i=1}^{|V|} x_i^s U(R_i) \tag{4.4}$$

$$\text{s.t.} \quad x_i^s \sum_{j \in N(i)} x_j^s = 0, \quad \forall\ i = 1, ..., |V| \tag{4.5}$$

$$\sum_{i \in \mathcal{T}} y_{ij}^s + \sum_{k \in \mathcal{R}} y_{jk}^s \leq 1,\ \forall j \in \mathcal{R} \tag{4.6}$$

$$R_i = k' p_t d_i^{-\gamma} / (\eta + x_j^s \alpha p_t \sum_{j=1, j \neq i}^{|V|} d_{ji}^{-\gamma}), \quad i = 1, \ldots, |V| \tag{4.7}$$

The optimization problem $(\mathcal{P})$ can be interpreted as follows. For slot $s$, the link set, $\{x_i^s : i = 1, \ldots, |V|\}$ that maximizes the total utilities defined in (4.4) under the constraints of (4.5)-(4.7) is the optimum scheduling. Constraint (4.5) ensures that any link scheduled in slot $s$ is conflict-free. Constraint (4.6) restricts that a node can communicate with at most one node in a slot, where $\mathcal{T}$ represents the set of senders being scheduled in the same slot, and $\mathcal{R}$ is the set of corresponding receivers. Constraint (4.7) specifies the achievable throughput in the presence of MUI. Note that $(\mathcal{P})$ can be extended as a superframe utility optimization problem by modifying (4.4) as $\max \sum_{s=1}^{S} \sum_{i=1}^{|V|} x_i^s U(R_i)$, where $S$ is the total number of slots in a superframe.

Problem $(\mathcal{P})$ can be reduced to the maximum weighted independent set (MWIS)[7] problem as follows. First of all, we incorporate the constraint (4.6) into the graph $G$ by adding an edge between two nodes in $V$ if the corresponding links have common transmitters or receivers, and denote the resultant graph as $G'$. Let $\mathcal{K}$ denote the collection of independent sets in $G'$. Since the superset $\mathcal{K}$ satisfies the constraints

[6]Problem $(\mathcal{P})$ is NLIP because $U(R_i)$ is nonlinear, $x_i$ is integer, and $R_i$ is continuous.

[7]An independent set in a graph is the set of vertices such that no two vertices in the independent set share the same edge. Associating each vertex a weight, the independent set with the maximum total weights is the maximum weighted independent set.

(4.5) and (4.6), finding the optimal solution of problem ($\mathcal{P}$) is equivalent to finding the set with maximum weight in $\mathcal{K}$, denoted by $\kappa^* \in \mathcal{K}$ in $G'$. In other words,

$$\kappa^* = \arg\max_{\kappa \in \mathcal{K}} \sum_{i=1}^{|\kappa|} U(R_i), \tag{4.8}$$

where $U(R_i)$ corresponds to the weight of vertex $i \in V$ and the slot index $s$ has been dropped such that $\kappa^*$ represents the optimal scheduling for one slot. As MWIS is known to be NP-hard, there is no polynomial-time algorithm to solve ($\mathcal{P}$).

We relax the assumption of perfect distance information by discrete stochastic approximation as follows. Let $\tilde{U}(\cdot)$ denote the noisy version of $U(\cdot)$, *i.e.*, $\tilde{U}(\cdot)$ contains errors due to inaccurate distance estimation. Rewrite (4.8) as

$$\kappa^* = \arg\max_{\kappa \in \mathcal{K}} \tilde{U}(\kappa), \tag{4.9}$$

where $\tilde{U}(\kappa) = \sum_{i=1}^{|\kappa|} \tilde{U}(R_i)$. To mitigate the large deviation of the estimation errors, the above noisy objective function $\tilde{U}(\kappa)$ is approximated by $\mathbb{E}[\tilde{U}_m(\kappa)]$, where $\mathbb{E}[\cdot]$ is the expectation operator, and $\{\tilde{U}_m(\kappa), m = 1, 2, \ldots\}$ represent the sequence of noisy utilities associated with set $\kappa \in \mathcal{K}$. $\tilde{U}_m(\kappa)$ can be obtained from different distance estimations. For example, each communication link can perform multiple ranging procedures and reports the result to the PNC.

The approximated optimization is stated as

$$\kappa^* = \arg\max_{\kappa \in \mathcal{K}} \sum_{i=1}^{|\kappa|} \tilde{U}(R_i) \approx \arg\max_{\kappa \in \mathcal{K}} \mathbb{E}[\tilde{U}(\kappa)]. \tag{4.10}$$

For the unimodal objective function[8], optimization techniques such as golden-section search or gradient-based approaches may be used to find the maxima of the nonlinear

[8] A function $f(x)$ is unimodal if for some value $m$, it is monotonically increasing (decreasing) for $x \leq m$ and monotonically decreasing (increasing) for $x \geq m$. The maximum (minimum) value of $f(x)$ is thus $f(m)$ and there is no other local maxima (minima).

function $\mathbb{E}[\tilde{U}(\kappa)]$. In our consideration of heterogeneous traffic classes, however, the optimization objective function is not necessary to be unimodal. Besides, deriving the distribution of $\tilde{U}(\kappa)$ is very difficult if not impossible, since $\tilde{U}(\kappa)$ is combinatorial, dependent on the element in $\kappa$. In situations where the objective function is difficult to derive analytically, discrete approximation is an applicable technique to solve the optimization problems with uncertainties. Let $\{\tilde{U}_m(\kappa), m = 1, 2, \ldots\}$ represent the sequence of noisy utilities associated with set $\kappa \in \mathcal{K}$, where $\tilde{U}_m(\kappa)$ is obtained from different distance estimations. The random sequence $\{\tilde{U}_m(\kappa)\}$ is assumed to be independently and identically distributed (i.i.d.) with finite mean and variance. By the strong law of large numbers (SLLN)[9], the sample mean of $\{\tilde{U}(\kappa)\}$, denoted by $\overline{U}(\kappa) = 1/M \sum_{m=1}^{M} \tilde{U}_m(\kappa)$, converges almost surely (a.s.) to $\mathbb{E}[\tilde{U}(\kappa)]$. Together with the finiteness of the set $\mathcal{K}$, it is implied that

$$\arg\max_{\kappa \in \mathcal{K}} \overline{U}(\kappa) \rightarrow \arg\max_{\kappa \in \mathcal{K}} \mathbb{E}[\tilde{U}(\kappa)] \quad a.s.$$

Therefore, instead of using one biased utility value to solve the optimization problem $(\mathcal{P})$, we take the series $\{\tilde{U}_m(\kappa)\}$ in approximating the noisy objective function to avoid trapping into a local optimum [46].

The above stochastic optimization problem can be solved by several metaheuristics, such as simulated annealing, Tabu search, genetic algorithm, and global search algorithm (GSA) [10]. Different from direct search, these algorithms have been proposed to increase the search efficiency by defining "move sets", assisting the algorithm in moving from one solution to superior ones. Among these algorithms, we select GSA as the base to solve (4.11) since its convergence to a global optimum can be theoretically proved under asymptotic conditions. Next section presents the improved

[9]There are different forms of SLLN, *e.g.*, Kolmogorov's SLLN, the one for Markov chains, the one for non i.i.d. variable, the one for dependent variables, and so on. Here we use the SLLN that is defined based on the i.i.d. random variables with finite mean and variance.

GSA.

## 4.4 Algorithms for Solving Utility Maximization Problem

In this section, we present two algorithms for solving the optimization problem (4.8). We propose a metaheuristic method called exclusive-region based global search algorithm (ER-GSA) to efficiently find $\kappa^*$ in (4.11) for each time slot and prove its convergence. We also present an algorithm for updating the utility value in each algorithm iteration.

Consider the fact that we have a finite discrete solution space $\mathcal{K}$. Each element $\kappa$ in $\mathcal{K}$ represents a flow set, and the goal is to find the optimal $\kappa^*$ that achieves the maximum aggregate utility. With the noisy objective function involved, the basic idea of GSA [10] is to select the search direction taking into account the previous search trajectory. If a point in the solution space has been determined by the algorithm as a better solution than the one being visited previously, it has a higher chance to be the optimum. Such a strategy reduces the risk of staying on a false maximum. If we consider the search trajectory as a random sequence, this random sequence generated during the algorithm iterations can be modeled a Markov chain, where each state represents a point (equivalently, a flow set $\kappa$) in the solution space that has been visited by the algorithm. In each iteration, the transition of the Markov chain is determined by comparing the objective value of the current state with that of a randomly chosen point from the solution space.

We apply the basic principle of GSA and present the improved algorithm called ER-GSA. We use the following notations in the algorithm. At the $m$-th iteration, $\kappa_m$ is the current state ($\kappa_m \in \mathcal{K}$), $W_m(\kappa)$ is the number of times the algorithm has

**Algorithm 4.1** ER-GSA

**Input:** $\mathcal{K}$

**Step 1.** Randomly select an initial user subset $\kappa_0 \in \mathcal{K}$ and let $\kappa_0^* = \kappa_0$. Set $W_0(\kappa_0) = 1$ and $W_0(\kappa) = 0$ for all $\kappa \in \mathcal{K} \backslash \{\kappa_0\}$. Calculate $\overline{U}_0(\kappa_0)$. Let $m = 0$, and go to step 2.

**Step 2.** Randomly select another user subset $\kappa_m^{'} \in \mathcal{K} \backslash \{\kappa_m\}$. Compute the corresponding $\overline{U}_m(\kappa_m^{'})$ using the variable-sample method. Go to step 3.

**Step 3.** If $\overline{U}_m(\kappa_m) > \overline{U}_m(\kappa_m^{'})$, let $\kappa_{m+1} = \kappa_m$, and go to step 5. Otherwise, go to step 4.

**Step 4.** Sort $\overline{U}_m(\kappa_m^{'})$ in descending order. Denote $S_i$ the $i$-th flow in the sorted set, and $S^{'}$ an empty set.

**for** $i = 1$ to $|\kappa_m^{'}|$ **do**
  **if** $S_i \notin ER_l \quad \& \quad l \notin ER_{S_i} \quad \forall\, l \in S^{'}$ **then**
    $S^{'} = S^{'} \cup \{S_i\}$
  **end if**
**end for**

If $\overline{U}_m(S^{'}) > \overline{U}_m(\kappa_m^{'})$, let $\kappa_m^{'} = S^{'}$; otherwise $\kappa_m^{'}$ remains unchanged. Update $\kappa_{m+1} = \kappa_m^{'}$ and go to step 5.

**Step 5.** Let $m = m + 1$, $W_m(\kappa_m) = W_{m-1}(\kappa_m) + 1$, and $W_m(\kappa) = W_{m-1}(\kappa)$ for all $\kappa \in \mathcal{K} \backslash \{\kappa_m\}$. If $W_m(\kappa_m) > W_m(\kappa_{m-1}^*)$, then let $\kappa_m^* = \kappa_m$. Otherwise, let $\kappa_m^* = \kappa_{m-1}^*$. Go to step 2.

**Output:** $\kappa_m^*$

visited state $\kappa$, and $\kappa_m^*$ is the state being visited by the algorithm most frequently up to the $m$-th iteration. ER-GSA starts by randomly selecting an initial flow subset $\kappa_0$. During each iteration $m$, a new subset $\kappa_m^{'}$ is randomly selected (step 2), and the sample mean $\overline{U}_m(\kappa_m^{'})$ is computed. Different from GSA where a fixed sample size is used, $\overline{U}_m(\kappa_m^{'})$ is calculated according to a *variable-sample mean method*, with the number of samples equal to $|\kappa_m^{'}|$, *i.e.*, $\overline{U}_m(\kappa_m^{'}) = \frac{1}{|\kappa_m^{'}|} \sum_{i=1}^{|\kappa_m^{'}|} \tilde{U}_m(\kappa_i), \kappa_i \in \kappa_m^{'}$. As will be shown in Remark 2, the use of variable sample size in calculating the sample mean ensures the algorithm convergence.

The algorithm then decides whether to update $\kappa_m^*$ by comparing the sample mean $\overline{U}_m(\kappa_m^{'})$ with that of the current point $\kappa_m$ (step 3). If the current state is superior than the newly selected one, *i.e.*, $\overline{U}_m(\kappa_m) > \overline{U}_m(\kappa_m^{'})$, the algorithm proceeds to step

`5`, updating the best subset $\kappa_m^*$ according to $W_m(\kappa)$.

Before updating the counter $W_m(\kappa)$, we notice that the algorithm convergence depends on how fast the optimum can be located. The more frequent the optimum can be hit by the algorithm, the faster the algorithm can converge to the global optimum. To increase the chance of selecting the optimum flow set $\kappa^*$, we use a local enhancement based on the concept of exclusive region that refines the chosen subset such that the optimal subset can be located faster. In `step 4`, we first sort $\overline{U}_m(\kappa_m^{'})$ in descending order. Then we remove those flows violating the condition of concurrent transmissions defined in (3.6). This is performed by the for-loop in `step 4`, where the set $ER_l$ contains those flows within the exclusive region of flow $l$, and $S_i$ denotes the $i$-th flow in the sorted set. The conditions ($S_i \notin ER_l$) and ($l \notin ER_{S_i}$) imply that flows $S_i$ and $l$ are allowed to transmit concurrently, according to (3.6). The resulting set, $S^{'}$, is compared with the randomly chosen set $\kappa_m^{'}$ at the current iteration. If $S^{'}$ generates a higher aggregate utility, it will replace $\kappa_m^{'}$. At the end of one iteration, the algorithm updates the counter $W_m(\kappa_m)$. If $W_m(\kappa_m) > W_m(\kappa_{m-1}^*)$, the algorithm will update the best solution $\kappa_m^*$ by $\kappa_m$. By tracking the counter $W_m(\kappa)$, the algorithm can further avoid falsely retaining on a local optimum due to the biased objective function value.

### 4.4.1 Utility update

The utility-maximization formulation can be further modified such that the scheduling decision satisfies the fairness criteria in the long-term. We employ a simple rule based on weighted fair queuing (WFQ) [47], which normalizes the instantaneous utility of flow $i$ in slot $s$, denoted by $U_i(s)$, to the total utility this flow has obtained, $\sum_{t=1}^{s-1} U_i(t)$. In addition, users can be discriminated based on the cost of bandwidth usage. To

this end, a control parameter denoted as $\rho_i^{(s)}$ for flow $i$ in slot $s$ is define as:

$$\rho_i^{(s)} := \frac{c_i}{(\sum_{t=1}^{s-1} U_i(t) + \epsilon)^s}, \tag{4.11}$$

where $c_i$ is a predefined parameter representing the revenue contribution or importance of a particular user, and $\epsilon > 0$ is a small nominal constant to avoid zero denominator. Consequently, the weighted utility for flow $i$ in slot $s$ is given by

$$U(R_i) = \rho_i^{(s)} U_i(s), \tag{4.12}$$

where $U_i(s)$ is a function of $R_i$ and traffic class corresponding to each flow. (4.12) has the following properties: 1) the weighted utility being an exponential function of slot index $s$ ensures the flows with less sum utility a higher priority; 2) a flow is opportunistically scheduled if it has higher utility value in the current slot than others; 3) different level of protections to traffic classes can be achieved by adjusting the parameter $c_i$.

### 4.4.2 Convergence of ER-GSA

The convergence of ER-GSA is based on Theorem 2.1 in [10], which proves that the sequence $\{\kappa_m^*\}$ converges almost surely to an element of $\mathcal{S}$.

**Theorem 4.1** (Convergence of GSA [10, Theorem 2.1])**.** *Consider the sequence* $\{\kappa_m, m = 1, 2, \ldots\}$ *generated by the algorithm as a Markov chain on the state space* $\mathcal{K}$*. Let* $\mathcal{S} \in \mathcal{K}$ *denote the set of global optimizers of the function* $\overline{U}(\kappa)$ *(it is likely that multiple user subsets can maximize the objective function). For each* $i, j \in \mathcal{K}$*, denote a random variable* $Y^{(i \to j)} = \overline{U}(j) - \overline{U}(i)$*. If* $Y^{(i \to j)} > 0$*, let* $j$ *be the next state. If* $Y^{(i \to j)} \leq 0$*, let* $i$ *be the next state. For all* $\kappa^* \in \mathcal{S}$*,* $\kappa \notin \mathcal{S}$*, and* $\nu \in \mathcal{K} \backslash \{\kappa, \kappa^*\}$*, if*

(C1) $\mathbb{P}\left\{Y^{(\kappa \to \kappa^*)} > 0\right\} > \mathbb{P}\left\{Y^{(\kappa^* \to \kappa)} > 0\right\}$,

(C2) $\mathbb{P}\left\{Y^{(\nu\to\kappa^*)} > 0\right\} > \mathbb{P}\left\{Y^{(\nu\to\kappa)} > 0\right\}$,

*then the sequence* $\{\kappa_m^*\}$ *converges almost surely to an element of* $\mathcal{S}$ *(That means there exisits a set* $A$ *such that* $\mathbb{P}(A) = 1$ *and for all* $w \in A$*, there exists* $M_w > 0$ *such that* $\kappa_m^*(w) \in \mathcal{S}$ *for all* $m \geq M_w$*).*

*Proof.* Based on conditions (C1) and (C2), the algorithm convergence can be proved following Theorem 2.1 in [10]. □

If conditions (C1) and (C2) hold, Algorithm 4.1 states that ER-GSA converges to the global optimum with probability one. Thus we need to verify the validness of these two conditions in ER-GSA to concluded the convergence of ER-GSA. Since the exact distribution of $\tilde{U}(\kappa)$ is difficult to obtain in practice, the following assumption is employed.

**Assumption 4.1.** $\{\tilde{U}(\kappa)\}$ *can be approximated as a sequence of Gaussian random variables with different means but the same variance.*

Based on Assumption 4.1, the following two lemmas verify (C1) and (C2), respectively.

**Lemma 4.1** (Validation of (C1))**.** *The Markov chain* $\{\kappa_m\}$ *is more likely to move from a non-optimal state to an optimal state than the reverse direction, as stated in condition* (C1)*.*

*Proof.* Based on Assumption 4.1, the distribution of $\tilde{U}(\kappa)$ can be described as $\tilde{U}(\kappa) \sim \mathcal{N}(\mu_\kappa, |\kappa|^2\sigma^2)$. By using a variable-sample method, $|\kappa|$ realizations of $\tilde{U}(\kappa)$ are taken and averaged in each iteration. Thus the distribution of $\overline{U}(\kappa)$ is given by $\overline{U}(\kappa) \sim \mathcal{N}(\mu_\kappa, \sigma^2)$. Similarly, we can write $\overline{U}(\kappa^*) \sim \mathcal{N}(\mu_{\kappa^*}, \sigma^2)$. As as result, $Y^{(\kappa\to\kappa^*)}$ and $Y^{(\kappa^*\to\kappa)}$ are Gaussian random variables with distributions $Y^{(\kappa\to\kappa^*)} \sim \mathcal{N}(\mu_{\kappa^*} - \mu_\kappa, 2\sigma_2)$,

and $Y^{(\kappa^* \to \kappa)} \sim \mathcal{N}(\mu_\kappa - \mu_{\kappa^*}, 2\sigma^2)$, respectively. Since $\kappa^* \in \mathcal{S}$ and $\kappa \notin \mathcal{S}$, $\mu_\kappa < \mu_{\kappa^*}$. Because $Y^{(\kappa \to \kappa^*)}$ and $Y^{(\kappa^* \to \kappa)}$ have the same variance and $\mu_* - \mu > 0 > \mu - \mu_*$, it follows that $\mathbb{P}\left\{Y^{(\kappa \to \kappa^*)} > 0\right\} > \mathbb{P}\left\{Y^{(\kappa^* \to \kappa)} > 0\right\}$. Thus, condition (C1) is satisfied. □

**Lemma 4.2** (Validation of (C2))**.** *If the current state of the Markov chain is not an optimal one, the transition will be more likely made toward the optimal one than any other states, as stated in condition* (C1)*.*

*Proof.* Similar to the proof of Lemma 4.1, we have $Y^{(\nu \to \kappa^*)} \sim \mathcal{N}(\mu_{\kappa^*} - \mu_\nu, 2\sigma^2)$ and $Y^{(\nu \to \kappa)} \sim \mathcal{N}(\mu_\kappa - \mu_\nu, 2\sigma^2)$. Since $\kappa^* \in \mathcal{S}$, $\kappa \notin \mathcal{S}$, and $\nu \in \mathcal{K} \backslash \{\kappa, \kappa^*\}$, we have $\mu_{\kappa^*} - \mu_\nu > \mu_\kappa - \mu_\nu$. It readily shows that $\mathbb{P}\left\{Y^{(\nu \to \kappa^*)} > 0\right\} > \mathbb{P}\left\{Y^{(\mu \to \kappa)} > 0\right\}$, which proves (C2). □

*Remark* 4.2. Condition (C2) is fairly restrictive, and it may not hold using the original GSA. In Original GSA, only one sample of $\tilde{U}(\kappa)$ is taken in each iteration. Thus the variance of $\tilde{U}(\kappa)$ depends on the number of flows in $\kappa$ that denies (C1) and (C2) in general. Since the estimated utilities of each flow are asymptotically independent Gaussian random variables with variance $\sigma^2$, we can satisfy condition (C2) using a *variable-sample method*: in each iteration (`step 2`), we take $|\kappa|$ samples and calculate the sample mean $\overline{U}(\kappa)$. Consequently, the variance of $Y^{(i \to j)}$ is independent of the number of flows in the subsets. On the other hand, (C2) is more loose, even the variance of each $\tilde{U}(\kappa)$ is not identical.

## 4.5 Performance Evaluation and Discussion

Simulations are conducted to evaluate the performance of the proposed utility based scheduling algorithm. The two heuristic algorithms presented in Chapter 3, namely

RaA and PaA, and TDMA are considered as the benchmark for performance comparison. We start by tracking the instantaneous behavior of the proposed utility-based scheduling algorithm. We then use three metrics to study the long-term performance of our proposed solutions, namely the cumulative utilities, the minimum utility among flows, and the fairness support. Finally the complexity and efficiency of the ER-GSA are evaluated.

### 4.5.1 Experimental setting

The simulated network consists of 20 nodes uniformly distributed in a square area of $10 \times 10$ m$^2$. Each sender arbitrarily chooses another node as the receiver, forming 10 peer-to-peer communication flows. The data rate of each flow is estimated as: $R = k \cdot W \log_2(1 + \text{SINR})$, where $0 < k \leq 1$ reflects the efficiency of the transceiver design, $W$=500 MHz, the UWB power spectrum density of the transmission and noise are $-41$ dBm/MHz, $-114$ dBm/MHz, respectively, and the pathloss exponent is set to 4. The distance between two nodes $\tilde{d}$ is modeled by $d = \tilde{d} + \delta$ where $d$ is the actual distance, and $\delta$ is the estimation error approximated. We model $\delta$ by a normal distributed random variable according to [48, Eq. (15)], *i.e.*, $\delta \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = 0.05$. The cross-correlation of the target signal and the interfering signals is assumed to be 0.1. Table 4.1 lists the considered three traffic classes and their utility functions. Each superframe contains 10 slots. The size of exclusive region, denoted as $d_{ER}$, is set to 2 m, except in Section 4.5.3 where we vary the size of exclusive region to study its impact on the aforementioned three performance metrics.

Table 4.1: Traffic parameters used in simulations.

| Traffic Class | B.W. Requirement | Utility Function |
|---|---|---|
| I | 1 Mbps | $\begin{cases} 1 & \text{if } r \geq 1, \\ 0 & \text{if } r < 1. \end{cases}$ |
| II | 1-20 Mbps | $1 - e^{\frac{-0.03r^2}{20+r}}$ |
| III | 0-250 Mbps | $\begin{cases} 1 & \text{if } r \geq 250, \\ \sin^{0.3}\left(\frac{\pi}{2} \cdot \frac{r}{250}\right) & \text{if } r < 250. \end{cases}$ |

### 4.5.2 Utility-based scheduling

The total utility that each flow has gained by the time of interest is defined as the cumulative flow utility, and we use two scenarios to demonstrate the long-term performance of the proposed utility-based scheduling. The first scenario consists of ten Class-3 flows with maximum bandwidth requirement of 250 Mbps. This corresponds to the situation where several high bandwidth-demand users are active at the same time. Figure 4.2 shows the cumulative flow utility. For this single traffic case, equal utility should be ensured among users. Hence the parameter $c_i$ in (4.11) is set to be 1 for all flows. In Figure 4.2, a flat segment on the performance curve means that the flow is idle in that particular slot; the curve rises when a new slot is assigned. It is clear that each flow has different scheduling pattern, but their cumulative utility, i.e. users' satisfactory level about the assigned bandwidth, is about the same in the long term. The effectiveness of the control parameter $\rho$ in (4.11) can also be inferred from the figure. As time elapses, the parameter $\rho$ of those users with more aggregate utilities are reduced, and thus the users with less utility and higher value of $\rho$ have higher priority to be scheduled for transmissions in the subsequent slots.

The second scenario contains three traffic classes: three Class-1 flows demanding minimum bandwidth of 1 Mbps, three Class-2 flows with maximum bandwidth requirement of 20 Mbps, and four Class-3 flows with maximum bandwidth of 250 Mbps. The parameter $c_i$'s are set to be $[2,\ 10,\ 1]$ for Class-1, Class-2, and Class-3, respec-

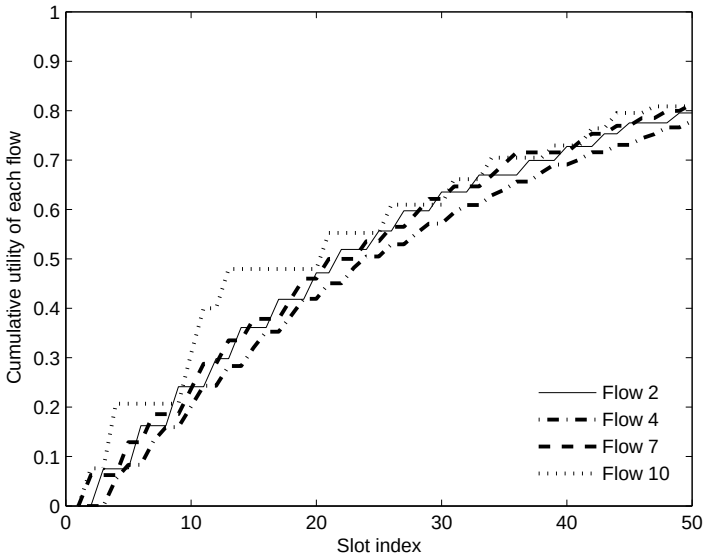



Figure 4.2: Scheduling performance of single-class case: 10 Class-3 flows. Four sample flows are shown with link distance 5.73 m, 8.61 m, 6.93 m, and 3.89 m, respectively.

tively. Figure 4.3 shows the cumulative flow utility in 50 slots. It can be seen that the flows within the same class achieve resembling utilities. Class-1 flows can achieve cumulative utility of one in the first few slots, because of their relatively lower bandwidth requirements and the use of step utility function.

### 4.5.3 Utility vs. fairness

As explained earlier, utility maximization and fairness are two conflict objectives in network optimization. We compare the ER-GSA with two heuristic algorithms proposed in Chapter 3: proportional allocation algorithm (PaA) and repeating allocation algorithm (RaA).

Three performance metrics are compared: 1) the *total utility* of all flows; 2) the

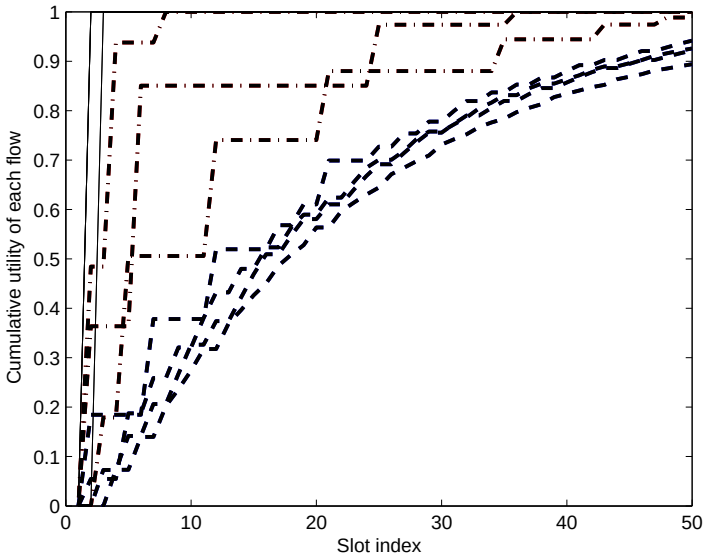
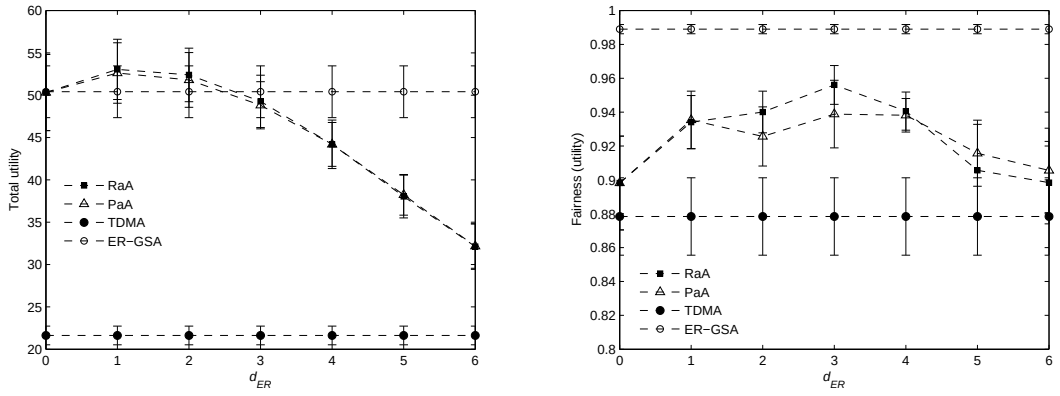



Figure 4.3: Scheduling performance of multi-class case: three Class-1 flows (solid line), three Class-2 flows (dash-dot line), and four Class-3 flows (dashed line).

*minimum per-flow utility* among all flows; and 3) the Jain's *fairness index* [36]. Each point in the figures presented in this section indicates the result at the end of the 10-th slot. The 95% confidence interval from 10 different random topologies is plotted as error bars. Other parameters follow the default setting defined at the beginning of this section.

### Total Utility vs. Fairness

Utility maximization and fair allocation are known to be two conflicting objectives. For instance, if we always choose the flows with better channel quality to transmit, we can achieve higher overall throughput and utility, with the consequence of starving some flows. Our proposed solution is to maximize the total utility under the fairness



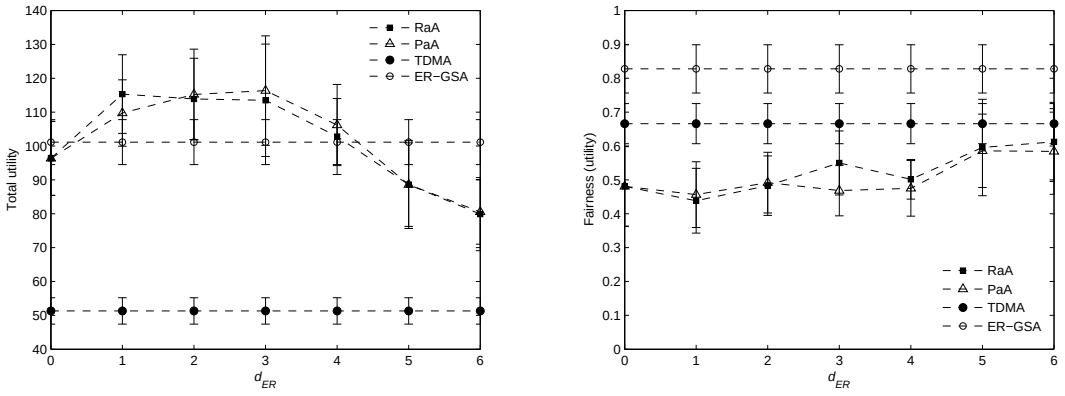

(a) Total utility.

(b) Utility fairness.

Figure 4.4: Comparisons among different scheduling algorithms in single-class case.

constraint. In Figure 4.4(a), the total utility generated by ER-GSA, PaA, RaA, and TDMA among ten Class-3 flows are shown. For PaA and RaA, the total utility varies significantly with respect to the size of exclusive region. ER-GSA achieves comparable total utility as that achieved by PaA and RaA with the best $d_{ER}$. On the other hand, ER-GSA supports much higher level of fairness than PaA and RaA, as shown in Figure 4.4(b). TDMA maintains the best fairness in terms of the number of time slots allocated to each flow, but its achieved utility is much lower (about 58% less) than that achieved by ER-GSA.

We further evaluate the performance of three traffic classes case, using the same configuration as that in Section 4.5.2. To measure fairness for heterogeneous traffic, the Jain's fairness index is computed as $\sum_{i=1}^{N}(u(i)/c(i))^2/(N\cdot\sum_{i=1}^{N}(u(i)/c(i))^2)$, where $u(i), c(i)$ are the utility and the corresponding weighting factor of flow $i$. The total utility and the fairness index at the end of 10th slot are shown in Figure 4.5(a) and Figure 4.5(b), respectively. For total utility, ER-GSA achieves about 90% total utility of that achieved by PaA or RaA, but the later two fail to maintain fairness in the




(a) Total utility.

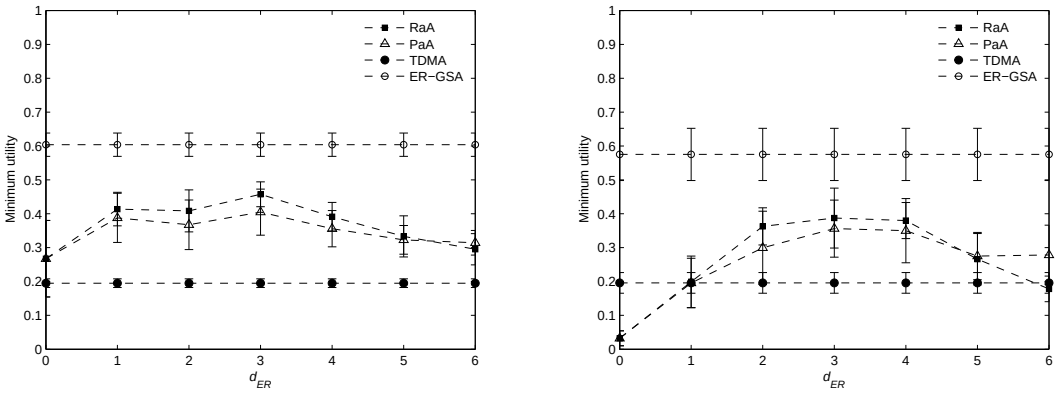
(b) Utility fairness.

Figure 4.5: Comparisons among different scheduling algorithms in three-class case.

presence of multi-class traffic. Together with the results in the single-class case, we conclude that ER-GSA can maintains a good balance between utility maximization and fairness, while PaA and RaA PaA and RaA can achieve high network throughput with less computation time.

### Minimum utility

The maximum aggregate utility is often achieved by maximizing certain users utility while starving others. Hence it is important to consider the performance of the worst user. From Figures 4.6(a) and 4.6(b), it can be seen that the achieved minimum utility using ER-GSA is much higher than that of other algorithms. Note that RaA can be considered as a max-min fair scheduler in terms of number of time slots. In RaA, the user with minimum number of assigned slots is chosen at each scheduling iteration. Thus, RaA performs slightly better than PaA by increasing the minimum utility. TDMA performs the worst.




(a) One-class case.

(b) Three-class case.

Figure 4.6: Comparisons of the minimum utility.

### 4.5.4 Algorithm efficiency and stability

Finally, we study the computation efficiency and convergence issue of ER-GSA. Basically the ER-GSA converges asymptotically. In practice, limited amount of time is allowed for executing the algorithm. To ensure that the algorithm reaches the optimal point promptly and retains on the optimum steadily, ER-GSA modifies the original GSA in two aspects: a) in step 4, the strong interferers in the randomly selected flow set are further removed once a better set of flows is found, so that the resultant flow set is more likely to be the optimum; and b) ER-GSA uses a variable-sample method in step 2 to ensure the convergence.

To evaluate the improvement in step 4, we compute the likelihood that the subset $S'$ in step 4 in ER-GSA generates higher utility than that by $\kappa'_m$ in step 2, *i.e.*, $\mathbb{P}\left[U_m(S') > U_m(\kappa'_m)\right]$. The amount of improvement depends on the network density and $d_{ER}$, as shown in Figure 4.7. By increasing $d_{ER}$, more flows will be removed from $\kappa'_m$, but the chance that $U_m(S') > U_m(\kappa'_m)$ is lower. Thus, a proper exclusive region

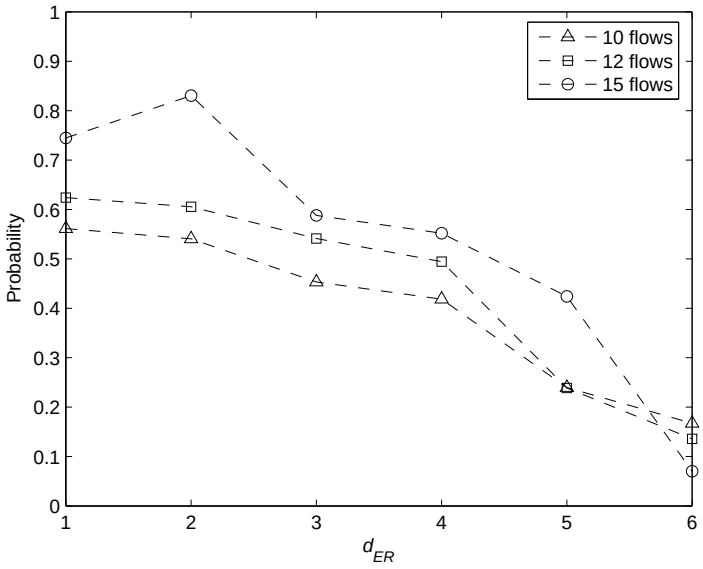



Figure 4.7: $\mathbb{P}\left[U_m(S') > U_m(\kappa'_m)\right]$ versus $d_{ER}$ in different network densities.

size can ensure that $\mathbb{P}\left[U_m(S') > U_m(\kappa'_m)\right]$ is larger than 0.5 such that the optimum may be located faster. The improvement is more significant when there are more flows to be scheduled. Furthermore, Figure 4.8 snapshots the utility achieved in a slot using GSA, and ER-GSA with $d_{ER} = 1$ m. The trajectory of a single run and that of the average of 100 runs are shown. The utility of the optimal set of flows is also plotted for comparison. The proposed ER-GSA can reach the optimal set with much fewer iterations than GSA. On the other hand, it is also important that the algorithm can stay at the optimum state against measurement and estimation errors (noise), which is measured by the stability factor as defined below. Denote $m_i$ and $m'_i$ the number of iterations that the algorithm reaches and leaves the global optimum state at the $i$-th time, respectively. The stability factor $\xi$ of the searching algorithm

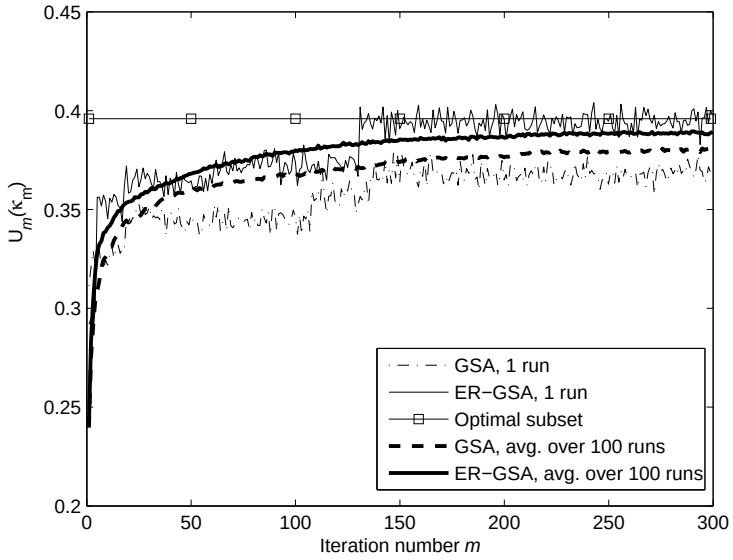



Figure 4.8: Scheduling trajectory.

is defined as

$$\xi = \frac{\sum_i (m_i^{'} - m_i)}{M - m_1},$$

where $(m_i^{'} - m_i)$ represents the number of iterations that the algorithm stays at the optimal point consecutively since the optimal point is reached at the $i$-th time, and $M$ is the total number of iterations. Generally, the smaller value of $\xi$ implies that the algorithm is more sensitive to disturbance. The impact of disturbance may be overcome by increasing the sample size adaptively, as suggested in the variable-sample path random search (SPRS) algorithm [49] (See Appendix C). When the sample size grows sufficiently fast, the SPRS algorithm will stay at the optimal point once it is reached, at the cost of increased computation load. We compare the stability and computation cost of all algorithms being discussed. All algorithms are coded in $C$ language and executed on a Pentium-4 2.8 GHz CPU. We schedule 10 flows for 10

Table 4.2: Execution time and stability $\xi$

| Algorithm | GSA | ER-GSA | SPRS | RaA | PaA |
|---|---|---|---|---|---|
| Execution Time (ms) | 2.63 | 4.28 | 290.89 | 0.438 | 0.114 |
| | 2.61 | 3.44 | 362.06 | 0.560 | 0.110 |
| | 2.47 | 3.42 | 136.66 | 0.594 | 0.115 |
| | 2.58 | 3.43 | 206.73 | 0.706 | 0.078 |
| | 2.29 | 3.80 | 379.75 | 0.579 | 0.121 |
| | 2.66 | 3.52 | 261.06 | 0.589 | 0.076 |
| | 2.63 | 3.32 | 351.34 | 0.570 | 0.158 |
| | 2.58 | 3.30 | 325.87 | 0.595 | 0.122 |
| | 2.55 | 3.68 | 313.74 | 0.601 | 0.105 |
| | 2.71 | 3.34 | 309.51 | 0.593 | 0.106 |
| Stability factor $\xi$ (%) | 73 | 84 | 100 | - | - |

slots, and repeat the algorithm 10 runs where each run contains $1,000$ iterations. Table 4.2 shows the execution time and stability factor $\xi$ corresponding to each algorithm[10]. It can be seen that SPRS has the best stability property, but it is not feasible to implement for real-time scheduling. Considering the typical superframe length of 65 to 90 ms, the complexity of the other four algorithms are acceptable. Note that although the total computation time for the ER-GSA algorithm is slightly higher than that for the GSA algorithm, as shown in Figure 4.7, ER-GSA has better chance to converge to the optimum. In other words, if the algorithm is terminated prematurely due to insufficient computation time, the search results of the ER-GSA is generally better. On the other hand, a sufficient number of iterations is needed to ensure the convergence of ER-GSA and GSA. A simple rule is to let it be at least the size of the entire solution space, so each point can be checked once statistically. More

[10]Since the heuristic RaA and PaA take only one sample of the distance information as the algorithm input, we do not consider their stability performance.

complicated termination rules, such as performing a paired $t$-test or a more general Wilcoxon test (which does not require the assumption of normal distribution in the test data) after a certain number of iterations to decide the termination (as in SPRS), may introduce excessive computation overhead, and thus they are not recommended. Other techniques such as robust trimming to eliminate significantly biased data may also be used to improve the algorithm efficiency.

## 4.6 Summary

Envisioning the need of supporting heterogeneous services in UWB networks, in this chapter, we explored the scheduling design for multi-class traffic. The main accomplishments of this chapter can be summarized as follows:

- We formulated the optimal scheduling for concurrent UWB transmission with heterogeneous bandwidth requirements as a weighted utility maximization problem. Such a formulation achieves efficient and fair resource allocation for UWB networks with heterogeneous users.

- The assumption of perfect distance information is relaxed by factoring the possible estimation errors in the objective function, leading to a stochastic optimization problem.

- We proposed a meta-heuristic algorithm tailed for UWB networks called ER-GSA to solve the NP-hard utility maximization problem and proved its convergence.

- The proposed algorithm has been comprehensively evaluated via simulations. For single traffic class, the proposed scheduling algorithms can provide more

than 200% utility gain over TDMA. For multiple traffic classes, the ER-GSA algorithm can achieve high total utility and both inter-class fairness and intra-class fairness.

In the present chapter and Chapter 3, we have assumed the existence of a central coordinator in UWB networks. With the aid of a central coordinator, we addressed the issue of efficient resource allocation for UWB networks. In the remaining parts of this thesis, we consider pure ad hoc UWB networks without central control.

Table 4.3: List of local notations for Chapter 4

| **Notation** | **Definition** |
|---|---|
| $b$ | MUI factor |
| $c_i$ | differentiation parameter of flow $i$ |
| $m$ | iteration index |
| $p_N$ | noise power |
| $U_i(s)$ | instantaneous utility of flow $i$ in slot $s$ |
| $U(R_i)$ | utility value with respect to bandwidth $R$ of link $i$ |
| $\tilde{U}$ | noise version of $U(\cdot)$ |
| $\{\tilde{U}_m\}$ | a sample sequence of $\tilde{U}$ |
| $\overline{U}$ | sample mean of $\{\tilde{U}_m\}$ |
| $W_m(\kappa)$ | number of times that the algorithm has visited state $\kappa$ up to the $m$th iteration |
| $\alpha$ | path-loss exponent |
| $\sigma_G$ | shadowing parameter |
| $\kappa^*$ | optimal schedule per slot |
| $\kappa_m^*$ | the state being visited by the algorithm most frequently up to the $m$th iteration |
| $\kappa_m$ | the state being visited at the $m$th iteration |
| $\kappa_m^{'}$ | a randomly chosen state at the $m$th iteration |
| $\rho_i^{(s)}$ | utility update parameter or low $i$ in slot $s$ |
| $\mathcal{K}$ | the union of independent sets in a graph |

# Chapter 5

# Prioritized Channel Access for Distributed UWB Networks with Bursty Multimedia Traffic

## 5.1 Introduction

As we have reviewed in Chapter 2, a number of distributed MAC protocols have been proposed [4–6, 50] to exploit the advantages of UWB in distributed ad hoc networks. However, these protocols either involve heavy signaling overhead or do not support heterogeneous traffic. On the other hand, the WiMedia Alliance[11] defines two distributed channel access mechanisms, namely prioritized channel access (PCA) and the distributed reservation protocol (DRP), in their specification for UWB-based HDR WPAN [19]. The PCA employs the same differentiation mechanisms as the enhanced distributed channel access (EDCA) in IEEE 802.11e, while DRP can be

[11]The WiMedia Alliance is an industry organization established in 2005, aiming at promoting UWB technology for future WPANs.

deemed as a distributed TDMA protocol. Although these protocols are not completely new, their performance and potential issues when used in UWB networks remain untouched. This chapter focuses on performance analysis for PCA, and we will study DRP in the next chapter.

The PCA combines the CSMA/CA with the use of different contention parameters associated with different traffic classes to achieve prioritized channel access in a distributed manner. Most previous work on the analysis of CSMA/CA protocol and its variants assumes saturation stations and independent interarrival times between packets. Multimedia applications, however, exhibit strong burstiness/correlations between interarrivals that violate the above assumptions. As UWB is envisioned to support various multimedia applications with stringent delay requirement, it is important to consider the characteristics of multimedia traffic in analyzing the delay performance of PCA.

Multimedia traffic with bursty and correlated interarrival time forms a nonrenewal arrival process that significantly deviates from the Bernoulli or Poisson process commonly assumed in the literature. The nonrenewal arrival process has a profound impact on the queueing statistics, as has been confirmed by many studies (see [51] and the reference therein). Although a nonrenewal process is more accurate in capturing the real characteristic of multimedia applications than the renewal counterpart, the exact queueing analysis is quite difficult and generally incurs a high computational burden. An alternative is to seek some acceptable approximations with close enough performance characteristics to those of the original system.

In this chapter, we study the performance of PCA in which the arrival process reveals bursty and correlated characteristics. The arrival process is described by a Markov Modulated Poisson Process (MMPP), for its versatility of modeling various traffic sources and the capability of capturing the burstiness and correlation in the

arrival stream. The network consists of two classes whereby multimedia traffic such as voice or video streaming has higher priority to access the channel, and data traffic such as file transfer has lower priority. We model the backoff and channel access behaviors of a tagged user in each class, and obtain the probability generating function (PGF) of the MAC service time distribution. For the mean frame waiting time (*i.e.*, queueing delay), we resort to the queueing analysis and model the tagged station buffer as a $MMPP/G/1$ queue. The $MMPP/G/1$ modeling presented in [52] takes the essential characteristics of MMPP process into account and thus guarantees the model accuracy. However, the required computation exponentially increases with the number of states of the MMPP. Alternatively, approximation approaches have been developed to reduce computational burden with acceptable accuracy. The $G/G/1$ approximation is commonly used and quite general, but it can be possibly accurate as the queue utilization approaches to one [53]. Recently, Jagerman et al. (2004) propose a renewal approximation, which is capable of capturing the statistical characteristic of autocorrelated arrival process. This method has a comparable accuracy to the $MMPP/G/1$ modeling but largely reduces computational effort [51]. We obtain the mean waiting time through the above three approaches, and comprehensively study their accuracy.

The remainder of this chapter is organized as follows. Section 5.2 reviews the related work. In Section 5.3, we introduce the PCA protocol, the traffic and the network model considered in this work. We present the analysis of service time in Section 5.4, followed by that of waiting time in Section 5.5. We validate our analytical model and discuss the protocol performance under different system parameters through numerical results in Section 5.6. Section 5.7 summarizes this chapter.

## 5.2 Related Work

The considered PCA is a CSMA/CA based MAC protocol with traffic prioritization. There has been a tremendous amount of research studying the performance of CSMA/CA protocols and its variants, such as the DCF in IEEE 802.11 and the EDCA in IEEE 802.11e. Two major approaches have been employed in deriving the average MAC service time, namely the discrete Markov modeling [54–59] and the mean value analysis [60,61]. Most of the work is concerned with the asymptotic performance where each station in the network is saturated with traffic arrivals, thereby the mean service time can be found equal to the reciprocal of the throughput. In practice, however, the station queues may not always be full, thus the inverse relation between average service time and throughput does not exist. Another approach has been proposed in [61] where the mean service time for both saturated and unsaturated stations can be successfully captured based on renewal theory.

Recently, the emergence of multimedia applications in the wireless domain has drawn much attention on studying the QoS provisioning for delay-sensitive traffic. Besides the service time, the waiting time (*i.e.*, queueing time) of a MAC frame has a significant impact on the delay performance, which is not only dependent on its service time that the network provides, but is also effected by the incoming traffic characteristics. Several research works on queueing analysis for DCF and EDCA have appeared, where the arrival process is always assumed uncorrelated [59,62]. For multimedia traffic, however, packet interarrivals are typically correlated and bursty in nature. In [63], a nonrenewal MMPP arrival process is considered, resulting in a $MMPP/G/1/K$ queueing model. These studies obtain the collision probability as a function of the station idle probability (*i.e.*, when the MAC buffer at the tagged station is empty), which is dependent on both the service time distribution and the

Table 5.1: User priority to access categories mapping

| Priority | Traffic type | $CW_{min}$ | $CW_{max}$ | AIFSN |
|---|---|---|---|---|
| lowest | Background | 15 | 1023 | 7 |
| ↓ | Best effort | 15 | 1023 | 4 |
| | Video | 7 | 511 | 2 |
| highest | Voice | 3 | 255 | 1 |

characteristics of arrival process. Thus, the studies rely on certain recursive algorithms to find the collision probability, and the resultant computation is normally high. In addition, the impacts of burstiness and correlation in interarrival streams have not been explored, thus their results may not be so useful for assessing the delay performance of multimedia traffic with bursty/correlated arrivals.

## 5.3 Preliminaries

### 5.3.1 PCA Protocol

PCA provides several mechanisms, which can be used stand-alone or as an integrated approach to achieve prioritized channel access. User traffic is differentiated into four access categories (ACs), as shown in Table 5.1, where each AC is associated with different contention parameters [19]. The first differentiation mechanism is the contention window (CW) size. When a station[12] has a frame at the MAC sublayer buffer, it will first sense the channel. If the channel is busy, it performs the backoff procedure by first setting the backoff counter to an integer sampled from the CW. The CW size starts with an initial value ($CW_{min}$) and is doubled after every unsuccessful transmission. A higher priority AC with a smaller value of $CW_{min}$ may spend less time on backoff and thus gain higher priority in a statistical sense.

[12]Throughout this chapter, the terms "station" and "user" are used interchangeably.

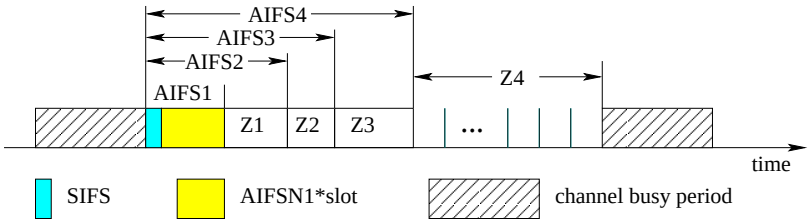



Figure 5.1: Illustration of prioritized channel access for different ACs.

The second differentiation mechanism is the use of arbitrary interframe space (AIFS). After the channel becomes idle for an AIFS time, stations can count down the backoff counter at the beginning of each idle slot, and also the first slot of a channel busy period. By assigning the higher priority ACs with shorter AIFS, they obtain higher chances to access the channel than low priority ACs. The length of AIFS is determined by AIFS = SIFS + AIFSN $\times \sigma$, where SIFS represents the short interframe spacing, AIFSN is an integer between $[1, 7]$, and $\sigma$ is the slot time duration. Figure 5.1 shows an example of four ACs, where $AC_1$ has the highest priority. To illustrate the effect of different AIFS lengths, the time between two busy period except $AIFS_1$ is divided into four contention zones, $\mathsf{Z}_i$, $i = 1, 2, 3, 4$. In $\mathsf{Z}_1$, only $AC_1$ stations are allowed to contend for channel access, while in $\mathsf{Z}_2$ the contentions are between $AC_1$ and $AC_2$, *i.e.*, contentions in $\mathsf{Z}_i$ involve $AC_j, j \leq i$. Consequently, each AC encounters different contentions in its different contention zones.

After successful contention, the station can transmit for a duration up to that specified by the parameter called transmission opportunity (TXOP). A larger TXOP allows the station to transmit a burst of packets without contending for the channel between transmissions.

### 5.3.2 MMPP Model for Bursty Traffic

Multimedia streams such as voice, coded video, and Internet traffic usually possess correlated and bursty characteristics which may significantly affect system performance (*e.g.*, delay outage probability and throughput) [64, 65]. By burstiness it is meant that one can observe the clustering phenomenon of arrivals on the time line [66]. Let $X = \{X_j\}_{j=0}^{\infty}$ denote the interarrival time of a traffic source with common distribution $A(x)$[13], and finite mean and variance, where $X_j$ represents the $j$-th sample of $X$ taken at the time $t_j$. The burstiness can be characterized by the squared coefficient of variation of the interarrival time denoted as $\mathsf{c}^2$, given by [67],

$$\mathsf{c}^2 = \frac{\mathbb{V}(X)}{\mathbb{E}^2(X)}, \tag{5.1}$$

where $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ denote the mean and variance operator, respectively. A highly bursty arrival process tends to have a higher value of $\mathsf{c}^2$, while a non-bursty Poisson process has a $\mathsf{c}^2$ value of 1.0. Another important feature of multimedia traffic, particularly the variable bit rate (VBR) streams, is the high correlation between interarrival times that produces long range dependence into the arrival process, and hence cumulative effect on the queueing system. The degree of correlation between interarrival times can be measured by the autocorrelation function $\mathsf{r}_X(j)$ of the interarrival time process $X$, namely,

$$\mathsf{r}_X(j) = \frac{\mathbb{E}(X_k X_{k+j}) - \mathbb{E}^2(X)}{\mathbb{V}(X)}, \quad j = 0, 1, \cdots, \ \forall\ k. \tag{5.2}$$

In this study, the arrival process of multimedia traffic is represented by MMPP. The MMPP model is a nonrenewal doubly stochastic process where the rate process is determined by the state of a continuous-time Markov chain. An $m$-state MMPP is

[13]Here $X$ is assumed to be stationary.

characterized by two elements: the infinitesimal generator $\mathbf{Q}$ given by

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & -\sigma_2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & -\sigma_m \end{bmatrix}, \tag{5.3}$$

where $\sigma_i = \sum_{j=1, j\neq i}^{m} \sigma_{ij}$ and $\sigma_{ij}$ governs the transition rate from state $i$ to state $j$; and the Poisson arrival rate matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_m), \tag{5.4}$$

where $\lambda_i$ is the rate of a Poisson arrival process at state $i$ of the Markov chain. The steady-state probability vector $\mathbf{\Pi}$ of the Markov chain can be determined using the following equations:

$$\mathbf{\Pi Q} = \mathbf{\Pi}, \ \mathbf{\Pi e} = 1, \tag{5.5}$$

where $\mathbf{e} = (1, 1, \cdots, 1)^T$. The reason for choosing MMPP is two-fold. Firstly, many studies have shown that MMPP has enough flexibility to describe a wide variety of traffic with correlated and bursty arrival processes [68]. Secondly, the queueing related results of MMPP have been well-studied [69–71]. Therefore, the use of MMPP offers versatility in the modeling environment and allows to achieve analytical tractability while preserving the actual traffic characteristics [51].

### 5.3.3 MAC Model

We consider a distributed network consisting of two classes of users for simplicity. The proposed analytical model can be generalized to more than two classes with minor modifications. Let $N_i$ denote the number of users in $\mathrm{AC}_i$ for $i = 1, 2$, and $\mathrm{AC}_1$ users

have higher priority than $AC_2$ in accessing the channel. The operation of beacon group specified in the standard ensures there are no hidden node problems [19, 72].

The analysis is based on a discrete-time system, where the basic time unit is called *generic slots* denoted by $\phi$. The length of a generic slot is varied, depending on the channel status including idle, successful transmission and collision. If there is no transmission during a physical time slot, the length of genetic slot is identical to that of a physical time slot $\sigma$. On the other hand, the channel may be busy in two cases. If there is a successful transmission, it takes $T_s$ time during which the channel is sensed to be busy. In the basic access mode, $T_s = T_{(\texttt{DATA})} + \text{SIFS} + T_{(\texttt{ACK})} + \text{DIFS}$. In the RTS/CTS mode, $T_s = T_{(\texttt{RTS})} + T_{(\texttt{CTS})} + T_{(\texttt{DATA})} + 3\text{SIFS} + T_{(\texttt{ACK})} + \text{DIFS}$. If there is a collision, it takes $T_c$ time, where $T_c = T_{(\texttt{DATA})} + \text{ACK-timeout} + \text{DIFS}$ for basic mode, and $T_c = T_{(\texttt{DATA})} + \text{CTS-timeout} + \text{DIFS}$ for RTS/CTS mode. In the above. $T_{(\texttt{x})}$ represents the transmission time of the frame type `x`.

An ideal wireless channel without transmission error is assumed so that all transmitted frames may be lost only due to collisions caused by simultaneous transmissions from multiple users. The effect of imperfect channels can be embedded in our analysis following the approach presented in [73]. For simplicity, all MAC frames are assumed to have the same fixed length. The case of different frame lengths (equivalently, the consideration of TXOP) can be incorporated in our model following the work in [74].

The main performance metric of interest is the average frame delay, consisting of the average waiting time and the average service time. The former counts the time that a data frame stays in the MAC sublayer buffer before becoming the head of the buffer, and the latter is the time from the moment that a frame reaches the head of the MAC sublayer buffer to the moment that it is successfully transmitted or discarded. Since the frame waiting time is a function of the frame service time, we first derive the service time in the next section. The result will be used in Section 5.5

for waiting time analysis.

## 5.4 MAC Service Time Analysis

The modeling of frame service time heavily relies on two key probabilities, namely, the station transmitting probability and the frame collision probability, conditioned on there is at least one frame in the user's buffer to be served. Considering an $AC_i$ user, the former is denoted by $\tau_i$, and the latter is denoted by $p_i$. For a lossless queueing system, the probability of non-empty buffer is given by the server utilization factor $\rho = \lambda_X \cdot \bar{Z}$ where $\lambda_X$ is the mean frame arrival rate and $\bar{Z}$ is the mean frame service time. The probability that an unsaturated $AC_i$ station transmits in a randomly chosen generic slot thus equals $\tau_i \rho_i,\ \rho_i \in (0, 1]$. We follow the approach proposed in [61] to obtain $p_i$ and $\tau_i$. With the probabilities $\tau_i$ and $p_i$, we then proceed to derive the PGF (equivalently, the Z-transform) of the MAC service time for both classes. By numerical inversion of the Z-transform, the probability mass function (PMF) and the corresponding moments can then be obtained.

### 5.4.1 Transmission and Collision probabilities

We assume that the transmission probability, *i.e.*, the probability of a station to initiate a transmission in a given backoff slot is constant in all its backoff slots [55,58]. Since the channel access procedure of the tagged station regenerates itself for each new MAC frame, the complete service periods for MAC frames form renewal cycles in the renewal process. The average length of the renewal cycle is thus equivalent to the average frame service time [61]. According to the renewal reward theorem, in a randomly chosen slot, the transmitting probability $\tau_i$ of an $AC_i$ station can be

obtained as the average reward during the renewal cycle, given by

$$\tau_i = \frac{\mathbb{E}[R_i]}{\mathbb{E}[R_i] + \mathbb{E}[B_i]}, \tag{5.6}$$

where $\mathbb{E}[R_i]$ is the expected number of transmission trials for a frame and $\mathbb{E}[B_i]$ is the expected number of total backoff slots experienced by the frame. Assuming a frame of $\text{AC}_i$ station may incur an average collision probability of $p_i$, $R_i$ follows a truncated geometric distribution and $\mathbb{E}[R_i]$ is give by

$$\mathbb{E}[R_i] = \sum_{j=0}^{m-1} (p_i)^j. \tag{5.7}$$

Similarly, $\mathbb{E}[B_i]$ can be obtained as

$$\mathbb{E}[B_i] = \sum_{j=0}^{m-1} b_j (p_i)^j, \tag{5.8}$$

where $b_j = CW_j/2$ is the average number of backoff slots in backoff stage $j$, $j = 0, \cdots, m$, and $m$ is the retry limit. Notice that the class-dependent CW parameters have been included in the analysis. The collision probability of $\text{AC}_2$ can be obtained as

$$p_2 = 1 - (1 - \rho_1\tau_1)^{N_1}(1 - \rho_2\tau_2)^{N_2 - 1}, \tag{5.9}$$

given by the fact that the $\text{AC}_2$ station can only transmit in zone 2 with possible collisions with one or more of the other stations from any class. The computation of collision probability of $\text{AC}_1$ is more involved, as its transmissions may take place in either zone 1 or zone 2 with collision probabilities $p_{1,1}$ and $p_{1,2}$, respectively, where

$$p_{1,1} = 1 - (1 - \rho_1\tau_1)^{N_1 - 1}, \tag{5.10}$$

$$p_{1,2} = 1 - (1 - \rho_1\tau_1)^{N_1 - 1}(1 - \rho_2\tau_2)^{N_2}. \tag{5.11}$$

For the tagged $\text{AC}_1$ station, its frame transmission may occur in either zone 1 or zone 2. Suppose zone 1 contains $M$ slots (*i.e.*, the difference between $\text{AIFSN}_1$ and

$\text{AIFSN}_2$), the frame transmission may take place in zone 2 if neither itself nor any of the other $\text{AC}_1$ stations transmit in zone 1 with probability denoted by $\theta_2$,

$$\theta_2 = \left((1-\tau_1\rho_1)^{N_1-1}(1-\tau_1)\right)^M. \tag{5.12}$$

Otherwise, the transmission occurs in zone 1 with probability $\theta_1 = 1-\theta_2$. The average collision probability of an $\text{AC}_1$ station can thus be given by

$$p_1 = \theta_1 p_{1,1} + \theta_2 p_{1,2}. \tag{5.13}$$

By jointly solving Equations (5.6), (5.9) and (5.13) we can obtain $(\tau_1, \tau_2, p_1, p_2)$.

### 5.4.2 PGF of Frame Service Time

As mentioned in the previous section, the mean waiting time is a function of the service time distribution. We derive the PGF of frame service time in this section. Similar to [75], we work on a discrete-time system where the time interval in our analysis is approximated as multiples of a common quantity, representing the smallest granularity that can be observed by our model. We refer to this time unit as the *generic slot*. Thus the frame service time is a discrete random variable leading to a Z-transform based analysis. For a tagged station of $\text{AC}_i$, it spends an amount of time $Z_i = B_i + R_i$ to transmit a frame successfully, where $B_i$ ($R_i$) is the random variable representing the amount of time attributed to backoff (transmission trials). Moreover, the introduction of AIFS differentiation causes further delay to $\text{AC}_2$ stations as explained in Section 5.3.1. This additional amount of time is referred to as the "pre-backoff waiting" period [61], denoted by $Z'$. Therefore, the PGF of frame service time can be written as

$$\begin{aligned} G_{Z_1}(z) &= G_{B_1}(z)G_{R_1}(z), \\ G_{Z_2}(z) &= G_{B_2}(z)G_{R_2}(z)G_{Z'}(z). \end{aligned} \tag{5.14}$$

In the following, we derive each component in $G_{Z_i}(z)$, where the subscript $i$ will be omitted for notation brevity.

**Generic Slot** $G_\phi(z)$

The time unit is measured in generic slot $\phi$ as defined in 5.3.3. For a randomly chosen slot, the channel status may be in one of the following three mutually exclusive events: being idle ($\mathsf{i}$), having a successful transmission ($\mathsf{s}$), or having a collision ($\mathsf{c}$). The length of generic slot $\phi$ can be expressed as:

$$\phi = \Omega_\mathsf{i}\sigma + \Omega_\mathsf{s}T_s + \Omega_\mathsf{c}T_c, \tag{5.15}$$

where $\Omega_\mathsf{e}$ is a binary variable which takes the value of one if the event $\mathsf{e} \in \{\mathsf{i}, \mathsf{s}, \mathsf{c}\}$ occurs, and zero otherwise. Thus the PGF of $\phi$ takes the following form:

$$G_\phi(z) = p_\mathsf{i} z^\sigma p_\mathsf{s} G_{T_s}(z) p_\mathsf{c} G_{T_c}(z), \tag{5.16}$$

where $p_\mathsf{i}$, $p_\mathsf{s}$, and $p_\mathsf{c}$ are class-dependent as given by [61]

$$p_{\mathsf{i},1} = (1-\rho_1\tau_1)^{N_1}, \tag{5.17}$$

$$p_{\mathsf{i},2} = (1-\rho_1\tau_1)^{N_1}(1-\rho_2\tau_2)^{N_2}, \tag{5.18}$$

$$p_{\mathsf{s},1} = N_1\rho_1\tau_1(1-\rho_1\tau_1)^{N_1-1}, \tag{5.19}$$

$$p_{\mathsf{s},2} = N_1\rho_1\tau_1(1-\rho_1\tau_1)^{N_1-1}(1-\rho_2\tau_2)^{N_2} + N_2\rho_2\tau_2(1-\rho_2\tau_2)^{N_2-1}(1-\rho_1\tau_1)^{N_1}, \tag{5.20}$$

$$p_{\mathsf{c},i} = 1 - p_{\mathsf{i},i} - p_{\mathsf{s},i}, \quad i = 1, 2. \tag{5.21}$$

**Backoff Period** $G_B(z)$

Between two successful transmissions, the time contributed by the backoff procedure $B$ is

$$B = \sum_{j=1}^{N_C} \phi_j, \tag{5.22}$$

where $N_C$ is the overall number of generic slots between two successful transmissions given that a frame transmission undergoes $C$ trials ($C \in \{1, 2, \cdots, m\}$) before success, and $\phi_j$ is the length of the $j$th generic slot (we assume $\phi$ is an i.i.d. random variable). Using the conditional expectation, the PGF of $B$ can be written as

$$G_B(z) = \mathbb{E}[z^{\sum_{j=1}^{N_C} \phi_j}] = \sum_{c=1}^{m} \mathbb{E}[z^{\sum_{j=1}^{N_c} \phi_j}]\mathbb{P}[C = c]. \tag{5.23}$$

Similar to the argument of $R_i$ in (5.7), $C$ is a geometric random variable with successful probability $1-p$, i.e., $\mathbb{P}[C = c] = p^{c-1}(1-p)$. For the first term in (5.23), the sum of a random number $N_C$ of i.i.d. random variables $\phi$, represented by $S_{N_C} = \sum_{j=1}^{N_C} \phi_j$, has the following property: $G_{S_{N_C}}(z) = G_{N_C}(G_\phi(z))$. Using this property, we can obtain $G_B(z)$ as

$$G_B(z) = (1-p) \sum_{c=1}^{m} p^{c-1} G_{N_C}(G_\phi(z), c), \tag{5.24}$$

where we use the notation $G_{N_C}(G_\phi(z), c)$ to indicate that it is a function of $c$, which can be derived according to [75] as follows. Let $x_j$ be the number of generic slots contained in the backoff stage $j$, $j = 0, \cdots, C-1$. According to the exponential binary backoff, $x_j$ is uniformly distributed over $[0, CW_j - 1]$, where $CW_j = \min\{2^j CW_0, 2^{m'} CW_0\}$, $m'$ being the maximum backoff stage. The PGF of $x_j$ can be derived as

$$G_{x_j}(z) = \sum_{k=0}^{CW_j - 1} \frac{z^k}{CW_j} = \frac{1 - z^{CW_j}}{(1-z)CW_j}. \tag{5.25}$$

The random variable $N_C$ can be expressed as

$$N_C = \sum_{j=0}^{C-1} x_j, \tag{5.26}$$

such that the corresponding PGF is given by

$$G_{N_C}(z, c) = \frac{1}{(1-z)^{c+1}} \prod_{j=0}^{c-1} \frac{1 - z^{CW_j}}{CW_j}. \tag{5.27}$$

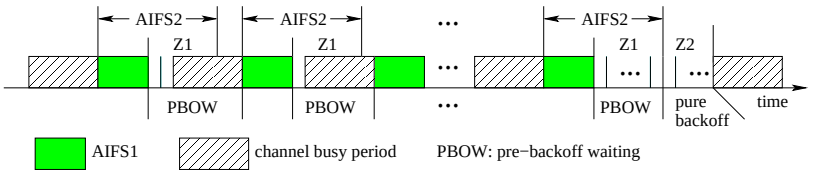



Figure 5.2: Illustration of pre-backoff waiting periods of $AC_2$.

**Retry Period** $G_R(z)$

Given that there are $C$ transmission trials encountered before a successful frame transmission, the random variable $R$ representing the total time contributed by transmission trials can be written as

$$R = \sum_{j=1}^{C} \phi_j. \tag{5.28}$$

Therefore, the PGF of $R$ is given by

$$G_R(z) = G_C(G_\phi(z)). \tag{5.29}$$

According to the fact that $C$ follows a truncated geometric distribution, its PGF is derived as

$$G_C(z) = \sum_{k=1}^{m} z^k \mathbb{P}[C = k] + p^m z^m = (1-p) \sum_{k=1}^{m} z^k p^{k-1} + (pz)^m.$$

**Pre-backoff Period** $G_U(z)$

For $AC_2$, it undergoes extra delays than $AC_1$ as illustrated in Figure 5.2. When the tagged $AC_2$ station is backing off in $\mathsf{Z}_2$, its backoff procedure is interrupted if any other stations (either $AC_1$ or $AC_2$) transmit. Let us call such a backoff procedure during which the tagged station counts down its backoff counter as the "pure backoff". Because any transmissions of $AC_1$ stations in $\mathsf{Z}_1$ will stop the tagged $AC_2$ station from possibly entering $\mathsf{Z}_2$, it takes a certain amount of time before the tagged $AC_2$ station

can start the pure backoff. We call such a delay as the "pre-backoff waiting period" (PBOW), denoted by $U$. Using a similar argument, we can derive $G_U(z)$ as (see Appendix B),

$$G_U(z) = (1-p)\sum_{c=1}^{m} p^{c-1} G_{N_U(c)}(G_\eta(z)) + (pz)^m z^\Delta \tag{5.30}$$

Thus far, we have derived the PGF of the frame service time for each priority class. However, it is often very difficult, or even impossible, to analytically invert the Z-transform of a discrete probability distribution. Several numerical inversion algorithms have been proposed to address this difficulty. Next we employ the approach in [76, 77] to obtain the $n$th moment of a discrete random variable from its PGF.

### 5.4.3 Numerical Evaluation of the Frame Service Time

The PMF of the frame service time $Z_i$, $i = 1, 2$, derived in the previous subsection, can be obtained by the numerical algorithm reported in [77]:

$$Z_i(k) = \frac{1}{2klr^k}[\beta_0(k,l,r) + (-1)^k \beta_k(k,l,r) + 2\sum_{j=1}^{k-1}(-1)^j Re(\beta_j(k,l,r))], \tag{5.31}$$

where $\beta_j(k,l,r) = \sum_{j_1=0}^{l-1} e^{-\pi i j_1/l} Z(re^{\pi \mathrm{j}(j_1+lj_2)/lk})$, $\mathrm{j} = \sqrt{-1}$, $1 \le j_2 \le k$ for real $r$ and integer $l$. As suggested in [77], the algorithm can achieve low error estimate (less than $10^{-8}$) by setting $l = 1$ and $r = 10^{-4/k}$, reducing to the simplified formula:

$$Z_i(k) = \frac{1}{2kr^k}[Z_i(r) + (-1)^k Z(re^{\pi i}) + 2\sum_{j=1}^{k-1}(-1)^j Re(Z(re^{\pi ij/k}))] \tag{5.32}$$

The $n$th moment, $\mu_n$, is obtained by numerically inverting $Z(z')$, $z' = e^z$ [76]:

$$\mu_n = \frac{n!}{2nlr_n^n}\{Z(r_n) + (-1)^n Z(-r_n) + 2\sum_{j=1}^{nl-1} Re[Z(r_n e^{\pi ij/nl})e^{\pi ij/l}]\} - \bar{e}. \tag{5.33}$$

## 5.5 Mean Waiting Time Analysis

Our mean waiting time analysis is obtained by modeling each station as a $G/G/1$ queue. It is well known that there is no exact expression for the mean waiting time of the $G/G/1$ queue. In what follows we consider three approximate queueing systems and summarize them in Table 5.2.

### 5.5.1 *MMPP/ G/ 1*

The *MMPP/ G/ 1* model is parameterized by the service time distribution and its Laplace-Stieltjes transform $h(s)$; the arrival process is parameterized by $\mathbf{Q}$ and $\mathbf{\Lambda}$ (see Section 5.3.2). The mean waiting time $\overline{W}$ can be found as [78, Sec. 3.1.4.1]

$$\begin{aligned}\overline{W} =& \frac{1}{\rho}\Bigg[\frac{1}{2(1-\rho)}[2\rho + \lambda_a h^{(2)} - 2h\left((1-\rho)\boldsymbol{g} + h^{(1)}\mathbf{\Pi\Lambda}\right) \\ & (\mathbf{Q}+\mathbf{e}\Pi)^{-1}\boldsymbol{\lambda}] - \frac{1}{2}\lambda_a h^{(2)}\Bigg],\end{aligned} \tag{5.34}$$

where $h^{(2)}$ is the second moment of $h(s)$, and $\boldsymbol{g}$ is a vector that can be obtained by the iterative algorithm provided in [78, Sec. 3.2]. We consider two approximations for the service time: the exponential and the gamma distributions. They are considered because the Laplace transform of either has a closed-form expression, which is required in computing the vector $\boldsymbol{g}$ in (5.34). In addition, these two distributions are representative in the sense that they rely on different orders of moments to model the distribution. That is, the exponential distribution can be modeled just by the first-order statistics, while the gamma distribution needs the first two moments to describe its distribution. Thus we consider two queueing systems, namely, MMPP arrival process with gamma service time, denoted as $Q^{\Gamma}_{MMPP}$, and MMPP arrival process with exponential service time, denoted as $Q^{M}_{MMPP}$. Their distributions and corresponding Laplace transforms are listed in Table 5.2, where $m_1$ and $m_2$ represent

Table 5.2: Summary of queueing Systems.

| Notation | Queue | Service time distribution | Mean waiting time |
|---|---|---|---|
| $Q^{\Gamma}_{MMPP}$ | $MMPP/\,\Gamma/\,1$ | Gamma<br>$f_{\Gamma}(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$<br>$\tilde{h}_{\Gamma}(s) = (\frac{\beta}{s+\beta})^{\alpha}$ | Eq. (5.34) |
| $Q^{M}_{MMPP}$ | $MMPP/\,M/\,1$ | Exponential<br>$f_{M}(x) = 1/m_1 e^{-x/m_1}$<br>$\tilde{h}_{\Gamma}(s) = (\frac{\beta}{s+\beta})^{\alpha}$ | Eq. (5.34) |
| $Q^{\Gamma}_{heavy}$ | $MMPP/\,\Gamma/\,1$ $(\rho \to 1)$ | Gamma | Eq. (5.35) |
| $Q^{M}_{heavy}$ | $MMPP/\,M/\,1$ $(\rho \to 1)$ | Exponential | Eq. (5.35) |
| $Q^{M}_{PMRQ}$ | $PMRQ/\,M/\,1$ | Exponential | Eq. (5.38) |

the first two moments of the service time obtained from (5.33), with $\alpha = \frac{m_1^2}{m_2 - m_1^2}$ and $\beta = \frac{m_1}{m_2 - m_1^2}$.

### 5.5.2 Heavy traffic approximation

For the heavy-traffic case (or saturated stations), *i.e.*, $\rho \to 1$, $\overline{W}$ can be approximated by [79]

$$\overline{W} \approx \frac{\rho}{1-\rho} \frac{h(c_X^2 + c_Y^2)}{2}, \tag{5.35}$$

where $c_X = \lambda_X \sigma_X$ and $c_Y = \lambda_Y \sigma_Y$ denote the coefficient of variation of the interarrival time and service time, respectively. Combining with the exponential service time approximation, we obtain the queueing systems $Q^{M}_{heavy}$. Likewise, $Q^{\Gamma}_{heavy}$ represents the heavy-traffic approximation with gamma-distributed service time.

### 5.5.3 *PMRQ* Approximation

The exact analysis of the queueing system with autocorrelated arrival processes (thus nonrenewal) is generally hard and incurs a high computational burden. Approxi-

mating the nonrenewal arrival process by a renewal counterpart is a commonly used approach to deal with complex queueing system. Recently, Jagerman et al. [51] proposed a renewal approximation to analyze delay systems with autocorrelated arrival processes. The property of the correlated interarrival time is first captured by the peakedness function as defined in [80]. By mapping a $G/\,G/\,1$ queue to an approximating $GI/\,G/\,1$ queue called *PMRQ* (Peakedness Matched Renewal Queue), which preserves the peakedness of the original arrival process and its arrival rate, it is shown that the approximate $GI/\,G/\,1$ queue achieves close enough performance measures to those of the original system. In this chapter we adopt the peakedness matching technique proposed in [51] to estimate the mean waiting time in our system, leading to the $Q^{\Gamma}_{PMRQ}$ approximation with gamma service time distribution and the $Q^{M}_{PMRQ}$ with exponential service time distribution. Note that the *PMRQ* approximation has also been applied to a recent work [81], studying the impact of correlated wireless channel variations to queueing systems. In the following, we give the gist of the *PMRQ* approximation relevant to our study.

The purpose of the *PMRQ* approximation is to approximate a general arrival process $X$ by a renewal process $X'$, considering the fact that $X'$ is generally analytically simpler than $X$. The approximation is achieved by matching the peakedness function of $X$, denoted as $z_X(s)$, to that of $X'$, denoted as $z_{X'}(s)$. It has been shown that the Laplace transform of $X'$ takes the following form:

$$\tilde{a}_{X'}(s) = \frac{\lambda_X \alpha_E + (\lambda_X + A_E \alpha_E)s}{\lambda_E \alpha_E + (\lambda_E + \alpha_E + A_E \alpha_E)s + s^2}, \quad s \geqslant 0, \tag{5.36}$$

where $\lambda_X$ is the average arrival rate of $X$, $A_E$ and $\alpha_E$ are estimated from $z_X(s)$. To obtain the mean waiting time, first consider the complementary stationary distribution of the waiting time, $W$, asymptotically approximated as

$$\mathbb{P}[W > t] \approx \Gamma_W e^{-\theta_W t}, \quad t \geq 0, \tag{5.37}$$

where $\Gamma_W$ is referred to as the asymptotic coefficient, and $\theta_W$ is called the critical decrement. The corresponding mean waiting time can be approximated by

$$\overline{W} \approx \frac{\Gamma_W}{\theta_W}. \tag{5.38}$$

Given the Laplace transform of the approximate renewal process $\tilde{a}_{X'}(s)$ and that of the service time distribution $\tilde{h}(s)$, one can compute $\theta_W$ as the smallest positive root of the equation

$$\tilde{a}_{X'}(\theta)\tilde{b}(-\theta) = 1, \quad \theta > 0. \tag{5.39}$$

On the other hand, the asymptotic coefficient, $\Gamma_W$, can be found from the formula

$$\Gamma_W = 2\frac{\tilde{k}_+(0) - \tilde{k}_+(\theta_W)}{\tilde{k}_+(-\theta_W) - \tilde{k}_+(\theta_W)}, \tag{5.40}$$

where $\tilde{k}_+$ can be obtained by decomposing the kernel transform

$$\tilde{k}(s) = \tilde{a}_{X'}(-s)\tilde{h}(s) \tag{5.41}$$

into $\tilde{k}(s) = \tilde{k}_-(s) + \tilde{k}_+(s)$. By inserting $\tilde{a}_{X'}(s)$ and $\tilde{h}(s)$ into (5.41) and using the partial fraction decomposition technique, one can obtain the decomposition of $\tilde{k}(s)$

$$\tilde{k}_+(s) = \tilde{k}(s) - \tilde{k}_-(s) \tag{5.42}$$

$$\begin{aligned}\tilde{k}_-(s) &= \frac{\lambda_X\alpha_E - (\lambda_X + A_E\alpha_E)r_1}{r_1 - r_2} \cdot \frac{\tilde{h}(r_1)}{s - r_1} \\ &+ \frac{\lambda_X\alpha_E - (\lambda_X + A_E\alpha_E)r_2}{r_2 - r_1} \cdot \frac{\tilde{h}(r_2)}{s - r_2},\end{aligned} \tag{5.43}$$

where $(r_1, r_2)$ are the roots of the quadratic function $\lambda_X\alpha_E - (\lambda_X + \alpha_E + A_E\alpha_E)s + s^2 = 0$. Notice a typo in [51, Eq. (9.5)] has been fixed here.

## 5.6 Numerical Results and Discussions

In this section, we first validate the efficacy of our analytical results through simulations. We then study the effects of traffic characteristics, namely the burstiness

and correlations, to the MAC layer performance. We focus on the temporal performance metrics, *i.e.*, frame service time and waiting time, while other metrics such as throughput and efficiency can be readily obtained.

The traffic arrival process is modeled by the two-state MMPP, which has been widely used as a building block for the construction of various multimedia sources such as voice, video and Internet traffic [82, 83, Sec. IV-A]. The use of the two-state MMPP model also enables simple and explicit forms of important parameters that facilitate our demonstration. A two-state MMPP is characterized by the infinitesimal generator $\mathbf{Q} = [q_{ij}]$, given by

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{bmatrix}, \tag{5.44}$$

a diagonal matrix $\mathbf{\Lambda}$ of the Poisson arrival rates given by

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \tag{5.45}$$

and the initial probability vector

$$\boldsymbol{\pi_0} = \frac{1}{\lambda_1\sigma_2 + \lambda_2\sigma_1} \begin{bmatrix} \lambda_1\sigma_2 & \lambda_2\sigma_1 \end{bmatrix}. \tag{5.46}$$

The steady-state probability vector $\mathbf{\Pi}$ is given by

$$\mathbf{\Pi} = (\pi_1, \pi_2) = \frac{1}{\sigma_1 + \sigma_2}(\sigma_2, \sigma_1), \quad \lambda_a = \frac{\sigma_1\lambda_2 + \sigma_2\lambda_1}{\sigma_1 + \sigma_2}. \tag{5.47}$$

The mean arrival rate $\lambda_a$ is given as

$$\lambda_a = \frac{\sigma_1\lambda_2 + \sigma_2\lambda_1}{\sigma_1 + \sigma_2}. \tag{5.48}$$

In addition, two parameters are used to describe the burstiness and the correlation of the arrival process. The burstiness is characterized by the squared coefficient of

Table 5.3: Parameters used in the performance evaluation

| | | | |
|---|---|---|---|
| Channel rate | 110 Mbps | Retry limit $[m]$ | 7 |
| Slot time $[\sigma]$ | 9 $\mu s$ | Max. backoff stage $[m']$ | 6 |
| SIFS | 10 $\mu s$ | Min. contention window | 32 |
| PHY header | 13.125 $\mu s$ | Frame payload | 500 bytes |

variation of the interarrival time, $\mathsf{c}^2$, as defined in (5.1) and has the following form [84]

$$\mathsf{c}^2 = 1 + \frac{2\sigma_1\sigma_2(\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^2(\lambda_1\lambda_2 + \lambda_1\sigma_2 + \lambda_2\sigma_1)}. \tag{5.49}$$

The one-step correlation coefficient, $\mathsf{r}_1$, is used to describe the correlation between interarrival times, as given by [84]

$$\mathsf{r}_1 = \frac{\lambda_1\lambda_2(\lambda_1 - \lambda_2)^2\sigma_1\sigma_2}{c^2(\sigma_1 + \sigma_2)^2(\lambda_1\lambda_2 + \lambda_1\sigma_2 + \lambda_2\sigma_2)^2}, \tag{5.50}$$

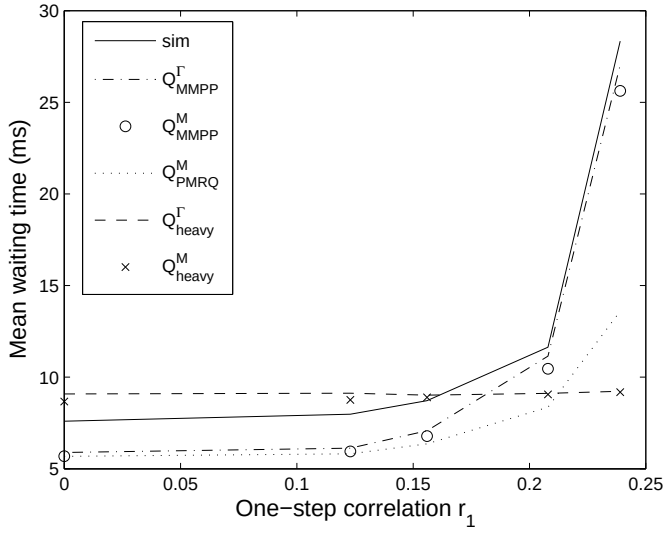
where $t_n$ denotes the $n$th interarrival time. Based on the interrelation between $\mathsf{c}^2$ and $\mathsf{r}_1$, we can generate the arrival processes with the same mean arrival rate but different bursty/correlation characteristics, as suggested in [84]. In our experiments, we fix $\lambda_1 = 1$ and find the corresponding MMPP parameters $(\sigma_1, \sigma_2)$ as a function of $\lambda_2$ from (5.48) and (5.49). Subsequently, the relation between $\mathsf{r}_1$ and $\lambda_2$ can be obtained by (5.50). The value of $\mathsf{c}^2$ is chosen from $\{2, 10, 20\}$, representing different degrees of burstiness. It is reported in [85] that $\mathsf{c}^2 = 18.1$ is very large compared to that of a Poisson process which has a $\mathsf{c}^2$ value of 1.0. The corresponding correlation $\mathsf{r}_1$ is then obtained as long as the inequality $\lambda_i > \sigma_i$ is satisfied.

The PCA protocol in [19] is simulated using our event-driven simulator. All the numerical results reported here are obtained based on the PHY and MAC parameters listed in Table 5.3. Both RTS/CTS handshake and the contention free burst (CFB) functionality [19] are disabled. Because of the space limitation, we fix the minimum contention window size for all ACs and only report the results relevant to the impact of AIFS. In all experiments we consider the following setting: the number of stations

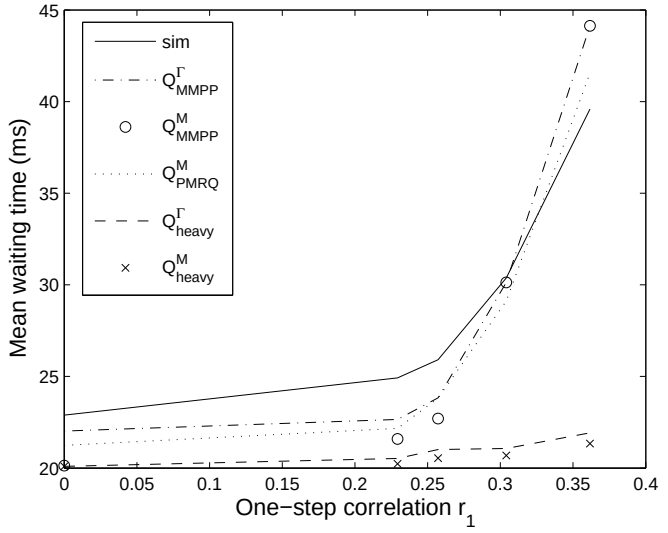
$N_1 = N_2 = 5$; each $AC_1$ station carries a traffic flow driven by the two-state MMPP with the same parameters $\mathsf{r_1}$ and $\mathsf{c^2}$; $AC_2$ stations are saturated such that there are always frames in their MAC buffers. Such a setting mimics the scenario where the station carrying multimedia traffic has higher priority, and the traffic delivered by other stations is considered as the background traffic with low priority. We are interested in the mean waiting time and mean service time of the high-priority multimedia traffic.

### 5.6.1 Model Validation

To verify the efficacy of our analysis, we compare the mean waiting time obtained from simulations and those obtained from the aforementioned approximation methods as summarized in Table 5.2. We fix the mean arrival rate of $\lambda_a = 0.6$, and consider two burstiness levels: $\mathsf{c^2} = 2$ and $\mathsf{c^2} = 10$. Figure 5.3(a) displays the result of a low bursty case, with $\mathsf{c^2} = 2$. The simulation results show that the mean waiting time tends to increase as the one-step correlation increases, and a rapid increase can be found for $\mathsf{r_1} > 0.2$. It can be seen that $Q^{\Gamma}_{MMPP}$ and $Q^{M}_{MMPP}$ can reasonably capture this increasing trend, while the $Q^{\Gamma}_{MMPP}$ slightly outperforms the $Q^{M}_{MMPP}$. The $Q^{M}_{PMRQ}$ performs similar to the previous two approximations for low and medium correlation $\mathsf{r_1}$, and loses its accuracy for high correlation range. The heavy-traffic approximation performs close to the simulation results for low and medium $\mathsf{r_1}$, but the flat curve indicates that this approximation cannot properly reflect the impact of correlation (here the server utilization factor $\rho$ is about 0.7). Figure 5.3(b) displays the results of higher bursty traffic, with $c^2 = 10$. Similar to the low-bursty case, the $Q^{\Gamma}_{MMPP}$ and $Q^{M}_{MMPP}$ well approach the simulated mean waiting time curve for all ranges of correlation $\mathsf{r_1}$. The $Q^{M}_{PMRQ}$ performs very close to the previous two approximations in this setting. Again, the heavy-approximation does not effectively reflect the impact




(a) Low bursty traffic $c^2 = 2$.




(b) Highly bursty traffic $c^2 = 10$.

Figure 5.3: Comparisons of mean waiting time obtained from simulations and approximated queueing systems (listed in Table 5.2) in different degrees of burstiness for $\lambda_a = 0.6, M = 2$.

of correlation in interarrival times. The above results suggest that the $Q^{\Gamma}_{MMPP}$ and $Q^{M}_{MMPP}$ can capture the impact of traffic characteristics to the mean waiting time with reasonable accuracy. In particular the former performs slightly better than the latter. For brevity, in what follows we will only report the results for the gamma approximation.

**Remark**: It can be found that the $Q^{M}_{PMRQ}$ is also effective in responding to the effect of bursty/correlation in the arrival process. Although the $Q^{M}_{PMRQ}$ tends to underestimate the mean waiting time for low and medium levels of correlation, the use of *GI* arrival approximation does help in reducing the computational burden. Its inaccuracy should be due to the relatively lossy peakedness function obtained from the exponential service time approximation.

### 5.6.2 Mean Service Time

We first consider the saturation case, where each station is backlogged with constant bit rate traffic. In Figure 5.4 we show the delay ratio ($Z_2/Z_1$) under different $M$, where $\text{AIFSN}_1 = 2$. The accordance of analysis results with simulation ones suggests that our model is fairly accurate. It can be seen that, the delay ratio increases as $M$ increases. The reason is clearly, since larger $M$ implies longer graceful period the high-priority AC could observe and in turn longer *pre-backoff waiting periods* the low-priority AC may incur. The increasing trend is much significant for larger number of flows. Figure 5.5 reports the per flow throughput for $N_1 = N_2 = 10$, which indicates the low-priority $\text{AC}_2$ is particularly sensitive to the increase of $M$ than the high-priority $\text{AC}_1$. As $M$ changes from 1 to 6, the throughput loss of $\text{AC}_2$ is about 780%, and the throughput increase of $\text{AC}_1$ is about 136%.

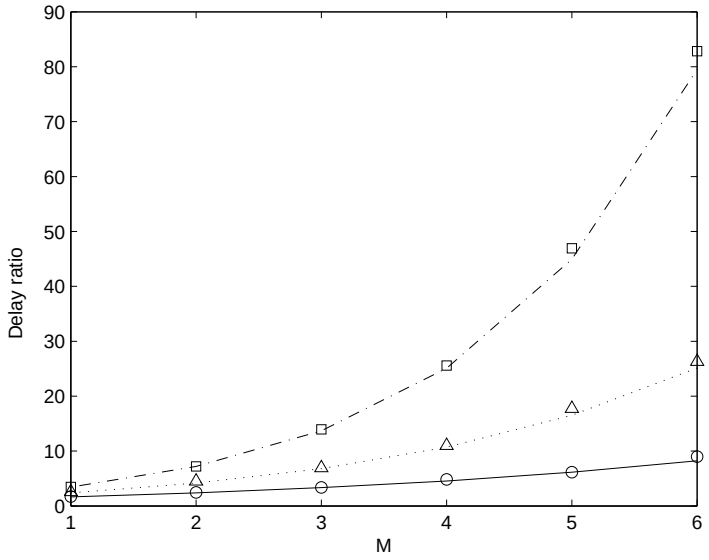



Figure 5.4: Delay ratio $Z_2/Z_1$ *vs.* $M$ for saturated stations.

### 5.6.3 Burstiness/Correlation vs. Mean Waiting Time

To further explore this performance characteristic, we present the results of different traffic densities, *i.e.*, $\lambda_a = 0.3$ in Figure 5.6(a) and $\lambda_a = 0.6$ in Figure 5.6(b), respectively. Comparing the effects of burstiness $\mathsf{c}^2$ and correlation $\mathsf{r}_1$, both figures show that, for low and medium correlation $\mathsf{r}_1$, the traffic burstiness dominates the mean waiting time. For highly correlated traffic, the mean waiting time grows exponentially, which implies that the correlation $\mathsf{r}_1$ between interarrival times has stronger effects on the mean waiting time. These results confirm the importance of taking into account the second-order statistics (*e.g.*, burstiness/correlation) of the multimedia traffic in estimating the mean waiting time. According to our simulation results, on the other hand, the mean service time is not sensitive to the burstiness/correlation properties of interarrivals. Hence the assumption of Poisson arrival process, which

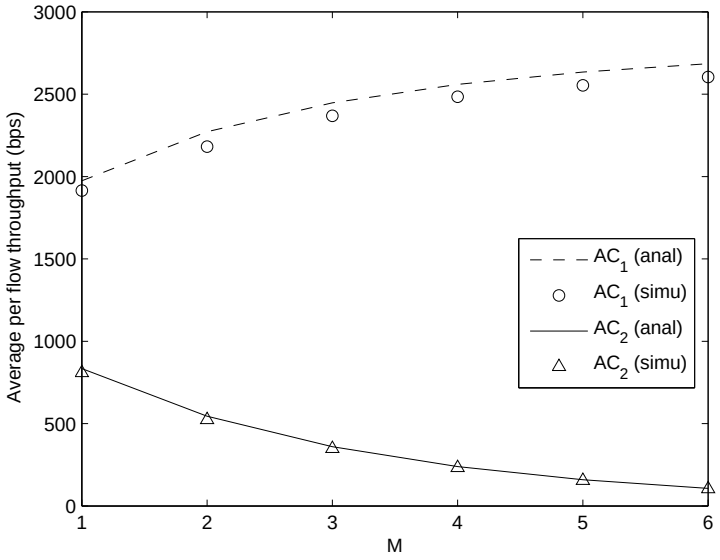
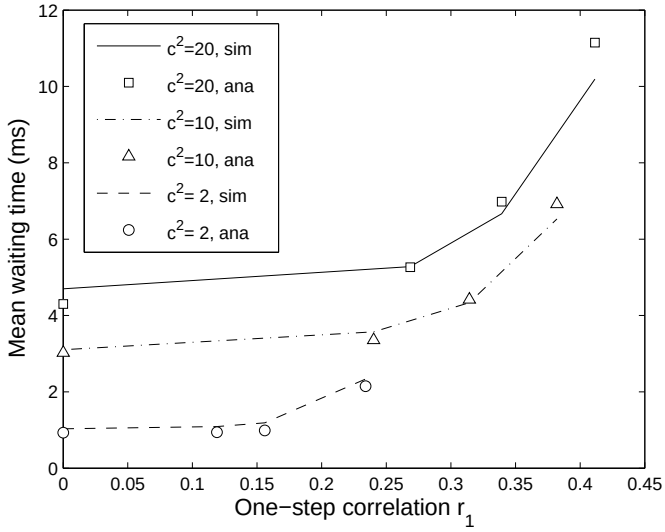
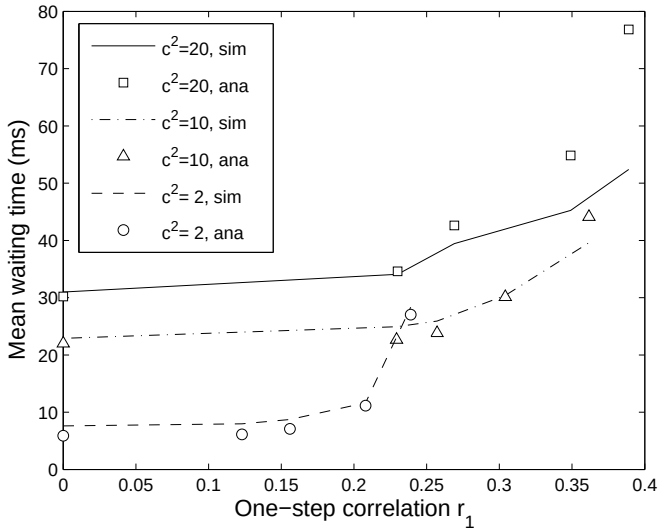



Figure 5.5: Per flow throughput for saturated stations.

has $\mathsf{r}_1 = 0$ and $\mathsf{c}^2 = 1$, is reasonably valid for obtaining the mean service time estimation. However, this assumption greatly underestimates the mean waiting time of the incoming traffic with bursty/correlated arrivals, and thus compromises its usage in evaluating the multimedia traffic performance. For instance, video traffic generally has strict delay bound, where a video frame may become useless if it cannot arrive at the decoding buffer in time. Proactively dropping the video frame that has high probability of exceeding the deadline has been an effective approach to improve the video quality and bandwidth utilization in wireless transmissions [86]. In this context, an accurate estimate about the mean frame waiting time can assist in designing an effective transmission policy.




(a) Low traffic load $\lambda_a = 0.3$.

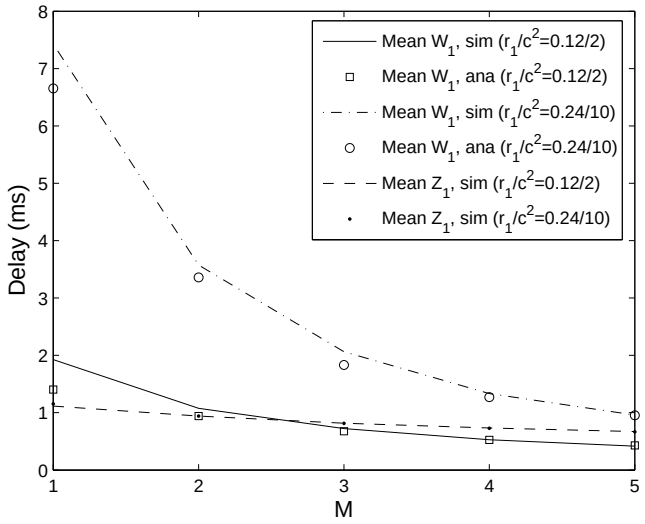



(b) High traffic load $\lambda_a = 0.6$.

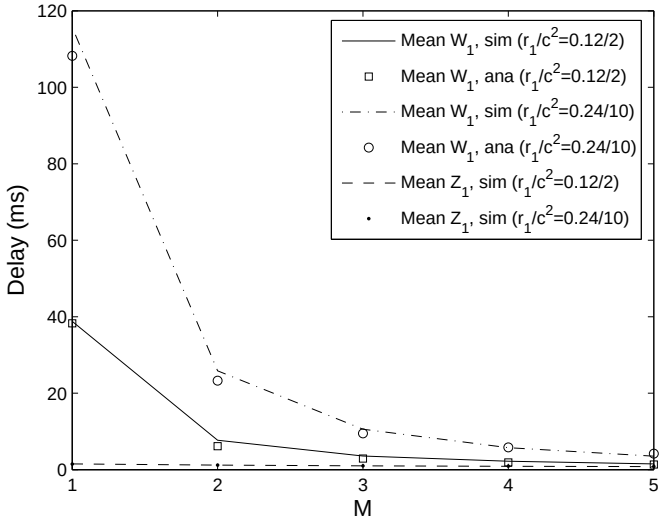
Figure 5.6: The impact of burstiness and correlation in interarrival times to the mean waiting time of $AC_1$ in different traffic loads for $M = 2$.

### 5.6.4 Impact of AIFS

We show the impact of AIFS on the mean waiting time in Figures 5.7(a) and 5.7(b) for $\lambda_a = 0.3$ and 0.6, respectively. The label $M$ of the $x$-axis represents the difference between $\text{AIFS}_1$ and $\text{AIFS}_2$, and a larger $M$ provides more protections to the $\text{AC}_1$ transmissions. We compare the mean waiting time $W_1$ of $\text{AC}_1$ resulting from two scenarios: low-correlated/bursty interarrivals (*i.e.*, $\mathsf{r_1} = 0.12$, $\mathsf{c^2} = 2$) and high correlated/bursty interarrivals (*i.e.*, $\mathsf{r_1} = 0.24$, $\mathsf{c^2} = 10$). We also report the mean service time $Z_1$ of $\text{AC}_1$, obtained from simulations, to demonstrate the effect of AIFS differentiation. We have the following observations. 1) The descending trend in both figures shows that, although setting a larger $M$ can help to reduce the mean waiting time of $\text{AC}_1$, yet the achieved gain is most significant when $M$ is increased from 1 to 2 and its strength is reduced for larger $M$. On the other hand, the results in [61] have shown that, increasing $M$ could remarkably degrade the throughput of low-priority $\text{AC}_2$ stations while the increase of $\text{AC}_1$'s throughput is minor. Hence this gross observation suggests that a conservative setting of AIFS should be considered in differentiating the transmission opportunity of high-priority traffic from low-priority traffic. 2) Traffic with higher bursty/correlation levels is more sensitive to the AIFS differentiation. Take the low traffic load case in Figure 5.7(a) for example, $W_1$ drops by about 87% from $M = 1$ to $M = 5$ for the high correlated/bursty scenario, while the reduction is by about 78% for the low correlated/bursty scenario. 3) The effect of AIFS is magnified for higher traffic load. Consider the highly correlated/bursty interarrivals case for instance. Increasing $M$ from 1 to 2 yields about 52% decrease in mean waiting time $W_1$ for $\lambda_a = 0.3$ [Figure 5.7(a)], while it is about 78% for $\lambda_a = 0.6$ [Figure 5.7(b)]. The above observations indicate that dynamically changing the contention parameters should be beneficial to improving QoS provisioning for multimedia traffic using the PCA protocol.




(a) Low traffic load $\lambda = 0.3$.

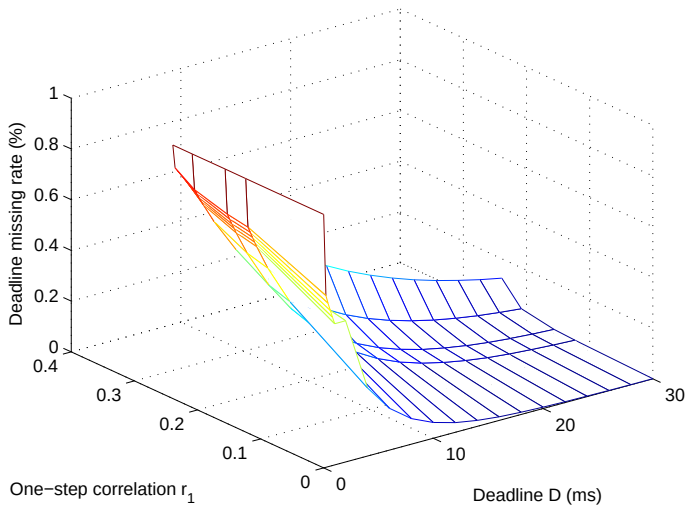



(b) High traffic load $\lambda = 0.6$.

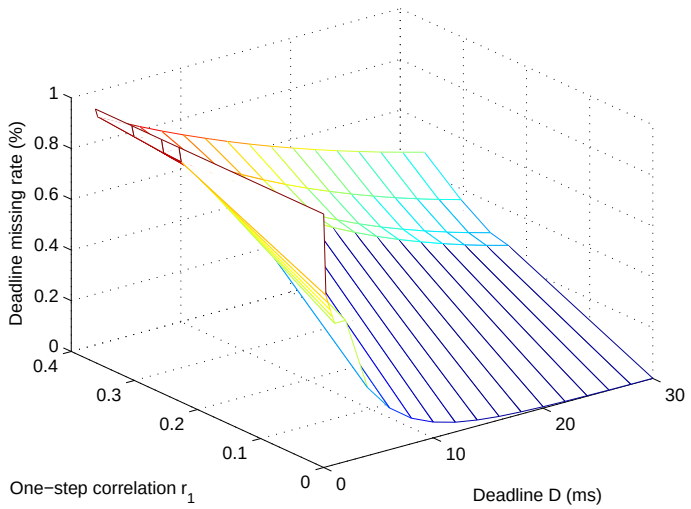
Figure 5.7: The impact of AIFS on mean waiting time for $N_1 = N_2 = 5$.

### 5.6.5 A Potential Application

Finally we present a potential application of our analysis. For multimedia traffic sensitive to delay, the deadline missing ratio (DMR) is a useful temporal metric in characterizing the QoS provisioning. DMR is defined as the probability that the frame waiting time in the MAC buffer exceeds a predefined deadline, i.e., $\mathbb{P}[W > D]$. Direct computation of this tail probability is generally difficult, since the exact waiting time distribution may not exist in explicit form. Alternatively, we can adopt the approximation of (5.37), as suggested in [51], to obtain the analytical value of DMR, by taking advantage of its adequate accuracy in most cases as we have discussed above. Here we fix $M = 2$ and $\lambda_a = 0.6$, and vary the degrees of burstiness and correlation. The results are shown in Figures 5.8(a) and 5.8(b) for $\mathsf{c}^2 = 2$ and $c^2 = 10$, respectively. For the low bursty traffic $\mathsf{c}^2 = 2$, we can see that the DMR drops quickly for $D$ less than 10 ms, and the tail becomes quite flat for larger $D$, since the correlation $\mathsf{r}_1$ has minor impact on the mean waiting time when the traffic load is light. For highly bursty traffic, as shown in Figure 5.8(b), not only the DMR is higher compared to that of a low correlated one, but the correlation $\mathsf{r}_1$ also has a dramatic impact on the DMR. Furthermore, if we compare the DMR curve for $\mathsf{r}_1 = 0$ in both figures, we can find that they are nearly the same. However, as the correlation $\mathsf{r}_1$ increases, the DMR surface of the high-burstiness traffic ($\mathsf{c}^2 = 10$) is clearly different from that of low-burstiness traffic ($\mathsf{c}^2 = 2$). For the high-burstiness traffic, a deadline (say $D = 10$) which is sufficient to ensure low DMR for the low-correlated interarrivals is not applicable to the high-correlated interarrivals, where additional protections such as smaller minimum contention window and longer TXOP may be cooperatively used with AIFS to ensure a desired low DMR.




(a) Low bursty traffic $c^2 = 2$.




(b) High bursty traffic $c^2 = 10$.

Figure 5.8: Dead line rate (DMR) vs. one-step correlation ($r_1$) for different degrees of burstiness, for $N_1 = N_2 = 5, M = 2, \lambda_a = 0.6$.

## 5.7 Summary

We have presented a simple yet accurate model for performance study of the distributed PCA protocol in the WiMedia MAC specification. The main accomplishments of this chapter can be summarized as follows:

- We successfully modeled the complex contention behavior between prioritized stations and derived the PGF of service time distribution, which is then used to obtain the service time distribution. We used this result to find important parameters (the first two moments of the service time distribution) for the queueing analysis.

- The traffic arrival process is modeled by MMPP, which is general enough to represent a wide variety of multimedia streams with bursty and correlated characteristics in interarrivals. We solved the resulting queueing system by three approaches, including the exact MMPP arrival process, its *GI* counterpart, and the asymptotical heavy-traffic approximation, as summarized in Table 5.2.

- We comprehensively studied the performance of these three approaches. Although none of these methods is clearly the best in all cases, the $Q^{\Gamma}_{MMPP}$ and $Q^{M}_{MMPP}$ approximations provide reasonable accuracy and adequately reflect the impact of burstiness/correlation in interarrivals. In particular, our results indicate that the $Q^{M}_{PMRQ}$ comparable accuracy as the exact MMPP modeling in most cases with much less computation effort.

Through numerical results, we have several important observations. For example, the effect of AIFS tends to be magnified when the traffic load is high, or interarrivals are highly bursty and correlated. Since increasing AIFS exponentially reduces the chance of low priority users to access the channel, the parameter AIFSN used to

control AIFS may be chosen moderately as long as the high priority user is well protected. Our analysis also suggests that, dynamically adjusting the contention parameters in response to the traffic characteristics and the network condition may need to be considered to support delay-sensitive multimedia traffic. This provides an important guideline for configuring UWB networks in supporting multimedia traffic with stringent delay requirement. How to choose appropriate contention parameters according to the locally collected information should deserve further research.

Table 5.4: List of local notations for Chapter 5

| **Notation** | **Definition [unit]** |
|---|---|
| $\mathsf{c}^2$ | squared coefficient of variation of the arrival process |
| $m$ | retry limit |
| $m^{'}$ | maximum retry limit |
| $m_1(m_2)$ | first (second) moment of the service time |
| $p_i$ | average frame collision probability of $\mathrm{AC}_i$ |
| $B_i$ | amount of time attributed to backoff for a successful transmission of an $\mathrm{AC}_i$ frame [$\phi$] |
| $\mathsf{r}_1$ | one-step correlation coefficient of the arrival process |
| $x_i$ | number of generic slots in backoff stage $i$ |
| $C$ | number of transmission trials |
| $CW_i$ | contention window size of $\mathrm{AC}_i$ |
| $N_U(C)$ | number of pre-backoff segments given $C$ transmission trials |
| $Q$ | number of pre-backoff waiting periods in one segment |
| $\mathbf{Q}$ | infinitesimal generator of MMPP |
| $R_i$ | amount of time attributed to transmission trials for a successful transmission of an $\mathrm{AC}_i$ frame [$\phi$] |
| $T_s$ | successful transmission time |
| $U$ | overall pre-backoff period [$\phi$] |
| $\overline{W}$ | frame mean waiting time |
| $X$ | interarrival time process |
| $\mathsf{Z}_i$ | contention zone $i$ |
| $Z_i$ | frame service time of $\mathrm{AC}_i$ |
| $\eta$ | length of one pre-backoff period |
| $\theta_j$ | frame transmission probability in zone $j$ |
| $\lambda_a$ | average frame arrival rate |
| $\rho$ | server utilization factor |
| $\sigma$ | time slot length [ms] |
| $\tau_i$ | frame transmitting probability of $\mathrm{AC}_i$ in a random slot |
| $\phi$ | generic slot length |
| $\mathbf{\Lambda}$ | Poisson arrival rate matrix of MMPP |

# Chapter 6

# Distributed Reservation Protocol for UWB Networks with Shadowing Effect

## 6.1 Introduction

To enable ad hoc connectivity in UWB networks, the WiMedia specification defines two distributed channel access mechanisms, the prioritized channel access (PCA) and the distributed reservation protocol (DRP). The PCA is a contention-based protocol, which provides differentiated channel access via the same differentiation mechanisms as the enhanced distributed channel access (EDCA) in IEEE 802.11e. Devices need to sense the idle channel before accessing the medium, and various class-dependent contention parameters are used to prioritize traffic classes. It is known that QoS guarantee in EDCA is achieved in a statistical sense, which is difficult to satisfy the delay requirement of real-time traffic [87]. The detailed analysis of PCA can be found

in Chapter 5.

The DRP is a distributed time division multiple access (TDMA) protocol, by which users have exclusive right of transmission during the reserved time slots. Different from the ordinary TDMA, negotiation of channel time is carried out between peer users through distributed negotiation procedure with the aid of beacon exchange[14]. Such a reservation based protocol provides guaranteed channel access and thus is preferable to real-time traffic with stringent delay requirement. However, real-time traffic generally reveals bursty characteristics, where the packet interarrival times are highly non-uniform. Consequently, the discrepancy between the reserved bandwidth and the dynamic traffic arrival rate leads to low channel utilization. This problem can be alleviated by using DRP with soft reservation, which means the unused slots can be accessed by other users using the access rule of PCA. In hard reservation only the owner of the reservation access the medium.

While the centralized TDMA protocol and its variants have been thoroughly studied [7–9], the distributed slot reservation problem has been less explored. The main difference of a distributed reservation protocol from its centralized counterpart lies in the resulting reservation pattern. To reduce the delay variation, it is desired to reserve contiguous time slots per scheduling cycle, which is possible with the aid of the centralized coordinator. However, maintaining such a uniform reservation pattern (uniform interval between two reservations) is difficult in a distributed environment, where the available time slots within a scheduling cycle may reveal an arbitrary pattern. In [88], the delay performance of DRP under a given reservation pattern is analyzed using a two-dimensional embedded Markov chain based on the assumption

[14]The beacon issued by each user allows distributed network synchronization and the correct operation of the network. Beacon collisions are possible and can be resolved by a beacon collision resolution protocol (BCRP) specified in the standard.

of Poisson arrivals. Although the model is straightforward, it is limited to deal with the hard reservation only and is not extendable to real-time traffic with bursty arrivals because the PASTA (Poisson arrivals see time average) principle is not valid. The assumption of error-free wireless channel also lead to underestimate the effect of time-varying wireless channel to the delay performance, which have been noticed in many recent studies [89–91].

In indoor environments, users with UWB devices are more likely to be stationary while nearby persons may frequently walk through the line-of-sight (LOS) of an ongoing link resulting in short-term shadowing. For UWB communications with several hundred megabits per second data rate, even a half second shadowing can affect more than hundred megabits data and thus degrade the quality of the ongoing flows severely, even with guaranteed channel time using DRP.

In this chapter, we propose an analytical model to study the performance of DRP, taking into account the arbitrary reservation pattern and time-varying UWB channel. In indoor environments, the time-variation of wireless channels is generally caused by two sources: movement of the transmitter (Tx) or the receiver (Rx) (or both), and/or the movement of scatters. For typical UWB indoor applications, the Tx and Rx are stationary, thus the time-variation in the received signal power is caused by the shadowing rather than fast fading. Based on the measurement results suggested by the standardization group, the time-varying UWB channel with shadowing is modeled by a finite-state discrete-time Markov chain at the packet level. In addition to the time-varying channel behavior, another factor that impacts the protocol performance is the reservation method. Given a reservation pattern, we are interested in the impact of the reservation method to the DRP performance. The proposed analytical model is developed based on the vacation queueing model, where the interval between two reservation period is deemed the vacation period of the tagged user. Generally such

a queueing system is complex and difficult to solve. By using the well-established matrix-geometric approach, we are able to solve the complex queueing system and obtain several important performance metrics such as mean service time, waiting time, and throughput. Through numerical results, we get insight into the performance of DRP in a realistic environment and provide useful guidelines in improving the protocol. To the best of our knowledge, this is the first work analyzing standardized UWB MAC protocols considering the time-varying UWB channel.

The remainder of this chapter is organized as follows. Section 6.2 presents protocol description and modeling method for DRP, followed by the packet-level UWB shadowing channel model. We formulate and solve the queueing system in Section 6.3. In Section 6.3 we derive various performance metrics for both hard and soft reservations. Numerical results and discussions are given in Section 6.5. The chapter is concluded in Section 6.6.

## 6.2 Preliminaries

The system under investigation is a distributed UWB WPAN where users reserve transmission times using DRP. Because of user movement, an ongoing link may be frequently shadowed off leading to performance degradation of DRP. Our goal is to quantitatively assess the protocol performance under shadowing effect when different reservation methods are used. This section first gives an overview of DRP and explains how we model the protocol behavior given a specific reservation pattern. Then we present the methodology of modeling the UWB shadowing channel at the packet level.

### 6.2.1 Overview of DRP and its modeling

In WiMedia MAC, the channel time is divided into superframes composed of 256 media access slots (MAS). The superframe consists of a beacon period (BP) and a data transfer period (DTP), as depicted in Figure 2.3. In DTP, a number of MASs should be retained for PCA traffic. The remaining MASs are available for reservation via DRP. To reserved MASs, a user needs to first sends a request indicating the set of MASs of interest to its receiver. The request can be encoded in the beacon frame or sent via PCA or DRP. Upon receiving the request, the receiver checks the availability of the intended MASs according to its locally collected information. The reservation can be successfully made only when resources are available at both the sender the receiver sides. If the reservation is accepted, the receiver announces the result in its beacon. Consequently, other users become aware of the reservation and the sender can start the transmission from the next superframe using the reserved MASs. The reservation can be reserved in a hard or soft manner. By *hard* reservation, the channel is occupied by the owner throughout the reserved MASs. If *soft* reservation is used, the unused MASs can be accessed by other users using PCA rule. The reservation type is indicated in the beacon frame.

A tagged user is said to be in *service* during its reserved time slots, otherwise it is said to be on *vacation*. Therefore, the vacation time of a tagged user is equivalent to the service periods of other users, the beacon time and the PCA period in a superframe. Since the latter two periods are simply constant, we pay more attentions on the service period. The length of a service period depends on the reservation method. When hard reservation is used, the service time duration is identical to the reservation period duration, which is deterministic. On the other hand, the service time duration may be less than or equal to the reservation period, depending on the buffer status.

In principle, the vacation period can be described by the discrete PH type distributions. To elaborate this, we first briefly review the discrete PH type distribution. Generally, a discrete PH distribution characterizes the time until absorption into state $K$ in a Markov chain on the state space $\{0, \cdots, K\}$ with initial probability vector $(\overrightarrow{\eta}, \eta_K)$, where $\eta_K = 1 - \overrightarrow{\eta}\,\overrightarrow{1}$ with $\overrightarrow{1}$ denoting a column vector of 1's, and transition matrix structured by

$$\begin{bmatrix} \mathbf{V} & \overrightarrow{V^0} \\ \overrightarrow{0} & 1 \end{bmatrix}, \tag{6.1}$$

where $\overrightarrow{V^0} = \overrightarrow{1} - \mathbf{V}\overrightarrow{1}$. Thus knowledge of $\overrightarrow{\eta}$ and $\mathbf{V}$ is sufficient to specify the distribution of a PH distribution, denoted by $PH_d(\overrightarrow{\eta}, \mathbf{V})$ with order $K$. The discrete PH type distribution is general to represent every discrete distribution with finite support. For example,a geometric distribution with parameter $p$ can be represented by $PH_d(1, 1-p)$ with order 1.

If the service period of individual user is of discreet PH distribution, then the vacation period of the tagged user, which is the sum of the service periods of other users, can also be represented by the convolution of several PH distributions. For the case of hard reservation, the reservation period of each user is deterministic, which is a special case of the discrete PH distribution. For soft reservation, the reservation period is stochastic, depending on the queue status of the tagged user. Once the buffer of the tagged user is empty, the channel becomes idle and thus other user can immediately access the channel using PCA. In this case, the tagged user queue can be modeled by vacation queue with a time-limited policy. Such a vacation queue has been studied in [92], where the service period under the time-limited vacation policy is shown to be of discrete PH type. Notice if we further consider the beacon period and the time allocated to the PCA in each superframe, the resulting vacation time remains PH type distribution, considering the facts that the deterministic variable is

a special case of discrete PH type distribution, as well as the closure property of PH type distribution. In the attempt to study the interaction between the time-varying channel and the queueing performance, we focus on modeling the joint behavior of these two dynamics and assume that the vacation time distribution of the arbitrary reservation pattern is known. The derivation of vacation period distribution can be found in Appendix D.

### 6.2.2 Packet-Level Channel Model for UWB Shadowing

In an indoor environment, the moving obstacles, such as people, may frequently penetrate the LOS of an ongoing wireless link, thus shadow off the most significant power contribution. For UWB communications with several hundred megabits per second data rate, even a half second shadowing due to moving obstacles can affect more than hundred megabits data and thus degrade the quality of the ongoing flows severely.

To evaluate the actual protocol performance for UWB applications, a channel model which captures the multipath and time-varying characteristics of UWB propagation is a crucial component in the entire system modeling. The IEEE standardization group has suggested the UWB indoor channel models for system test and validation. However, such a waveform simulation needs to simulate the transmission of every bit over the wireless channel, which is time-consuming for network protocol analysis. Rather, a packet-level channel model which provides the packet error rate (PER) information according to statistical channel properties can greatly facilitate a complex system modeling. In [93], a packet-level channel model for UWB shadowing due to user movement is proposed. In the following, we briefly introduce how to obtain the UWB shadowing channel model and the corresponding Markov chain model at the packet level.

There are four types of channel model based on different measurement environments. Channel model 1 (CM1) describe a LOS scenario with a separation between Tx and Rx of less than 4 m. CM2 and CM3 describe the NLOS scenario with Tx-Rx distance $1-4$ m and $4-10$ m, respectively. CM4 describes an environment with strong delay dispersion. Based on the UWB channel model suggested in IEEE 802.15.3a [94], we first generate the channel IR without shadowing. Normally 100 random channel realizations should be sufficient, and they are dumped into a trace file for further processing. Following the steps described in [93], we can obtain the average shadowed received power given a specific location of the obstacle.

The effect of shadowing can be estimated based on the angular power spectrum (APS), which has been used in [95] to study the time-varying UWB channel impulse response (IR) with the stationary transmitter/receiver and the LOS blocked by a moving obstacle. The APS refers to the distribution of power over the incident angle. For a UWB channel, multipath components arrive with clustering phenomenon, where each cluster consists of a random number of rays with different power delay profile [96]. Therefore, the APS of UWB signal should be a function of delay. Based on the practical measurement, the APS of UWB channel denoted by $\mathbb{P}(\theta,\tau)$ has been given in [95],

$$\mathbb{P}(\theta,\tau)=\begin{cases}\text{rect}(\frac{2\pi\tau}{\tau_{\max}}), & 0<\tau<\tau_{\max}\\ \frac{1}{2\pi}, & \tau>\tau_{\max}\end{cases} \tag{6.2}$$

where $\theta$ is the incident angle with respect to the LOS component and $\tau$ is the delay of the tap. Based on the APS of UWB propagation, we can obtain the shadowed channel impulse response (IR). Figure 6.1 shows a simple scenario suggested in [97] for measuring the shadowing effect of a single obstacle. A person, which is emulated as a cylinder with radius of $r$, moves through the LOS between the transmitter and the receiver at an average speed of $v$. The moving path is assumed to be perpendicular

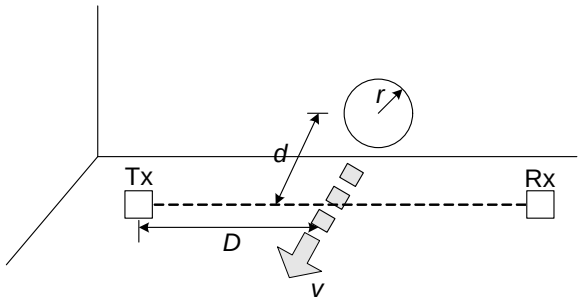



Figure 6.1: Illustration of blocking scenario.

to the LOS, with a distance of $D$ from the receiver. Based on this scenario, the basic steps to estimate the shadowing effect are summarized as follows:

1) Create a random channel realization, with a certain power ascribed to each tap. The CM1 model reported in [94] is used because it describes the scenario of LOS.

2) For each delay tap, compute $\mathbb{P}(\theta, \tau)$.

3) Place the obstacle at a certain location on the moving path. Perform geometric computations to get the angular range that the obstacle is shadowed off.

4) The remaining power for each delay tap is calculated, and thus a new channel IR can be obtained.

The above procedure generates the normalized received power in the presence of the obstacle at the prescribed position. By allocating the obstacle at the position along the moving path, we can draw a curve representing the average shadowed received power due to a moving obstacle. This will need to repeat the entire procedure with sufficient spatial sampling intervals. To facilitate the computation, the result generated in step 1 can be dumped into a file for further reuse.

Our goal is to construct a Markov chain for describing the dynamics of the UWB shadowing channel at the packet level, based on the shadowed channel IR. The idea is similar to the finite-state Markov chain (FSMC) channel model widely used in

modeling the Rayleigh fading channel, where the Markov channel model is characterized by two statistics, the state transition probability and the state-dependent error rate. First, we divide the area near the LOS into $N$ zones. The spatial partition is symmetric with respect to the LOS, resulting in a Markov chain with a state space $\{x, x = 0, \cdots, 2N\}$, where state 0 corresponds to the spatial zone without shadowing. The average duration that the channel state remains at state $n$, $t_n$, is decided by the size of the zone and the average mobility speed $v$. In the corresponding discrete-time Markov Chain model, each channel state will last for $N_n = t_n/t_s$ slots, where $t_s$ is the slot duration, e.g., packet transmission time. Figure 6.2 illustrates how the average received SNR curve is partitioned and the resulting Markov chain. Assuming that the time of a person staying within each zone is geometric distributed with mean $N_n$, the transition probability of the discrete-time Markov chain can thus be approximated as

$$h_{x,y} = \begin{cases} \lambda_0, & x = 0, y = 1, \\ 1 - \lambda_0, & x = y = 0, \\ \frac{1}{N_x+1}, & x \in [1, \cdots, 2N], y = mod(x + 1, 2N + 1) \\ \frac{N_x}{N_x+1}, & x = y \in [1, \cdots, 2N], \\ 0, & \text{otherwise}, \end{cases} \tag{6.3}$$

where $\lambda_0$ is the arrival rate of a obstacle, determined by the density and mobility of the obstacle entering the shadowed area, and *mod* represents the modulo operation. Note that the Markov channel model can be formulated as either discrete or continuous ones. Here we use a discrete one to facilitate our queueing analysis.

To obtain the average error rate of each state, we approximate the average received SNR of state $n$ by $(\Gamma_n + \Gamma_{n+1})/2$, where $\Gamma_n$ corresponds to the SNR at the boundary of zone $n$. Given the average received SNR, $\gamma_b$, the PER associated with state $n$, $\theta_n$, is calculated by $\theta_n = 1 - (1 - BER_n)^L$, where $BER_n = Q(\sqrt{2\gamma_b})$ for antipodal

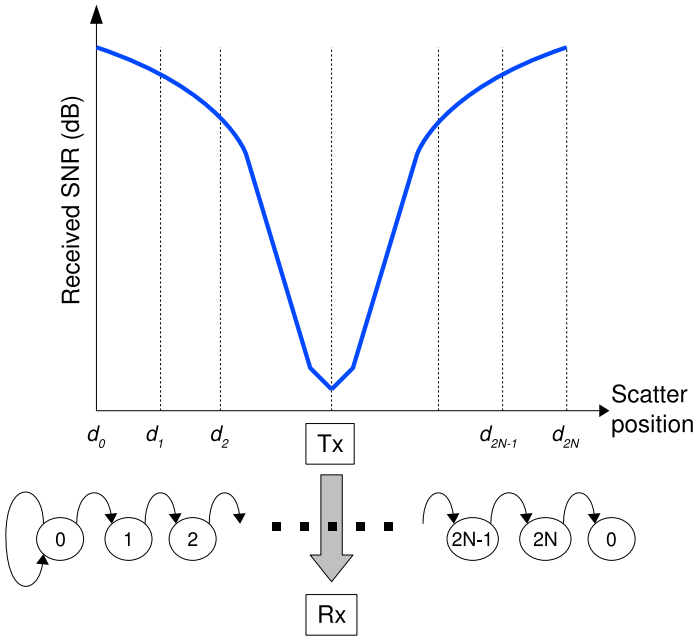



Figure 6.2: Partition method of the Markov channel model.

signals [98].

## 6.3 Queueing Model and Analysis

To capture the evolution of the joint service/vacation and channel process with the use of different reservation methods, we develop the analytical model based on the following assumptions and notations.

1. We consider a fixed reservation pattern. A user can reserve time slots using either hard or soft reservation, but not both. The mixed use of both reservations is not considered.

2. A reservation period contains $S$ slots. For simplicity, at most one packet can be successfully transmitted per slot.

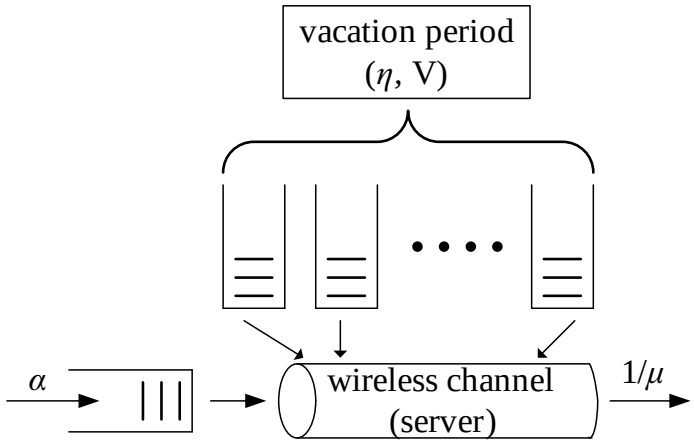



Figure 6.3: Illustration of the queueing model.

3. An infinite buffer size is assumed. The term "buffer" and"queue" will be used interchangeably.

4. The tagged user's queue is modeled as a discrete-time queue with vacation, as depicted in Figure 6.3. The length of vacation period $V$, consisting of multiple service periods of other users, follows the discreet PH type distribution denoted by $PH_d(\overrightarrow{\eta}, \mathbf{V})$ with order $K$, where the parameters $\overrightarrow{\eta}$ and $\mathbf{V}$ can be derived using the approach shown in Appendix D.

5. A packet may arrive at the beginning of the slot with probability $\alpha$, and leave at the end of the slot (*i.e.*, successfully transmitted) depending on the channel quality, represented by the average received SNR $\gamma_b$.

6. The UWB channel variation due to shadowing is captured by a discrete-time Markov chain with transition probability matrix $\mathbf{H} = [h_{x,y}]$ in (6.3).

7. The average service rate is denoted by $1/\mu$, where $\mu$ is the average service time.

The analysis of DRP protocol over UWB shadowing channel is conducted by modeling the joint behaviors of queue length, channel state, and service/vacation period as a discrete-time Markov chain $\chi$ with state space $\{(i,(0,k,x)\cup(s,x)), i \geq 0; k=1,\cdots,K; x=1,\cdots,N; s=1,\cdots,S\}$, where $i$ is the number of packets in the queue during vacation (service); in the tuple $(0,k,x)$, 0 represents the tagged user is in the vacation period, $k$ represents the phase of the vacation, $x$ is the channel state; the tuple $(s,x)$ refers to the service state with $s$ denoting the slot index within each allocation initiated by one. By ordering the state space of the Markov chain lexicographically, the corresponding transition probability matrix $\mathbf{P}$ describing the number of packets in the buffer can be obtained and partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{A_0} & \mathbf{C_0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \cdots \\ \mathbf{B_0} & \mathbf{A_1} & \mathbf{C_1} & \mathbf{0} & \mathbf{0} & \cdots & \cdots \\ \mathbf{0} & \mathbf{B_1} & \mathbf{A_1} & \mathbf{C_1} & \mathbf{0} & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B_1} & \mathbf{A_1} & \mathbf{C_1} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \tag{6.4}$$

where each component is a block matrix describing the evolution of the queue length. It is observed that $\mathbf{P}$ has the exact geometric structure as the QBD process [99]. In this context the matrix-geometric solution can be readily used to solve $\mathbf{P}$. To differentiate the underlying reservation pattern, the Markov chain corresponding to hard and soft reservation is denoted by $\chi_h$ and $\chi_s$, respectively. Meanwhile, the matrix blocks in $\mathbf{P}$ depends on the use of reservation method, thus we use $\mathbf{P}_h$ and $\mathbf{P}_s$ to denote the transition probability matrices for hard reservation and soft reservation, respectively.

We note that the above Markov chain model is not limited to the Bernoulli arrival process. For example, a more general Markovian arrival process (MAP) can

be represented by two substochastic matrices[15] and incorporated in the embedded Markov chain $\chi$ [92]. In addition, the consideration of one successful transmission per slot implies the QBD process has one-step memory, as can be observed from (6.4). Generalization to multiple transmissions per slot only leads to multi-step memories in the QBD process. The model can be also extended to the finite buffer case using the approach in [100].

Before deriving the matrix blocks in $\mathbf{P}_h$ and $\mathbf{P}_s$, we introduce some auxiliary variables. We use $i$ and $j$ to denote the queue length, $k$ and $l$ for the vacation phase, $x$ and $y$ for the channel state. Meanwhile, the following matrix shall be frequently used,

$$\mathbf{\Theta} = \text{diag}(\theta_1, \theta_2, \cdots, \theta_{2N}), \quad \mathbf{\Theta}^{'} = \mathbf{I}_{2N+1} - \mathbf{\Theta}, \tag{6.5}$$

and $\mathbf{I}_{2N+1}$ is an identity matrix of dimension $2N+1$. In addition, $\overrightarrow{1_x}$ is a column vector of zeros except the $x$th position being one.

### 6.3.1 Markov Chain for DRP with Hard Reservation

For DRP with hard reservation, the tagged user persistently occupies the service period regardless the buffer status. Each block in $\mathbf{P}_h$ is a square matrix of order $(2N+1)(K+S) \times (2N+1)(K+S)$ as detailed in the following.

$\mathbf{A_0}$: The queue length remains zero in the following situations.

a) During vacation period, the system state switches from $\{0,(0,k,x)\}$ to $\{0,(0,l,y)\}$ with probability $(1-\alpha)v_{k,l}h_{x,y}$. The transition probability of dimension $(2N+1)K \times (2N+1)K$ is given as

$$\mathbf{A_0^3} = (1-\alpha)\mathbf{V} \otimes \mathbf{H}, \tag{6.6}$$

[15] A matrix with the sum of each row less than or equal to one is called a substochastic matrix; it is a stochastic matrix if the sum of each row is equal to one.

where $\otimes$ denotes the Kronecker product.

b) The transition from vacation period to service period, i.e., from state $\{0,(0,k,x)\}$ to $\{0,(s,y)\}$ occurs with probability $(1-\alpha)v_k^0 h_{x,y}$ for $s=1$, otherwise it is zero. The transition probability matrix of dimension $(2N+1)K \times (2N+1)$ for the former case can be represented as

$$\mathbf{A_0^2} = (1-\alpha)\overrightarrow{V^0} \otimes \mathbf{H}. \tag{6.7}$$

c) The transition during service period only occurs from state $\{0,(s,x)\}$ to $\{0,(s+1,y)\}$ for $0 \le s \le S-1$ with probability $(1-\alpha)h_{x,y} + \alpha(1-\theta_x)h_{x,y}$, which can be written in matrix form of dimension of $(2N+1)\times(2N+1)$ as

$$\mathbf{A_0^0} = (1-\alpha)\mathbf{H} + \alpha\mathbf{\Theta}'\mathbf{H}. \tag{6.8}$$

d) At the end of service period, the system state $\{0,(S,x)\}$ may transit to $\{0,(0,k,y)\}$ with probability $(1-\alpha)\eta_k h_{x,y} + \alpha(1-\theta_x)\eta_k h_{x,y}$. Thus we have the block matrix of dimension $(2N+1)\times(2N+1)K$,

$$\mathbf{A_0^1} = (1-\alpha)\,\overrightarrow{\eta} \otimes \mathbf{H} + \alpha\,\overrightarrow{\eta} \otimes \mathbf{\Theta}'\mathbf{H} \tag{6.9}$$

The resultant transition probability matrix which describes the queue remaining zero in two consecutive slots has the following structure:

$$\mathbf{A_0} = \begin{bmatrix} \mathbf{A_0^3} & \mathbf{A_0^2} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{A_0^0} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \mathbf{A_1^0} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \mathbf{A_0^0} \\ \mathbf{A_0^1} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \end{bmatrix}. \tag{6.10}$$

Using the similar argument, we can obtain $\mathbf{A_1}$ with same structure as $\mathbf{A_0}$ but

different components given by

$$\mathbf{A_1^3} = (1-\alpha)\mathbf{V} \otimes \mathbf{H},\ \mathbf{A_1^2} = (1-\alpha)\overrightarrow{V^0} \otimes \mathbf{H}, \tag{6.11}$$

$$\mathbf{A_1^1} = (1-\alpha)\overrightarrow{\eta} \otimes \mathbf{\Theta H} + \alpha\overrightarrow{\eta} \otimes \mathbf{\Theta}^{'}\mathbf{H}, \tag{6.12}$$

$$\mathbf{A_1^0} = (1-\alpha)\mathbf{\Theta H} + \alpha\mathbf{\Theta}^{'}\mathbf{H}. \tag{6.13}$$

In addition, $\mathbf{C_0}$ is given by

$$\mathbf{C_0^3} = \alpha\mathbf{V} \otimes \mathbf{H},\ \mathbf{C_0^2} = \alpha\overrightarrow{V^0} \otimes \mathbf{H}, \tag{6.14}$$

$$\mathbf{C_0^1} = \alpha\overrightarrow{\eta} \otimes \mathbf{\Theta H},\ \mathbf{C_0^0} = \alpha\mathbf{\Theta H}. \tag{6.15}$$

The transition probability matrix $\mathbf{B_1}$, corresponding to the queue length decreasing by one, has the following structure

$$\mathbf{B_1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B_1^0} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \mathbf{B_1^0} \\ \mathbf{B_1^1} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \end{bmatrix}, \tag{6.16}$$

with the components given by

$$\mathbf{B_1^0} = (1-\alpha)\mathbf{\Theta}^{'}\mathbf{H},\ \mathbf{B_1^1} = (1-\alpha)\overrightarrow{\eta} \otimes \mathbf{\Theta}^{'}\mathbf{H}. \tag{6.17}$$

Notice $\mathbf{B_0}$ is identical to $\mathbf{B_1}$ in hard reservation case.

**Stationary Probability**

If the Markov chain $\chi_h$ is positive recurrent, *i.e.*, the stability condition holds, the matrix equation

$$\mathbf{R} = \mathbf{C}_0 + \mathbf{R}\mathbf{A}_1 + \mathbf{R}^2\mathbf{B}_0 \tag{6.18}$$

has the minimum nonnegative solution, $\mathbf{R}$, with all the eigenvalues in the unit disk. The stationary distribution $\overrightarrow{\pi} = [\overrightarrow{\pi_0}\ \overrightarrow{\pi_1}\ \overrightarrow{\pi_2}\ \cdots]$ can be expressed as the matrix geometric form

$$\overrightarrow{\pi_k} = \overrightarrow{\pi_0}\mathbf{R}^k, \quad k \geq 0, \tag{6.19}$$

where $\overrightarrow{\pi_k}$ is the stationary probability of $k$ packets in the queue. From $\overrightarrow{\pi}\mathbf{P} = \overrightarrow{\pi}$ we can obtain a set of linear equations

$$\overrightarrow{\pi_0}(\mathbf{A_0} + \mathbf{RB_0}) = \overrightarrow{\pi_0}. \tag{6.20}$$

Furthermore, the normalization condition $\overrightarrow{\pi}\overrightarrow{1} = 1$ can be reduced to

$$\overrightarrow{\pi_0}(\mathbf{I} - \mathbf{R})^{-1}\overrightarrow{1} = 1. \tag{6.21}$$

The stationary probability $\overrightarrow{\pi_0}$ can be obtained by jointly solving (6.20) and (6.21).

### 6.3.2 The Markov Chain for DRP with Soft Reservation

The main difference of soft reservation from hard reservation is, whenever the tagged users buffer becomes empty, the unused time slots in the allocation section will be used by other users following the PCA access rule. In other words, the tagged user enters vacation period if the service period ends or the queue becomes empty, whichever happens first. Consequently, the system behavior of soft reservation differs from that of hard reservation when it involves the queue length being zero, namely, $\mathbf{A_0}$, $\mathbf{B_0}$ and $\mathbf{C_0}$ in $\mathbf{P_s}$.

$\mathbf{A_0}$: The system remains on state 0, i.e., the queue length equals zeros, only when the server is on vacation and there is no packet arrival. The corresponding transition matrix can be derived as $(1-\alpha)(\mathbf{V} \otimes \mathbf{H})$ before the vacation phase reaches the absorbing state. For the case that the system enters to the absorbing state, the

vacation phase will be renewed, since the queue is still empty that forces the system to proceed with another vacation time. This will happen with probability determined by $\overrightarrow{V^0}$, while the new vacation phase is determined by $\overrightarrow{\eta}$. Overall, the block matrix $\mathbf{A_0}$ with dimension $(2N+1)K \times (2N+1)K$ is given as

$$\mathbf{A_0} = (1-\alpha)(\mathbf{V} \otimes \mathbf{H} + \overrightarrow{V^0}\overrightarrow{\eta} \otimes \mathbf{H}) \tag{6.22}$$

$\mathbf{C_0}$: The queue length may increase from zero to one during the vacation period with corresponding transition matrix $\alpha\mathbf{V} \otimes \mathbf{H}$. It may also occur when a packet arrives at the end of the vacation period with transition matrix $\alpha\overrightarrow{1_1}^T \otimes \overrightarrow{V^0} \otimes \mathbf{H}$ with dimension $(2N+1)K \times (2N+1)S$, where $T$ denotes the matrix transpose. As a result, the block matrix $\mathbf{C_0}$ can be given by

$$\mathbf{C_0} = \left[\alpha\mathbf{V} \otimes \mathbf{H} \quad \alpha\overrightarrow{1_1}^T \otimes \overrightarrow{V^0} \otimes \mathbf{H}\right], \tag{6.23}$$

with dimension $(2N+1)S \times (2N+1)(K+S)$.

$\mathbf{B_0}$: During the vacation period, the probability that the queue length changes from one to zero is zero, and it is $(1-\alpha)(1-\theta_i)\delta_i h_{x,y}$ when the system state switches from $(1,(s,x))$ to $(0,(0,v,y))$. Thus we have the block matrix $\mathbf{B_0}$ of dimension $(2N+1)(K+S) \times (2N+1)K$

$$\mathbf{B_0} = \begin{bmatrix} \mathbf{0} \\ (1-\alpha)\overrightarrow{\eta} \otimes \mathbf{\Theta}'\mathbf{H} \\ (1-\alpha)\overrightarrow{\eta} \otimes \mathbf{\Theta}'\mathbf{H} \\ \vdots \\ (1-\alpha)\overrightarrow{\eta} \otimes \mathbf{\Theta}'\mathbf{H} \end{bmatrix}. \tag{6.24}$$

**Stationary Probability**

Since the number of states for the boundary level is different from that for the non-

boundary levels, the infinitesimal generator $\mathbf{P}_s$ has the complex boundary behavior. In this context, we resort to the following matrix-geometric solution to find the stationary distribution [99, p. 24], which can be expressed as the matrix geometric form

$$\overrightarrow{\pi_k} = \overrightarrow{\pi_1}\mathbf{R}^{k-1}, \quad k \geq 2. \tag{6.25}$$

The above equation together with $\overrightarrow{\pi}\mathbf{P} = \overrightarrow{\pi}$ gives a set of linear equations as

$$\overrightarrow{x}B[\mathbf{R}] = \overrightarrow{x}, \tag{6.26}$$

where $\overrightarrow{x} = [\overrightarrow{\pi_0}\ \overrightarrow{\pi_1}]$, and

$$B[\mathbf{R}] = \begin{bmatrix} \mathbf{A_0} & \mathbf{C_0} \\ \mathbf{B_0} & \mathbf{A_1} + \mathbf{RB_1} \end{bmatrix}. \tag{6.27}$$

Since $B[\mathbf{R}]$ is singular, we need other independent equations, which can be obtained from the normalization condition $\overrightarrow{\pi}\overrightarrow{1} = 1$ and can be reduced to

$$\overrightarrow{\pi_0}\overrightarrow{1} + \overrightarrow{\pi_1}(\mathbf{I} - \mathbf{R})^{-1}\overrightarrow{1} = 1. \tag{6.28}$$

Then we can solve (6.26) and (6.28) to find $\overrightarrow{\pi_0}$ and $\overrightarrow{\pi_1}$. Other $\overrightarrow{\pi_i}, i \geq 2$ can be obtained from (6.25).

## 6.4 Analysis for Hard and Soft Reservation

### 6.4.1 Stability

The stability condition of the vacation queue has been given by Alfa [92] for time-limited polling system. The queueing system modeled in this chapter is analogous to that in [92], where the system is stable under the following condition.

**Proposition 6.1.** *The system is stable if*

$$\alpha(S+\overline{V})\mu < S. \tag{6.29}$$

*Proof.* Define one cycle as the duration between the end of one vacation period to the end of next vacation. Because the user can occupy the channel for at most $S$ slots regardless the use of reservation scheme, the average length of one cycle time is $S+\overline{V}$. During one cycle period, the average number of arrivals is $\alpha(S+\overline{V})$. Given the average service time $\mu$, the average time to serve the arrivals during one cycle is $\alpha(S+\overline{V})\mu$, which has to be less than the maximum service duration $S$, for the system to be stable. □

### 6.4.2 Mean Service Time

Define packet transmission process from the time that a packet becomes the head-of-line (HOL) of the buffer to the time it is successfully received by the receiver. The service time is thus the duration of one packet transmission process. In this subsection we derive the average service time of a packet. To simplify the analysis, we assume that the channel quality remains the same during the packet transmission period. This is justified by the fact that channel variation process due to shadowing is much slower than the packet transmission process. We will verify this assumption in Section 6.5 by simulations.

The average service time can be derived as follows. Suppose a packet becomes the HOL at slot $s$ when the channel is in state $x$, for $1 \le s \le S$ and $1 \le x \le N$, with probability $p_{\mathrm{H}}(s,x)$. Because a packet may become the HOL at one of the slot in allocation section with equal chance, which is independent of the channel state, we have

$$p_{\mathrm{H}}(s,x) = \frac{1}{NS}. \tag{6.30}$$

Assuming that the channel remains on the same state $x$ during the packet transmission process, the number of time slots consumed until a successful transmission follows a geometric distribution with parameter $\theta_x$. In other words, a packet may undergo $m$ transmission trails for a successful transmission with probability $\theta_x^{m-1}\theta_x$, and it consumes $m + \lfloor \frac{m-1}{S} \rfloor \overline{V}$ slots in total. Therefore, we obtain the average service time as

$$\begin{aligned}\mu &= \sum_{s=1}^{S}\sum_{x=1}^{N} p_{\mathrm{H}}(s,x) \sum_{m=1}^{\infty} \theta_x^{m-1}(1-\theta_x)(m + \lfloor \frac{m-1}{S} \rfloor \overline{V}) \\ &= \frac{1}{NS}\sum_{s=1}^{S}\sum_{x=1}^{N}(\frac{1}{1-\theta_x} + \nu),\end{aligned} \tag{6.31}$$

where $\nu = \sum_{m=1}^{\infty} \theta_x^{m-1}(1-\theta_x)\lfloor \frac{m-1}{S} \rfloor \overline{V}$.

### 6.4.3 Average Throughput

The average throughput, denoted as $\Omega$, is defined as the average amount of successfully transmitted information in bits per time unit, which can be computed as

$$\Omega = \sum_{i=1}^{\infty}\sum_{x=1}^{N}\underbrace{\sum_{s=1}^{S}\pi_{i,s,k}}_{(\star)} c_x(1-\theta_x), \tag{6.32}$$

where $\pi_{i,s,k}$ represents the joint probability of queue length $i$, channel state $x$, and service epoch $s$; $c_x$ represents the maximum number of successful packets per slot given the channel state $x$. We assume $c_x = 1$ for simplicity. In (6.32), $(\star)$ represents the joint probability that there are $i$ packets in queue with channel state being $n$, given by

$$(\star) = \overrightarrow{\pi_i}(\overrightarrow{q} \otimes \overrightarrow{1_x}),\ \overrightarrow{q} = [\underbrace{0\ \cdots\ 0}_{r}\ \underbrace{1\ \cdots\ 1}_{s}]^T. \tag{6.33}$$

Using (6.33), (6.32) can be written as

$$\Omega = \sum_{i=1}^{\infty} \overrightarrow{\pi_i} \sum_{x=1}^{N} (\overrightarrow{q} \otimes \overrightarrow{1_x})(1-\theta_x) = \sum_{i=1}^{\infty} \overrightarrow{\pi_i}\overrightarrow{q} \otimes \hat{\mathbf{\Theta}}, \tag{6.34}$$

where $\hat{\mathbf{\Theta}} = [1-\theta_0 \; 1-\theta_1 \; \cdots \; 1-\theta_{2N}]^T$. For DRP with hard reservation, the average throughput is given by

$$\Omega_h = \overrightarrow{\pi_0}\mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}\overrightarrow{q} \otimes \hat{\mathbf{\Theta}}. \tag{6.35}$$

Similar, the average throughput of DRP with soft reservation can be given as

$$\Omega_s = \overrightarrow{\pi_1}(\mathbf{I}-\mathbf{R})^{-1}\overrightarrow{q} \otimes \hat{\mathbf{\Theta}}. \tag{6.36}$$

### 6.4.4 Mean Waiting Time

For DRP with hard reservation, the average queue length at an arbitrary time, denoted as $L_h$, is derived as

$$L_h = \sum_{k=1}^{\infty} k\overrightarrow{\pi_k}\overrightarrow{1} = \overrightarrow{\pi_0}\mathbf{R}(\mathbf{I}-\mathbf{R})^{-2}\overrightarrow{1}. \tag{6.37}$$

Similarly, for DRP with soft reservation, the average queue length at an arbitrary time, denoted as $L_s$, is given by

$$L_s = \overrightarrow{\pi_1}(\mathbf{I}-\mathbf{R})^{-2}\overrightarrow{1}. \tag{6.38}$$

The mean waiting time, $\overline{w}$, is obtained by Little's law as

$$\overline{w}_h = \frac{L_h}{\lambda}, \quad \overline{w}_s = \frac{L_s}{\lambda} \tag{6.39}$$

## 6.5 Numerical Results and Discussions

This section starts by describing the simulation setup, including the methodology and parameters. The analytic and simulation results are compared to validate the

Table 6.1: Simulation parameters.

| | |
|---|---|
| packet size | 1500 bytes |
| slot length ($t_s$) | 256 $\mu$s |
| body width ($r$) | 30 cm |
| mobility speed ($v$) | 0.5 m/s |
| obstacle arrival rate ($\lambda_0$) | 0.3 |
| obstacle-receiver distance ($d$) | 1 m |
| number of channel states ($N$) | 21 |

mathematical modeling. The interaction between various system parameters and protocol performance will also be discussed.

We first simulate the shadowing scenario of UWB channel depicted in Figure 6.1. Following the procedures described in Section 6.2.2, we obtain the normalized average shadowed power at the receiver end, as shown in Figure 6.4. Table 6.1 lists the relevant parameters. The actual curve is obtained by a spatial sampling interval of 10 cm. We further partition the spatial area into 21 zones, resulting in a discrete-time Markov chain with 21 states, and channel transition matrix $\mathbf{H}$. The received power of each state is approximated by the mean of the boundary values, as indicated by the approximated curve. Based on the obtained packet-level channel model, we simulate the discrete-time system shown in Figure 6.3. The tagged user can access the channel during its reserved allocation section of length $S$ slots, while the interval between two allocation sections is considered as the vacation time with the length determined by a d-PH distribution. The simulation run time is $5 \times 10^6$ time slots. A error-free feedback channel for signaling the acknowledgement is assumed such that the sender can get the reception status of the transmitted packet at the end of the time slot.

In the sequel we demonstrate the analytic results based on the equations presented in the previous section. The accuracy of the analytic model heavily relies on the rate matrix $\mathbf{R}$, which is computed iteratively until $\|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\| \leq 10^{-8}$, where $k$ denotes the iteration index.

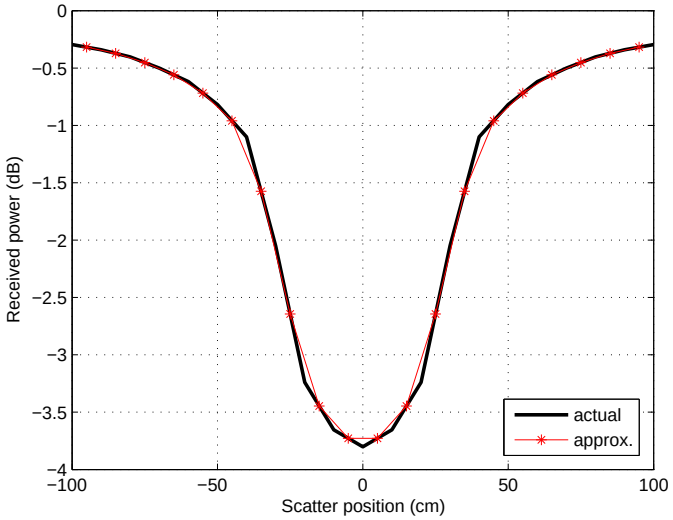
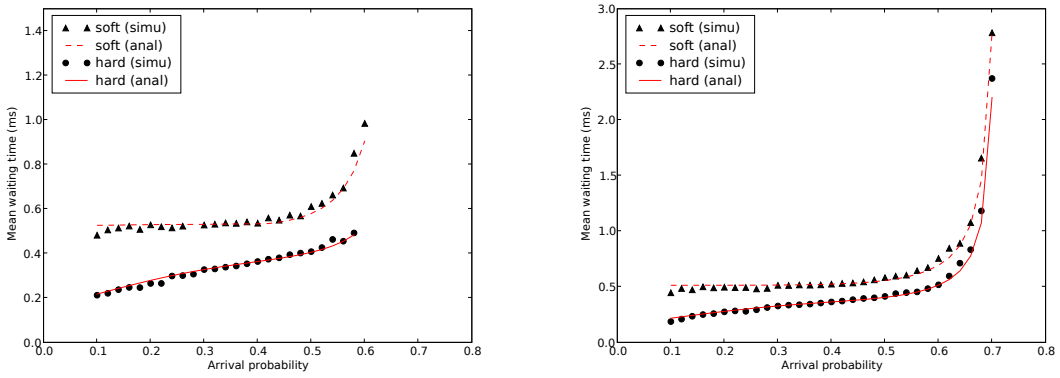



Figure 6.4: Normalized received power shadowed by moving people using parameters specified in Table 6.1.

### 6.5.1 Mean Waiting Time

The mean waiting time of DRP with soft and hard reservation versus arrival probability under different received SNR are shown in Figures 6.5(a) and 6.5(b), where the service duration $S = 7$ slots, and the vacation period $V = 3$ slots. As seen from the figures, the analytical results match the simulation ones very well. The figures also show that, a higher received SNR reduces the mean waiting time as expected. Comparing the performance of DRP with different reservation schemes, DRP with soft reservation incurs higher mean waiting time than hard reservation, and the difference is more significant for lower received SNR. From the network standpoint, soft reservation is advantageous to improving the channel utilization by releasing the idle slots to the neighbors. However, this is achieved with the cost of reducing the possible transmission time of individual user. Hence, there is a tradeoff between the network




(a) Average SNR = 20 dB.

(b) Average SNR = 25 dB.

Figure 6.5: Comparisons of reservation methods: mean waiting time *vs.* average arrival probability for $S = 7$ and $V = 3$.

resource utilization and the QoS of the traffic flow. While configuring the optimal reservation pattern might not be trivial in a distributed network, a rule of thumb may be to use the soft reservation in a conservative way. A user may release the unused slot if the traffic incoming rate slows down for a period of time and the number of contending nodes is large. Otherwise the reserved time slots should be used persistently even the buffer is empty. On the other hand, the two mechanisms perform similarly for high arrival probability region. The reason is obvious, since the queue is seldom being empty when the packet arrives densely. As a result, slot release can rarely take place, implying that soft reservation reduces to hard reservation.

Finally, we show the average delay under different reservation patterns. For illustration purpose, two d-PH distributions $V_1 \sim (\overrightarrow{\eta_1}, \mathbf{V}_1)$ and $V_2 \sim (\overrightarrow{\eta_2}, \mathbf{V}_2)$ with the

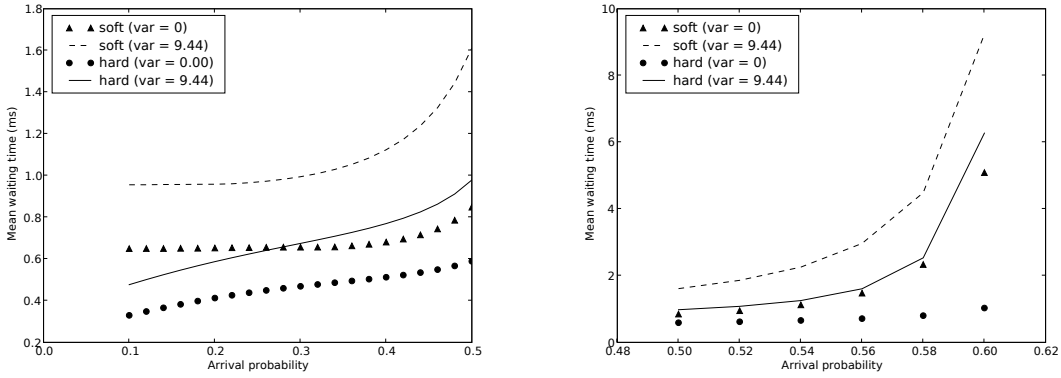
same mean but different variance are used to represent the vacation period:

$$\overrightarrow{\eta_1} = [1,0,0,0],\ \mathbf{V}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \overrightarrow{\eta_2} = [0.4, 0.25, 0.2, 0.15],\ \mathbf{V}_2 = \begin{bmatrix} 0.2 & 0.3 & 0.25 & 0.25 \\ 0 & 0.7 & 0.3 & 0 \\ 0 & 0 & 0.5 & 0.3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.40}$$

and $\overline{V}_1 = \overline{V}_2 = 4$, $\text{var}[V_1] = 0$, $\text{var}[V_2] = 9.44$. The vacation time with a larger variance implies that the reserved pattern is more interleaved. In Figure 6.6, the service duration is 7 slots long, and the average SNR equals 20 dB. For presentation brevity, only analytical results are shown. As seen from Figure 6.6(a) for low-medium packet arrival probability, a larger gap between the two mean waiting time curves of soft reservation indicates that soft reservation is more sensitive to the variance of vacation time. On the other hand, the mean waiting time of hard reservation reveals a larger slope implying that hard reservation is more sensitive to the traffic intensity. Figure 6.6(b) shows that the effect of variance of the vacation time to both schemes tends to be magnified as the traffic load increases. For both reservation methods, a uniformly slot reservation yields lower packet delay than that resulting from a unequally spaced reservation pattern. This observation basically agrees with previous studies on analyzing the slot allocation pattern in TDMA systems [8,9,88].

### 6.5.2 Mean Service Time

The mean service time measures the time that a packet is served by the wireless channel. A higher received SNR provides a lower PER and thus, a packet can be successfully received with less retransmissions. Table 6.2 shows the mean service time of DRP under different channel conditions and vacation distributions. In simulations, we fix the traffic intensity $\alpha = 0.5$, the service period $S = 7$ slots, and vary the received SNR and the vacation distributions following (6.40). It can be seen that the




(a) $\alpha = 0.1 \sim 0.5$. (b) $\alpha = 0.5 \sim 0.6$.

Figure 6.6: Comparisons of reservation pattern: mean waiting time *vs.* arrival probability for $S = 7$, $\overline{V} = 4$, and received SNR = 20 dB.

mean service time linearly decreases as the increase of the received SNR. Moveover, the reservation pattern, whether uniform or interleaved, has nearly null impact to the mean service time. The agreement between the simulation and analytical results validates the approximated derivation of service time in Section 6.4.2, based on the assumptions the channel quality remains the same during the packet transmission period.

### 6.5.3 Average Throughput

This subsection presents the effects of reservation methods and patterns on average throughput. We fix the arrival intensity $\alpha = 0.5$ and service period $S = 7$ slots, and consider two vacation distributions following (6.40), both with the mean vacation period of 4 slots. In Fig. 6.7, the horizontal axis represents the average SNR, and the vertical axis represents the average throughput normalized by the offered load. Both reservation methods achieve similar throughput and reach their performance limit

Table 6.2: Comparison of mean service time (in microseconds) with different vacation distributions.

| SNR (dB) | simulation (var=0) | simulation (var=9.44) | analytical |
|---|---|---|---|
| 12 | 0.4539 | 0.4578 | 0.4924 |
| 14 | 0.4463 | 0.3833 | 0.3929 |
| 16 | 0.3532 | 0.3489 | 0.3511 |
| 18 | 0.3080 | 0.3077 | 0.2999 |
| 20 | 0.2781 | 0.2794 | 0.2699 |
| 22 | 0.2662 | 0.2667 | 0.2611 |

when the average SNR is sufficiently high. On the other hand, the throughput loss due to shadowing is significant for low SNR range. In fact, the average throughput without shadowing effect is close to one even for the received SNR = 12 dB, which is not shown in the figure for presentation brevity.

Both reservation schemes reach its performance limit when the average SNR is sufficiently high. Particularly soft reservation achieves better average throughput than hard reservation. This is because for soft reservation, the user releases the unused time slots whenever the buffer is empty even the reservation period has not been expired. For hard reservation, however, the channel becomes idle when there is no packet to send during the reservation period. As can be seen in Figure 6.7, hard reservation leads to a certain loss on channel efficiency even the average SNR is sufficiently high to produce nearly zero PER. While soft reservation achieves better channel efficiency than hard reservation, the mean waiting time resulted from soft reservation tends to be longer than that of hard reservation, as shown in Figure 6.5. Comparing the results of mean waiting time and average throughput, the loss of channel efficiency is about 20% when hard reservation is used, while the mean packet waiting time of hard reservation is reduced about 50% in comparison with that of soft reservation. This

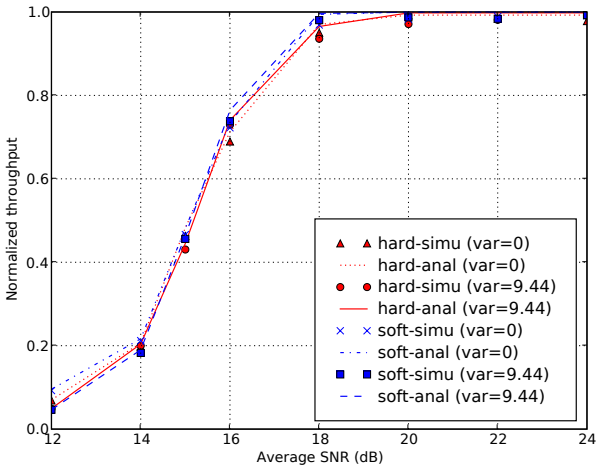



Figure 6.7: Normalized throughput of DRP *v.s.* average SNR with different reservation methods, for $\alpha = 0.5$, $S = 7$, and the vacation periods follow (6.40) with the same mean $\overline{V} = 4$ but different variances.

implies a tradeoff between the channel efficiency and the packet delay, depending on the use of reservation schemes. On the other hand, both performance metrics can be improved by increasing the received SNR, with the cost of higher transmission power, modulation order or coding rate. Alternatively, in multi-band OFDM, where the maximum transmission power is used, time and frequency domain spreading[16] may be used to increase the SNR of received data.

[16]Time-domain spreading is achieved by transmitting the same information across two consecutive OFDM symbols. Frequency-domain spreading entails transmitting the same information (complex number) on two separate subcarriers within the same OFDM symbol.

## 6.6 Summary

In this chapter, we have proposed an analytical model for the DRP defined in WiMedia MAC for UWB-based WPAN. By modeling the sender buffer as a discrete-time queue with vacation and incorporating the dynamics of UWB channel owing to shadowing, we have obtained various important performance metrics under different reservation methods. The proposed analytical model has been validated through simulations, and the interactions between several system parameters have been observed from numerical results. It has been shown that the short-term shadowing effect significantly degrades the protocol performance. Comparing the hard and soft reservations, the soft reservation incurs longer packet delay than that of hard reservation. Both reservation methods achieve similar throughput performance. As to the effect of reservation pattern, the soft reservation is more sensitive to the variation in the reservation pattern. Moreover, the non-uniform reservation pattern is more vulnerable to the increase of traffic intensity. Since the short-term shadowing notably degrades the DRP performance, one of our future works is to find appropriate adaptive strategies to smoothly support isochronous traffic. We also expect that the proposed model can be applied to analyze other upper-layer protocols such as automatic repeat request (ARQ) over UWB channels.

Table 6.3: List of local notations for Chapter 6

| **Notation** | **Definition (unit)** |
|---|---|
| $\overrightarrow{1}$ | column vector of all ones |
| $\overrightarrow{1_x}$ | column vector of all zeros except the $x$th position being one |
| $\alpha$ | packet arrival probability per slot |
| $\overrightarrow{\eta}$ | initial probability vector of the phase type distribution |
| $\gamma_b$ | average received SNR (dB) |
| $\mu$ | average service time |
| $i$ | number of packets in queue |
| $n$ | channel state |
| $r$ | number of vacation phases |
| $s$ | slot index in allocation section |
| $v$ | vacation phase |
| $N$ | number of channel states |
| $S$ | length of allocation section (slot) |
| $V$ | vacation length |
| $\overrightarrow{V^0} = [v_k^0]$ | probability vector describing the transition from a transient state to the absorbing state |
| $\mathbf{V} = [v_{k,l}]$ | transition probability matrix for the transient states of the phase type distribution |

# Chapter 7

# Conclusions and Future Work

In this thesis we have addressed two fundamental questions surrounding UWB and its potential applications: how the bandwidth can be efficiently utilized, and how the QoS requirement of multimedia traffic can be guaranteed. We have tackled these issues using a MAC-centric approach. In the following, we summarize our research results, followed by a discussion of possible future work.

## 7.1 Major Research Contributions

We consider two types of UWB networks, centrally controlled and distributed controlled ones, respectively.

### 7.1.1 Efficient Resource Allocation for Centralized UWB Networks

The inherent wide spectrum of UWB signals allows concurrent transmissions within close proximity. However, arbitrary concurrent transmission may lead to lower aggre-

gate throughput than that of TDMA. Therefore, it is necessary to strategically select the concurrent transmitting links in order to efficiently utilize the bandwidth. We have shown the optimal scheduling for concurrent transmission is NP-hard, and proposed low-complexity algorithms for different design objectives, including throughput maximization and multi-class services provisioning. The main contributions are:

- Deriving the sufficient condition when concurrent UWB transmission yields higher network throughput than TDMA scheduling. Based on this condition, low-complexity scheduling algorithms for improving the network throughput in UWB networks have been proposed.

- Proposing a utility-based scheduling algorithm which can provide differentiated services among different traffic classes with different bandwidth requirements. The proposed solution also achieves fairness in the presence of heterogeneous traffic.

- Taking advantage of UWB's ranging capability in aiding the scheduling decision. While the perfect distance information is commonly assumed, we resort to stochastic optimization taking into account the ranging errors. We have proposed a meta-heuristic algorithm incorporating the sufficient condition (for concurrent transmission) to solve the NP-hard scheduling problem. Simulation results show that the proposed approach is promising in improving the algorithm efficiency with affordable computational complexity.

### 7.1.2 Modeling and Analysis for MAC protocols in Distributed UWB Networks

Despite several distributed MAC protocols have been proposed for UWB networks, they usually require a significant amount of signalings and are difficult to implement. On the other hand, the WiMedia Alliance has included two simple distributed MAC protocols, namely PCA and DRP, in their specification for future UWB-based WPANs. Our goal is to understand the capabilities of these two protocols in supporting multimedia applications for UWB networks such that possible improvement guidelines can be obtained. To achieve this, we mathematically model and analyze these two protocols.

**Contention-based PCA**

While many analytical models have been proposed for theoretically analyzing contention-based MAC protocols, rarely the characteristics of multimedia traffic are taken into account. In addition, analyzing MAC protocols with prioritized traffic classes is usually difficult. Since multimedia traffic, which reveals bursty and correlated interarrival times, will be the key application of UWB networks, it is important to develop a suitable analytical model taking into account the traffic characteristics of multimedia traffic. We have proposed an analytical model for studying the performance of PCA, considering the bursty and correlated interarrival times of multimedia traffic. This is the first work exposing the significance of the second-order characteristics (burstiness/correlation) of multimedia traffic on the delay performance of the PCA protocol. The advantages of the proposed analytical model are:

- It can accurately capture the contention behaviors between high and low priority classes. Comparing to the Markov-chain based models, the proposed model is simple and more realistic.

- Both the average frame service time and frame waiting time can be obtained.

Based on the proposed analytical model, we have obtained several important observations: (1) The frame waiting time is very sensitive to the second-order characteristics (burstiness and correlation) of traffic arrival process. On the other hand, the frame service time is not sensitive to bursty/correlation properties of interarrival times; (2) The effect of differentiation parameter AIFS in protecting the high priority traffic tends to be magnified when the traffic load increases, and/or the interarrival times are more bursty and correlated; and thus, (3) the value of AIFS should be moderate, as long as the high priority multimedia traffic can be protected. The above observations suggest that, dynamically adjusting the contention parameters should be beneficial to improving the QoS provision for multimedia traffic.

**Reservation-based DRP**

To satisfy the stringent delay requirement of multimedia traffic, the WiMedia MAC employs a reservation-based protocol called DRP to achieve bandwidth guarantee. However, several factors may impact the effectiveness of DRP in bandwidth guarantee. Firstly, the nearby persons may frequently walk through the LOS of an ongoing link leading to short-term shadowing. With a high speed transmission rate, even a half-second shadowing can affect a significant amount of data. Secondly, the DRP is different from the centralized counterpart, as the available time slots is non-uniformly distributed over the superframe. Thirdly, the reservation method (*i.e.*, soft and hard), introduces a tradeoff between channel efficiency and delay performance. We have proposed an analytical model for analyzing the performance of DRP considering the above factors. Our main achievements are:

- The proposed analytical model is the first one taking into account the time-varying behavior of the UWB multipath channel in analyzing MAC protocols.

- The proposed analytical model is unified in the sense that it can capture the exact protocol behavior under different reservation methods (soft and hard), given a specific reservation pattern. With minor modifications, various traffic models can be incorporated in the analytical model.

- The proposed model provides important insights into the protocol performance in realistic UWB networks, and can be applied to the analysis of other upper-layer protocols over UWB channels, such as TCP, ARQ.

We have observed that the short-term shadowing significantly degrades the protocol performance. Comparing the hard and soft reservations, the soft reservation incurs longer packet delay than that of hard reservation. Both reservation methods achieve similar throughput performance. As to the effect of reservation pattern, the soft reservation is more sensitive to the variation in the reservation pattern. Moreover, the non-uniform reservation pattern is more vulnerable to the increase of traffic intensity. One of the most interesting finding is that, the intuitive advantage of soft reservation in improving the channel utilization is not obvious according to our numerical results. In fact, soft reservation tends to introduce longer average waiting time that can be a negative impact to multimedia traffic sensitive to the interarrival time.

## 7.2 Future Work

In this section, we briefly outline some potential research directions from this thesis.

### 7.2.1 Interference-aware MAC Protocols for UWB Networks

- *Inter-piconet interference* – In the first part of this thesis, we assume there is no inter-piconet interference in centralized UWB networks. The inter-piconet

interference may take place when multiple geographically overlapped piconets use the same channel and are not synchronized in time. The effect of inter-piconet interference can be severe to users locating within the overlapping area. If PNCs belonging to different WPANs can hear each other, it is possible for overlapped piconets to achieve synchronization in MAC layer, and further merge their superframes [1]. When piconets are partially overlapped, PNCs might be hidden from each other. The issue may be tackled from the physical layer to reduce the symbol collision probability [39]. How to extend our work taking the possible inter-piconet interference into account should deserve further studying.

- *Interference mitigation through Cognitive UWB Radio* – Since UWB signals are spread over a wide spectrum, they unavoidably overlap with existing radio systems such as global system for mobile communication (GSM), universal mobile communication system (UMTS), and WLANs. How to mitigate interference from/to these coexisting systems is a critical issue for UWB networks. Moreover, the inter-piconet interference and the shadowing effect discussed in Chapter 6 require proper adaptive mechanisms to reduce their effect to UWB networks. Cognitive radio technology allows devices to rapidly adapt their transmission schemes (*e.g.*, modulations and protocols) to the environment [101]. Despite the original objective of cognitive radio is to improve spectrum efficiency, its application to resolve the above-mentioned issues in UWB networks is an interesting research issue.

### 7.2.2 Low-complexity Algorithms for Utility Maximization with Heterogeneous Traffic

In Chapter 4, bandwidth allocation in centralized UWB networks is formulated as the network utility maximization (NUM) problem. This formulation has found many applications in solving networking problems, such as TCP congestion control [102], bandwidth allocation in wireless networks [44] and wireline networks [103], as well as video streaming [104]. The main difficulty of NUM lies in the possible non-concavity of the objective function. Previous research mainly focuses on continuous and concave utility functions for elastic traffic. Very recently, discontinuous and nonconcave utility functions have received attentions due to the explosion of interest in multimedia traffic (*e.g.*, real-time, delay-adaptive, or streaming traffic). The consideration of nonconcave utility functions, however, turns NUM into nonconvex optimization problems [105]. Since exact solutions for nonconvex optimization problems are difficult, current research attempts to find in what conditions the traditional approach for solving the concave NUM problems can be still applied to nonconcave ones. In this thesis aiming at WPAN, the solution space is relatively small such that search-based approaches may be applicable. Envisioning the increasing demand of providing multimedia services in future wireless networks, *e.g.*, WiMAX, it is critical to design low-complexity algorithms for solving nonconcave NUM problems. Additionally, fair bandwidth allocation for networks with heterogeneous traffic also requires further investigation.

### 7.2.3 Packet-level UWB channel model

To evaluate the protocol performance, both mathematical modeling and simulation/emulation are widely accepted approaches. Comparing to the latter one, an

accurate mathematical model can greatly facilitate the early stage of system design in reducing the time and cost of building the system of interest. In the context of wireless communications, the close interactions between the time-varying channel behavior due to multipath fading and upper-layer protocols have been evidenced from previous research. Examples include TCP [89, 90], auto repeat request (ARQ) [28, 106], and link scheduling [91, 107]. In the literature, the finite state Markov channel (FSMC) has been widely used to characterize the variation of the flat fading channel. For UWB channels with frequency-selectivity, there is a lack of convenient channel model for characterizing both the fast and slow fading. To describe the channel variation at packet-level, we need certain statistical information about the time-varying channel, *e.g.*, the amplitude distribution of the channel gain and/or the level crossing rate of the fading process. Such information is typically obtained from measurement results. Currently only limited channel measurements specific for UWB propagation channels have been conducted. In Chapter 6 we develop an analytical model based on the measurements from a specific scenario, where a single obstacle that moves across the LOS is considered [97]. How to extend this model to include more generalized scenarios, *e.g.*, different mobility patterns (speed and direction) and multiple obstacles require further research.

# Appendix A

# NP Hardness of Optimal Concurrent Scheduling

We show that the optimal scheduling for concurrent transmissions that optimizes throughput is NP-hard using the notion of conflict graph [34]. Consider a UWB peer-to-peer network as a directed graph $G = (V, E)$ where the nodes correspond to the UWB devices and the links correspond to the peer-to-peer links between the devices (e.g., a directed link $l_{ij}$ from device $i$ to $j$). We use the terms "node" and "link" in reference to the directed graph $G$, and reserve the terms "vertex" and "edge" for the *conflict graph* $H$. Under graph $G$, the conflict graph $H$ is defined such that a vertex corresponds to the link in the graph $G$. An edge, denoted as $(l_{ij}, l_{pq})$, is drawn if the links $l_{ij}$ and $l_{pq}$ can not be scheduled concurrently. According to the exclusive region with radius $r$, there is an edge between vertex $l_{ij}$ and $l_{pq}$ if either $d_{iq} \leq r$ or $d_{pj} \leq r$. Given the conflict graph $H$, an independent set is a set of vertices, such that there is no edge between any two of the vertices. In other words, an independent set consists of the links that can be scheduled concurrently. If each link has identical capacity, the problem of finding the largest independent set in $H$, known as NP-hard,

can be reduced to the throughput maximization problem. Since the link throughput depends on the propagation distance and shadowing effect and thus is not identical, the optimal scheduling leading to maximum throughput can be seen as a weighted version of the aforementioned throughput optimization problem. This results in the hardness of optimal scheduling problem in peer-to-peer UWB networks.

To solve the maximum weighted independent set problem, several approximation algorithms based on greedy strategy have been proposed [108, 109]. In [109], the proposed greedy algorithms can achieve $1/\triangle(H)$ of the weight of the maximum independent set, where $\triangle(H)$ is the maximum degree of the graph. However, such approximation may be loose for a dense network. In addition, these algorithms rely on iteratively configuring the graph until it is empty, and thus are considered computation-expensive for complete graphs in our case.

# Appendix B

# Derivation of $G_U(z)$

To derive the PGF of the pre-backoff period $G_U(z)$, consider the fact that when the tagged $AC_2$ station is backing off in $\mathsf{Z}_2$, the pure backoff procedure is interrupted if any other stations transmit, which occurs with probability $1-\gamma$, where $\gamma = (1-\tau_1)^{N_1}(1-\tau_2)^{N_2-1}$. Hence, when the tagged station experiences $x_i$ backoffs in stage $i$, there are $(1-\gamma)\sum_{i=1}^{C-1} x_i$ interruptions or *segments*, given that there are $C$ transmission trials before one successful transmission. In each segment, the backoff counter can be decremented only when $\mathsf{Z}_1$ is idle with probability $\theta_2$. With probability $1-\theta_2$, one or more $AC_1$ stations may transmit in any of the slots in $\mathsf{Z}_1$ and the current uncompleted $\mathsf{Z}_1$ is ended immediately due to this transmission. Therefore, for each segment, the tagged $AC_2$ station may experience several *pre-backoff waiting periods* (PBOW) prior to the pure backoff. We denote $Q$ as the number of PBOWs in each segment. Since $Q$ itself is a geometric random variable with parameter $\theta_2$, to simplify the analysis, we let $Q = 1/\theta_2$. As a result, the overall number of PBOWs for one successful transmission of $AC_2$ station (with $C$ transmission trials), denoted as $N_U(C)$, is equal

to $N_U(C) = (1-\gamma)/\theta_2 \sum_{i=1}^{C-1} x_i$, and its PGF is given by

$$G_{N_U(C)}(z) = \mathbb{E}[z^{\frac{1-\gamma}{\theta_2}\sum_{j=0}^{C-1} x_j}] = \prod_{j=0}^{C-1} G_{x_j}(z^{(1-\gamma)/\theta_2}). \tag{B.1}$$

To compute the length of PBOW, denoted by $\eta$, we use an argument similar to that for computing the length of a generic slot. Assume $\eta$ is an i.i.d. random variable. If any $\mathrm{AC}_1$ stations transmit at the $\tilde{N}$th slot in $\mathsf{Z}_1$, the length of PBOW equals $\eta = (\tilde{N}-1)\Delta + T_s$. Therefore,

$$G_\eta(z) = G_{\tilde{N}}(z^\Delta) z^{-\Delta} G_{T_s}(z), \tag{B.2}$$

where

$$G_{\tilde{N}}(z) = \sum_{k=1}^{M} z^{\tilde{N}} P[\tilde{N}=k] = (1-p_{\mathrm{i},1}) \sum_{k=1}^{M} z^k (p_{\mathrm{i},1})^{k-1}, \tag{B.3}$$

and $p_{\mathrm{i},1}$ is given by Equation (5.17). The total pre-backoff period, $U = \sum_{i=1}^{N_U(C)} \eta_i$, has the PGF given by

$$\begin{aligned} G_U(z) &= (1-p) \sum_{c=1}^{m} p^{c-1} \mathbb{E}[z^{\sum_{i=1}^{N_U(c)} \eta_i}] + p^m z^{m\Delta} \\ &= (1-p) \sum_{c=1}^{m} p^{c-1} G_{N_U(c)}(G_\eta(z)) + (pz)^m z^\Delta. \end{aligned} \tag{B.4}$$

# Appendix C

# SPRS Algorithm

**Algorithm C.1** SPRS Algorithm [49]

**Input:** $\{\kappa_m\} \in \mathcal{K}$, $K \in \mathbb{N}$

**Step 1** Randomly select an initial user subset $\kappa_0 \in \mathcal{K}$; $N_0 :=$ initial sample size; $m := 0$. Go to step 2.

**Step 2** Given $\kappa_m$, calculate the corresponding $U_m(\kappa_m)$. Randomly select another user subset $\kappa'_m \in \mathcal{K} \backslash \{\kappa_m\}$. Compute the corresponding $U_m(\kappa'_m)$. Go to step 3.

**Step 3** If $U_m(\kappa_m) > U_m(\kappa'_m)$, then let $\kappa_{m+1} = \kappa_m$. Otherwise, let $\kappa_{m+1} = \kappa'_m$. Go to step 4.

**Step 4** Perform a paired t-test between $U_m(\kappa_m)$ and $U_m(\kappa'_m)$ to test the null hypothesis $H_0 : U_m(\kappa_m) = U_m(\kappa'_m)$. If $p$-value>0.2, then $N_{m+1} := N_m$; otherwise $N_{m+1} := N_m + \delta_N$. Every $K$ iterations, do $N_{m+1} := N_m + \delta_N$. $m := m + 1$. Go to step 2.

# Appendix D

# Derivation of Vacation Period Distribution

In Section 6.3, the vacation period is represented by a d-PH type distribution $PH_d(\boldsymbol{\eta}, \mathbf{V})$ with order $K$. Based on the derivation in [100], we elaborate how the parameters $\boldsymbol{\eta}$ and $\mathbf{V}$ can be obtained using a simple example, where there are five users $\{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}\}$ contending the channel and the reservation pattern is given as $\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{a}, \mathsf{e}, \mathsf{c}, \mathsf{d}$. The generalization to arbitrary number of users and reservation pattern should be straightforward. We focus on the user $\mathsf{a}$. Therefore the entire reservation pattern can be divided into two subcycles $\phi_j,\ j = 1, 2$, where $\phi_1 = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ and $\phi_2 = \{\mathsf{a}, \mathsf{e}, \mathsf{c}, \mathsf{d}\}$. Let $S_i$ denote the length of the reservation period of user $i,\ i = \mathsf{a}, \cdots, \mathsf{e}$.

We first derive the vacation period length in each subcycle, which is identical to the sum of the service period of other users in the same subcycle. For hard reservation, the service period is constant with length $S_i$ and can be readily characterized by

$PH_d(\boldsymbol{\gamma}_i, \mathbf{F}_i)$ with order $S_i$ given by

$$\boldsymbol{\gamma}_i = \begin{bmatrix} 1 & \mathbf{0}_{S_i-1} \end{bmatrix}, \quad \mathbf{F}_i = \begin{bmatrix} \mathbf{0}_{S_i-1} & \mathbf{I}_{S_i-1} \\ 0 & \mathbf{0}_{S_i-1}^T \end{bmatrix}, \tag{D.1}$$

where $\mathbf{0}_{S_i-1}$ is a column vector of all zeros with dimension $S_i-1$. For soft reservation, the service period may be shorter than the reservation period when the buffer becomes empty prior to the expiration of the reservation period. In such a case the service period is still of d-PH type but its derivation is lengthy. We refer the readers to the appendix in [92] for the details. In this thesis we assume the service period distribution is known and denoted by $PH_d(\boldsymbol{\gamma}_i, \mathbf{F}_i)$ for user $i$. Therefore, the vacation period in subcycle $j$ can be represented by $PH_d(\boldsymbol{\delta}_j, \mathbf{L}_j)$ with dimension $\sum_{i\in\{\phi_j\setminus \mathsf{a}\}} S_i$. In the first subcycle, user $\mathsf{a}$ takes vacation after its service period according to the initial probability vector $\boldsymbol{\delta}_1$, which is give by $\boldsymbol{\delta}_1 = \begin{bmatrix} \boldsymbol{\gamma}_{\mathsf{b}} & \mathbf{0} \end{bmatrix}$, since the service period of user $\mathsf{a}$ is succeeded by that of user $\mathsf{b}$. The channel will be occupied by user $\mathsf{b}$ according to the service period of user $\mathsf{b}$, given by $\mathbf{F}_{\mathsf{b}}$. At the end of the service period of user $\mathsf{b}$, user $\mathsf{c}$ will access the channel. Therefore, we have absorption according to $\mathbf{F}_{\mathsf{b}}^0$ and beginning of service of user $\mathsf{c}$ according to $\boldsymbol{\gamma}_{\mathsf{c}}$. Then the channel is occupied by user $\mathsf{c}$ according to $\mathbf{F}_{\mathsf{c}}$. After serving user $\mathsf{c}$, first subcycle is over according to $\mathbf{F}_{\mathsf{c}}^0$. As a result, the vacation period in the first subcycle is completely characterized by

$$\mathbf{L}_1 = \begin{bmatrix} F_{\mathsf{b}} & \mathbf{F}_{\mathsf{b}}^0\boldsymbol{\gamma}_{\mathsf{c}} \\ 0 & F_{\mathsf{b}} \end{bmatrix}, \quad \mathbf{L}_1^0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{F}_{\mathsf{c}}^0 \end{bmatrix}, \quad \boldsymbol{\delta}_1 = \begin{bmatrix} \boldsymbol{\gamma}_{\mathsf{b}} & \mathbf{0} \end{bmatrix}. \tag{D.2}$$

Similarly, the vacation period in the second subcycle $PH_d(\boldsymbol{\delta}_2, L_2)$ is given by

$$\mathbf{L}_2 = \begin{bmatrix} F_{\mathsf{e}} & \mathbf{F}_{\mathsf{e}}^0\boldsymbol{\gamma}_{\mathsf{c}} & 0 \\ 0 & F_{\mathsf{c}} & \mathbf{F}_{\mathsf{c}}^0\boldsymbol{\gamma}_{\mathsf{d}} \\ 0 & 0 & F_{\mathbf{d}} \end{bmatrix}, \ \mathbf{L}_2^0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{F}_{\mathsf{d}}^0 \end{bmatrix}, \ \boldsymbol{\delta}_2 = \begin{bmatrix} \boldsymbol{\gamma}_{\mathsf{e}} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{D.3}$$

From the vacation period distribution in each subscycle we can construct the vacation period distribution of the tagged user. It can be seen that the tagged user may take vacation in each subscycle for certain amount of time and than move to the succeeding subscycle, but can not skip the proceeding subcycle before returning to the current one, and vice versa. The transition between two subsycles can be described by four matrices $\mathbf{U}_0$, $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$. The matrix $\mathbf{U}_0$ specifies the transition during the vacation period in the current subcycle, and $\mathbf{U}_1$ determines the end of the current vacation period and which is the next vacation period, given by

$$\mathbf{U}_0 = \begin{bmatrix} \mathbf{L}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_2 \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{L}_1^0 \boldsymbol{\delta}_2 \\ \mathbf{L}_2^0 \boldsymbol{\delta}_1 & 0 \end{bmatrix} \tag{D.4}$$

The above distribution is interpreted as follows. $\mathbf{U}_0(1,1)$ indicates that the tagged user is taking vacation in the first subcycle according to $(\boldsymbol{\delta}_1, \mathbf{L}_1)$. The first subcycle ends according to $\mathbf{L}_1^0$, and the tagged user enters the second one where the vacation period is determined by $(\boldsymbol{\delta}_2, \mathbf{L}_2)$, as stated by $\mathbf{U}_1(1,2)$. Similar arguments lead to positions $\mathbf{U}_0(2,2)$ and $\mathbf{U}_1(2,1)$. The matrix $\mathbf{U}_2$ is used for describing the transitions from the end of vacation period of the tagged user to the start of the service period, which is given by

$$\mathbf{U}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{L}_1^0 \\ \mathbf{L}_2^0 & \mathbf{0} \end{bmatrix} \tag{D.5}$$

The matrix $\mathbf{U}_3$ defines the transitions from the service period of the tagged user to the beginning of the vacation period, as given by

$$\mathbf{U}_3 = \begin{bmatrix} \mathbf{0} & \boldsymbol{\delta}_2 \\ \boldsymbol{\delta}_1 & \mathbf{0} \end{bmatrix} \tag{D.6}$$

Based on the above derivation, We modify the matrix blocks in (6.4) as follows:

- The matrix $\mathbf{V}$ is replaced with the matrix $\mathbf{U}_0$.

- The matrix $\mathbf{V}^0\boldsymbol{\eta}$ is replaced with the matrix $\mathbf{U}_1$.
- The vector $\mathbf{V}^0$ is replaced with the matrix $\mathbf{U}_2$.
- The vector $\boldsymbol{\eta}$ is replaced with the matrix $\mathbf{U}_3$.

Unavoidably the rate matrix $\mathbf{R}$ becomes larger. Furthermore, the service period duration in soft reservation is dependent on the vacation period distribution that involves more computation. The iterative algorithm in [100] with moderate modification can be used to obtain the service duration of the tagged user, which will be one of the future works.

# List of Acronyms

| | |
|---|---|
| **AC** | access category |
| **AIFS** | arbitrary interframe space |
| **APS** | angular power spectrum |
| **ARQ** | auto repeat request |
| **BP** | beacon period |
| **BCRP** | beacon collision resolution protocol |
| **BPST** | beacon period start time |
| **CAP** | contention access period |
| **CDMA** | code division multiple access |
| **CSMA/CA** | carrier sense multiple access with collision avoidance |
| **CTA** | channel time allocation period |
| **CTS** | clear to send |
| **CW** | contention window |
| **DCF** | distributed coordination function |
| **DIFS** | DCF interframe space |
| **DRP** | distributed reservation protocol |
| **DSSS** | direct-sequence spread-spectrum |

| | |
|---|---|
| **DTP** | data transfer period |
| **EDCA** | enhanced distributed channel access |
| **FSMC** | finite state Markov channel |
| **GSA** | global search algorithm |
| **HDTV** | high-definition television |
| **LOS** | line-of-sight |
| **MAC** | medium access control |
| **MAI** | multiple access interference |
| **MAS** | media access slot |
| **MMPP** | Markov-modulated Poisson process |
| **MWIS** | maximum weighted independent set |
| **NLOS** | non-line-of-sight |
| **OFDM** | orthogonal frequency division multiplexing |
| **PaA** | proportional allocation algorithm |
| **PCA** | prioritized channel access |
| **PNC** | piconet coordinator |
| **PSD** | power spectral density |
| **QoS** | quality of service |
| **RTS** | request to send |
| **RX** | receiver |
| **SIFS** | short interframe space |
| **SINR** | signal to interference plus noise ratio |
| **TDMA** | time division multiple access |
| **TX** | transmitter |

| | |
|---|---|
| **TXOP** | transmission opportunity |
| **UWB** | Ultra Wideband |
| **VoIP** | voice over Internet Protocol |
| **VBR** | variable bit rate |
| **WiMAX** | worldwide interoperability for microwave access |
| **WLAN** | wireless local area network |
| **WMAN** | wireless metropolitan area network |
| **WPAN** | wireless personal area network |

# Bibliography

[1] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks WPANs*, IEEE standard part 15.3 Std. IEEE Std 802.15.3, Sept. 2003.

[2] R. Negi and A. Rajeswaran, "Capacity of power constrained ad hoc networks," in *IEEE INFOCOM*, Hong Kong, China, Mar. 2004.

[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] F. Cuomo, C. Martello, A. Baiocchi, and F. Captriotti, "Radio resource sharing for ad hoc networking with UWB," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, pp. 1722–1732, 2002.

[5] M.-G. D. Benedetto, L. D. Nardis, M. Junk, and G. Giancola, "UWB$^2$: Uncoordinated, wireless, baseborn medium access for UWB communication networks," *Mobile Networks and Applications*, vol. 10, no. 5, pp. 663–674, 2005.

[6] H. Jiang, K.-H. 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, 2007.

[7] M. Hofri and Z. Rosberg, "Packet delay under the golden ratio weighted TDM policy in a multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 33, no. 3, pp. 341–349, 1987.

[8] I. Rubin and Z. Zhang, "Message delay and queue-size analysis for circuit-switched TDMA systems," *IEEE Trans. Commun.*, vol. 39, no. 6, pp. 905–914, 1991.

[9] M. K. Khan and H. Peyravi, "Delay and jitter analysis of generalized demand-assignment multiple access (DAMA) protocols with general traffic," in *Proc. The 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, Hawaii, USA, Mar. 2005.

[10] S. Andradóttir, "A global search method for discrete stochastic optimization," *SIAM Journal on Optimization*, vol. 6, no. 6, pp. 513–530, 1996.

[11] Federal Communications Commission (FCC), "First report and order-02048," http://www.fcc.gov/Bureaus/Engineering_Technology/Orders/2002/fcc02048.pdf, 2002.

[12] M. Z. Win and R. A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, 2000.

[13] R. J. Fontana, "Recent system applications of short-pulse ultra-wideband (UWB) technology," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 9, pp. 2087–2104, 2004.

[14] *DS-UWB Physical Layer Proposal*, IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a Std. IEEE P802.15-04/0137r4, Jan. 2005.

[15] *Multiband OFDM Physical Layer Proposal*, IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a Std. IEEE P802.15-03/268r1, Sept. 2003.

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

[17] M. Z. Win and R. A. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 51–53, 1998.

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

[19] *High rate ultra wideband PHY and MAC standard*, ECMA International 1st Edition ECMA-368, Dec. 2005. [Online]. Available: http:// www.ecma-international.org/publications/standards/Ecma-368.htm

[20] Y. Chu and A. Ganz, "A centralized MAC protocol for QoS support in UWB-based wireless networks," *Wireless Personal Commun.*, vol. 34, no. 1, pp. 45–66, 2005.

[21] B. Radunovic and J.-Y. L. Boudec, "Optimal power control, scheduling, and routing in UWB networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 7, pp. 1252–1270, 2004.

[22] Y. Shi, Y. T. Hou, H. D. Sherali, and S. Kompella, "Cross-layer optimization for UWB-based ad hoc networks," in *Proc. Milcom'06*, Washington, DC, Oct. 2006, pp. 1–7.

[23] R. Negi and A. Rajeswaran, "Scheduling and power adaptation for networks in the ultra wide band regime," in *Proc. IEEE Globecom*, Dallas, TX, Nov. 2004, pp. 139–145.

[24] J. Cai, K.-H. Liu, X. Shen, J. W. Mark, and T. D. Todd, "Power allocation and scheduling for ultra-wideband wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1103–1112, 2008.

[25] K. Lu, D. Wu, Y. Fang, and R. C. Qiu, "On medium access control for high data rate ultra-wideband ad hoc networks," in *Proc. IEEE WCNC*, vol. 795-800, New Orleans, LA, Mar. 2005.

[26] S. S. Kolenchery, J. K. Townsend, and J. A. Freebersyser, "A novel impulse radio network for tactical military wireless communications," in *Proc. IEEE MILCOM'98*, Oct. 1998, pp. 59–65.

[27] *Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: MAC Enhancements for Quality of Service (QoS)*, IEEE 802.11 WG Std. IEEE Std 802.11e/D3.3.2, Nov. 2002.

[28] K.-H. Liu, H. Rutagemwa, X. Shen, and J. W. Mark, "Efficiency and goodput analysis of Dly-ACK in IEEE 802.15.3," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3888–3898, 2007.

[29] H. Chen, Z. Guo, R. Yao, X. Shen, and Y. Li, "Performance analysis of delayed acknowledgment scheme in UWB-based high-rate WPAN," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 606–621, 2006.

[30] Y. Xiao, X. Shen, and H. Jiang, "Optimal ACK mechanisms of the IEEE 802.15.3 MAC for ultra-wideband systems," *IEEE J. Select. Areas Commun.*, vol. 24, no. 4, pp. 836–842, 2006.

[31] M. Win and R. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–689, 2000.

[32] J. G. Proakis, *Digital Communication*, 4th ed. New York: McGraw-Hill, 2001.

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

[34] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. Mobicom*, San Diego, CA, Sept. 2003, pp. 66–80.

[35] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, 2002.

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

[37] B. Hu and N. C. Beaulieu, "Accurate evaluation of multiple-access performance in TH-PPM and TH-BPSK UWB systems," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1758–1766, 2004.

[38] N. Kumar, S. Venkastesh, and R. M. Buehrer, "A spread-spectrum MAC protocol for impulse-radio networks," in *Proc. IEEE VTC'05 Fall*, Dallas, TX, Sept., 2005, pp. 665– 669.

[39] P. Gong, P. Xue, and D. K. Kim, "Intelligent MAC protocol for efficient support of multiple SOPs in UWB-based sensor networks," *Lecture notes in Control and Information Sciences*, vol. 344, pp. 194–201, 2006.

[40] V. Rakocevic, J. Griffiths, and G. Cope, "Analysis of bandwidth allocation schemes in multiservice IP networks using utility functions," in *Proc. International Teletraffic Congress (ITC17')*, Salvador da Bahia, Brazil, Dec. 2001.

[41] J. W. Lee, R. R. Mazumdar, and N. B. Shroff, "Non-convex optimization and rate control for multi-class services in the Internet," *IEEE/ACM Trans. Networking*, vol. 13, pp. 827–840, 2005.

[42] M. Dianati, X. Shen, and S. Naik, "Cooperative fair scheduling for the downlink of CDMA cellular networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1749–1760, 2007.

[43] Y. Cao and V. Li, "Utility-oriented adaptive QoS and bandwidth allocation in wireless networks," in *Proc. IEEE ICC*, New York, NY, May 2002, pp. 3071–3075.

[44] N. Lu and J. Bigham, "Utility-maximization bandwidth adaptation for multi-class traffic QoS provisioning in wireless networks," in *Proc. 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks (Q2SWinet'05)*, Montreal, Canada, Oct. 2005, pp. 136–143.

[45] J. W. Lee, R. R.Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class CDMA wireless networks," in *Proc. IEEE INFOCOM*, New York, NY, June 2002, pp. 1480–1489.

[46] G. Pflug, *Optimization of Stochastic Models: The Interface between Simulation and Optimization*. Dordrecht, Boston: Kluwer Academic Publishers, 1996.

[47] A. Demers, S. Keshav, and S.Shenker, "Analysis and simulation of a fair queueing algorithm," in *Proc. SIGCOMM'89*, Austin, TX, 1989, pp. 1–12.

[48] B. Denis, J.-B. Pierrot, and C. Abou-Rjeily, "Joint distributed synchronization and positioning in UWB ad hoc networks using TOA," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 4, pp. 1896–1911, 2006.

[49] T. Homem-De-Mello, "Variable-sample methods for stochastic optimization," *ACM Trans. Modeling and Computer Simulation (TOMACS)*, vol. 13, no. 2, pp. 108–133, 2003.

[50] 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," *Wireless Commun. and Mobile Compu.*, vol. 5, no. 5, pp. 567–580, 2005.

[51] D. L. Jagerman, B. Balcioglu, T. Altiok, and B. Melamed, "Mean waiting time approximations in the G/G/1 queue," *Queueing Systems*, vol. 46, no. 3, pp. 481–506, 2004.

[52] W. Fischer and K. Meier-Hellstern, "The markov-modulated poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1992.

[53] L. Kleinrock, *Queueing Systems*. New York: Willy Interscience, 1976, vol. Volume II: Computer Applications.

[54] C. Hu, H. Kim, J. Hou, D. Chi, and S. S. Nandagopalan, "Provisioning quality controlled medium access in ultrawideband-operated WPANs," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 23-29 2006.

[55] J. W. Robinson and T. S. Randhaw, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 22, no. 5, pp. 917–928, 2004.

[56] Z.-N. Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE J. Select. Areas Commun.*, vol. 22, no. 10, pp. 2095–2106, 2004.

[57] Y. Xiao, "Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1506–1515, 2005.

[58] J. Hui and M. Devetsikiotis, "A unified model for the performance analysis of IEEE 802.11e EDCA," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1498–1510, 2005.

[59] X. Chen, H. Zhai, X. Tian, and Y. Fang, "Supporting QoS in IEEE 802.11e wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2217–2227, 2006.

[60] Y. Lin and V. Wong, "Saturation throughput of IEEE 802.11e EDCA based on mean value analysis," in *Proc. IEEE WCNC*, Las Vegas, NV, Apr. 2006, pp. 475–480.

[61] X. Ling, K.-H. Liu, Y. Cheng, X. Shen, and J. W. Mark, "A novel performance model for distributed prioritized MAC protocols," in *Proc. IEEE Globecom*, Washington, DC, Nov. 2007.

[62] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE INFOCOM*, Hong Kong, China, Mar. 2004, pp. 1404–1413.

[63] C. G. Park, H. S. Jung, and D. H. Han, "Queueing analysis of IEEE 802.11 MAC protocol in wireless LAN," in *Proc. International Conference on Systems and International Conference on Mobile Communications and Learning Technologies(ICN/ICONS/MCL'06)*, Mauritius Island, Apr. 2006.

[64] K. Khalil and Y. Sun, "The effect of bursty traffic on the performance of local area networks," in *Proc. IEEE Globecom*, Orlando, FL, Dec. 1992, pp. 597–603.

[65] C.-J. Chang, C.-H. Lin, D.-S. Guan, and R.-G. Cheng, "Design of a power-spectrum-based ATM connection admission controller for multimedia communications," *IEEE Trans. Ind. Electron.*, vol. 45, no. 1, pp. 52–59, 1998.

[66] S. Molnar and G. G. Miklos, "On burst and correlation structure of teletraffic models," in *Proc. 5th IFIP Workshop on Performance Modelling and Evalution of ATM Networks, Ilkley, UK, July*, 1997.

[67] I. Habib and T. N. Saadawi, "Multimedia traffic characteristics in broadband networks," *IEEE Commun. Mag.*, vol. 30, no. 7, pp. 48–54, 1992.

[68] D. P. Heyman and D. Lucantoni, "Modeling multiple IP traffic streams with rate limits," *IEEE/ACM Trans. Networking*, vol. 11, no. 6, pp. 948–958, 2003.

[69] S. Shah-Heydari and T. Le-Ngoc, "MMPP models for multimedia traffic," *Telecommunication Systems*, vol. 15, no. 3, pp. 273–293, 2000.

[70] A. Heindl, "Decomposition of general queueing networks with MMPP inputs and customer losses," *Performance Evaluation*, vol. 51, no. 2-4, pp. 117–136, 2003.

[71] S. A. AlQahtani and S. A. Dhahran, "A simulation-based comparison of multimedia traffic prioritization schemes for high-performance input-queued packet switches," *Journal of Computer Science*, vol. 2, no. 4, pp. 347–354, 2006.

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

[73] Y. Zheng, K. Lu, D. Wu, and Y.Fang, "Performance analysis of IEEE 802.11 DCF in imperfect channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1648–1656, 2006.

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

[75] A. Zanella and F. D. Pellegrini, "Statistical characterization of the service time in saturated IEEE 802.11 networks," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 225–227, 2005.

[76] G. L. Choudhury and D. M. Lucantoni, "Numerical computation of the moments of a probability distribution from its transform," *Operations Research*, vol. 44, no. 2, pp. 368–381, 1996.

[77] J. Abate, G. L. Choudhury, and W. Whitt, "An introduction to numerical transform inversion and its application to probability models," *International Series in Operations Research and Management Science*, vol. 24, pp. 257–324, 2000.

[78] W. Fischer and K. Meier-Hellstern, "The markov-modulated poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1993.

[79] D. Gross and C. Harris, *Fundamentals of Queueing Theory.* New York, NY: John Wiley & Sons Inc., 1985.

[80] A. E. Eckberg, "Generalized peakedness of teletraffic processes," in *Proc. 10th International Teletraffic Congress ITC'10*, vol. session 4.4b, Montreal, Canada, 1983.

[81] A. Ekbal and J. M. Cioffi, "Effect of wireless channel process on queueing delay-approximate analysis using peakedness function," in *Proc. IEEE ICC*, Seoul, Korea, 2005, pp. 468–47.

[82] S. Li and C. Hwang, "On the convergence of traffic measurement and queueing analysis: a statistical-matching and queueing (SMAQ) tool," *IEEE/ACM Trans. Networking*, vol. 5, no. 1, pp. 95–110, 1997.

[83] L. Muscariello, M. Mellia, M. Meo, M. A. Marsan, and R. L. Cigno, "Markov models of internet traffic and a new hierarchical MMPP model," in *Proc. IEEE ICC*, Paris, France, 2004, pp. 2143–2147.

[84] Z. Cui and A. Nilsson, "The impact of correlation on delay performance of high speed networks," in *Proc. 26th Southeastern Symposium on System Theory*, Athens, OH, Mar. 2004, pp. 371–374.

[85] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 833–846, 1986.

[86] J. She, F. Hou, and P.-H. Ho, "An application-driven MAC-layer buffer management with active dropping for real-time video streaming in 802.16 networks," in *Proc. IEEE International Conference on Advanced Networking and Applications (AINA)*, Ontaio, Canada, 2007, pp. 451–458.

[87] D. Chen, D. Gu, and J. Zbang, "Supporting real-time traffic with QoS in IEEE 802.11e based home networks," in *Proc. IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, Jan. 2004, pp. 205–209.

[88] H. Wu, Y. Xia, and Q. Zhang, "Delay analysis of DRP in MBOA UWB MAC," in *Proc. IEEE ICC*, Istanbul, Turkey, June 2006.

[89] M. Zorzi, A. Chockalingam, and R. R. Rao, "Throughput analysis of TCP on channels with memory," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1289–1300, 2000.

[90] H. Rutagemwa and X. Shen, "Modeling and analysis of WAP performance over wireless links," *IEEE Trans. Mobile Comput.*, vol. 2, no. 3, pp. 221–232, 2003.

[91] L. Le, E. Hossain, and A. Alfa, "Queuing analysis for radio link level scheduling in a multi-rate TDMA wireless network," in *Proc. IEEE Globecom*, Dallas, TX, Nov. 2004.

[92] A. S. Alfa, "A discrete MAP/H/1 queue with vacations and exhaustive time-limited service," *Operation Research Letters*, vol. 18, no. 1, pp. 31–40, 1995.

[93] R. Zhang and L. Cai, "A Markov model for UWB indoor channel with shadowing," in *Proc. International Conference on Quality of Service in Heterogenous Wired/Wireless Networks(Q'shine*, Vancouver, Canada, Aug. 2007.

[94] I. 802.15.SG.3a, "Channel modeling sub-committee report final," *IEEE 802.15-02/490rl-SG3a*, Feb. 2005.

[95] A. F. Molisch, "Time variance for UWB wireless channels," IEEE P802.15-02/461-SG3a and IEEE 802.15-02/462-SG3a.

[96] A. F. Molisch, J. R. Foerster, and M. Pendergrass, "Channel models for ultrawideband personal area networks," *IEEE Wireless Commun.*, vol. 10, no. 6, pp. 14–21, 2003.

[97] S. V. Schell, "Analysis of time variance of a UWB propagation channel," IEEE P802.15-02/452-SG3a and IEEE 802.15-02/453-SG3a.

[98] W.-C. Liu and L.-C. Wang, "BER analysis in a generalized UWB frequency selective fading channel with random arriving clusters and rays," in *Proc. IEEE ICC*, Glasgow, UK, 2007, pp. 4281–4286.

[99] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models-An Algorithmic Approach.* Mineola, NY: Dover Publications, 1994.

[100] I. Frigui and A. S. Alfa, "Analysis of a discrete time table polling system with MAP input and time-limited service discipline," *Telecommunication Systems*, vol. 12, no. 1, pp. 51–77, 1999.

[101] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[102] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. Networking*, vol. 10, no. 2, pp. 272–286, 2002.

[103] H. Yaïche, R. R. Mazumdar, and c. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 667–678, 2000.

[104] C. E. Luna, L. P. Kondi, and A. K. Katsaggelos, "Maximizing user utility in video streaming applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 2, pp. 141–148, 2003.

[105] M. Chiang, S. Zhang, and P. Hande, "Distributed rate allocation for inelastic flows: optimization frameworks, optimality conditions, and optimal algorithms," in *Proc. IEEE INFOCOM*, Miami, FL, Mar. 2005, pp. 2679–2690.

[106] H. Shen, L. Cai, and X. Shen, "Performance analysis of TFRC over wireless link with truncated link level ARQ," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1479–1487, 2006.

[107] M. Dianati, X. Shen, and S. Naik, "Scheduling with base station diversity and fairness analysis for the downlink of CDMA cellular networks," *Wireless Commun. Mobile Comp.*, vol. 7, no. 5, pp. 569–579, 2007.

[108] R. Boppana and M. M. Halldorsson, "Approximating maximum independent sets by excluding subgraphs," *BIT Numerical Mathematics*, vol. 32, no. 2, pp. 180–196, 1992.

[109] S. Sakai, M. Togasaki, and K. Yamazaki, "A note on greedy algorithms for maximum weighted independent set problem," *Discrete Applied Mathematics*, vol. 126, pp. 313–322, 2003.

[110] K.-H. Liu, X. Ling, X. Shen, and J. W. Mark, "Performance analysis of prioritized MAC in UWB-WPAN with bursty multimedia traffic," *IEEE Trans. Veh. Technol.*, to appear.

[111] K.-H. Liu, L. Cai, and X. Shen, "Multi-class utility-based scheduling for UWB networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1176-1187, 2008.

[112] K.-H. Liu, L. Cai, and X. Shen, "Exclusive-region based scheduling algorithms for UWB WPAN," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 933-932, 2008.

[113] K.-H. Liu, X. Shen, R. Zhang, and L. Cai, "Performance analysis of distributed reservation protocol for UWB-based WPAN," *IEEE Trans. Veh. Technol*, to appear

[114] K.-H. Liu, X. Shen, R. Zhang, and L. Cai, "Delay analysis of distributed reservation protocol with UWB shadowing channel for WPAN," in *Proc. IEEE ICC*, Beijing, China, May 19-23, 2008.

[115] K.-H. Liu, L. Cai, and X. Shen, "Effective scheduling for UWB networks using discrete stochastic optimization," in *Proc. IEEE CCNC*, Las Vegas, NV, Jan. 11-13, 2007, pp. 624-628.

[116] K.-H. Liu, L. Cai, and X. Shen, "Performance enhancement of medium access control for UWB WPAN," in *Proc. IEEE Globecom*, San Francisco, CA, Nov. 27-Dec. 1, 2006.