

Distributed Reservation Algorithms for Video Streaming over WiMedia UWB Networks

by

Maryam Daneshi

B.Sc., Sharif University of Technology, 2007

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

in the Department of Computer Science

© Maryam Daneshi, 2009

University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Distributed Reservation Algorithms for Video Streaming over WiMedia UWB Networks

by

Maryam Daneshi

B.Sc., Sharif University of Technology, 2007

Supervisory Committee

Dr. Jianping Pan, Co-Supervisor
(Department of Computer Science)

Dr. Sudhakar Ganti, Co-Supervisor
(Department of Computer Science)

Dr. Kui Wu, Departmental Member
(Department of Computer Science)

Supervisory Committee

Dr. Jianping Pan, Co-Supervisor
(Department of Computer Science)

Dr. Sudhakar Ganti, Co-Supervisor
(Department of Computer Science)

Dr. Kui Wu, Departmental Member
(Department of Computer Science)

ABSTRACT

Ultra-wideband (UWB) technologies with higher data rates and lower transmission power over shorter ranges, have enabled a new set of applications in Wireless Personal Area Networks (WPANs). For example, UWB can offer data rates 50 to 500 times higher than the current WPAN technologies such as Bluetooth. This property makes UWB a primary candidate for indoor high-speed multimedia applications such as whole-house Internet Protocol Television (IPTV) and Personal Video Recorder (PVR). Lower power emission brings less interference to other devices, and larger bandwidth makes UWB less affected by interference from others, which are very attractive attributes in a household environment.

However, the effective and efficient utilization of such high data rate wireless channel represents a new challenge to WPAN Media Access Control (MAC), especially for high quality video streaming applications. To meet the minimum bandwidth and maximum delay requirement for Quality-of-Service (QoS) guarantee, high-definition

IPTV and PVR services usually need to reserve a certain amount of channel time for exclusive access in a dynamic manner, since the number of video flows may change over time in a piconet. WiMedia Alliances MAC protocols for UWB-based WPANs have become an international standard. The Distributed Reservation Protocol (DRP) is part of this standard that reserves the wireless channel on a slot-by-slot basis for different flows. However, not much work has been done on DRP reservation algorithms and their performance.

In this research, we propose, analyze and evaluate two application-aware reservation algorithms. One algorithm allocates time slots based on the first-fit idea whereas the other takes one step further by doing a best-fit reservation according to the maximum tolerable delay bound. Our proposed algorithms try to find the best possible time slots for any requests with respect to the existing reservations in the piconet and those arriving later. With these algorithms, devices in the same piconet that have data to transmit can negotiate and reserve time slots based on their traffic specification and QoS requirement while following WiMedia MAC reservation policies. We analyze the reservation algorithms and policies with a tiered overflow model, and evaluate their performance with Network Simulator (*NS-2*) and an MPEG-4 video traffic generator. We further discuss the ways of improving video streaming quality and system resource utilization in UWB networks.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	v
List of Tables	viii
List of Figures	ix
Acknowledgements	xi
Dedication	xii
1 Introduction	1
2 Background and Related Work	6
2.1 IPTV In-Home Distribution	6
2.1.1 Wireless Solutions	7
2.2 Reservation-based Wireless MAC	9
2.2.1 IEEE 802.15.3	9
2.2.2 Time Division Multiple Access	10
2.2.3 Video Streaming for UWB	11
2.2.4 UWB WiMedia DRP	11

3	WiMedia UWB Distributed Reservation Protocol	14
3.1	Prioritized Contention Access (PCA)	15
3.2	Distributed Reservation Protocol (DRP)	16
3.2.1	DRP Reservation Policies	17
3.3	Reservation Approaches	21
3.3.1	Network Calculus	22
3.3.2	Equivalent Bandwidth Theory	23
3.3.3	A Reservation Example	25
4	Reservation Algorithms	27
4.1	Algorithm Setup	28
4.2	First-Fit Algorithm	29
4.3	Best-Fit Algorithm	32
4.4	Reservation Examples	34
5	System Model	37
5.1	Simplified Model	38
5.1.1	Simplified Model Solution	40
5.2	Best-Fit Model	42
5.2.1	Best-Fit Model Solution	44
5.3	Homogeneous and Heterogeneous Traffic	47
5.4	Summary	48
6	Performance Evaluation of Reservation Algorithms	49
6.1	Simulation Setting	49
6.2	Homogeneous Traffic	51
6.2.1	Simulation Methodology	52
6.2.2	Model Validation	53

6.2.3	Performance Comparison of Algorithms	56
6.3	Heterogeneous Traffic	58
6.3.1	Simulation Methodology	58
6.3.2	Model Validation	59
6.3.3	Performance Comparison of Algorithms	66
6.4	Summary	68
7	Conclusion and Future Work	70
7.1	Further Research Issues	72
	Bibliography	74

List of Tables

Table 3.1	WiMedia UWB reservation policies	19
Table 3.2	List of notation	23
Table 6.1	WiMedia UWB parameters used in simulation	51
Table 6.2	Token bucket filter parameters	52
Table 6.3	Token bucket filter parameters for heterogeneous traffic	59

List of Figures

Figure 3.1	WiMedia UWB superframe.	15
Figure 3.2	A two dimensional view of the WiMedia UWB superframe. . .	18
Figure 3.3	Fluid twin token bucket shaper model.	22
Figure 3.4	An example of reservation approaches for a flow.	25
Figure 4.1	Reservation example of the first-fit algorithm.	34
Figure 4.2	Reservation example of the best-fit algorithm.	35
Figure 5.1	The simplified system model.	39
Figure 5.2	The best-fit system model.	43
Figure 6.1	Simulation scenario.	50
Figure 6.2	Blocking probability for homogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	53
Figure 6.3	System utilization for homogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	54
Figure 6.4	Blocking probability for homogeneous traffic with variable delay bounds: analysis and simulation.	55
Figure 6.5	System utilization for homogeneous traffic with variable delay bounds: analysis and simulation.	56
Figure 6.6	Blocking probability of homogeneous traffic: first-fit vs best-fit.	57
Figure 6.7	System utilization of homogeneous traffic: first-fit vs best-fit. .	58

Figure 6.8	Blocking probability for heterogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	60
Figure 6.9	System utilization for heterogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	61
Figure 6.10	Best-fit, flow-based blocking probability for heterogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	62
Figure 6.11	First-fit, flow-based blocking probability for heterogeneous traffic with a delay bound of 50 <i>ms</i> : analysis and simulation.	63
Figure 6.12	Blocking probability for heterogeneous traffic with variable delay bounds: analysis and simulation.	64
Figure 6.13	System utilization for heterogeneous traffic with variable delay bounds: analysis and simulation.	65
Figure 6.14	Best-fit, flow-based blocking probability for heterogeneous traffic with a variable delay bounds: analysis and simulation.	65
Figure 6.15	First-fit, flow-based blocking probability for heterogeneous traffic with a variable delay bounds: analysis and simulation.	66
Figure 6.16	Blocking probability of heterogeneous traffic: first-fit vs best-fit.	67
Figure 6.17	System utilization of heterogeneous traffic: first-fit vs best-fit.	68

ACKNOWLEDGEMENTS

I would like to thank all people who have helped and inspired me during my studies. I would like to thank my supervisors, Dr. Jianping Pan and Dr. Sudhakar Ganti, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

I want to express my gratitude to my thesis committee member, Dr. Kui Wu and Dr. Xiaodai Dong for their valuable suggestions.

Last but not least, I wish to thank my family specially my mother and beloved brothers. Without them, I would never go this far.

DEDICATION

To the loving memory of my father
and to my mother

Chapter 1

Introduction

The increasing worldwide demand for rapid, low-latency and high-volume data communications to homes and businesses has made economical information distribution and delivery increasingly important. Although there are various existing technologies such as: Asymmetric Digital Subscriber Line (ADSL), Symmetric Digital Subscriber Line (SDSL), Very High Bitrate DSL (VDSL), Fiber to the home (FTTH), Wireless local loop (WLL), ... that distribute traffic to homes, there is still much ongoing work on how to effectively deliver traffic within an access network's infrastructure.

The new barrier to end-to-end broadband service provisioning is the home network, the so-called "last meter problem". Last meter problem is the final step of delivering connectivity from a communications provider to a customer. A large number of houses have network enabled appliances such as multimedia entertainment sets, TV sets and video recorders, yet most of them are not equipped to support their interconnection.

The most trivial solution is a cable-based home network, though the cost and inconvenience of a large-scale home rewiring are prohibitive factors for most users. Therefore, there has been a large focus on "no-new-wires" solutions that would lever-

age the existing in-home cable or wireless technologies. There are various candidates such as Bluetooth, IEEE 802.11 and IEEE 802.15.3 for wireless home networking solutions. These “no new wires” solutions are justifiably expected to play a key role in the wide adoption of the digital home.

Wireless Personal Area Network (WPAN) is a network for interconnecting devices using a wireless technology that are within about 10 meters range of each other. This group of devices that are connected in an ad-hoc fashion is called a piconet.

In the category of WPANs, Ultra-WideBand (UWB) is a new technology for high data rate and short-range wireless services. UWB offers data rates from 50 to 500 times higher than the current WPAN technologies such as Bluetooth, and due to its low transmission power (-41 dBm/MHz), its current application range is limited to 0.5 Gbps in 10 meters range, which makes it suitable for indoor high-speed applications such as Internet Protocol Television (IPTV) and Personal Video Recorder (PVR). UWB also works on large bandwidths, so it is less affected by signal interference from other devices. Furthermore, UWB technologies have low power emissions such that their electrical interference with other devices is not a major problem. Utilizing such a high data rate wireless channel effectively and efficiently has become a new challenge to WPAN Media Access Control (MAC).

There are currently two major approaches to wireless MAC: contention-based and contention-free (polling or reservation-based). To guarantee Quality-of-Service (QoS) for an application, a minimum bandwidth and a maximum delay requirement need to be satisfied. High-definition IPTV and PVR services usually need to reserve a certain amount of channel time for exclusive access in a dynamic manner, since the number of video devices and flows may change in a piconet.

The IEEE 802.15.3 MAC [1], a new specification designed for WPAN to support ad-hoc networking and multimedia QoS guarantee, allows wireless devices to have

exclusive access to medium in a time division manner. The medium time is managed by a centralized coordinator. The coordinator decides which device can send during what period of time. This centralized architecture introduces several drawbacks to piconet, i.e., the reliance on the coordinator. That led to the introduction of a new distributed MAC by WiMedia Alliance [2]. Unlike IEEE 802.15.3, WiMedia MAC has no centralized management device, while it still offers exclusive access to the medium in a distributed manner. This makes WiMedia MAC a suitable specification for ad-hoc networking specially multimedia traffic.

In WiMedia MAC, the timeline is divided into fixed timeframes called superframes. The superframe duration is around $65,536 \mu s$ [2]. One superframe has 256 Medium Access Slots (MAS) of duration $256 \mu s$ each. MAS is the unit time of superframe that each device uses for reservation. Each superframe has two main parts, a Beacon Period (BP) and a Data Transfer Period (DTP). The availability Information Element (IE) is transmitted in a beacon packet during BP. It indicates a device's current utilization of MASs. There are two protocols for data transmission during the DTP of the superframe: Prioritized Contention Access (PCA) and Distributed Reservation Protocol (DRP). PCA is similar to the Enhanced Distributed Channel Access (EDCA) defined in IEEE 802.11e standard. It provides differentiated channel access to frames with different priorities. On the other hand, DRP is used by the devices to negotiate and reserve bandwidth. A reservation guarantees a period of time for transmission during which the reservation owner has exclusive access to the medium. For DRP channel access, a source node will observe the availability IE of all the neighbours, including the receiver, to find out which MAS can be reserved for exclusive access. Unreserved MASs are available for contention-based access by all stations with PCA.

A two dimensional structure of superframe has been proposed in [3] to elaborate the UWB MAC policies. Each column of the 16×16 superframe matrix is called an

allocation zone. Allocation zones of a WiMedia superframe excluding the BP zones are grouped into four sets called “isozones”. The MAS columns within the same isozone are distributed evenly across the superframe. More specifically, the MAS slots located in the same row and adjacent allocation zones within an isozone are separated from each other by a fixed interval that depends on the isozone in which the MAS slots are located. Such an interval is referred to as the native Service Interval (SI) of the isozone.

In this work, two application-aware reservation algorithms are proposed, analyzed and evaluated. With our algorithms, devices in the same piconet that have data to transmit will negotiate and reserve time slots based on their traffic specification and QoS requirement and follow WiMedia MAC reservation rules. The proposed algorithms are called first-fit and best-fit. The algorithms try to find the best possible time slots for any application with respect to existing reservations in the piconet and those coming later.

To extend the research to generic scenarios, we used both performance analysis and network simulation approaches to study the system utilization and blocking probability of video streams over UWB WPAN networks. We proposed a simplified model for superframe to analyze the performance of reservation algorithms. Also, based on the isozone structure of the WiMedia standard, we divided time slots into tiers and further proposed a tiered overflow model for the best-fit algorithms. This model considers the algorithm’s properties by incorporating a preprocessing module.

We also built a model for the network simulation. In the simulation, we added a UWB MAC module to the existing wireless package of the Network Simulator version 2 (*ns-2*) [4] and integrated the simulator with an MPEG-4 video traffic generator. The simulation results further validate the analytical model, providing a way of comparing the performance of proposed algorithms.

In this research, we first identified the issues of distributed reservation algorithms by studying the existing wireless medium access methods. We further investigated ways of making resource reservation based on the traffic specification to guarantee QoS for each flow. We compared the two reservation algorithms for two types of scenarios, homogeneous and heterogeneous, and showed the advantages of one over the other. The performance study of algorithms showed that the best-fit algorithm outperforms the first-fit algorithm. This observation was true for both types of scenarios.

The contributions of this thesis have three main aspects. First, we studied different wireless technologies and realized that UWB is a promising wireless technology for home networks. Second, we proposed two models for UWB-MAC DRP and incorporated the reservation algorithms into these models. Our models were suitable to justify and analyze the performance of the reservation algorithms. Third, we performed simulations for video streaming in WPANs. Simulation was used to first verify the system models and compare the performance of the proposed algorithms.

The rest of the thesis is organized as follows. In Chapter 2, we review the existing wireless technologies for home networks, summarize the related work on reservation-based MAC and outline the existing work on UWB technology. In Chapter 3, we describe the structure of WiMedia UWB MAC with special focus on the DRP part and present our reservation approach.

In Chapter 4, we present our proposed reservation algorithms in detail. In Chapter 5 the system model and the detailed analytical approach will be presented. In Chapter 6, we validate our analytical and simulation models. We evaluate and compare the performance of homogeneous and heterogeneous scenarios separately. In Chapter 7, we conclude our work followed by the discussion on the ways of further improving reservation algorithms and WiMedia UWB MAC superframe utilization.

Chapter 2

Background and Related Work

In this chapter, we provide a brief background of in-home video distribution. Then we review current Wireless MAC protocols, specifically reservation-based MAC protocols and the existing work on UWB WiMedia, with special focus on the DRP protocol.

2.1 IPTV In-Home Distribution

Service providers, with the help of current telecommunication technologies, both in backbone and access networks, can deliver IPTV services to the doorsteps of subscribers. Video distribution among all rooms in a household environment is still a challenge for home networks [5]. Ethernet is often suggested by service providers as the default Local Area Network (LAN) technology. Unfortunately, ethernet cables are not available in most homes. Due to this problem, both customers and service providers are looking at other alternatives such as “no-new-wires” and wireless solutions to deliver high-quality video and audio. Today, more than 50 candidate technologies, working groups and standard specifications exist for home networks [6].

One solution for home networks is reutilizing the existing household cablelines, phonelines and powerlines for data communications [6]. These technologies minimize

the installation cost but their achievable throughput limitation along with their shared medium nature, limits this reuse option.

On the other hand, wireless technologies such as IEEE 802.11 b/e/g, Bluetooth and UWB are expected to be the next technology for digital houses. Wireless options provide a simple and cheap solution. The high transmission rate over short distances of some of these wireless technologies, makes them a candidate for in-home high-definition video streaming. In the following subsection we outline some of the existing wireless technologies and specifications for home networks.

2.1.1 Wireless Solutions

Varieties of wireless technologies have been proposed as solutions for in-door video streaming. We first introduce some of the wireless technologies and their properties for in-home networks and the standards designed to support these technologies. Later we show which of these technologies are more suitable for video streaming.

Wireless Technologies and Standards

IEEE 802.11 [7] is the most mature wireless protocol for Wireless Local Area Networks (WLAN) communications and is one of the major candidate technologies for home networks. New versions and the enhancements made to 802.11 standard make it suitable for QoS support. IEEE 802.11b PHY [8] layer has been adopted to support data rates up to 11 Mbps. The 802.11g PHY [9] layer specification achieves data rates up to 54 Mbps. The major drawback of IEEE 802.11 is its lack of QoS and isochronous transmission slots. IEEE 802.11e [10] provides QoS via the new MAC layer.

Bluetooth [11] is a wireless radio system designed for short-ranges and is intended to serve as an air-interface to replace cables between personal devices. With a maxi-

mum data rate of 1 Mbps in 10 *cm* to 10 *m* range, Bluetooth is ideal for small home environment. Its low bandwidth capability only permits limited and dedicated usage and prevents it from being used for in-house multimedia [6].

IEEE 802.15.3 [1] is a another specification designed to support ad-hoc networking and multimedia QoS guarantees. It has been designed to achieve high data rates (up to 54 Mbps). This makes the standard a suitable candidate for high-definition video distribution. IEEE 802.15.3 works based on a centralized architecture. All communications are carried through or enabled by the centralized node. This architecture may cause several problems for peer-to-peer mobile applications.

The wireless technologies described so far are widely used for home networks, but due to the limitations mentioned, are not well suited for high quality in-home video streaming.

On the other hand, Ultra-Wide-Band (UWB) is a new wireless technology that has attracted attention for in-home video distribution. Its data rate can go up to hundreds Mbps and even Gbps within 10 meters distance making it a promising wireless technology for in-home multimedia networking [12]. One of the newest PHY and MAC standards for UWB WPAN has been developed by WiMedia Alliance. This standard was suggested to resolve the problem of 802.15.3. It's high physical data rate (up to 480 Mbps) makes it an excellent candidate for video distribution. Promising properties of WiMedia UWB technology make it one of the best candidates for in-home video streaming. Before focusing our work on WiMedia UWB MAC, we compare the existing wireless MAC protocols (specially reservation-based) and their properties in the following section.

2.2 Reservation-based Wireless MAC

Existing wireless MAC protocols mainly follow three approaches: contention-based, contention-free, or a hybrid of both. Polling and reservation are two strategies for contention-free MAC protocols, and the reservation-based ones can be more efficient and are often used to provide QoS guarantees for in-home distribution of high-definition IPTV and PVR services. In this work, our main focus is on the reservation-based MAC protocols. In the following subsections we describe a couple of contention-free MAC protocols.

2.2.1 IEEE 802.15.3

As mentioned before, IEEE 802.15.3 is one of the wireless standards and its main target is to enable high data-rate multimedia applications in WPANs. Its MAC layer follows a reservation-based strategy. The 802.15.3 piconet is formed in an ad-hoc manner, where devices may join or leave the network at any time. All communications within a piconet work in a peer-to-peer manner. A key property of the IEEE 802.15.3 MAC is the reliance on a piconet controller (PNC). The IEEE 802.15.3 MAC uses a master and slave model where a PNC has the job of coordinating channel access between devices, advertising device capabilities and coordinating sleep and wake-up schedules [1]. This architecture has some drawbacks and does not scale well with a large or dynamic system, and the performance of such a system highly depends on the coordinator. The PNC is elected dynamically and any device may become the PNC. All communications between all devices can only be carried out through or enabled by the PNC. This causes several problems when trying to support peer-to-peer (P2P) mobile applications. Any failure of the coordinator would disable the entire piconet for a considerable amount of time and may require several seconds before the rest of

devices recognize and elect a new PNC [13].

2.2.2 Time Division Multiple Access

Time Division Multiple Access (TDMA) MAC protocols are another set of reservation-based channel access methods. In TDMA protocols, the time is divided into slots of fixed durations, which are then grouped into frames. A fixed portion of each frame is dedicated to data traffic (slots reserved for data traffic) and a scheduling algorithm establishes a transmission schedule, which repeats in every frame until new traffic demands can be incorporated into the schedule. Through this protocol, nodes have exclusive access to reserved time slots.

A lot of work has been done on the TDMA-based MAC. Most of them are focused on centralized TDMA MAC. For example, [14] proposed a centralized TDMA which consists of an admission control policy, traffic conditioning mechanism and bandwidth allocation policy to provide fair bandwidth distribution among bursty data flows.

While the centralized TDMA protocol and its variants have been thoroughly studied, the distributed versions have been much less explored. [15] has proposed a TDMA-based distributed reservation algorithm for guaranteed link bandwidths in ad-hoc mesh networks. The bandwidth of each link is the number of time slots assigned to it in the frame. Their algorithm consists of two parts: first, it uses a distributed Bellman-Ford and iteratively finds locally feasible schedules by exchanging link scheduling information between nodes. Second, nodes work independently to converge the local feasible schedules to a global schedule and notify all other nodes of availability of a new schedule.

One important drawback of these works is that existing TDMA systems mainly allocate one block of time slots, often fixed, to each node per frame. Even though this block could have variable lengths, it may not be flexible enough to meet the

maximum delay requirement in QoS provisioning.

2.2.3 Video Streaming for UWB

There is a limited work reported on multimedia applications over the UWB technology. [16] has proposed an adaptive MAC protocol that provides QoS support for multimedia applications in UWB-based wireless networks. The resource allocation algorithm in this work assigns the transmission parameters to nodes based on the current traffic condition. The features of the algorithm are: 1) taking advantage of the flexibility of adjusting the QoS requests such as bit rates and 2) reserving a certain amount of channel capacity when assigning the power and bit rate. This work is limited to fixed transmission rate for the whole transmission time of a node, and doesn't consider peak rate and traffic burstiness.

Also, [17] introduces a wireless video streaming system over UWB radio and gives a demonstration system that uses a Nokia 7710 phone as a UWB transmitter and a projector as a UWB receiver. The data collected in this paper is yet another proof that UWB technology offers a great opportunity for short-range wireless multimedia applications.

2.2.4 UWB WiMedia DRP

There are also some works reported on WiMedia DRP MAC. One group focused on the analytical study of WiMedia DRP MAC [18, 19]. On the other hand, there is a limited experiment-based performance analysis of WiMedia UWB MAC.

[18] has studied the performance of DRP channel access delay for MBOA-UWB MAC. It has categorized delay analysis according to slot allocation into three main groups: single slot per station per superframe, multiple continuous slots per station per superframe and multiple non-continuous slots per superframe. This paper has

generalized its analysis to the third group. It proposed a bi-dimensional Markov chain for analysis and the slot allocation pattern can be arbitrary. Various periodical slot reservation patterns have been analyzed in this work and it has been shown that evenly distributed reservation patterns have less impact on delay constraints than others. The error-free wireless channel assumption maintained the analysis tractable.

[19] has further studied the delay impact of different reservation patterns with the consideration of a shadowing channel propagation model. This work was one of the first papers analyzing upper layer protocols by taking into account the time-varying UWB channel. It proposes a discrete-time model in which the tagged user proceeds on a vacation during the time slots reserved by other users. This model is used to analyze the impact of different reservation patterns on the delay performance of DRP. This work has also limited its effort to predefined reservation patterns.

[20] is the first work with experiment-based performance evaluation of video streaming over UWB networks. It analyzed the tradeoffs in UWB physical and MAC layers with regard to transmission rate, retry limit, reservation percentage and pattern. They have observed that at a given SNR, the choice of TxRate and retry limit has to be balanced properly with the achieved throughput and reliability. To reduce the turnaround overhead for higher throughput, clustered reservation is preferred whereas lower latency is a result of scattered reservation.

Here it is worth mentioning that the problem of reserving time slots in MAC layer for exclusive access is similar to the traditional dynamic memory allocation problem in the operating systems which is the allocation of memory storage or distributing the ownership of limited memory among computer programs. There are several methods and algorithms such as first-fit, best-fit and worst-fit, proposed for memory allocation [21]. In all these algorithms, memory is allocated from a large pool of unused memory based on the selection strategy. These algorithms only deal with the request

size and suffer from problems such as internal or external fragmentation.

Another reported research on the analytical model of WiMedia UWB MAC, specially the contention-based part is by [22]. The focus of this work is to develop an analytical model for WiMedia UWB MAC protocols using the renewal reward theorem framework and quantify the video streaming performance considering all practical features (PCA, Hard DRP, Soft DRP, TXOP) of WiMedia MAC protocols.

Since WiMedia UWB standards and products have been available only recently, there is not much effort reported in the literature on WiMedia DRP reservation algorithms. Compared to the existing work, we model the superframe as a two dimensional matrix and introduce two application-aware reservation algorithms that assign time slots to UWB devices based on their traffic specification and QoS requirements. Our algorithms take into account the minimum bandwidth and maximum delay requirements of the flow and follow a service interval based allocation strategy, and they also consider the reservation policies of WiMedia MAC standard. We also evaluate our work by proposing an analytical model for WiMedia DRP and verify the correctness of our proposed algorithms and model using simulation.

Chapter 3

WiMedia UWB Distributed Reservation Protocol

As discussed in Chapter 1 and Chapter 2, due to the high bit rate and low transmission power in short range, UWB is claimed to be the next networking technology to appear on the market. WiMedia Alliance, a large industry consortium, is developing a new specification for the physical and MAC layers of UWB systems. To maximize flexibility, the functionality of this MAC is distributed among devices. These functions include distributed coordination to avoid interference between different groups of devices by appropriate use of channels and distributed medium reservations to ensure QoS.

As discussed earlier, in WiMedia UWB MAC, the timeline is divided into fixed time frames called the superframe. Figure 3.1 shows WiMedia UWB's superframe structure. Its duration is around $65,536 \mu s$. One superframe has 256 Medium Access Slots (MAS) of duration $256 \mu s$ each. Superframe has two main parts: Beacon Period (BP) and Data Transfer Period (DTP). DTP could be a combination of Prioritized Contention Access (PCA) and Distributed Reservation Protocol (DRP).

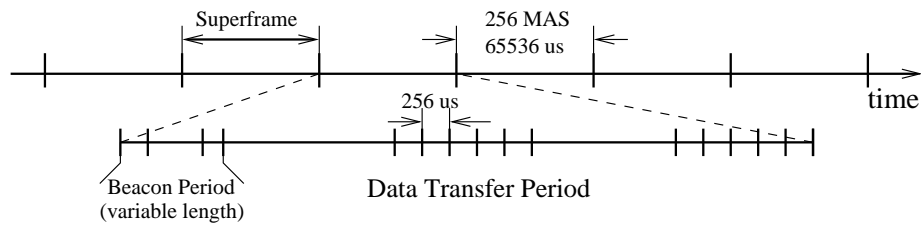


Figure 3.1: WiMedia UWB superframe.

The MAC sublayer provides two schemes for both asynchronous and isochronous data transfer. As mentioned before, these schemes are PCA and DRP. DRP is used to reserve the medium for TDMA-like isochronous access. PCA’s functionality is based on CSMA for network scalability. The MAC has policies that ensure equal sharing of the bandwidth.

In this chapter we first outline the MAC data transfer mechanisms and discuss WiMedia MAC policies governing DRP reservation. Later we discuss the reservation approaches and terminologies considered in this work.

3.1 Prioritized Contention Access (PCA)

As described earlier, data frames are transmitted during the DTP. WiMedia MAC provides both asynchronous and isochronous data communication services. The asynchronous service is provided by a prioritized Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol called PCA. This means that, if a device wishes to transmit, it has to first listen to the channel. If the channel is sensed busy, then the transmission is deferred for a “random” interval. Also, during the packet transmission if a packet collision is detected, a random backoff is applied, which forces a device to defer its access to the channel for an extra period.

PCA is similar to Enhanced Distributed Channel Access (EDCA) defined in IEEE 802.11e. Packets with different priorities are transmitted using different CSMA/CA

contention parameters. This reduces the probability of collisions on the channel. The basic difference between PCA and EDCA is the properties of UWB PHY and the existence of DRP. There are four Access Categories (ACs) for frames buffered in a device for transmission. A device employs a prioritized contention procedure for each AC to obtain a transmission opportunity for the frames belonging to that AC.

3.2 Distributed Reservation Protocol (DRP)

The other isochronous transmission method of WiMedia MAC is DRP. DRP enables devices to reserve one or more MASs which they can use to communicate with one or more neighbours. A reservation, defined by a subset of MASs during the superframe, guarantees a period of time for transmission during which the reservation owner has exclusive access to the medium. All devices that use DRP for transmission should announce their reservations by including this information in their beacons. Information included in a beacon frame is coded as Information Element (IE). One of the IEs defined in UWB MAC is the DRP IE. Information included in DRP IE contains the number of MAS, Target/Owner Device Address and Owner Reservation Type.

Reservation negotiation is always initiated by the device that will initiate frame transactions in the reservation, referred to as the reservation owner. The request includes the set of MASs that the transmitter intends to reserve for transmission. The device that will receive information is referred to as the reservation target. In case of receiving a request, the receiver analyzes the channel time utilization and sends a response indicating whether the reservation is accepted or not.

There are four types of reservations: hard, soft, private and PCA reservation. In a hard reservation, devices other than the reservation owner and target(s) shall not transmit or initiate a frames transaction. If there is any remaining time in a

reservation block that will not be used, the reservation owner and target(s) should release the reservation by sending a special message informing other neighbours about the early release of the reservation block. A device may transmit any type of frames in a hard reservation.

In a soft reservation, devices access the medium using PCA rules. The reservation owner may access the medium with the highest priority and without performing backoff. It may begin transmission at the beginning of each reservation block. The reservation owner may transmit any type of frames without backoff. Neighbours of a reservation owner shall follow PCA rules to access the medium.

In a private reservation, devices other than the reservation owner and target(s) shall not transmit frames. If there is a time slot that the reservation owner and target(s) are not using, then the time slot should be released. A device may consider the released reservation block as available. Finally, during a PCA reservation, any device may access the medium using PCA rules [2]. Once a reservation is successfully negotiated, the reservation is announced in the beacon messages via the DRP IEs. Other devices become aware of the reservation by receiving the beacons, and therefore defer access to the medium during the reserved MASs [2].

3.2.1 DRP Reservation Policies

WiMedia MAC standard [2] has specified a set of rules and constraints for distributed channel reservation. These rules limit both the size of the reservation blocks and their possible locations in a superframe. WiMedia [3] has proposed a two dimensional view of the superframe to better illustrate these rules. The superframe structure is shown in Figure 3.2. 256 medium access slots of the superframe have been arranged as a 16 by 16 matrix. Up to 32 MAS slots at the beginning of the superframe can be used for the BP and the rest is reserved for DTP.

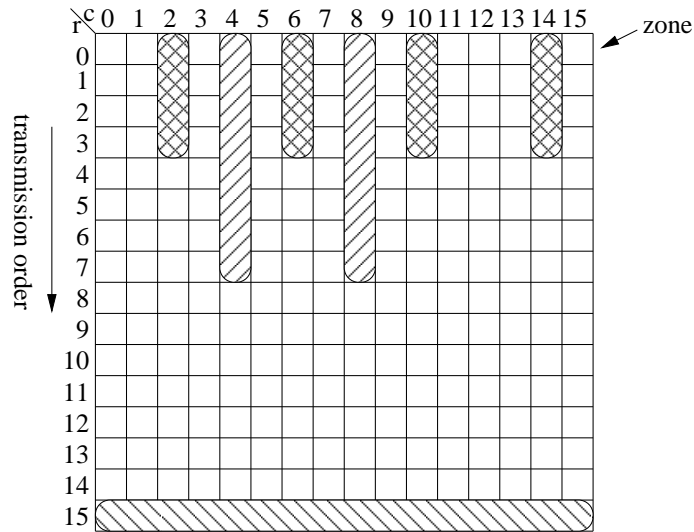


Figure 3.2: A two dimensional view of the WiMedia UWB superframe.

Each column of the superframe matrix is called an allocation zone. Allocation zones of a WiMedia superframe excluding the BP zones are grouped into four sets called isozones, as shown in Table 3.1. The MAS columns within the same isozone are distributed evenly across the superframe. The MAS slots located in the same row and “adjacent” allocation zones within an isozone are separated from each other by a fixed interval that depends on the isozone in which the MAS slots are located. Such an interval is referred to as the native service interval of the isozone, e.g., isozone 2 that includes zone 2, 6, 10 and 14 has a native service interval of 16 *ms*.

As mentioned above, the size of a reservation is the total medium time reserved in terms of the number of MAS slots. The size of each block should not exceed certain threshold values specified in the standard. For example, a reservation block cannot be greater than eight MAS slots, no matter which zone they are in. The location of the block also limits the number of MAS slots in that block. For example, a block starting from row 0 cannot have a size greater than eight MAS slots in any zone, or four MAS slots if starting from row 8, as listed in Table 3.1.

Those higher-indexed isozones are capable of supporting smaller service intervals,

Table 3.1: WiMedia UWB reservation policies

Zone Index (Column)	0	1	2	3	4	5	6	7
Isozone Index	BP	BP,3	2	3	1	3	2	3
MAS Index (Row)	0	1	2	3	4	5	6	7
Max Block Size	8	7	6	5	4	4	4	4
Zone Index (Column)	8	9	10	11	12	13	14	15
Isozone Index	0	3	2	3	1	3	2	3
MAS Index (Row)	8	9	10	11	12	13	14	15
Max Block Size	4	4	4	4	4	3	2	1

hence tighter delay bounds. As mentioned in the MAC specification, there are two types of reservations, row component and column component. A row component is the portion of a reservation that includes an equal number of MASs at the same offset(s) within every zone, optionally excluding zone 0. For this type of reservation, the native service interval is close to the duration of an allocation zone (4.096 *ms*). MAC policies on reservation locations require a row component to be located at as high-indexed MAS locations as possible within all allocation zones. A column component is defined as the portion of the reservation that is not a row component. According to the MAC specification, in the column component of a reservation, the reservation owner shall select reservation blocks that meet its requirements such that each block is located within the first eight MASs of its zone, if possible. If not, the reservation owner shall select reservation blocks that meet its requirements and minimize the highest MAS number selected in any zone. In Figure 3.2 the forward-slashed block shows a row component. This reservation is at the highest column index as indicated in the specification. The crossed reservation is a column component. The reservation is within the first eight MASs of the zone as required by the the specification.

According to the standard, if multiple potential zone locations meet the requirement, the reservation owner shall select reservation blocks in zones such that the latest used set is as early as possible in the order of isozone.

If multiple possible zones are available in the same isozone, the reservation owner should pick a zone that has the smallest zone index. The reservation owner shall place each reservation block at the earliest available location within its zone.

In order to make room for subsequent reservations that may request smaller service interval or tighter delay bounds, MAC policies on reservation locations require the selection of reservation blocks in isozones with the smallest possible isozone index, provided that their locations meet the applications latency requirement.

Figure 3.2 shows three ways of reserving 16 MAS slots assuming that the superframe is empty at the beginning. WiMedia MAC policies require the reservation to start from the lowest possible isozone; therefore one possible reservation is to have two reservation blocks of eight MAS slots each, starting from row 0 in zone 8 and zone 4, respectively, as shown in Figure 3.2. With this reservation, the maximum service interval will be 48 *ms*, i.e., if a packet misses the MAS at row 7 and column 8, it has to wait 184 MAS slots to meet the first reserved MAS at row 0 and column 4 in the next superframe. In addition, the remaining available MAS slots will become irregular, which makes the follow-on reservations harder to satisfy.

Another possible reservation is to have four reservation blocks of four MAS slots each, starting from row 0 in zone 2, 6, 10 and 14, respectively. In this case, the service interval will be about 16 *ms*, i.e., if a packet misses the last MAS of a block, it only has to wait 60 MAS slots to meet the first MAS in the next block. Further, the remaining available MAS slots are still symmetric to accommodate future requests.

An extreme case is to have 16 blocks of 1 MAS each with a service interval of 4 *ms*, at the same row offset in all zones, which is known as a row component reservation and should be located at the lower portion of the matrix, as shown in Figure 3.2. The other reservations given in the figure are types of column component reservations.

3.3 Reservation Approaches

Making a reservation of network resources for video flows over WiMedia UWB channels requires the determination of the number of MAS slots needed and their location within the superframe. The number of slots needed per superframe can be derived from the flow's Traffic Specification (TSpec). Delay requirement of each flow specified in the TSpec is defined as the maximum delay each packet of the flow can tolerate from the time it has been generated till it gets transmitted in the MAC layer. Satisfying the delay requirement for video traffic then depends on the service rate and the location of the reserved slots.

To make the reservation easy to handle, we associate each flow with a traffic shaper. In this work we used the fluid twin bucket shaper to regulate the video source traffic. This shaper was also proposed in WiMedia draft [3].

The fluid twin token bucket is composed of two token buckets: 1) The size of the first token bucket is always set to 0; 2) the second token bucket size is set to the value of the maximum burst size (b). The shaper is shown in Figure 3.3. The tokens arrive at the first bucket, which has bucket size 0, at the peak data rate (p) and the tokens arrive at the second bucket at the mean data rate (r). If a packet with size l arrives at the first token bucket, it needs to wait till l tokens arrive from the first generator. If it does not find a token upon its arrival it needs to wait a maximum of l/p time units before the tokens are generated and it is passed to the second token bucket shaper. If there are at least l tokens in the second token bucket, it is immediately sent out or it needs to wait for a maximum time of l/r before it is sent to the network.

A token bucket injects packets into the network only if there is an equivalent amount of tokens available and when a packet is transmitted, it consumes and removes exactly the same number of tokens from the bucket. With the help of a traffic shaper, any arbitrary traffic can be bounded by the token bucket model characterized by

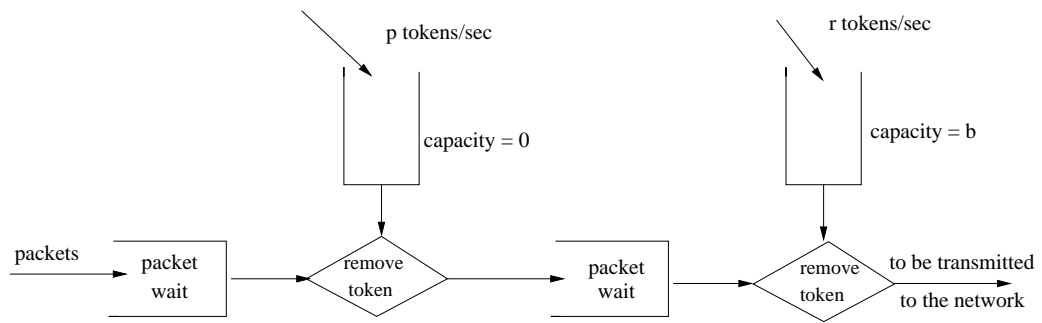


Figure 3.3: Fluid twin token bucket shaper model.

$\{r, p, b\}$.

After a basic introduction to the properties of the traffic shaper, we need to describe the relationship between the shaper's parameters and traffic's service rate. There are many methods to provide QoS for flows in the network. Here we introduce two main approaches: network calculus and equivalent bandwidth theory. Table 3.2 is giving a list of notations that we used through out this work.

3.3.1 Network Calculus

Network calculus is a theoretical framework for analyzing performance guarantees of network systems. In this framework, traffic flows are described by a cumulative function R which is the number of bits seen on the flow in any given time interval. In a given system S , we can describe receiving data with a cumulative function R . The system output can also be described with another cumulative function R' . Depending on the system structure, packets will be delivered after a variable delay.

Providing guarantees to data flows requires QoS support from network, i.e., limit the traffic sent by the source. This support is done by using the concept of arrival

Table 3.2: List of notation

Notation	description
$R(t)$	input arrival function
$R'(t)$	output arrival function
$a(t)$	arrival curve
$b(t)$	service curve
C	channel capacity
r	flow's mean rate
σ^2	flow's variance
p	flow's peak rate
b	flow's burst size
g	service rate
D	flow's maximum tolerable delay
B	token bucket buffer size
s	maximum service interval
d_q	token bucket queueing delay
MAC_OH	protocol MAC overhead

curve. The arrival curve is the basis of traffic characterization using any traffic shaper. From definition, arrival curve $a(t)$ is defined to give upper bounds on the arrival functions, where $a(t_2 - t_1) \geq R(t_2) - R(t_1)$ for all $t_2 \geq t_1 \geq 0$.

The service that is offered by network on an outgoing link can be characterized by a minimum service curve, denoted by $b(t)$. A network element with input arrival function $R(t)$ and output departure function $R'(t)$ is said to offer the service curve $b(t)$ if for all t_2 with $t_2 \geq t_1 \geq 0$, $R'(t_2) - R(t_1) \geq b(t_2 - t_1)$ holds [23]. Bandwidth management and allocation is the basic objective of resource allocation in QoS provisioning which allows high utilization of network resources.

3.3.2 Equivalent Bandwidth Theory

In order to make reservation, network nodes need to offer some guarantees to flows by reserving required resources, e.g., bandwidth. This is achieved by using a function for a flow called the Equivalent Bandwidth (EBW). This function characterizes the

bit rate required to be reserved for a given flow. More precisely, for a flow with cumulative function R as described in previous subsection; for a fixed, but arbitrary delay D , we can define the effective bandwidth g of the flow as the bit rate required to serve the flow in a work conserving manner, with a delay requirement D .

In this work we use a computationally simple approximation for the equivalent bandwidth. The required bandwidth allocation (g) for a traffic source for a given overall flow loss ratio P_{loss} is suggested as [24]:

$$g = a_1 * r + a_2 * \frac{\sigma^2}{C} \quad (3.1)$$

where C is the channel capacity, r and σ^2 are the mean and variance of the traffic rate. therefore:

$$\sigma^2 = r(p - r) \quad (3.2)$$

where p is the source peak rate. To find the coefficients a_1 and a_2 , [24] has suggested the following empirical approximation:

$$a_1 = 1 - \frac{\log_{10} P_{loss}}{50} \quad (3.3)$$

$$a_2/a_1 = -6 * \log_{10} P_{loss} \quad (3.4)$$

P_{loss} can be chosen based on the accepted packet loss ratio for the type of traffic in hand. In this work, we have chosen $P_{loss} = 0.01$ for our flow loss ratio. With this approximation for service rate, a flow is guaranteed to be transmitted between its minimum and maximum requested rate. Note that P_{loss} is not the intent of study in this work.

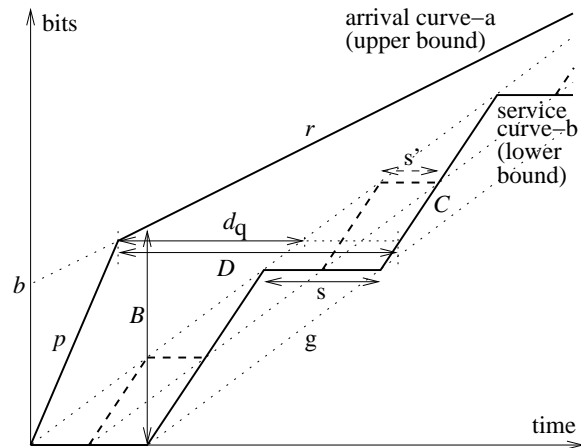


Figure 3.4: An example of reservation approaches for a flow.

3.3.3 A Reservation Example

We introduced two approaches for providing QoS for flows. In this section we show an example with different reservation patterns for a given flow with certain bandwidth and delay requirements. The maximum tolerable delay (D) for each flow includes the token bucket queueing delay (d_q) and the service interval (s), i.e., $D \leq d_q + s$. Figure 3.4 shows the relationship among the token bucket TSpec, the service rate and the delay bound for a given flow. Service rate can be computed from TSpec parameters $\{r, b, p\}$, using Equation (4.1). Service interval is due to the periodical arrangement of time slots in the superframe with service rate g .

The figure shows two possible reservation patterns with the same service rate g , the first one (solid service line) with a service interval (s) twice of the second one (dashed line, s'). Even though they have the same queueing delay d_q , have the same g , and are serviced at the same channel capacity C during the reserved slots, their delay bound is different due to the different service intervals. The first reservation can reduce its D by either reducing s or increasing g (i.e., reducing d_q). This example shows the impact of choosing service rate and also service interval on the overall performance of the system. In order to provide QoS guarantee for a flow, certain

number of time slots need to be reserved for the flow per superframe. This service rate determines how much delay the flow is going to experience in the token bucket shaper queue. Based on the delay requirement of the flow and the methodology used to compute the service rate, the number of time slots required can be computed. The next step is the arrangement of time slots in the superframe. This determines the service interval delay that a certain flow experiences during transmission. The example provided here illustrated the relationship between flow's arrival curve, service rate and the delay the flow experiences. The next step in solving the problem is how to arrange the time slots in the superframe for a given service rate to guarantee the QoS parameters.

In this chapter we briefly described the problem we are trying to solve, its challenges and our approach for solving the problem. Next, we are going to introduce two reservation algorithms for the DRP part of WiMedia UWB MAC.

Chapter 4

Reservation Algorithms

As discussed in the previous chapter, we need to design a reservation strategy for WiMedia UWB MAC to assign time slots to flows based on their QoS requirements. Here we introduce two reservation algorithms called *first-fit* and *best-fit*, respectively. Using the TSpec and QoS parameters of each flow discussed earlier, first and best-fit algorithms try to find a reservation for the flow, including the number of reserved slots and their locations in the superframe, that satisfies those requirements of the flow. UWB nodes can know the existing reservations by exchanging the DRP availability IE in beacon messages, so each node can apply the reservation algorithm individually. If a reservation was found, the reservation owner updates the reservation bitmap so that other UWB nodes can identify the MASs that are no longer available. The reservation shall guarantee a service rate between the minimum and maximum value requested by the flow and meet its maximum tolerable delay.

In the rest of this chapter, we describe the initial computation for both algorithms and then briefly describe first and best-fit algorithms followed by some examples.

4.1 Algorithm Setup

Both of the reservation algorithms use flow's TSpec (r, b, p) along with physical data rate achieved by the UWB to make the reservation. There is also some overhead introduced by the protocol in the MAC layer, which needs to be considered.

First the algorithm needs to know how many time slots a flow needs to guarantee the requested service rate. In the algorithms, we use the equivalent bandwidth approximation introduced in Chapter 3 to approximate the service rate (g) and compute the number of slots required per superframe (m) for such a service rate. To compute g , we use the following equations and parameters:

$$g = a_1 * r + a_2 * \frac{\sigma^2}{C} \quad (4.1)$$

As discussed in Chapter 3, with a packet loss ratio of 0.01 the coefficients a_1 and a_2 will be 1.04 and 12.48 respectively. Knowing the service rate, m is computed as:

$$m = \lceil \frac{g * TotalSlots}{C * (1 - MAC_OH)} \rceil$$

where channel capacity (C) could be any of the data rates supported by UWB devices as indicated in the standard and $TotalSlots$ is the number time slots in the superframe excluding the beacon period. In this work we assume the entire DTP of the superframe (240 MASs) is dedicated to DRP. MAC_OH is the percentage of overhead the protocol introduces in the MAC layer for each packet.

Besides the number of time slots per superframe, the algorithms also need to know the delay bound of the flow (D). As discussed before, the tolerable delay of a flow is the sum of the queueing delay (d_q) and Service Interval (s). Depending on the location of the reservation in the superframe, SI delay could take any value in the

range of $[0, 65.536]$ *ms*. There is a trade-off between queueing delay, service interval delay and service rate. To have a smaller queueing delay bound (hence a larger limit for SI delay), we need to have a larger service rate. This requires more MASs per superframe.

For a given service rate, we can compute the queueing delay as [3]:

$$d_q = \frac{(p - g) * b}{g * (p - r)}$$

These parameters (the number of MAS slots per superframe and delay bound) help us find an allocation for a given flow to support QoS guarantee. Next we introduce and discuss the proposed reservation algorithms briefly.

4.2 First-Fit Algorithm

First-fit algorithm is a reservation strategy that explicitly follows the WiMedia MAC reservation rules mentioned in the previous chapter. Pseudocode of first-fit algorithm is given in Algorithm 4.1.

Knowing how many time slots (m) a flow needs to reserve and also the delay requirement it has ($d - d_q$), the reservation algorithm can start searching through the superframe for a possible reservation. First-fit algorithm starts from the first empty place in the superframe in the increasing order of isozones, i.e., columns 8, 4, 12, 2, ..., 15.

It computes bs as the length of the first empty block that it finds in the superframe (see Initialization step of the pseudocode). If this is placed in row i and column j of the superframe, then bs is computed as:

$$bs = \min\{\text{free block size at } (i, j), \text{ maximum allowable block size at row } i\}$$

The algorithm then computes the number of such blocks this reservation requires as k . k is computed based on b , the length of the first empty place that is found in the superframe.

$$k = \lceil m/bs \rceil$$

The algorithm searches the superframe to find k empty blocks of size b each. The search is done in the order of isozones. These blocks could be anywhere in the superframe. If the algorithm could find k blocks of size bs , then it goes to the next step (see DelayCheck step of the pseudocode), which checks the delay of this potential reservation. If the delay is less than what the flow has requested, then this is a good allocation, otherwise the algorithm drops this reservations and tries to find another reservation (see Allocation step of the pseudocode).

The search for the next potential reservation is in the increasing order of rows and columns of isozones. This means, the algorithm will jump to the next empty place in the order of isozones (both row-wise and column-wise) and repeat Allocation step with this new starting location.

The algorithm searches through the entire superframe for a reservation. If it couldn't find an allocation for the given flow, it will search for another service rate between the minimum and maximum service rate (see BinarySearch step of pseudocode).

In all steps of the algorithm, we follow the reservation rules of the WiMedia standard. This means that at any row of the superframe, we never reserve more than what the standard has specified. The algorithm never reserves more than 8 consecutive time slots for a block per flow.

Algorithm 4.1 First-fit Algorithm

Input:

r : Mean Data Rate, b : Maximum Burst Size, p : Peak Data Rate, D : Delay Bound

ts : Total Slots, C : Channel capacity

Output:

reservation bitmap

{ s : Service Interval, d_q : Queueing Delay}

$$g_{min} = \frac{p}{1+d_q*\frac{p-r}{b}}, g_{max} = p$$

let $a_1 = 1.04$ and $a_2 = 12.48$

let (i, j) be the first empty place in the superframe

let bs be the number of consecutive empty blocks at (i, j)

$$g = \frac{1.04*r+12.48*r*(p-r)}{C}$$

$$d_q = \frac{(p-g)*b}{g*(p-r)}$$

Initialization:

Compute d_q for g :

$$\text{Let } m = \frac{g*ts}{C}$$

Allocation:

Find (i, j)

Compute bs

Find $k = \lceil \frac{m}{bs} \rceil$

if (k blocks exist) **then**

 goto DelayCheck

if No more place to check **then**

 goto BinarySearch

end if

else

 compute new (i, j) goto Allocation

end if

DelayCheck:

Reserve these blocks temporarily

Compute the maximum Service Interval (s) for this allocation

if $s + d_q < d$ **then**

 Commit the allocation

else

 Remove the temporarily allocation

 goto BinarySearch

end if

BinarySearch: {Binary search between g_{min} and g_{max} for the value of g }

if Another g value exists **then**

 with the new g goto Initialization

else

 return *No Reservation Possible*

end if

4.3 Best-Fit Algorithm

The other reservation algorithm proposed here is called best-fit. The main difference between the first-fit and best-fit algorithm is the way they search the superframe to find an allocation. The pseudocode of the best-fit algorithm is given in Algorithm 4.2. First-fit always starts from the first empty slot whereas the best-fit algorithm starts from the isozone that has the closest natural service interval to the flow's requested delay (excluding the queueing delay) and searches through the superframe from there.

Another difference between the best-fit and first-fit algorithm is the slots in the superframe they try to find those k blocks of size bs . Best-fit algorithm tries to keep the superframe well-structured. It does not allocate at any empty slot, but it allocates the blocks of the same flow at the same row of different columns. If a flow leaves and releases its reserved time slots, the superframe will still be well-structured. By keeping reservations symmetric, future flows can fit easily and the superframe will have less fragmentation. In all steps of this algorithm, we also follow the reservation rules of the WiMedia standard.

We discuss the convergence property of both the first-fit and best-fit algorithms in this part. If the input variables of both of the algorithms, i.e, request size, are valid, they will eventually converge. This means that as long as the request size is greater than 0 and smaller than 240 (which is the maximum capacity of the superframe), both algorithms will converge and return a reservation, otherwise a “no reservation possible” message stops the search. These conditions should be checked upon a request arrival.

Algorithm 4.2 Best-fit Algorithm

Input:

r : Mean Data Rate, b : Maximum Burst Size, p : Peak Data Rate, D : Delay Bound

ts : Total Slots, C : Channel capacity

Output:

reservation bitmap

{ s : Service Interval, d_q : Queueing Delay}

$$g_{min} = \frac{p}{1+d_q*\frac{p-r}{b}}, g_{max} = p$$

let $a_1 = 1.04$ and $a_2 = 12.48$

let (i, j) be the first empty place in isozone with delay $\leq (d - d_q)$

let bs be the number of consecutive empty blocks at (i, j)

$$g = \frac{1.04*r+12.48*r*(p-r)}{C}$$

$$d_q = \frac{(p-g)*b}{g*(p-r)}$$

Initialization:

Compute d_q for g :

Let $m = \frac{g*ts}{C}$

Allocation:

Find (i, j)

Compute bs

Find $k = \lceil \frac{m}{bs} \rceil$ starting at row i

if (k blocks exist) **then**

 goto DelayCheck

if No more place to check **then**

 goto BinarySearch

end if

else

 compute new (i, j) goto Allocation

end if

DelayCheck:

Reserve these blocks temporarily

Compute the maximum Service Interval (s) for this allocation

if $s + d_q < d$ **then**

 Commit the allocation

else

 Remove the temporarily allocation

 goto BinarySearch

end if

BinarySearch: {Binary search between g_{min} and g_{max} for the value of g }

if Another g value exists **then**

 with the new g goto Initialization

else

 return *No Reservation Possible*

end if

4.4 Reservation Examples

In this section we will try to describe the proposed algorithms using some examples. Assume a flow requires 16 times slots to guarantee its QoS requirements and its delay requirement is not more than 50 *ms* of SI delay. We show different reservation strategies each algorithm uses to accommodate this flow.

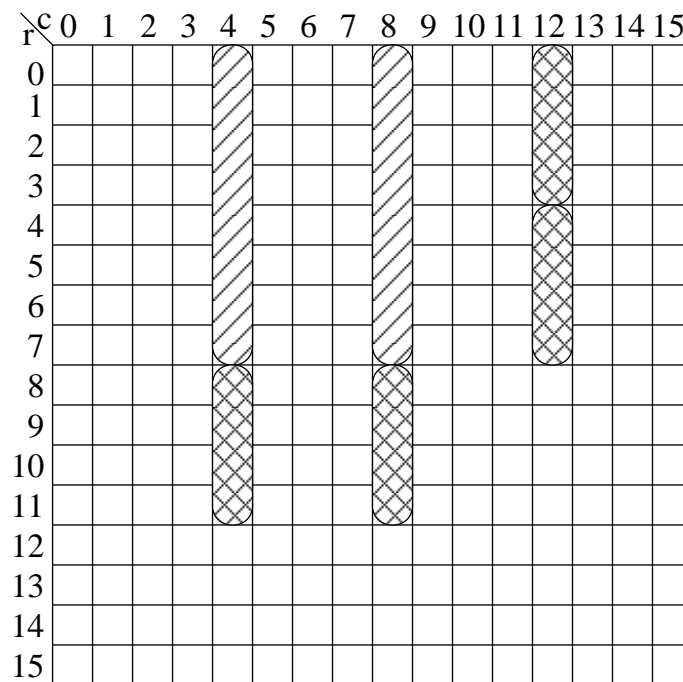


Figure 4.1: Reservation example of the first-fit algorithm.

Figure 4.1 shows the first-fit reservation strategy. In this example we assume there is a flow in the system and it has reserved 16 time slots (forward-slashed).

First-fit algorithm will find the cross-slashed reservation for the new flow. It finds the first empty slot in the superframe (column 8, row 8). The maximum block size a flow can have at this position according to the MAC standard is 4. As we can see in Fig. 4.1, the algorithm follows the first-fit strategy mentioned in Section 4.2 to find 4 blocks of size 4. This reservation will introduce maximum SI of around 33 *ms* (between column 12, row 8 and column 4 and row 8) which is within the specified

delay requirement of the flow. This reservation will be accepted and the time slots will be reserved for this flow.

Figure 4.2 shows the best-fit reservation strategy for the same flow. In this example again we assume that there is a flow in the system and it has reserved 16 time slots (forward-slashed).

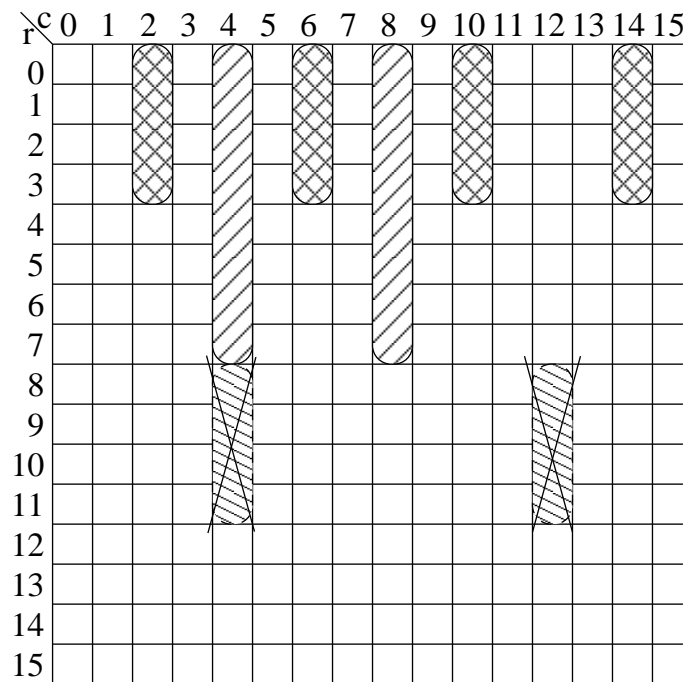


Figure 4.2: Reservation example of the best-fit algorithm.

Based on the delay requirement of the flow, the best-fit algorithm will choose the zone with a natural SI of less than 50 *ms*. It will choose column 4 to start. The block size is computed to be 4 (back-slashed). The algorithm tries to find 4 blocks of size 4 in the same zone (column 4 and 12) starting at the same row. This reservation is not feasible so the algorithm ignores this zone and tries to fit the request in the next zone. Four blocks of size 4 in column 2, 6, 10 and 14 will be reserved for this flow. This reservation will introduce a maximum SI of around 15 *ms* (between column 2, row 3 and column 6 and row 0) which is within the range of requested delay. This reservation will be accepted and time slots will be reserved for this flow.

In this chapter we introduced two reservation algorithms and described their properties briefly. First-fit only considers the WiMedia MAC reservation rules whereas best-fit takes one step further and takes in to account each flow's delay requirement and tries to keep the superframe well-structured at any time.

We also showed some examples to clarify the properties of the proposed algorithms. In the next chapter we model the algorithms to further analyze and justify the correctness of the proposed algorithms.

Chapter 5

System Model

In the previous chapter, we introduced two reservation algorithms for the DRP part of WiMedia UWB MAC. In this chapter, we will propose two analytical models for the MAC superframe with respect to the reservation algorithms. These models follow the philosophy of reserving circuits (or channels) for arriving calls in a circuit switching system. The calls in our system are multi-rated as opposed to single rate calls in a telephone system. We can compute various performance parameters such as call blocking, system utilization and overflow traffic characteristics with this model. The first model is a simplified model for the superframe and follows the a replication idea which will be described later in this chapter. The model's structure is derived from the delay requirement of flows. The second model for the superframe is constructed for the best-fit algorithm. This model follows a four-tiered overflow system to capture the properties of the best-fit algorithm.

For both models, call requests (flows) are offered to the system (superframe) requiring a certain number of time slots with a specific delay requirement. A request could either be carried (time slots reserved) or dropped. To study the performance of our model, we use two performance metrics: blocking probability (β) and system

utilization (u). Blocking probability shows the portion of the flows in the steady state that cannot reserve their requested time slots and therefore cannot transmit. System utilization shows the percentage of superframe that is reserved in the steady state.

We use Beshai’s approach [26] to derive the desired performance metrics for both models. The best-fit model also uses Glabowskis model [28] to approximate overflowed traffic parameters. For both models we will briefly give the mathematical approach to solve them.

Later in this chapter, we will show that these models are valid for both homogeneous (all flows request the same number of time slots and have fixed value delay requirement) and heterogeneous (the mixture of flows with different request sizes and different delay requirements) traffic.

The rest of this chapter is organized as follows. First, we present our proposed models and then outline the solution approach for each model to derive the metrics of interest. Later we justify why these models work for both homogeneous and heterogeneous scenarios.

5.1 Simplified Model

To analyze the performance of reservation algorithms for WiMedia DRP MAC, we propose a simplified model. This model only considers the size and delay requirement of a flow’s request. As discussed in the previous chapter, the two most important parameters when making reservation for a flow in DRP MAC, are the number of time slots and the delay requirement of the flow. Let’s assume that the superframe’s duration is T and it has C time slots. Then if a flow requests m slots per superframe of duration T and has a delay limit of d , it is equivalent to having a sub-flow requesting $\lceil m/k \rceil$ slots in a sub-frame of duration d , where $k = \lceil T/d \rceil$. Sub-frames will have

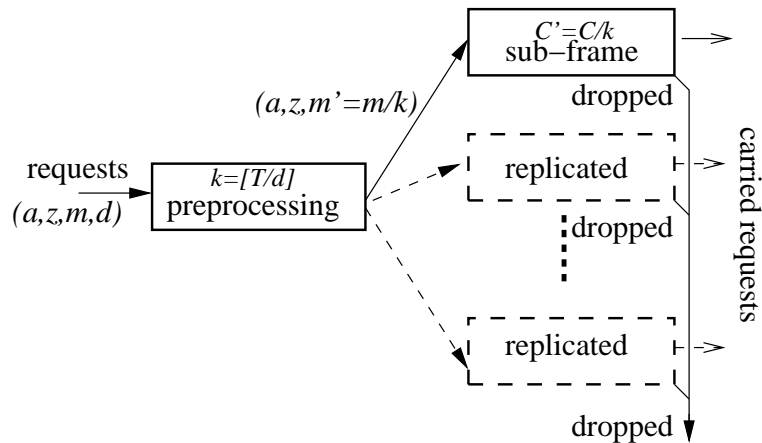


Figure 5.1: The simplified system model.

$C' = \lfloor C/k \rfloor$ time slots each. This means, sub-frames just replicate the reservation in the first sub-frame. With this model, a request is carried when all its sub-requests are carried by all sub-frames. Figure 5.1 shows the structure of this model. The model has a preprocessing unit which takes flows and apply the mentioned formulas to each request and build the appropriate replication model.

As an example, let's assume a flow with a request size of 16 and a delay requirement of 30 *ms* is offered to the superframe. Based on the model, sub-flow will have a request of $\lceil m/k \rceil = \lceil 16/3 \rceil = 6$, where $k = \lceil 65/30 \rceil = 3$. This request will be offered to a server (sub-frame) of size $C' = \lfloor 256/3 \rfloor = 85$. This means to carry a request of size 16 with delay requirement of 30 *ms*, after every 30 *ms* of superframe, this flows needs to have at least one time slot dedicated to it. We divided the flow's request evenly across the superframe and after every 30 *ms*, flow will have 1/3 of its requested size, i.e. 6. In the next subsection we will show the strategy to solve this model.

This model only takes request size and delay requirement into account, not the WiMedia MAC reservation policies. The model can predict the system behavior, but it is expected that the performance metrics computed from this model only provide an optimistic bound for the results.

5.1.1 Simplified Model Solution

To solve the proposed simplified model, we model the flows offered to the system as the mixture of non-Poissonian multi-channel requests. There are $i = 1, \dots, n$ flows in the piconet and each request is characterized by (a, z, m, d) , where $a = \lambda/\mu$ is the offered load (λ is the mean arrival rate and μ is the mean service rate), z is the peakedness of the flow, i.e., the variance to mean traffic intensity ratio that shows how much the offered traffic exceeds the system's actual capacity. Offered traffic is deemed peaky, regular, or smooth according as z is greater than, equal to, or less than one. m is the number of the requested time slots and d shows the maximum tolerable delay of the flow. m is the number of requested time slots and d shows the maximum tolerable delay of the flow. In this model, the preprocessing unit applies the replication strategy mentioned in the previous section to each flow and compute the new parameters for each sub-flow and sub-frame.

To solve this model, we use Beshai's two-component representation of the mixture of non-Poissonian multi-channel traffic model [26]. This solution is based on Delbrouck's approximation [25]. His solution replaced a non-Poissonian input by a series of Poissonian streams which unlike the Fourier series is not unique. Beshai's approximation only takes the first two components of this series and each sub-flow is approximated with an equivalent set of two Poissonian streams of m and $2\zeta m$ slots per call:

$$\{(w_{i,1} = a_i \times \frac{\zeta_i - z_i}{\zeta_i - 1}, 1, m_i), (w_{i,2} = a_i \times \frac{z_i - 1}{\zeta_i - 1}, 1, \zeta_i m_i)\},$$

In the numerical evaluation of this representation the value of ζ_i is approximated as an integer greater than z_i (2 has been used here). $P_x = P(X = x)$ is the state

probability vector and it is determined from:

$$xP_x = \sum_{i=1}^n w_{i,1}P_{x-m} + w_{i,2}P_{x-\zeta m}, \quad 0 < x \leq C' \quad (5.1)$$

P_x shows the steady-state probability of exactly x simultaneously busy time slots in the system.

$$P_k = 0 \text{ for } k < 0, \text{ and } \sum_{k=1}^{C'} P_k = 1 \quad (5.2)$$

k is a notation here and the condition $k < 0$ is used for computational purposes.

$Q_x = P(X > x) = 1 - P(X \leq x)$ is defined as cumulative complementary distribution function. We have the following equation set

$$Q_{x-1} = Q_x + P_{x-1}, \quad C' \geq x \geq 1 \text{ and } Q_{C'} = P_{C'} \quad (5.3)$$

The blocking probability of each stream is given by

$$b_i = (w_{i,1} \times Q_{C'-m_i+1} + w_{i,2} \times Q_{C'-\zeta_i m_i+1})/a_i.$$

Thus, the carried traffic of each stream is computed as

$$c = \sum_{x=0}^{C'} xP_x = \sum_{x=0}^{C'} x(Q_{x-1} - Q_x) = \sum_{x=0}^{C'} Q_x. \quad (5.4)$$

System utilization is computed by using the ratio of the total traffic carried by the system to the total system capacity, i.e.,

$$u = \frac{c}{C'}.$$

Blocking probability of the system is computed as the summation of blocking

probability of each stream:

$$\beta = \sum_{i=0}^n b_i \quad (5.5)$$

5.2 Best-Fit Model

To evaluate the performance of the best-fit algorithm, we specifically model the Wi-Media UWB superframe as a four-tier non-Poissonian loss servers (time slots) serving (reserved by) the flows. In this model we use the isozone structure of the superframe for tier mapping; therefore each tier is representing one class (isozone) of the superframe. For example, the first class (C_0) has one column that has two blocks of eight slots each for reservation, while the second class (C_1) has two columns, each of which has four blocks of four slots each, available for four requests of eight slots each. Because of this selection, each class of time slots, has a natural service interval. This service interval is the minimum delay a flow would experience if it reserves time slots from a particular class.

A flow requests for a certain number of time slots with a specified delay requirement. This request is offered to an appropriate class based on the availability of the time slots and the flow's requested delay. This means the preprocessing unit looks at the flow's requested delay and the native SI each of the classes (class 0 has a native SI of 65 *ms*, class 1 has a native SI of around 33 *ms* and so on) and forward the request to the correct class. If enough slots are available and they match the flow's requirement, they will be occupied for the duration of the flow in the piconet. Otherwise the request is going to be dropped to the next class. From the structure of isozones in the superframe, the next class always has more capacity and a smaller native SI; therefore there may be a feasible reservation for a flow dropped from the upper class.

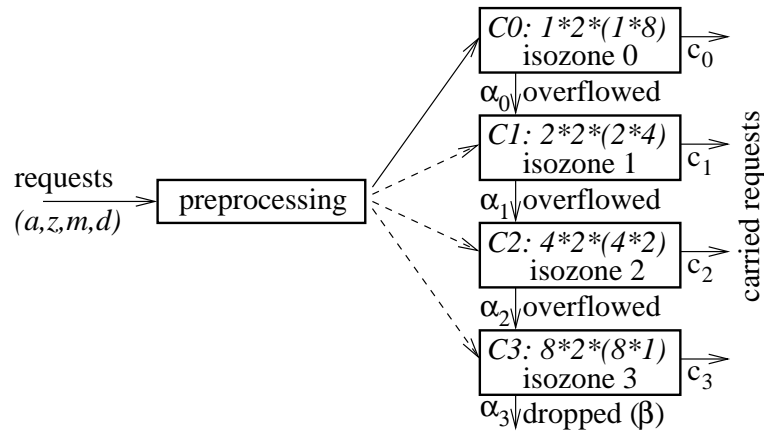


Figure 5.2: The best-fit system model.

As shown in Figure 5.2, to incorporate the reservation strategies and the flow's TSpec/QoS parameters, a preprocessing module has been introduced. This module forwards requests to appropriate classes based on their TSpec/QoS parameters. This means, whenever a request arrives, this module analyzes it and directs to an appropriate class.

Initially the preprocessing module forwards the request to an appropriate class based on its delay. In that class, if there are enough empty time slots to carry this flow's need, the request will be carried, otherwise the request is overflowed to the next class. There is no waiting queue for requests, so if a flow cannot be carried at the request time, it will be dropped.

To explain the model, we use the same example given for the simplified model, assuming a flow with a request size of 16 and a delay requirement of 30 ms is offered to our system. The preprocessing unit will consider the delay requirement of the flow and the natural SI of each class. This request cannot be carried by neither class zero (native SI = 65 ms) nor class one (native SI = 33 ms), so the request should be forwarded to class two with a native SI of 16 ms . C_2 can carry the flow as 4 blocks of size 4 each. The preprocessing unit will offer the request to C_2 . The way this

request will be handled in this class depends on the availability of time slots. If there are not enough unreserved time slots in this class, the request will be dropped to C_3 subsequently. In the next subsection we will present the mathematical model used to solve this model and compute the performance metrics.

5.2.1 Best-Fit Model Solution

To solve this model, we use Beshai's two-component representation of the mixture of non-Poissonian multi-channel traffic model [26].

We assume, there are $i = 1, \dots, n$ flows in the piconet and each flow request is characterized by (a, z, m, d) , where m is the request size and $a = \frac{m*\lambda}{\mu}$ is the mean load intensity (λ is the mean arrival rate and μ is the mean service rate), z is the peakedness of the flow, i.e., the variance to mean traffic intensity ratio that shows how much the offered traffic exceeds the system's actual capacity. d is the maximum tolerable delay of the flow.

The portion of the requests that is carried by each class is denoted as c_j ($j = 0, \dots, 3$) and the overflowed portion of them to the next class is denoted as α_j . The blocking probability is the portion of the requests that are dropped from the last class and it is defined as $\beta = \alpha_3$.

$P_{x,j} = P_j(X = x)$ shows the steady-state probability of exactly x simultaneously busy time slots in class j and it is determined from:

$$xP_{j,x} = \sum_{i=1}^n w_{i,1}P_{x-m,j} + w_{i,2}P_{x-\zeta m,j}, \quad 0 < x \leq C_j \quad (5.6)$$

$$P_{j,k} = 0 \text{ for } k < 0, \text{ and } \sum_{k=1}^{C_j} P_{j,k} = 1. \quad (5.7)$$

k is a notation here and the condition $k < 0$ is used for computational purposes.

$Q_{x,j} = P_j(X > x) = 1 - P_j(X \leq x)$ is defined as cumulative complementary distribution function of each class. We have the following equation set

$$Q_{x-1,j} = Q_x + P_{x-1,j}, \quad 1 \leq x \leq C_j \quad \text{and} \quad Q_{C_j,j} = P_{C_j,j} \quad (5.8)$$

The blocking probability of each stream for each class is

$$b_{i,j} = (w_{i,1}Q_{c_j-m_i+1,j} + w_{i,2}Q_{c_j-\zeta_i m_i+1,j})/a_i,$$

if flow i is offered to class j , otherwise, $b_{i,j} = 0$. Thus, the total blocking probability of this class is computed as:

$$\alpha_j = \sum_{i=0}^n b_{i,j} \quad (5.9)$$

Accordingly, the carried traffic of each stream for each class is computed as

$$c_j = \sum_{x=0}^{C_j} x P_{x,j} = \sum_{x=0}^{C_j} x (Q_{x-1,j} - Q_{x,j}) = \sum_{x=0}^{C_j} Q_{x,j}. \quad (5.10)$$

When the flow cannot be served by the initial class, the class with a higher index can be used as an alternative for the overflow traffic. Therefore, the traffic to the next class is no longer Poisson. Wilkinson initially modeled the overflow traffic based on a theory called Equivalent Random Theory (ERT). For each class, if the blocking probability is greater than zero, the overflowed portion of the requests is offered to the next class. The theory is to approximate the traffic by a Pascal (negative binomial) distribution. After that, Delbrouck [25] found the parameter similarities between Pascal and Bernoulli distributions and of their limiting relationships to Poisson distributions. It is possible to implement a unified approximating procedure to estimate the main congestion functions of peaky traffics. To determine the parameters of the overflowed traffic from a primary class to an alternative we use Wilkinson's approx-

imation. In this model which is based on Riordan's formula [27], the moments of marginal distribution is derived from the Erlang's B recursive formula:

$$E_x(a_{i,j}) = \frac{aE_{x-1}(a_{i,j})}{x + aE_{x-1}(a_{i,j})}, \quad E_0(A) = 1 \quad (5.11)$$

where a is the offered load and x could have any value between 0 and the maximum server capacity. Based on the Erlang's B formula, the approximation for the mean and peakedness of the overflowed portion is,

$$\alpha_{i,j} = E_{C_j}(a_{i,j}) * a_{i,j} = b_{i,j}a_{i,j}, \quad i = 1, \dots, n \quad (5.12)$$

$$Z_{i,j} = 1 - a_{i,j}\{E_{C_j-1}(a_{i,j}) - E_{C_j}(a_{i,j})\}, \quad (5.13)$$

$$Z_{i,j} = 1 - \alpha_{i,j} + \frac{a_{i,j}}{C_j + 1 - a_{i,j} + \alpha_{i,j}} \quad (5.14)$$

Glabowski's model [28] has extended the Wilkinson's formula to match the multi-rate traffic.

$$V_{i,j} = C_j - \sum_{l=1, l \neq i}^{C_j} c_{l,j}, \quad j = 0, \dots, 3, \quad (5.15)$$

$$z'_{i,j} = 1 - \alpha_{i,j} + \frac{a_{i,j}}{V_{i,j}/m_{i,j} + 1 + \alpha_{i,j} - a_{i,j}}, \quad (5.16)$$

Where $V_{i,j}$ is defined as the part of class C_j that is not occupied by the carried requests.

The same method used from (5.6) to (5.10) can be used in the in class(es) to compute the carried traffic and blocking probability of each class. The total blocking probability of the system can be easily computed from the probability of dropping of the last class.

System utilization is computed by using the ratio of the total traffic carried by

the system and the total system capacity, i.e.,

$$u = \frac{\sum_{j=0}^3 c_j}{C0 + C1 + C2 + C3}.$$

System blocking probability β is equal to α_3 .

5.3 Homogeneous and Heterogeneous Traffic

In this section we show under what traffic scenarios the proposed models are valid. As mentioned before, we are studying a piconet-based structure for UWB MAC. There could be a single type of traffic transmitted between UWB nodes (homogeneous scenario) in the piconet or a mixture of traffic (heterogeneous). There are two parameters that distinguish homogeneous and heterogeneous traffic: request size and delay requirement of each traffic flow.

As shown in the previous sections, Beshai's approximation used to solve these models can take care of the mixture of requests with different request sizes; therefore both models can be used for both homogeneous and heterogeneous traffic.

The simplified model structure is based on the delay requirement of flows. The sub-frame structure is constructed based on the delay requirement of flows; therefore the model is limited to work for a fixed-value delay for all flows. To tackle this problem, we assume that the delay for building the model can be the average of all delay requirements of the traffic flows.

On the other hand, the best-fit model is able to handle heterogeneous traffic with different delay requirements for each flow. This model has a fixed structure for the superframe (isozone-based) and handles flows separately based on their QoS parameters.

5.4 Summary

In this chapter we proposed two models for WiMedia UWB MAC and considered the reservation algorithms in our models. The simplified model generally showed the structure of the reservation algorithms by simply considering the superframe as a set of servers that carries requests. The model used the delay requirement of flows to simplify the problem into a replication model and uses Beshai's model to solve it.

The other model specifically focused on the properties of the best-fit model and used a four-tier overflow structure to model the superframe and the reservation algorithm. We presented strategies used for computing the desired performance metrics, i.e., system utilization and blocking probability. Both models were able to handle homogeneous and heterogeneous traffic in a piconet.

In the next chapter we will present a mathematical solution for solving the model and a simulation result gathered to validate the proposed algorithms and models.

Chapter 6

Performance Evaluation of Reservation Algorithms

In this chapter, we present and evaluate the analytical and simulation results of the proposed reservation algorithms for WiMedia UWB MAC. We first validate our system models by evaluating the utilization and blocking probability through analysis as well as through simulations. Afterwards, we compare the performance of the proposed algorithms. In the simulations, we consider video as the source of traffic. We consider two types of traffic in our WPAN, homogeneous and heterogeneous. We show and evaluate the results for both cases under various settings.

6.1 Simulation Setting

In this work, we used *ns-2* version 2.33 as our simulation tool and added a WiMedia DRP MAC module to it. For reservation purposes, we added a virtual scheduler to the simulation package. The scheduler handles all requests from all UWB devices. It maintains the current state of the superframe reservation. As mentioned before, UWB devices can use beaconing methods to have the most recent bitmap of the

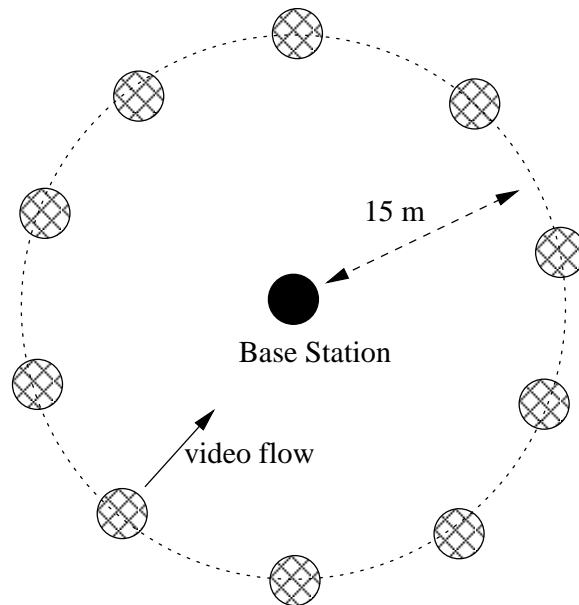


Figure 6.1: Simulation scenario.

reservation on superframe.

We also adopted the NOAH [29] extension to our wireless simulation module for the static routing scheme. For the physical layer, we used the existing wireless code and matched its parameters to those of WiMedia's. Table 6.1 shows a list of MAC and PHY parameters used in the simulation. These parameters have been extracted from the ECMA-368 standard for the physical and MAC layer of WiMedia UWB [2]. Symbol Rate presented in this table is used to compute the symbol interval ($1/\textit{Symbol Rate}$). This parameter is used to compute the number of symbols that can be transmitted in each packet. Any packet transmitted through UWB MAC has the Physical Layer Convergence Protocol (PLCP) header and the PLCP Protocol Data Unit (PPDU). Some values related to these parameters are also listed in Table 6.1.

In our simulation we used video as our main source of traffic. A contributed MPEG-4 video traffic generator in *ns-2* [30] is used in the simulations. A set of parameters, i.e., initial seed and rate factor, are available in this module to generate video flows with various mean/peak rates.

Table 6.1: WiMedia UWB parameters used in simulation

Parameter	Value
Physical bit rate	480 <i>Mbps</i>
Symbol Rate	3.2 <i>MHz</i>
Medium Access Slot	256 μs
SIFS Time	10 μs
Guard time	12 μs
Preamble Length	144 bits
Rx/Tx Turnaround Time	10 μs
PLCP Header Length	48 bits
PPDU Packet Size	4125 bytes
PPDU Maximum Payload Length	4095 bytes

To regulate the traffic each video traffic generator is attached to a twin token bucket filter (i.e., a traffic shaper). We extended the single token bucket filter module in *ns-2* to capture the properties of a twin token bucket introduced in Chapter 3. The filter follows the traffic parameters of the video source. We gathered these parameters from the video trace file and adopted the traffic shaper to match those parameters.

For the simulation, a two dimensional topology is used to simulate a WPAN. As shown in Figure 6.1, 10 UWB nodes are located on the circumference of a circle with a radius of 15 meters. A base station is located at the center of this circle. All devices are in a one-hop transmission range of the base station. All UWB nodes are transmitting video traffic to the base station.

6.2 Homogeneous Traffic

In this section, we evaluate the performance of homogeneous video streaming using simulation and analysis. We first outline the simulation methodology, verify the analytical models and later compare the performance of the proposed algorithms for the specified setting.

Table 6.2: Token bucket filter parameters

TSpec	MPEG-4 video traffic
Mean Rate (r)	11,497,884 bps
Peak Rate (p)	21,291,874.96 bps
Burst Size (b)	340,670 bits

6.2.1 Simulation Methodology

Video streaming in a home distribution network has an upper limit of 100 ms on the maximum tolerable delay [31]. In this section, the performance of homogeneous video traffic is evaluated. We gathered the video's parameters from the video trace file, which are listed in Table 6.2. These parameters are also used by the token bucket filter. Using the formulas given in Chapter 3, a flow with these parameters for its TSpec requires around 8 time slots to guarantee the QoS requirements.

We examined the video flows for three types of delay requirements: 50 *ms*, 100 *ms* and variable delay. The 50 and 100 *ms* delay emulate two-way and one-way video streaming scenarios, respectively. In the case of variable delay, each video flow picks its delay requirement from an interval of 50 to 100 *ms* with a uniform distribution. During the simulation, each UWB node will start video flows based on a Poisson arrival process, and each flow will stay in the system for a time duration following an exponential distribution.

As mentioned before, we used two performance metrics to compare the behaviour of each algorithm and their efficiency: blocking probability and system utilization. Another note about the simulation setup is that the best-fit algorithm, as mentioned in the previous chapter, tries to fit a request in one isozone whereas first-fit does cross zone allocation.

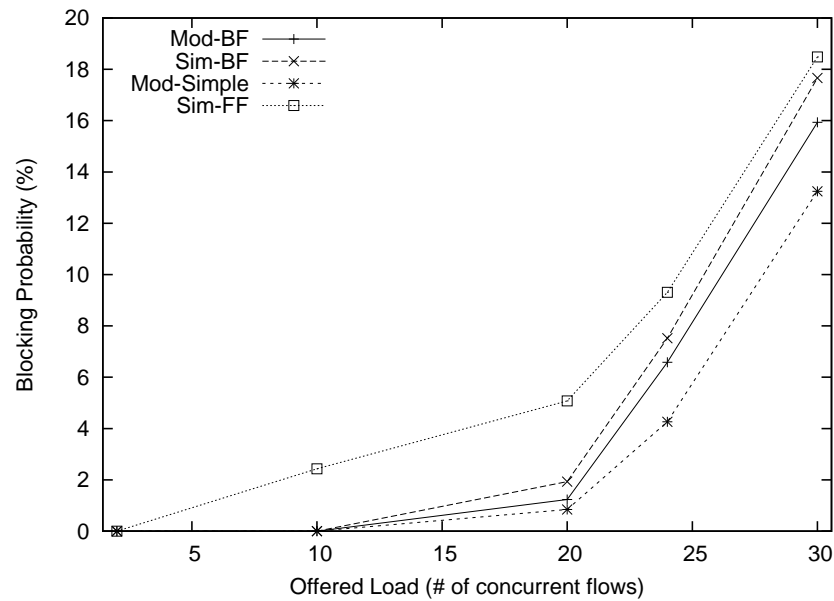


Figure 6.2: Blocking probability for homogeneous traffic with a delay bound of 50 *ms*: analysis and simulation.

6.2.2 Model Validation

We first obtain the analytical results following the approaches in Chapter 5, and compare them with the simulation results, to validate the analytical model. Requests are assumed to have a fixed size, and only the service interval is considered and not the queueing delay. All simulation results presented in this work is computed as the average of 10 runs. If the error bars are too small, they are not shown in the results. The request size is determined by the traffic generator and the search algorithms. Both the queueing delay and service interval are considered for the maximum delay bound.

Fixed Delay

Figure 6.2 shows the blocking probability for best-fit (BF) and first-fit (FF) by analysis (Mod) and simulation (Sim) as well as the simplified model (Simple) when the delay

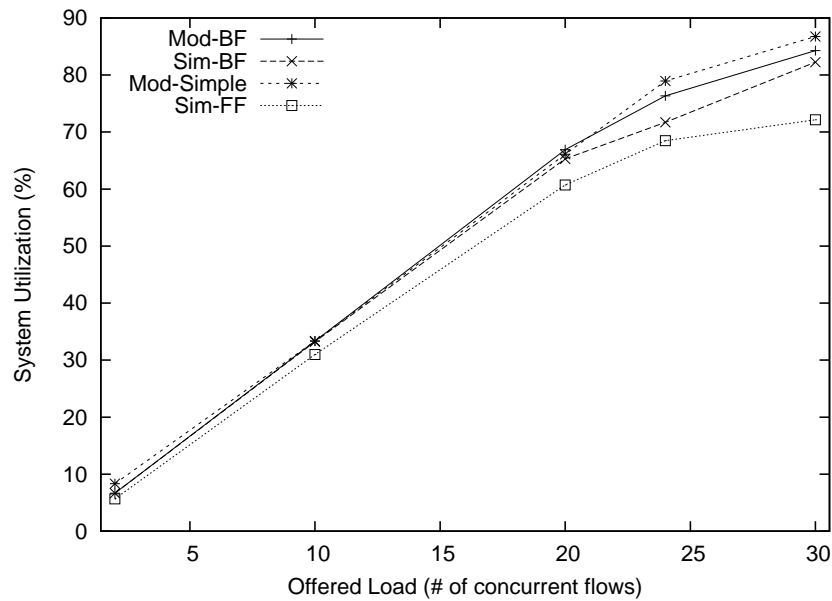


Figure 6.3: System utilization for homogeneous traffic with a delay bound of 50 *ms*: analysis and simulation.

bound is 50 ms. As it can be seen from the figure, the simplified model gives a lower bound for the blocking probability of the algorithms whereas the best-fit analytical model captures the real performance closely even when the offered load is increased to 30 requests. At this load, if all requests were accepted, the system would be fully utilized. However, due to the delay bound constraints, not all requests can be accommodated, leading to a considerable blocking probability.

Figure 6.3 shows the system utilization by both analysis and simulation. As discussed above, the simplified model analysis gives an upper bound for the system utilization. However, the best-fit analytical model can capture the real behaviour. Although the utilization increases with an increased offered load, after 20 concurrent flows, any further increase in offered load actually increases blocking probability rapidly, which leads to a much slower increase in the system utilization. By Figure 6.2 and Figure 6.3, it is shown that best-fit has a better performance in terms of a lower blocking probability and a higher system utilization than the first-fit.

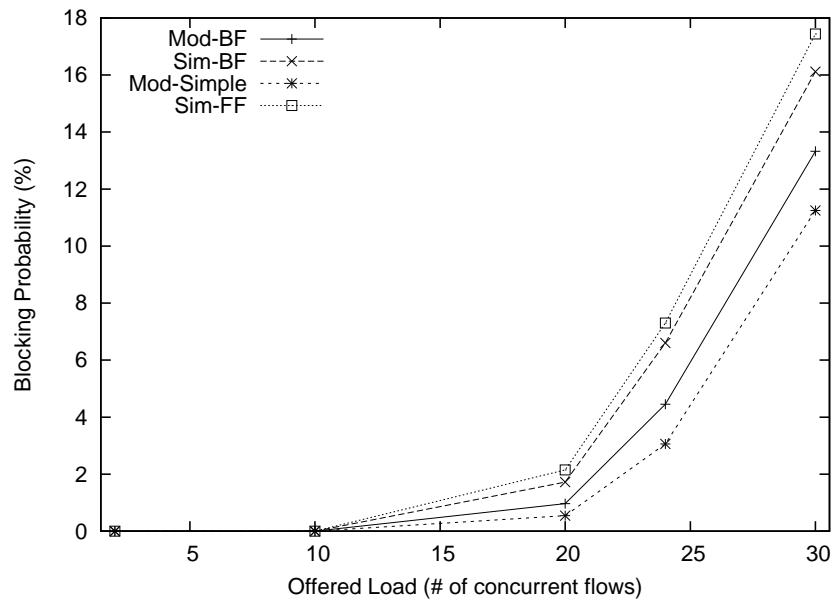


Figure 6.4: Blocking probability for homogeneous traffic with variable delay bounds: analysis and simulation.

Variable Delay

Here we show the results for the case where flows have variable delays. As mentioned before, flows pick delay bound based on a uniform distribution between 50 and 100 *ms* values. For the simplified model, as indicated in the previous chapter, the model structure requires a known value for the delay bound. Therefore, the combined average delay requirement of all flows is used for this model with an average delay requirement of 75 *ms*.

Figure 6.4 shows the blocking probability for best-fit and first-fit by analysis and simulation. Here, the simplified model again gives a lower bound for blocking probability of algorithms whereas the best-fit analytical model captures the real performance better when flows have variable delays. This figure shows that the best-fit model can capture the properties of the system.

Figure 6.5 shows both the analytical and simulation results of the system utiliza-

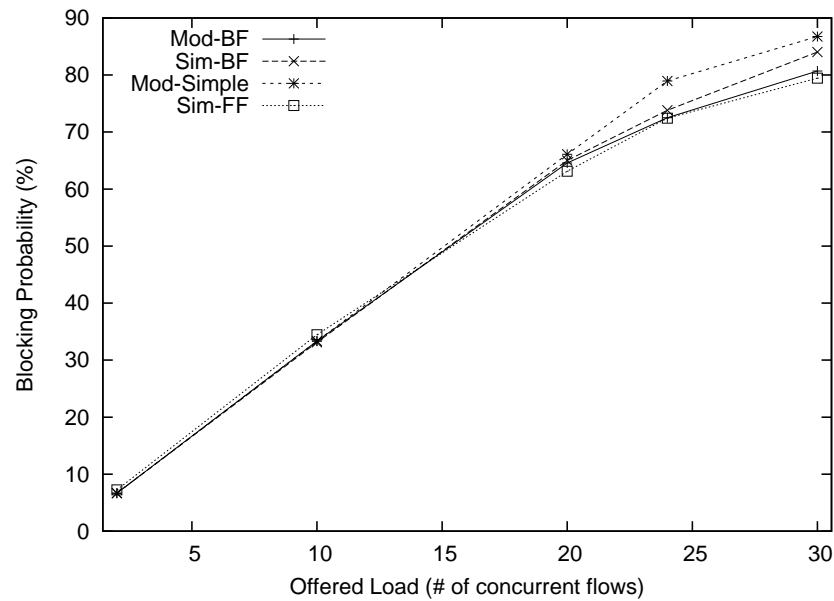


Figure 6.5: System utilization for homogeneous traffic with variable delay bounds: analysis and simulation.

tion. As mentioned above, the simplified model analysis gives an upper bound for the system utilization. However, the best-fit analytical model can capture the real behaviour better.

6.2.3 Performance Comparison of Algorithms

We further compare the best-fit and first-fit algorithms with different delay bounds. Here, only the simulation results are given, as it has been demonstrated that the best-fit analysis model has very good predictability on the real performance achieved by the simulation, and normally best-fit, which is our focus, outperforms first-fit.

Figure 6.6 shows the blocking probability for best-fit and first-fit with 50 and 100 *ms* and variable delay bound respectively. Tighter the delay bound, it is intuitive that it is more likely that the request cannot be accommodated, which leads to a higher blocking probability. This behaviour has been clearly indicated in Figure 6.6.

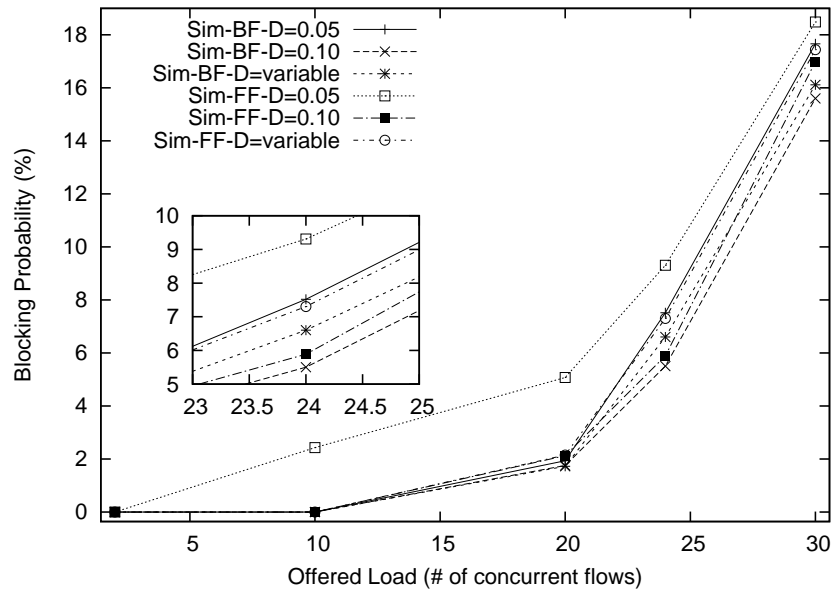


Figure 6.6: Blocking probability of homogeneous traffic: first-fit vs best-fit.

Furthermore, the first-fit algorithm has a higher blocking probability than best-fit, especially for tighter delay bounds, since best-fit tries to maintain the symmetry of the superframe, which leaves more flexibility to fit the follow-on requests. For the variable delay case, it is trivial that the performance of the algorithms is between the tighter, i.e., 50 *ms*, and the larger bound, i.e., 100 *ms*.

Figure 6.7 also shows the system utilization for the above cases. The best-fit algorithm allows the system to be more efficiently utilized than the first-fit, especially for tighter delay bounds.

In this section we presented simulation and analytical results for the case of homogeneous traffic in the given simulation scenario. We studied the algorithms using various delay requirements for video flows and verified the proposed models through simulations. In the next section we will study the same simulation scenario for heterogeneous traffic and will study the performance parameters of the system in various cases.

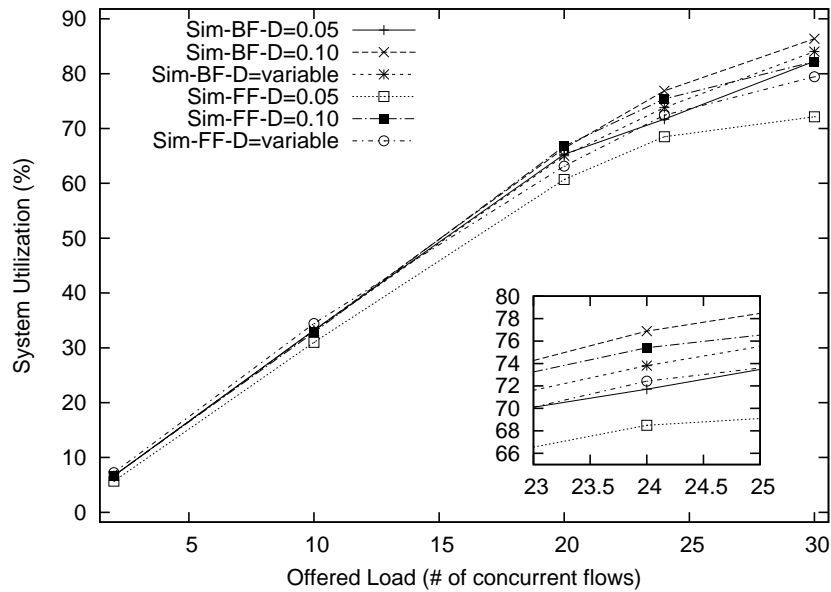


Figure 6.7: System utilization of homogeneous traffic: first-fit vs best-fit.

6.3 Heterogeneous Traffic

To further generalize the properties of the proposed algorithms and models, in this section, we evaluate the performance of heterogeneous video streaming using simulation and analysis.

6.3.1 Simulation Methodology

This section examines the performance of heterogeneous video traffic. This means that in the scenario described in Section 6.1, each UWB node will have more than one type of video flow for transmission.

Three different video flows are used for this purpose. We gathered video parameters by generating their trace files and calculating the parameters of interest. Table 6.3 shows the TSpec parameters for video flows (*Type 1-3*). We have chosen the video streams such that their QoS/TSpec requirements need a reservation of size 4, 8 and 12 MAS slots per superframe, respectively. These parameters are used by

Table 6.3: Token bucket filter parameters for heterogeneous traffic

Video flow	Type-1
TSpec	MPEG-4 video traffic
Mean Rate (r)	5,096,722.16 bps
Peak Rate (p)	9,289,374.96 bps
Burst Size (b)	148,630 bits
Video flow	Type-2
TSpec	MPEG-4 video traffic
Mean Rate (r)	11,497,884 bps
Peak Rate (p)	21,291,874.96 bps
Burst Size (b)	340,670 bits
Video flow	Type-3
TSpec	MPEG-4 video traffic
Mean Rate (r)	15,320,891.30 bps
Peak Rate (p)	38,632,500 bps
Burst Size (b)	618,120 bits

the token bucket filter as well.

Similar to the homogeneous case, we examined video flows for three types of delay requirements, 50, 100 *ms* delay limit and a case with a uniform distribution of delay values between 50 and 100 *ms*. Each node will uniformly choose one of these three videos and transmit it to the base station. During the simulation, each node would start its selected video flow based on a Poisson arrival process. Each flow will stay in the system for a time duration following an exponential distribution.

6.3.2 Model Validation

Similar to the homogeneous scenario, in this section, we first obtain the analytical results following the approaches for heterogeneous traffic explained in the previous chapter, and compare them with the simulation results, to validate the analytical models. In analysis, requests are assumed to have a fixed size, and only the service interval is considered, not the queueing delay. In simulation, the request size is determined by the traffic generator and the search algorithms. Both the queueing

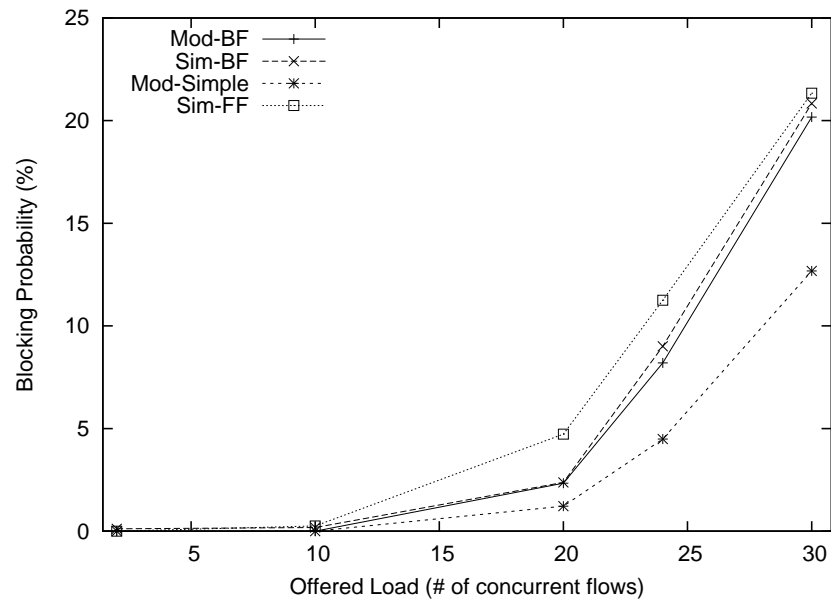


Figure 6.8: Blocking probability for heterogeneous traffic with a delay bound of 50 *ms*: analysis and simulation.

delay and service interval are considered for the maximum delay bound. We used two delay types for algorithms validation, 50 *ms* delay requirement for all flows and the case where flows pick a delay requirement value uniformly between 50 and 100 *ms*.

Fixed Delay

Figure 6.8 shows the blocking probability for the best-fit and first-fit algorithms by calculation and simulation and the simplified model when the delay bound is 50 *ms*. The blocking probability shown in this figure is the total blocking probability for all piconet regardless of the flow type. As it can be seen in the figure, the simplified model gives a lower bound for the blocking probability of the algorithms whereas the best-fit analytical model captures the real performance. Nodes pick their flow type with a uniform distribution, so on average, each node will reserve around 8 MAS slots per superframe to carry its flows. In that case, at the offered load of 30, the system should be fully utilized. However, due to the delay bound constraints, not all requests

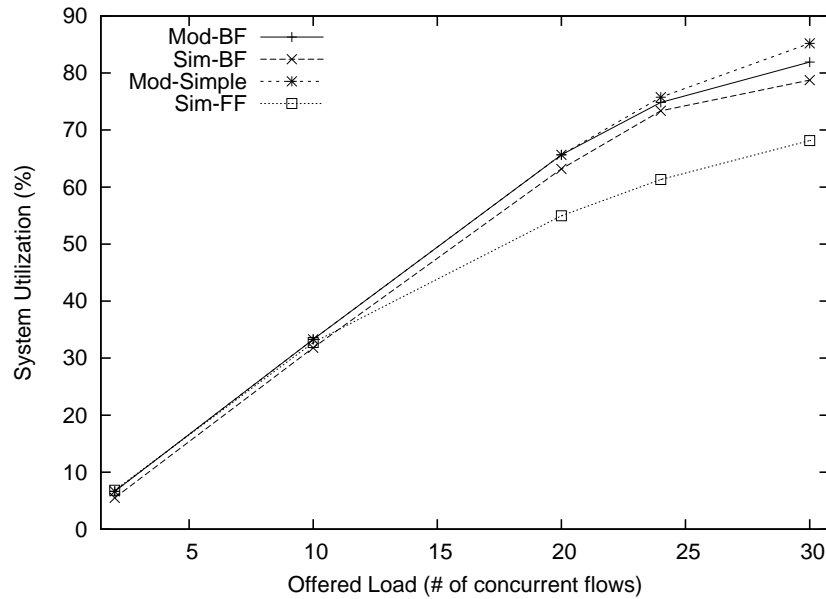


Figure 6.9: System utilization for heterogeneous traffic with a delay bound of 50 *ms*: analysis and simulation.

can be accommodated, leading to a considerable blocking probability.

Figure 6.9 shows analytical and simulation results for the system utilization. The simplified model analysis gives an upper bound of the system utilization. However, the best-fit analytical model can capture the real behaviour closely. Although the utilization increases with an increased offered load, after 20 concurrent flows, any further increase in the offered load actually increases blocking probability rapidly, which leads to a much slower increase in the system utilization. By Figure 6.8 and Figure 6.9, it is shown that best-fit always has a better performance in terms of a lower blocking probability and a higher system utilization than the first-fit. We also show the analytical and simulation results of the blocking probability of each flow (*Type1-3*) for both algorithms. Figure 6.10 shows the results of the best-fit algorithm. With an increase in the offered load, the blocking probability of all types of flows increases. It can also be seen that the blocking probability of flows increases as the request size increases from 4 to 12. Another observation is that the model gives a lower bound

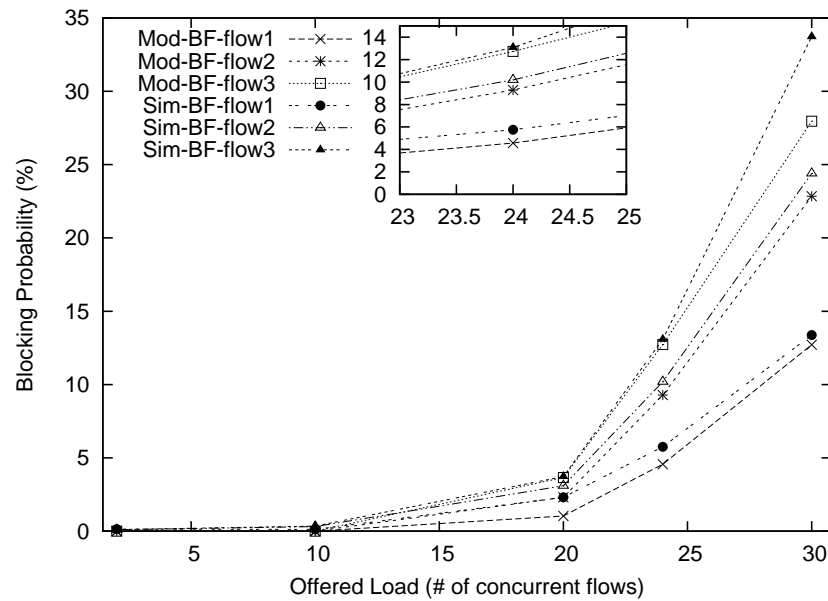


Figure 6.10: Best-fit, flow-based blocking probability for heterogeneous traffic with a delay bound of 50 *ms*: analysis and simulation.

for each flow's blocking probability. The same case holds for the first-fit algorithm as shown in Figure 6.11. We have used the simplified model's results to validate the simulation results of the first-fit algorithm. Again the model gives a lower bound for the blocking probability of each flow.

The performance of both the first-fit and the best-fit algorithm seems to be close. One reason is the request size of heterogeneous flows. Due to the symmetric structure, the best-fit algorithm performs well but on the other hand, since the first-fit algorithm reserves time slots wherever it finds empty time slots, it can utilize the small gaps between various allocations and perform well.

Variable Delay

In this section we show the results for the case of flows with variable delay. As mentioned, flows are assigned delay requirements based on a uniform distribution with the minimum value of 50 and the maximum value of 100 *ms*. The simplified

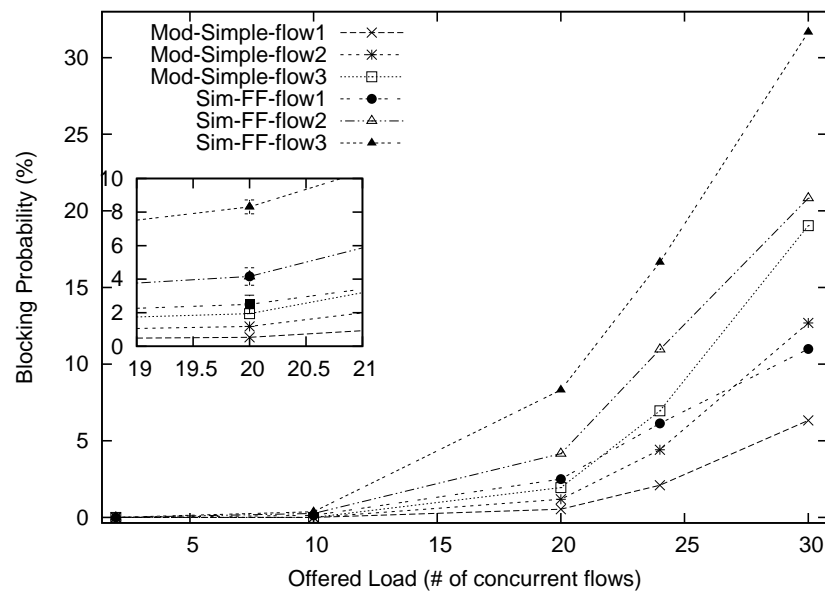


Figure 6.11: First-fit, flow-based blocking probability for heterogeneous traffic with a delay bound of 50 ms: analysis and simulation.

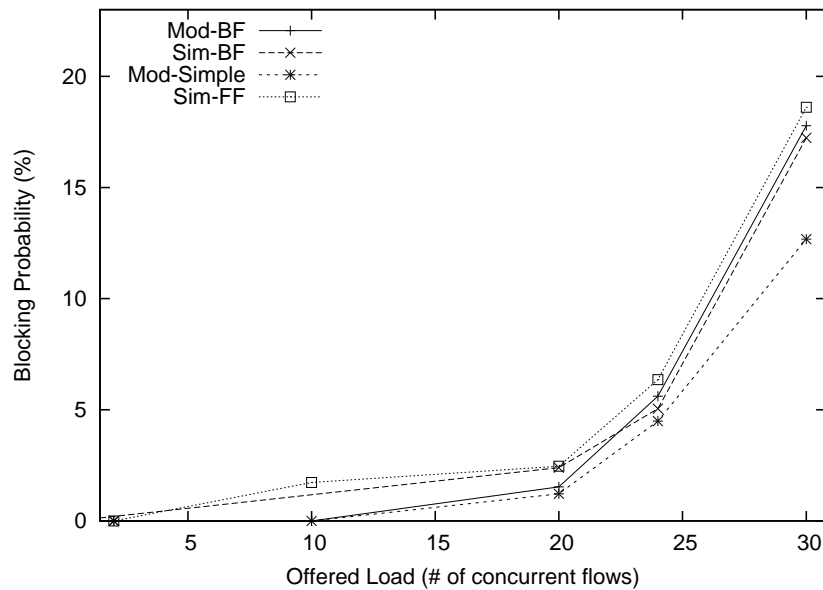


Figure 6.12: Blocking probability for heterogeneous traffic with variable delay bounds: analysis and simulation.

model, as discussed in the previous chapter, can only work with one delay value for all flows. The structure of the model is based on this delay value. To analyze the heterogeneous traffics with variable delay, we use the average of delay values to build the model; therefore in this case we assume all requests have a delay request of 75 ms .

Figure 6.12 shows the analytical and simulation results of the blocking probability for best-fit and first-fit. In this case, the simplified model gives a lower bound for the blocking probability of the algorithms whereas the best-fit analytical model captures the real performance of the best-fit algorithm closely even when the flows have variable delays.

Figure 6.13 also shows the system utilization from both the analysis and simulation. As mentioned above, the simplified model analysis has an upper bound on the system utilization. However, the best-fit analytical model can capture the real behaviour closely.

We also show the simulation and analysis results for the blocking probability with

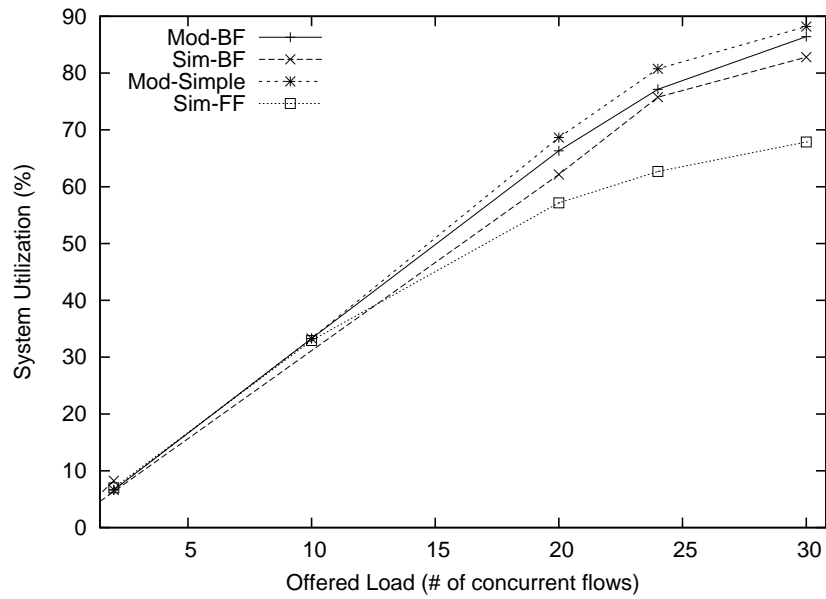


Figure 6.13: System utilization for heterogeneous traffic with variable delay bounds: analysis and simulation.

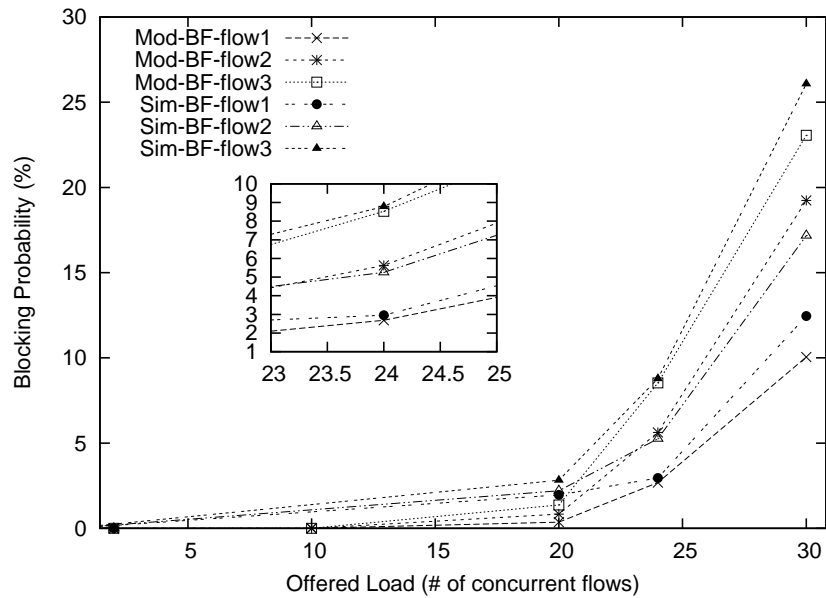


Figure 6.14: Best-fit, flow-based blocking probability for heterogeneous traffic with a variable delay bounds: analysis and simulation.

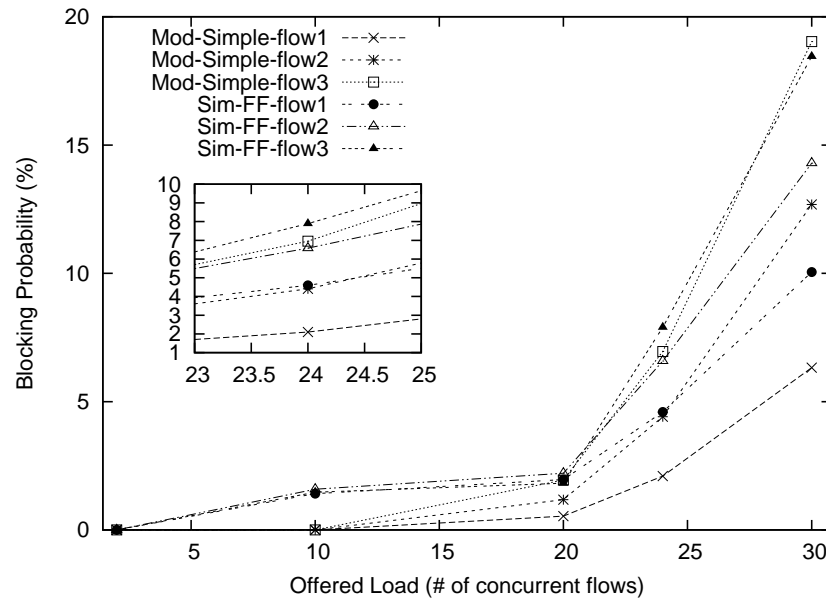


Figure 6.15: First-fit, flow-based blocking probability for heterogeneous traffic with a variable delay bounds: analysis and simulation.

best-fit and first-fit of each flow (sizes 4, 8 and 12) in the case of variable delay in Figure 6.14 and Figure 6.15, respectively. Similar to the case of flows with a fixed delay of 50 *ms*, blocking probability for each flow increases with an increase in the offered. Also, increasing the size of the flow will result in an increase of the blocking probability accordingly.

As mentioned for the fixed delay scenario, the performance of both first-fit and best-fit algorithms seems to be close here too. Again the wide range of request sizes, i.e., 4, 8 and 12, helps the first-fit algorithm utilize the superframe's time slots better and has a close performance to the best-fit algorithm.

6.3.3 Performance Comparison of Algorithms

In the previous section we showed the simulation and analytical results to validate the algorithms and their models. In this section, we compare the best-fit and first-

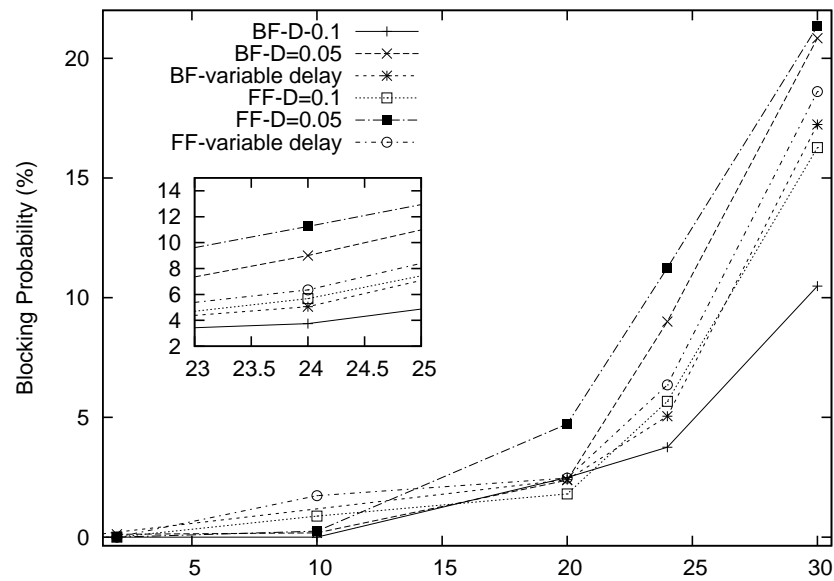


Figure 6.16: Blocking probability of heterogeneous traffic: first-fit vs best-fit.

fit algorithms with different delay bounds with heterogeneous traffic. Figure 6.16 shows the blocking probability for best-fit and first-fit with 50, 100 *ms* and variable delay bounds respectively. It is intuitive that tighter the delay bound, it is a higher chance that the request cannot be accommodated, which leads to a higher blocking probability.

This behaviour can be clearly observed from these figures. Furthermore, the first-fit algorithm has a higher blocking probability than best-fit, especially for tighter delay bounds. This is due to the fact that best-fit is trying to maintain the symmetry of the superframe that leaves much more flexibility to follow-on requests. For the variable delay case, it is trivial to observe that the performance of the algorithms is between the tighter, i.e., 50 *ms*, and the looser, i.e., 100 *ms* delay cases.

Figure 6.17 shows the system utilization for the above cases. The best-fit algorithm can allow the system to be more efficiently utilized than the first-fit, especially for tighter delay bounds.

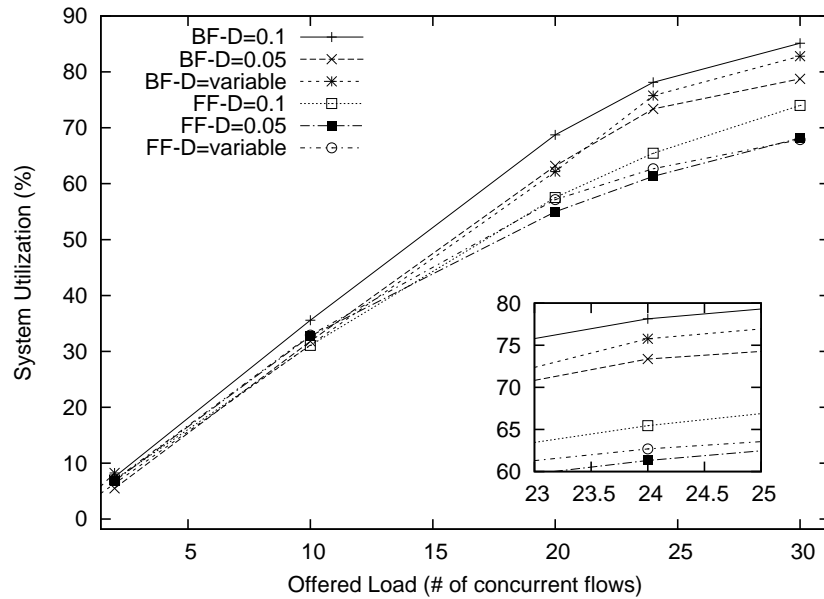


Figure 6.17: System utilization of heterogeneous traffic: first-fit vs best-fit.

In this section we presented the simulation and analytical results for the case of heterogeneous traffic in piconet. We studied and checked the validity of the proposed models and algorithms using four different types of flows as the main source of traffic in the piconet. We also studied the cases with various delay requirements for the flows. We compared the blocking probability of flows both individually and collectively. Also we compared the performance of the algorithms for various delay requirements using simulation results.

6.4 Summary

In this chapter, we have investigated the performance of MPEG-4 video streaming over UWB wireless networks by simulation. We extended *ns-2*'s wireless package to match WiMedia UWB properties and added DRP MAC to it. We used simulation for two main reasons: first to validate the system models proposed in Chapter 5 and also to investigate the validity of the proposed algorithms introduced in Chapter 4.

Compared to first-fit, both simulation and analytical results of the best-fit algorithm confirmed the predicted behaviour. Best-fit algorithm seemed to be a promising reservation algorithm for UWB DRP. Under the two main scenarios studied in this work, i.e., homogeneous and heterogeneous traffic, the best-fit algorithm had a better performance in terms of superframe utilization and blocking probability. The best-fit algorithm is specially appropriate for the cases where flows have tighter delay bounds.

The model proposed for the best-fit algorithm also captured the properties of the algorithm closely. The results observed through analysis matched closely with the simulation results. Compared to the simplified model, the best-fit model was able to incorporate various properties of the system. The model also worked very well for the heterogeneous traffic with individual delay requirements for the flows.

Chapter 7

Conclusion and Future Work

Ultra-wideband (UWB) technologies, with high data rates and low transmission power over shorter ranges, have enabled a new set of applications in Wireless Personal Area Networks (WPANs). WiMedia standard for UWB MAC has proposed a distributed reservation-based protocol called DRP. In this research, we first identified the needs of distributed reservation algorithms by studying the existing wireless medium access methods. We proposed two distributed reservation algorithms for WiMedia UWB MAC. Our algorithms considered both the traffic specification and QoS requirement of video flows along with WiMedia MAC reservation policies. One algorithm allocated time slots based on the first-fit idea whereas the other took one step further by doing a best-fit reservation according to the maximum tolerable delay bound. The second algorithm could utilize the medium time well and had a lower blocking probability compared with the first one.

We also proposed two models to capture the properties of the reservation algorithms and WiMedia UWB MAC structure. The simplified model only considered the request size and delay requirement of flows whereas the best-fit model was a more specific tiered overflow model that closely captured the properties of both the algorithm

and superframe. We performed simulations to further verify the proposed algorithms, and models and demonstrated the difference between the proposed algorithms in a more realistic WPAN environment. The simulation was conducted under two types of settings, one was for homogeneous traffic and the other was for heterogeneous traffic.

The main contributions of this thesis have two main aspects. First, we proposed a model for UWB-MAC DRP and incorporated the reservation algorithms into this model. Our model was suitable to justify and analyze the performance of the reservation algorithms. Second, we performed simulation for video streaming in WPANs. The simulation was used to first investigate the validity of the system models and to compare the performance of the algorithms.

The main conclusions are as follows:

1. WiMedia UWB is a promising wireless technology for IPTV in-home distribution. The DRP part of this standard can be greatly utilized for QoS guarantee of video flows.
2. Among the proposed reservation algorithms in this work, simulation and analytical results of the best-fit algorithm implied that it is a better reservation strategy. Even though the first-fit algorithm's performance was close to best-fit, overall the strategies the best-fit algorithm follows led to better results compared to the first-fit algorithm.
3. The simplified model is a good analytical tool to predict and justify the behaviour of the reservation algorithms. Compared to the simplified model, the model proposed for the best-fit algorithm captured the properties of both the algorithm and the WiMedia UWB MAC structure closely.

7.1 Further Research Issues

To have a complete reservation methodology for DRP WiMedia MAC, there are many related interesting issues that need further investigation. There are three main open issues that need to be addressed:

1. In our proposed reservation algorithms, flows used the same time slots for transmission during their existence in the system. It will be interesting if our reservation algorithms can also try to reallocate the reserved blocks. If we compact the reserved blocks, we can avoid segmentation and therefore reduce the blocking probability and better utilize the superframe.
2. The best-fit algorithm only makes reservations in one isozone of the superframe. It can take one step further and make reservations between multiple isozones. Instead of just dropping requests to higher isozones, request can be split to lower isozones as well. By cross-isozone allocation, delay requirement can still be met while achieving a higher system utilization. This improvement also requires proper adjustments to the best-fit system model.
3. The main focus of this work was on the DRP-only MAC. In the future, it will be interesting to combine DRP with PCA for a better utilization of the channel. In our calculations, peak rate was an important factor in computing the number of time slots a flow needed for successful transmission. Flows are not always transmitting at their peak rate. Flows will have more reserved time slots than the ones actually needed for data transmission in those slots. In future, it will be interesting to try to reserve time slots at the average data rate of video flows in DRP and the rest carried through the PCA part of WiMedia MAC.

Based on the above conclusions, the proposed reservation algorithms, specially the best-fit algorithm, is a promising reservation methodology for UWB WiMedia MAC.

Many opportunities for further improvement still exist. We outlined a few key issues in this chapter in the hope of inspiring further research interests in the this new area.

Bibliography

- [1] IEEE Computer Society, *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)*, 2003.
- [2] Standard ECMA-368, *High Rate Ultra Wideband PHY and MAC Standard*, 2007.
- [3] Wimedia Alliance, *Wimedia Logical Link Control Protocol*, 2007.
- [4] “Network Simulator Version 2,” <http://www.isi.edu/nsnam/ns>.
- [5] K. Kerpez, D. Waring, G. Lapiotis, J. Lyles, and R. Vaidyanathan, “IPTV Service Assurance,” *Communications Magazine, IEEE*, vol. 44, no. 9, pp. 166–172, 2006.
- [6] T. Zahariadis, K. Pramataris, and N. Zervos, “A Comparison of Competing Broadband in-home Technologies,” *Electronics & Communication Engineering Journal*, vol. 14, no. 4, pp. 133–142, 2002.
- [7] IEEE 802.11 Working Group, “IEEE Std. 802.11-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications,” 1999.
- [8] IEEE 802.11 Working Group, “IEEE Std. 802.11b-1999: Higher Speed Physical Layer Extension in the 2.4GHz band,” 1999.

- [9] IEEE 802.11 Working Group, "IEEE Std. 802.11g-2003: Further Higher Data Rate Extension in the 2.4 GHz band," 2003.
- [10] IEEE 802.11 Working Group, "IEEE Std. 802.11e-2005, Amendment 8: Media Access Control (MAC) Quality of Service Enhancements," 2005.
- [11] R. Shepherd, "Bluetooth wireless technology in the home," *Electronics and Communication Engineering Journal*, vol. 13, no. 5, pp. 195–203, 2001.
- [12] J. S. Lee, Y. W. Su, and C. C. Shen, "A Comparative Study of Wireless Protocols: Bluetooth, UWB, Zigbee, and Wi-fi," *IEEE Conference of the Industrial Electronics Society*, pp. 46–51, Nov. 2007.
- [13] J. Del Prado Pavon, S. Shankar N, V. Gaddam, K. Challapali, and C.-T. Chou, "The Mboa-WiMedia Specification for Ultra Wideband Distributed Networks," *IEEE Communication Magazine*, vol. 44, pp. 128–134, 2006.
- [14] E. Hossein and V. K. Bhargava, "A Centralized TDMA-based Scheme for Fair Bandwidth Allocation in Wireless IP Networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 2201–2214, 2001.
- [15] P. Djukic and S. Valaee, "Distributed Link Scheduling for TDMA Mesh Networks," *IEEE International Conference on Communications Proceeding*, 2007.
- [16] Y. Chu and A. Ganz, "Adaptive Mac Protocol for QoS Support in UWB-based Wireless Networks," in *IEEE International Conference on Communications*, vol. 9, pp. 4040–4045, 2004.
- [17] W. Cui, P. Ranta, T. A. Brown, and C. Reed, "Wireless Video Streaming Over UWB," *IEEE Conference on Ultra-Wideband*, pp. 933–936, 2007

- [18] H. Wu, Y. Xia, and Q. Zhang, "Delay Analysis of DRP in MBOA UWB Mac," *IEEE International Conference on Communications Proceeding*, pp. 229–233, 2006
- [19] K. H. Liu, X. Shen, R. Zhang, and L. Cai, "Delay Analysis of Distributed Reservation Protocol with UWB Shadowing Channel for WPAN," *IEEE International Conference on Communications Proceeding*, pp. 2769–2774, 2008.
- [20] R. Ruby, Y. Liu, and J. Pan, *Evaluating Video Streaming Over Wireless Networks, ACM workshop on Wireless Multimedia Networking and Performance Modeling*, 2008.
- [21] P. Wilson, M. Johnstone, M. Neely, and D. Boles, "Dynamic Storage Allocation: A Survey and Critical Review," Springer, vol. 989, pp. 1–116, 1995
- [22] R. Ruby, "Performance Analysis of WiMedia UWB Protocols," M.Sc. thesis, University of Victoria, Victoria, Canada, 2009
- [23] L. B. Jean-Yves and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. Springer, 2004.
- [24] U. M. Jame Roberts and J. Virtamo, eds., *Broadband Network Teletraffic, Final Report of Action COST 242*, vol. 1155. Springer, 1996.
- [25] L. Delbrouck, "A Unified Approximate Evaluation of Congestion Functions for Smooth and Peaky Traffics," *IEEE Transactions on Communications*, pp. 85–91, 1981.
- [26] M. Beshai, "The Poissonian-Spectrum Method for Treating a Loss System Serving Non-Poissonian Multi-Bit-Rate Traffic," *Annual Joint Conference of the IEEE Computer and Communications Proceeding*, pp. 1010–1020, 1989.

- [27] R. Wilkinson, “Theories for Toll Traffic Engineering in the U.S.A,” *Bell Syst. Tech.J.*, vol. 35, no. 2, pp. 421–514, 1956.
- [28] M. Glabowski, K. Kubasik, and M. Stasiak, “Modelling of Systems with Overflow Multi-rate Traffic and Finite Number of Traffic Sources,” *International Symposium on Communication Systems, Networks and Digital Signal Processing*, pp. 196–199, 2008.
- [29] “NO Ad-Hoc Routing Agent (NOAH) for NS-2,”
<http://icapeople.epfl.ch/widmer/uwb/ns-2/noah>.
- [30] A. Matrawy, I. Lambadaris, and C. Huang, “Mpeg4 Traffic Modeling Using the Transform Expand Sample Methodology,” *IEEE International Workshop on Networked Appliances*, pp. 249–256, 2002.
- [31] International Telecommunication Union, *ITU-T Recommendation G.9954, Home Networking Transceivers - Enhanced Physical, Media Access, and Link Layer Specifications*, 2007.