QUALITY OF SERVICE SUPPORT IN IEEE 802.11 WIRELESS LAN

by

Wasan Pattara-Atikom

M.B.A, University of Pittsburgh, 2003

M.S. in Telecommunications, University of Pittsburgh, 2000

B.Eng. in Computer Engineering, Khon Kaen University, 1995

Submitted to the Graduate Faculty of

the School of Information Sciences and Telecommunications in

partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2005

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES AND TELECOMMUNICATIONS

This dissertation was presented

by

Wasan Pattara-Atikom

It was defended on

March 31st 2005

and approved by

Prof. Sujata Banerjee

Prof. Prashant Krishnamurthy

Prof. David Tipper

Prof. Nitin Vaidya

Prof. Martin Weiss

Dissertation Advisors: Prof. Sujata Banerjee,

Prof. Prashant Krishnamurthy (Co-Chair)

QUALITY OF SERVICE SUPPORT IN IEEE 802.11 WIRELESS LAN

Wasan Pattara-Atikom

University of Pittsburgh, 2005

Wireless Local Area Networks (WLANs) are gaining popularity at an unprecedented rate, at home, at work, and in public hot spot locations. As these networks become ubiquitous and an integral part of the infrastructure, they will be increasingly used for multi-media applications. The heart of the current 802.11 WLANs mechanism is the Distributed Coordination Function (DCF) which does not have any Quality of Service (QoS) support. The emergence of multimedia applications, such as the local services in WLANs hot-spots and distributions of entertainment in residential WLANs, has prompted research in QoS support for WLANs.

The absence of QoS support results in applications with drastically different requirements receiving the same (yet potentially unsatisfactory) service. Without absolute throughput support, the performance of applications with stringent throughput requirements will not be met. Without relative throughput support, heterogeneous types of applications will be treated unfairly and their performance will be poor. Without delay constraint support, time-sensitive applications will not even be possible. The objective of this dissertation is, therefore, to develop a comprehensive and integrated solution to provide effective and efficient QoS support in WLANs in a distributed, fair, scalable, and robust manner.

In this dissertation, we present a novel distributed QoS mechanism called Distributed Relative/Absolute Fair Throughput with Delay Support (DRAFT+D). DRAFT+D is designed specifically to provide integrated QoS support in IEEE 802.11 WLANs. Unlike any other distributed QoS mechanism, DRAFT+D supports two QoS metrics (throughput and delay) with two QoS models (absolute and relative) under two fairness constraints (utilitarian and temporal fairness) in the same mechanism at the same time a fully distributed

manner. DRAFT+D is also equipped with safeguards against excessive traffic injection. DRAFT+D operates as a fair-queuing mechanism that controls packet transmissions (a) by using a distributed deficit round robin mechanism and (b) by modifying the way Backoff Interval (BI) are calculated for packets of different traffic classes. Fair relative throughput support is achieved by calculating BI based on the throughput requirements. Absolute throughput and delay support are achieved by allocating sufficient shares of bandwidth to these types of traffic.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3G	Third-Generation
ABAC-RT	Adaptive Bandwidth Auto-Control for Relative Throughput Flows
AC	Access Category
ACK	Acknowledgment
AD	Absolute Delay Support
AD-MS	Mobile Station with Absolute Delay Requirement
AIFS	Arbitrary IFS
AIFSN	Arbitrary IFS Number
AP	Access Point
AQ	Absolute QoS Support
ARME	Assured Rate MAC Extension
AT	Absolute Throughput Support
AT-MS	Mobile Station with Absolute Throughput Requirement
ATM	Asynchronous Transfer Mode
BI	Backoff Interval
BS	Base Station
BSS	Basic Service Set
CAGR	Compounded Annual Growth Rate
CBR	Constant Bit Rate
COV	Coefficient of Variation
CSMA	Carrier Sense Multiple Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CW	Contention Window

CWD	Contention Window Differentiation
CWS	Contention Window Separation
DC	Deficit Counter
DCF	Distributed Coordination Function
DDFC	Decentralized Delay Fluctuation Control Mechanism
DFS	Distributed Fair Scheduling
DIFS	DCF IFS
DRAFT+D	Distributed Relative/Absolute Fair Throughput with Delay Support
DRR	Deficit Round Robin
DSG-AT	Distributed SafeGuard for Absolute QoS Support in Absolute Throughput Flows
DSG-RT	Distributed SafeGuard for Absolute QoS Support in Relative Throughput Flows
DWFQ	Distributed Weighted Fair Queuing
EDCA	Enhanced Distributed Channel Access
EDCF	Enhanced DCF
ETSI	European Telecommunications Standards Institute
EWMA	Exponentially Weighted Moving Average
HCF	Hybrid Coordination Function
HIPERLAN	HIgh PErformance Radio Local Area Network
HoQ	Head-of-Queue
IBSS	Independent Basic Service Set
IDC	International Data Corp
IFS	InterFrame Space
LAN	Local Area Network
MAC	Medium Access Control
MS	Mobile Station
MSDU	MAC Service Data Unit
NIC	Network Interface Card
PCC	Priority-Based Contention Control
PCF	Point Coordination function
PER	Packet Error Rate

PF	Persistence Factor
PIFS	PCF IFS
QoS	Quality of Service
RT	Relative Throughput Support
RT-MS	Mobile Station with Relative Throughput Requirement
RTS/CTS	Request to Send/Clear to Send
SCFQ	Self-Clocked Fair Queue
SIFS	Short Inter-frame Space
TDMA	Time Division Multiple Access
TXOP	Transmission Opportunity
VBR	Variable Bit Rate
VIFS	Variable Length IFS
VoIP	Voice over IP
VoWLAN	Voice over WLAN
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Networks

PREFACE

At every stage of our lives, there are people who teach us, help us, and inspire us. First, I would like express my deepest appreciation to my advisors, Dr. Sujata Banerjee and Dr. Prashant Krishnamurthy. I am truly grateful to Dr. Banerjee for her insight, guidance, and encouragement throughout the course of my study. Even after having accepted a new position at Hewlett-Packard Laboratory in California, she continued to work with me and provided valuable advice without which this study would not have been completed. I am greatly indebted to Dr. Krishnamurthy who, despite his busy schedule, always took the time to meet and consult with me and provided substantive feedback and advice on various aspects of my study. I am sincerely appreciative for his constant inspiration and his academic, financial, and emotional support. I was quite fortunate to have two advisors. I cannot think of a better combination than both of them. I would not have been able to sustain work on this study without their light to guide me.

I would like to thank the members of my dissertation committee, Dr. Martin Weiss, Dr. David Tipper, and Dr. Nitin Vaidya, for their valuable feedback during the development of the dissertation proposal and at the dissertation defense. Their suggestions help me to understand more completely the implications and importance of my dissertation research.

I also want to thank all my friends for many fruitful discussions. A special thank to Dr. Richard Donato for his time, his countless language consultations, and his numerous personal and professional advice. Lastly, I would like to extend my appreciation to my parents and to my brother for their love and constant support which inspired me to continue my work.

As many people have said, a dissertation is not an easy task. However, what is often not stated is that, with supportive advisors, good friends, and a loving family, writing a dissertation can be an exciting and enlightening experience.

1.0 INTRODUCTION

Wireless Local Area Networks (WLANs) have emerged as an untethered alternative for access networks. As the price of WLAN's Network Interface Cards (NICs) dropped from $695 in 1999 to $24.99 in 2004 [1], WLANs have rapidly become the networks of choice. The strong and growing demand for WLANs in both consumer markets such as residential networks [2, 3] and industrial markets such as retail, education, health care and wireless hot-spots in hotels, airports, and restaurants [1, 4] has been documented repeatedly in business, industry, and education.

One of the many reasons for the widespread popularity of WLANs is that WLANs allow users to access networks without physically being attached to them. Users can work wherever, in their offices, warehouses or backyards, and whenever they want without being constrained by a physical tethered network connection. Other advantages of WLANs [1] are reduced errors in health-care facility, improved profitability, flexibility, time savings, and cost savings. Increased productivity by as much as 22% was also reported in [5]. The author in [1] suggested that pervasive high-speed wireless data services are both compelling and inevitable. It is just a question of how and when. And if we know the answer to how, then it is only a matter of time until WLANs will become a ubiquitous reality.

Recent statistics from Business Communications Review [6] reported that the worldwide WLAN market grew more than 200 percent from 2000 to 2002. Infonetics Research [7] reported that the worldwide market revenue of WLAN is expected to grow to $786.2 million by 2Q05. WLAN revenues are expected to grow to $2.6 billion in 2005 [8]. This number is estimated to reach $8.6 billion in 2008 [9]. According to Infonetics Research, the worldwide units will grow at 131% per annum [7]. It reported that there were 39 million WLAN users at the end of 2004, compared to 1.185 billion fixed lines and 1.5 billion mobile phones. The number of WLAN users is expected to grow threefold and reach 120 million by 2008 [9].

International Data Corp (IDC) [10] also reported that the market for hotspots grew 13 times in term subscribers during the year of 2002 and 2003 and it is also expected to reach 7 million subscribers by 2008 with total revenues of $600 million. In-Stat/MDR [11] reported that WLAN segment is the fastest growing segment of the networking products industry. The explosive growth in WLANs in the past few years has put Atheros Communications, Inc., a leading developer of WLAN chipsets, as the fastest growing private company in Silicon Valley [12].

In addition to data applications, WLANs has begun to attract attention from the cellular community. Cellular service providers now perceive WLANs and Third-Generation (3G) cellular technologies as complementary rather than competitive. The cellular service provides seamless coverage and mobility, while WLANs provide high speed connections in selected areas [13, 14]. Integrated WLAN mobile handsets have been introduced by many leading companies, such as SyChip [15], NEC [16], and Cisco [17]. The first WLAN/cellular handsets are already available since the last quarter of 2004. The adoption of WLAN/cellular handsets is expected to grow dramatically by 2009 [18]. Using 802.11 WLANs as the last mile alternative has also been suggested in [19]. Several market forecasts [4, 13, 20, 21] report that the WLANs market will continue to grow tremendously in the next decade.

1.1 BACKGROUND TO THE STUDY

Applications that have been traditionally used in wired-LANs are also gradually and increasingly being used in WLANs. The benefits of WLANs are not limited to data networking among computers. As home networking begins to gain popularity, home entertainment (audio/video) will become an important application making the need to unwire the living room as the next target. In the years ahead, residential wireless networks will become integral parts of digital homes and will bridge the data network interconnecting desktop PCs, mobile laptops and handhelds, and the consumer electronics network handling multimedia distribution among High-Definition TVs (HDTVs), DVD payers, digital cameras, camcorders, videogame consoles, cellphones and emerging Internet appliances [22]. There is recent interest in using Voice over WLAN (VoWLAN) [23, 24] as a low cost alternative for voice communications, or as extension to cellular networks, e.g. Voice over IP (VoIP) over WLANs [25, 26]. Some applications such as email, web access, and data transfer can tolerate high loss and delay, but other applications such as multimedia and Quality of Service (QoS)-sensitive (delay-constraint or throughput-specific) applications cannot. The recent survey in [1] confirmed that Quality of Service (QoS) support is among the top 10 most important features of WLANs. This has created an urgent need for QoS support in WLAN.

In the past decade, there have been tremendous efforts to enhance and incorporate QoS over the Internet. Specifically, two main approaches have been proposed, namely Integrated Service (IntServ) [27] and Differentiated Service (DiffServ) [28]. While the main idea of IntServ is to provide end-to-end per-flow QoS guarantees, the main idea behind DiffServ is to provide simple service differentiation according to specified service types. Since we have potential QoS support at the IP layer, one might question why we still need QoS support mechanisms at the MAC layer. Although, IntServ can potentially provide the QoS support without a need for QoS mechanisms at the MAC layer, the interest in using IntServ has subsided due to issues of scalability, complexity, and practicality. In contrast, DiffServ is simple and does not require end-to-end signaling. However, it is only applicable within the core IP networks. In its present form, DiffServ does not address the issue of QoS in edge networks, such WLANs[29].

The bandwidth of WLANs is typically orders of magnitude lower than traditional wired-LANs (54 Mbps compared to 10 Gbps). Due to the significantly limited bandwidth, congestion is more likely in WLANs. Although there is no QoS support in widely adopted wired-LANs (such as Ethernet, Fast Ethernet, and Gigabit Ethernet), the emerging local services, particularly multimedia services, both in WLANs hotspots and residential networks make limited use of the wired backbone and hence QoS support over the wireless link becomes important. Even though end-to-end QoS may not be available, several services in WLANs have local traffic (e.g., from a DVD player in a living room to an HDTV set in the bedroom). This traffic may not depend on wired-QoS since they may never leave the wireless portion of the network. However, this traffic will require QoS support in the WLAN.

Without QoS support, the performance of multimedia and QoS-sensitive applications will be severely degraded and eventually become jeopardized. The next generation killer applications are not yet determined but they are expected to be multimedia based. *The provision of QoS support in WLANs is the focus of this research study.*

1.2 DESIRED PROPERTIES OF QOS MECHANISMS IN WLANs

Based on our observations of the current mechanisms available in the state of the art and proposed in the literature, we compiled the following list of desired properties for any QoS support mechanism in WLAN. Often, there is a trade-off between these properties. While some properties are achieved in some mechanisms, other properties are sacrificed. We believe a balance of these properties is required to achieve full benefits of QoS support is one of the most crucial research tasks to effectively and efficiently deliver QoS support in WLANs.

- The mechanism should provide QoS support with a *high degree of fairness*, particularly when it provides relative throughput support. Without fairness, traffic classes may suffer unfair bandwidth allocation at minimum to more serious problems like starvation.
- The mechanism should provide QoS support with *no changes or minimum changes* to the IEEE 802.11 MAC mechanism, i.e., the Distributed Coordination Function (DCF) mode of IEEE 802.11. More specifically, rather than coming up with a completely brand new MAC mechanism, a new way in which parameters of DCF are calculated and selected is a better alternative. These parameters are discussed in chapter 2. With this approach, no changes or minimum changes to the underlying mechanism should be needed. The benefit of this approach is that the proposed mechanism is likely to be compatible with the current mechanism and easier to implement and adopt.
- The mechanism should provide QoS support with *high scalability*. By scalability, we mean that the QoS support will not degrade even for a high number of connections with different data rates from a large number of Mobile Stations (MSs) active at the same time. As the data rate of WLANs continue to increase (e.g., 156 Mbps WLAN has already been proposed in [30] and 100-500 Mbps WLAN is being standardized under

IEEE 802.11n [31]), we expect a high number of connections and MSs to exist in a WLAN. In particular, revenue-critical low data rate connections such as voice connections from cellular networks or Voice over WLAN (VoWLAN) connections within the WLANs are likely to be numerous. Therefore, the ability of the proposed mechanism to maintain a good performance as the number of participant MSs or connections increases is a vital feature to the success and continuing adoption of WLANs.

- The mechanism should provide QoS support with *very low variation of throughput*. The performance of most multimedia applications depends largely on the stability and low variation of the available bandwidth. Without low throughput variation, the performance of multimedia and real-time applications will be compromised.
- The mechanism should provide QoS support with *minimum requirements of admission control*. To provide absolute throughput support, a simple admission control and resource reservation mechanism to partition and allocate the available bandwidth is needed. However, admission control adds overhead and complexity and it should only be needed when it is absolutely necessary, e.g. for partitioning bandwidth between absolute and relative throughput classes.
- The mechanism should provide QoS support with *ease of user requirement translation*. In most QoS support mechanisms in the current literature, the relationship between the user requirements, such as throughput or delay, and QoS parameters has not been addressed. This relationship can be complicated making it difficult to translate from the user's QoS requirements to the network QoS parameters. Without a proper translation between the user's QoS requirement and the network QoS parameters, the QoS mechanism in the network or Medium Access Control (MAC) layers may not understand the demand of the user correctly; therefore, the demand from the user may not be fulfilled adequately. In our mechanism, we will consider the user requirement translation as an important issue. The mechanism should provide QoS support with a direct and easy translation from the user's requirement.
- The mechanism should provide QoS support in a *highly distributed manner*. By distributed manner, we mean the mechanism should not require a central control. Rather, the mechanism should reside in the MS and operate independently from the mechanisms

in other MSs. As the history of Ethernet and 802.11 suggest, a highly distributed nature is one of the most attractive features. Many researchers believe that being distributed in nature was the key to the success and wide adoption of Ethernet and current WLAN technologies.

- The mechanism should provide QoS support with the *ability to adapt to unexpected network loads.* Two scenarios of network load are under-loaded and overloaded. In the under-loaded scenario, the total throughput requirements are less than the effective channel capacity. In this case, the requested throughput of a traffic class that demands only relative throughput[1] should be fully satisfied. However, in the overloaded scenario, the total requirement from all traffic streams is higher than the effective channel capacity. In this scenario, the mechanism should adapt so that the throughput provided to traffic classes demanding relative throughput remains fair while the throughput of traffic classes demanding absolute throughput should be fully satisfied in spite of the network load.
- The mechanism should provide QoS support with *minimum computational complexity.* In the past, many QoS mechanisms were proposed and promised a hard guaranteed service; however these mechanisms usually suffered from high computational complexity and became impractical to implement. History suggests that the lack of simplicity inhibits the wide adoption of these mechanisms. Therefore, it is important to keep the computational complexity of the proposed mechanism to be as low as possible.
- The mechanism should provide QoS support with *robustness.* In the future, heterogeneous applications ranging from those with low bandwidth requirements and high tolerance to loss and delay, e.g. email, to those with high bandwidth requirement and sensitivity to delay, e.g. VoIP and real-time video, are to be proliferated in WLANs. The performance of the mechanism under various scenarios such as the presence of hidden nodes and usage of request-to-send/clear-to-send (RTS/CTS - discussed in Chapter 2) and heterogeneous traffic types, should be maintained equally well.
- The mechanism should be an *integrated solution.* By integrated solution, we mean the

[1] Throughput support can be classified into two flavors: relative throughput and absolute throughput. By relative throughput support, we mean the ability to allocate different bandwidths to different traffic classes in proportion to their requirements. By absolute throughput support, we mean the ability to deliver a specific throughput. Media distribution and voice telephony are the examples for applications that requires absolute throughput support.

mechanism should support several QoS objectives (throughput, variation of throughput, delay constraint), throughput requirements (absolute or relative), and priorities (low, medium, or high) at the same time in the same mechanism.

Research on distributed QoS support mechanisms for 802.11 networks is a relatively new area of study, motivated by emerging applications requiring QoS support. The current mechanisms use well known QoS schemes (based on priority and fair queuing) from wired networks and map QoS requirements onto 802.11 MAC parameters. However, it remains unclear what are the right set of MAC parameters for priority-based mechanisms like Enhanced Distributed Channel Access (EDCA) to achieve a satisfactory QoS support. Most schemes in the research literature have focused on only throughput delivery; however other QoS metrics, such as delay, need more attention. The majority of current mechanisms provide only QoS differentiation, but not specific QoS support. The future applications are expected to be multimedia-based and used commonly in WLANs. These applications are time-sensitive and will require stringent QoS support. The next generation QoS mechanism in WLANs, therefore, should be designed to provide QoS to support both throughput and delay assurance.

Without QoS support, applications with drastically different requirements may receive the same yet potentially unsatisfied service. Without absolute throughput support, the performance of applications with stringent throughput requirement such as VoIP will not be met and eventually will be jeopardized. Without relative throughput support, heterogeneous types of applications will be treated unfairly and their performance may be poor. Without delay constraint support, time-sensitive applications will not even be possible. We argue that these QoS features should be all supported. The criteria to evaluate how well a mechanism can provide these services was given in the desired properties discussed above. These properties include fairness, scalability, robustness, integration, and other important requirements such as changes needs from based mechanism, user's translation, and distributed control. To the best of our knowledge, there is no distributed mechanism that addresses and achieves these goals.

1.3 RESEARCH CONTRIBUTIONS

Three objectives of this study are to 1) review and identify the deficiencies of the existing distributed QoS mechanisms in WLANs, 2) implement and compare the performance of selected priority-based and fair-schedule-based mechanisms, 3) develop and evaluate a novel QoS mechanism in WLANs. According to the review of literature, we found that QoS support in WLANs can be classified into: a) priority-based and b) fair-scheduling-based approaches. Most research efforts were invested in priority-based approaches and focused only on throughput delivery. Other issues, such as delay support, throughput variation, and QoS translation, remain largely unaddressed. Performance comparison of four QoS mechanisms in WLANs showed that it remains unclear what the right set of MAC parameters to achieve a robust QoS support in EDCA are. While, the existing fair-scheduled-based mechanisms provide only fair throughput support, they also suffer from complicated QoS translation, high variation of throughput and delay, and performance degradation in certain scenarios, e.g., without employing RTS/CTS.

The main contribution of this dissertation is to a novel distributed QoS mechanism in WLANs, called Distributed Relative/Absolute Fair Throughput with Delay Support (DRAFT+D). DRAFT+D provides integrated QoS support and exhibits several desired properties. The main features of DRAFT+D include:

- **Integrated QoS Support:** DRAFT+D supports two QoS metrics (throughput and delay) with two QoS models (absolute and relative) under two fairness constraints (utilitarian and temporal fairness) with safeguard against excessive traffic in the same mechanism at the same time.
- **Fully Distributed in Nature:** All QoS support including relative throughput, absolute throughput, absolute delay, and safeguard support are achieved in a fully distributed manner without requiring the exchange or collection of any information from other MSs.
- **Several Desired Properties:** With modification of only one WLAN parameter, DRAFT+D achieved QoS support with high degree of fairness, high scalability, low variation of throughput and delay, ease of QoS translation, and high robustness in various and diverse scenarios.

1.4 ORGANIZATION

The remainder of this dissertation is organized as follows. Chapter 2 presents a review of the available literature in the field of QoS in WLANs. A hierarchical taxonomy of distributed MAC mechanisms in IEEE 802.11 WLANs is presented. The review is organized according to the classification presented in [32]. The goal is to provide the reader with an overview of the state-of-the-art in the field, and to identify the major shortcomings in the existing work. The research problems raised in this chapter will be used to propose a novel QoS support solution in Chapter 3. We will present baseline performance and sensitivity analysis in Chapter 4 and comprehensive simulation results in Chapter 5. We will show that DRAFT+D can provide three types of QoS support: a) relative throughput support, b) absolute throughput support, and c) absolute delay support, in a robust manner under several conditions. Finally, we will summarize the contribution and discuss future work in Chapter 6.

2.0 BACKGROUND AND LITERATURE REVIEW

In this chapter, state-of-the-art approaches to provide Quality of Service (QoS) support in Wireless Local Area Networks (WLANs) are presented. Although centralized protocols, such as Point Coordination function (PCF) and HIgh PErformance Radio Local Area Network (HIPERLAN)/2, promised precise QoS guarantee, their adoption has been limited. On the contrary, distributed protocols such as the Distributed Coordination Function (DCF) of IEEE 802.11 that is not equipped with any QoS support have been widely and continuously adopted. The reasons are that these distributed protocols are easier to implement, require smaller overheads, incur less complexity, and more robust than the centralized ones. From the evolution in data networks, it is clear that distributed protocols are more favored. We believe that they will continue to dominate data networks compared to centralized protocols. For the purpose of this study, we will only address distributed QoS mechanisms based on IEEE 802.11 at the Medium Access Control (MAC) level in WLANs. We will illustrate the overview, taxonomy, essential concepts, and difficulties in providing QoS support in WLANs.

Despite research attention, the literature survey suggests that most research efforts were invested in priority-based approaches, largely due to the standardization of Enhanced Distributed Channel Access (EDCA) of IEEE 802.11. However, the current mechanisms often provide only QoS differentiation or only throughput support, while other important issues, such as throughput variation, delay, and QoS translation, remain largely unaddressed. To date, no mechanism has an integrated solution to the problem of multidimensional QoS support in WLANs. We believe that what we need is an integrated mechanism that can provide both fair and specific QoS support, for both throughput and delay requirements, under different fairness criteria with small overhead, high robustness, scalability, and is distributed in nature.

2.1 NOTION OF QUALITY OF SERVICE

Quality of Service (QoS) is the ability to provide a level of assurance for data delivery over the network. For example, traffic of different classes or traffic with different requirements receive different levels of QoS assurance. Therefore, we will use the term QoS support mechanism to refer to any mechanism that is equipped by any kind of QoS support. The term QoS guarantee will be refereed to a mechanism that can provide guaranteed support. A QoS system has several components including the QoS mechanism, QoS mapping, admission control, and resource allocation. QoS mapping refers to the translation of the QoS representation from one layer to the next. Admission Control is used to determine whether a network is able to support the requested traffic with the requested network level QoS parameter. Resource allocation involves the allocation of suitable network resources according to the requested QoS.

Concurrent with the gradual migration of traditional applications from wired-networks to WLANs is an increasing use of multimedia-based applications in WLANs. Multimedia-based applications are time-sensitive and require specific QoS support with small variation. Based on anticipated future applications in WLANs, we can categorize the objectives and the approaches to accomplish QoS in this network as follows. The objectives of QoS provision can be categorized into: a) prioritized QoS support and b) parameterized QoS support [33]. Prioritized QoS support aims at providing "different" level of QoS support for different classes of traffic, e.g., high priority traffic receives better throughput and delay than low priority class traffic. Prioritized QoS support is also known as differentiated QoS support. Parameterized QoS support aims at providing a "specific" level of QoS support, e.g., at least 64 Kbps and delay less than 30 ms, on average. Parameterized QoS support is also known as specific QoS support. Under prioritized QoS support, scheduling mechanisms classify packets into different priority classes. Under parameterized QoS support, scheduling mechanisms consider the requirement of a particular packet and provide the appropriate treatment .

Approaches to accomplish QoS support can be categorized into: a) priority-based scheduling and b) fair scheduling. Priority-based scheduling provides "better" performance for high priority traffic while fair-scheduling provides "proportionally fair" performance based on a weight. We will discuss these two approaches in more detail in Section 2.5 - 2.6. It is important to note that, without admission control, only prioritized QoS can be supported, either "differently" via prioritized-based scheduling or "fairly" via fair scheduling. Admission control is inevitably required for any parameterized QoS support.

2.1.1 QoS Metrics

The important QoS metrics for multimedia applications are delay, jitter, and throughput. End-to-end delay is the time between the arrival of a packet and its successful delivery to the receiver. Jitter is the variation of delay and is an important metric for multimedia applications. Jitter is measured by the difference between the previous delay and the current delay. Throughput is the amount of data successfully transmitted and received in unit time. In this study, throughput support is further classified into two flavors, i.e. relative throughput and absolute throughput support. Relative throughput support is the term used to describe the ability to allocate different bandwidths to different traffic classes in proportion to their requirements. Web access and files sharing are examples of applications that require relative throughput support. All best-effort applications can take advantage of relative throughput support. Absolute throughput support is the term used to describe the ability to deliver a specific throughput. Media distribution and voice telephony are examples of applications that require absolute throughput support. In WLANs, a higher bandwidth may be needed to achieve a particular throughput due to the error in the wireless channel.

Because the wireless channel is time-varying, two types of scheduling can be consider in wireless network: a) non-opportunistic scheduling and b) opportunistic scheduling. In non-opportunistic scheduling, such decisions do not consider the time-varying characteristics of the wireless channel. The network resources are considered constant and do not change with time. By opportunistic, we mean the ability of scheduler to allow packet transmission based on favorable channel conditions [34] . In opportunistic scheduling, such decisions depend on the conditions of the wireless channel since the characteristics of channel wireless change with time and each user may perceive the quality of the channel differently. For opportunistic scheduling, two types of fairness constraints were suggested [34]: a) temporal fairness and b) utilitarian fairness. In the temporal fairness scheme, time is used as the criterion to maintain fairness among Mobile Stations (MSs). In utilitarian fairness scheme, the achieved performance values, e.g. throughput, are used as the criterion to maintain the fairness among MSs. Therefore, the advantage of utilitarian fairness is that it ensures that a certain level of performance. However, a user with an extremely poor channel could have a detrimental impact on the overall system performance. A review of opportunistic scheduling in wireless networks and multi-rate ad hoc networks can be found in [34] and [35], respectively. In this study, we will focus on a non-opportunistic scheduler where both temporal and utilitarian fairness will be supported.

2.2 DISTRIBUTED MAC PROTOCOLS

Centralized protocols, such as reservation Time Division Multiple Access (TDMA) or polling and scheduling schemes have received much attention from the research community, since they promise precise QoS guarantees. Examples of centralized protocols are the PCF of IEEE 802.11 [36], HIPERLAN/2 of European Telecommunications Standards Institute (ETSI) [37], numerous wireless Asynchronous Transfer Mode (ATM) [38] and Hybrid Coordination Function (HCF) which is an extension of IEEE 802.11 [39]. Comprehensive reviews of centralized protocols can be found in [40, 38, 41]. With centralized protocols, each MS requests the right to access the channel from a single point of coordination. The coordination point, usually called Base Station (BS) or Access Point (AP), performs admission control, bandwidth and parameters assignment, and channel access control. The major advantage of these centralized protocols is that they can guarantee QoS assurance once admitted to the network. However, history shows that the adoption of these mechanisms has been limited due to several disadvantages such as high overhead [42], high cost/complexity [43] and issues in scalability, practicality [43] and flexibility [44].

In contrast to centralized protocols, distributed protocols, which are the focus of this study, do not require a central control, are simple to implement, require smaller overhead, and incur less complexity. Examples of distributed protocols are the DCF of IEEE 802.11 and the elimination-yield mechanism in HIPERLAN/1. All distributed protocols are based on the principles of Carrier Sense Multiple Access (CSMA). Carrier sensing refers to MSs listening to the physical channel to detect any ongoing transmissions and backing off in case it detects any transmission. Although, these protocols are currently not equipped with QoS support, i.e., no priority mechanisms [44] and no delay-bounds or throughput support [45], they are widely adopted. To address this limitation, there are several schemes proposed to incorporate QoS mechanisms in distributed protocols. Next, we will present a detail description of the DCF mechanism and demonstrate the lack of QoS support. Then, we will present existing QoS support mechanisms that have been proposed for 802.11 WLANs over DCF. Their advantages, disadvantages, and limitations will be discussed.

2.3 DISTRIBUTED COORDINATION FUNCTION (DCF) OF IEEE 802.11

The DCF mode of IEEE 802.11 was designed for data applications [46] and is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) [36]. The channel contention procedure begins when an MS senses the channel to determine whether or not another MS is transmitting. The collision avoidance mechanism employs two techniques: InterFrame Space (IFS) and backoff mechanism. The IFS is the period of time an MS is required to wait after it senses an idle channel and enters the transmission process. If the channel is idle for a time equal to DCF IFS (DIFS), the MS can transmit a packet (Fig. 2.1).

If the channel is busy, the MS waits until the medium becomes idle, waits for an additional DIFS and enters a deferral period (backoff period) to reduce the chance of collisions with other contending MSs. In the deferral period, the MS selects a Backoff Interval (BI) that is uniformly distributed between zero and a Contention Window (CW). CW is initially set to CW_{min} and is doubled every consecutive collision until CW reaches CW_{max}, as shown in Eq. (2.1). The deferral period is divided into time slots, each a duration of $aSlotTime$ seconds. The length of the backoff interval is calculated as a multiple of $aSlotTime$ as shown in Eq. (2.2).

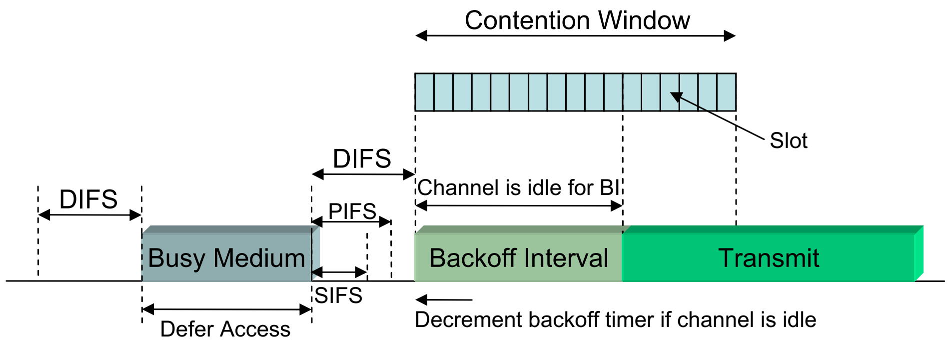


Figure 2.1: Transmission Procedure of DCF Mechanism

$$CW = min[((CW_{min} + 1) \times 2^c) - 1, CW_{max}] \tag{2.1}$$

$$BI = Uniform\_Random(0, CW) * aSlotTime \tag{2.2}$$

While the medium is idle, a backoff timer counts down from the randomly selected BI value every $SlotTime$, as shown in Fig. 2.2, adapted from [47]. When the timer reaches zero and the medium is still idle for additional $DIFS$, the MS can transmit. The backoff timer is frozen when a transmission is detected and continues to elapse when the channel becomes idle again. After receiving a frame, the receiving MS waits for a Short Inter-frame Space (SIFS) and responds with an Acknowledgment (ACK) to confirm a successful transmission. The SIFS parameter is smaller than DIFS to allow acknowledgments to have the highest priority in accessing the channel. A collision occurs when the backoff timer of two or more MSs reach zero at the same time. To reduce the probability of collisions, the CW is doubled every time a collision repeatedly occurs until the maximum value of the CW (CW_{max}) is reached, i.e. 1023 timeslots. This procedure is called *exponential backoff*.

The hidden node problem occurs when a MS can hear only some but not all MS's transmissions. To overcome this problem, an optional mechanism employing Request To Sent (RTS) and Clear To Send (CTS) messages is used. In the Request to Send/Clear to Send (RTS/CTS) mode, the MS first transmits an RTS message and waits for a CTS

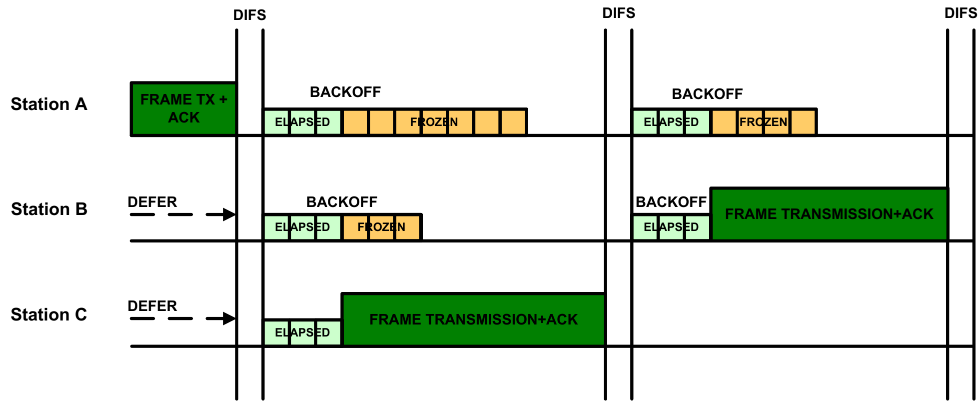
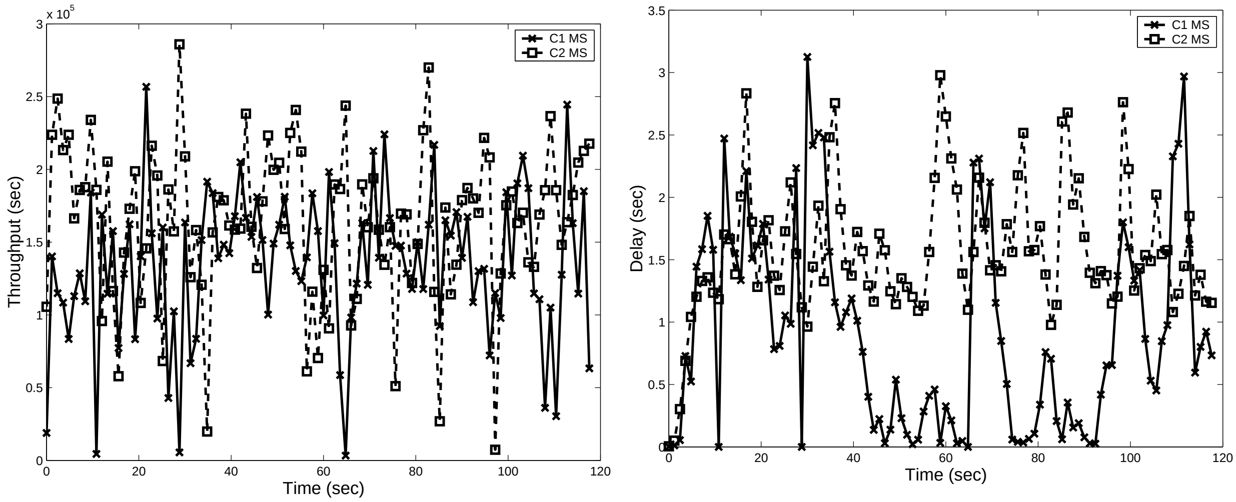


Figure 2.2: Backoff Procedure of DCF Mechanism

message from the recipient before beginning data transmission. With this method, all MSs in the range of sender and receiver will be aware of the frame transmission. While these two extra frames present additional overheads, Anastasi and Lenzini [48] reported that this method significantly alleviates the hidden node problem and reduces average access delay, particularly for large data frames [47]. As reported in [49] by Bianchi, the performance of DCF depends primarily on CW_{min} and the number of active MSs, and is only marginally dependent on these parameters when RTS/CTS is used. The IEEE standard also defines a PCF IFS (PIFS) that is between the SIFS and DIFS values and is used with a centralized polling mechanism. In the infrastructure mode, an AP transmits beacon frames periodically to deliver management information that is necessary for the association process (where a MS associates with an AP) [47].

Because DCF was originally designed for data applications [46], its main weaknesses is the lack of QoS support (absolute throughput, relative throughput or delay support). The lack of any QoS support also means that DCF provides no fairness among different traffic classes or among different transmitting frame sizes. Additionally, high variation of throughput and delay is inherited from the exponential backoff mechanism, particularly in overloaded situations. To demonstrate the limitations of DCF, a simple scenario with ten



2.3 (a): Throughput

2.3 (b): MAC Delay

Figure 2.3: Performance of DCF MSs with requirement of 150 Kbps and 300 Kbps

DCF MSs in a 2 Mbps[1] WLAN was simulated using OPNET 9.1. The MSs operate in an ad-hoc mode (that is sufficient to evaluate the MAC characteristics). Eight MSs transmit at a low data rate (150 Kbps) and two MSs at a high data rate (300 Kbps). The total aggregated offered load is 1.8 Mbps. With a raw channel rate of 2 Mbps, the MAC overhead makes the offered load higher than the effective aggregated throughput (overloaded situation).

Figure 2.3 (a) shows the instantaneous throughput of two representative DCF MSs. Here, C1 (Class 1) indicates a MS with low data rate and C2 (Class 2), a MS with high data rate. The average throughput of C1 is around 138 Kbps and C2 around 160 Kbps, even though, C1 requires only *half* the throughput of C2. This demonstrates unfair bandwidth allocation in DCF. The instantaneous throughput and access delay widely fluctuate as shown in Fig. 2.3 (a) and Fig. 2.3 (b). The coefficients of variation[2] (COV) of throughput and access delay are both very high ranging from 0.3 to 1.6. The problems of lack of fairness, and high variation of throughput and access delay degrade the performance of QoS-sensitive applications. Clearly, this evidence suggests deficiency of DCF for QoS provision in WLANs.

[1]Data rates of 5.5 Mbps and 11 Mbps have also been considered in this work.
[2]This is the ratio of the standard deviation of the quantity to its mean

In summary, the DCF mode of IEEE 802.11 has advantages of simplicity, ease of implementation, and suitability for data applications. However, DCF does not support QoS requirements or provide specific delay/throughput support. Moreover, several studies report high variation of throughput and delay [32, 50], relatively poor performance for voice transmission, and the inability to provide low delay variation [51] in overload conditions. Next, we will describe several mechanism for QoS support in 802.11 proposed in the literature and standards process.

2.4 TAXONOMY OF DISTRIBUTED QOS MECHANISMS IN WLANS

A hierarchical taxonomy of distributed MAC mechanisms based on IEEE 802.11 WLANs is depicted in Figure 2.4. At the highest level, MAC schemes can be categorized into distributed and centralized control protocols. In the class of distributed MAC protocols, we consider the DCF mode of 802.11. The 802.11 MAC protocol parameters, such as IFS, BI, CW_{min}, CW_{max}, and Persistence Factor (PF) have been suggested for QoS support as described below. The approaches to provide QoS support can be classified into priority based and fair scheduling based approaches. For each approach, essential concepts, advantages, disadvantages, and limitations will be discussed.

2.5 PRIORITY-BASED QOS SUPPORT

The objective of the majority of QoS support mechanisms proposed in the literature is to provide service *differentiation* by allowing faster access to the channel to traffic classes with higher priority. Faster access can be provided by allocating a smaller waiting time (IFS) [52, 53, 54, 45, 55] or a smaller contention window (CW) [44, 56, 57, 58, 59, 60] that results in a smaller backoff interval (BI) on average. The values of IFS and CW are fixed once they are assigned; therefore, we call the QoS mechanisms employing these assignments static priority-based mechanisms. The common drawback of priority-based mechanisms is that the

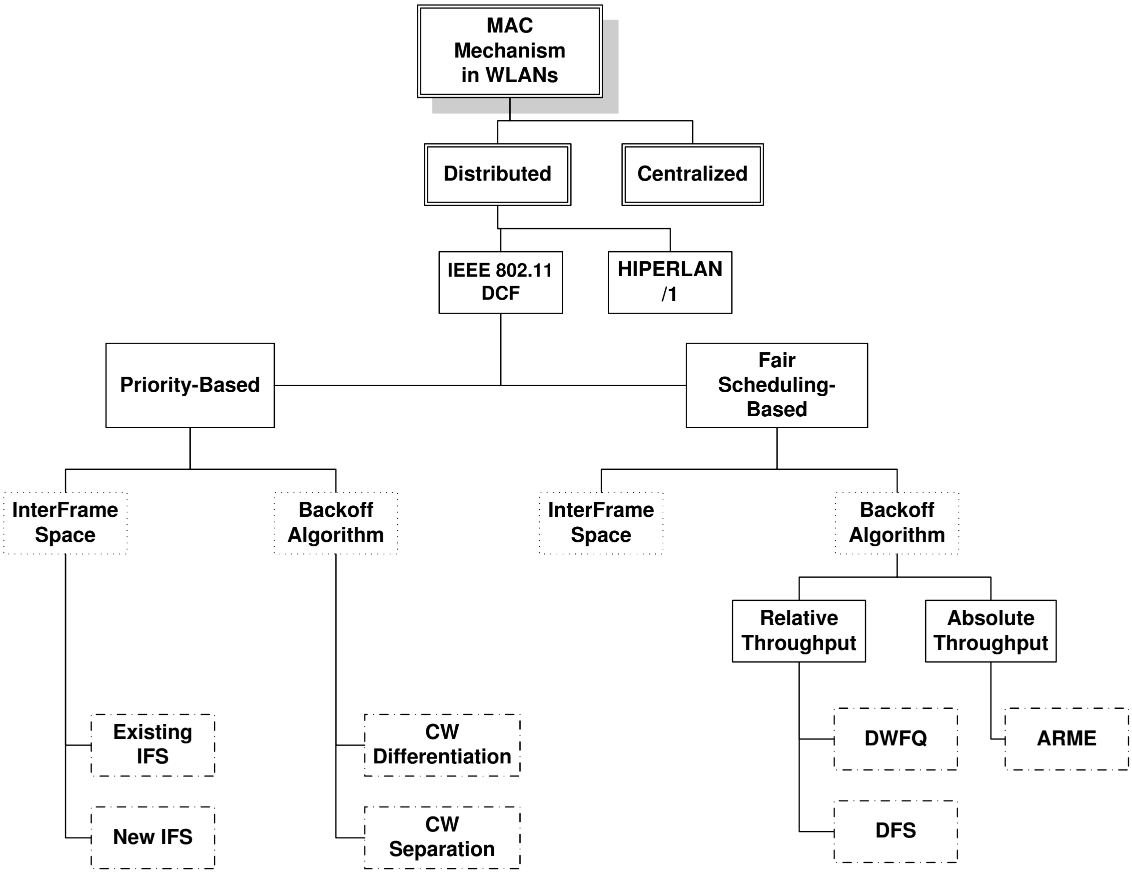


Figure 2.4: Taxonomy of Distributed QoS Mechanisms in WLAN

higher throughput or smaller average delay of high priority traffic is achieved at the expense of low priority, i.e., reduced throughput or increased average delay [61]. In short, binding the priority to channel access makes these QoS-support mechanisms unfair. As the number of high priority traffic MSs increases, they tend to monopolize the channel preventing access for low priority traffic. The selection of a random backoff interval and exponential expansion of the contention window also introduce large variations in delay and throughput. Moreover, the standard value of CW_{min} is usually too small for common networks as reported in [49].

2.5.1 Using the Inter-frame Space (IFS)

The idea behind QoS-support mechanisms that exploit different waiting times after the busy period is to assign a smaller IFS value to higher priority traffic. A higher priority frame needs to wait for a shorter duration so it can seize the channel sooner than a low priority frame. The low priority frame finds the channel busy and has to wait until the high priority traffic has completed transmission. Moreover, it has to enter a backoff period after the channel becomes free. During such a backoff period, a smaller IFS also decreases the waiting time for high priority traffic.

2.5.1.1 Using existing IFS values for priority Many researchers have proposed using IFS values that are *already available* from the 802.11 standard to differentiate between low priority and high priority traffic. In the 802.11 standard, three different IFS values specified are the SIFS, PIFS and DIFS. Sheu and Sheu [45], Banchs et al. [55], and Deng et al. [62] suggest using the PIFS and DIFS values to differentiate between high priority (real-time) and low priority (non real-time) traffic. However, like DCF, this mechanism shows increases in average access delay and packet losses under high load conditions. Further, using only DIFS and PIFS, however, allows for only two priority classes. To differentiate between more than two priority classes, other alternatives have to be explored. Deng et al. [62] has employed two different backoff algorithms after the IFS waiting time to support more traffic classes. Another alternative is to use new IFS values as discussed next.

2.5.1.2 Using new IFS values for priority In contrast to using only the DIFS and PIFS values, Benveniste [63] and Aad et al. [53] proposed using multiple IFSs to differentiate among priority classes where each priority class is given a different IFS. The values of IFS are assigned such that the IFS_i of low priority frames are longer than the IFS_j of high priority frame ($IFS_i > IFS_j$) where priority of class $i <$ priority of class j. Therefore, the higher priority frame will get access to the channel sooner than the lower priority frames. Aad et al. also add a small random time at the end of IFS to avoid the collision among frames in the same priority class. The idea of new IFS values has led to the on-going standard to provide QoS called Enhanced Distributed Channel Access (EDCA), to be described in Section 2.5.3.

2.5.1.3 Discussion and Summary Although priority-based mechanisms can help differentiate the throughput for traffic with different classes of priority, fairness between different traffic classes is neglected. Additionally, specific throughput and delay are not assured. In some cases, where there is only high priority class (e.g. voice), differentiation may not even be witnessed. Without the help of an admission control, a specific level of throughput and/or delay, which is an ultimate goal of providing QoS in WLAN, cannot be achieved [32, 64].

2.5.2 Using Backoff Algorithms

As described previously, the BI is an integer value which corresponds to the number of timeslots that a MS needs to wait *after the IFS* before it can transmit data. This section presents the mechanisms that modify the backoff algorithm. Based on whether or not the ranges of CW_{min} and CW_{max} among priority classes overlap, two main approaches by assigning different contention windows are a) Contention Window Differentiation (CWD) and b) Contention Window Separation (CWS).

2.5.2.1 Contention Window Differentiation (CWD) In the case of CWD, given two classes of traffic A and B, there are two ranges of the CW namely, CW_A (between 0 and $CW_{min,A}$) and CW_B (between 0 and $CW_{min,B}$). Ayyagari et al. [56], Benvenisite et al. [54] proposed mechanisms to modify the minimum and maximum value of the CWs. The values of CW are assigned such that the CW_{min} and CW_{max} values of low priority frames are higher than that of high priority frame ($CW_{min,i} > CW_{min,j}$ and $CW_{max,i} > CW_{max,j}$ where priority of class i < priority of class j). Since BI is a random number that is uniformly distributed between 0 and CW_{min}, the two traffic classes are differentiated by the average BI values. The drawback of CWD is that the two contention windows overlap; thus, a low priority traffic can sometimes access the network sooner than high priority traffic. In an overloaded condition, it is critical for the high priority traffic to get faster access to the network than the low priority traffic. The lower priority frame selects a longer BI on average whereas the higher priority frame selects a smaller BI on average. Therefore, the higher priority frame is likely to get access to the channel earlier than the lower priority frames. The quality of differentiation depends on the amount of overlap between the contention windows of different traffic classes.

To dynamically adjust the range of CW as the number of active MSs changes, Chen et al. [65] proposed a scheme called Priority-Based Contention Control (PCC) where a priority reference value (called priority limit) is piggy-backed with transmitted frames to help each MS calculate its CW according to Eq. (2.3). In this equation, the parameter S is a scaling factor, PT is the priority of the traffic flow being served in each MS and PL is the current priority reference value. Eventually, a MS with a high priority flow will receive a smaller CW. To increase the probability of getting access to the medium, for real-time traffic the value of PT is increased when the first collision occurs. This method helps reduce potential variation of throughput and delay.

$$CW = S \times \frac{1}{PT - PL + 1} \tag{2.3}$$

In the work by Barry et al. [58], the CW for a high priority traffic class is between a CW$_{min}$ of [8,32] and a CW$_{max}$ of 64. The CW for low priority traffic is between a CW$_{min}$ of [32,128] and a CW$_{max}$ of 1024. Because the overlap of the CW$_{min}$ and CW$_{max}$ of the lower and higher priority traffic is small, the delay between low priority traffic and high priority traffic is clearly differentiated as shown in their simulation results. In the extreme case, there is no overlap between the CW$_{min}$ and CW$_{max}$ of different traffic classes, which to be discussed next.

2.5.2.2 Contention Window Separation (CWS) As in the case of CWD, higher priority traffic in CWS receives a CW that results in a smaller BI whereas lower priority traffic receives a CW that results in a longer BI. The CW$_{min}$ and CW$_{max}$ are completely separated and, thus, the traffic from higher priority classes is more likely to be transmitted before traffic from lower priority classes, if they arrive at the same time. However, CWS does not guarantee that the high priority traffic will always transmit sooner. The reason is that the CW starts from 0 regardless of the value of CW$_{min}$ and CW$_{max}$. Therefore, although the ranges of CW$_{min}$ and CW$_{max}$ are completed separated, the possible value of BI can still overlap. The implication of this will be discussed in more details in Section 2.5.4.

An example of CWS is the algorithm proposed by Deng et al. [52] described by Eq. 2.4 and 2.5 below, where $\mu(a, b)$ is a random number generated from a uniform distribution between a and b. i is the number of consecutive collisions.

$$\begin{aligned} CW_{high} &= \left\lfloor \mu(0,1) * 2^{i+1} \right\rfloor & (2.4) \\ CW_{low} &= 2^{i+1} + \left\lfloor \mu(0,1) * 2^{i+1} \right\rfloor & (2.5) \end{aligned}$$

Although this mechanism completely separates the CW initially, the separation may not be valid in time because of the following reason. This mechanism is employed in each MS independently. As the number of consecutive collisions in each MS can be different, the CW of high priority traffic CW$_{high}$ and the CW of low priority traffic CW$_{low}$ can overlap and create an inconsistency among frames in the same priority class among MSs. Furthermore, the scheme proposed by Deng et al. supports only up to two classes of priority. Therefore, Deng et al. combine their CWS proposal with the IFS mechanism described earlier to support more priority classes. Simulation results in [52] however do show improvements in reducing delay and packet loss and increasing throughput for higher priority traffic.

A similar scheme called *distributed priority scheduling* is proposed by Kanodia et al. [60] where the priority of every MSs' head-of-line packet is piggybacked onto RTS, CTS, data and ACK frames. Using this information, each MS can create a table of frames that are expected to be transmitted along with their priorities in a ranked list. Frames with the highest rank choose a smaller CW interval while those with lower rank have an additional waiting time and select the BI from a larger CW. The CW can be calculated as shown in Eq. (2.6) where $\mu(a, b)$ is a random number generated from a uniform distribution between a and b, r_j is the rank of node j's packet, l represents the number of retransmission attempts, m is the maximum attempt, α is the separation of BI given to the highest priority traffic and γ is the separation of BI between the first and second attempt. The analysis and simulations in [60] show that this mechanism reduced mean end-to-end delays of high priority packets.

$$BI = \begin{cases} \mu(0, 2^l CW_{min} - 1), & r_j = 1, l < m \\ \alpha CW_{min} + \mu(0, \gamma CW_{min} - 1), & r_j > 1, l < 0 \\ \mu(0, 2^l \gamma CW_{min} - 1), & r_j = 1, l \geq 1 \end{cases} \tag{2.6}$$

2.5.2.3 Discussion and Summary

Because the CW in DCF always starts from zero, no matter the values of CW_{min} and CW_{max} of each priority are, the range of CW among classes will always overlap. The effort to separate the range between CW_{min} and CW_{max} via CWS only provides clearer QoS differentiation than that of CWD on the average case. However, it does not assure that low priority traffic will always wait longer than higher priority traffic. Because of this, the low priority traffic can gain a faster access than the high priority traffic even though both classes of traffic arrive to the head of queue at the same time. This behavior is not desirable for a QoS mechanism.

Moreover, after each consecutive collision, the value of CW is doubled, as in Eq. (2.7) below. In this equation, i is the number of consecutive times a MS attempts to send a frame and the initial $\mathrm{CW}_{min}{}^{3+0}$is $2^3 = 8$.

$$CW = 2^{3+i} - 1 \tag{2.7}$$

As a result of this binary exponential backoff, the probability of waiting time or backoff time increases in direct proportion to the amount of time a MS has been waiting. This behavior is undesirable for time-sensitive traffic. Also, the BI selection of the system (for all MSs combined) is not uniformly distributed, but rather exponentially distributed where smaller BI values are more likely than longer ones. A long CW occurs only if multiple consecutive collisions occur. For example, BIs ranging from 0 to 7 timeslots appear as choices for selection every time a new packet needs transmission. In contrast, a backoff interval of 1023 timeslots appears as a choice only when the packet transmission has failed on 8 consecutive trials. So an MS that has recently entered into contention could potentially transmit earlier than a MS that has faced several collisions. Moreover, the initial value of CW, which is CW_{min} = 15), is too small to avoid unnecessary collisions, especially in an overloaded network [49]. The work of Romdhani et al. [66] was proposed to partially address this problem by decreasing the value of CW slowly rather than resetting back to CW_{min} immediately after a successful transmission.

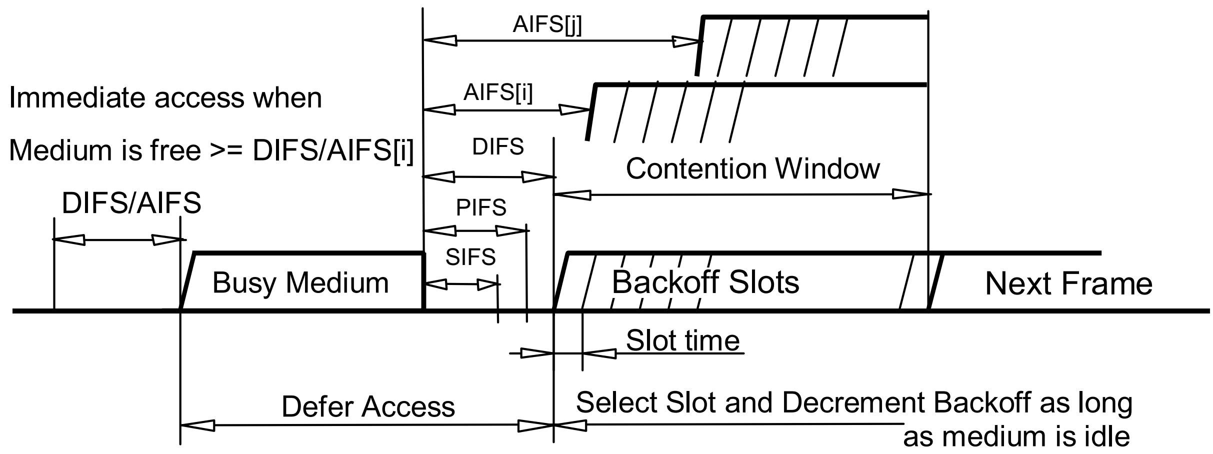


Figure 2.5: Transmission Procedure of EDCA Mechanism

2.5.3 Enhanced Distributed Coordination Function (EDCF) of 802.11e

The Enhanced DCF (EDCF) is currently being standardized by the IEEE 802.11 working group E [67, 33]. Recently, the nomenclature was changed from EDCF to Enhanced Distributed Channel Access (EDCA); however most of the key features remain the same. In this study, we will refer to both EDCA and EDCF interchangeably. EDCA is also a static priority-based mechanism. EDCA consists of 8 priority class but only 4 priority queues. Priorities 0-2 are mapped into Access Category (AC) 0, or the lowest priority queue, and designated for best effort traffic. Priority 3 is mapped into AC 2 and designated for video probe traffic. Priorities 4-5 are mapped into AC 2 and designated for video application while priorities 6-7 are mapped into AC 3 and designated for voice applications.

An important new feature of 802.11e is the concept of Transmission Opportunity (TXOP). A TXOP is an interval of time indicating when a station has the right to initiate transmission. A TXOP is composed of a starting time and a maximum allowable duration ($TXOPlimit$). $TXOPlimit$ is distributed via beacon frames. A new type of IFS named Arbitrary IFS (AIFS) has been added. The value of AIFS depends on the priority class of traffic as shown in Figure 2.5 [67]. Each priority class has its own queue and backoff

Table 2.1: Proposed Values of Parameters in EDCA

AC	AIFSN	CW_{min}	CW_{max}	Application
0	2	15	1023	Best Effort
1	1	15	1023	Video Probe
2	1	7	15	Video
3	1	3	7	Voice

counter. $AIFS_i$ can be calculated according to Eq. (2.8) where i represents the index of the traffic class. The BI of each priority class is chosen according to a uniform distribution over $[0, CW_i]$ as shown in Eq. (2.9). The values of Arbitrary IFS Number (AIFSN), CW$_{min}$, and CW$_{max}$ for certain types of application as proposed are shown in Table 2.1 [68].

$$AIFS_i = SIFS + aAIFS_i \times SlotTime \tag{2.8}$$

$$BI_i = \mu(0, CW_i) \times SlotTime \tag{2.9}$$

2.5.4 Drawbacks of EDCA

In this section, we will present the drawbacks and potential inadequacy of QoS support of EDCA. One may find EDCA attractive due to its simplicity, prioritized-QoS support, and decentralized nature. Several studies report good service differentiation (e.g., improved throughput, access delay and dropped rate for high priority traffic [69]).

Nonetheless, recent studies and performance analyses indicate several drawbacks of EDCA. Many improvements have been proposed to remedy these drawbacks and limitations. However, these improvements often alleviate these issues at the expense of complicating the base mechanism. The following list is the summary of the drawbacks compiled from our observations and what is reported in the literature.

Static Priority-based Mechanism: In static priority mechanisms, the values of AIFSN, CW_{min}, and CW_{max} are fixed based on the priority class. Fixing these values with a priority class makes EDCA difficult to adapt in a dynamic environment where the amount of traffic, traffic patterns, and the number of MSs can change with time [70]. As the number of MSs increases, high priority traffic with very small values of CW_{min} and CW_{max} may result in high probability of collision as reported, especially in highly contentious conditions [71, 66]. Moreover, the priority of a traffic class is statically combined with the right to access the medium leading to several problems ranging from unfair occupation of the bandwidth to starvation of low priority traffic.

Exponential Backoff: Although exponential backoff is a beneficial mechanism to avoid further collisions, it may not be suitable for QoS sensitive applications. The colliding packets are supposed to be transmitted successfully sooner than any newly contending packets. However, the colliding packets are penalized by the exponential backoff with a longer waiting time while newly contending packets are given a small waiting time. Exponential backoff can cause high variation of throughput and delay in overloaded situations. Moreover, in the exponential backoff, the probability of waiting time or backoff time increases in direct proportion to the amount of time an MS has been waiting [72]. In exponential backoff, the longer a frame has waited, the longer it is likely that a packet will wait. This property is undesirable for time-sensitive traffic.

Fairness: EDCA is not fair in both short-term and long-term. The short-term unfairness is a result of the overlapping range of CW between high priority and low priority traffic whereas the long-term unfairness stems from the assignment of initial CW. To understand this, let us define CW_{lower} and CW_{upper} as a lower bound and an upper bound from which BI is selected. In other words, CW is simply the difference between CW_{lower} and CW_{upper}. According to this definition, CW_{min} is the initial CW_{upper} while CW_{max} is the maximum value of CW_{upper}. The problem is that, although, the CW_{upper} of high priority traffic is smaller than that of low priority traffic, CW_{lower} is always 0. Because of this, the range of CW of high priority and low priority always overlaps no matter what the values of CW_{min} and CW_{max} are, as shown in Figure 2.5.4. Occasionally, low priority traffic is transmitted sooner than high priority traffic due to a smaller drawn value of BI.

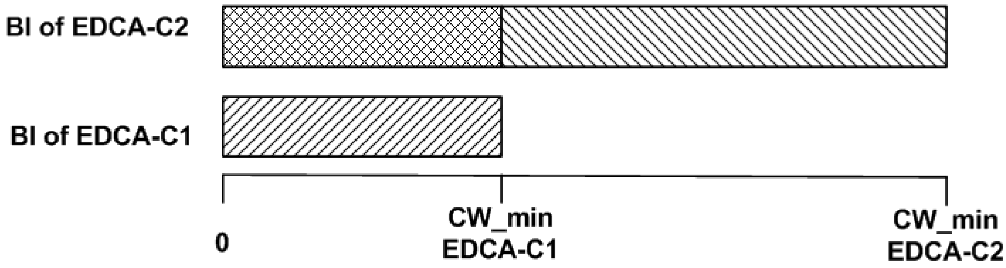


Figure 2.6: Overlapped Ranges of CW among Different Priority Classes in EDCA

The long-term unfairness is due to the range of CW_{min} and CW_{max}. The long-term share of bandwidth of a priority class directly depends on the values of CW_{min} and CW_{max} of the priority class. Because the assignment of CW_{min} and CW_{max} in EDCA is fixed – i.e., CW_{min} and CW_{max} of high priority traffic are smaller than those of low priority traffic, the shares of bandwidth among priority classes are not in proportion. The share of high priority is simply "more" than that of low priority; however, it is unclear how much more the bandwidth of high priority class is in comparison to that of low priority class. Because of the lack of fairness in general, neither utilitarian fairness nor temporal fairness is supported in EDCA. Under utilitarian fairness, the achieved performance values, e.g. throughput, are used as a criterion to maintain the fairness while time is used as a criterion to maintain fairness under temporal fairness. These fairness problems can be resolved by adopting fair-scheduling mechanisms, as discussed later.

CW Differentiation: The differentiation through CW provides relative throughput; however the analysis by Robinson and Randhawa [73] confirms that high priority traffic can suffer performance degradation from heavy load of low priority traffic. At high loads, EDCA cannot ensure specific performance of traffic of the same priority because the probability of collision is a function of the size and composition of the set of contending MSs. More specifically, even though the offered load remains the same, the performance of high priority traffic can be different if the number of contending MSs varies. Moreover, Xiao [74] reported that the delay depends on the overlap portion within and among priority classes. The reason is that the CW of each priority class can overlap and the size of such an overlap can change dynamically.

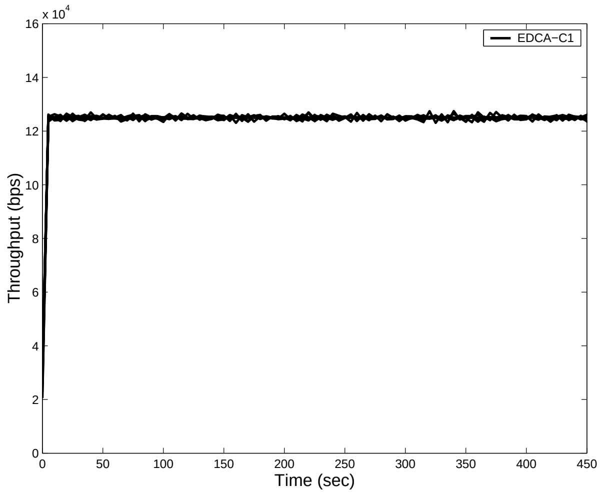
IFS Differentiation: Differentiation through IFS may be effective for high priority traffic; however it leaves lower priority traffic susceptible to starvation [50, 71, 75]. There are also some conflicting findings related to the influence of IFS. While Robinson and Randhawa [73] suggest that AIFS helps reduce the probability of collision for high priority traffic, Xiao [74] suggested that AIFS does not. Compounding the drawbacks from CW differentiation, EDCA is faced with an unpleasant situation for QoS differentiation. The differentiation through IFS leaves lower priority traffic susceptible to starvation whereas the differentiation through CW leaves high priority traffic susceptible to performance degradation under heavy load situation.

QoS Translation: In EDCA, the translation of QoS from the user level, i.e., throughput and delay, to EDCA parameters, i.e., IFS, CW_{min}, or CW_{max}, is not obvious. It is not clear what the values of $AIFS$, CW_{min}, and CW_{max} should be to provide say a throughput of 64 Kbps with a 50 ms target delay. Without proper QoS translation, the user's demand may not be adequately satisfied.

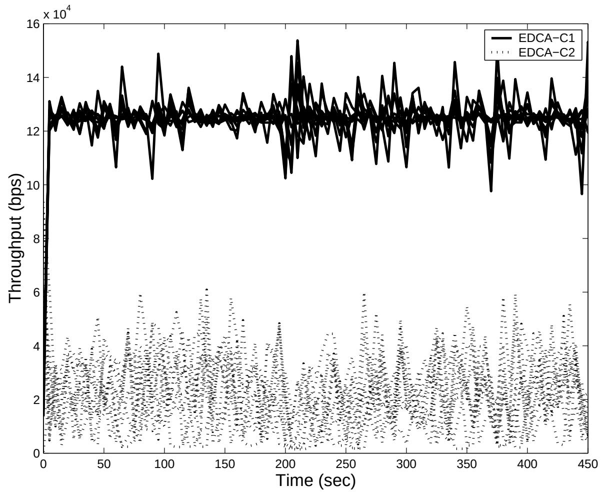
Parameters Sensitivity: In a dynamic environment, the amount of traffic, traffic pattern, and the number of MSs can change frequently. The quality of QoS support of EDCA is highly sensitive to the given set of parameters, i.e. AIFS, CW_{min}, CW_{max}. With small changes of parameters, e.g., CW_{min}, drastically different QoS performance can be observed, as shown later in the simulation result. In EDCA, different values of CW_{min} and CW_{max} are used to provide differentiation among priority classes. However, it is not easy to find a right set of parameters that provides satisfactory performance for high priority, acceptable performance for low priority, and robustness for all priority classes in a complex and diverse scenarios. The reason is that the right set of parameters depend upon many factors such as load [76, 77] and the number of active MSs. A demonstration via simulation is presented next.

We will demonstrate this problem by considering two scenarios. The first scenario (EDCA-10C1) consists of MSs only from high priority class and the second scenario consists MSs from both high priority and low priority class. In scenario EDCA-10C1, there are 10 high priority class MSs (EDCA-C1) with an offered load of 400 Kbps in 2 Mbps WLAN. The $AIFSN$, CW_{min}, and CW_{max} of EDCA-C1 MSs are set to 1 ($PIFS$), 16, and 64, respectively. In the second scenario (EDCA-10C1+10C2), an additional 10 low priority class MSs (EDCA-C2) with an offered load of 400 Kbps are added into the network, i.e., in addition to 10 EDCA-C1 MSs from EDCA-C1 scenario. The $AIFSN$, CW_{min}, and CW_{max} of EDCA-C2 MSs are set to 2 ($DIFS$), 256, and 1024, respectively.

In EDCA-C1 scenario, when there is no traffic from a low priority class, Figure 2.7 (a) shows that all EDCA-C1 MSs receive the exact throughput as needed with very little variation. However, by adding traffic from low priority MSs, the throughput variation of EDCA-C1 MSs increases relatively significantly as shown in Figure 2.7 (b). Although the



2.7 (a): 10 EDCA MSs at 125 Kbps

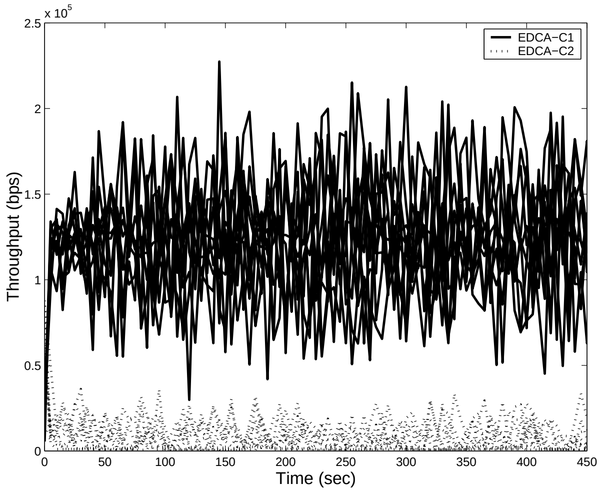


2.7 (b): 20 EDCA MSs at 125 Kbps

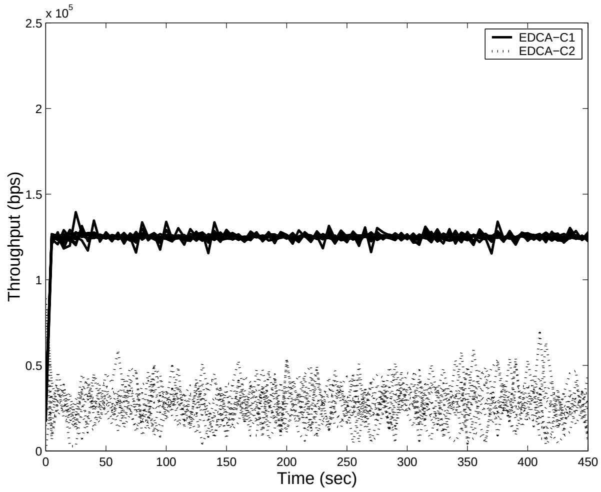
Figure 2.7: Deterioration of QoS Support in EDCA

total offered load of the high priority class remains the same and is still under the effective aggregate throughput of the WLAN (1.5 Mbps), the experienced throughput of the high priority MSs fluctuates increasingly. This result confirms the analyses of the influence of low priority traffic on high priority traffic from Robinson et al. [73] who suggests that the traffic from high priority classes of EDCA is not protected from the traffic from low priority classes even though the values of $AIFSN$ and CW_{min} of high priority class is very small, or as small as possible.

Moreover, we found that the results of EDCA-10C1+10C2 scenario can be changed tremendously, if only the value of CW_{min} of EDCA-C2 MSs is changed from 256 to 128 or 512. When $CW_{min} = 128$, the variation increases enormously as shown in Figure. 2.7 (a) whereas the variation decreases substantially when $CW_{min} = 512$ as shown in Figure. 2.7 (a). This result confirms the sensitivity of QoS support of EDCA. It is worth nothing that it might seem that EDCA can provide a good QoS support if the right parameter is assigned. However, we believe that the right parameter is not easy to assign, since different values of parameters yield vastly different results. The the values of CW_{min} and CW_{max} represent the tradeoff between the differentiated ability and the number of supported MSs. That is, the smaller the values of CW_{min} and CW_{max} are, the better the differentiation. However, the smaller the values of CW_{min} and CW_{max} are, the smaller the number of support MSs. Many sets of values have been suggested and used to demonstrate differentiated QoS support with EDCA [66, 68]. However, it not clear if any fixed set of values will be suitable for all scenarios.



2.8 (a): $CW_{min} = 128$



2.8 (b): $CW_{min} = 512$

Figure 2.8: Parameter Sensitivity of QoS Support in EDCA

In conclusion, several studies including analyses, performance evaluations, and limitations of EDCA, have been reported in past a few years. These publications suggest that EDCA can provide service differentiation support. However, EDCA may suffer from issues, such as degraded QoS support of high priority traffic, starvation of low priority traffic, uncontrolled delay, lack of admission control, lack of fairness, complex QoS translation, and sensitivity to parameter tunings. Several proposals [66, 76] have been suggested in the literature to alleviate these problems and improve the performance; however, the improved performance is achieved at the expense of increased complexity of the mechanism making it less attractive.

2.6 FAIR SCHEDULING BASED QOS SUPPORT

To overcome the unfair apportioning of bandwidth, recently there have been proposals that use fair queuing mechanisms as part of the channel access. It is important to emphasize that priority is no longer applied in fair scheduling. One flow maybe require higher throughput than the other flows. The flow with higher throughput does not have a higher priority than the other flows with lower throughput. Each flow simply requires different throughput and the scheduler attempts to provide a fair resource allocation according to the requirements.

Consider two traffic classes that need 200 Kbps and 100 Kbps respectively. A mechanism is considered to be fair if the experienced throughput for these classes are in the ratio 2:1 on average. Fair scheduling algorithms [78] attempt to partition the network resource fairly among flows in proportion to a given flow weight. They work by regulating the waiting time so that traffic in each class has a fair opportunity to be sent, which is different from the schemes that bind channel access to priority. In this case, the bandwidth is fairly apportioned between different traffic classes. Next, we will briefly describe three QoS mechanisms based on fair queue scheduling. The first two mechanisms aim at providing relative throughput support, i.e., ability to allocate fair throughput among different traffic classes, while the last mechanism considers absolute throughput support, or ability to deliver specific throughput.

2.6.1 Distributed Weighted Fair Queuing (DWFQ)

Due to the fact that the length of CW is inversely proportional to the average throughput, Banchs et al. [79] proposed a modification to the backoff algorithm called Distributed Weighted Fair Queuing (DWFQ). The mechanism is based on fair scheduling and provides fair access to shared bandwidth, in proportion to *flow weights*. In DWFQ, each MS specifies its weight. The default value of weight for best-effort traffic is equal to 1. Any number that is lower than one is not allowed while a number that is higher than 1 indicates better than best-effort service. All flows of all MSs are constrained by having the same ratio (TW_i) between experienced throughput (R_i) and a weight (W_i), i.e. $TW_i = \frac{R_i}{W_i}$. The weights are used to differentiate and apportion the bandwidth between traffic classes. By comparing its own TW_i to those of other MSs, a given MS can adjust its CW accordingly. The CW is decreased if its TW_i is smaller than those of other MSs and it is increased. However, the randomness associated with using the CW increases the variability of throughput and delay, especially in overloaded condition. To demonstrate this issue, we consider a scenario with 10 DWFQ MSs. Two MSs with low data rate (C1) demand throughput support of 150 Kbps and eight MSs high data rate (C2) demand relative throughput support of 300 Kbps. The total offered load is 1.8 Mbps which is higher than the effective channel capacity of a 2 Mbps WLAN. Figure 2.9 shows the instantaneous throughput of two representative DWFQ MSs. We can see that the instantaneous experienced throughput of both representative MSs fluctuates wildly. Moreover, DWFQ requires an additional field in the frame header in the MAC layer to exchange the values of TW_i among MSs. Requiring additional fields in the frame header is often not desirable. Finally, it is not clear what the appropriate values of W_i for each traffic classes are or how to map this value to different types of traffic class.

2.6.2 Distributed Fair Scheduling (DFS)

Rather than having fixed ranges of CW for the low priority frame and the high priority frame as in Section 2.6.1, Vaidya et al. [80] proposed a mechanism called Distributed Fair Scheduling (DFS). In DFS, each MS specifies its weight similarly to that of DWFQ, except the weight values of all traffic classes in every MSs are required to sum to 1. A packet with

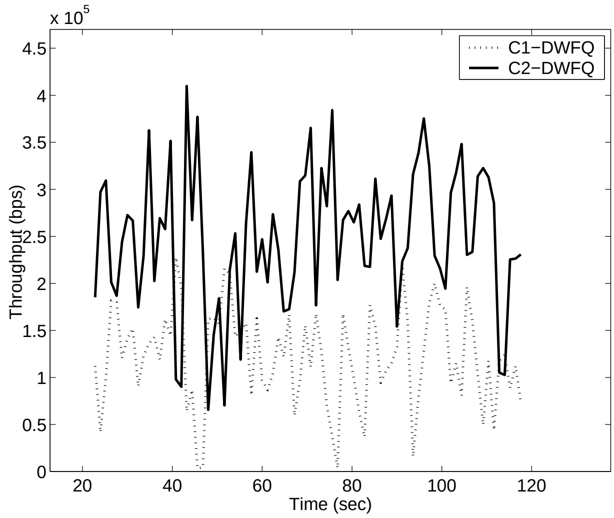


Figure 2.9: Instantaneous Throughput of DWFQ, 10 MSs at 300 Kbps

the smallest ratio between its length and its weight receives an opportunity to transmit first. The weight represents a value associated with the priority class. A higher priority traffic class has an associated higher weight. This mechanism picks a backoff interval proportional to a finish tag. The finish tag is the ratio between the packet length and the weight of a frame given by:$B_i = \left\lfloor \left\lfloor Scaling\_Factor * \frac{L_i}{\phi_i} \right\rfloor * \rho \right\rfloor$, where B_i is the backoff interval, L_i is the packet length, ϕ_i is the weight, and ρ is a random variable uniformly distributed in the range of [0.9, 1.1]. This random number with mean 1 is introduced to prevent a collision when two or more MSs count down to zero simultaneously. With the combination of weight and packet length in backoff calculation, traffic with different throughput classes can be treated differently. This mechanism is based on Self-Clocked Fair Queue (SCFQ) (Ref. 22 in [78]) which has $O(log(v))$ complexity where v is the number of flows. Additionally as the authors themselves note, the experienced throughput are quite sensitive to the choice of frame lengths and right value of weights making it complicated to map the QoS requirement into the weight, as reported in [71]. For example, it is not clear what is the meaning of a weight value of 0.3 to the users. To demonstrate the effect of the values of weight in DFS. We consider a scenario where there are two classes of traffic, five class 1 (C1) MSs and five class 2 (C2) MSs. The value of weight of C1 MSs is twice as high as the value of weight of C2 MSs.

Then, we varied the value of weight of C1 MSs between 0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2, 0.4, 0.8, and 1.0. Figure. 2.10 shows that a right value of weight is required to achieve the highest throughput. Otherwise, different values of weights for MSs will result in different experienced throughput (even though the throughput allocation is still fair). Finally, DFS supports only one type of throughput model, i.e. the relative throughput support.

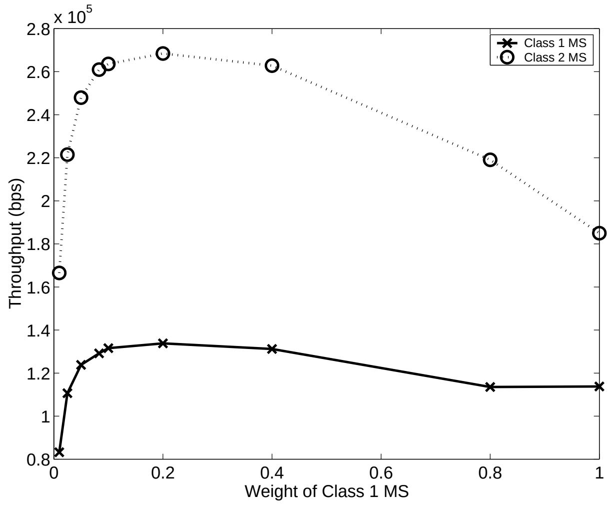


Figure 2.10: Throughput vs. weight for 10 DFS MSs

2.6.3 Assured Rate MAC Extension (ARME)

Banchs et al. [81] proposed Assured Rate MAC Extension (ARME) which is based on fair queue scheduling and aims at providing fair access to shared bandwidth, in proportion to a *rate of token bucket*. The principle of fair queuing can be used to regulate the wait time of traffic according to its class so that traffic in each class has an equal opportunity to be sent and the bandwidth is fairly apportioned between different traffic classes. The rates of token are used to differentiate between traffic classes and to apportion the bandwidth between them. In ARME, each MS receives a rate of token bucket as specified or desired throughput. By comparing the available token in the bucket, a given MS can adjust its own CW. The CW is increased if 1) the network is detected as being in an overloaded situation, 2) the queue of frame is empty or 3) the number of available tokens is smaller than the minimum requirement. The network is declared overloaded if the average number of collisions is higher than a certain threshold. If any of these 3 situations does not occur, the CW will be increased. The priority access of MS using this mechanism results in a smaller CW than that of DCF MS. This mechanism of QoS support is achieved at the expense of the scalability, another desired property of QoS support mechanism in WLANs. Additionally, the randomness associated with using the CW increases the variation of throughput and delay, especially in an overloaded situation. To demonstrate this issue, we consider a scenario with 10 ARME MSs with throughput requirement of 300 Kbps. The total offered load is 3 Mbps which is higher than the effective channel capacity of a 2 Mbps WLAN. The first MS starts transmission at time 0, the second MS at time 20, the third MS at time 40, and so on. Figure 2.11 shows that as additional ARME MSs becomes active, the ability to provide absolute throughput is significantly deteriorated whereas the instantaneous experienced throughput of each MS wildly fluctuates.

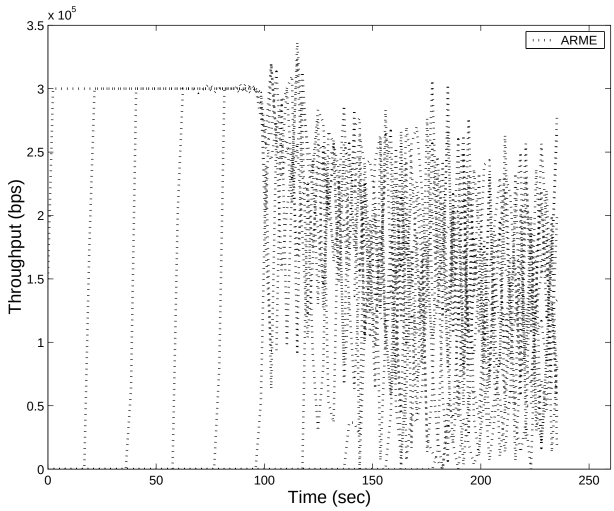


Figure 2.11: Instantaneous Throughput of ARME, 10 MSs at 300 Kbps

2.6.4 Discussion and Summary

Although, DWFQ, DFS, and ARME mechanisms are based on fair scheduling and share some similarities, these mechanisms are different in the several aspects such as the choice of the fair queuing mechanism, the translation of user requirements, the WLAN parameters used to differentiate service and the type of QoS support. The choice of the fair queue scheduling mechanism will affect the protocol complexity and computational cost. For example, DFS is based on SCFQ [82] which has $O(\log(n))$ complexity where n is the number of flows. DWFQ is based on the fairness ratio between experienced throughput and its weight while ARME is based on credit-based scheduling. However, DWFQ requires exchanging a special frame to disseminate the fairness ratio among MSs. Requiring additional frames increases the complexity and overhead of the protocol.

The method or parameter used to translate users' requirements into the parameter of the fair queuing mechanisms are different. The users' requirements are for example throughput and delay. The fair queuing parameter could be a weight as used in DWFQ and DFS. The weight parameter can provide fairness but it is difficult to map the throughput or delay to the right value of weight because it depends on other factors such as the number of flows, total channel capacity, and so on. In terms of the WLAN parameters that are used to provide QoS support, DFS modifies the value of BI whereas ARME and DWFQ modify the value of CW. Each parameter has its own differentiation ability and effectiveness that vary with scenario. The BI is directly proportional to the expected throughput while CW is less specific and random. Lastly, in term of QoS support, all three mechanisms provide only one type of QoS support, either relative throughput or absolute throughput. ARME aims at providing absolute throughput while DWFQ and DFS provide relative throughput. Although an integrated solution is a desired property of QoS support in WLANs, no mechanism has been successful in achieving a satisfactory level of integration. Thus, the purpose of this study is to develop an integrated QoS mechanism in WLANs.

2.7 DELAY-CONSTRAINT SUPPORT IN WLANS

Although priority-based mechanisms provide better access, i.e. higher throughput and smaller delay for high priority traffic, the delay still remains unbounded. Most QoS-sensitive applications require delay bounds. However, only a small number of research efforts have considered delay support.

In Decentralized Delay Fluctuation Control Mechanism (DDFC) [83], Yamada et al. proposed a mechanism to calculate the value of CW according to the buffering delay, i.e., $CW = \frac{(CWmin+1)\times 2^c \times t_0}{t-(t_s-t_0)}$, where c is number of consecutive collisions, t_s is a buffering delay threshold, and t_0 is a scaling factor. According to DDFC, the longer the buffering delay, the smaller the value of CW. As a result, the longer the time a packet has waited, shorter will be the backoff the packet will endure. Although, DDFC may reduce the average delay via a smaller calculated CW, the problem of high probability of collision due to the smaller value of CW has not been addressed. Moreover, the CW remains expanding exponentially for every consecutive collision and randomly selected from (0,CW), as before. Therefore, time-sensitive traffic may suffer from the high variation of delay. Wong and Donaldson [84] proposed another mechanism to reduce delay by calculating the Persistence Factor (PF) according to the buffering delay. PF parameter determines the degree of increase of CW when collisions occur. In DCF, $PF = 2$, which means CW is doubled for every consecutive collision. If $PF = 3$, CW is tripled instead and results in a longer waiting time. The longer the buffering delay, the smaller the value of PF. PF increases from 1 to 2 in the first half of the lifetime of a packet to avoid high probability of collisions and decreases from 1 to 0 as the buffering delay increases to the maximum time life time of a packet to decrease the long delay.

In [85], Jayaparvathy et al. proposed a mechanism to exchange the experienced delay and adjust the backoff time accordingly. If the experienced delay of a transmitting node is longer than its current experienced delay, the backoff time will be reduced, otherwise the backoff time will be doubled. Chen et al. [86] proposed another mechanism to generate a jamming signal to give an advantage to access the channel for real-time re-retransmitting MSs. This mechanism is similar to that of blackburst proposed by Sobrinho and Krishnakumar in [87].

Kanodia et al. [60, 88] attempted to address delay support in multi-hop ad hoc networks by considering three schemes, time-to-live (TTL) allocation, fixed per-node allocation (FPN), and uniform delay budget (UDB). While the TTL scheme gives preference to packets that have traveled several hops, the FPN scheme discriminates against packets with long paths. Although the UDB scheme allocates the delay budget uniformly among nodes, this scheme still does not guarantee a delay bound.

In conclusion, despite the imminent need of delay support for time-sensitive applications like voice and video, research on this topic has been surprisingly limited and largely unexplored. Only a small number of research studies has addressed delay support by giving a smaller backoff timer to time-sensitive traffic. No mechanism has accomplished closing the gap in providing delay bounds in a WLAN, which remains a fertile research area.

2.8 CONCLUDING REMARKS

Research on distributed QoS support mechanisms for 802.11 networks is a relatively new area of study, motivated by emerging applications requiring QoS support. Current mechanisms use well known QoS schemes (based on priority and fair queuing) from wired networks and map QoS requirements onto 802.11 MAC parameters. However, for priority-based mechanisms like EDCA, it remains unclear what are the right set of MAC parameters to achieve a satisfactory and robust QoS support. Most schemes in the research literature have focused on only throughput delivery. Other QoS metrics, such as delay and jitter, need more attention. The current distributed mechanisms provide only QoS differentiation, but no specific QoS level is supported. A new protocol that provides specific QoS support that is tailored for WLANs is proposed in Chapter 3. Simulation results validating the performance of the protocol will be presented in Chapter 4 and Chapter 5.

3.0 SUPPORTING RELATIVE/ABSOLUTE THROUGHPUT AND DELAY IN WLANS

In this chapter, we will present a novel Quality of Service (QoS) mechanism called Distributed Relative/Absolute Fair Throughput with Delay Support (DRAFT+D). DRAFT+D is designed specifically to provide integrated QoS support in IEEE 802.11 Wireless Local Area Networks (WLANs). The QoS support is achieved by modification of only one 802.11 Medium Access Control (MAC) parameter with ease of implementation and straight-forward QoS translation. Unlike any other distributed QoS mechanisms, DRAFT+D supports two QoS metrics (throughput and delay) with two QoS models (absolute and relative) under two fairness constraints (utilitarian and temporal fairness). While achieving the aforementioned QoS support, DRAFT+D is also equipped with a safeguard against excessive traffic to maintain predictable performance.

DRAFT+D is based on fair queue scheduling mechanisms [80], which can take advantage of carrier sensing and event synchronization in a WLAN. Carrier sensing and synchronization of backoff timers in a WLAN allow Self-Clocked Fair Queue (SCFQ) to be implemented in a fully distributed through the Backoff Interval (BI). The correct relative and absolute bandwidth allocation is achieved by calculating of BI according to the specified QoS requirements, i.e., throughput or delay. In the next section, we will present the assumptions and limitations of the research, and follow with detailed descriptions and justifications of the proposed mechanisms.

3.1 ASSUMPTIONS

The main objective of DRAFT+D is to provide QoS support in WLANs. The assumptions under which the mechanism operates are listed as follow:

- The MAC layer of each Mobile Station (MS) can only see its own network segment.
- We consider QoS support only within a single-hop WLAN. The Multiple-hop or end-to-end QoS support is not considered in the proposed mechanism; therefore, the proposed mechanism addresses QoS support only within a Basic Service Set (BSS) of a WLAN.
- Each MS independently specifies the throughput and delay requirements and operates collectively according to the provided set of rules.

3.2 OVERVIEW OF DRAFT+D

In this section, we will present an overview of DRAFT+D including its components, a working flowchart, and bandwidth allocation in DRAFT+D. This overview will provide a conceptual roadmap for the rest of this chapter. The components of QoS support in DRAFT+D are depicted in Figure 3.1. Three main objectives of DRAFT+D are to provide: a) fair QoS support, b) specific QoS support, and c) excessive traffic control. Fair QoS refers to relative throughput support while specific QoS support refers to absolute throughput and absolute delay. Both relative and absolute throughput support are achieved via calculation of BI according to the value of a flow weight. The value of this weight is a simple map from the throughput or delay requirements in a form of quantum rate. Two criteria of fairness (see Chapter 2) for relative throughput support are provided: a) utilitarian fairness and b) temporal fairness. Two types of absolute QoS are supported by allocating sufficient share of the channel capacity to these types of traffic. To simultaneously maintain absolute QoS support and prevent excessive traffic, two mechanisms are incorporated: a) Deficit Round Robin (DRR) and b) Distributed SafeGuard for Absolute Throughput Support in Relative Throughput Flows (DSG-RT). Both DRR and DSG-RT are employed independently in each MS to achieve predictable performance. DRR is used to control the load submitted by an

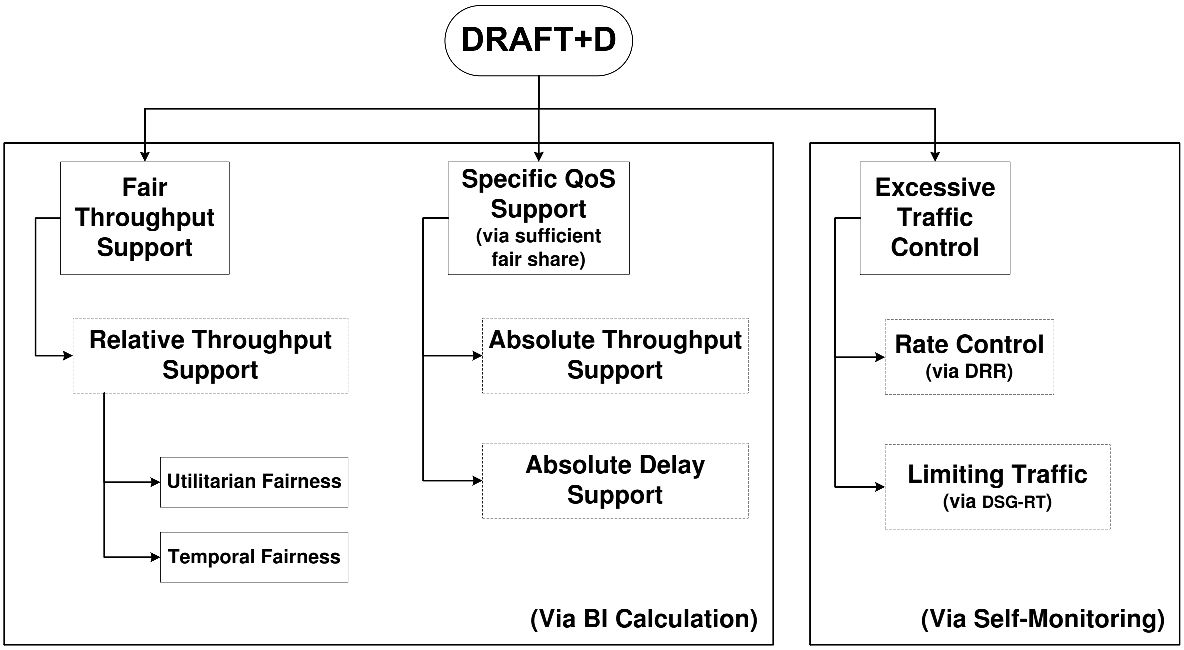


Figure 3.1: Components of QoS Support in DRAFT+D

MS to the air. DSG-RT is used to limit new flows if the network is overloaded. The main advantage of DRR and DSG-RT is that they are implemented locally and independently without requiring exchange of any information from other MSs. Unlike any other QoS mechanism in a WLAN, DRAFT+D achieves QoS support including safeguard against excessive traffic in a fully distributed manner.

One of the properties of fair-scheduling-based mechanisms is the ability to provide specific average delay [89]. Intuitively, if a fair-scheduling-based mechanism can provide a specific throughput support according to a flow weight, it will be able to provide a specific delay (on average) as well. Since DRAFT+D emulates fair scheduling in a WLAN, DRAFT+D has the advantage of the ability to provide specific delay, on average.

A flowchart in Figure 3.2 shows the five-step procedure from QoS specification to packet transmission in DRAFT+D. The first step is the QoS requirement specification, i.e., throughput (λ) or delay (δ) required. The second step is the translation of these QoS requirements into a common denominator, called quantum rate (λ). The third step is to calculate an

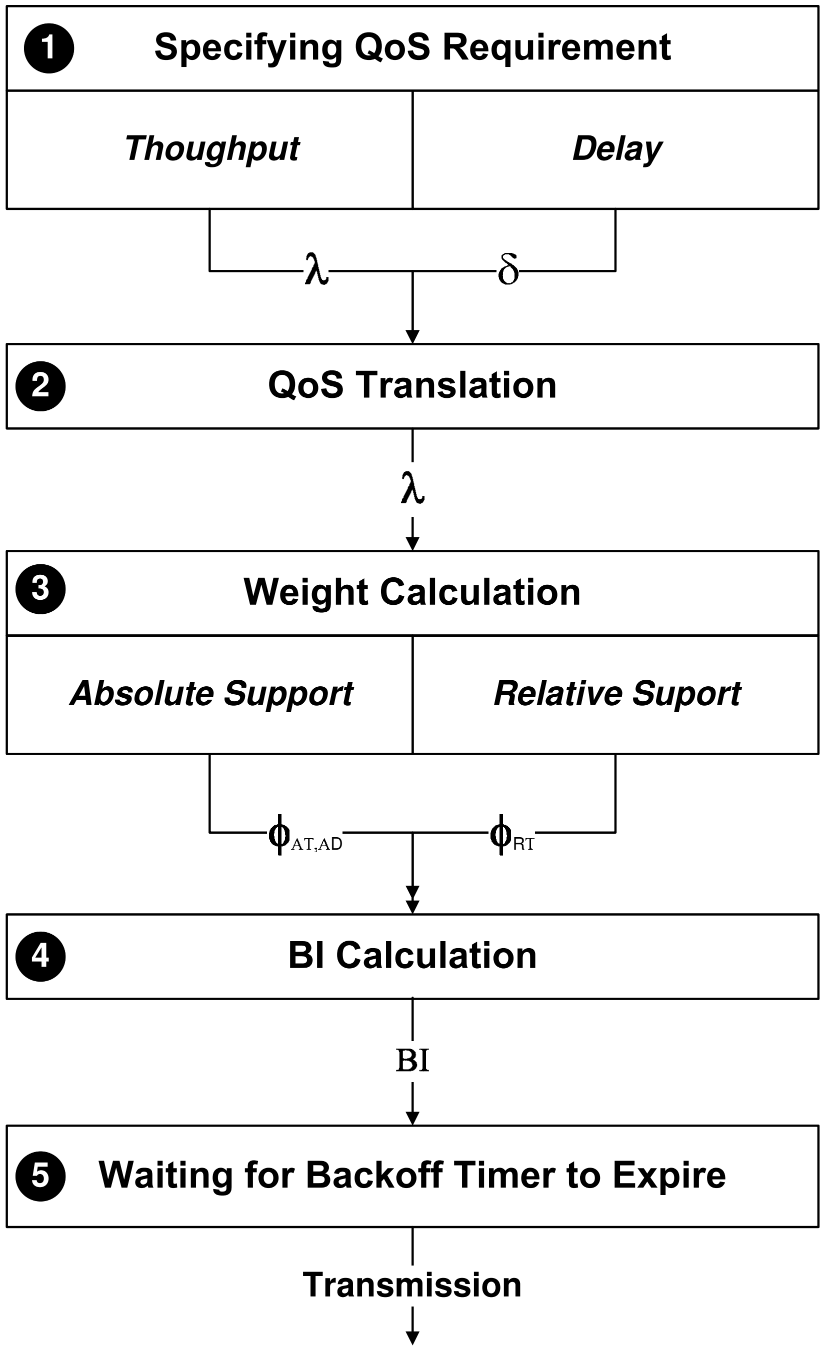


Figure 3.2: 5-step (QoS-Specification-to-Transmission) Procedure in DRAFT+D

appropriate value of a flow weight (ϕ) from the translated quantum rate according to the QoS model, i.e., absolute or relative. The fourth step is to calculate the value of BI based on the value of weight. The final step is to wait until the calculated backoff timer expires and then begins the packet transmission.

Figure 3.3 demonstrates how DRAFT+D partitions bandwidth for different QoS requirements of different classes of traffic (demand vs. supply). From the right side of the figure (supply side), we assume a WLAN with data rate RAW has an effective channel capacity of λ_e, i.e., after accounting for overhead from physical layer, protocol headers, and waiting times. Then, DRAFT+D partitions the effective channel capacity in the form of fair shares ($\overline{\lambda}$) into two QoS classes: a) absolute class ($\overline{\lambda}_{AT,AD}$), and b) relative class ($\overline{\lambda}_{RT}$).

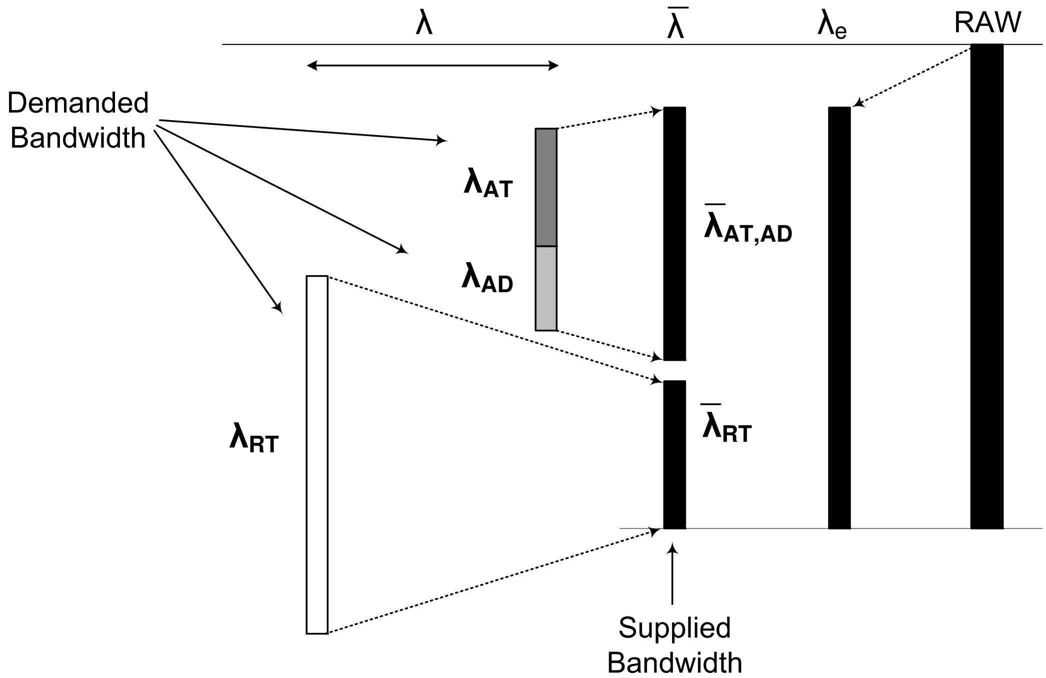


Figure 3.3: Bandwidth Allocation in DRAFT+D

From the left side of the figure (demand side), requirements from absolute and relative QoS classes in the form of quantum rates ($\lambda_{AT}, \lambda_{AD}, \lambda_{RT}$) are specified independently and handled separately. Because absolute throughput and absolute delay support share the same underlying mechanism; their quantum rates are grouped and handled together. According to these requirements, DRAFT+D, then partitions sufficient bandwidth to flows in the absolute QoS class while fairly allocating bandwidth to flows in the relative QoS class. Via a simple manipulation of weight, the share of bandwidth of flows in a absolute QoS class ($\overline{\lambda}_{AT,AD}$) is escalated to sufficiently provide specific throughput and delay support while the share of bandwidth of flows in the relative throughput class ($\overline{\lambda}_{RT}$) is de-escalated to accommodate all the traffic flows of this class. We show that as long as the sum of quantum rates of all the flows from the relative throughput class is smaller than the escalated remaining bandwidth, absolute QoS will be fully supported and maintained.

We start by describing the components in DRAFT+D DRAFT+D: a) the way DRR is applied in WLANs, b) QoS translation, and then c) the calculation of unbiased weight.

3.3 THE BASICS OF DRAFT+D

3.3.1 Applying DRR in WLANs

Traffic at each MS can be categorized into one of three supported QoS classes, i.e., a) absolute throughput, b) relative throughput, and c) absolute delay. A traffic class i at MS j ($TC^j[i]$) with throughput requirement ($\lambda^j[i]$) or with target delay ($\delta^j[i]$) (i.e., specified independently by each MS) is allocated a *service quantum* of Q bits every $t^j[i]$ second. Without loss of generality, we assume that there is only one traffic class per MS and omit the superscript in what follows to simplify the presentation. The quantum rate is set equal to the throughput requirement (from either throughput or delay). Traffic requiring a higher throughput or lower target delay is given a higher quantum rate. We will explain later how DRAFT+D translates a given delay requirement to a corresponding throughput requirement. Therefore, we will use the same notation for throughput requirement and the quantum rate, i.e., $\lambda[i]$. In time, each traffic class in each MS *independently* accumulates service quanta according to its specified quantum rate. The accumulated quanta of $TC[i]$, which represent unused resources corresponds to a *Deficit Counter* ($DC[i]$). $DC[i]$ is increased continuously with time at a rate of $\lambda[i]$ up to the maximum limit (DC_{max}) as shown in Eq. (3.1). $DC[i]$ is decreased by the size of the frame in bits ($l[i]$) whenever a frame is successfully transmitted as shown in Eq. (3.2).

$$DC[i](t) = DC[i](t') + \lambda[i] \times (t - t'), \quad DC[i] < DC_{max} \tag{3.1}$$

$$DC[i](t) = DC[i](t) - l[i](t) \tag{3.2}$$

The DRR mechanism acts as a traffic regulator to control the flow rate, as shown in Figure 3.4. During the test process, if $DC[i]$ is below a minimum limit (DC_{min}), $TC[i]$ is not eligible to contend for network access. Without a traffic regulator, it is difficult, if not impossible, to provide absolute QoS support since traffic can be injected into the network unpredictably and/or excessively. This embedded rate control is the first advantage of DRAFT+D.

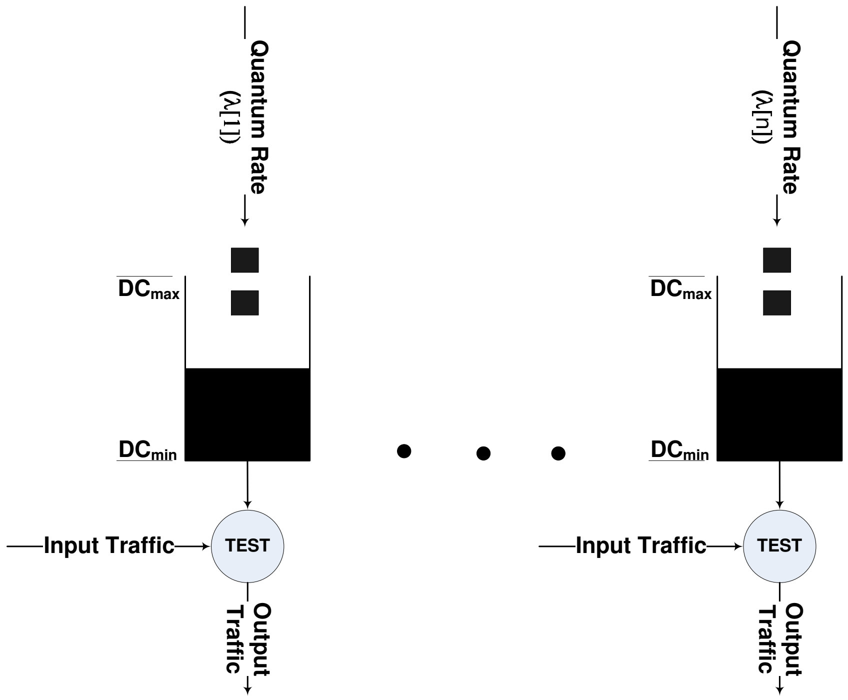


Figure 3.4: Token Bucket - A Rate Control Mechanism in DRAFT+D

DRR can be used to further reduce the probability of collision by calculating a Variable Length IFS (VIFS) according to the value of $DC[i]$. However, this is not the main focus of the proposed mechanism. A description and preliminary simulation results of VIFS mechanism can be found in Appendix A.

3.3.2 QoS Translation

In Enhanced Distributed Channel Access (EDCA), it is not obvious what the values of $AIFS$, CW_{min}, and CW_{max} should be to provide specific throughput and delay targets. In DRAFT+D, QoS translation is straight-forward. Let us suppose that the desired throughput is a throughput requirement parameter and the target delay is a delay parameter. To simplify the QoS translation, both throughput and delay requirements are mapped into a common denominator, i.e., quantum rate. Under throughput support, the quantum rate is set equal to the throughput requirement. For example, if the throughput requirement of a flow is 200 Kbps, we simply let the quantum $\lambda[i]$ be 200 Kbps. Under delay support, the quantum rate is set either to: a) the throughput requirement or, b) the delay-specific quantum rate, which ever is higher. The delay-specific quantum rate (λ_d) is calculated from the ratio between the maximum packet size (L_{max}) and the specified target delay ($\delta[i]$), as shown in Eq. (3.3).

$$\lambda_d[i] \quad = \quad \frac{L_{max}}{\delta[i]} \tag{3.3}$$

For example, the delay-specific quantum rate of a traffic flow with 64 Kbps throughput and 40 ms target delay is 200 Kbps, assuming $L = 1$ Kbyte ($\frac{8000}{0.04}$). Therefore, a bandwidth requirement of 200 Kbps (rather than 64 Kbps) is picked as the quantum rate. In this study, target delay is the Head-of-Queue (HoQ) delay. The HoQ delay is the duration from the time a packet arrives at the head-of-queue until the reception of acknowledgement. Other types of delay metric, e.g., MAC delay, are also applicable at the price of higher throughput requirement. With this QoS mapping, the rest of the mechanism makes use of $\lambda[i]$ (either from throughput requirement or delay requirement). $\lambda[i]$ is then used in the calculation of weight ($\phi[i]$) to provide fairness, as discussed next.

3.3.3 Calculating Unbiased Weight

An unbiased weight is a representation of quantum rate in the form of a flow weight. The unbiased weight is from the ratio between the quantum rate and a reference data rate, as shown in Eq. (3.4).

$$\phi[i] = \frac{\lambda[i]}{R} \tag{3.4}$$

$\phi[i]$ denotes the weight of class i, $\lambda[i]$ denotes the quantum rate, and R denotes a reference data rate which is set to 1 Mbps. According to Eq. (3.4), a flow with a quantum rate of 200 Kbps receives a weight of 0.2 ($\frac{200Kbps}{1Mbps}$) while a flow with a quantum rate of 400 Kbps receives a weight of 0.4 ($\frac{400Kbps}{1Mbps}$).

The required quantum rate for delay support was discussed earlier in Section 3.3.2. The unbiased weight for delay support can be calculated simply by plugging the required quantum rate into Eq. (3.4). Collapsing all QoS requirements into quantum rates helps simplify the weight calculation and entire QoS support mechanism. Next, we will present the ways to calculate the weights and backoff intervals to provide absolute and relative QoS support, and the rationale for the calculation in detail.

3.4 RELATIVE AND ABSOLUTE QOS SUPPORT

3.4.1 Calculating Weight

The way in which the weight parameter is calculated is crucial to provide fairness and QoS support. We will now present the way to calculate weight to provide relative throughput, absolute throughput and absolute delay support. While, relative throughput support aims at providing fair bandwidth allocation to different flows, absolute throughput support aims at providing specific bandwidth allocation. To provide both relative throughput and absolute throughput support, we have to calculate the corresponding weights differently.

For relative throughput support, the unbiased weight is multiplied by a deescalating factor called θ ($0 < \theta \leq 1$), as shown in Eq. (3.5). $\phi_{rt}[i]$ denotes the de-escalated weight of a relative throughput flow of class i. For example, if $\theta = 0.5$ and $L = 1$ Kbyte, a flow requiring relative throughput of 200 Kbps receives a relative weight of 0.1 ($0.5 \cdot \frac{200Kbps}{1Mbps}$) while a flow with relative throughput of 400 Kbps receives a relative weight of 0.2 ($0.5 \cdot \frac{400Kbps}{1Mbps}$).

$$\phi_{rt}[i] \quad = \quad \theta \cdot \phi[i] \tag{3.5}$$

For absolute throughput and absolute delay support, the unbiased weight is multiplied by an escalating factor called ω ($\omega \geq 1$), as shown in Eq. (3.6). $\phi_{at}[i]$ denotes the weight of absolute class i. For example, assuming $\omega = 5$ and $L = 1$ Kbyte, a flow requiring absolute throughput of 200 Kbps receives a weight of 1 ($5 \cdot \frac{200Kbps}{1Mbps}$) while a flow requiring absolute throughput of 400 Kbps receives a weight of 2 ($5 \cdot \frac{400Kbps}{1Mbps}$).

$$\phi_{at}[i] \quad = \quad \omega \cdot \phi[i] \tag{3.6}$$

Under utilitarian fairness, MSs with the same throughput requirements should receive the same experienced throughput, irrespective of the data rate at which they are connected. However, the network resources which each user utilizes may be drastically different depending on the data rate at which they are connected. Under temporal fairness, each MS with a different data rate should receive a different weight. To achieve temporal fairness, a MS connecting with a lower data rate should receive a lower weight than a MS connecting with a high data rate so that both MSs will utilize the same channel time. In DRAFT+D, temporal fairness can be enforced simply by multiplying a temporal factor (γ) to the weight calculations in Eq. (3.5) and Eq. (3.6). The temporal factor can be calculated as shown in Eq. (3.7).

$$\gamma \quad = \quad \frac{R_{connected}}{R_{max}} \tag{3.7}$$

For example, under temporal fairness, the weight value of a MS-1 requiring 64 Kbps and connected at 2 Mbps is equal to 0.036 whereas the weight value of a MS-2 requiring 64 Kbps and connected at 11 Mbps is equal to 0.2. We can see that the weight value of MS-1 is around 5.5 times *less* than that of MS-2 because MS-1 requires 5.5 times *more* time to send the same amount of data than MS-2. Simulation results to demonstrate the use of temporal fairness will be presented in Chapter 5.

The four parameters, ϕ_{rt}, ϕ_{at}, θ, and ω, are important in achieving absolute QoS and relative throughput support. Flows from an absolute throughput class expect to receive as much bandwidth as requested while the flows from relative throughput classes will fairly share the rest of the bandwidth according to their weights. Next, we will present the way the weight parameter in DRAFT+D is mapped into an 802.11 MAC parameter, namely the backoff interval.

3.4.2 Calculating Backoff Interval

In EDCA, two 802.11 MAC parameters, i.e., InterFrame Space (IFS) and BI, are modified to provide differentiated but unfair QoS support. In DRAFT+D, only one parameter, the BI, is modified to provide both relative and absolute QoS support. The modification of BI is simple yet effective, and can be used to provide both relative and absolute QoS support.

To provide relative QoS support, we calculate the values of Contention Window (CW) and BI in a manner similar to Distributed Fair Scheduling (DFS), proposed by Vaidya et al. [80]. Eq. (3.8) - Eq. (3.11) show how the values of CW and BI are calculated. We will explain the equations below.

$$CW_{center}[i] = 2^{\kappa} \cdot \frac{L}{\phi[i]} \tag{3.8}$$

$$CW[i] = \frac{R_{max}}{\lambda[i]} \tag{3.9}$$

$$CW_{lower}[i] = \left\lceil CW_{center}[i] - \frac{CW[i]}{2} \right\rceil \tag{3.10}$$

$$CW_{upper}[i] = \left\lceil CW_{center}[i] + \frac{CW[i]}{2} \right\rceil \tag{3.11}$$

$$BI[i] = \left\lceil CW_{lower}[i], CW_{upper}[i] \right\rceil \tag{3.12}$$

According to Eq. (3.8), $CW_{center}[i]$ is the mean value of CW of traffic class i with weight $\phi[i]$. $\phi[i]$ is calculated from the specified throughput or target delay requirement; L is the packet size in Kbytes. The parameter κ is a constant used in the calculation of CW_{center}. A higher value of κ results in a higher value of CW_{center} and a smaller probability of collision.

A discussion and the sensitivity analysis of κ will be given in Section 4.3.1. According to Eq. (3.8), the value of CW_{center} is inversely proportional to the value of weight. That is, the larger the value of weight, the smaller is the value of $CW_{center}[i]$. For example, if $\kappa = 5$ and $L = 1$ Kbyte, a weight of 0.2 yields a $CW_{center}[1]$ of 160 ($\frac{2^5}{0.2}$) while a weight of 0.4 yields a $CW_{center}[2]$ of 80 ($\frac{2^5}{0.4}$).

$CW[i]$ represents the number of possible values of BI of traffic class i. Note that in Eq. (3.9), the maximum data rate of the WLAN R_{max} is used to calculate CW because this value will result in the largest range of $CW[i]$ helping to avoid potential unnecessary collisions. For example, $CW[i]$ of a flow with 500 Kbps is 22 ($\frac{11Mbps}{500kbps}$). Therefore, there are 22 spaces regardless of the connected data rate. $CW[i]$ is doubled for every consecutive collision.

A backoff slot ($BI[i]$) is uniformly and randomly selected from a range between $CW_{lower}[i]$ and $CW_{upper}[i]$, as shown in Eq. (3.12). That is, $CW_{lower}[i]$ is the lower bound of $BI[i]$ while $CW_{upper}[i]$ is the upper bound of $BI[i]$ of a particular weight value ($\phi[i]$). By calculating the value of BI based on weight, the collisions among different classes with different values of weight are reduced automatically. The reason is that the range of $BI[i]$ of traffic with different weights can be separated from each other, as shown in Figure 3.5. For example, the

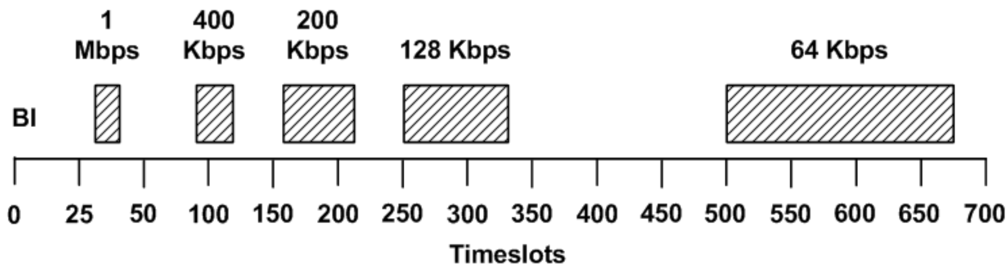


Figure 3.5: Non-Overlapping Ranges of $BI[i]$ in DRAFT+D

BI of MS-1 with $\lambda[1] = 200Kbps$, is between 132 and 188 (160 ± 28) whereas the BI of MS-2 with $\lambda[2] = 400Kbps$ is between 66 and 94 (80 ± 14), assuming $\kappa = 5$ and $L = 1$ Kbytes. Under normal circumstances, the values or steps of the required throughput of applications may be quite different from one another, e.g. 1 Mbps, 400 Kbps, 200 Kbps, 128 Kbps, and 64 Kbps.

Further, the calculation of CW and BI based on weight also provides fairness. According to Eq. (3.8), a flow of the same weight will receive the same range of possible BI. *Because the waiting time for a packet transmission is directly proportional to the value of selected BI, flows with the same weight will receive the same experienced throughput and/or experienced delay, on average*, as show in Figure 3.6. Flows with proportionally larger or smaller weight wait for proportionally shorter or longer periods. Moreover, because each MS independently calculates its own CW and BI and all MSs share a common clock via the idle and busy states, and the waiting time is directly proportional to the value of weight, fairness can be achieved in a fully distributed manner without requiring the exchange or collection of any information from other MSs. Next, we will provide an analysis of the condition under which absolute QoS can be supported and maintained.

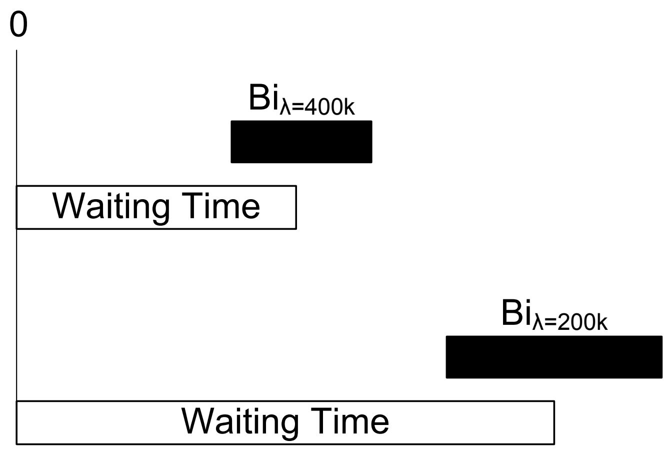


Figure 3.6: Proportional Waiting Time via $BI[i]$ in DRAFT+D

3.4.3 Analysis of Absolute QoS Support

Let us define the fair share of channel capacity as the proportion of effective channel capacity that a flow should receive according to its weight. Let $\lambda^k[x]$ denote a traffic flow of class x at MS k with a weight of $\phi[x]$ and λ_e denote the effective channel capacity of a WLAN. We once again assume one traffic class per MS and omit the superscript k in what follows. The fair share of $\lambda[x]$ (denoted by $\overline{\lambda}[x]$) is equal to the product of the effective channel capacity and the normalized weight, as shown in Eq. (3.13). The normalized weight represents the percentage of the weight value in comparison to the sum of weights of all flows in all MSs. The normalized weight is thus the ratio between the weight of the MS and the sum of all weights of all flows in the network. It is important to note that, in practice, the share of a flow can be observed via the experienced throughput without the need to know the sum of all weights. The reason is that every MS in a WLAN listens to the idle/busy period, and elapses or freezes its backoff timer accordingly. Therefore, their waiting times are directly proportional to the value of weights and fair shares.

$$\overline{\lambda}[x] \quad = \quad \lambda_e \cdot \frac{\phi[x]}{\sum\limits_{\forall i} \phi[i]} \tag{3.13}$$

In what follows, AT, AD, and RT represent all flows in absolute throughput, absolute delay, and relative throughput classes respectively.

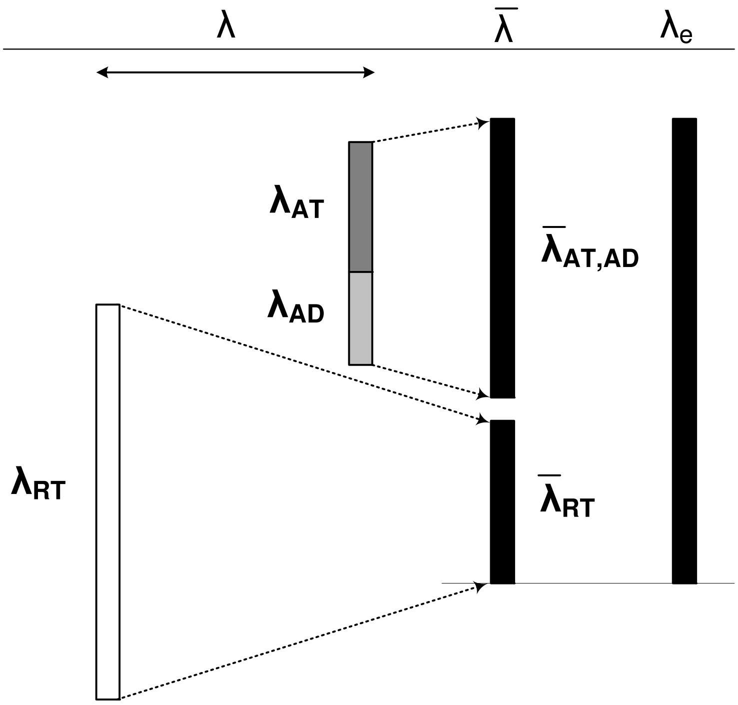


Figure 3.7: Fair Share Allocation in DRAFT+D

The parameter ω escalates the fair share of the flows requiring absolute QoS support while the parameter θ de-escalates the fair share of flows requiring relative throughput support.

Consequently, the fair share of flows from the absolute QoS class ($\overline{\lambda}[x \in AT, AD]$) are escalated and the fair share of flows from the relative throughput class ($\overline{\lambda}[x \in RT]$) are de-escalated compared to the unbiased fair share calculated from the unbiased weight ($\phi[i]$), as shown in Eq. (3.14) and Eq. (3.15), and graphically in Figure 3.7.

$$\overline{\lambda}[x \in AT, AD] = \lambda_e \cdot \omega \cdot \frac{\phi[x]}{\sum_{\forall i} \phi[i]} \tag{3.14}$$

$$\overline{\lambda}[y \in RT] = \lambda_e \cdot \theta \cdot \frac{\phi[y]}{\sum_{\forall i} \phi[i]} \tag{3.15}$$

In Eq. (3.14), we can substitute the weight by the quantum rate and show this relationship in terms of quantum rate, as shown in Eq. (3.16).

$$\overline{\lambda}[x \in AT, AD] = \lambda_e \cdot \frac{\omega \cdot \lambda[x]}{\omega \cdot \sum_{\forall i \in AT, AD} \lambda[i] + \theta \cdot \sum_{\forall j \in RT} \lambda[j]} \tag{3.16}$$

To achieve absolute QoS support, we assume that the values of ω and θ are chosen such that the fair share of absolute QoS class ($\overline{\lambda}_{x \in AT, AD}$) is equal or higher than the specified throughput requirement ($\overline{\lambda}[x \in AT, AD] \geq \lambda[x]$). Then, the desired throughput (given by the quantum rate) will be obtained. Eq. (3.17) shows the condition under which absolute QoS can be supported.

$$\begin{aligned}
\lambda[x] &\leq \overline{\lambda}[x \in AT, AD] \\
\lambda[x] &\leq \lambda_e \cdot \frac{\omega \cdot \lambda[x]}{\omega \cdot \sum_{\forall i \in AT, AD} \lambda[i] + \theta \cdot \sum_{\forall j \in RT} \lambda[j]} \\
\sum_{\forall j \in RT} \lambda[j] &\leq \frac{\omega}{\theta} \cdot (\lambda_e - \sum_{\forall i \in AT, AD} \lambda[i])
\end{aligned} \tag{3.17}$$

From Eq. (3.17), as long as the sum of quantum rates of all the flows from the relative throughput class (the left side of the equation) is smaller than the right side of the equation, which we shall refer to as "the escalated upper bound" of the remaining bandwidth, absolute QoS will be fully supported. The remaining bandwidth is the difference between the effective channel capacity and the sum of the throughput requirement from flows in the absolute QoS class. Obviously, the total traffic from absolute QoS flows must be less than the effective channel capacity. We will call the condition in Eq. (3.17) where absolute QoS can be supported as the *overload condition* and the ratio ω/θ as the *overload ratio*. This ratio indicates how much the network can be overloaded while maintaining absolute QoS support.

3.4.4 Calculating Appropriate Values of ω and θ

While parameter ω is applied only to the traffic from absolute QoS class, parameter θ is applied only to the traffic from relative throughput class. Both parameters affect the ability to provide absolute QoS support. Appropriate values of these parameters are crucial to the performance of DRAFT+D in providing QoS support. Once appropriate values of ω and θ are calculated, we assume that they are distributed via a beacon frame to be used as standard values for all MSs in the network. Next, we will demonstrate the impact and suggest ways to compute the appropriate values of ω and θ.

Absolute QoS can be supported if the value of ω is set large enough such that the fair share of absolute QoS is higher than the specified or required quantum rate. Using algebraic manipulation of Eq (3.17), appropriate ω and θ can be computed from an estimate of the expected total throughput requirements of flows in the relative throughput class and the absolute QoS class, as shown in Eq. (3.18).

$$\frac{\omega}{\theta} \geq \frac{\sum\limits_{\forall j \in RT} \lambda[j]}{\lambda_e - \sum\limits_{\forall i \in AT,AD} \lambda[i]} \tag{3.18}$$

For example, let us suppose that we have a hypothetical WLAN *W1* with a raw data rate of 2 Mbps and $\lambda_e = 1.5$. If this WLAN is expected to support 0.5 Mbps of traffic from the absolute throughput class and 5 Mbps traffic from the relative throughput class, the appropriate value of ω/θ is $(\frac{5}{1.5-0.5}) = 5$. In this hypothetical WLAN, if the offered load is not overly excessive, absolute throughput will be fully supported. If the offered load is excessive (more than 5.5 Mbps), absolute QoS support will deteriorate gracefully to relative throughput support.

To understand the impact of ω and θ, we assume $\theta = 1$ for now. When $\theta = 1$, the overload ratio is simply equal to ω. In this case, the network can support a total quantum (from flows in the relative throughput class) ω times as much as the remaining bandwidth (i.e., the bandwidth after flows from absolute QoS class are accommodated). For example, if $\omega = 5$, $\lambda_e = 1.5$ Mbps, and the traffic from absolute throughput class is 0.5 Mbps, the amount of supported traffic from relative throughput class will be 5 Mbps, or 5 times as much as the remaining bandwidth (1 Mbps).

However, in practice, the value of ω cannot be increased indefinitely to accommodate an infinite amount of traffic. The condition when ω is no longer valid can be determined by the difference between the initial CW_{center} and the next CW_{center}. If the difference of the two CW_{center}'s is less than 1, the new value of ω is not valid. This condition can be derived from the relationship between CW_{center} and ω, as shown in Eq. (3.19) which is modified from Eq. (3.18).

$$CW_{center}[i] = \frac{2^k \cdot R}{\omega \cdot \lambda[i]} \tag{3.19}$$

For example, if $\lambda[i] = 500$ Kbps, $\kappa = 5$, and $R = 1$ Mbps, the maximum value of ω of 7 can be calculated as shown in Eq. (3.20).

$$\begin{aligned}
CW_{center,\omega 1} - CW_{center,\omega 2} &\leq 1 \\
\frac{2^5}{0.5 \cdot \omega_1} - \frac{2^5}{0.5 \cdot \omega_2} &\leq 1 \\
\omega_1^2 + \omega_1 - 64 &\geq 0 \\
\omega &\leq 7.52 \qquad (3.20)
\end{aligned}$$

That is, the valid value of ω to provide absolute throughput with for $\lambda = 500$ Kbps is between 1 and 7. Depending on the design of the expected total throughput and the maximum throughput, an appropriate value of ω can be chosen. Next, we will present the impact and the way to calculate the appropriate value of parameter θ.

From Eq. (3.18), the ratio ω/θ represents the number of times the amount of traffic from relative throughput class can be supported beyond the remaining bandwidth. Although, the range of valid value of ω is limited, we can manipulate the value of θ to support a higher the amount of traffic from relative throughput class while maintaining absolute QoS support. The smaller the value of θ, the higher the amount of traffic from relative throughput class can be supported. For example, in the same hypothetical WLAN *W1*, if we change the value of θ

from 1.0 to 0.5, the acceptable total offered load increases from 5.5 Mbps to 10.5 Mbps. This is, the total offered load in a 2 Mbps WLAN with effective throughput of only 1.5 Mbps. Out of 10.5 Mbps acceptable offered load, 0.5 Mbps is from the absolute throughput class and 10 Mbps is from the relative throughput class. This amount of offered load is exceptionally high since it is more than 650% higher than the effective channel capacity. Under most circumstances, this amount of offered load should be sufficient for any operational WLAN. We can further decrease the value of θ to accommodate a larger amount of traffic, if necessary. The trade-off of using a small value of θ is the limitation of maximum achievable throughput of traffic from the relative throughput class. The smaller the value of θ, the smaller is the actual throughput of flows in from the relative throughput class.

So far, we have described the mechanisms and analyses of DRAFT+D to provide relative throughput support, absolute throughput support, and absolute delay support. Absolute throughput and absolute delay support can be maintained as long as the amount of traffic is within the estimated value. However, if the amount of injected traffic of relative throughput class is overly excessive, the absolute throughput and absolute delay support will be deteriorated. Next, we will present a safeguard mechanism against excessive traffic from relative throughput class.

3.5 EXCESSIVE TRAFFIC CONTROL WITH DSG-RT

The objective of Distributed SafeGuard for Absolute Throughput Support in Relative Throughput Flows (DSG-RT) is to control and limit the excessive amount of traffic from relative throughput class while maintaining absolute QoS support. DSG-RT provides a set of rules that each MS applies independently to achieve the goal of predictable performance. This goal is achieved by comparing the actual experienced throughput of a new Mobile Station with Relative Throughput Requirement (RT-MS) with an expecting throughput threshold. If the experienced throughput is higher than a threshold, which we call *overload threshold*, a new flow can continue sending traffic. However, if the experienced throughput is smaller than the overload threshold, a new flow should not be allowed to continue. The overload condition can be determined locally, as shown in Eq. (3.21). This equation is derived by multiplying both sides of (3.17) with the quantum rate of class k ($\lambda[k]$).

$$\begin{aligned}\lambda[k] \cdot \sum_{\forall j \in RT} \lambda[j] &\leq \lambda[k] \cdot \frac{\omega}{\theta} \cdot \Big[\lambda_e - \sum_{\forall i \in AT,AD} \lambda[i]\Big] \\ \frac{\theta}{\omega} \cdot \lambda[k] &\leq \frac{\lambda[k]}{\sum\limits_{\forall j \in RT} \lambda[j]} \cdot \Big[\lambda_e - \sum_{\forall i \in AT,AD} \lambda[i]\Big]\end{aligned} \tag{3.21}$$

According to Eq. (3.21), as long as the experienced throughput of flows in the relative throughput class (the right side of (3.21)) equals or is higher than the product between the quantum rate and $\frac{\theta}{\omega}$ ratio (the left side of the equation), the ability to provide absolute QoS will be maintained. In Eq. (3.21), the overload threshold can be calculated according to the specified quantum rate whereas the fair share can observed from the actual experienced throughput. Because, in practice, the experienced throughput can fluctuate and its measurement may not be completely accurate, to prevent deterioration of absolute support, the original overload threshold is multiplied by a safety factor (β), as shown in Eq. (3.22).

$$\overline{\lambda}[k] \geq \beta \cdot \frac{\theta}{\omega} \cdot \lambda[k] \tag{3.22}$$

Using Eq. (3.22), we propose a safeguard mechanism against excessive traffic based on the measurement of experienced throughput during the testing period. The testing period consists of two phases: a) transient period and b) decision period. During the transient period, a new flow will transmit $N1$ packets into a WLAN. The purpose of the packets during transient period is to make sure that the experienced throughput is stable and representative. The experienced throughput is calculated from the Exponentially Weighted Moving Average (EWMA) of instantaneous throughput. After the transient period, the new flow will continue transmitting $N2$ packets into a WLAN. For every successful transmission, the experienced throughput will be measured against the overload threshold. During the decision period, if the experienced throughput is smaller than the overload threshold, the new flow will stop transmission. If not, the flow can continue. Appropriate values of the number of packets during the transient and decision period ($N1, N2$) have been determined in the sensitively analysis section in Chapter 4. Note that a number of MSs may probe the channel around the same time causing additional loads. We consider the effect of the number Ξ of MSs that probe the channel in Section 4.3.7

It is important to note that five parameters required for DSG-RT are $\overline{\lambda}[k]$, $\lambda[k]$, β, θ, and ω. These parameters are available locally at each MS. $\overline{\lambda}[k]$ can be measured independently at each MS while $\lambda[k]$ is the throughput requirement specified by each MS. The parameter β, θ and ω are assumed to be standard values in a given WLAN. The simplicity of the scheme is that these parameters are available *locally* at each MS. Therefore, each MS can independently perform the test by itself in a fully distributed way. Additionally, the overload condition in Eq. (3.22) can also be applied as an adaptive bandwidth control mechanism. Rather than stopping a new flow, the adaptive mechanism will throttle the fair share of flows from relative throughput class (see discussion in Appendix B). Finally, a distributed safeguard mechanism against excessive traffic from absolute throughput and absolute delay class is also possible in DRAFT+D with a similar method (See Appendix C for analyses, preliminary protocol description, and simulation results).

3.6 DISCUSSIONS

In WLANs, an Access Point (AP) is commonly used to relay traffic between the wired network, e.g., Internet, and the MSs. The amount of downlink traffic from the wired network often represents the majority of traffic in the WLAN [90]. If the downlink traffic destined to a particular MSs requires absolute throughput or absolute delay support, appropriate quantum rates will be assigned by the AP to such flows according to the requirement. We assume that the appropriate quantum rates are communicated by the destination MSs to the AP by a reservation mechanism in an upper layer. Once the AP receives appropriate quantum rates for the downlink from the destination MSs, the AP will use the same basic function of DRAFT+D to provide QoS support accordingly.

It is to be noted that the values of BI in DRAFT+D are likely to be larger that those of Distributed Coordination Function (DCF). Therefore, MSs with DRAFT+D will be at disadvantage in competing for bandwidth to those of DCF. However, a larger value of BI does not incur a significant delay and overhead under both light and heavy loads. The reason is that, in the light load situation, the channel is idle most of the time. The backoff timer of each MS counts down at the rate of the number of idle timeslot multiplied by the value of BI. The value of each timeslot is around 8-50 μs [91], depending on the physical layer. Therefore, the total amount of waiting time due to backoff is very small. For example, MS 1 receives a value of BI of 30 and MS 2 receives a value of BI 300. We assume that there is no other transmission in the channel and timeslot lasts 10 μs, MS 1 will wait during the backoff for 300 μs while MS 2 will wait for 3 ms. We can notice that the difference in waiting time is on the order of a few ms which is very small in comparison to other delays such transmission delay. As shown in Chapters 4 and 5, DRAFT+D significantly outperforms EDCA under heavy loads.

In DRAFT+D, we assume that each MS independently specifies the throughput and delay requirements and operates according to the provided set of rules. Selfish MSs can gain advantages in two ways. The first way is to specify unreasonably high QoS requirements. The second way is to arbitrarily calculate an unreasonably small value of BI. To prevent MSs from specifying a QoS requirement that is higher than necessary, a monetary incentive mechanism where users need to pay more if they request higher QoS requirements can be used. However, this mechanism should be implemented in a higher layer. To prevent MSs from calculating a small value of BI, a monitoring mechanism can be used to observe the backoff duration of previous transmitting packets. If an unreasonably small backoff duration is detected, selfish MSs could face penalty in accessing the network [92, 93, 94].

3.7 CONCLUDING REMARKS

In this chapter, we presented a detailed description of DRAFT+D for providing a) relative throughput, b) absolute throughput, and c) absolute delay with safeguard against excessive traffic. In EDCA, two MAC parameters are modified to provide differentiated but unfair QoS support. DRAFT+D modifies only the way BI is calculated to provide both relative and absolute QoS support. Relative throughput support is achieved by calculating fair BI.

Absolute throughput and absolute delay support is achieved by allocating sufficient fair share for these types of traffic. Absolute throughput and absolute delay can be supported as long as the overload condition is maintained. The overload condition can be maintained in a fully distributed manner via a safeguard mechanism called DSG-RT. The advantage of DSG-RT is that it can independently limit a new flow from relative throughput class without requiring any information from other MSs. In Chapter 4, we will present baseline performance evaluation, performance comparison with selected QoS mechanisms, and sensitivity analysis of DRAFT+D. In Chapter 5, comprehensive performance evaluation of DRAFT+D in several diverse and realistic scenarios will be presented.

4.0 BASELINE PERFORMANCE EVALUATION, COMPARISON, AND SENSITIVITY ANALYSIS

In Chapter 3, we presented a description of Distributed Relative/Absolute Fair Throughput with Delay Support (DRAFT+D) protocol that provides support for relative throughput, absolute throughput, and absolute delay with a distributed safeguard mechanism. In this chapter, we will present a baseline performance evaluation of DRAFT+D for providing QoS support. The important findings are summarized as follows. Fair bandwidth allocation according to weight can be achieved with very low variation of throughput and delay. Specific bandwidth allocation is also achieved by absolute throughput support via the introduction of ω and θ parameters with very low variation. The simulation results confirm that the analysis in Section 3.4.3 of fair share and overload condition are correct. Specific delay support for the Head-of-Queue (HoQ) packet is also achieved via absolute delay support. The mechanism that limits excessive traffic from new RT-MSs can be accomplished in a distributed manner without requiring exchange of any information among MSs. Finally, the results confirm the effectiveness and robustness of DRAFT+D in an extremely saturated condition (10 Mbps in 2 Mbps WLAN).

To demonstrate the performance of DRAFT+D, we conducted our simulations using OPNET 9.1. The model reused the 802.11 DCF model available in OPNET. Unless otherwise specified, the following assumptions and parameters are used:

- Each MS works independently and cooperatively according to the provided specification. We assume that no MS attempts to cheat or gain illegal advantages over the other MSs.
- MSs operate in Ad-Hoc mode which is sufficient and appropriate to evaluate the performance of the MAC mechanism. In Chapter 5, we will relax this assumption and show

the result in infrastructure mode where an AP is used to relay traffic among MSs. If there is the presence of AP in the network, the AP will use the same set of rules as DRAFT+D. Additionally, the AP is also responsible for relaying traffic among MSs or between WLANs and wired-networks.

- In this simulation, each MS has only one class of traffic per MS, however, our mechanism works for many classes of traffic per MS as well.
- All MSs are located within a single Independent Basic Service Set (IBSS) where every MS is able to detect a transmission from other MSs. We assume that hidden-terminals are not present and the Request to Send/Clear to Send (RTS/CTS) mechanism is not used. In Chapter 5, we will relax this assumption and show the results when RTS/CTS is used.
- Time variant behavior of the wireless channel is not considered. MSs are assumed to operate in a WLAN where the characteristics of the wireless channel is static. The channel is also assumed error-free. In Chapter 5, we will relax this assumption and evaluate the performance in erroneous channels.
- We perform most of the simulations on a 2 Mbps WLAN to strategically minimize the simulation times into manageable durations. In Chapter 5, we will relax this assumption and evaluate the performance in 5.5 and 11 Mbps WLANs.
- We assume that all MSs are connected to the network at the same data rate. Later, we will relax this assumption and consider multi-rate environments. However, multi-rate is considered only in the static sense. By this, we mean each MS can connect to the WLAN at different data rates. However, once an MS is connected, the data rate does not change.
- MSs can be located anywhere within the IBSS. However, they remain at the particular location during the course of simulation.
- The traffic from each MS is sent to random destinations.
- All flows are of constant bit rate. The constant bit rate traffic is to clearly demonstrate the variation of throughput and delay (or the lack of) created by the mechanism. We will relax this assumption in Chapter 5.

- The frame size is fixed at 1000 bytes. Later, we will relax this assumption and consider variable length packets.

- Buffer size is set to 256,000 bits, $IFS = DIFS$, $\kappa = 5$, $\omega = 5$.

During the course of performance evaluation, the following metrics and their measurements are used:

- **Throughput:** Throughput requirement and quantum rate refers to the offered load of input traffic. The experienced throughput denotes the rate of successful data transmission over time.

- **Delay:** The HoQ delay denotes the waiting time since a packet becomes head-of-queue until the reception of acknowledgement. The HoQ delay comprises the durations of packet transmission, packet retransmission (if any), acknowledgement of transmissions, and waiting times. The queuing delay is the duration for which the packet is waiting in the queue before it becomes the head of the queue. The MAC delay refers to the period between the time when a packet is received into the queue and the time when an acknowledgement packet is received to confirm a successful transmission.

- **Jitter:** We consider jitter only in the MAC layer. This metric measures the difference between the previous HoQ delay and the current HoQ delay.

- **Collision Rate:** Collision occurs when the backoff timer of two or more MSs reach zero at the same time. Collision rate is calculated from the number of collisions per second.

- **Aggregate Throughput:** The aggregate throughput is the sum of throughput from all MSs

- **Backoff Interval:** The backoff interval is recorded from the actual selected values of BI for each packet.

In the next section, we will present an overview of the results and follow with sensitivity analysis for important parameters of DRAFT+D, i.e., κ, ω, θ, $N1$, $N2$, β and Ξ.

4.1 OVERVIEW OF RESULTS

In this section, sets of results that highlight the main functionalities of DRAFT+D are presented. We will show that DRAFT+D provides excellent QoS support, namely relative throughout support, absolute QoS support, and DSG-RT. For relative throughput support, fair throughput allocation is achieved with respect to the two fairness constraints (utilitarian fairness and temporal fairness). For absolute QoS support, specific QoS requirement with respect to throughput or delay are achieved. Finally, DRAFT+D also provides good performance of DSG-RT to prevent injection of excessive amounts of traffic. All QoS support is provided via simple calculation of BI in a fair and fully distributed manner.

To demonstrate the performance of DRAFT+D, we will evaluate 7 scenarios where the number of Mobile Stations with Relative Throughput Requirement (RT-MSs), the number of Mobile Stations with Absolute Throughput Requirement (AT-MSs), the number of Mobile Stations with Absolute Delay Requirement (AD-MSs), the total throughput requirement (Σ), and DSG-RT flags vary as shown in Table 4.1. In scenario 10RT, we will demonstrate the ability to DRAFT+D provide fair bandwidth allocation via relative throughput support. In scenario 1AT+9RT, we will demonstrate the ability to provide specific bandwidth allocation via the absolute throughput support. In scenario 1AT+19RT, we will show that the analysis of fair share and overload condition in Chapter 3 is confirmed with the simulation results. Scenario 1AT+199RT-05TT will show the impact of parameter θ while scenario 1AT+19RT-DSG will show the ability to limit excessive traffic using DSG-RT. Scenario 1AD+19RT will show the ability to provide absolute delay support of DRAFT+D and scenario 1AT+1AD+18RT will show the ability of DRAFT+D to simultaneously support the aforementioned three classes of QoS.

In these scenarios and in what is to follow, a new MS starts transmitting traffic 10 seconds (s) after the previous MS. The first MS starts transmission at time $t = 10s$, the second MS at time $t = 20s$, the third MS at time $t = 30s$, and so on. Each MS remains active for 240 seconds. Thus, the first MS is active during $10s < t < 250s$, the second MS during $20s < t < 260s$, the third MS during $30s < t < 270s$, and so on. During the course of simulation of 500 seconds, all MSs are active simultaneously during $100s < t < 250s$.

Table 4.1: Simulation Parameters in the Overview of Results

#	Name	AT-MSs	RT-MSs	AD-MSs	Σ	DSG-RT
1	10RT	-	10RT@500Kbps	-	5 Mbps	no
2	1AT+9RT	1AT@500Kbps	9RT@500Kbps	-	5 Mbps	no
3	1AT+19RT	1AT@500Kbps	19RT@500Kbps	-	10 Mbps	no
4	1AT+19RT-05TT	1AT@500Kbps	19RT@500Kbps	-	10 Mbps	no
5	1AT+19RT-DSG	1AT@500Kbps	19RT@500Kbps	-	10 Mbps	yes
6	1AD+19RT	-	19RT@500Kbps	1AD@0.016s	10 Kbps	no
7	1AT+1AD+18RT	1AT@100Kps	18RT@500Kbps	1AD@0.02s	10 Kbps	no

4.1.1 Providing Relative Throughput Support

We will first demonstrate the ability of DRAFT+D to provide fair throughput support. In scenario 10RT, a WLAN comprises of 10 RT-MSs with throughput requirement (λ) of 500 Kbps, making the total throughput requirement 5 Mbps in a 2 Mbps WLAN. Figure 4.1 (a) shows that RT-MSs receive as much throughput as required (500 Kbps) when the network is not saturated. Network saturation is the condition where the total throughput requirement, in the form of offered load or quantum rate, is higher than the effective channel capacity. The effective channel capacity is around 1.5 Mbps throughout the simulation (Figure 4.1 (b)). In scenario 10RT, the saturated condition begins when MS 4 becomes active at time $t = 40s$. At this point, the total offered load adds up to 2 Mbps while the effective channel capacity is around 1.5 Mbps.

Under saturated conditions, each RT-MS fairly shares the effective channel capacity according to its weight. Since all RT-MSs in scenario 10RT have the same weight, they receive the same experienced throughput. We also observe that the variation of throughput is very small even in a highly saturated condition where the total throughput requirement is more than 3 times the effective channel capacity (5 Mbps vs. 1.5 Mbps). The result is comparable to that of DFS which is a similar mechanism.

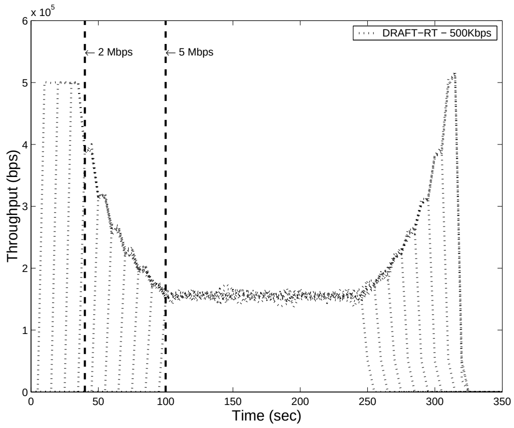
Figure 4.1 (c) shows the MAC delay of RT-MSs. As expected, the MAC delay and its variation are very small when the total throughput requirement is less than the effective channel capacity. We see that the MAC delay is directly proportional to the value of weight.

Figure 4.1 (d) and Figure 4.1 (e) show that jitter and collision rates are also relatively small. The jitter is within ±40 ms and the mean collision rate is only around 0.13. The collision rates of all MSs are under 0.3 most of the time. Finally, Figure 4.1 (f) shows the selected backoff interval. We observe that the variation of selected backoff interval is very small, between 61 and 66. The initial result of fair experienced throughput with low variation, low collision rate, and stable effective channel capacity indicates the efficiency and effectiveness of DRAFT+D. These are important properties of the QoS mechanism. We will show in Section 5.1 that these properties remain intact even in more complicated scenarios.

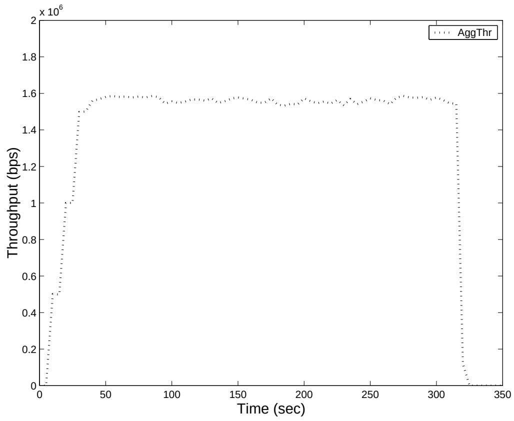
4.1.2 Providing Absolute Throughput Support

In the previous section, we demonstrated the excellent performance of DRAFT+D to provide fair throughput support. In this section, we will demonstrate the unique ability of the mechanism to provide absolute throughput support. Unlike any other distributed QoS mechanism, DRAFT+D is specifically designed to *recognize* and *support* different QoS requirements for different traffic flows at the same time in the same mechanism. DRAFT+D delivers the exact throughput (on average) to Mobile Station with Absolute Throughput Requirements (AT-MSs) and allocates fair shares of the remaining bandwidth to RT-MSs. It is important to note that, without admission control, absolute throughput support can not be guaranteed. In this scenario, we assume that the total offered load from AT-MSs and RT-MSs are within reason. We will relax this assumption in later cases where the distributed safeguard mechanism is employed.

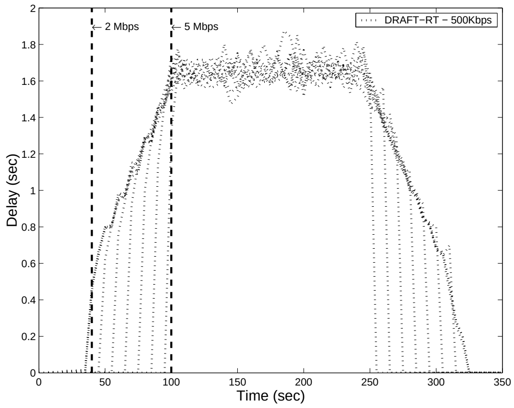
In this scenario (1AT+9RT), there are 1 AT-MS and 9 RT-MSs. Other simulation parameters remain the same as the scenario 10RT. Figure 4.2 (a) shows that the AT-MS receives as much throughput as its demand throughout the simulation even in a heavily saturated condition while RT-MSs fairly share the remaining bandwidth. The remaining bandwidth (λ_r) is defined as the difference between the effective channel capacity (λ_e) and the total throughput requirement of AT-MSs ($\sum_{\forall i \in AT} \lambda[i]$), $\lambda_r = \lambda_e - \sum_{\forall i \in AT} \lambda[i]$. Figure 4.2 (b) shows the delay performance where the MAC delay of AT-MS is around 100 times smaller than that of RT-MSs. While the MAC delay of RT-MSs is as high as 2.5 seconds, the MAC



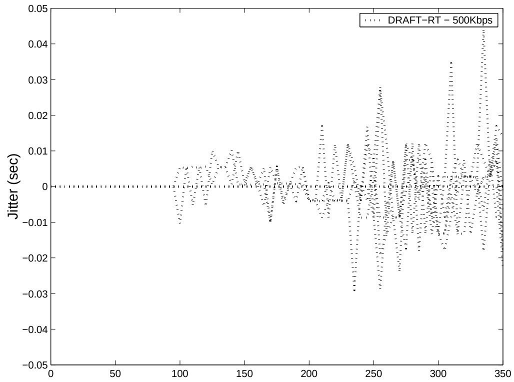
4.1 (a): Throughput



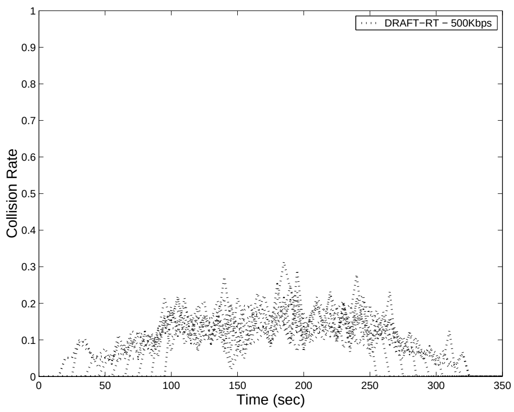
4.1 (b): Aggregate Throughput



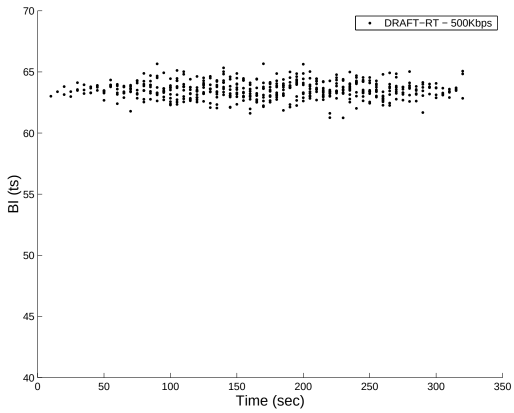
4.1 (c): MAC Delay



4.1 (d): Jitter



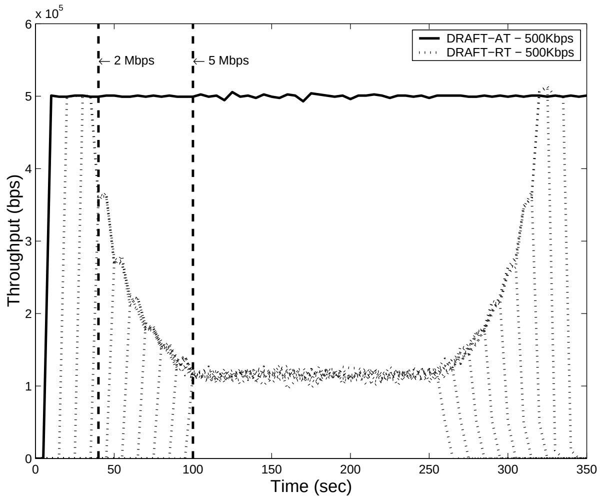
4.1 (e): Collision Rate



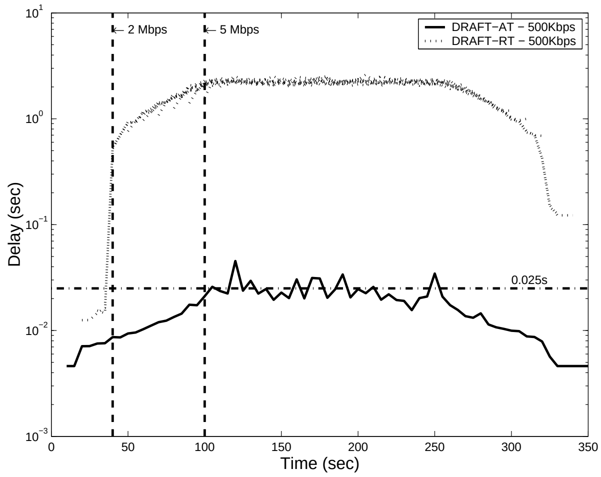
4.1 (f): Backoff Interval

Figure 4.1: Relative Throughput Support with 10 RT-MSs @ 500 Kbps

delay of AT-MS is only around 0.025 second. The ability of DRAFT+D to delivery specific throughput and low delay in a highly saturated condition suggests that absolute throughput support of DRAFT+D can be viable for QoS-sensitive applications.



4.2 (a): Throughput



4.2 (b): Delay

Figure 4.2: Absolute Throughput Support with 1 AT-MS & 9 RT-MSs @ 500 Kbps

According to the analysis in Chapter 3, absolute throughput can be supported as long as the overload condition (Eq. 3.17) is satisfied. If the overload condition is violated, the experienced throughput of AT-MSs will deteriorate (gracefully) to relative throughput. We will demonstrate this behavior by adding 10 more RT-MSs to the scenario 1AT+9RT. In the new scenario (1AT+19RT), there are 1 AT-MS and 19 RT-MSs, i.e., the total throughput requirement is 10 Mbps. As the number of active MSs increases, the condition needed to maintain support of absolute throughput will be violated at some point. Based on Eq. (3.17), the maximum number of RT-MSs (RT_{max}) that DRAFT+D can support before losing the ability to provide absolute throughput can be calculated as follows (assuming all RT-MSs have the same throughput requirement):

$$RT_{max} \leq \frac{\omega}{\theta} \times \frac{\lambda_e - \sum\limits_{\forall i \in AT} \lambda[i]}{\lambda_{j \in RT}}$$

For the RT-MSs with throughput requirement of 500 Kbps ($\lambda_{j\in RT} = 500$ Kbps), the sum of throughput requirement of AT-MS of 500 Kbps ($\sum\limits_{\forall i \in AT} \lambda[i] = 500$ Kbps), effective channel capacity of 1.5 Mbps ($\lambda_e = 1.5$ Mbps), $\omega = 5$, and $\theta = 1$, the maximum number of RT-MSs where the absolute throughput can be maintained is 10. At this point (Figure. 4.3), the total throughput requirement is 5 Mbps, 4.5 Mbps is from RT-MSs and 0.5 Mbps is from AT-MS. Figure. 4.3 shows that the analytical estimate of the maximum number of RT-MSs, i.e., 10, matches exactly the result in the simulation. We conclude that the analyses of fair share and the condition under which DRAFT+D can provide absolute throughput support in Chapter 3 are correct.

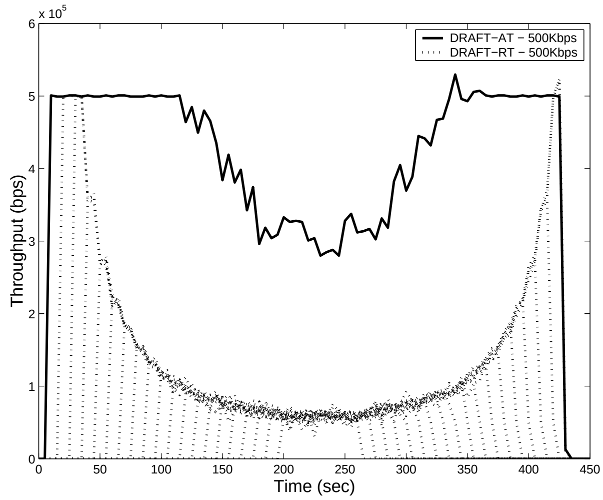


Figure 4.3: Absolute Throughput Support with 1 AT-MS & 19 RT-MSs @ 500 Kbps, $\theta = 1.0$

It important to point out that, prior to the deterioration of absolute throughput support (Figure. 4.3), the throughput required by AT-MS is fulfilled and the experienced throughput of RT-MSs remains very stable even though the network is more than 3 times higher than the effective channel capacity. However, we believe that an even better performance can be achieved. Next, we consider the two potential solutions employed in RT-MSs to provide better absolute throughput support in the situation where additional traffic from RT-MSs is injected into the network.

1. Calculating an appropriate value of θ

2. Employing DSG-RT mechanism in RT-MSs

4.1.3 The Effect of Parameter θ on Absolute Throughput Support

Parameter θ is a deescalating factor that is used to reduce the fair share of RT-MSs in comparison to the fair share based on the normal weight calculation. The range of θ is between 0 and 1 ($0 < \theta \leq 1$). The smaller the value of θ, the smaller is the fair share of RT-MSs. According to Eq. (3.17), the higher the value of overload ratio (ω/θ), the larger is the amount of traffic from RT-MSs that can be supported while absolute throughput is maintained. Therefore, to achieve a higher amount of supported traffic from RT-MSs, we can simply increase the value of the overload ratio. Two ways that the overload ratio can be increased: a) decreasing θ (discussed below) or b) increasing ω (discussed in Section 4.3).

The first method to prolong the absolute throughput support is to calculate an appropriate value of θ. The appropriate value of θ can be calculated by Eq. (4.1) during the phase of network design where the maximum throughput requirement of RT-MS can be projected or estimated. Assume that we would like to support the traffic from RT-MSs up to 9.5 Mbps and traffic from AT-MSs up to 0.5 Mbps, i.e., total of 10 Mbps in 2 Mbps WLAN. According to Eq. (4.1), the value of θ must be equal to or less than 0.5 to achieve this objective.

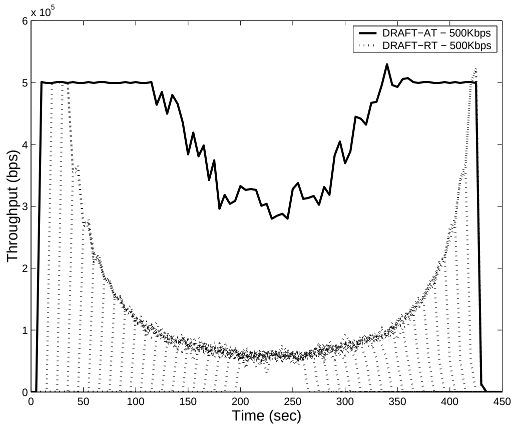
$$\theta \quad \leq \quad \omega \cdot \frac{\lambda_e - \sum\limits_{\forall i \in AT, AD} \lambda[i]}{\sum\limits_{\forall j \in RT} \lambda[j]} \tag{4.1}$$

To demonstrate the effectiveness of θ, we consider a new scenario 1AT+19RT+05TT. Scenario 1AT+19RT+05TT is similar to scenario 1AT+19RT, except the value of θ is changed from 1.0 to 0.5. In scenario 1AT+19RT+05TT, Figure 4.4 (a) shows that the AT-MS receives as much throughput as it requires for the entire simulation. During $200 < t < 250$, the total throughput requirement in the WLAN is 10 Mbps. This result matches the calculation discussed previously. We believe that this level of aggregate throughput requirement of 10 Mbps in a 2 Mbps WLAN is extremely high and sufficient in most situations. However, without admission control, absolute throughput support cannot be guaranteed. Additional traffic can still be injected into the network, since there is no mechanism to prevent it from doing so. If the amount of throughput requirement from RT-MSs is higher than 10 Mbps (although unlikely), the ability to provide absolute throughput will deteriorate. The results in terms of aggregate throughput, MAC delay, jitter, collision rate, and backoff interval are also provided in Figure 4.4 (b) - Figure 4.4 (f). Due to the higher offered load, the MAC delay, jitter, collision rate, and backoff interval of 1AT+19RT scenario are higher than those of 10RT scenario. However, the aggregate throughput remains the same. Note that, the aggregate throughput of scenario 10RT drops sooner than that of scenario 1AT-19RT because there are fewer MSs in scenario 10RT than scenario 1AT+19RT. Since each MS remains active for the same 240s, the last MS of scenario 10RT becomes inactive sooner than the last MS of scenario 1AT+19RT. Next, we will evaluate the second method, namely DSG-RT, to maintain the ability to provide absolute throughput support by limiting offered load in a distributed manner.

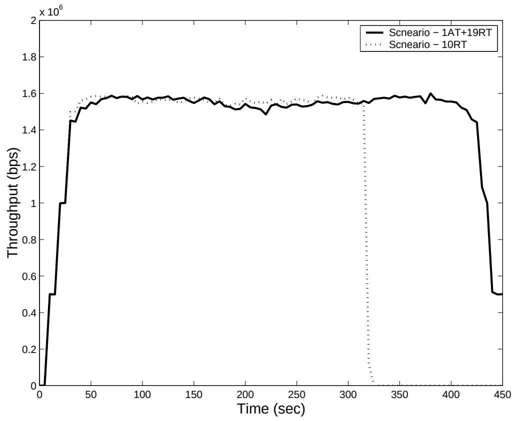
4.1.4 Distributed SafeGuard for Absolute Throughput Support in Relative Throughput Flows

The objective of DSG-RT is to limit the excessive traffic from RT-MSs to prevent deterioration of absolute QoS support. This objective is achieved by each RT-MS independently comparing the experienced throughput with the overload threshold. A new flow or RT-MS will not be allowed to continue if the overload threshold is violated. The overload threshold is the throughput level that can be calculated independently by each MS from the specified quantum rate, ω, and θ at each RT-MS, Eq. (3.22).

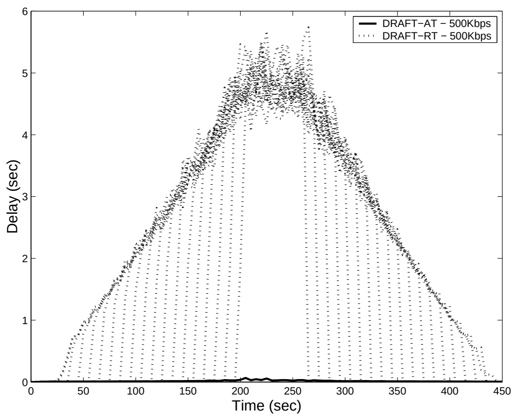
To demonstrate the performance of DSG-RT, we consider the scenario 1AT+19RT-DSG where θ is kept at 1.0. Without DSG-RT, the throughput of AT-MS begins to deteriorate at $t = 110s$ or when the 11^{th}RT-MS becomes active (Figure 4.3). With DSG-RT, absolute throughput support can be maintained since additional traffic from RT-MSs after $t > 110s$ will not be allowed to continue (Figure 4.5). The experienced throughput of AT-MS remains equal to the specified throughput requirement while the experienced throughput of accepted RT-MSs is proportional according to its weight. The overload threshold is set at 200 Kbps. If the experienced throughput of RT-MSs is smaller than the overload threshold, a new MS will not be allowed. The main advantage of DSG-RT is that the mechanism is *fully distributed* and does not require exchanging any information from other MSs. Each RT-MS monitors the experienced throughput and accepts or rejects a new flow independently.



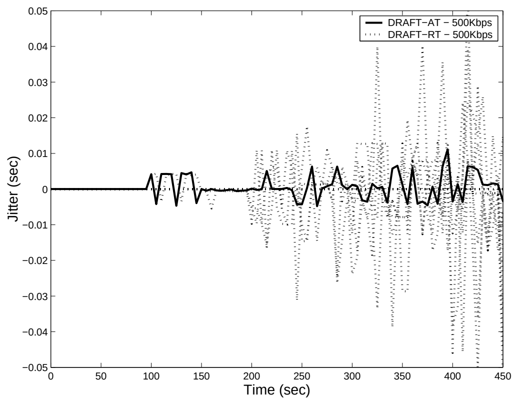
4.4 (a): Throughput



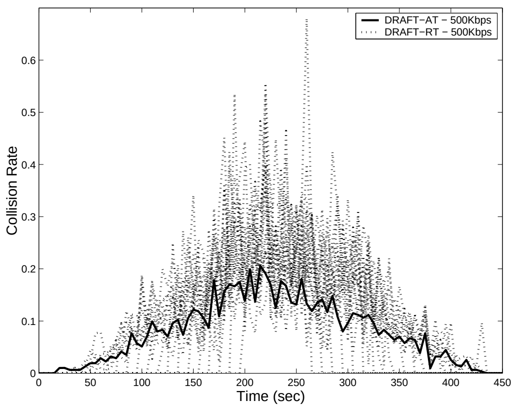
4.4 (b): Aggregate Throughput



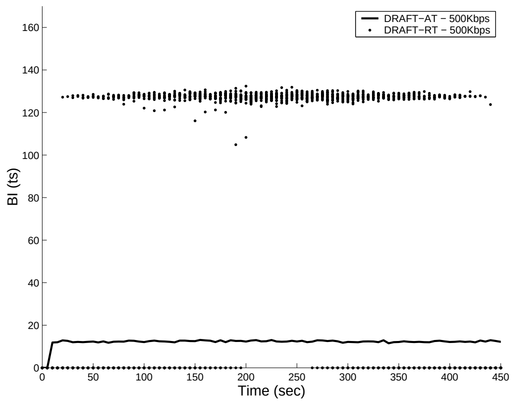
4.4 (c): MAC Delay



4.4 (d): Jitter



4.4 (e): Collision Rate



4.4 (f): Backoff Interval

Figure 4.4: Absolute Throughput Support with 1 AT-MS & 19 RT-MSs @ 500 Kbps

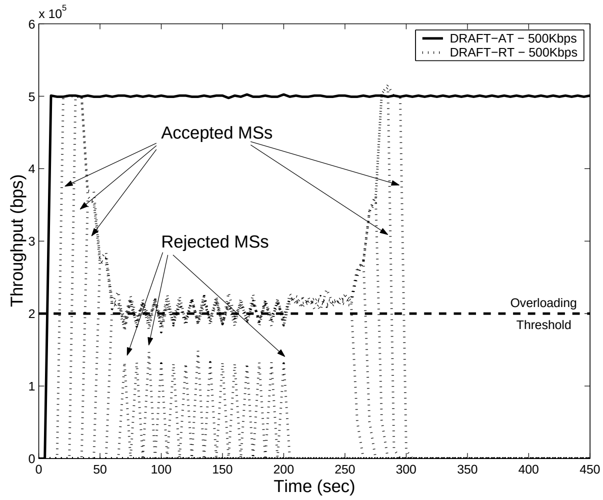
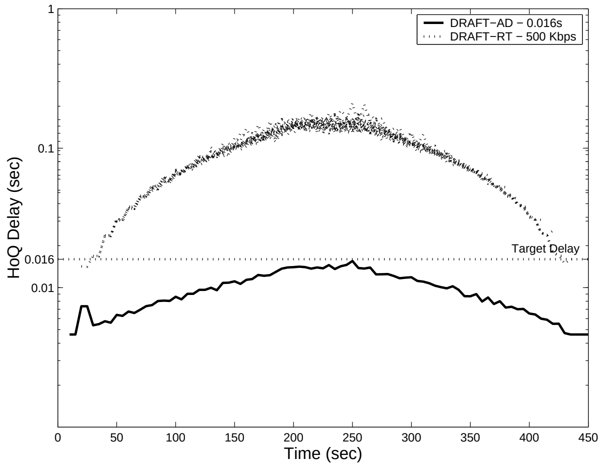


Figure 4.5: Absolute Throughput Support with DSG-RT, 1 AT-MS & 19 RT-MSs

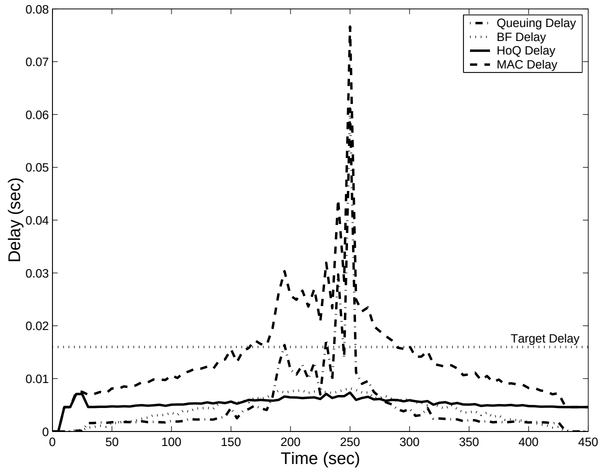
4.1.5 Providing Absolute Delay Support

In this section, we will demonstrate the ability of DRAFT+D to provide absolute delay support. The objective of absolute delay support is to provide an experienced delay (on average) within the specified target delay to the HoQ packet. To the best of our knowledge, there is no other mechanism has achieved this objective in a distributed manner. Only a few proposals shed some light on better-than-best-effort delay support, as discussed in Chapter 2.

To demonstrate absolute delay support, we consider the scenario 1AD+19RT which is composed of 1 Mobile Station with Absolute Delay Requirement (AD-MS) requiring 500 Kbps with 0.016s target delay and 19 RT-MSs. The throughput requirement of RT-MSs is also 500 Kbps. Therefore, the total throughput requirement is 10 Mbps in a 2 Mbps WLAN. The target delay of 0.016s represents the worst case target delay requirement for 500 Kbps with a maximum packet size of 1000 bytes. The reason is that 500 Kbps is the least amount of capacity needed to support a 0.016s target delay. Figure 4.6 (a) shows that the HoQ delay of AD-MS is within the specified target delay throughout the simulation even though the network is extremely saturated and total throughput requirement is 10 Mbps.



4.6 (a): HoQ Delay of AD-MS and RD-MSs



4.6 (b): Decomposition of MAC Delay of AD-MS

Figure 4.6: Delay Support with 0.02s Target Delay, 1 AD-MS & 19 RT-MSs

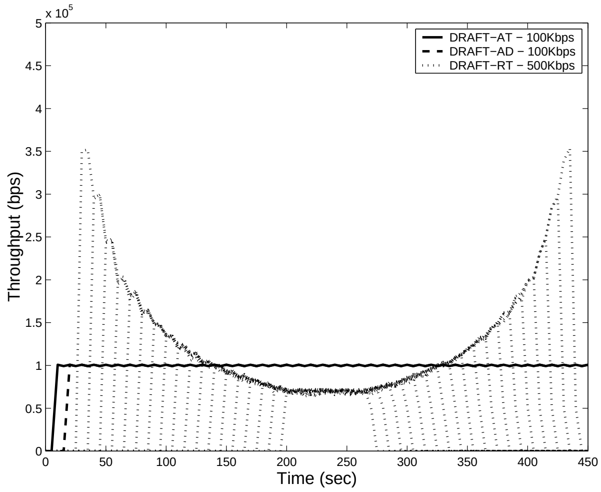
We can decompose the MAC delay into: a) queuing delay, b) backoff delay, and c) HoQ delay. Figure 4.6 (b) shows that the queuing delay increases significantly while the backoff delay increases slightly as a function of offered load. On the contrary, the HoQ delay stays relatively constant throughout the simulation as the load increases. Depending on several factors such as burstiness of input traffic, network load, etc, the MAC delay can become higher than the target delay. The reason is that the MAC delay ($D*$) is not bounded by the target delay. Rather, it is approximately bounded by Eq. 4.2, adapted from [95]. Here, σ is the size of token bucket, L_{max} is the maximum packet size, and λ is the quantum rate.

$$D* \quad \leq \quad \frac{\sigma + L_{max}}{\lambda} \tag{4.2}$$

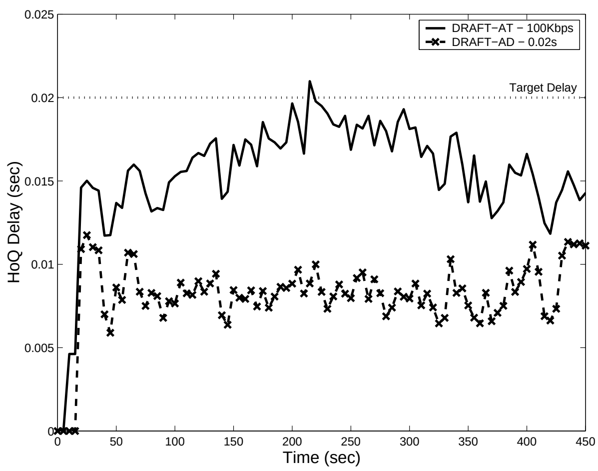
The absolute delay support of DRAFT+D is designed to provide a specific target delay for the head-of-queue packet; however it does not provide assurance on the MAC delay. Absolute delay support is a step towards delay support and could lead to provision of MAC delay support at the price of amount of traffic from AT-MSs and AD-MSs (via Eq. 4.2). However, research on this topic is beyond the scope of this study. We will present comprehensive results, e.g., multiple delay requirements, aperiodic input traffic, large numbers of MSs, later in Section 5.3.

4.1.6 WLAN with Heterogeneous MSs

In practice, a WLAN may consist of MSs with different QoS requirements. In this section, we will demonstrate the performance of DRAFT+D when the network consists of MSs requiring both throughput and delay requirements, we consider a scenario (1AT+1AD+18RT) where there are 1 AT-MSs requiring 100 Kbps, and 1 AD-MSs requiring 100 Kbps with 0.02s target delay, and 18 RT-MSs requiring 500 Kbps. The translation of target delay into throughput makes the required quantum rate of AD-MS equals to 400 Kbps. The result shows that both AT-MS and AD-MS receive the same experienced throughput that is as much as they are needed, as shown in Figure 4.7 (a). The HoQ delay of AD-MS is smaller than that of AT-MSs and well below the specific target delay for the entire simulation, as shown in Figure 4.7 (b). This result suggests that throughput and delay can be supported simultaneously without any degradation or interference from one another.



4.7 (a): Throughput



4.7 (b): HoQ Delay of AT-MS and AD-MS

Figure 4.7: Performance in a WLAN with 1 AT-MS, 1 AD-MS, & 18 RT-MSs

So far, we have demonstrated the ability of DRAFT+D to provide QoS support in terms of: a) relative throughput, b) absolute throughput, and c) absolute delay. The initial results show that DRAFT+D can provide excellent performance with and without safeguard mechanism. The performance of throughput support surpasses any other QoS mechanisms while the delay support and safeguard mechanism are unique features of DRAFT+D. In the following sections, we will show the performance of other selected QoS mechanisms, i.e., Distributed Weighted Fair Queuing (DWFQ), DFS, and EDCA. We will end this chapter with an evaluation of sensitivity analysis and parameter tuning.

4.2 COMPARATIVE SIMULATION RESULTS

In Section 4.1.1-4.1.3, we demonstrated the ability of DRAFT+D to provide absolute throughput and relative throughput support. A question one might ask is how well other state-of-the-art QoS mechanisms perform in comparison with DRAFT+D. In this chapter, we will demonstrate the performance of three state-of-the-art mechanisms: one is based on priority mechanism and the other two are based on fair-scheduling mechanism.

For priority-based mechanisms, EDCA is chosen since it has been receiving a lot of attention from the research community. EDCA is a priority-based mechanism which could support both requirements. However, the supported circumstances are highly dependent on the number MSs, the total load, the configuration of IFS, CW_{min} and CW_{max} of each traffic class [73]. For fair-scheduling based mechanisms, DWFQ and DFS (as discussed in Chapter 2) are chosen since both of them are designed to provide fair throughput support, comparable to the relative throughput support of DRAFT+D.

We consider the 1AT-19RT scenario where there are 1 AT-MS and 19 RT-MSs, each with throughput requirement of 500 Kbps as a base line for comparison. For EDCA scenario, AT-MS is configured with $AIFS = 1$, $CW_{min} = 16$ $CW_{max} = 32$ while the RT-MSs is configured with $AIFS = 2$, $CW_{min} = 32$, and $CW_{max} = 1024$. We give EDCA the advantage of configuration such that the $AIFS$ of AT-MS is smaller than that RT-MSs. Moreover, CW_{min} and CW_{max} of AT-MS and RT-MSs are completely separated. The weights of AT-MS and RT-MSs in DWFQ scenario is 2 and 1, respectively, while the weights of AT-MS and RT-MSs in DFS are is 0.05 ($\frac{1}{20}$) since they require the same throughput of 500 Kbps.

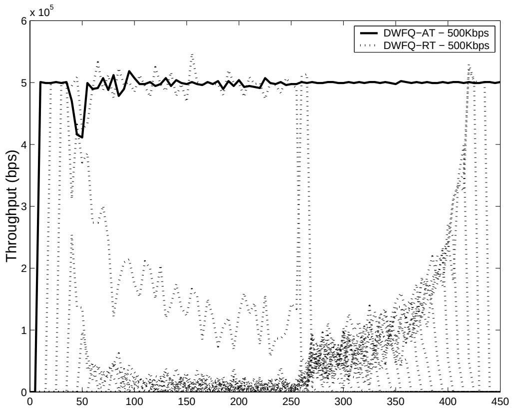
Figure 4.8 (a) shows the performance of DWFQ mechanism. Although DWFQ is supposed to provide fair bandwidth allocation, the fairness performance is not as good as expected. According to Figure 4.8 (a), DWFQ appears to work in favor of existing MSs or currently active MSs. Therefore, fairness among RT-MSs is significantly compromised. It is important to point out that DWFQ may seem to provide absolute throughput support; in fact, DWFQ does not. This result is due to the fact that DWFQ works in favor of existing DWFQ and the AT-MS is the first MS to transmit. To clarify this obscurity, we re-configured the simulation so that the AT-MS started at $t = 55$. This hypothesis is confirmed (that DWFQ cannot provide absolute throughput support) from the new result, as shown in Figure 4.8 (b). The performance of DWFQ based on MAC Delay, collision rate, Backoff Interval, and aggregate throughput is also worse than that of DRAFT+D. Figure 4.9 shows that DFS can provide excellent fairness. However, DFS is unable to provide the absolute throughput requirement since it was designed to support only relative throughput requirement. Figure 4.10 shows that EDCA can provide absolute throughput and relative throughput under relatively saturated conditions, e.g, under 3 Mbps of total throughput requirement. This amount of throughput requirement is significantly lower than the amount of throughput requirement that DRAFT+D can support, i.e., 10 Mbps. Moreover, the variation of experienced throughput of RT-MSs is larger than that of DRAFT+D.

In the next sections, we will investigated the operating ranges of seven parameters that are important to the overall performance of DRAFT+D. The first three parameters, i.e., κ, ω, and θ, directly influence the ability to provide relative and absolute QoS support. The last four parameters, i.e., $N1$, $N2$, β, and Ξ[1], influence the performance of DSG-RT.

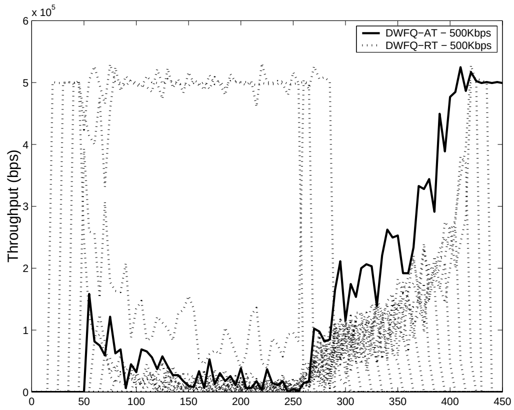
4.3 SENSITIVITY ANALYSIS AND PARAMETERS TUNING

In this section, the effect of seven parameters, i.e., κ, ω, θ, $N1$, $N2$, β and Ξ, will be investigated. The results showed that the right set of these parameters can be calculated and assigned relatively easily. Wide ranges of values of these parameters are available to provide a good and predictable performance. In short, the values of κ around 6, the values of ω between 1 and 7, the values of θ from 1 to 0.16, the values of $N1$ larger than 30, the values of $N2$ larger than 10, the values of β higher than 1.0, and any value of Ξ, are found to be within a good range of operation. The details of tuning of each parameter, the rationales, and simulation results will be presented next.

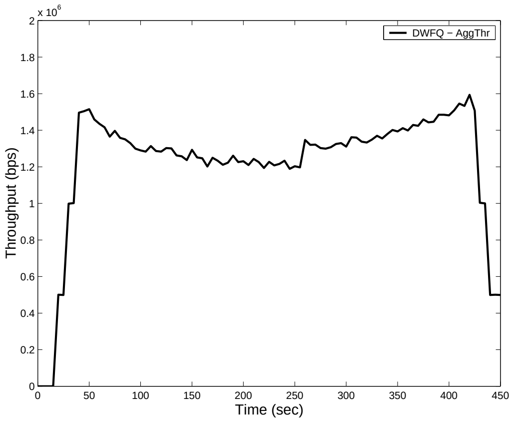
[1]the number of simultaneously probing MSs.



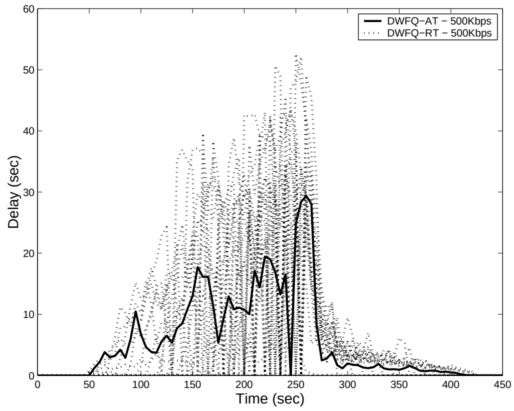
4.8 (a): Throughput, AT-MS starts at 10s



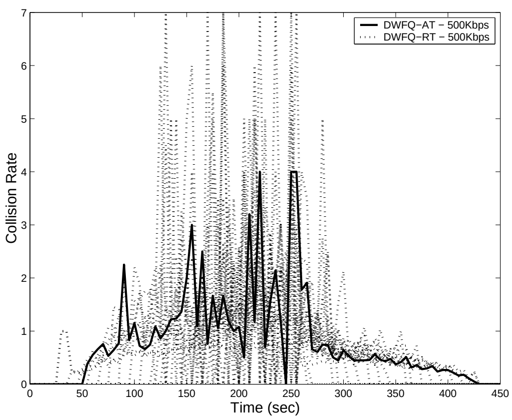
4.8 (b): Throughput, AT-MS starts at 55s



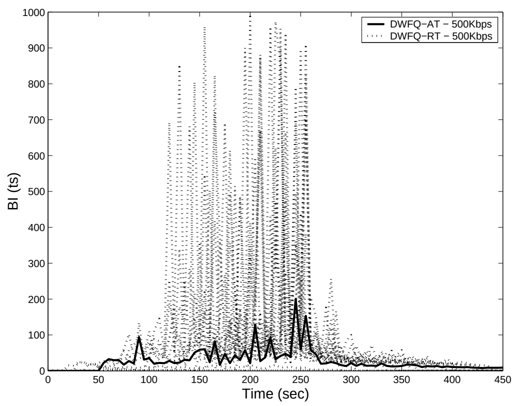
4.8 (c): Aggregate Throughput



4.8 (d): MAC Delay

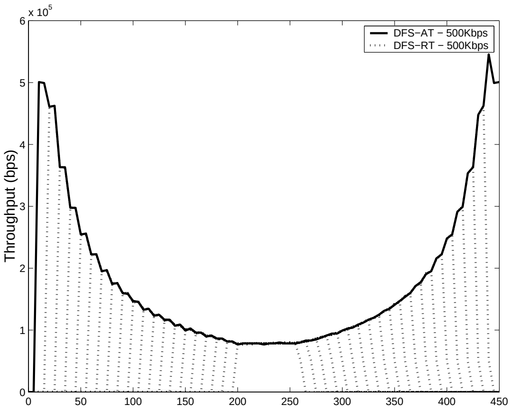


4.8 (e): Collision Rate

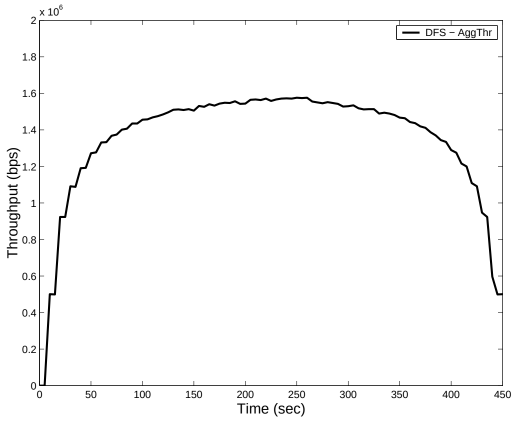


4.8 (f): Backoff Interval

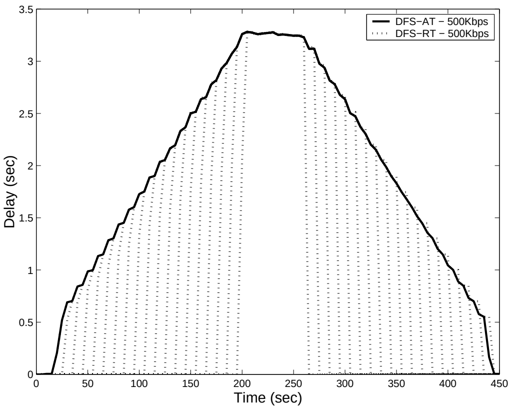
Figure 4.8: Performance of DWFQ in providing Relative Throughput Support



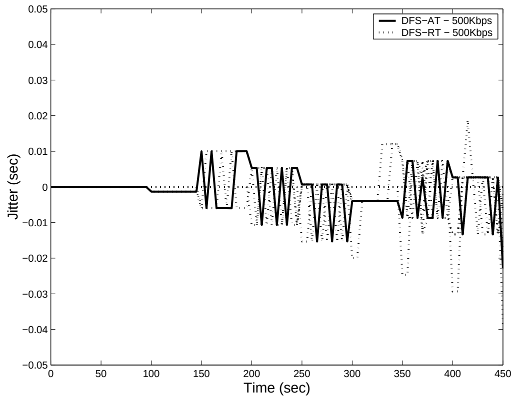
4.9 (a): Throughput



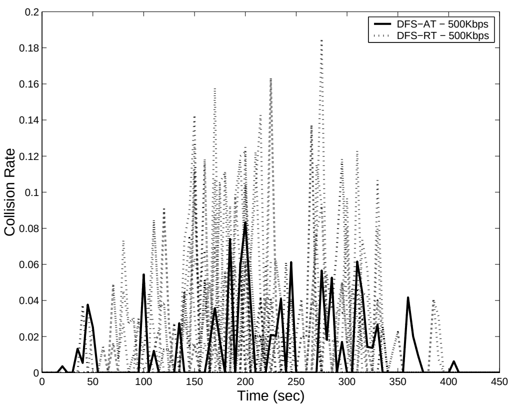
4.9 (b): Aggregate Throughput



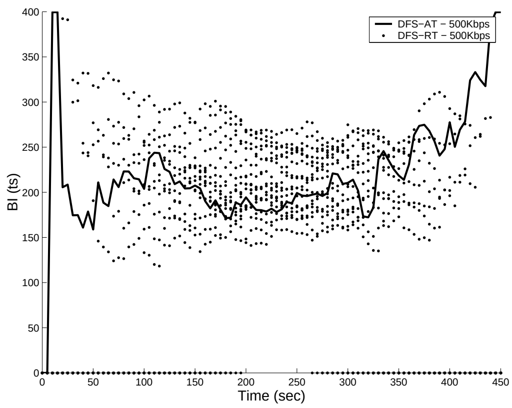
4.9 (c): MAC Delay



4.9 (d): Jitter

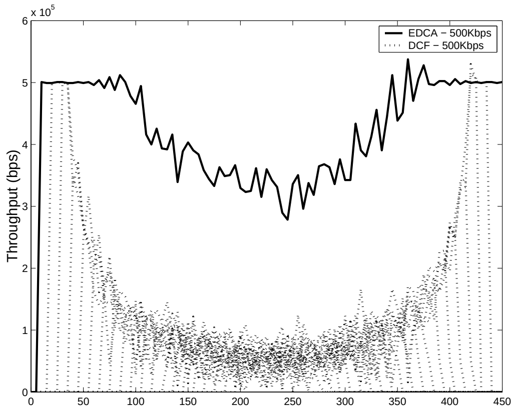


4.9 (e): Collision Rate

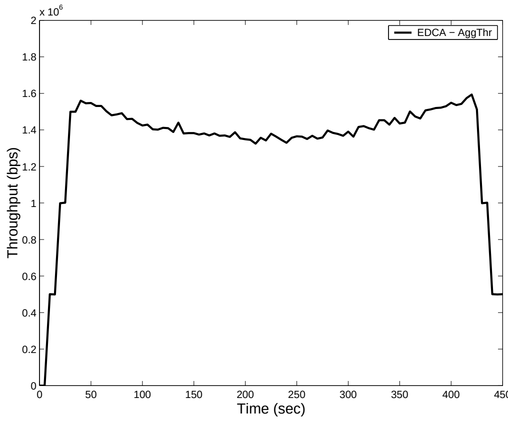


4.9 (f): Backoff Interval

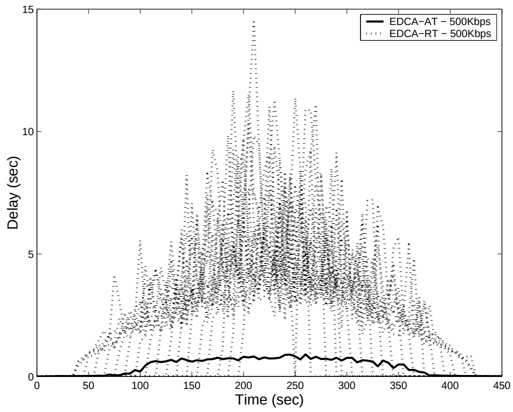
Figure 4.9: Performance of DFS in Providing Throughput Support



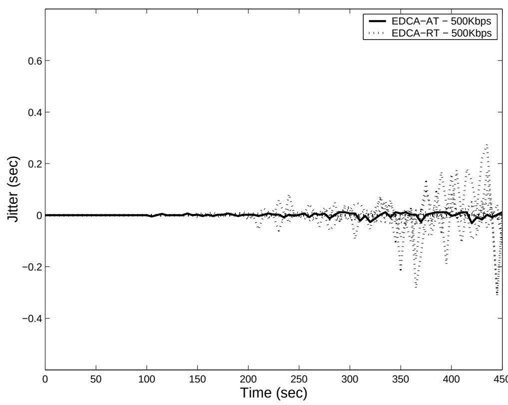
4.10 (a): Throughput



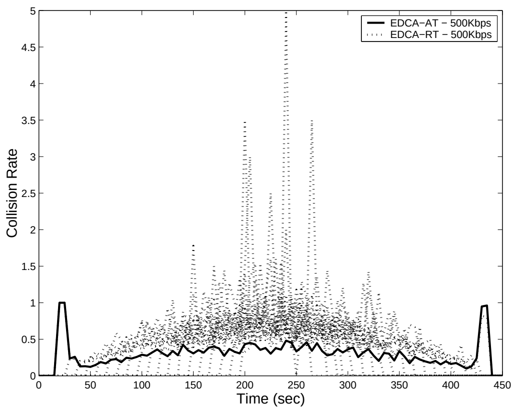
4.10 (b): Aggregate Throughput



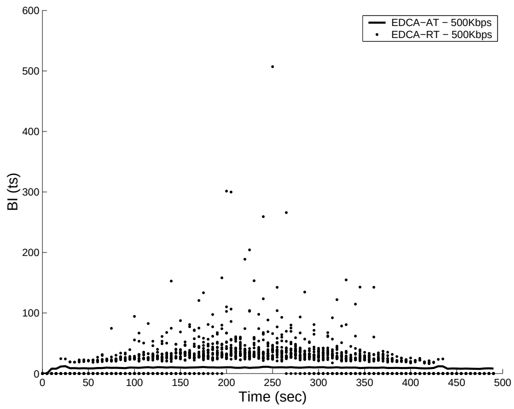
4.10 (c): MAC Delay



4.10 (d): Jitter



4.10 (e): Collision Rate



4.10 (f): Backoff Interval

Figure 4.10: Performance of EDCA in Providing Throughput Support

Table 4.2: Comparison of $CW_{center}[i]$ for Different Values of κ

κ	$CW_{center}[64Kbps]$	$CW[64Kbps]$	$CW_{center}[200Kbps]$	$CW[200Kbps]$
1	32	172	10	55
2	62	172	20	55
3	125	172	40	55
4*	250	172	80	55
5	500	172	160	55
6	1000	172	310	55

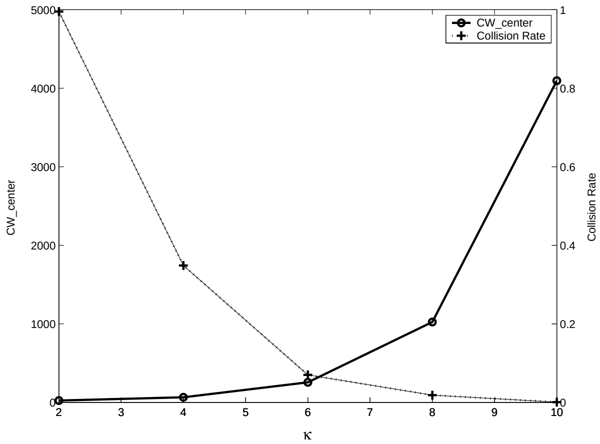
4.3.1 The Effect of Parameter κ

Parameter κ is a constant used in the calculation of average value of BI or CW_{center}, as shown in Eq. 4.3. The value of κ directly influences the calculated value of CW_{center}, which, in turn, influences the value of $BI[i]$. A higher value of κ results in a higher value of CW_{center} and a smaller probability of collision. However, the higher value of CW_{center} may result in a smaller effective channel capacity due to the unnecessary waiting time.

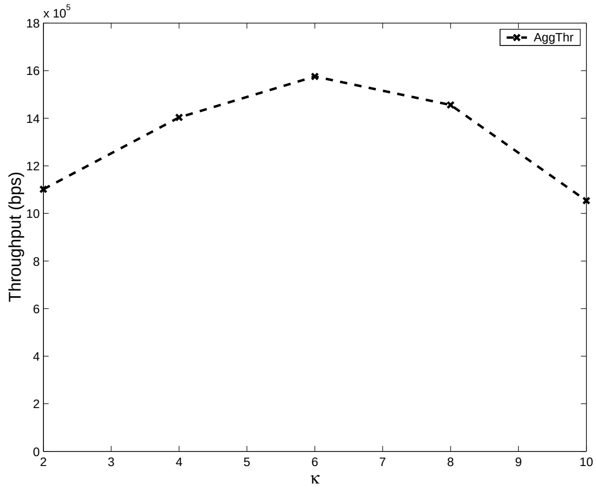
$$CW_{center}[i] = \frac{2^{\kappa}}{\phi[i]} \tag{4.3}$$

In practice, CW_{center} needs to be chosen large enough to accommodate the corresponding Contention Window ($CW[i]$). Table 4.2 shows the values of $CW_{center}[i]$ for different values of κ, i.e., 1) $\lambda[i] = 64Kbps$ and $CW[1] = 172$, and 2) $\lambda[2] = 200Kbps$ and $CW[2] = 55$. We can see from Table 4.2 that if $\kappa < 4$, the $CW_{center}[i]$ is smaller than the corresponding $CW[i]$; therefore these values of κ are not appropriate. Too small of a value of κ can result in unnecessary collisions, poor aggregate throughput, and increased variations of throughput and delay. A sufficiently high value of κ will provide enough room to avoid unnecessary collisions.

Choosing an appropriate value of κ is a tradeoff between probability of collision and the maximum effective channel capacity. To study the influence of κ, we consider the 1AT+19RT scenario, change the value of κ from 2 to 10, and measure the value of CW_{center}, collision rate, and aggregate throughput. As expected, the result in Figure 4.11 (a) shows that as the value of CW_{center} increases as a function of κ while the collision rate decreases as a function of κ. This result confirms our intuition about the effect of κ on the probability of collision. That is, the higher the value of κ, the smaller is the probability of collision.



4.11 (a): CW_{center} vs Collision Rate



4.11 (b): Aggregate Throughput

Figure 4.11: The Effect of Parameter κ

Figure 4.11 (b) shows the aggregate throughput as the value of κ changed from 2 to 10, with a step of two. A significantly low value of κ, e.g. $\kappa = 2$, results in high collision rate and deterioration of the effective channel capacity. On the contrary, a significantly high value of κ, e.g. $\kappa = 10$, results in excessive overheads (in terms of waiting time) which, in turn, decreases the effective channel capacity as well. According to Figure 4.11 (b), we see that an appropriate value of κ, i.e. around 6, provides good balance between waiting time and probability of collision. The value of κ around 6 provides the *highest and relatively constant* effective channel capacity throughput the entire simulation. This property – relatively constant effective channel capacity – indicates the effectiveness of DRAFT+D in calculating appropriate value of BI and regulating traffic in extremely saturated condition. The effective channel capacity remained relatively constant across simulation scenarios previously presented where the number of simulation parameters such as number of MSs, loads, traffic types, and WLAN's parameters were changed. Constant effective channel capacity is an important and unique properties of DRAFT+D.

4.3.2 The Effect of Parameter ω

Parameter ω is introduced to support absolute throughput. We will look at parameter ω from two perspectives: a) the effect of ω on supported amount of traffic from RT-MSs and b) the effect of ω on the value of CW_{center}. These two perspectives represent the lower bound and the upper bound of the appropriate value of ω.

From the first perspective, we look back to the basic for ω. Parameter ω is an escalating factor used to increase the fair share of AT-MSs from the normal weight. According to this definition, the appropriate value of ω can be calculated by the ratio between the deescalated aggregate quantum rate from RT-MSs and the remaining bandwidth, as shown in Eq. (4.4). In other words, an appropriate value of ω represents the ratio of supported amount of traffic from RT-MSs above the actual remaining bandwidth. The higher the value of ω, the higher is amount of traffic from RT-MSs that can be supported.

$$\omega \quad \geq \quad \frac{\theta \cdot \sum\limits_{\forall j \in RT} \lambda[j]}{\lambda_e - \sum\limits_{\forall i \in AT,AD} \lambda[i]} \tag{4.4}$$

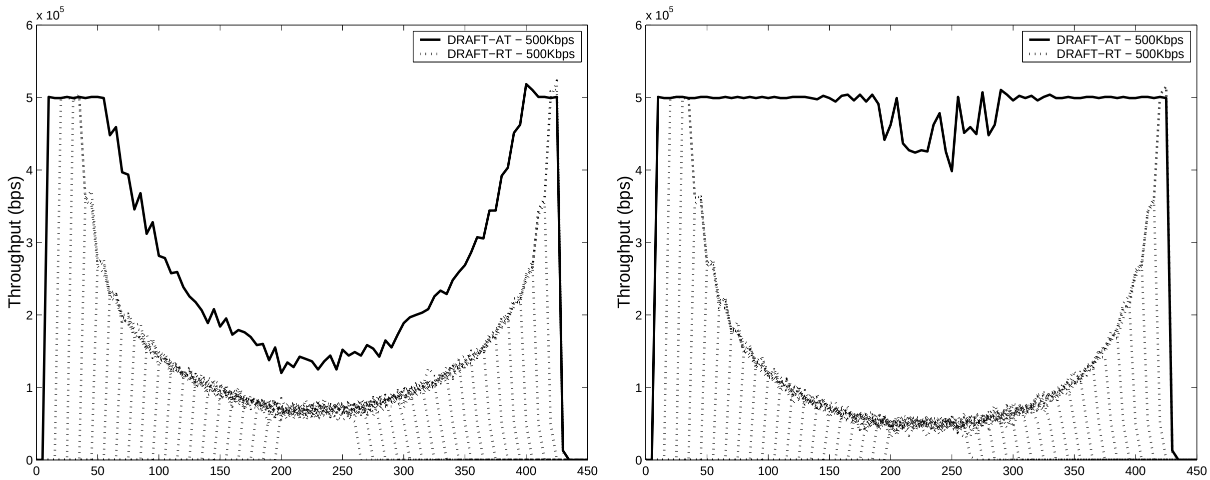
Let $\theta = 1$, then the increased fair share of channel capacity is achieved by decreasing the value of CW_{center}. Therefore, we look at the calculation of CW_{center} and the relationship between CW_{center} and ω in the second perspective. According to the calculation of CW_{center}, see Eq. (3.8), and calculation of weight of AT-MSs, see Eq. (3.6), the relationship between CW_{center} and ω can be derived and expressed in Eq. (4.5).

$$CW_{center}[i] = \frac{2^k \cdot R}{\omega \cdot \lambda[i]} \tag{4.5}$$

According to Eq. (4.5), the higher the increased fair share of channel capacity, the smaller is the value of CW_{center}. However, the fair share cannot be increased indefinitely. An important question is what and how to calculate the maximum value of ω is. We know that increasing the value of ω increases the value of calculated weight. If we assume that ω is an integer number. The condition where ω is no longer valid is when the higher value of ω does not result in an integer difference between the original CW_{center} and the new CW_{center}. According to Eq. (4.5), we can calculate the value of ω that makes the difference between the CW_center smaller than 1 as shown in Eq. (4.6), assuming $\lambda[i] = 500$ Kbps, $\kappa = 5$, and $R = 1$ Mbps.

$$\begin{aligned}
CW_{center,\omega 1} - CW_{center,\omega 2} &\leq 1 \\
\frac{2^5}{0.5 \cdot \omega_1} - \frac{2^5}{0.5 \cdot \omega_2} &\leq 1 \\
\omega_1^2 + \omega_1 - 64 &\geq 0 \\
\omega &\leq 7.52 \qquad (4.6)
\end{aligned}$$

According to Eq. (4.6), the maximum value of ω for $\lambda = 500$ Kbps is less than or equal to 7. This means, the values of ω between 1 and 7 is valid and able to provide absolute throughput support according to Eq. (4.4). However, Eq. 4.4 and absolute throughput support do not hold for the value of ω above 7. To confirm this analysis, we consider the 1AT+19RT scenario and use the values of ω: 2 and 10. Figure 4.12 shows that the results in both cases match our analysis. When ω was set to 2, Figure. 4.12 (a) shows that absolute throughput is supported as long as the amount of throughput requirement from RT-MSs is less 2 Mbps, i.e., twice as much as the remaining bandwidth as calculated from Eq. (4.4). When ω was set to 10, Figure 4.12 (b) shows that the ability to support absolute throughput is not maintained as our analysis predicted.



4.12 (a): $\omega = 2$

4.12 (b): $\omega = 10$

Figure 4.12: The Effect of ω to Throughput Support

The significant finding is the bound of ω. We found that there exists a range of values of ω that can be used to provide absolute throughput support. The absolute throughput is fully and predictably supported within this range. However, if ω is set outside this range, the absolute throughput support is longer abided by our analysis. Therefore, it is important to calculate this range in a priori to ensure to absolute throughput support. The range of valid ω can be simply calculated via by Eq. (4.6). Next, we will discuss the effect of parameter θ in maintaining absolute throughput support.

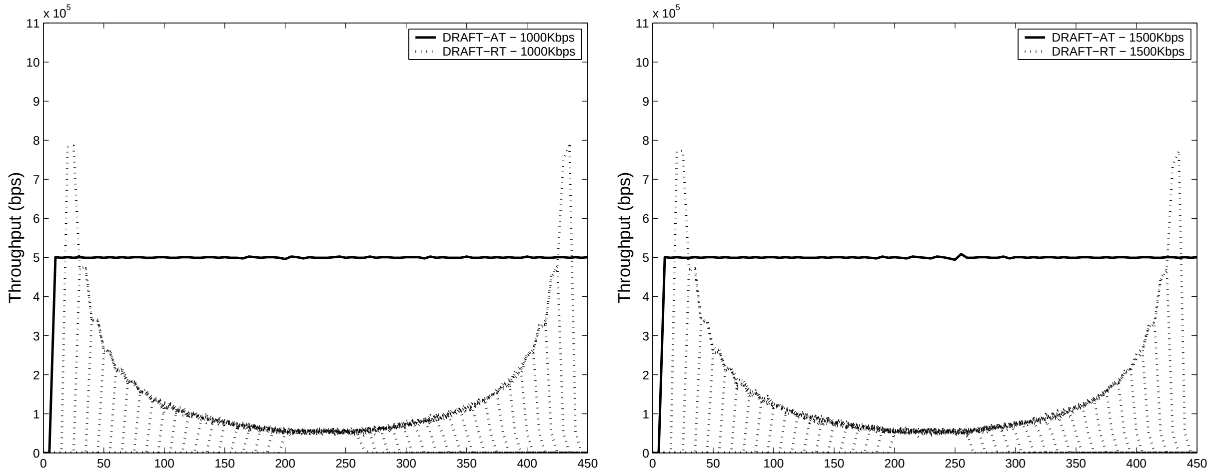
4.3.3 The Effect of Parameter θ

Parameter θ is a factor introduced to work in conjunction with ω to provide absolute throughput support. While ω increases, the fair share of traffic of absolute throughput class, θ decreases the fair share of traffic of relative throughput class. Because ω can be assigned within a certain range, another way to increase the supported offered load is to adjust θ. That is, we can decrease θ to increase the amount of supported offered load in a WLAN while maintaining absolute throughput support.

To demonstrate the effect of θ, we consider 1AT+19RT scenario and increase the offered load of RT-MSs in two cases: a) from 9.5 Mbps (19×500 Kbps) to 19 Mbps (19×1000 Kbps) and b) from 9.5 Mbps to 28.5 Mbps (19×1500 Kbps). In order to support the increased offered loads, the parameter θ needs to be adjusted accordingly. The required value of θ for these offered loads can be calculated by Eq. (3.18), i.e., 0.25 and 0.16, respectively.

Figure 4.13 shows the absolute throughput support can be maintained as the maximum offered load increases by decreasing θ by the same factor. More specifically, Figure 4.13 (a) shows that, by decreasing θ from 0.5 to 0.25, the maximum supported offered load increases from 9.5 Mbps to 19 Mbps and the absolute throughput support is still maintained. Similarly, Figure 4.13 (b) shows that, by decreasing θ from 0.25 to 0.16, the maximum supported offered load increases from 19 Mbps to 28.5 Mbps. Again, 19 Mbps and 28.5 Mbps of offered loads are in a 2 Mbps WLAN with effective channel capacity of only 1.5 Mbps. This result suggests the robustness of DRAFT+D in providing absolute throughput.

Now, let turn out attention to parameters that influence the performance of DSG-RT.



4.13 (a): $\theta = 0.25$

4.13 (b): $\theta = 0.16$

Figure 4.13: The Effect of θ on Throughput Support

The procedure of DSG-RT begins by sending out a number of packets at the desired rate. During this period, the experienced throughput is calculated. DSG-RT compares the experienced throughput with the overload threshold. The experienced throughput is calculated from the EWMA of instantaneous throughput. The overload threshold is calculated from the product of the safety factor, the overload ratio, and the throughput requirement. During the testing procedure, if the experienced throughput is smaller than the overload threshold, the new flow will be immediately stopped. Otherwise, a new flow can continue.

According to this procedure, three parameters are important to the performance of DSG-RT: 1) the number of packets during the transient period ($N1$), 2) the number of probing packets during the decision period ($N2$), and 3) the level of safety factor (β). The effect of the number of probing MSs (Ξ) that become active at the same time will be also evaluated next.

4.3.4 The Effect of the Number of Packets During the Transient Period

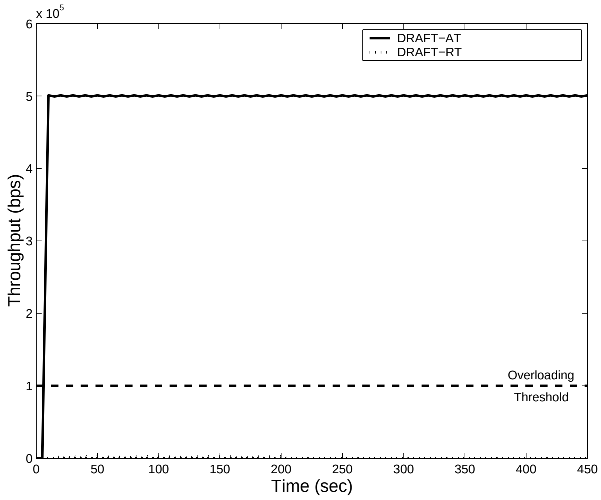
The experienced throughput of a new RT-MS often takes some time (or a number of packets - $N1$) to register until it can accurately reflect the true fair share. If DSG-RT begins the testing too early, a flow is likely to be stopped because the experienced throughput starts from zero. In this section, we will determine the number of packets during the transient period that are required for representative experienced throughput values. We consider scenarios (1-6) where there are 1 AT-MS and 19 RT-MSs employing DSG-RT. The throughput requirement, the time at which the MSs become active, and other simulation parameters remain the same as in 1AT+19RT scenario. The only difference is that RT-MSs in these scenarios are subjected to the test of DSG-RT.

Figure 4.14 shows that if the mechanism does not wait long enough ($N1 > 30$), 2 issues can arise. First, a new RT-MS will be stopped immediately (Figure 4.14 (a)) or prematurely (Figure 4.14 (b)). Second, a new RT-MS might be inaccurately accepted or rejected due to the immaturity of the estimate of experienced throughput, as shown in Figure 4.14 (c). In both cases, these issues impact the accuracy of the supported traffic from RT-MSs; however they do not have any negative impact on the AT-MSs. Figure 4.14 (d) shows that when DSG-RT waits sufficiently long ($N1 = 30$), the mechanism works accurately in accepting and rejecting a new RT-MS.

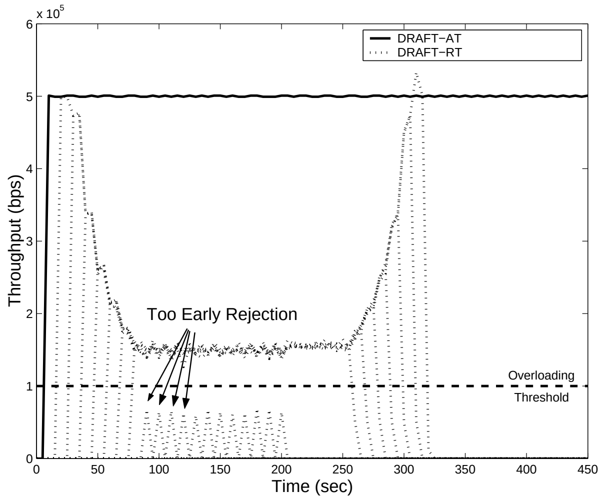
These results suggest that the estimate of experienced throughput needs some time to become mature. Since there is no significantly negative impact on waiting a little longer (Figure 4.15), a higher number of packets, e.g., 40 or 50 packets, is recommended. In this set of scenarios, we fixed the number of probing packets during the testing procedure and safety factor for overload threshold at 10 and 2, respectively. In the next sections, we will examine the effect of them.

4.3.5 The Effect of the Number of Probing Packets During a Decision Period

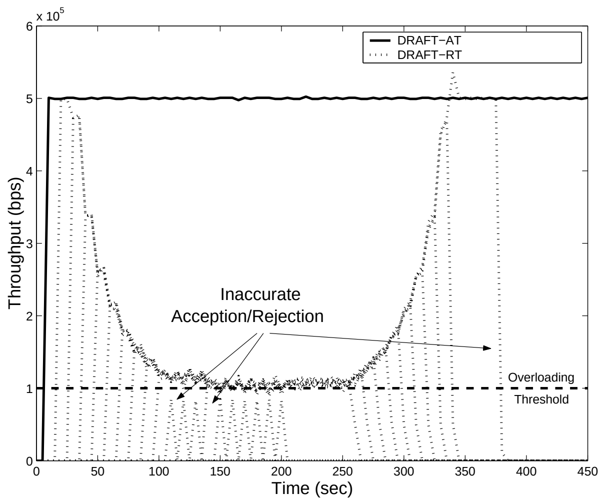
During a decision period, the experienced throughput is being compared with the overload threshold for every successful transmission. During this period, if the experienced throughput drops below the overload threshold, a new RT-MS will be stopped immediately. In this section, we will determine the number of necessary probing packets ($N2$) that can provide an accurate estimate of the experienced throughput.



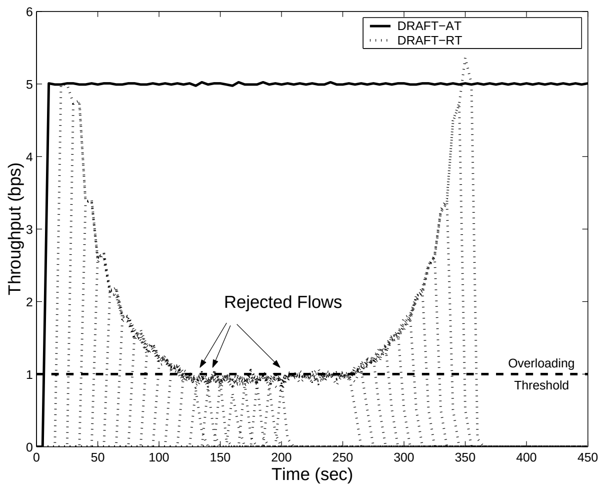
4.14 (a): $N1 = 0$



4.14 (b): $N1 = 10$



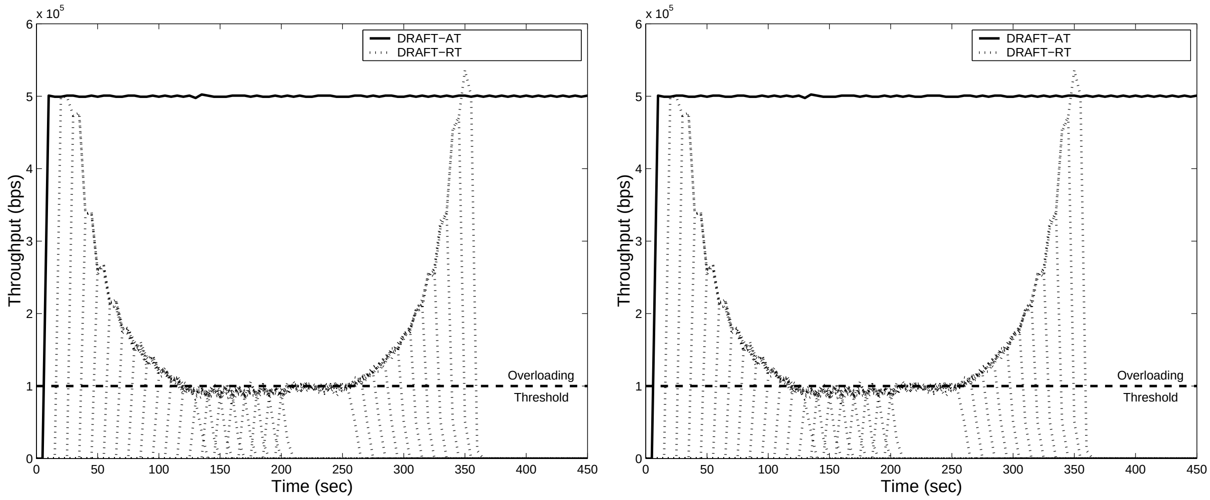
4.14 (c): $N1 = 20$



4.14 (d): $N1 = 30$

Figure 4.14: The Effect of the Number of Packets during the Transient Period

At the first glace, it might seem that a high number of probing packets may increase the accuracy of the throughput estimate. However, this is not always true. According to Figure 4.14, we see that the mechanism always errs on the conservative side. That is, the mechanism tends to stop flows a little earlier rather than later. This behavior is, in fact, desirable. If the mechanism already stops flows during the testing procedure with 10 testing packets, the higher number of packets will not yield much advantage at all. To prove this,

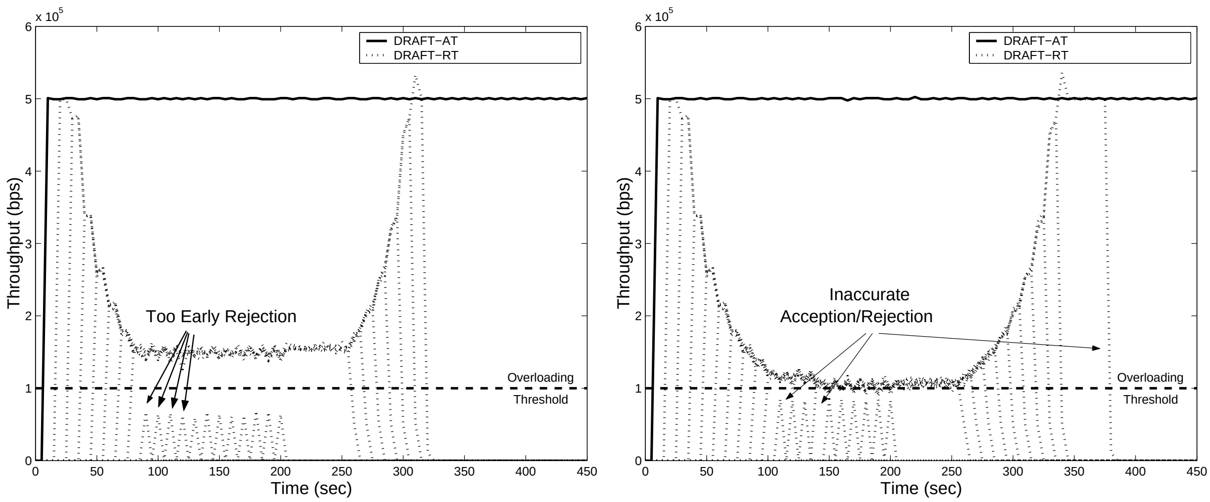


4.15 (a): $N1 = 40$

4.15 (b): $N1 = 50$

Figure 4.15: The Effect of the Number of Packet During the Transient Period

we increased the number of probing packets ($N2$) from 10 to 40 and kept the number of packets during the transient period ($N1$) to 10. Figure 4.16 shows that the result remain exactly the same as Figure 4.14 (b) and Figure 4.14 (c). This result matches our intuition.

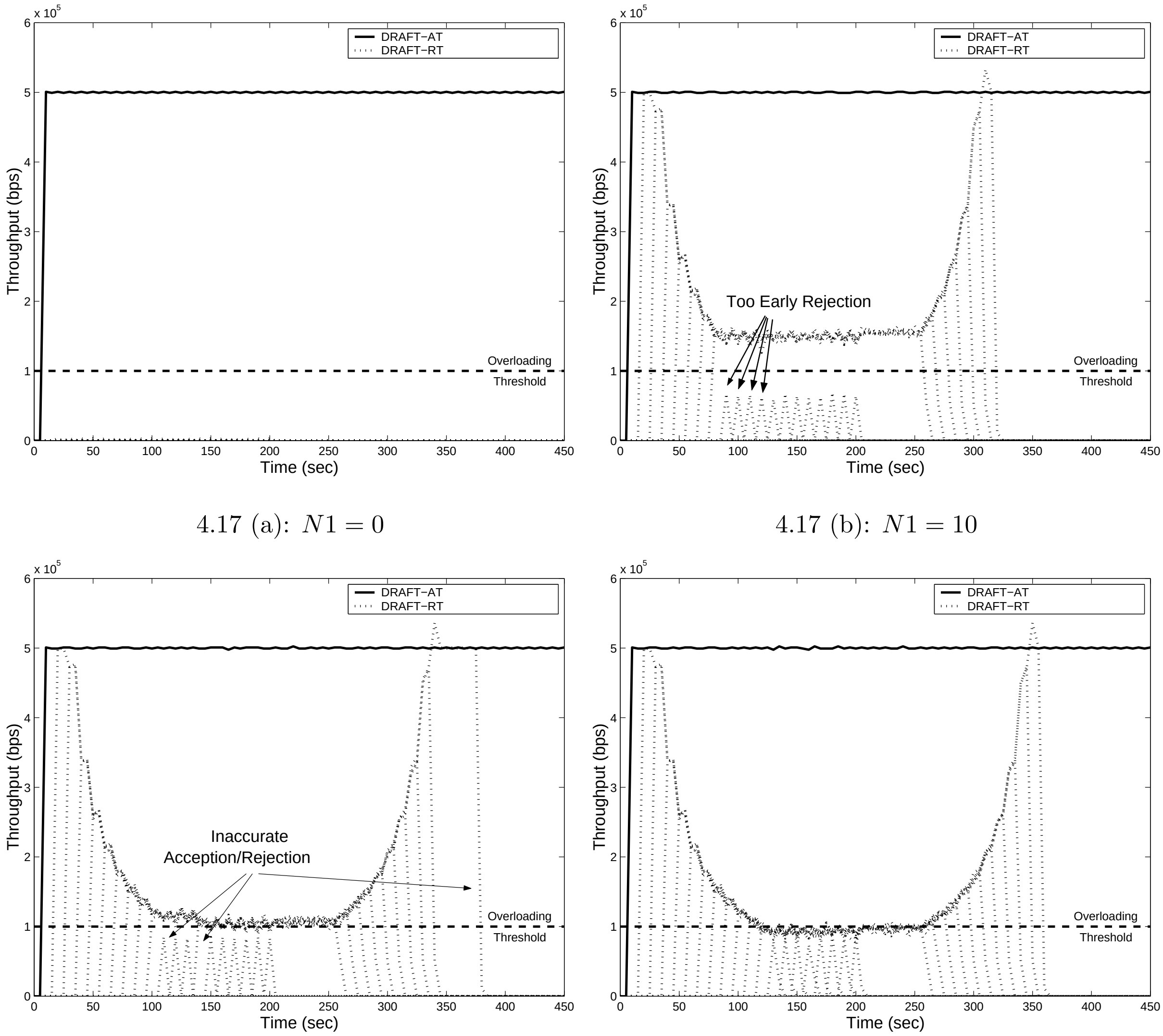


4.16 (a): $N1 = 10$

4.16 (b): $N1 = 20$

Figure 4.16: The Effect of the Number of Probing Packets, $N2 = 40$ packets

On the contrary, it is more interesting to examine whether a smaller number of probing packets can provide a good estimate of experienced throughput. Figure 4.17 shows that the performance of DSG-RT is exactly the same if 5 packets are used during the testing period. Based on these results, we conclude that once the experienced throughput is mature (e.g., at $N1 = 30$, only a smaller number of probing packets are necessary to provide a good estimate of current fair share and good performance of DSG-RT.



4.17 (a): $N1 = 0$

4.17 (b): $N1 = 10$

4.17 (c): $N1 = 20$

4.17 (d): $N1 = 30$

Figure 4.17: The Effect of the Number of Probing Packets, $N2 = 5$ packets

4.3.6 The Effect of Safety Factor

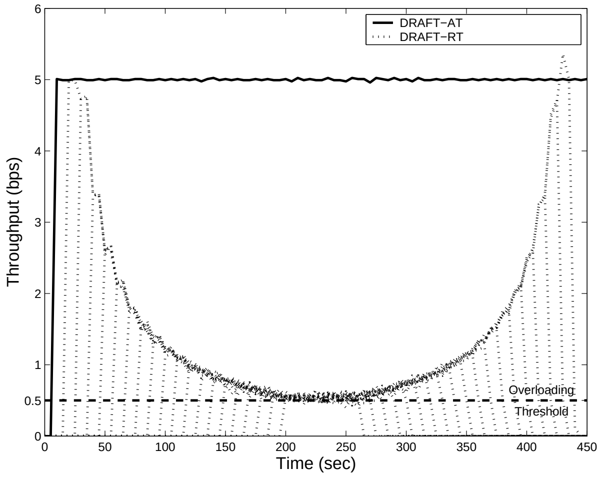
According to our analysis in Chapter 3, absolute QoS support can be maintained as long as the experienced throughput of relative QoS is higher than the overload threshold, see Eq. (3.22). However, in practice, the experienced throughput can fluctuate and its estimate may not be 100% accurate. To prevent deterioration of absolute QoS support, the original overload threshold is multiplied by a safety factor (β). In this section, we examine the effect of different values of the safety factor on experienced throughput and experienced delay.

Figure 4.18 shows that the safety factor does not have any impact on absolute throughput support. However, the safety factor represents a tradeoff between the number of accepted RT-MSs (or amount of traffic from RT-MSs) and the experienced throughput of RT-MSs. The higher the value of the safety factor, the smaller the number of accepted RT-MSs. With the same remaining bandwidth to RT-MSs, the smaller number of accepted RT-MSs results in a higher experienced throughput for each active RT-MS.

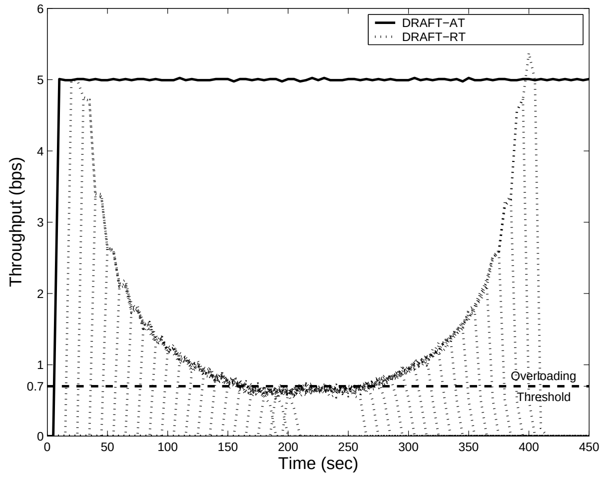
For experienced delay, the opposite effect is expected and witnessed. That is, the higher the value of the safety factor, the smaller is the experienced delay. The reason is that a higher value of safety factor results in a smaller amount of total throughput requirements, or offered load. According to Figure 4.19, the experienced delay of RT-MSs decreases by one eighth while the experienced delay of AT-MS decreases by one-half. Next, we will examine the effect of the number of probing MSs that become active at the same time.

4.3.7 The Effect of the Number of Probing MSs

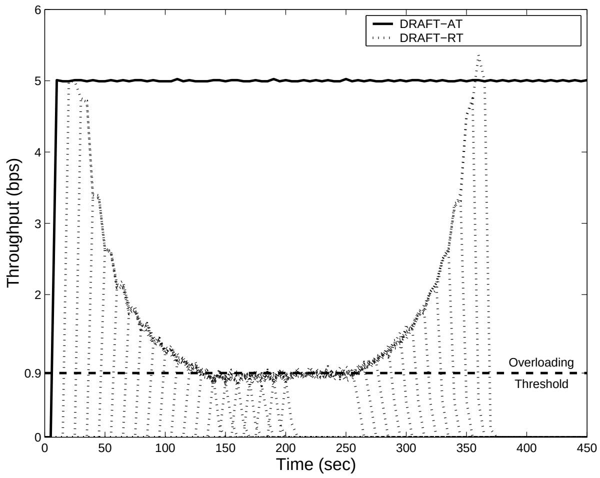
In the previous scenarios, each RT-MS is set to start sending out traffic in 10 seconds intervals. In this section, we will investigate the effect when more than one MSs starts to transmit traffic at the same time. Based on 1AT+19DSAC-RT scenario, there can be total of 11 RT-MSs accepted in the network. Figure 4.20 shows that the mechanism still performs well in this scenario. If the number of simultaneously active MSs is smaller than the maximum acceptable MSs, all the MSs will be accepted. However, if the number of simultaneously active MSs is higher than the maximum acceptable MSs, all the MSs will be rejected. These results are as expected and demonstrated the robustness of DSG-RT.



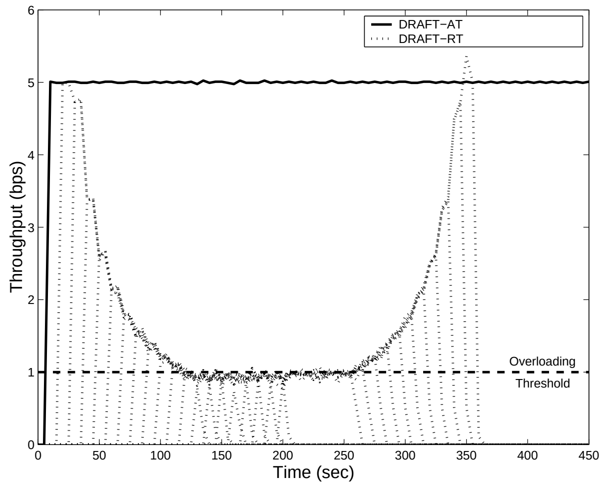
4.18 (a): $\beta = 1.0$



4.18 (b): $\beta = 1.4$



4.18 (c): $\beta = 1.8$

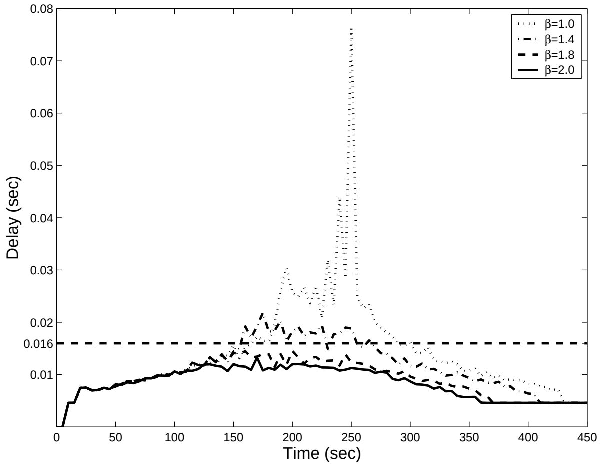


4.18 (d): $\beta = 2.0$

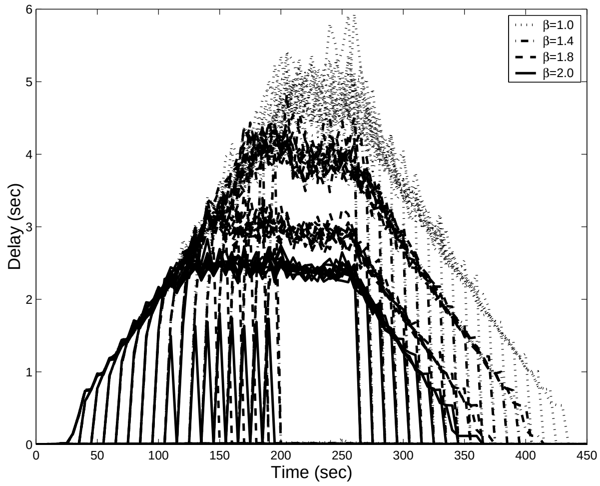
Figure 4.18: The Effect of the Safety Factor on the Experienced Throughput

4.4 CONCLUDING REMARKS

In conclusion, the simulation results confirmed that DRAFT+D can provide QoS support, i.e., a) relative throughput support, b) absolute throughput support, and c) absolute delay support, even in highly overloaded conditions. The variation of throughput and delay is small throughout the course of simulations. The safeguard mechanism can prevent excess



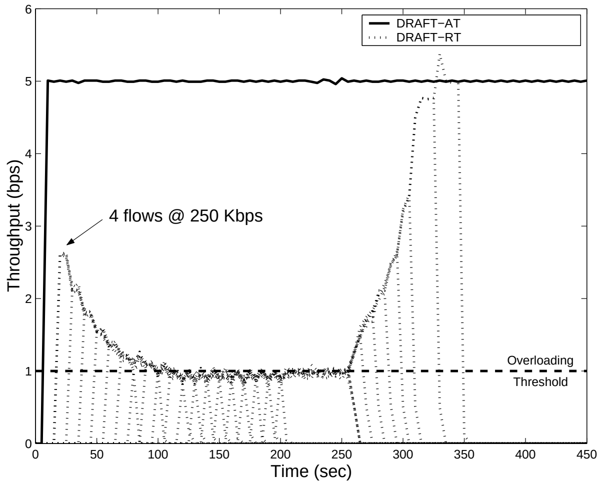
4.19 (a): AT-MSs



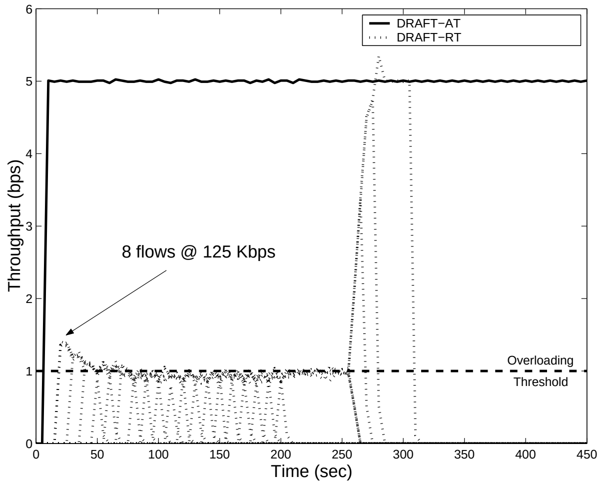
4.19 (b): RT-MSs

Figure 4.19: The Effect of the Safety Factor on the Experienced Delay

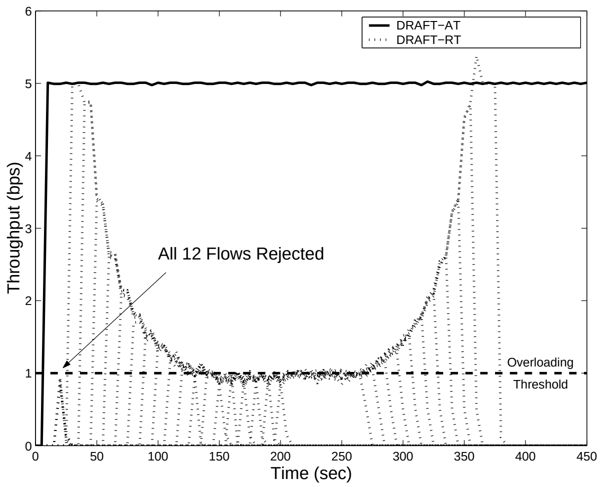
traffic from relative throughput class expected. Finally, relatively wide ranges of values of κ, ω, θ, $N1$, $N2$, β, and ξ can provide a good performance. The right set of parameters within DRAFT+D can be calculated relatively easily and in advance. In the next chapter, we will present performance of DRAFT+D in various scenarios where the effect of levels of throughput requirement, the number of levels of throughput requirement, the time at which MSs become active, number of active MSs, traffic type, mode of operation, the use of RTS/CTS, data rate, and fairness criteria will be evaluated.



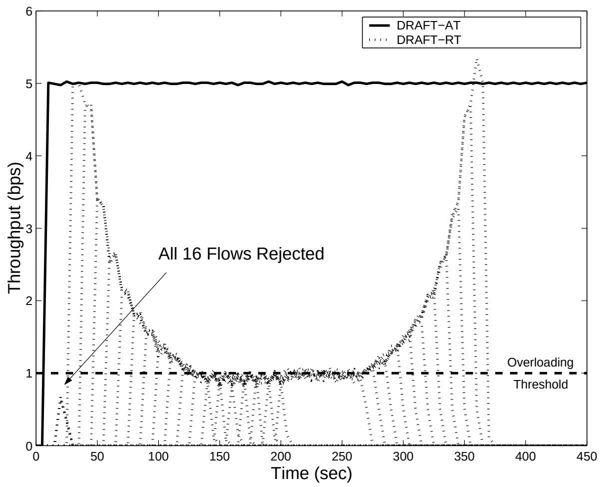
4.20 (a): $\Xi = 4$ Flows



4.20 (b): $\Xi = 8$ Flows



4.20 (c): $\Xi = 12$ Flows



4.20 (d): $\Xi = 16$ Flows

Figure 4.20: The Effect of Simultaneously Active MSs

5.0 COMPREHENSIVE PERFORMANCE EVALUATION

In this chapter, we will present comprehensive performance evaluation of DRAFT+D. Three main components of DRAFT+D, i.e., throughput support, delay support, and safeguard against excessive traffic support, will be evaluated in several scenarios. In these scenarios, performance of DRAFT+D with different levels of throughput requirement, the number of levels of throughput requirement, the time at which MSs become active, number of active MSs, traffic type, mode of operation, the use of RTS/CTS, data rate, and fairness criteria, will be evaluated. We will begin with a comprehensive evaluation of the throughput support, then follow with an evaluation of DSG-RT. We will end this chapter with a comprehensive evaluation of the delay support.

5.1 COMPREHENSIVE EVALUATION OF THROUGHPUT SUPPORT

Let us recap the objectives of the throughput support. The objectives of throughput support are twofold: 1) to provide specific or absolute throughput support and 2) to provide fair or relative delay support. In section 4.1, we have demonstrated these features with the some simulation scenarios. In this section, we will examine the ability to provide absolute throughput and relative throughput support with under several conditions to gain a deeper understanding and more confidence in the performance of the proposed mechanism.

To achieve this objective, we will evaluate 12 scenarios where 8 factors are varied (see Table 5.1). These factors are the levels of throughput requirements, the types of input traffic (Constant Bit Rate (CBR) vs. Variable Bit Rate (VBR)), channel error, the mode of operation (Ad-Hoc vs. Infrastructure), the use of RTS/CTS mechanism, the number of

Table 5.1: Simulation Parameters for Evaluation of Throughput Support

#	Name	AT-MSs	RT-MSs	Traffic	Error	Mode	RTS	Fairness	Speed (Mbps)
1	3AT+19RT:DTR	4x100Kbps 5x150Kbps 10x250Kbps	4x100Kbps 5x150Kbps 10x250Kbps	CBR	No	Adhoc	No	Util.	2
2	1AT+19RT:VBR	1x500Kbps	19x500Kbps	VBR	No	Adhoc	No	Util.	2
3	1AT+19RT:E-2	1x500Kbps	19x500Kbps	CBR	10^{-2}	Adhoc	No	Util.	2
4	1AT+19RT:E-3	1x500Kbps	19x500Kbps	CBR	10^{-3}	Adhoc	No	Util.	2
5	1AT+19RT:AP	1x500Kbps	19x500Kbps	CBR	No	Infra	No	Util.	2
6	1AT+19RT:RTS	1x500Kbps	19x500Kbps	CBR	No	Adhoc	Yes	Util.	2
7	10AT+20RT	10x100Kbps	20x200Kbps	CBR	No	Adhoc	No	Util.	2
8	20AT+40RT	20x25Kbps	40x50Kbps	CBR	No	Adhoc	No	Util.	2
9	1AT+19RT:5SR	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Util.	5.5
10	1AT+19RT:11SR	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Util.	11
11	1AT+19RT:MRTF	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Util.	5.5+11
12	1AT+19RT:MRUF	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Temp.	5.5+11

MSs (30 vs. 60), data rates (5.5 vs. 11 Mbps), and fairness criteria (utilitarian vs. temporal fairness). The results suggest that absolute throughput can be supported in all scenarios as long as the overload condition is satisfied and the effective channel capacity remains relatively constant even in a situation where network is extremely saturated up to 6 times as much as the effective channel capacity and the number of MSs is as high as 60. The effective channel capacity decreases significantly in two scenarios: a) the error is high, PER $> 10^{-2}$, and b) the MSs operate in infrastructure mode. In these scenarios, if the effective channel capacity is calculated or estimated appropriately, QoS provision with DRAFT+D remains supported. That is, the AT-MSs will receive as much throughput as required while RT-MSs will share the remaining bandwidth according to their throughput requirements and fairness criteria. Next, we will examine the effect of different levels of throughput requirement in providing relative throughput and absolute throughput support.

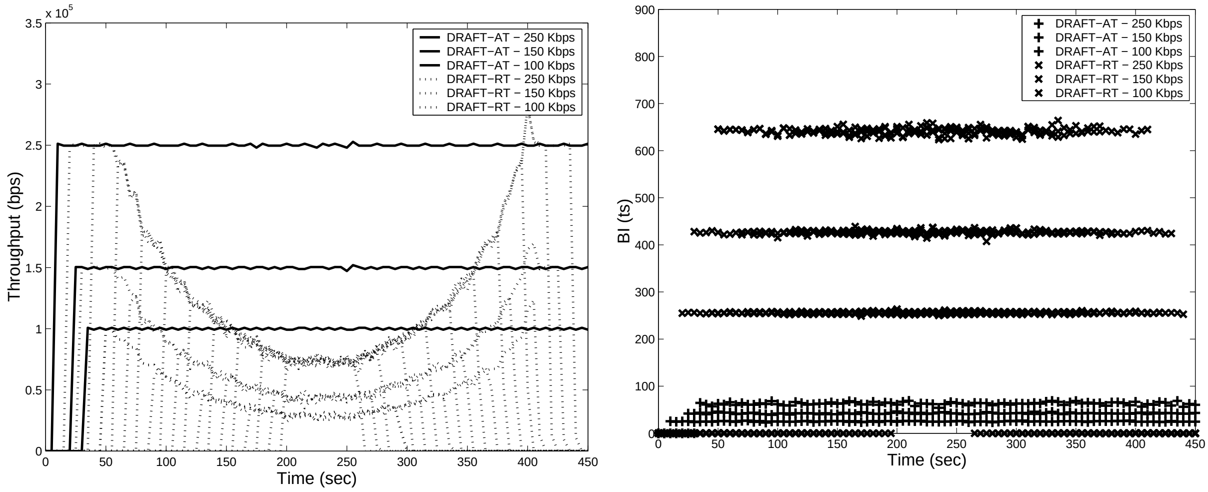
5.1.1 The Effect of Different Levels of Throughput Requirements

Thus far, all MSs have *the same* throughput requirement of 500 Kbps. In this section, we will examine the performance of DRAFT+D when there are multiple flows with *different* throughput requirements, i.e., 250 Kbps, 150 Kbps, 100 Kbps. We consider a scenario (3AT+19RT:DTR) where there are 3 AT-MSs where AT-MSs require throughput of 250 Kbps, 150 Kbps and 100 Kbps, respectively, and 19 RT-MSs where 10 RT-MSs have throughput requirement of 250 Kbps, 5 RT-MSs have throughput requirement of 150 Kbps second, and 4 RT-MSs have throughput requirement of 100 Kbps. For the sake for comparison, we kept both the total throughput requirements of AT-MSs and RT-MSs the same as 1AT-19RT scenario, i.e., 500 Kbps for AT-MSs and 4.35 Mbps for RT-MSs. We used $\omega = 5$ here.

The result shows that the higher number of levels of throughput does not appear to have any impact on the ability of DRAFT+D to provide absolute throughput and relative throughput support. Figure 5.1 (a) shows that AT-MSs receive as much throughput as required for the entire simulation while the RT-MSs share the remaining bandwidth according to the specified throughput requirements. That is, the higher the specified throughput requirement, the higher is the experienced throughput. The reason is that the MSs with higher throughput requirement receive smaller values of BI, as shown in Figure 5.1 (b). It is noteworthy that although AT-MSs and RT-MSs become active at different times, the throughput support is still maintained. Therefore, it appears that neither the levels of throughput requirement nor the time at which MSs begin transmission have any impact on the ability of DRAFT+D to provide absolute throughput and relative throughput support. The finding matches the analysis in Chapter 3 which indicated that only the aggregate throughput requirement from AT-MSs and RT-MSs, ω, θ, and effective channel capacity have an impact on absolute throughput and relative throughput support. Next, we will examine the effect of variable bit rate characteristic of input traffic.

5.1.2 The Effect of Variable Bit Rate (VBR) Traffic

So far, we chose CBR as input traffic to demonstrate the variation of the experienced throughput created by the mechanism. The results in the previous sections show that DRAFT+D

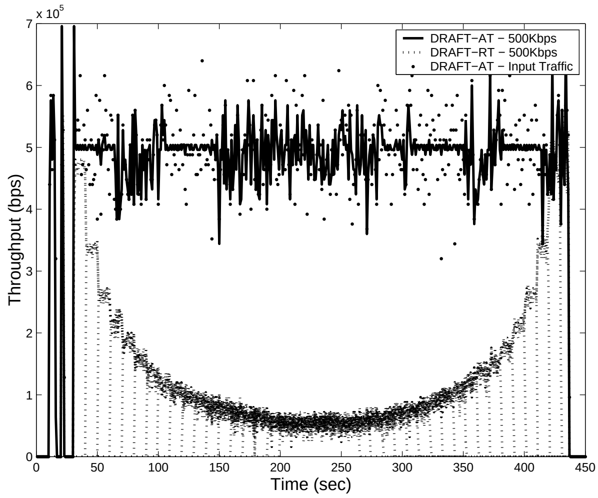


5.1 (a): Throughput

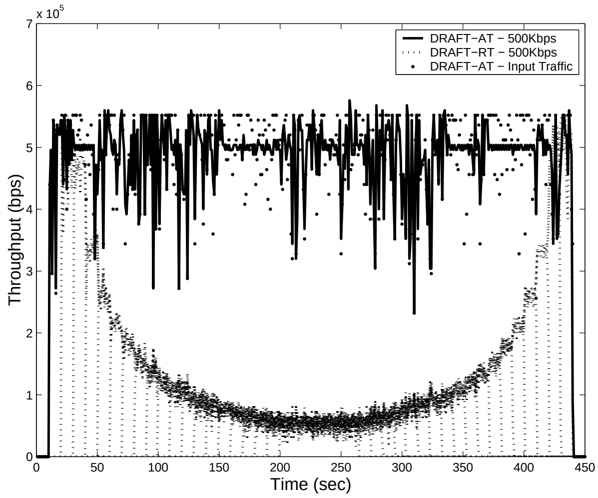
5.1 (b): BI

Figure 5.1: The Effect of Different Throughput Requirements on Throughput Support

introduces only negligible variations to the experienced throughput. This is a good property of DRAFT+D over EDCA and DWFQ. To study the effect of aperiodic input traffic, we consider a scenario (1AT+19RT:VBR) where the input traffic is changed to VBR. Two types of VBR traffic are considered: a) VBR traffic with exponentially distributed inter-arrival time and b) ON-OFF traffic. The inter-arrival time for each packet is exponentially distributed with mean 0.016 second, i.e., 500 Kbps on average. The duration of ON-period is exponentially distributed with mean 0.9 second and the duration of OFF-period is exponentially distributed with mean 0.1 second. Traffic is generated at the rate of 550 Kbps during the ON-period whereas no traffic is generated during the OFF-period, i.e., 555 Kbps on average. The results in Figure 5.2 shows that the variation of experienced throughput of AT-MS increases in both cases. However, this variation is caused by the aperiodicity of inter-arrival time in the offered load. In fact, the variation of experienced throughput is smaller than the original input traffic. This reduced variation is a by-product of DRAFT+D's inherent ability to regulate traffic flow of the embedded token bucket.



5.2 (a): Exponential Inter-Arrival Time



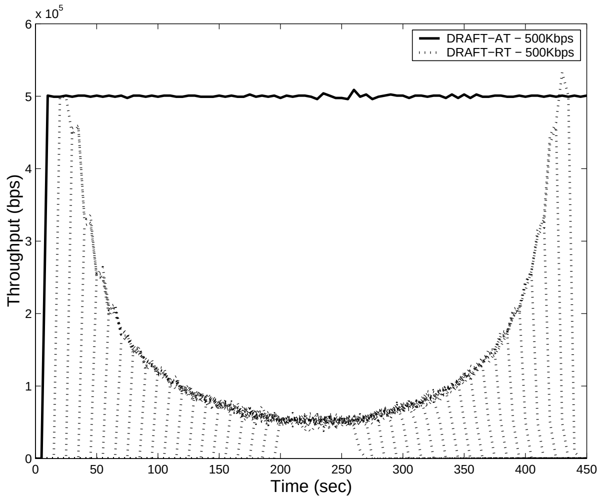
5.2 (b): ON-OFF

Figure 5.2: The Effect of VBR Traffic on Throughput Support

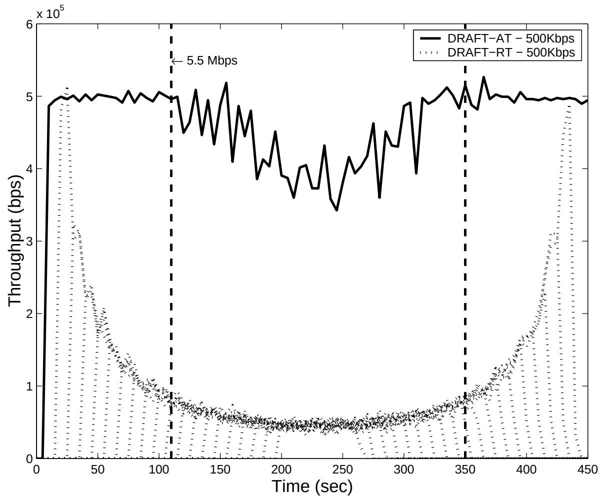
5.1.3 Presence of Channel Error

In WLANs, high error rates are one of the inherited characteristics of a wireless channel. So far, error-free channels have been assumed. In this section, we relax this assumption and consider two scenarios where errors are presented in the wireless channel. For this purpose, we use an ON-OFF (Gilbert-Elliott channel) model [96] to generate errors at the receiver MSs. During the ON-period (or bad-state), the received frames will be assumed as erroneous frames. During the OFF-state (good-state), the received frames will be accepted as they are reported from the physical layer, either as collided frames or as successfully received frames. The duration of ON-period and OFF-period are randomly generated from an exponential distribution with means 0.126×10^{-2} and 0.125 second, respectively. These durations can be translated into a Packet Error Rate (PER) of approximately 10^{-2} (or 1% of the time the channel is bad). Figure 5.3 (a) shows that this level of PER does not have any significant impact on the performance.

Next, we increased the duration of the ON-period to 0.0138 second while we kept the duration of OFF-period at 0.125 second. With these duration-pairs, the PER increases to



5.3 (a): $PER = 10^{-2}$



5.3 (b): $PER = 10^{-1}$

Figure 5.3: The Effect of Channel Error on Throughput Support

approximately 10^{-1} (or 10% of the time the channel is bad). The impact of this PER is depicted in Figure 5.3 (b). According to this figure, we can separate the impact into two segments. The first segment is when the overload threshold is still satisfied, i.e., $0s > t > 110s$ and $t > 350s$. During these periods, the absolute throughput support is maintained. Although, the variation of throughput increases slightly with Coefficient of Variation (COV)[1] of 0.01, the experienced throughput of AT-MS remains approximately at 500 Kbps. The second segment is when the overload threshold is violated, i.e., $110s > t > 350s$. During this period, the variation of throughput increases substantially with COV of 0.12 and the ability to provide absolute throughput is no longer supported. The reason for the dip in the experienced throughput of AT-MS and deterioration of ability to maintain absolute throughput support results from the decrease in effective channel capacity from 1.5 Mbps to 1.25 Mbps due to the increase PER. At this level of effective channel capacity, the fair share of AT-MS is less than the specified throughput requirement; thus absolute throughput cannot be supported.

[1]This is the ratio of the standard deviation of the quantity to its mean

In practice, different MSs may suffer different levels of error rates. Some may suffer even a higher error rate than 10%. However, we conclude from this result that absolute throughput those using DRAFT+D can tolerate error without any degradation as long as there is no significant loss of effective channel capacity. However, in a situation where the error is high enough to cause a significant loss of channel capacity, the ability to provide absolute throughput support can be degraded to relative throughput support. If the network is not overly saturated, the ability to provide absolute throughput can be maintained. For example, if the total offered load in smaller than 5.5 Mbps in a 2 Mbps WLAN, the experienced throughput of AT-MS remains supported. This result indicates a trade-off between the ability to tolerate error and the amount of traffic from AT-MSs to be supported. The higher the ability to tolerate error, the smaller is the amount of traffic from AT-MSs that can be supported. In conclusion, the effect of high error rate on the absolute throughput support can be alleviated by two methods: a) the effective channel capacity should be estimated with an average error rate in mind to maintain absolute throughput support in a real network and b) a new estimate of effective channel capacity can be used in conjunction with DSG-RT to limit the traffic from RT-MSs appropriately and effectively. Next, we will examine the effect of the mode of operation.

5.1.4 The Effect of Presence of Access Point

So far, we assume that the WLAN operates in the Ad-hoc mode where traffic can be sent from a source MS to a destination MS directly. The Ad-hoc mode is sufficient and appropriate to evaluate the behavior of the mechanism because it represents the fundamental operation of the MAC layer. However, the commonly used mode of WLAN is the infrastructure mode where the AP plays an important role in relaying traffic among MSs within the WLAN and/or relaying traffic from/to the outside networks, e.g., Internet. In this section, we consider a scenario (1AT+19RT:AP) where the MSs operate in the infrastructure mode. In 1AT+19RT:AP scenario, there is 1 AT-MS and 19 other RT-MSs, each MS requiring 250 Kbps. This throughput requirement makes up the total offered load to 5 Mbps in a 2 Mbps WLAN. In 1AT+19RT:AP scenario, all traffic from each MS must be relayed via an AP. Figure 5.4 shows that the performance of DRAFT+D remains the same for both AT-MS and

RT-MSs. We also consider another scenario where traffic among MSs and AP is asymmetric. In this scneario, there are 20 MSs. One AT-MS which requires 100 Kbps uplink and 150 Kbps downlink. Among 19 RT-MSs, 9 RT-MSs require 600 Kbps and 10 RT-MSs require 400 Kbps. Figure 5.5 shows that the performance of DRAFT+D remains the same for both that AT-MSs receive throughput as much as required for both uplink and downlink and RT-MSs share the rest of the available bandwidth according their requirements.

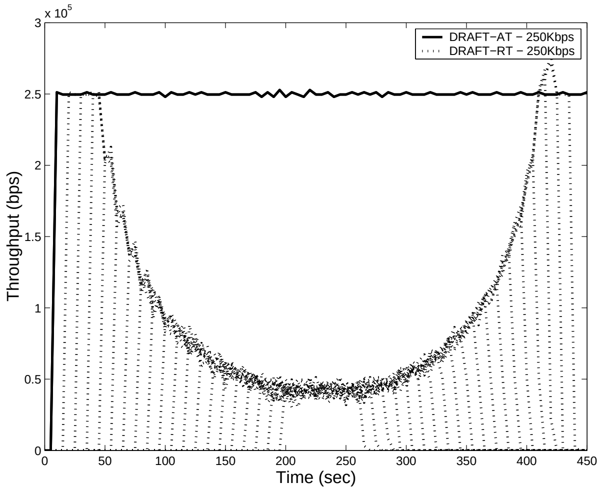


Figure 5.4: The Effect of AP on Throughput Support

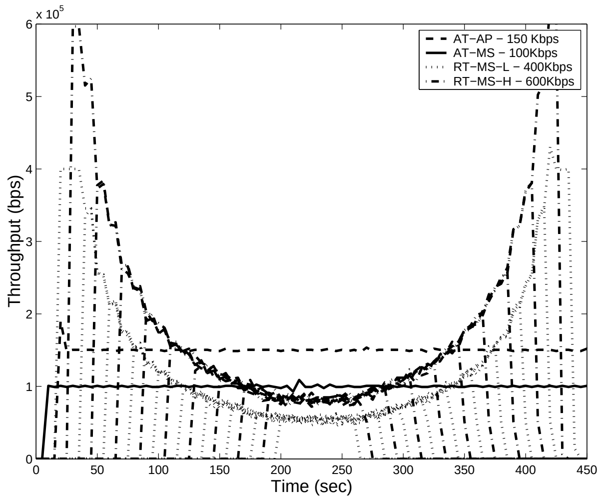


Figure 5.5: The Effect of AP in Asymmetric Traffic WLAN

One important observation is that the effective channel capacity of a WLAN operating in infrastructure mode decreases from 1.5 Mbps to approximately 550 Kbps. However, this is not a result of DRAFT+D. Rather, it is a result of the overhead in relaying all the traffic between source MSs and destination MSs via the AP. For every successful packet transmission, two packets will be transmitted. One is from the source MS to AP and the second is from AP to the destination MS plus all the acknowledgement and re-transmission packets. Consequently, the effective channel capacity of a WLAN is reduced by more than half. However, if most of the traffic in a WLAN comes from the outside networks (a common traffic pattern), the negative effect of AP will be reduced. That is, the AP will be perceived as another MS with high volume of traffic and the effective channel capacity of a WLAN will not deteriorate significantly.

5.1.5 Presence of RTS/CTS

The hidden-terminal situation occurs when a set of MSs cannot hear the transmissions of the other set of MSs. To prevent this potential problem, an exchange of control frames called Request to Send/Clear to Send (RTS/CTS) prior to the packet transmission has been suggested. Although this operation is optional in the IEEE standard, it is usually beneficial to employ this in a WLAN [49]. In this section, we first evaluate the effect of hidden terminals. Then we evaluate the performance of DRAFT+D when the RTS/CTS mechanism is used.

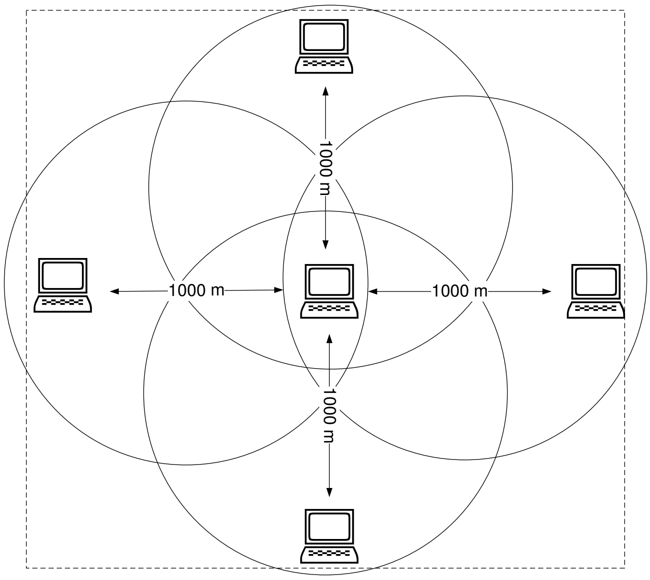
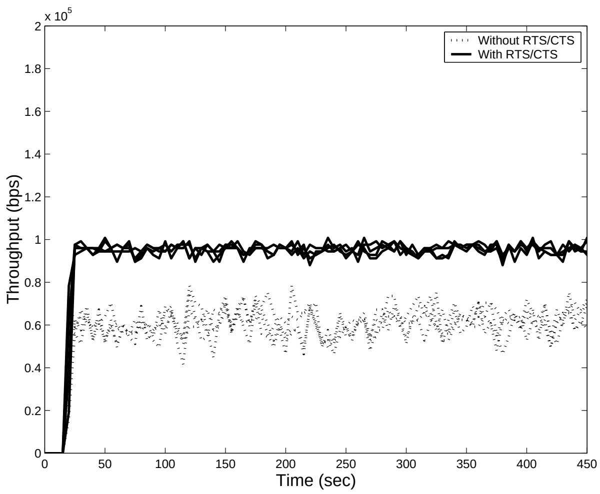


Figure 5.6: Hidden Terminals Scenario

To demonstrate the effect of hidden terminals, we consider a scenario which is composed of 5 AT-MSs - four on the centers of the four edges of a square and one in the center of the square. The four AT-MSs on the centers of the edges of the square cannot hear the transmissions from one another. These four AT-MSs have traffic requirements of 100 Kbps each of which is to be transmitted to the fifth MS in the center which can hear the transmissions from all MSs, as shown in Figure 5.6. This scenario represents the worst-case scenario since all four MSs are hidden from one another. If additional MSs are presence in the network and are transmitting to the MS in the center, they will not be hidden from



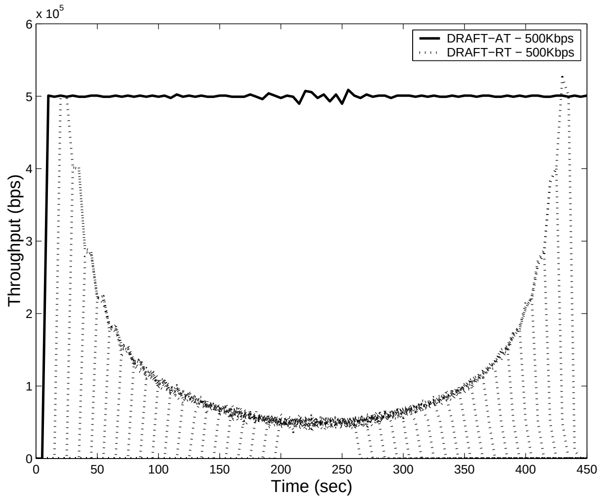


Figure 5.7: The Effect of Hidden Terminals

Figure 5.8: The Effect of RTS/CTS on Throughput Support

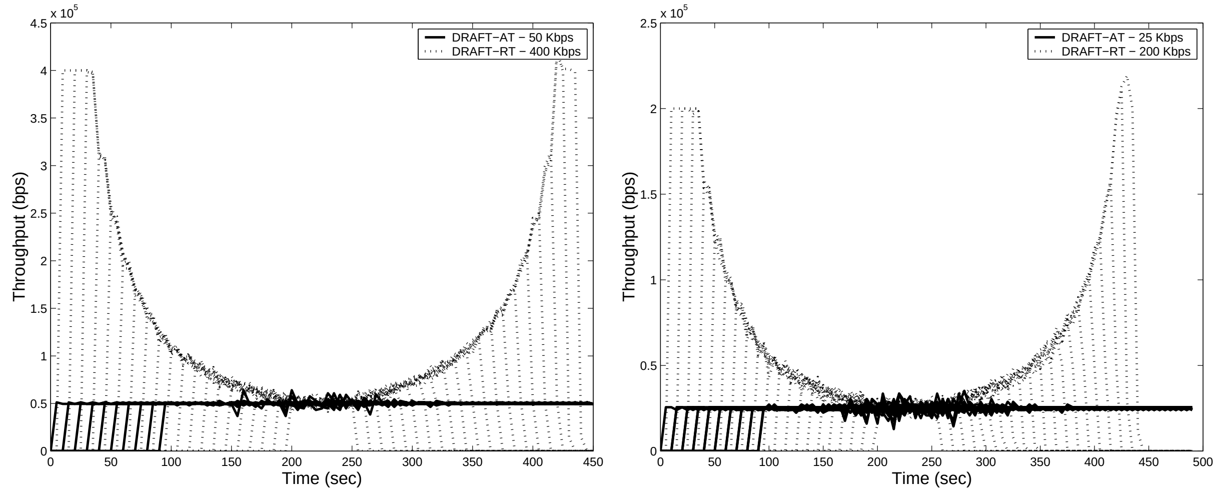
one of these four MSs. Due to the significantly increased number of collisions from hidden terminals, the performance of all MSs degrades in this environment when RTS/CTS is not used, as shown in Figure 5.7. However, when RTS/CTS is used, the problem of RTS/CTS is significantly alleviated at the expense of additional overhead and slightly increased variation of throughput. It is important to note that in a highly loaded network, the collision rate could increase due to the collisions among RTS/CTS frames. Based on this result, we assume that the RTS/CTS mechanism indeed mitigates much of the hidden-terminal problem. In the next scenario, we demonstrate the performance of DRAFT+D of the basic scenario when RTS/CTS is used and no hidden terminal is explicitly placed in the network. Figure 5.8 shows that the RTS/CTS mechanism by itself has no significant impact on the performance of DRAFT+D. The experienced throughput of both AT-MS and RT-MSs remains similar to the scenario without when RTS/CTS was not employed. Next, we will examine the effect of the number of MSs.

5.1.6 The Effect of the Number of Mobile Stations

Thus far, we limited the number of MS of AT-MS to one and the number of RT-MS to 19 (total of 20 MSs). The number of MSs often reported as a important factor that deteriorate the effectiveness of distributed mechanisms in WLAN [97]. In WLANs using DCF, the number of MSs has a significant impact on the performance. The higher the number of active MSs, the higher the probability of collision and the lower the effective channel capacity. In this section, we evaluate the performance of DRAFT+D two scenarios when the number of AT-MSs and RT-MSs increases. In the first scenario (10AT+20RT), the number of AT-MSs increases from 1 to 10 while the number of RT-MSs increases to 20, total of 30 MSs. The throughput requirement of AT-MSs is 50 Kbps while the throughput requirement of RT-MSs is 400 Kbps. In the second scenario (20AT+40RT), the number of AT-MSs increases to 20 while the number of RT-MSs increases to 40, total of 60 MSs. The throughput requirement of AT-MSs is 25 Kbps while the throughput requirement of RT-MSs is 200 Kbps. In both scenarios, Figure 5.9 shows that the variations of throughput of AT-MSs increase slightly. However, the AT-MSs still receive as much throughput as the specified throughput requirement. This result suggests that the number of MSs has only a small impact on the throughput support. At this level of network saturation and number of MSs, DRAFT+D performs very well (robust and relatively unaffected) in comparison with other QoS mechanisms in WLAN. Next, we will present the performance of DRAFT+D where the data rate is changed to 5.5 and 11 Mbps.

5.1.7 The Effect of Data Rates

In all the previous scenarios, the data rate of all MSs was set to 2 Mbps. However, other data rates are also available in WLAN; for instance, 11 Mbps, 5.5 Mbps, 2 Mbps, and 1 Mbps in 802.11b. In this section, we will investigate the performance of DRAFT+D in two scenarios where the data rate is changed to 5.5 Mbps and 11 Mbps. In both scenarios, the throughput requirement of all MSs is set to 1 Mbps to match the increased effective channel capacity in both scenarios. Figure 5.10 shows that the experienced throughput of AT-MS and RT-MSs remains the same as the scenario with 2 Mbps data rate. That is, AT-MS receives as much

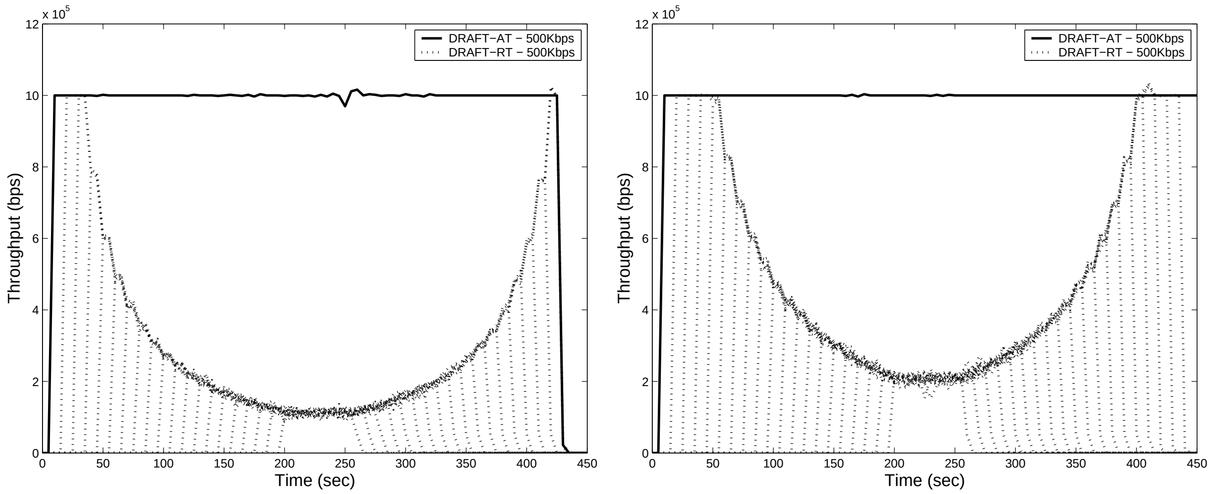


5.9 (a): 10AT-MSs & 20aRT-MSs

5.9 (b): 20AT-MSs & 40aRT-MSs

Figure 5.9: The Effect of Number of MSs on Throughput Support

throughput as required and RT-MSs share the remaining bandwidth. This suggests that the performance of DRAFT+D is unaffected by the data rate.



5.10 (a): 5.5 Mbps WLAN

5.10 (b): 11 Mbps WLAN

Figure 5.10: The Effect of Data Rate on Throughput Support

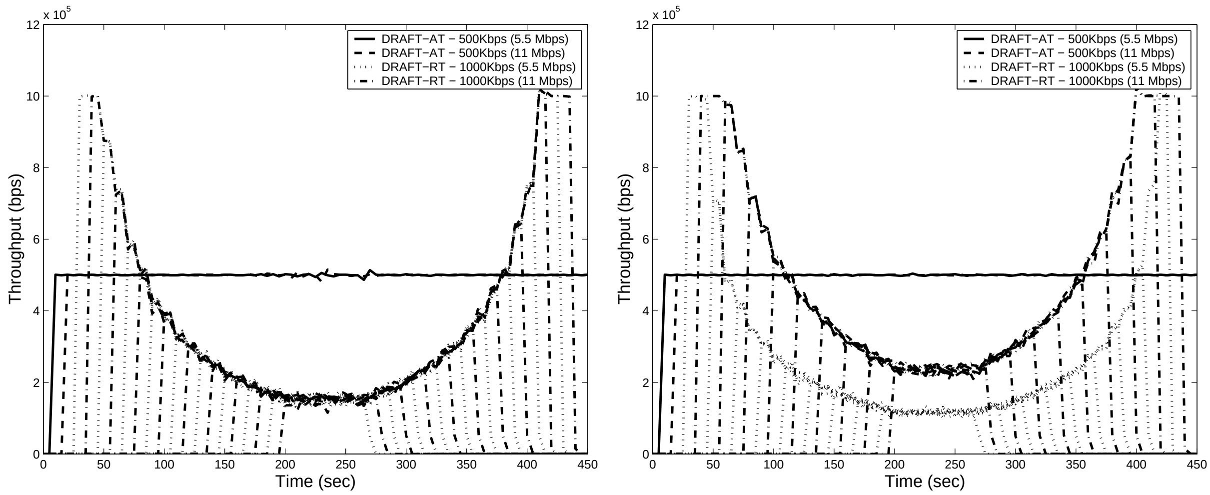
5.1.8 The Effect of Multi-Rate Environment

In addition to availability of multiple data rates, different MSs can connect to a WLAN at different data rates in a multi-rate environment. A multi-rate environment raises a new issue of fairness because providing the same throughput or delay to MSs connecting at different data rates requires a different amount of network resources. For example, an MS connecting at 5.5 Mbps requires twice as much time to achieve the same throughput as an MS connecting at 11 Mbps. Therefore to provide fairness in terms of network resources or temporal fairness, DRAFT+D is equipped with an alternative way to calculate the value of weight accordingly. To demonstrate the effect of different fairness criteria, we consider a scenario (2AT-18RT:MR) where half of the MSs connected at 5.5 Mbps while the other half of the MSs connected at 11 Mbps. More specifically, 1 AT-MS and 9 MSs connect to the WLAN at 11 Mbps while 1 AT-MS and 9 MSs connect at 5.5 Mbps. Based on 2AT-18RT:MR scenario, two experiments were conducted. One experiment used utilitarian fairness and the other used temporal fairness.

The result in Figure 5.11 shows that AT-MSs receive as much throughput as they required regardless of their connection rates or their fairness criteria. This result is as expected and desirable for absolute throughput support. On the contrary, RT-MSs with the same throughput requirement receive the same throughput based on utilitarian fairness, (Figure 5.11 (a)) while RT-MSs connecting at a lower speed receive lower throughput than MSs connecting at a higher speed based on the temporal fairness. Figure 5.11 (b) shows that RT-MSs connecting at 11 Mbps receives twice throughput as much as the MSs connecting at 5.5 Mbps when temporal fairness was enforced. Depending on the objective of the protocol designer or network provider, either fairness criteria can be properly enforced.

In short, we have showed that DRAFT+D is equipped to and performs well in supporting both fairness criteria: a) utilitarian fairness and b) temporal fairness. AT-MSs receive as much throughput as required regardless of their connected data rate or fairness criteria while RT-MSs fairly share the remaining bandwidth depending on the appropriate fairness criteria.

To this end, we have demonstrated the performance of DRAFT+D in various scenarios. The simulation results showed that DRAFT+D performed well regardless of the levels



5.11 (a): Utilitarian Fairness

5.11 (b): Temporal Fairness

Figure 5.11: The Effect of Temporal Fairness in Multi-Rate WLAN on Throughput Support

of throughput requirement, the number of levels of throughput requirement, the time at which MSs become active, number of active MSs, traffic type, mode of operation, the use of RTS/CTS, data rate, and fairness criteria. Moreover, the performance of DRAFT+D remains unchanged in moderately erroneous environment, e.g, PER is smaller than 10^{-2} where the effective channel capacity is not significantly degraded. A higher tolerance to errors can be achieved at the expense of a smaller amount of supported traffic. Next, we will demonstrate the performance of distributed safeguard mechanism of DRAFT+D, namely DSG-RT.

5.2 COMPREHENSIVE EVALUATION OF DSG-RT

The objective of DSG-RT is to limit the amount of traffic from new RT-MSs to prevent excessive traffic that causes the deterioration of the ability of DRAFT+D to provide absolute QoS support. DSG-RT is achieved by comparing the experienced throughput of probing packets with an overload threshold. A new RT-MS will not be allowed to continue transmitting

Table 5.2: Simulation Parameters for Evaluation of DSG-RT

#	Name	AT-MSs	RT-MSs	Traffic	Error	Mode	RTS	Fairness	Speed (Mbps)
1	3AT+19DSG-RT:DTR	4x100Kbps 5x150Kbps 10x250Kbps	4x100Kbps 5x150Kbps 10x250Kbps	CBR	No	Adhoc	No	Util.	2
2	1AT+19DSG-RT:VBR	1x500Kbps	19x500Kbps	VBR	No	Adhoc	No	Util.	2
3	1AT+19DSG-RT:E-2	1x500Kbps	19x500Kbps	CBR	10^{-2}	Adhoc	No	Util.	2
4	1AT+19DSG-RT:E-3	1x500Kbps	19x500Kbps	CBR	10^{-3}	Adhoc	No	Util.	2
6	1AT+19DSF-RT:RTS	1x500Kbps	19x500Kbps	CBR	No	Adhoc	Yes	Util.	2
7	10AT+20DSG-RT	10x100Kbps	20x200Kbps	CBR	No	Adhoc	No	Util.	2
8	20AT+40DSG-RT	20x25Kbps	40x50Kbps	CBR	No	Adhoc	No	Util.	2
9	1AT+19DSG-RT:5SR	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Util.	5.5
10	1AT+19DSG-RT:11SR	1x500Kbps	19x500Kbps	CBR	No	Adhoc	No	Util.	11

if experienced throughput is smaller than the expecting overload threshold. The overload threshold is the throughput level that can be calculated independently in each RT-MS from the specified quantum rate, ω, and θ, Eq. (3.22).

The procedure of DSG-RT begins by sending out a number of packets at the desired rate. During this period, the experienced throughput is calculated. DSG-RT compares the experienced throughput with the overload threshold. The experienced throughput is calculated from the EWMA of instantaneous throughput. The overload threshold is calculated from the product of the safety factor, the overload ratio, and the throughput requirement. During the testing procedure, if the experienced throughput is smaller than the overload threshold, the new flow will be immediately stopped. Otherwise, a new flow can continue.

In the following section, we will evaluate 10 scenarios where 8 factors are varied (see Table 5.2). We will demonstrate that DSG-RT performs well in scenarios regardless of the levels of throughput requirement, the time at which MSs become active, traffic type, error rate, the use of RTS/CTS, number of active MSs, and data rate.

5.2.1 The Effect of Different Levels of Throughput Requirements

In scenario 3AT+19DSG-RT:DTR, we examine the performance of DSG-RT when multiple flows with different throughput requirements are presented. We consider a scenario where there are 3 AT-MSs with throughput requirements of 500 Kbps, 250 Kbps and 125 Kbps, and 19 RT-MSs where 10 RT-MSs with throughput requirement of 500 Kbps, 5 RT-MSs with throughput requirement of 250 Kbps second, and 4 RT-MSs with throughput requirement of 150 Kbps. Figure 5.12 shows that DSG-RT performs well under this circumstance.

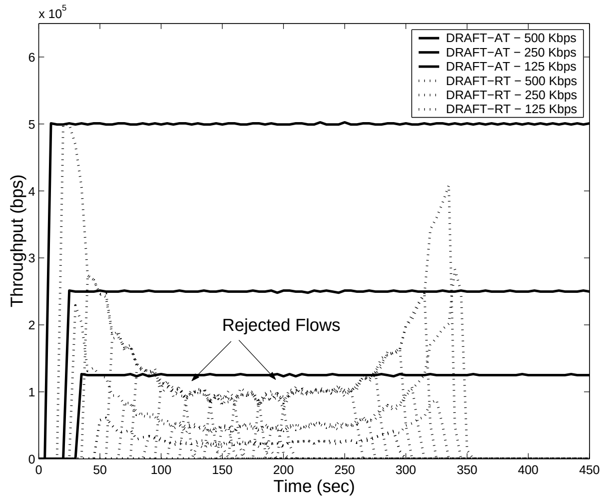
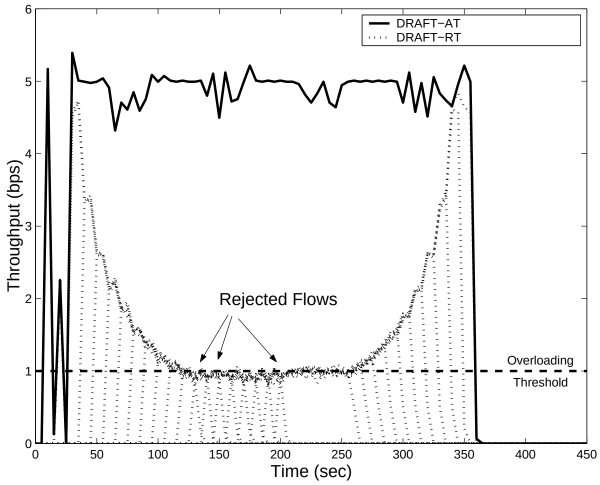


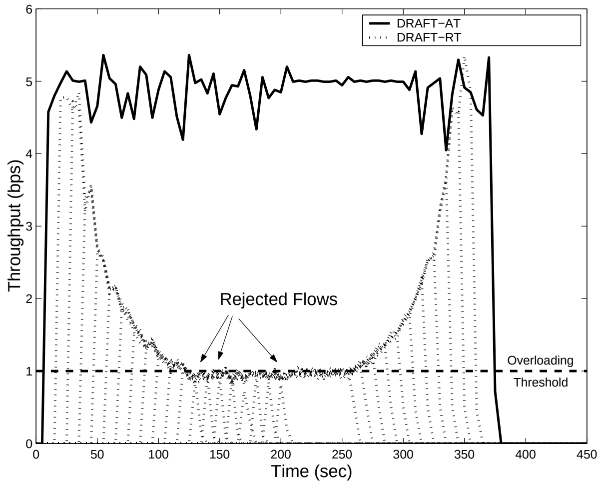
Figure 5.12: The Effect of Multiple Flows with Different Requirements

5.2.2 The Effect of Variable Bit Rate Traffic

In scenario 1AT+19DSG-RT:VBR, the VBR traffic is considered. We consider 2 types of VBR traffic: a) VBR traffic with exponentially distributed inter-arrival time with the mean of 0.016 and b) ON-OFF traffic with 0.9 second in ON-period and 0.1 second in OFF-period. Figure 5.13 shows that performance of DSG-RT remains the same even when the inter-arrival time of traffic is aperiodic.



5.13 (a): Exponential Inter-Arrival Time

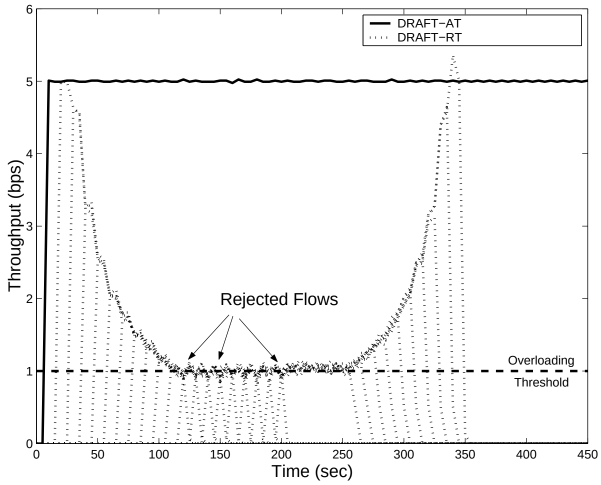


5.13 (b): ON-OFF

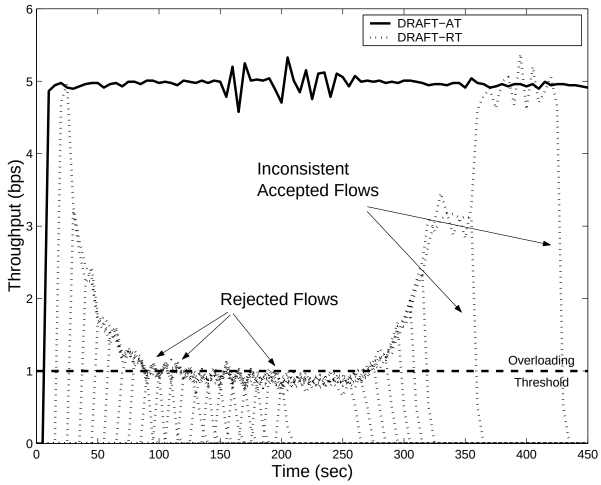
Figure 5.13: The Effect of Variable Bit Rate (VBR) Traffic on DSG-RT

5.2.3 The Effect of Channel Error

In this section, we examine the performance of DSG-RT in two scenarios (1AT+19DSG-RT:E-2 and 1AT+19DSG-RT:E-3) where Packet Error Rate (PER) of 10^{-2} and 10^{-3} are present in the wireless channel. The simulation results show the accuracy of DSG-RT in limiting new traffic remains undisturbed as long as the PERs is smaller than 10^{-2} (Figure 5.14 (a)). However, the performance of DSG-RT begins to deteriorate slightly as Packet Error Rate (PER) increases higher than 10^{-1}. Inaccurate admission and rejection can be observed. The reason is that the variation of experienced throughput increases directly as a function of error rate. Therefore, as the variation of experienced throughput increases, the accuracy of the estimate of the experienced throughput and the accuracy of the testing procedure decreases. This issue can be resolved by simply increasing the number of packets during the transient period as shown in Figure 5.15.

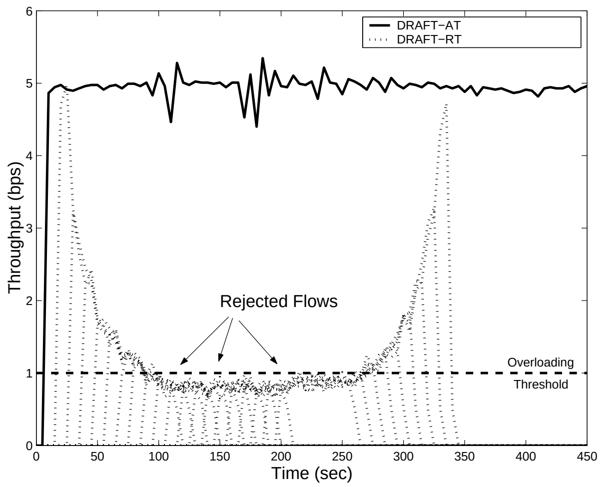


5.14 (a): $PER = 10^{-2}$



5.14 (b): $PER = 10^{-1}$

Figure 5.14: The Effect of Channel Error on DSG-RT

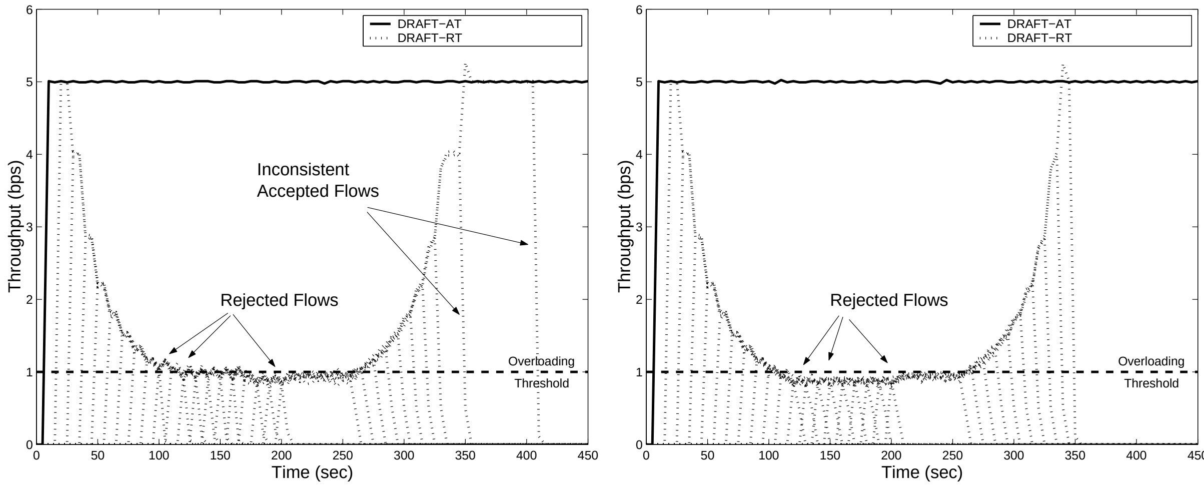


5.15 (a): $PER = 10^{-1}$

Figure 5.15: The Effect of Channel Error on DSG-RT, $N1 = 50$

5.2.4 The Effect of Request to Send/Clear to Send

In this section, we evaluate scenario 1AT+19DSF-RT:RTS when RTS/CTS mechanism is used to mitigate hidden-terminal problem. Figure 5.16 (a) shows that the accuracy of DSG-RT in limiting new traffic deteriorates slightly when RTS/CTS is used. This result suggests that the number of packets during the transient period should be set to 50 as shown in Figure 5.16 (b).

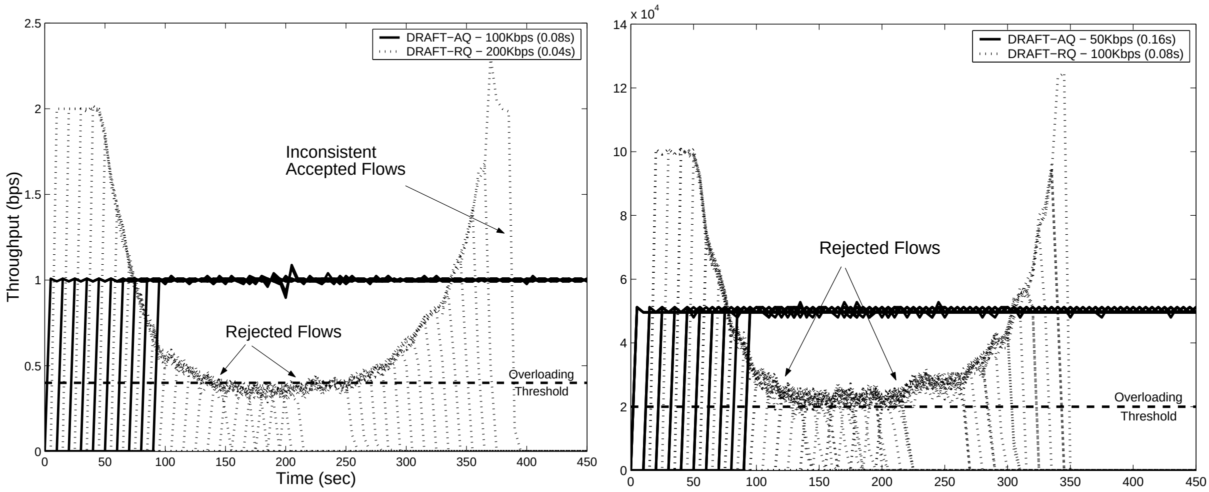


5.16 (a): $pkt^{th} = 30$ 5.16 (b): $pkt^{th} = 50$

Figure 5.16: The Effect of RTS/CTS on DSG-RT

5.2.5 The Effect of the Number of Mobile Stations

In this section, we evaluate the performance of DRAFT+D two scenarios when the number of AT-MSs and RT-MSs increases. In the first scenario (10AT+20DSG-RT), the number of AT-MSs increases from 1 to 10 while the number of RT-MSs increases from 10 to 20, or a total of 30 MSs. The throughput requirement of AT-MSs is 100 Kbps while the throughput requirement of RT-MSs is 200 Kbps. In the second scenario (20AT+40DSG-RT), the number of AT-MSs increases to 20 while the number of RT-MSs increases to 40, or a total of 60 MSs. The throughput requirement of AT-MSs is 50 Kbps while the throughput requirement of RT-MSs is 100 Kbps. Figure 5.17 shows that DSG-RT still performs well when the number of MSs increases. The experienced throughput of AT-MSs and RT-MSs remains stable and the testing procedure is still accurate.



5.17 (a): 10 AT-MSs, 20 RT-MSs

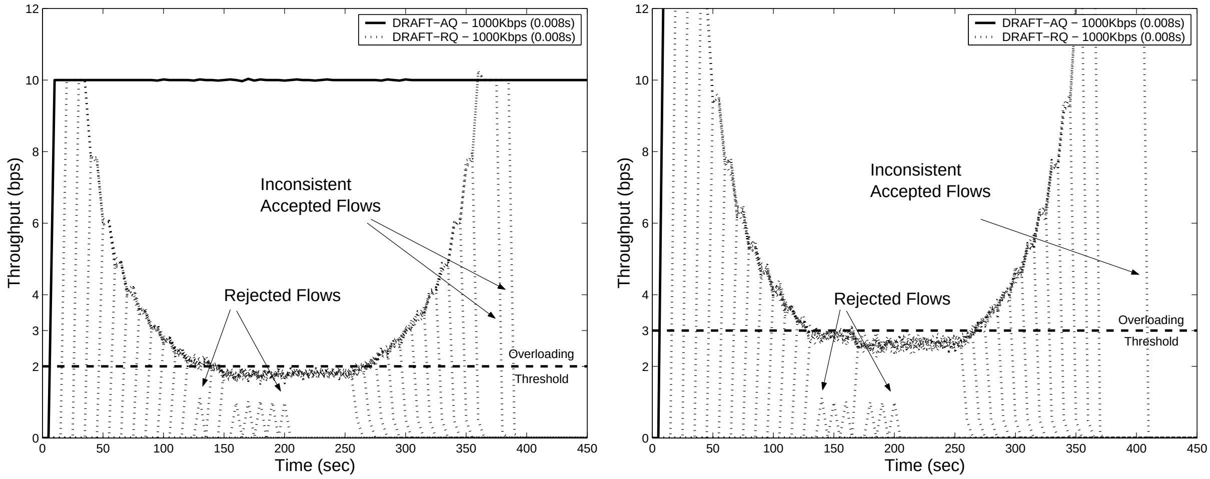
5.17 (b): 20 AT-MSs, 40 RT-MSs

Figure 5.17: The Effect of the Number of Mobile Stations

5.2.6 The Effect of Data Rate

In this section, we investigate the performance of DSG-RT in different data rates (scenario 1AT+19DSG-RT:5SR and 1AT+19DSG-RT:11SR). Figure 5.18 shows that the performance of DSG-RT remains about the same as the scenario with 2 Mbps data rate. There are some inaccurately rejected and admitted RT-MSs. The reason is that the estimate of experienced throughput needs some time to become mature. As the data rate increases, the number of packets during the transient period needs to be increased as well.

In conclusion, we demonstrated the performance of DSG-RT in various scenarios. The simulation results show that DSG-RT performed well regardless of the levels of throughput requirement, the number of levels of throughput requirement, the time at which MSs become active, the number of probing MSs, the number of active MSs, traffic type, mode of operation, the use of RTS/CTS, data rate, error rate, and fairness criteria. Moreover, three important parameters of DSG-RT are also evaluated. The results showed that the number of packets prior to the admission higher than 30 and the number of probing packet higher than 10



5.18 (a): 5.5 Mbps

5.18 (b): 11 Mbps

Figure 5.18: The Effect of Data Rate

are sufficient to provide good performance in all tested scenarios. The safety factor is a tradeoff between the number of accepted RT-MSs (or amount of traffic from RT-MSs) and the experienced throughput of RT-MSs. Finally, DSG-RT also performed well in situation where the number of simultaneously active MSs.

5.3 COMPREHENSIVE EVALUATION OF DELAY SUPPORT

The objective of absolute delay support in DRAFT+D is to provide an average HoQ delay. HoQ delay is a successful transmission delay of a HoQ packet. In Section 4.1, we showed a preview of absolute delay support of the basic scenario, 1AD+19RT scenario. In this section, we will evaluate the ability to provide absolute delay in 8 scenarios with different target delay requirements, traffic type, the mode of operation, the use of RTS/CTS, data rates, error rates, and fairness criteria, as shown in Table 5.3.

The results show that the absolute delay can be supported. The HoQ delay of AD-MSs is within the specified target delay as long as the condition to provide absolute throughput is

Table 5.3: Simulation Parameters for Evaluation of Delay Support

#	Name	AD-MSs	RT-MSs	Traffic	Error	Mode	RTS	Fairness	Speed (Mbps)
1	3AD+19RT:DTR	4x100Kbps:0.08 5x150Kbps:0.053 10x250Kbps:0.032	4x100Kbps 5x150Kbps 10x250Kbps	CBR	No	Adhoc	No	Util.	2
2	1AD+19RT:VBR	1x500Kbps:0.016	19x500Kbps	VBR	No	Adhoc	No	Util.	2
3	1AD+19RT:E-2	1x500Kbps:0.016	19x500Kbps	CBR	10^{-2}	Adhoc	No	Util.	2
4	1AD+19RT:E-3	1x500Kbps:0.016	19x500Kbps	CBR	10^{-3}	Adhoc	No	Util.	2
5	1AD+19RT:AP	1x500Kbps:0.016	19x500Kbps	CBR	No	Infra	No	Util.	2
6	1AD+19RT:RTS	1x500Kbps:0.016	19x500Kbps	CBR	No	Adhoc	Yes	Util.	2
7	10AD+20RT	10x100Kbps:0.08	20x200Kbps	CBR	No	Adhoc	No	Util.	2
8	20AD+40RT	20x25Kbps:0.32	40x50Kbps	CBR	No	Adhoc	No	Util.	2
9	1AD+19RT:5SR	1x500Kbps::0.016	19x500Kbps	CBR	No	Adhoc	No	Util.	5.5
10	1AD+19RT:11SR	1x500Kbps:0.016	19x500Kbps	CBR	No	Adhoc	No	Util.	11
11	1AD+19RT:MRTF	1x500Kbps:0.016	19x500Kbps	CBR	No	Adhoc	No	Util.	5.5+11
12	1AD+19RT:MRUF	1x500Kbps:0.016	19x500Kbps	CBR	No	Adhoc	No	Temp.	5.5+11

maintained. The important finding is that the factor that poses the heaviest burden on the delay support is aperiodicity of input traffic (Figure 5.20). Other factors affect the HoQ delay only slightly. Next, we will present the simulation results and the simulation parameters of each scenario.

5.3.1 The Effect of Multiple Flows with Different Target Delay Requirements

In 1AD+19RT scenario, we presented the performance of DRAFT+D in providing absolute delay support when the target delay was set to 0.016 second. In this scenario, we will investigate the performance of DRAFT+D when there are multiple target delay requirements. We consider scenario 3AD+19RT:DTR with different target delay requirement. The 3AD+19RT scenario consists of 3 AD-MSs with target delay requirements of 0.032, 0.054 and 0.08 second, respectively, and 19 RT-MSs. The throughput requirement of each AD-MS is 250 Kbps, 150 Kbps, and 100 Kbps, respectively, as we keep the total throughput requirement of AD-MSs at 500 Kbps. All AD-MSs become active at the same time, at $t = 10$. The the simulation parameters are kept the same as 1AD+19RT scenario. Figure 5.19 shows that the experienced delay of AT-MSs remains well within the specified target delay for the entire simulation. This result suggests that different levels of target delay do not have any impact on absolute delay support.

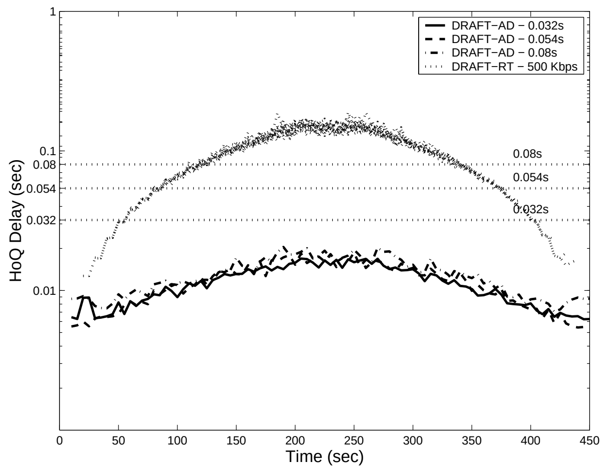
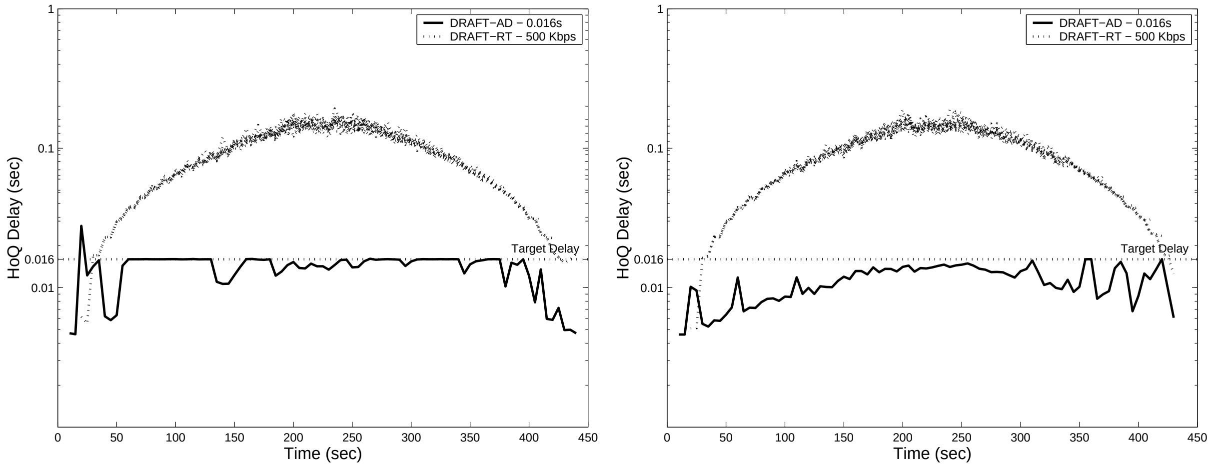


Figure 5.19: The Effect of Different Delay Requirements on Delay Support

5.3.2 Variable Bit Rate (VBR) Traffic

The characteristic of VBR traffic is the aperiodicity of inter-arrival time of the input traffic. This characteristic is found in many QoS-sensitive applications such as video applications. The aperiodicity of inter-arrival time poses a heavy burden on a mechanism that tries to meet target delay because the traffic may arrive back to back and cause a temporarily long delay. Additional bandwidth can be allocated to alleviate this potential problem as discussed earlier in Section 4.1.

In this section, we will examine the ability of DRAFT+D to provide delay support when the input traffic is changed from CBR to VBR. We considered 2 types of VBR traffic: a) traffic with exponentially distributed inter-arrival times and b) on-off traffic with 0.9 second in on-period and 0.1 second in off-period. Figure 5.20 shows that the HoQ delay of AD-MS can be as high as the target delay regardless of the network load. The increased HoQ delay is due to the back-to-back arrival of packets in VBR traffic (both types). Despite the back-to-back packet arrival, the HoQ delay of AD-MS is still within the specified target delay most of the time. This result suggests that DRAFT+D is able to provide a specific delay support (on average) for the HoQ packet.



5.20 (a): Exponential Inter-Arrival Time

5.20 (b): ON-OFF

Figure 5.20: The Effect of VBR on Delay Support

5.3.3 Presence of Channel Error

In this section, we will evaluate the performance when error is present in the wireless channel where the PER of wireless channel is 10^{-2}. Figure 5.21 shows that the experienced delay of AD-MS is still within the specified target delay.

5.3.4 The Effect of Presence of Access Point

In this section, we evaluate the performance of DRAFT+D when an Access Point (AP) is presence. We consider a scenario (1AD+19RT:AP) where there are 1 AD-MS and 19 RT-MSs with 250 Kbps throughput requirement. The AD-MS has the offered load of 250 Kbps and requires 0.032s target delay. Figure 5.22 shows that the experienced delay of AD-MS remains the same as the scenario where an AP is not used.

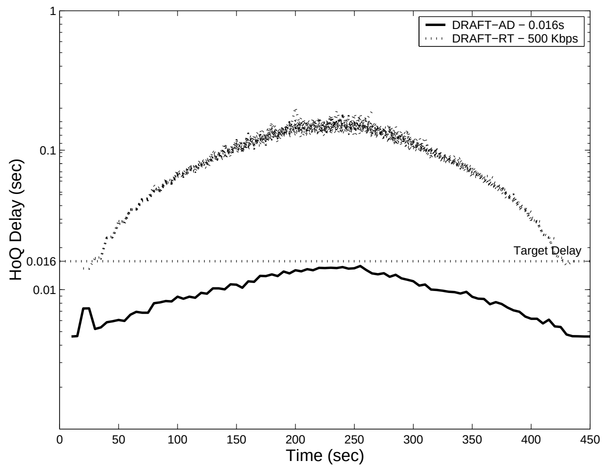


Figure 5.21: The Effect of Channel Error on Delay Support

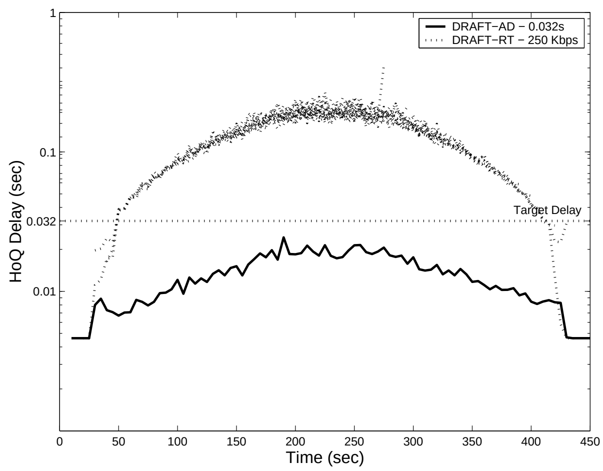


Figure 5.22: The Effect of Access Point on Delay Support

5.3.5 Presence of Request to Send/Clear to Send

We evaluate the performance of A-DRAFT when Request to Send/Clear to Send (RTS/CTS) mechanism is used at the MAC level. Figure 5.23 shows that the performance of AD-MS remains the same.

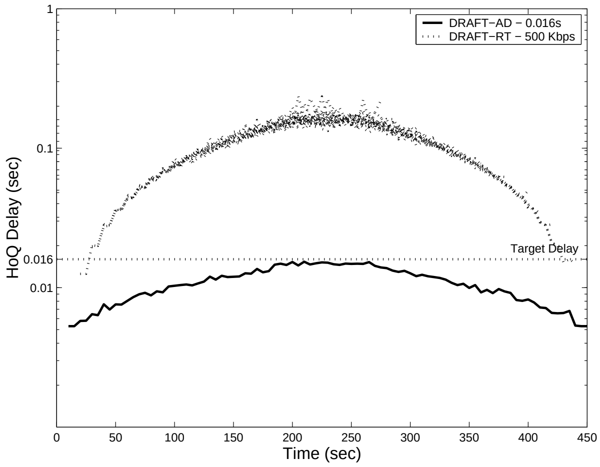


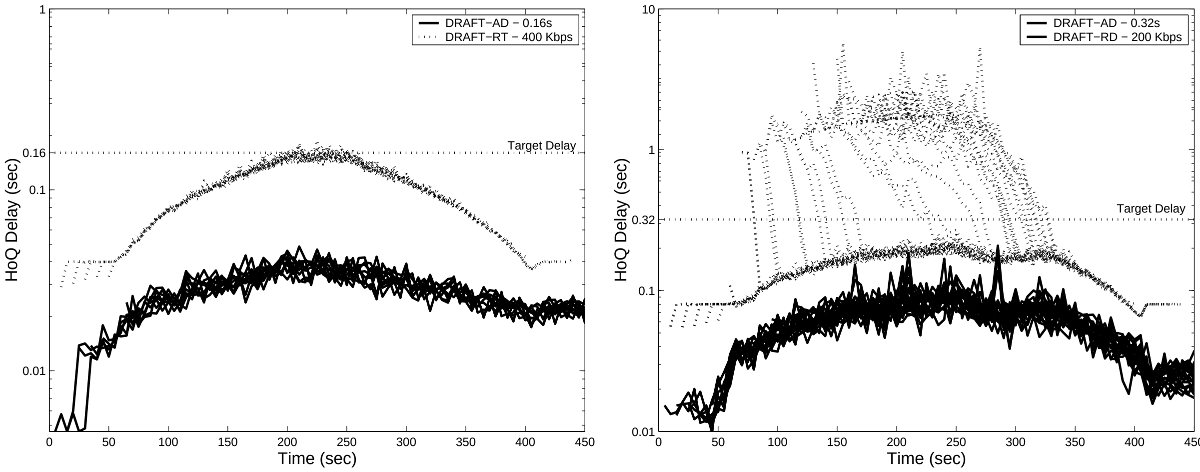
Figure 5.23: The Effect of RTS/CTS on Delay Support

5.3.6 The Effect of the Number of Mobile Stations

In this section, we will demonstrate the performance of DRAFT+D when the number of MSs increases. In the first scenario, there are 10 AD-MSs and RT-MSs. The AD-MSs have the target delay of 0.16 second and throughput requirement of 50 Kbps while the RT-MSs have the throughput requirement of 400 Kbps. In the second scenario, there are 20 AD-MSs have the target delay of 0.32 second and throughput requirement of 25 Kbps and 40 RT-MSs have the throughput requirement of 200 Kbps. The total throughput requirement of AD-MSs and RT-MSs of both scenarios is kept to same as 1AD+19RT scenario. Figure 5.24 shows that the HoQ delay of AD-MSs remains within the specified target delay for the entire simulation for both scenarios. This result suggests that the number of MSs does not appear to have any significant effect on the absolute delay support.

5.3.7 The Effect of Data Rates

In this section, we will investigate the performance of DRAFT+D in 5.5 Mbps and 11 Mbps WLAN. In these scenarios, both AD-MS and RD-MSs require the target delay of 0.008 second and have the throughput requirement of 1 Mbps. The total offered load in both scenarios is 20 Mbps. Figure 5.25 shows that the HoQ delay of AD-MS remains within the

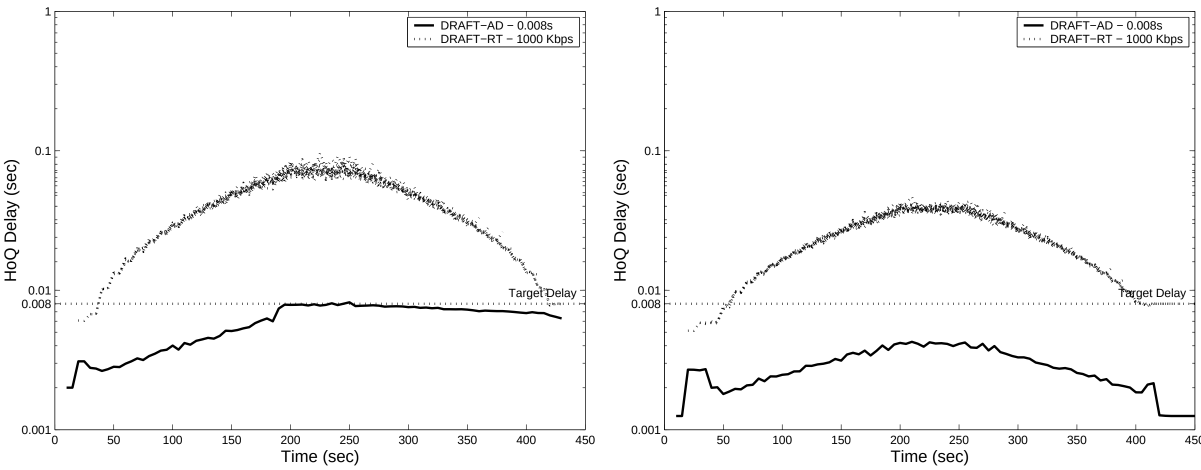


5.24 (a): 10AD-MSs & 20RT-MSs

5.24 (b): 20AD-MSs & 40aRD-MSs

Figure 5.24: The Effect of Number of MSs on Delay Support

specified target delay for the entire simulation in both scenarios.



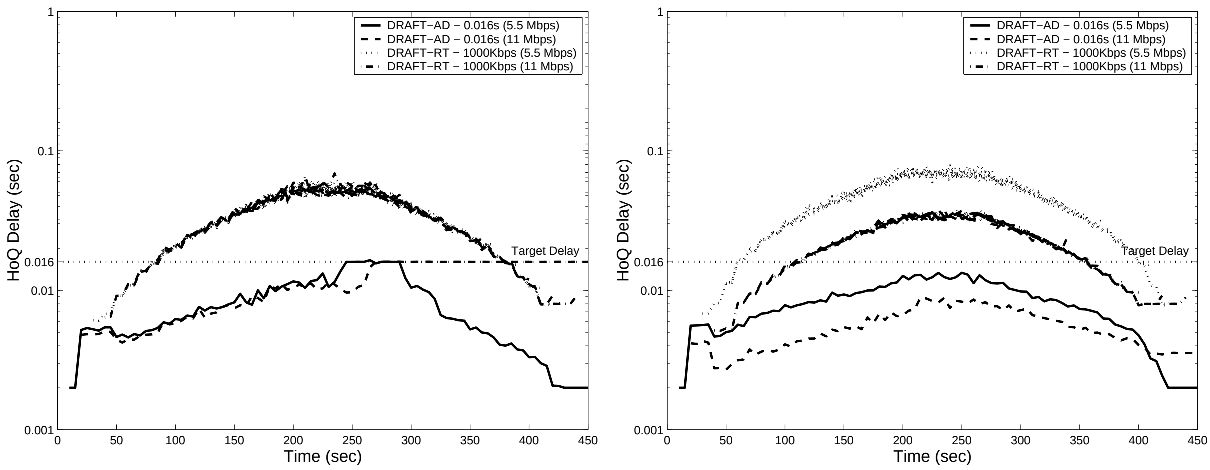
5.25 (a): 5.5 Mbps

5.25 (b): 11 Mbps

Figure 5.25: Performance of Delay Support in 5.5 Mbps

5.3.8 The Effect of Multi-Rate Environment

To demonstrate the effect of different fairness criteria on delay support, we consider a scenario (2AD-18RT:MR) where half of MSs connected at 5.5 Mbps while the other half of MSs connected at 11 Mbps. More specifically, 1 AD-MS and 9 RT-MSs connect to the WLAN at 11 Mbps while 1 AD-MS and 9 RT-MSs connect at 5.5 Mbps. Based on 2AD-18RT:MR scenario, two experiments were conducted. One experiment used utilitarian fairness and the other used temporal fairness. The result (Figure 5.26) shows that the HoQ delay of AT-MSs remains within the specified target.



5.26 (a): Utilitarian Fairness

5.26 (b): Temporal Fairness

Figure 5.26: The Effect of Temporal Fairness in Multi-Rate WLAN on Delay Support

5.4 CONCLUDING REMARKS

In this chapter, the comprehensive simulation results confirmed that DRAFT+D can provide QoS support, i.e., a) relative throughput support, b) absolute throughput support, and c) absolute delay support, in various scenarios in a predictable and robust manner. Many scenarios including the levels of throughput requirement, the number of levels of throughput requirement, the time at which MSs become active, number of active MSs, traffic type, mode of operation, the use of RTS/CTS, data rate, and fairness criteria have been evaluated. The ability of DRAFT+D to provide relative throughput support, absolute throughput support, absolute delay support, and admission control support remains unchanged while varying these factors. Three factors that impact the performance of DRAFT+D the most are: a) high error rate condition, b) the use of infrastructure mode. These factors affect the absolute throughput support due to the significant loss of effective channel capacity. This finding suggests that the effective channel capacity must be estimated appropriately as subject to the error rate in the channel and the overhead of relaying traffic in the infrastructure mode to maintain absolute throughput support. The performance of DSG-RT and absolute delay support remain relatively unaffected by these factors.

6.0 CONCLUSIONS AND FUTURE WORK

In this dissertation, we examined the limitations of Quality of Service (QoS) support. We developed an integrated QoS mechanism to support two QoS metrics (throughput and delay) with two QoS models (absolute and relative) under two fairness constraints (utilitarian and temporal fairness) with straight-forward QoS translation in a fully distributed manner. We believe that the integrated solution based on distributed fair scheduling is the answer for overcoming the lack and inadequacy of the provision of QoS in Wireless Local Area Networks (WLANs). We believe that pervasive high-speed wireless data services are at the same time compelling and inevitable. It is just a question of how and when. And if we know the answer to how, then it is only a matter of time until WLANs will become a ubiquitous reality.

The main contribution of this dissertation is a novel distributed QoS mechanism in WLANs, called Distributed Relative/Absolute Fair Throughput with Delay Support (DRAFT+D). Unlike any other QoS mechanism in WLANs, DRAFT+D provides both relative and absolute QoS support for both throughput and delay with safeguard against excessive traffic in a fully distributed manner in the same mechanism at the same time.

To the best of our knowledge, there are no other QoS mechanisms in WLANs that a) provide both fair (relative) and specific (absolute) QoS support, b) consider and support both throughput and delay requirements, c) equip with integrated safeguard against excessive traffic, d) consider and support both utilitarian and temporal fairness, e) achieve multiple types QoS support by modifying only the way to calculate Backoff Interval (BI), f) consider and provide ease of QoS translation, and g) provide all QoS support including safeguard against excessive traffic in a fully distributed manner in the same mechanism at the same time without requiring or collecting any information from or to any other Mobile Stations (MSs). The main features of DRAFT+D include:

- **Integrated QoS Support:** DRAFT+D supports two QoS metrics (throughput and delay) with two QoS models (absolute and relative) under two fairness constraints (utilitarian and temporal fairness) with safeguard against excessive traffic in the same mechanism at the same time. The relative throughput support is achieved by calculating the value of BI based on weight. The default fairness criteria is utilitarian fairness while the temporal fairness can be simply enforced by calculating the value of weight with temporal fairness factor. The absolute QoS support is achieved by allocating sufficient bandwidth to this type of traffic. Excessive traffic is prevented by a) limiting traffic of each flow with Deficit Round Robin (DRR) and b) limiting a new flow with Distributed SafeGuard for Absolute Throughput Support in Relative Throughput Flows (DSG-RT).
- **Fully Distributed in Nature:** All QoS support including relative throughput, absolute throughput, absolute delay, and safeguard support are achieved in a fully distributed manner without requiring the exchange or collection of any information from other MSs. DRAFT+D provides a set of rules that each MS applies independently to achieve the goal of QoS support and predictable performance.
- **Minimum Changes Required, Ease of Implementation, and Low Complexity and Overhead:** DRAFT+D requires minimum changes to the base protocol, i.e., Distributed Coordination Function (DCF), and ease of implementation by achieving all QoS support by modifying only one WLAN's parameter, i.e. BI. Possibility of hybrid flow/class-based queue implementation helps reduce the computational complexity of the protocol and also demonstrates flexibility and ease of implementation. Minimum protocol complexity is justified from the inherited O(1) computational cost on which the mechanism is based, i.e., DRR and Distributed Fair Scheduling (DFS). No additional fields in the frame headers are needed. The use of DSG-RT also reduces overheads from admission control for the entire network. The safeguard mechanism itself as well as other QoS support are very lightweight since it is fully distributed and does not require any collection or distribution of any information from or to any other MSs.
- **Ease of QoS Translation and Minimum Requirement of Admission Control:** To alleviate the problem of complex QoS translation and to ease the QoS translation, all QoS requirements are translated into a common denominator, i.e. quantum rate. The

quantum rate, which represents throughput, is an obvious and natural QoS parameter. The integrated safeguard against excessive traffic helps reduce the need and the burden of admission control and thereby minimizes the requirement of admission control.

- **Robustness, Scalability, Fairness, and Low Variation:** The simulation of diverse and realistic scenarios show promising results and demonstrates several desired properties of DRAFT+D. The performance of DRAFT+D is relatively insensitive to the values of protocol parameters. Relatively wide ranges of values to provide a good and predictable performance of these protocols parameters are easily found and assigned in DRAFT+D. High scalability (in terms of loads and the number of MSs), high robustness (against types of input traffic, WLAN's data rates of a WLAN, and WLAN's parameter), high degree of fairness, and low variation of throughput and delay are integral and critical properties of the mechanism that was developed and examined in this study.

6.1 FUTURE RESEARCH

In this section, we identify the potential directions for future research.

- **Implementation of DRAFT+D in a WLAN's Network Interface Card (NIC):** The next step to validate the functionality of DRAFT+D beyond simulation is its implementation in a real WLAN's NIC and evaluation of its performance on a test-bed network. Given the availability of WLAN's adaptor, which allows changes in WLAN's firmware and driver, e.g, from TEXAS Instrument Corp., DRAFT+D can be implemented. Promising results from research in real test-bed environments will affirm the attractiveness of DRAFT+D while any unexpected results will provide a better understanding of the protocol and enable protocol designers to make necessary adjustments.
- **QoS Provision for Newly Standardizing Protocols:** In the past, popular and widely adopted protocols such as Ethernet, DCF, and TCP/IP, were not equipped with QoS support. When attempting to integrate QoS support to these protocols, it was achieved at the expense of deterioration of protocol efficiency or with an increase in protocol complexity. This complication motivates the concept of QoS-ready mechanisms.

However, including QoS support in any new protocol must not introduce significant complexities and, thus, prevent the new mechanism from being widely adopted at all. We have shown in this research that DRAFT+D is simple to implement with the advantages of integrated QoS support. Newly standardizing protocols such as Ethernet over Power line [98], Ethernet Phoneline [99], and Wireless Personal Area Networks (WPAN) 802.15, could benefit from the simplicity and the integrated QoS support offered by DRAFT+D.

- **Opportunistic Scheduling:** In WLANs, network resources each MS perceives can be different and can change with time. This is due to the time-varying characteristic of wireless channels. In this study, we have not considered the impact of such time-varying characteristics. Further research could investigate a way to exploit good channel conditions in DRAFT+D to improve performance in terms of fairness and QoS support.
- **Multi-hop Networks and End-to-End QoS:** In this study, we consider QoS support in a single-hop WLAN. The next step is to extend the work to multiple hops toward end-to-end QoS support. In the past, multi-hop and end-to-end considerations often incur significant overhead and complexity. We may expect delay rather than throughput to be the major constraint in end-to-end QoS support. One possibility would be for flows to specify more stringent target delays depending on the delays in other hops.
- **Reduce Probability of Collision via Variable IFS:** Collision is among the most fundamental limitations and indicator of overall performance of any Carrier Sense Multiple Access (CSMA)-based mechanisms including a WLAN of 802.11. Currently, class-based InterFrame Space (IFS) has been proposed in Enhanced Distributed Channel Access (EDCA) and expected be become available in the next version of 802.11 WLANs. However, we believe that more research can be conducted to achieve a lower probability of collision in a WLAN. In the present study, we applied DRR as a traffic regulator via the value of Deficit Counter (DC). Further research could examine the way to map the value of DC into the value of IFS. Appendix A provides an initial protocol description and preliminary simulation results to demonstrate the advantage of such a Variable Length IFS (VIFS).
- **Bandwidth Control for Relative Throughput Flows:** Controlling the amount of traffic from the relative throughput class is very important to maintain the ability to

provide absolute throughput and delay support. Without such control, the ability to provide absolute throughput and delay can be jeopardized. In the study, we proposed DSG-RT to limit excessive traffic from the relative throughput class. In this case, the new flow will not be allowed to continue. However, other alternatives to manage a new flow might be preferable. Rather than limiting a new flow, the mechanism can throttle the share of the current flows and allow a additional new flow to be transmitted. Appendix B provides an initial description and preliminary simulation results to demonstrate the advantage of the adaptive bandwidth control mechanism.

- **Distributed Safeguard Against Excessive Traffic from Absolute QoS Class:** Without a safeguard against excessive traffic, specific QoS support is not possible because excessive traffic from the absolute QoS class can be introduced into the network. In the mechanism that we have proposed, DSG-RT is used to limit excessive traffic from the relative throughput class but not from excessive traffic from the absolute QoS class. We believe that further research on safeguarding against this situation is possible by using a method similar to the proposed DSG-RT. Understanding how to prevent excessive traffic from the absolute QoS class would provide a completely integrated and fully distributed QoS mechanism. Integrating these elements with DRAFT+D would provide the most comprehensive solution to QoS support in WLAN. Conducting this research is, therefore, critically important and can potentially resolve the majority of the QoS issues in WLAN. In Appendix C, a initial description, mathematical proof, and preliminary simulation results of a distributed safeguard mechanism against traffic from absolute throughput and delay class are presented.

APPENDIX A

APPLYING DRR ON THE CALCULATION OF VARIABLE IFS

According to [73], different values of IFS among MSs provide two advantages: 1) a lower average probability of collision and 2) a faster progressing of backoff timer. In each round, when a wireless channel becomes idle, a smaller value of IFS allows a flow to decrease its backoff timer sooner than flows with longer IFS. The different values of IFS can be used to reduce the number of relevant contending MSs; hence reducing the probability of collision. The relevant contending MSs are those which have the same value of IFS.

In the current literature, most studies proposed to assign different *fixed* values of IFS for different classes of traffic to provide differentiated QoS support. Essentially, a smaller value of IFS is assigned to higher priority traffic. Therefore, the higher priority traffic will have an advantage to access the channel than the lower priority traffic. This approach can be used to amplify the differentiation ability of static priority-based mechanisms. The advantage of this approach is simplicity and minimal changes to the legacy mechanism. Although this approach can help reduce the number of relevant MSs contending for the network access between high priority and low priority class traffic, the probability of collision can still be high if the number of MSs of the same class is high regardless of the number of MSs in the other classes.

To avoid this problem, i.e. high probability of collision when the number of MSs of the same class is high, we suggest a new mechanism to calculate variable length of IFS ($VIFS$) based on a fair-scheduling mechanism, namely DRR. The main advantages of VIFS are fairness and the significantly reduced probability of collision. This approach can provide fairness because the IFS is calculated based on fair scheduling mechanism. The IFS of MS with a higher value of accumulated resource, in terms of DC, will be shorter than the IFS of another MS with a smaller value of DC. The probability of collision is significantly reduced because the higher the number of possible lengths of IFS, the smaller the probability of collision. Moreover, the probability of collision is not fixed with the number of MSs in each class. Rather, the length of IFS is calculated dynamically.

A.1 MECHANISM DESCRIPTION

The value of IFS of class i (IFS_i) is calculated from the difference between $DIFS$ and the ratio between the value of deficit counter DC and the size of the quantum Q, as shown in Eq. (A.1). Because the value of DC of each MS is likely to be different, the value of the calculated IFS will also be different. According to this equation, four parameters directly influencing the value of IFS_i are $DIFS$, DC_i, α, and Q. Two additional parameters influencing DC_i thus indirectly influencing IFS_i are DC_{min}, and DC_{max}.

$$IFS[i] \quad = \quad DIFS - \alpha \frac{DC[i]}{Q} \tag{A.1}$$

The first term in Eq. (A.1) determines the maximum length of the calculated $IFS[i]$. The maximum length of $IFS[i]$ is obtained when DRAFT+D flows have used up their accumulated quanta, i.e. $DC[i] = DC_{min}$, assuming $DC[i]$ cannot be a negative value ($DC_{min} \geq 0$). Although any arbitrary value can be used, $DIFS$ is judiciously selected to give A-DFAFT+D flows an equal right to access the network in comparison with DCF flows, when $DC[i] = DC_{min}$. As long as, DRAFT+D has some available quanta ($DC[i] > DC_{min}$, DRAFT+D flows will have a smaller $IFS[i]$, thus providing a better advantage than DCF flows. However, once DRAFT+D flows use up their service quanta, other DCF flows can take turns to access the network. This selection provides a backward compatibility to the legacy 802.11.

$DC[i]$ represents unused allowance that is available for data transmission. At times, different traffic flows may have different values of $DC[i]$ depending on their rates of quantum and the rate of successful transmission. Different values of $DC[i]$ result in different values of $IFS[i]$. The higher the value of $DC[i]$ is, the smaller is the value of $IFS[i]$ is. A smaller $IFS[i]$ means a better advantage in accessing the wireless network. Unlike EDCA and other priority-based mechanisms where the value of IFS is fixed with traffic or priority class regardless of how much data a flow has sent, the value of $IFS[i]$ in DRAFT+D is calculated dynamically and instantaneously based on the current value of DC_i as given in Eq. (A.1). Therefore, the value of IFS_i directly reflects unused resources which, in turn, represents an eligibility level to access the network of each TC_i.

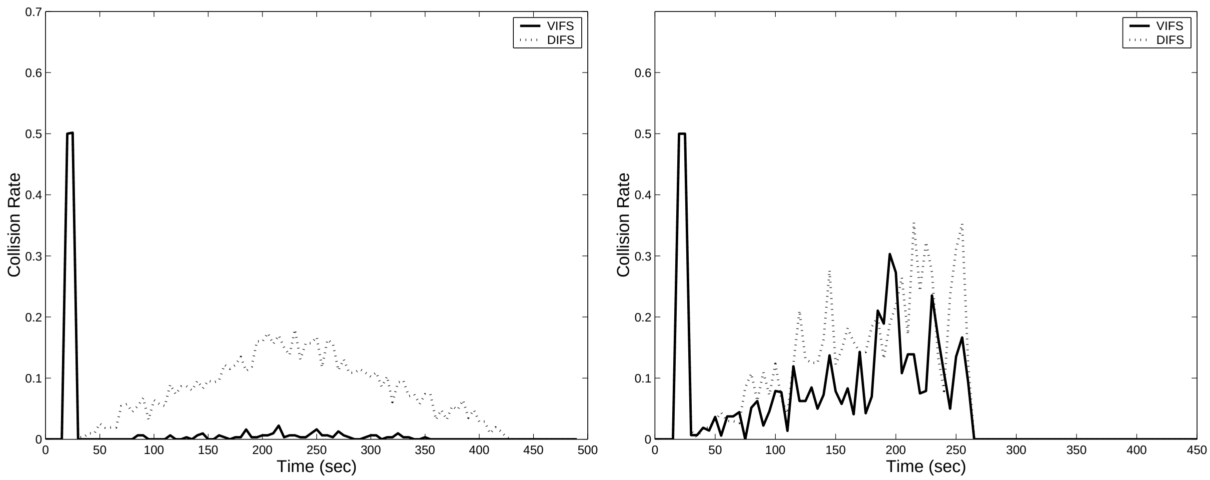
In DRAFT+D, the right to access the channel, $IFS[i]$, is separated from the QoS requirement, or the specified λ_i. Eq. (A.1) shows that $IFS[i]$ is dynamically and instantaneously determined by the current value of $DC[i]$ but not λ_i. According to our simulation results, variable length of $IFS[i]$, significant helps reduce the number of collisions and can be used to improve absolute delay support which is very sensitive to any performance variation or deterioration. The parameter α is a scaling factor that translates the ratio of $Q/DC[i]$ into an appropriate value of $IFS[i]$, i.e. between $PIFS$ and $DIFS$. The granularity of $IFS[i]$ can be appropriately adjusted according to the limitation of the transceiver via parameter α.

Two other parameters that indirectly influence $IFS[i]$ are the minimum DC limit (DC_{min}) and maximum DC limit (DC_{max}). DC_{min} is a parameter used to control the minimum value of $DC[i]$ required for data transmission. That is, the frame of a $TC[i]$ whose $DC[i]$ is lower than the DC_{min} will not be allowed to be transmitted until $DC[i]$ becomes higher than the DC_{min}. Moreover, DC_{min} is also used to control the maximum length of $IFS[i]$. That is, the smaller the value of DC_{min}, the longer the maximum length of $IFS[i]$ is. DC_{max} is a parameter indicating the maximum value of $DC[i]$. DC_{max} is used to control the maximum value of $DC[i]$. When the amount of accumulated quantum reaches DC_{max}, additional quanta will be discarded. Thus, in overloaded situations, it is possible that all stations have accumulated quanta up to DC_{max}. This may reduce the efficacy of VIFS

Ideally, with variable length IFS, we will have a continuous range of IFS for every level of eligibility. Therefore, the probability of collision will be significantly reduced. However, in practice, the sensitivity of transceiver and the latency will be limiting factors to determine the feasible range of operation. With continuous advancements in technology, we believe that the approach can become a viable and effective alternative to significantly reduce the probability of collision, which is a fundamental limitation of WLANs.

A.2 PRELIMINARY RESULTS

To demonstrate the advantage of VIFS, we consider scenario 1AT+19RT where there are 1 Mobile Station with Absolute Throughput Requirement (AT-MS) and 19 Mobile Station with Relative Throughput Requirements (RT-MSs). Each MS require 500 Kbps. In the first scenario, the IFS of all MSs is fixed at DCF IFS (DIFS). In the second scenario, all MSs employ VIFS based on the value of DC. Figure 1.1 shows that the collision rate of AT-MS reduces significantly from 0.18 to 0.02 (maximum value). However, the collision rate of RT-MSs remains the same (Figure 1.1). The reason is that all RT-MSs are backlogged and have the DC at the maximum bound. Consequently, all of their IFS values are approximately the same.



1.1 (a): AT-MS

1.1 (b): RT-MS at $t = 85$

Figure 1.1: Comparison of Collision Rate when Using VIFS and DIFS

APPENDIX B

ADAPTIVE BANDWIDTH AUTO-CONTROL FOR RELATIVE THROUGHPUT FLOWS

B.1 MECHANISM DESCRIPTION

The objective of Adaptive Bandwidth Auto-Control for Relative Throughput Flows (ABAC-RT) is to dynamically throttle the fair share of flows from relative throughput class while maintaining the absolute throughput support. The mechanism makes use of the same condition as DSG-RT. Rather than rejecting a new flow as in DSG-RT as described in Chapter 3, ABAC-RT reduces the value of weight; therefore throttling the fair share of all existing flows from relative throughput class. In this mechanism, weight of each flow in the relative throughput class is reduced to control their bandwidth when additional traffic from relative throughput class is added into the network. After that, the weight of each flow in relative throughput class will increase as the traffic decreases. This adaptive mechanism can be achieved with any standard feedback control mechanism. Here, we implement a simple threshold-based control mechanism and show that this mechanism works relatively well. Other sophisticated or advanced feedback control mechanisms can be substituted to improve the performance.

In the feedback control mechanism, parameter θ decreases by a factor of $\triangle$, e.g. $\triangle = 10^{-5}$, every time the experienced throughput ($\tilde{\lambda}[k] \in RT$) is lower than a threshold. Otherwise, the weight will be increased. This threshold is calculated from the product of a threshold factor (β), ratio of $\frac{\theta}{\omega}$, and the specified quantum rate. The threshold factor (β)

is simply a percentage, e.g. 120%, above the critical condition calculated from Eq. 3.21. The average experienced throughput is calculated using the standard exponential averaging method to smooth out the instantaneous experienced throughput. The instantaneous experienced throughput is calculated from the amount of data sent (in bits) and the time required to transmit these data. As the number of flows of the relative throughput class increases, the share of bandwidth of each flow gradually decreases. When the aggregate throughput of the relative throughput class reduces, the weight is slowly increased to the original value. The pseudo code is shown below.

1. **if** $\left(\tilde{\lambda}[k] < \beta \cdot \frac{\theta}{\omega} \cdot \lambda[k]\right)$
2. $\quad \theta = \theta - \triangle$
3. **else**
4. $\quad \theta = \theta + \triangle$
5. $\phi[k] = \frac{\lambda[k]}{R}$

Four important parameters to implement the adaptive feature are $\tilde{\lambda}[k]$, $\lambda[k]$, θ, and ω. $\tilde{\lambda}[k]$ can be measured independently at each MS while $\lambda[k]$ is the throughput requirement specified by each MS. The parameters θ and ω are assumed to be the standard values for a network. The simplicity of the scheme is that these parameters are available *locally* at each MS. Therefore, each MS can monitor its own experienced throughput and implement the control in a fully distributed way. Using this adaptive mechanism, relative throughput traffic flows can be freely added into the network. In the next section, we will present a preliminary result and sensitivity analysis of ABAC-RT. Additional results on the ABAC-RT can be found in [100]

B.2 PRELIMINARY RESULTS

The idea behind this mechanism is to allow each RT-MS to independently monitor its experienced throughput, compare it against an overload threshold, and then adjust the weight accordingly. If the experienced throughput is lower than the overload threshold, the value of weight (ϕ) will be reduced (via reduction of θ). Figure 2.1 shows that absolute throughput support can be maintained when ABAC is used; however, there is short-term unfairness among RT-MSs where the existing flows receive better throughput than the new flows.

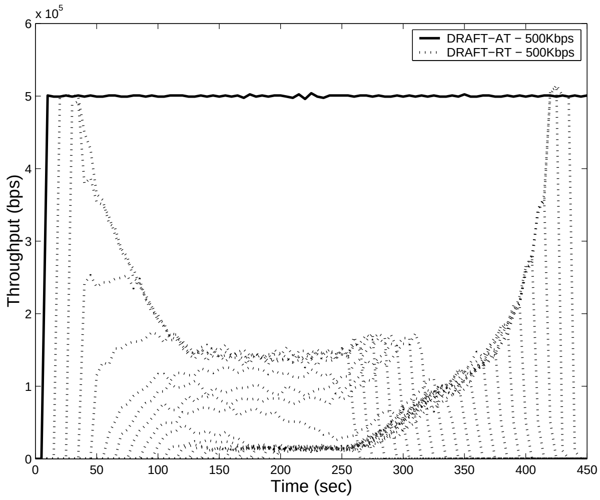
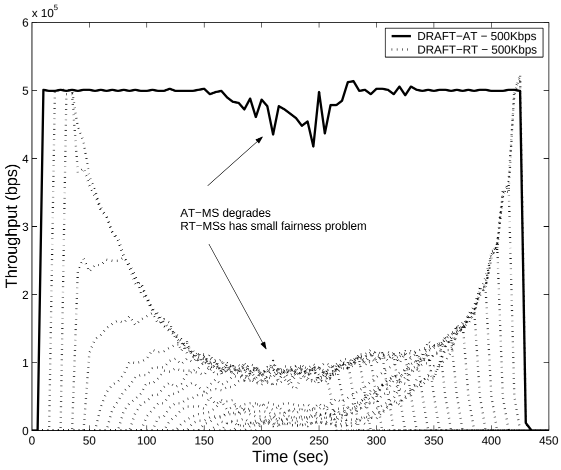


Figure 2.1: Absolute Throughput Support with Adaptive Mechanisms

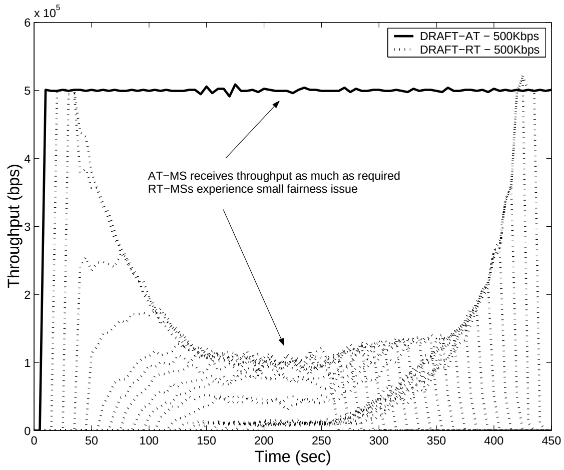
B.2.1 Effect of β and $\triangle$

To maintain the absolute QoS support, the experienced throughput of RT-MSs must be higher than the required threshold. The required threshold can be calculated from the product between the throughput requirement and the overload ratio. If the experienced throughput can be measured perfectly, we can simply use the required threshold as a point to activate the adaptation. However, in the real system, the experienced throughput is not measured perfectly and the adaptive mechanism may need some time to take effect. Because of this, some kind of safety factor should be applied to the required threshold to compensate the inaccuracy of the measurement and provide a buffer for the mechanism to adapt accordingly. Parameter β is a safety factor ($\beta > 1$) that multiplies to the required threshold. The new threshold is served as a point where the adaptive mechanism becomes active to adjust the value of weight.

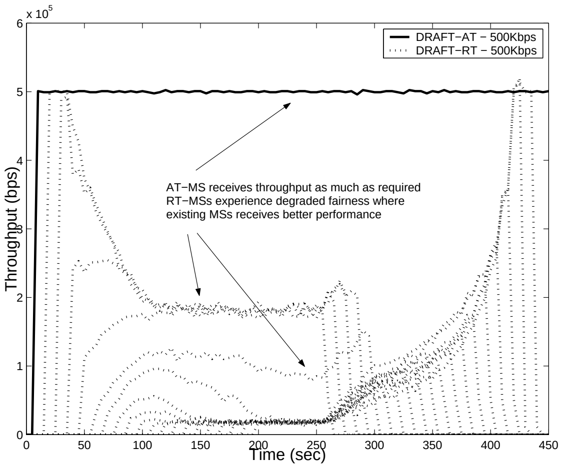
Figure 2.2 shows that β represents a trade-off between the variation of absolute throughput support and fairness of relative throughput support. The higher the value of β is , the smaller is the variation of absolute throughput and lower is the fairness of relative throughput are. The reason is that a higher value of β activates the adaptive mechanism sooner. The sooner the adaptive mechanism is activated, the better is the ability to provide absolute throughput to support. However, the sooner the adaptive mechanism is activated, the larger with the differences between the experienced throughput of existing MSs and new entering MSs be. According to out results, the performance of absolute throughput class is relatively insensitive to the value of β as long as $\beta \geq 1.2$. The value of $\beta = 1.2$ with step size of $10^{-}5$ appears to be sufficient to provide good performance for both absolute throughput and relative throughput class. The sensitivity of step size ($\triangle$) will be discussed next.



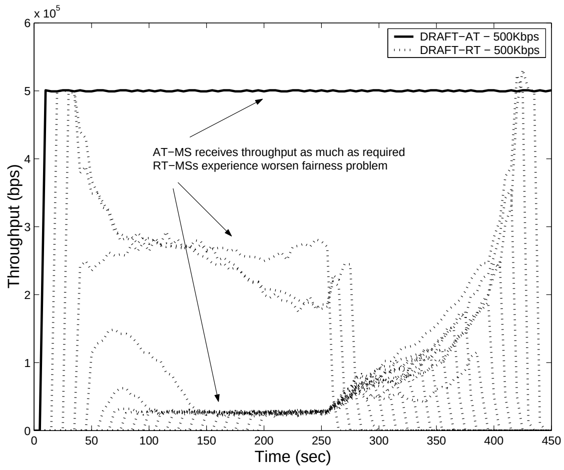
2.2 (a): $\beta = 1.0$



2.2 (b): $\beta = 1.2$



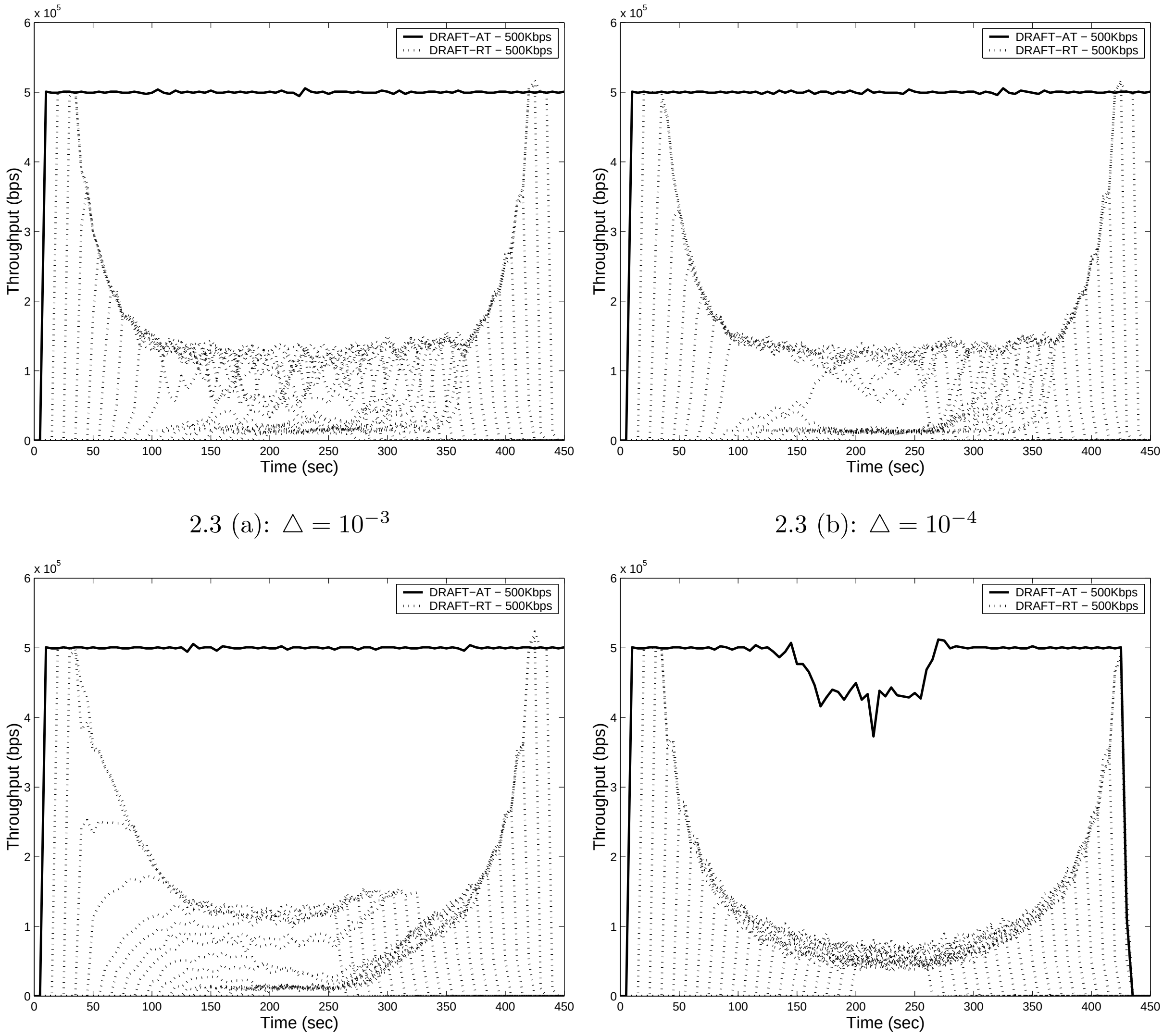
2.2 (c): $\beta = 2.0$



2.2 (d): $\beta = 3.0$

Figure 2.2: Effect of β in Adaptive Mechanism

Figure 2.3 shows that performance of absolute throughput class is relatively insensitive to the value of $\triangle$ as long as $\triangle \geq 10^{-5}$, i.e. sufficiently aggressive to adapt to the .



2.3 (a): $\triangle = 10^{-3}$

2.3 (b): $\triangle = 10^{-4}$

2.3 (c): $\triangle = 10^{-5}$

2.3 (d): $\triangle = 10^{-6}$

Figure 2.3: Effect of $\triangle$ in Adaptive Mechanism

APPENDIX C

DISTRIBUTED SAFEGUARD MECHANISM FOR ABSOLUTE THROUGHPUT FLOWS

C.1 MECHANISM DESCRIPTION

The objective of Distributed SafeGuard for Absolute Throughput Support in Absolute Throughput Flows (DSG-AT) is to limit excessive traffic from AT-MSs while maintaining the absolute throughput support to the current AT-MSs. To achieve this objective, we begin by considering two ways to calculate the bandwidth share of a relative throughput class. First, the share of a flow k with weight $\phi[k]$ from relative throughput class can be calculated from the product of the effective channel capacity (λ_e), de-escalating parameter (θ), and the normalized weight ($\frac{\phi[k]}{\sum_{\forall i} \phi[i]}$), as shown earlier in Eq. (C.1). Second, the fair share of a flow k with weight $\phi[k]$ from relative throughput class can be calculated from the product of the available channel capacity (λ_a), de-escalating parameter (θ), and the normalized weight of the relative throughput class, as shown in Eq. (C.2). The available channel capacity (λ_a) is the difference between the effective channel capacity and the bandwidth required for the absolute throughput class as shown in Eq. (C.3).

$$\overline{\lambda}[y \in RT] = \lambda_e \cdot \theta \cdot \frac{\phi[y]}{\sum_{\forall i} \phi[i]} \tag{C.1}$$

$$\overline{\lambda}[k \in RT] = \lambda_a \cdot \theta \cdot \frac{\phi[k]}{\sum_{\forall j \in RT} \phi[j]} \tag{C.2}$$

$$\lambda_a = \lambda_e - \sum_{\forall i \in AT} \lambda[i] \tag{C.3}$$

According to Eq. (C.2), there are 3 known variables and 2 unknown variables. $\phi[k]$ is calculated from a locally specified throughput requirement. The parameter θ is a standard value. $\overline{\lambda}$ is a local measurement of experienced throughput at each MS. Two unknown variables are the available bandwidth (λ_a) and the normalized weight of the relative throughput class ($\frac{\phi[k]}{\sum_{\forall j \in RT} \phi[j]}$). Due to the fairness property of DRAFT+D, we can solve for these two-unknowns by injecting two flows with different weight values into the network and measuring the corresponding experienced throughput.

Assume, two new flows, $k1, k2$ from RT-MSs are, one at a time, injected into a WLAN and each MS measure the experienced throughput. We assume that the available bandwidth and total sum of the weight from relative throughput class does not change during this process. The experienced throughput of these flows can be represented as show in Eq. (C.4) and Eq. (C.5). Let $\sum\limits_{\forall j \in RT} \phi[j]$ be the total sum of the weight from relative throughput class before addition of a new flow.

$$\overline{\lambda}[k1] = \lambda_a \cdot \theta \cdot \frac{\phi[k1]}{\sum\limits_{\forall j \in RT} \phi[j] + \phi[k1]} \tag{C.4}$$

$$\overline{\lambda}[k2] = \lambda_a \cdot \theta \cdot \frac{\phi[k2]}{\sum\limits_{\forall j \in RT} \phi[j] + \phi[k2]} \tag{C.5}$$

With few steps of algebraic manipulation to solve Eq. (C.4) and Eq. (C.5), we can calculate the sum of the weight from relative throughput class from Eq. (C.6), available bandwidth from Eq. (C.7), the sum of throughput requirement from absolute throughput class including the new flow from Eq. (C.8), and the sum of throughput requirement from relative throughput class from Eq. (C.9).

$$\sum_{\forall j \in RT} \phi[j] = \frac{\phi[k2] \cdot \left(\overline{\lambda}[k1] - \overline{\lambda}[k2]\right)}{\overline{\lambda}[k2] - \frac{\phi[k2]}{\phi[k1]} \cdot \overline{\lambda}[k1]} \tag{C.6a}$$

$$= \frac{\phi[k1] \cdot \left(\overline{\lambda}[k2] - \overline{\lambda}[k1]\right)}{\overline{\lambda}[k1] - \frac{\phi[k1]}{\phi[k2]} \cdot \overline{\lambda}[k2]} \tag{C.6b}$$

$$\lambda_a = \frac{\overline{\lambda}[k1] \cdot \Big(\sum_{\forall j \in RT} \phi[j] + \phi[k1] \Big)}{\theta \cdot \phi[k1]} \tag{C.7a}$$

$$= \frac{\overline{\lambda}[k2] \cdot \Big(\sum_{\forall j \in RT} \phi[j] + \phi[k2] \Big)}{\theta \cdot \phi[k2]} \tag{C.7b}$$

$$\sum_{\forall i \in RT} \lambda[i] = \frac{\lambda[k1]}{\overline{\lambda}[k1]} \cdot \lambda_a \tag{C.8a}$$

$$= \frac{\lambda[k2]}{\overline{\lambda}[k2]} \cdot \lambda_a \tag{C.8b}$$

$$\sum_{\forall i \in AT} \lambda[i] \quad = \quad \lambda_e - \lambda_a + \lambda[k] \tag{C.9}$$

In short, all the required parameters can be calculated independently and locally at each MS. We substitute the fair share of testing flows ($\overline{\lambda}$) in Eq. (3.17) with the measurement of the experienced throughput ($\tilde{\lambda}$) and evaluate whether the condition to provide absolute support can be maintained with an addition of a new absolute throughput flow. If Eq. (3.17) holds, the new absolute throughput flow will be allowed to remain active.

In conclusion, DSG-AT can determine whether a new AT-MS can become active without jeopardizing the ability to provide absolute throughput for the existing AT-MSs. The mechanism is achieved by injecting two small flows with different throughput requirements into the network, and measuring their experienced throughput, and determining the overload condition in Eq. (3.17). Unlike any traffic limiting mechanism, the main advantage of DSG-AT is that it is *fully distributed* and is performed *independently* at each MS.

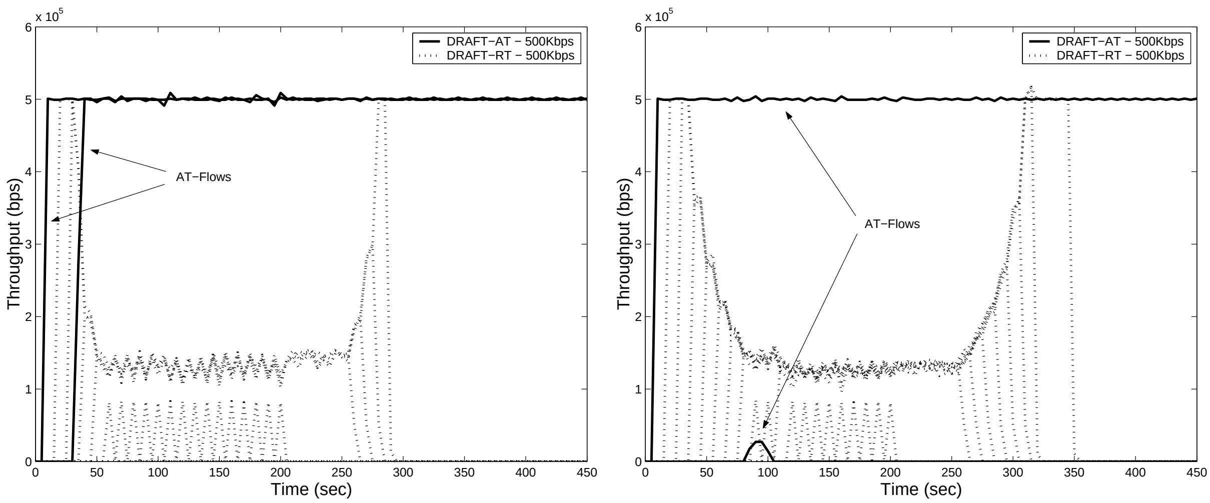
C.2 PRELIMINARY RESULTS

To demonstrate the performance DSG-AT, we consider two scenarios where a new AT-MS of 500 Kbps starts when the available bandwidth is: a) sufficient, at $t = 35$, and b) not sufficient, at $t = 85$. Figure 3.1 (a) shows that a new AT-MS starting at $t = 35$ can be accepted since available bandwidth at this time is still sufficient according to the calculation of available bandwidth by Eq. (3.17), as shown below.

$$\begin{aligned} \theta \cdot \sum_{\forall j \in RT} \lambda[j] + \omega \cdot \sum_{\forall i \in AT} \lambda[i] &\leq \omega \cdot \lambda_e \\ 1 \times (2 \times 500) + 5 \times (2 \times 500) &\leq 5 \cdot 1500 \\ 6000 &\leq 7500 \end{aligned}$$

When a new AT-MS starts a flow at $t = 85$, Figure 3.1 (b) shows that the new AT-MS is not accepted since the available bandwidth is no longer sufficient at this point according to the calculation of available bandwidth, shown below.

$$\begin{aligned} \theta \cdot \sum_{\forall j \in RT} \lambda[j] + \omega \cdot \sum_{\forall i \in AT} \lambda[i] &\leq \omega \cdot \lambda_e \\ 1 \times (7 \times 500) + 5(\times 2 \times 500) &\leq 5 \times 1500 \\ 13000 &\not\leq 7500 \end{aligned}$$



3.1 (a): New AT-MS at $t = 35$

3.1 (b): New AT-MS at $t = 85$

Figure 3.1: Safeguard against Excessive Traffic with DSG-AT

BIBLIOGRAPHY

[1] D. Molta, "Wi-Fi Tales'," *Network Computing*, pp. 44–62, 2004.

[2] A. Vance, "WLAN - Trands and Analysis," *Business Commiunications Reviews*, pp. 42–47, 2003.

[3] G. Vannuccini H.L. Truong, "The IEEE 802.11e MAC for Quality of Service in Wireless LANs," in *Proc. International Conference on Advances in Infrastructure for E-Electronic, e-Business, e-Education, e-Science, e-Medicine on the Internet (SSGRR 2003)*, 2003.

[4] Frost & Sullivan, "World Wireless LAN Markets," Tech. Rep., 2001.

[5] Cisco System, "Cisco Announces New Independent Study Demonstrating Productivity Gains Achieved Through Wireless LAN Use," Tech. Rep., 2001.

[6] A. Vance, "WLAN - Trands and Analysis," *Business Commiunications Reviews*, 2003.

[7] PR Newswire Association LLC., "Cisco Posts Strongest Wireless LAN Hardware Quarter Yet Overall Market Up," *PR Newswire Association*, 2004.

[8] Inc. Wireless News, "Dell'Oro Group: WLAN Market Will Grow to $4.3 Billion in 2009," *Wireless News*, 2005.

[9] Inc. Business Wire, "Shelron's ActivShopper Goes Wireless; Targeting 1.5 Billion Users Globally - 25% of the Planet's Population," *Business Wire*, 2005.

[10] Interanational Data Corp, "Stready Growth for Hotspot Market'," *New Straits Times (Malasia)*, 2003.

[11] In-Stat/MDR, "WiFi and Linksys Rule Home Networking Hardware Market in 2002," Tech. Rep., 2002.

[12] PR Newswire Association LLC., "Atheros Communications Ranked as Silicon Valley's Fastest Growing Private Company by the Silicon Valley," *PR Newswire*, 2004.

[13] J. Kaczman, "Wi-Fi hotspot networks sprout like mushrooms," *IEEE Spectrum*, September 2002.

[14] S. Weinstein, “Wireless LAN and Cellular Mobile – Competition and Cooperation,” Tech. Rep., Technical Talk: IEEE New Jersey Coast Section, 2002.

[15] Inc. Business Wire, “SyChip Introduces IEEE 802.11g Embedded Module for Mobile Handset Market,” *Business Wire*, 2005.

[16] Reed Elsevier Inc., “Terrible Twos? Dual cellular/Wi-Fi schemes face dual handoff, powering issues,” *Wireless Week*, 2005.

[17] Inc Crain Communications, “Cisco to make WLAN mobile handset,” *RCR Wireless News*, 2003.

[18] PR Newswire Association LLC., “WiFi VoIP Handset Market Hits $45M in 2004, Healthy Growth Projected Through 2009,” *PR Newswire*, 2005.

[19] C. Anderson, “The Wi-fi Revolution,” *Wired Magazine*, 2003.

[20] M. Paolini and R. Pow, “Public Wireless LAN Access: US Market Forecasts 2002-2007,” Tech. Rep., Analysys Research Limited, 2002.

[21] A. Kacker, “Opportunities in WLAN: pie in the sky?,” Tech. Rep., Analysys Research Limited, 2003.

[22] Intel Inc., “Meeting the Demands of the Digital Home with High-Speed Multi-Hop Wireless Networks,” *Intel Technology Journal*, Nevember 2002.

[23] M. Lewis, “VoWLAN: the WLAN killer app?,” *BluePrint Wi-Fi*, 2002.

[24] Vocera Communications, “Vocera Communications System: Wearable Instant Voice Communication,” Tech. Rep., 2003.

[25] M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, “Integration of 802.11 and Third-Generation Wireless Data Networks,” in *503-512*, March 2003, vol. 1.

[26] M. Kende, “VoIP - the opportunities and challenges,” Tech. Rep., Analysys Research Limited, 2003.

[27] IETF, “Integrated Services Working Group,” Tech. Rep.

[28] IETF, “Differentiated Services Working Group,” Tech. Rep.

[29] I. Mahadevan and K. M. Sivalingam, “Architecture and experimental framework for supporting QoS in wireless networks using differentiated services,” *MONET*, vol. 6, no. 4, pp. 385–395, 2001.

[30] G. Wu, M. Inoue, H. Murakami, and Y. Hase, “156 Mbps Ultrahigh-Speed Wireless LAN Prototype in the 38GHz Band,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, November 2001, vol. 6, pp. 3573 – 3578.

[31] Inc Computerworld, "802.11ns Pregame Show," *Computerworld*, 2004.

[32] W. Pattara-atikom, P. Krishnamurthy, and S. Banerjee, "Distributed Mechanisms for Quality of Service in Wireless LANs," *IEEE Wireless Communications Magazine*, vol. 10, no. 3, pp. 26 – 34, June 2003.

[33] D. Gu and J. Zhang, "QoS enhancement in IEEE 802.11 wireless local area networks," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 120 – 124, June 2003.

[34] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic Scheduling with QoS Requirements in Time-varying Wireless Channels," *Computer Networks*, vol. 4, no. 41, pp. 451–474, March 2003.

[35] A. Sabharwal B. Sadeghi, V. Kanodia and E. Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," in *in Proc. International Conference on Mobile Computing and Networking (Mobicom)*, Atlanta, Georgia, USA, September 2002, pp. 24 – 35.

[36] B. P. Crow, I. Widjaja, L. G. Kim, and P. T. Sakai, ""'IEEE 802.11 Wireless Local Area Networks"'," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 116–26, 1997.

[37] M. Johnsson, "HiperLAN/2 - The Broadband Radio Transmission Technology Operating in the 5 GHz Frequency Band," Tech. Rep., HiperLAN/2 Global Forum, 1999.

[38] D. Sobirk and J. Karlsson, "A Survey of Wireless ATM MAC Protocols," in *In Proc. of the IFIP TC6 / WG6.3 & WG7.3 International Conference on the Performance and Management of Complex Communication Networks (PMCCN)*, 1997, vol. 120, pp. 191 – 210.

[39] D. Qiao and K.G. Shin, "UMAV: A Simple Enhancement to the IEEE 802.11 DCF," in *In Proc. of the 36th Hawaii International Conference on System Sciences (HICSS)*, Big Island, Hawaii, January 2003.

[40] A. C. V. Gummalla and J. O. Limb, "Wireless Medium Access Control Protocols," in *IEEE Communications. Surveys and Tutorials*, Second Quarter 2000, vol. 3 of *2*.

[41] J. Sanchez, R. Martinez, and M. W. Marcellin, "A survey of MAC protocols proposed for wireless ATM," *IEEE Network Magazine*, vol. 11, no. 6, pp. 52–62, December 1997.

[42] E. Ayanoglu, K. Y. Eng, and M. J. Karol, "Wireless ATM: limits, challenges, and proposals," *IEEE Personal Communications*, vol. 3, no. 4, pp. 18–34, August 1996.

[43] P. L. Montessoro and D. Pierattoni, "Advanced research issues for tomorrow's multimedia networks," in *In Proc. International Conference on Information Technology: Coding and Computing (ITCC)*, Las Vegas, Nevada USA, April 2001, pp. 336–340.

[44] A. Veres, A.T. Campbell, M. Barry, and Li-Hsiang Sun, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 19, no. 10, pp. 2081 – 2093, October 2001.

[45] S. T. Shue and T. F. Shue, "A Bandwidth Allocation/Sharing/Extension Protocol for Multimedia Over IEEE 802.11 Ad Hoc Wireless LANs," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 19, no. 10, pp. 2065–2080, October 2001.

[46] T. Suzuki and S. Tasaka, "Performance evaluation of priority-based multimedia transmission with the PCF in an IEEE 802.11 standard wireless LAN," in *In Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, September - October 2001, vol. 2, pp. G–70 – G–77.

[47] A. Grilo and M. Nunes, "Performance Evaluation of IEEE 802.11e," in *In Proc. IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, September 2002, vol. 1, pp. 511 – 517.

[48] G. Anastasi and L. Lenzini, "QoS provided by the IEEE 802.11 wireless LAN to advanced data applications: a simulation analysis," *ACM Wireless Networks*, vol. 6, no. 2, pp. 99 – 108, 2000.

[49] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–47, March 2000.

[50] W. Pattara-atikom, S. Banerjee, and P. Krishnamurthy, "Starvation Prevention and Quality of Service in Wireless LANs," in *In Proc. IEEE International Wireless Personal Multimedia Communications (WPMC)*, Honolulu, Hawaii USA, October 2002, vol. 3, pp. 1078 – 1082.

[51] A. Kopsel and A. Wolisz, "Voice transmission in an IEEE 802.11 WLAN based access network," in *In Proc. ACM International Workshop on Wireless Mobile Multimedia (WoWMoM)*, Rome, Italy, July 2001, pp. 23 – 32.

[52] J. Deng and R. S. Chang, "A priority scheme for IEEE 802.11 DCF access method," *IEICE Transactions on Communications*, vol. E82-B, no. 1, pp. 96–102, 1999.

[53] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *In Proc. Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, Anchorage, AK, April 2001, vol. 1, pp. 209 – 218.

[54] M. Benvenisite, G. Chesson, W. Diepstraten, M. Hoeben, A. Singla, H. Teunissen, and M. Wentink, "EDCF Proposed Draft Text," Tech. Rep., IEEE working document 802.11-01/131r1, 2001.

[55] A. Banchs, X. Perez, M. Radimirsch, and H. J. Stuttgen, "Service differentiation extensions for elastic and real-time traffic in 802.11 wireless LAN," in *In Proc. IEEE*

Workshop on High Performance Switching and Routing (HPSR), Dallas, TX USA, May 2001, pp. 245–249.

[56] A. Ayyagari, Y. Bernet, and T. Moore, "IEEE 802.11 Quality of Service - White Paper," Tech. Rep., IEEE, February 2000.

[57] K. Kim, S. Shin, and K. Kim, "A Novel MAC Scheme for Prioritized Services in IEEE 802.11a Wireless LAN," in *In Proc. IEEE International Conference on ATM and high speed intelligent Internet (ICATM)*, Seoul South Korea, April 2001, pp. 196 – 199.

[58] M. Barry, A.T Campbel, and A. Veres, "Distributed Control Algorithms for Service Differentiation in Wireless Packet Networks," in *In Proc. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, Anchorage, AK USA, April 2001, vol. 1, pp. 582 – 590.

[59] M. Benveniste, "TCMA Proposed Draft Text," Tech. Rep., IEEE working document 802.11-01/117r2, 2001.

[60] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Distributed Multi-Hop Scheduling and Medium Access with Delay and Throughput Constraints," in *In Proc. International Conference on Mobile Computing and Networking (MOBICOM)*, Rome, Italy, July 2001, pp. 200 – 209.

[61] T. Janczak, J. Wozniak, O. Martikainen, J. Porras, and J. Hyvarinen, "Priority-based transmission in IEEE 802.11 networks," in *In Proc. IFIP TC6/WG6.8 Working Conference on Emerging Personal Wireless Communications*, Norwell, MA, August 2001, vol. 195, pp. 109 – 120, IFIP TC6/WG6.8 Working Conference on Personal Wireless Communications (PWC'2001). Kluwer Academic Publishers.

[62] J. Deng and R. S. Chang, "A priority scheme for IEEE 802.11 DCF access method," *IEICE Transactions on Communications*, vol. E82-B, no. 1, pp. 96–102, 1999.

[63] M. Benveniste, "Tiered contention multiple access' (TCMA), a QoS-based distributed MAC protocol," in *In Proc. The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2002, vol. 2, pp. 598–604.

[64] J. Thomas, "802.11e brings QoS to WLANs," *Network World*, 2003.

[65] W. T. Chen, Yu-Ming Lin, and Shou-Chih Lo, "Priority-Based Contention Control in IEEE 802.11 Wireless LANs," in *In Proc. International Workshop on Communications Quality & Reliability (CQR)*, Okinawa, Japan, May 2002, pp. 260–264.

[66] L. Romdhani, Q. Ni, and T. Turletti, "AEDCF: Enhanced Service Differentiation for IEEE 802.11 Wireless Ad-Hoc Networks," in *2003*.

[67] IEEE 802.11e/D5.0, "Draft Supplement: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)," Tech. Rep., IEEE 802.11 WG, Aug.

[68] S. Mangold, S. Choi, G. Hiertz, O. Klein, and B. Walke, "Analysis of IEEE 802.11 for QoS Support in Wireless LANs," *2003*, vol. 10, pp. 2–12, dec.

[69] S. Choi, J. del Prado, S. Shankar, and S. Mangold, "IEEE 802.11e Contention-Based Channel Access (EDCF) Performance Evaluation," in *In Proc. IEEE International Conference on Communications (ICC)*, Anchorage, Alaska, USA, May 2003, vol. 2, pp. 1151 – 1156.

[70] A. Duric, "VoIP on WLAN - QoS issues and VoIP Specifics," in *In Proc. ETSI Workshop on Speech Processing, Transmission and Quality Aspects (STQ)*, Sophia Antipolis, France, February 2003.

[71] Anders Lindgren, Andreas Almquist, and Olov Schelén, "Quality of Service Schemes for IEEE 802.11 Wireless LANs - An Evaluation," *In Special Issue of the Journal on Special Topics in Mobile Networking and Applications (MONET) on Performance Evaluation of Qos Architectures in Mobile Networks*, vol. 8, no. 3, pp. 223–235, June 2003.

[72] J. Weinmiller, H. Woesner, and A. Wolisz, "Analyzing and improving the IEEE 802.11-MAC protocol for wireless LANs," in *In Proc. International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunications System (MASCOTS)*, San Jose, CA, USA, February 1996, p. 200.

[73] J.W. Robinson and T.S. Randhawa, "Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function," *22*, , no. 5, pp. 917 – 928, 2004.

[74] Y. Xiao, "Enhanced DCF of IEEE 802.11e to Support QoS," in *In Proc. IEEE Wireless Communications and Networking Conference*, March 2003, vol. 2, pp. 1291– 1296.

[75] D. He and C. Shen, "Simulation Study of IEEE 802.11e EDCF," in *2003*.

[76] M. Malli, Q. Ni, T. Turletti, and C. Barakat, "Adaptive Fair Channel Allocation for QoS Enhancement in IEEE 802.11 Wireless LANs," in *In Proc. IEEE International Conference on Communications (ICC0*, Paris, France, June 2004.

[77] P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek, and X. Cheng, "Using IEEE 802.11e MAC for QoS over Wireless," in *In Proc. IEEE International Performance, Computing, and Communications Conference (IPCCC)*, Phoenix, USA, April 2003, pp. 1097–2641.

[78] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet Switching Networks," in *Proc. IEEE*, October 1995, vol. 83 of *10*, pp. 1374–1396.

[79] A. Banchs and X. Perez, "Distributed weighted Fair queuing in 802.11 wireless LAN," in *In Proc. IEEE International Conference on Communications (ICC)*, New York, NY USA, April 2002, vol. 5, pp. 3121– 3127.

[80] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed fair scheduling in a wireless LAN," in *In Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Boston, MA, August 2000, pp. 167 – 178.

[81] A. Banchs and X. Perez, "Providing throughput guarantees in IEEE 802.11 wireless LAN," in *In Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Orlando, FL, March 2002, vol. 1, pp. 130– 138.

[82] S.J. Golestani, "A self-clocked fair queueing scheme for broadband applications," in *In Proc. IEEE Networking for Global Communications (INFOCOM)*, Toronto, Ontario, Canada, June 1994, vol. 2, pp. 636–646.

[83] H.Yamada, H.Morikawa, and T.Aoyama, "Decentralized control mechanism suppressing delay fluctuation in wireless LANs," in *In Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Orlando, FL, October 2003, vol. 2, pp. 801– 805.

[84] G. W. Wong and R. W. Donaldson, "Improving the QoS Performance of EDCF in IEEE 802.11e Wireless LANs," in *In Proc. IEEE Pacific Rim Conference on Communications, Computers and signal Processing (PACRIM)*, Victoria, BC, Canada, August 2003, vol. 1, pp. 392– 396.

[85] R. Jayaparvathy and S. Srikanth, "Delay Performance Analysis of 802.11," in *Proc. 9th Asia Pacific Conference on Communication APCC*, Malaysia, 2003.

[86] W. T. Chen, B. B. Jian, and S. C. Lo, "An adaptive retransmission scheme with QoS support for the IEEE 802.11 MAC enhancement," in *In Proc. IEEE Vehicular Technology Conference (VTC-Spring)*, Birmingham, AL, September 2002, vol. 1, pp. 70– 74.

[87] J. L. Sobrinho and A. S. Krishnakumar, "Real-time traffic over the IEEE 802.11 medium access control layer," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 172–87, 1996.

[88] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Distributed Priority Scheduling and Medium Access in Ad Hoc Networks," *ACM Wireless Networksl (WINET)*, vol. 8, no. 5, pp. 455 – 466, September 2002.

[89] S. Keshav, *An Engineering Approach to Computer Network*, Addison-Wesley, 1997.

[90] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, "Understanding TCP fairness over wireless LAN," in *INFOCOM '03. Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, 2003, p. 863872.

[91] Jong-Deok Kim and Chong-Kwon Kim, "Performance analysis and evaluation of IEEE 802.11e EDCF," *Wireless Communications & Mobile Computing*, vol. 4, no. 1, pp. 55–74, February 2004.

[92] P. Kyasanur and N.H. Vaidya, "Detection and handling of MAC layer misbehavior in wireless networks," in *Proc. 2003 International Conference on Dependable Systems and Networks*, 2003, pp. 173 – 182.

[93] M. Raya, J.-P. Hubaux, and I. Aad, "DOMINO: A system to detect greedy behavior in IEEE 802.11 hotspots," in *ACM MobiSys*, Boston MA, USA, 2004.

[94] M. Cagalj, S. Ganeriwal, I. Aad, and J.-P. Hubaux, "On selfish behavior in CSMA/CA networks," in *INFOCOM '05. Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, Miami FL, USA, 2005.

[95] P. Goyal and H. M. Vin, "Generalized Guaranteed Rate Scheduling Algorithms: A Framework," *5*, , no. 4, pp. 561–571, 1997.

[96] J. Arauz, S. Banerjee, and P. Krishnamurthy, "MAITE a Scheme for Improving the Performance of TCP over wireless channels," in *In Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Atlantic City, NJ, October 2001, vol. 1, pp. 252 – 256.

[97] W. Pattara-atikom, P. Krishnamurthy, and S. Banerjee, "Comparison of Distributed Fair QoS Mechanism in Wireless LANs," in *In Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Francisco, CA, December 2003, vol. 1, pp. 553 – 557.

[98] M. K. Lee, R. E. Newman, H. A. Latchman, S. Katar, and L. Yonge 3, "HomePlug 1.0 powerline communication LANs - Protocol Description and Performance Results," *International Journal of Communication Systems*, vol. 16, no. 5, pp. 447–473, 2002.

[99] T. Zahariadis, K. Pramataris, and N. Zervos, "A comparison of Competing Broadband In-home technologies," *ELECTRONICS & COMMUNICATION ENGINEERING JOURNAL*, vol. 16, no. 5, pp. 447–473, 2002.

[100] W. Pattara-atikom, S. Banerjee, and P. Krishnamurthy, "A-DRAFT: An Adaptive QoS Mechanism to Support Absolute and Relative Throughput in 802.11 Wireless LANs," in *In Proc. ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, Venice, Italy, October 2004, pp. 117 – 126.