UNIVERSITY OF CINCINNATI

Date: _____

I, _____,
hereby submit this work as part of the requirements for the degree of:

in:

It is entitled:

This work and its defense approved by:

Chair: _____

Dynamic Adaptive Resource Allocation Scheme for Multimedia Services in Wireless and Mobile Networks

A thesis submitted to the
Division of Research and Advanced Studies
of

The University of Cincinnati
in partial fulfillment
of the requirements for the degree of
Master of Science

in the
Department of Electrical and Computer Engineering
and Computer Science
of

The College of Engineering
March, 2006

by

Pawan K. Lakshmanan

B.E. (Electrical & Electronics Engineering)

PSG College of Technology,

Bharathiar University, Coimbatore, India, May 2003

Thesis Advisor and Committee Chair: **Dr. Qing-An Zeng**

Abstract

As a wide variety of services are being made available for users in a wireless and mobile network, the type of traffic is not limited to voice anymore, rather applications such as web browsing, email, video messaging, etc., which generate non-voice traffic, are also gaining increasing popularity. With the growth of these bandwidth hungry multimedia services, there is an immediate need for a resource management scheme which exploits the flexible bandwidth requirements of these new traffic classes.

In this thesis, we present and analyze a novel resource management scheme, namely Dynamic Adaptive Resource Allocation (DARA), which is specially tailored to support the emerging classes of multimedia traffic in a wireless and mobile network. From the extensive simulation results, it is evident that our scheme achieves better performance with regard to traditional Quality of Service (QoS) parameters such as the call blocking probabilities of originating and handoff calls while significantly reducing the bandwidth reallocation frequency experienced by a call when compared to similar flexible resource allocation schemes. Higher priority handoff calls are also efficiently handled in our proposed scheme by a combined approach of channel reservation and bandwidth borrowing from ongoing calls at heavy traffic loads.

Acknowledgements

This work would not have been possible without the constant guidance and support of my thesis advisor, Dr. Qing-An Zeng. During the course of my graduate studies at the University of Cincinnati, he has been an invaluable mentor who was always available for helpful suggestions and critical comments to help me grow as a researcher. I would like to sincerely thank him for all the time and efforts that he devoted towards me.

I would also like to express my gratitude to the thesis defense committee members, Dr. Dharma P. Agrawal and Dr. Chia-Yung Han, for their constructive review of my research work. Thanks are also due to my colleagues at the Wireless and Mobile Networking Laboratory (WMN Lab.), for their camaraderie and professional assistance during my studies.

Last but definitely not the least; I would like to sincerely thank my parents and my sisters for their unconditional love and support throughout my educational years.

Contents

List of Figures	3
List of Tables	5
1 Introduction	6
1.1 History of Wireless and Mobile Networks	7
1.1.1 Early Wireless and Mobile Networks	7
1.1.2 First Generation (1G) Cellular Networks	8
1.1.3 Second Generation (2G) Cellular Networks	9
1.1.4 Third Generation (3G) Cellular Networks	11
1.1.5 Future Trends: Fourth Generation (4G) Cellular Networks	13
1.2 Multimedia Services in the 3G Cellular System	15
1.2.1 Existing Technologies for Multimedia Services	15
1.2.2 Classification of 3G Service Classes	15
1.2.3 Resource Management Issues for Multimedia Services	17
1.2.3.1 Challenges	17
1.2.3.2 Overview of Related Work and Our Contributions	18
1.3 Organization of the Thesis	20
2 Related Work	21
2.1 Introduction	21
2.2 Traditional Resource Allocation Schemes	22
2.2.1 Complete Sharing (CS) Scheme	22
2.2.3 Fixed Partitioning (FP) Scheme	23
2.3 Static Resource Allocation Schemes	24
2.4 Flexible Resource Allocation (FRA) Schemes	26
3 Preliminaries	29
3.1 Introduction	29
3.2 Important Probability Distributions	29
3.2.1 Exponential Distribution	30
3.2.2 Poisson Distribution	31
3.2.3 Pareto Distribtuion	33

3.3	Handoff Process	34
3.3.1	Types of Handoff	34
3.3.2	Handoff Criteria	35
3.3.4	Handoff Schemes	37
4	Dynamic Adaptive Resource Allocation (DARA) Scheme	39
4.1	Introduction	39
4.2	System Model	41
4.2.1	System Parameters	41
4.2.2	Traffic Model	45
4.3	DARA Bandwidth Allocation Algorithm	48
4.3.1	Originating Call Arrival Handling	48
4.3.2	Handoff Call Arrival Handling	49
4.3.3	Call Departure Handling	51
4.3.4	Bandwidth Degrade Process	54
4.3.5	Bandwidth Upgrade Process	54
4.4	QoS Metrics	55
4.4.1	Average Degraded Bandwidth (ADB)	55
4.4.2	Bandwidth Reallocation Frequency (BRF)	56
5	Performance Analysis	57
5.1	Introduction	57
5.2	Simulation Model	58
5.2.1	Programming Environment	58
5.2.2	Simulation Parameters	58
5.3	Simulation Results and Discussions	61
5.3.1	Simulation Set 1 (SS1) Results	62
5.3.2	Simulation Set 2 (SS2) Results	67
5.3.3	Simulation Set 3 (SS3) Results	72
5.3.4	Simulation Set 4 (SS4) Results	76
5.3.5	Simulation Set 5 (SS5) Results	79
6	Conclusions and Future Work	84
6.1	Conclusions	84
6.2	Future Work	85

List of Figures

1.1	Evolution of Cellular Systems	14
2.1	Complete Sharing (CS) Scheme	23
2.2	Fixed Partitioning (FP) Scheme	24
2.3	Dual Threshold Bandwidth Reservation (DTBR) Scheme	26
3.1	Example of Handoff in Cellular Networks	35
4.1	System Model Diagram	45
4.2	Flowchart for Call Arrival Handling	52
4.3	Flowchart for Call Departure Handling.....	53
5.1	(SS1) Comparison of the Blocking Probability for Class 1 Traffic.....	62
5.2	(SS1) Comparison of the Blocking Probability for Class 2 Traffic.....	63
5.3	(SS1) Comparison of the Blocking Probability for Class 3 Traffic.....	63
5.4	(SS1) Comparison of the ADB for Class 1 Traffic	64
5.5	(SS1) Comparison of the ADB for Class 2 Traffic	64
5.6	(SS1) Comparison of the ADB for Class 3 Traffic	65
5.7	(SS1) Comparison of the BRF for Class 1 Traffic	66
5.8	(SS1) Comparison of the BRF for Class 2 Traffic	66
5.9	(SS1) Comparison of the BRF for Class 3 Traffic	66
5.10	(SS2) Comparison of the Blocking Probability for Class 1 Traffic.....	68
5.11	(SS2) Comparison of the Blocking Probability for Class 2 Traffic.....	68
5.12	(SS2) Comparison of the Blocking Probability for Class 3 Traffic.....	68
5.13	(SS2) Comparison of the ADB for Class 1 Traffic	69
5.14	(SS2) Comparison of the ADB for Class 2 Traffic	69
5.15	(SS2) Comparison of the ADB for Class 3 Traffic	70
5.16	(SS2) Comparison of the BRF for Class 1 Traffic	70
5.17	(SS2) Comparison of the BRF for Class 2 Traffic	71
5.18	(SS2) Comparison of the BRF for Class 3 Traffic	71
5.19	(SS3) Comparison of the Blocking Probability for Class 1 Traffic.....	73
5.20	(SS3) Comparison of the Blocking Probability for Class 2 Traffic.....	73

5.21	(SS3) Comparison of the Blocking Probability for Class 3 Traffic.....	73
5.22	(SS3) Comparison of the ADB for Class 1 Traffic	74
5.23	(SS3) Comparison of the ADB for Class 2 Traffic	74
5.24	(SS3) Comparison of the ADB for Class 3 Traffic	74
5.25	(SS3) Comparison of the BRF for Class 1 Traffic	75
5.26	(SS3) Comparison of the BRF for Class 2 Traffic	75
5.27	(SS3) Comparison of the BRF for Class 3 Traffic	75
5.28	(SS4) Comparison of the Blocking Probability for Class 1 Traffic.....	76
5.29	(SS4) Comparison of the Blocking Probability for Class 2 Traffic.....	77
5.30	(SS4) Comparison of the Blocking Probability for Class 3 Traffic.....	77
5.31	(SS4) Comparison of the ADB for Class 1 Traffic	77
5.32	(SS4) Comparison of the ADB for Class 2 Traffic	78
5.33	(SS4) Comparison of the ADB for Class 3 Traffic	78
5.34	(SS4) Comparison of the BRF for Class 1 Traffic	78
5.35	(SS4) Comparison of the BRF for Class 2 Traffic	79
5.36	(SS4) Comparison of the BRF for Class 3 Traffic	79
5.37	(SS5) Comparison of the Blocking Probability for Class 1 Traffic.....	80
5.38	(SS5) Comparison of the Blocking Probability for Class 2 Traffic.....	81
5.39	(SS5) Comparison of the Blocking Probability for Class 3 Traffic.....	81
5.40	(SS5) Comparison of the ADB for Class 1 Traffic	82
5.41	(SS5) Comparison of the ADB for Class 2 Traffic	82
5.42	(SS5) Comparison of the ADB for Class 3 Traffic	82

List of Tables

1.1	Minimum Data Rates for 3G Systems	12
1.2	Third Generation Service Classes	16
5.1	Simulation Parameters	61

Chapter 1

Introduction

Traditionally, voice traffic has been the mainstay of cellular networks but with the rapid increase in data-centric services, it has been projected in a survey [1] that by the year 2010, about 60% of mobile traffic will consist of multimedia or non-voice traffic. This projection is consistent with the current demands of consumers who want to stay connected on the go [2] and expect unrestricted access to bandwidth hungry services such as email, broadband Internet and TV-quality video streaming. Since bandwidth is a scarce quantity in wireless and mobile networks, resource management and call admission schemes become critically important to efficiently meet the Quality of Service (QoS) requirements of the newer multimedia services while maximizing system resource utilization. The rest of the chapter is organized as follows – we first provide an overview of the evolution of wireless and mobile networks, next we discuss resource management issues for emerging multimedia services in current and future wireless networks. Finally, we present a brief study of relevant work and our contributions in the areas of resource management and call admission control for multimedia services in wireless and mobile networks.

1.1 History of Wireless and Mobile Networks

1.1.1 Early Wireless and Mobile Networks

The origin of wireless communications can be traced back to the year 1895, when the first radio-based wireless transmission was successfully achieved by Guglielmo Marconi in Italy over a distance of one and a half miles [3]. Over the next couple of decades, Marconi was a leading pioneer in establishing long range wireless communications standards and his efforts led to the deployment of first radio-based telephony system for conversations between ships in 1915.

In 1946, the first public mobile telephone network namely Mobile Telephone System (MTS) [4] was introduced in 25 cities in the United States. MTS was an analog system which relied on a single Base Station (BS) to provide service to the entire operating region of the network [5]. The mobile handsets provided to the customers of MTS were very bulky and therefore could only be used in cars. Also, due to the limitation on the number of available channels, only three voice calls could be carried out at the same time in a common service area. This placed a huge restriction on the number of customers that could be served by a MTS-based system. Some other drawbacks associated with this system were half duplex operation and the need for human operators to carry out manual switching of calls. Improved Mobile Telephone System (IMTS) was introduced as an alternative to MTS in the 1960s to improve spectrum efficiency [6]. It increased the capacity of the system by utilizing twenty three channels for voice calls instead of the three channels available in MTS-based systems. It also featured full duplex operation and automatic switching of calls thereby eliminating the need for human operators.

Both the MTS and IMTS systems used a single BS with a high power transmitter to provide service in the entire operating region. To expand the area of service, extensions were introduced by adding sufficiently separated BSs which utilized the same range of frequencies. As a result, interference was caused by the high power transmissions from the BS to the mobile users in neighboring service areas [5]. Furthermore, the limited number of users that could be serviced by these systems, made them impractical in densely populated areas.

1.1.2 First Generation (1G) Cellular Networks

A major breakthrough in wireless communications was achieved when researchers working at AT&T Bell Labs in the 1960s, proposed a mobile network based on the *cellular concept*. The cellular concept, first described in [7], was originally put forward in 1947 by D. H. Ring of AT&T Bell Labs in an unpublished company paper. It proposed replacing the high-coverage areas of legacy systems such as MTS and IMTS with a number of smaller low-coverage areas (cells) with non-overlapping boundaries, each with its own BS. The total channels available in the operating region were divided among the cells by carefully allocating different channel sets for neighboring cells. Thus, interference from adjacent cells was significantly reduced compared to the MTS/IMTS systems. Frequency reuse was also introduced by allowing sufficiently separated cells to utilize the same channel sets which resulted in a dramatic increase of the system capacity.

The commercial deployment of the first generation (1G) cellular networks, based on the cellular concept, began in the 1970s and 1980s. AT&T Bell Labs conducted the first service trials of an analog cellular system in Chicago in 1978 [5]. A fully operational cellular network was introduced under the label Advanced Mobile Phone Service (AMPS) in the United States in the early 1980s [6]-[10]. In other regions of the world, systems based on AMPS architecture were developed and put into operation. North European countries adopted the Nordic Mobile Telephony (NMT) system [9] in 1981 while the United Kingdom deployed an analog standard called Total Access Communications System (TACS) in 1985 [8][9]. In Japan, the first analog cellular system, the Nippon Telephone and Telegraph (NTT) system [8] went operational in 1979. All of these analog cellular networks were designed for voice only traffic and employed frequency modulation (FM) for speech transmission and frequency shift keying (FSK) for signaling purposes. It is interesting to note that the AMPS system is still operational in some parts of the United States with nearly 20 million customers [5][8].

1.1.3 Second Generation (2G) Cellular Networks

With advances in low-rate digital speech encoding and the increased availability of high density integrated circuits (ICs) in the 1980s, it became possible to migrate towards second generation (2G) *digital* cellular networks [8]. There were several factors fueling the development of 2G digital networks [5][11] over the analog implementation of 1G networks, some of which are outlined below:

- (a). Spectrally efficient radio access techniques such as Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) could be utilized as alterna-

tives to the Frequency Division Multiple Access (FDMA) technique used in 1G networks [6].

- (b). In 1G networks, each user was assigned a traffic channel regardless of whether he was *active* or *idle*. In a digital implementation, it would be possible to share the same traffic channel among several users by the use of different time slots (TDMA) or codes (CDMA) for each user leading to increased system capacity.
- (c). Digital encryption of voice data would lead to enhanced security and privacy for customers when compared to the total lack of encryption in 1G networks.
- (d). The application of digital error estimation and correction techniques on the digital voice stream would provide a better conversational quality for end users.

In Europe, work on the first commercial 2G network was started in 1982 with the formation of ‘Groupe Special Mobile’ which was later renamed Global System for Mobile Communications (GSM) [5][8]. The first commercial deployment of the GSM standard [12]-[14] was made in 1992 and used the 900 MHz band. Over the next few years, the GSM technology was extended to include three special variants: GSM900, DCS1800 and PCS1900 operating in the 900 MHz, 1800 MHz and 1900MHz bands respectively [15]. The GSM standard uses TDMA radio access technique and offers full-rate traffic speeds of up to 14.4 kbps which is suitable for voice transmission and low rate data services such as Short Messaging Service (SMS) [16]. Several extensions technologies are available for GSM networks to achieve higher data rates. High Speed Circuit Switched Data (HSCSD) [17][18], a simple upgrade to GSM, allots two, three or four slots per frame to

deliver speeds up to 57.2 kbps. A newer packet-switched technique, General Packet Radio System (GPRS) [19]-[21], supports variable data rates from 14.4 kbps to 115.2 kbps.

In 1993, another important 2G system based on CDMA namely IS-95 or cdmaOne [22][23], was standardized and the first commercial systems were introduced in South Korea and Hong Kong in 1995, followed by the United States in 1996 [5]. The systems based on cdmaOne architecture are backwards compatible with existing AMPS equipment and provide data traffic rates at 4.8 kbps and 14.4 kbps. An extension technology, IS-95b or cdmaTwo [24], offers data rates of up to 115.2 kbps by allowing eight parallel transmissions from a single handset.

An overview of the 2G networks will not be complete without mentioning some other 2G technologies that did not enjoy as much popularity as GSM or cdmaOne but were important nevertheless. In Japan, a digital system named Personal Digital Cellular (PDC) was established in 1991 [25]. Enhancements for the AMPS system in the United States, IS-54 [26] and IS-136 [27], were introduced in 1993 and 1996 respectively under the family name Digital AMPS (D-AMPS). This family of D-AMPS standards were backwards compatible with AMPS-based systems while providing a three fold increase in system capacity [5].

1.1.4 Third Generation (3G) Cellular Networks

The 2G systems have achieved tremendous success since their inception in the early 1990s and continue to enjoy widespread deployment throughout the world. The data rates

offered by the 2G networks (around 14.4 kbps) are suitable for voice transmission but they cannot support increasing demands for bandwidth hungry mobile applications such as broadband Internet, e-mail, on-demand TV and high-quality video conferencing.

In order to efficiently support these emerging services, the International Telecommunications Union (ITU), based in Geneva, started work on the third generation (3G) cellular networks in 1992 [28][29]. The standardization efforts of ITU resulted in the International Mobile Telecommunications – 2000 (IMT-2000) guidelines [30][31], which specify minimum data rates for 3G systems [6][9] as stated in Table 1.1.

Table 1.1: Minimum Data Rates for 3G Systems

<i>Mobility Criteria</i>	<i>Minimum Data Rate</i>
Vehicular movement	144 kbps
Pedestrian movement	384 kbps
Wireless access in local area	2 Mbps
Satellite access	9.6 kbps

A number of different 3G standards have evolved from the joint efforts of the ITU and the European Telecommunications Standards Institute (ETSI), the most important of which are listed below [5]:

- EDGE (Enhanced Data Rates for GSM Evolution) is a backwards compatible system with GSM and IS-136 standards, offering data rates of up to 473 kbps.
- WCDMA (Wideband CDMA), a new CDMA based system utilizing a new 5 MHz wide channel structure, offers data rates of up to 2 Mbps.
- cdma2000 is a descendant of the existing IS-95 standard and supports data rates of up to 2 Mbps.

As of December 2005, there were over 100 network operators in 42 countries offering 3G/WCDMA services to more than 40 million subscribers [32].

1.1.5 Future Trends: Fourth Generation (4G) Cellular Networks

The existence of several diverse 3G standards limits seamless *global roaming* between different cellular networks for a mobile user with a single handset. In addition, there is a fundamental difference between wireless cellular networks (1G, 2G or 3G) and wireless data networks such as WLANs and PANs. The difference is that wireless cellular systems are *circuit-switched* while wireless data networks are *packet-switched* [5]. Convergence issues for these differences between the wireless cellular systems and the wireless data networks will be addressed in the design of fourth generation (4G) cellular networks. It is envisioned [33]-[35] that 4G networks, possibly running on a common IP-based backbone, will provide users with seamless wireless access to voice, data and video services irrespective of which wireless network they belong to.

Even though there is a lot of vagueness about the expected data rates of 4G cellular system due to ongoing research, there have been reports suggesting deployment of 4G systems in 2010 and beyond, with expected bandwidth speeds of 50 Mbps [35][36] to 155 Mbps [34]. To handle these new data rates, Orthogonal Frequency Division Multiplexing (OFDM) [36][37] is quickly emerging as the access method of choice for 4G systems because it is spectrally more efficient than existing access methods such as TDMA, FDMA or CDMA. At this point in time, we can only speculate about the future of 4G systems

but one thing is certain – 4G systems will aim to provide smooth *interoperability* between diverse wireless networks, such as satellite, 2G/3G cellular, WLAN and other systems of wireless access to the underlying fixed network.

Figure 1.1: Evolution of Cellular Systems

Figure 1.1 represents a comprehensive snapshot of the different generations of cellular systems as they have evolved in the past and their expected growth in the coming years.

1.2 Multimedia Services in the 3G Cellular System

1.2.1 Existing Technologies for Multimedia Services

For the 2G cellular system, numerous extension technologies namely HSCSD, GPRS (for GSM networks) and cdmaTwo (for IS-95 networks), are already in place to support medium speed multimedia traffic, offering speeds of up to 173 kbps [38]. These families of extension technologies for 2G networks are collectively termed as 2.5G standards because of their significantly higher data rates. Furthermore, with the current ongoing deployment of 3G standards such as cdma2000, WCDMA and EDGE, users can expect to gain access speeds anywhere from 384 kbps to 2 Mbps [39] along with improved support for multimedia services with adaptive bandwidth requirements (variable transmission rate) [40].

As faster access speeds are made available to the end user, consumers will want to use their mobile handset for a variety of multimedia services, ranging from simple email browsing, file transfers, web browsing to bandwidth hungry services such as voice conferencing, TV-quality video streaming and broadband Internet. In the next section, we provide an outline of the different service classes that will be offered by the 3G system to accommodate the growing demands for these new multimedia applications.

1.2.2 Classification of 3G Service Classes

In general, the services that will be provided by the 3G cellular system can be distinguished into four classes [41]: (a). conversational class; (b). streaming class; (c). interac-

tive class and (d). background class. We can further group these distinct classes into two broad types of traffic [42]: real-time (RT) traffic and non-real-time (NRT) traffic, depending upon how *delay sensitive* the traffic is. The conversational and streaming classes (voice calls, streaming video, etc.) are highly delay sensitive classes and are grouped under RT traffic while the interactive and background classes (email, text messaging, etc.) belong to the lower priority NRT traffic because they have the least delay sensitivity.

The 3G applications and services are also sometimes classified in increasing order of their bandwidth requirements [43]: (a). voice and audio; (b) wireless messaging; (c). switched data; (d). medium multimedia; (e). high multimedia and (f). interactive high multimedia. Specific examples of these services, along with their expected capacity demand, are listed in Table 1.2.

Table 1.2: Third Generation Service Classes

<i>Class</i>	<i>Example</i>	<i>Uplink Speed</i>	<i>Downlink Speed</i>
Voice/Audio	voice calls	28.8 kbps	28.8 kbps
Wireless Messaging	e-mail, SMS	28.8 kbps	28.8 kbps
Switched Data	fax, dial-up internet access, file transfer	56 kbps	56 kbps
Medium Multimedia	web browsing, on-demand audio/video	20 kbps	2 Mbps
High Multimedia	Broadband Internet, TV- quality video streaming	20 kbps	2 Mbps
Interactive Multi-media	video conferencing, interac- tive TV	2 Mbps	2 Mbps

1.2.3 Resource Management Issues for Multimedia Services

1.2.3.1 Challenges

Bandwidth is a scarce resource in wireless and mobile networks and with the growth of bandwidth hungry multimedia applications, it becomes increasingly important to efficiently manage the system resources while satisfying the QoS requirements for the different classes of multimedia traffic.

Since the characteristics of multimedia or data traffic are different from traditional voice traffic, there are certain challenges involved for proper handling of non-voice traffic when transitioning from a voice-centric network [44]:

- Data traffic generally requires more traffic channels or bandwidth units (BUs) than voice traffic.
- Most of the multimedia traffic can be encoded at variable rates depending on the required quality and the bandwidth available for transmission [45]. In other words, its service requirements are *elastic* and can be made adaptive.
- A suitable priority scheme has to be implemented in order to distinguish the different classes of traffic and ensure fair bandwidth allocation while supporting the minimum QoS requirements.

In order to overcome these challenges, the Bandwidth Allocation Algorithm (BAA) plays a vital role in a wireless network for providing the necessary QoS guarantees for different classes of traffic while at the same time maintaining efficient bandwidth management in the system.

1.2.3.2 Overview of Related Work and Our Contributions

In this section, we briefly introduce some of the existing BAA schemes proposed in recent literature (detailed discussion in Chapter 2 – Related Work) and also highlight the salient features of our proposed scheme.

Traditional bandwidth management schemes, such as Complete Sharing (CS) and Fixed Partitioning (FP), lead to *sub-optimal* system performance in a wireless network with multiple classes of multimedia traffic [46]. The CS scheme exhibits *unfairness* because multimedia services which require less bandwidth monopolize the available resources in the system while starving the higher bandwidth services. In contrast, the FP scheme offers the benefit of comparable QoS performance for all the classes at the cost of *inefficient utilization* of system resources. In recent literature, numerous resource management schemes have been presented which are specially tailored for handling multiple classes of multimedia traffic in a wireless and mobile environment.

In [44], a CS based Dual Threshold Bandwidth Reservation (DTBR) scheme is proposed. The suggested DTBR scheme is *static* in nature and does not adaptively modify the bandwidth of ongoing multimedia calls to accept new calls resulting in high blocking probabilities for both originating and handoff calls. A more efficient approach lies with the class of Flexible Resource Allocation (FRA) schemes which *degrade* ongoing multimedia calls to accommodate new call requests at heavy traffic loads. One such FRA scheme is proposed in [47], in which inter-class fairness is achieved by dividing the system bandwidth among the different classes of multimedia traffic according to their arrival

rates. Since this scheme incorporates fixed resource partitions, derived from the traditional FP scheme, it may not perform well when there is a burst of new call arrivals belonging to an already overloaded class of traffic.

A graceful degradation bandwidth degradation strategy, for both RT and NRT multimedia traffic, is presented in [48]. The degradation and call admission policy is governed by a cost function formulated on the overall traffic carried by the system and the proportion of degraded calls in the system. In [49], another FRA scheme is proposed which changes the bandwidth of both ongoing and incoming calls according to the current system state. The scheme also introduces a dynamic prioritization scheme which ensures fair share bandwidth allocation for the different classes of multimedia traffic. But in the above work, the users experience high QoS fluctuation because of frequent bandwidth reallocation during the lifetime of a call resulting in an unpleasant user experience.

In our work, we describe a novel Dynamic Adaptive Resource Allocation (DARA) scheme [50] (an earlier version of which was presented in [51]), which focuses on improved handling of multimedia services with flexible bandwidth requirements in wireless and mobile networks. Our work improves upon the scheme proposed in [49] by significantly lowering the bandwidth reallocation frequency that a call undergoes during its lifetime while at the same time improving QoS metrics like call blocking probability for originating calls. We also extend our work to handle higher priority handoff calls by a combination scheme of channel reservation exclusively for handoff calls and during heavy traffic loads, allowing them to borrow bandwidth from ongoing originating calls.

In [52], two other schemes H1 and H2 (hereafter referred to as HMax and HMin), are proposed which always allocate the maximum and minimum number of resources respectively to incoming calls. Although these schemes are not FRA schemes in principle, they are also considered as they represent the upper and lower performance bounds for any FRA scheme [53].

1.3 Organization of the Thesis

The rest of the thesis is organized as follows – Chapter 2 provides an overview of related work in the areas of resource management and admission control for multiple classes of multimedia traffic in wireless and mobile networks. Our proposed scheme, DARA, is detailed in Chapter 3 with complete description of the system model, the traffic model and the bandwidth allocation rules which govern our scheme. Chapter 4 outlines our simulation model and presents the results of our extensive computer experiments. Finally, Chapter 5 lists down the conclusions of our research work and discusses potential future work.

Chapter 2

Related Work

2.1 Introduction

Numerous approaches have been published in recent literature proposing Bandwidth Allocation Algorithms (BAA) for wireless and mobile networks. In this chapter, we present an overview of some of the proposed BAA schemes which specifically address resource management issues for multiple classes of traffic in a wireless environment. Our discussion of these schemes is structured as follows, we first classify the schemes into three categories: (a). traditional schemes; (b). static resource allocation schemes and (c). flexible resource allocation (FRA) schemes. Traditional schemes are the simplest in nature and either share the whole system bandwidth among the different classes of traffic (CS — Complete Sharing) or allocate fixed exclusive bandwidth partitions for the different classes of traffic (FP — Fixed Partitioning). Static resource allocation schemes follow a more sophisticated approach either by implementing priority schemes or by using dynamic bandwidth partitions for the different classes of traffic. Finally, we conclude our discussion with the class of FRA schemes. The FRA schemes, in addition to using *differentiated* resource allocation like the static schemes, further improve the system capacity by degrading ongoing multimedia calls to free up bandwidth for new call requests. Our

proposed scheme, Dynamic Adaptive Resource Allocation (DARA) scheme, belongs to the class of FRA schemes.

2.2 Traditional Schemes

2.2.1 Complete Sharing (CS) Scheme

The CS scheme allows unrestricted access to the system bandwidth by any of the competing classes of traffic. The calls are treated on a *first come first serve* basis, i.e., as long as the required number of Bandwidth Units (BUs) requested by an incoming call are available in the system, the call will be accepted. As calls are serviced and they depart from the system, the released BUs by the departing calls are added back to the common system resources and made available to future incoming calls of all classes of traffic. A generic depiction of bandwidth allocation in a CS-based system for K different classes of traffic is shown in Figure 2.1, where C_i indicates the bandwidth available for calls belonging to class i traffic and C is the total system bandwidth.

Under the CS implementation, the whole of the system bandwidth is fully utilized at the cost of *unfair* distribution among the different classes of traffic. The reason being that the calls belonging to low bandwidth services, i.e., calls which require fewer BUs, monopolize the system resources while starving the calls belonging to higher bandwidth services, especially true in systems with asymmetric traffic loads.

Figure 2.1: Complete Sharing (CS) Scheme

2.2.1 Fixed Partitioning (FP) Scheme

In a FP scheme, the total system bandwidth C is divided into permanent and static partitions among the K classes of traffic with C_i portion of the bandwidth reserved exclusively for calls belonging to class i traffic. When a call belonging to class i traffic arrives in the system, if the required number of BUs are available in the C_i partition, the call is accepted otherwise the call is rejected. When a class i call departs, the released BUs are added back to the C_i partition. An illustration of the FP scheme for multiple classes of traffic is shown in Figure 2.2.

If the bandwidth partitions $C_1, C_2, \dots, C_K$ are sensibly chosen, the FP scheme provides comparable QoS performance for all the classes of traffic at the expense of reduced system resource utilization when compared to the CS scheme. An insightful study of both the CS and FP schemes, with extensive analytical and simulation models, is detailed in [46] which substantiate our above mentioned deductions.

Figure 2.2: Fixed Partitioning (FP) Scheme

2.3 Static Resource Allocation Schemes

As mentioned earlier, static resource allocation schemes differentially treat the multiple classes of traffic based on their varied QoS requirements and the current system state.

This differentiation is typically implemented by using priority schemes or by using dynamic bandwidth partitions for the different classes of traffic. It is to be noted here that this family of schemes does not modify the bandwidth of an ongoing call once it has been admitted into the system. Because of their static nature, these schemes are unable to take advantage of the flexible bandwidth requirements of multimedia traffic resulting in a *sub-optimal* system performance.

An example of a static resource allocation scheme is provided in [54], where a priority-based bandwidth allocation algorithm is implemented. In this scheme, a bandwidth partition is reserved exclusively for handoff pools and the rest of the system bandwidth is shared among all the different classes of traffic. In addition, a threshold-based fixed priority scheme is established between the different classes of traffic, which “admits high-priority traffic preferentially under heavy traffic loads”. Under this scheme, calls belonging to the low priority traffic will suffer from high blocking probabilities.

A CS-based Dual Threshold Bandwidth Reservation (DTBR) scheme is proposed in [44] which handles two broad classes of traffic — voice traffic (high priority) and data traffic (low priority). The total system bandwidth C is divided into three partitions by using two fixed thresholds T_1 and T_2 ($T_2 < T_1$). When the total occupied bandwidth is less than T_2 , both voice and data traffic can be serviced by the system. If the occupied bandwidth is more than T_2 but less than T_1 , only voice calls are serviced and when the occupied bandwidth is more than T_1 (heavy traffic load), only handoff voice calls are allowed. The bandwidth partitions for the DTBR scheme are shown in Figure 2.3. The drawback with

this scheme is that it is static in nature with fixed partitions; dynamic bandwidth allocation schemes can always achieve better performance than the proposed DTBR scheme.

Figure 2.3: Dual Threshold Bandwidth Reservation (DTBR) Scheme

2.4 Flexible Resource Allocation (FRA) Schemes

Due to the adaptive nature of multimedia services, most of the multimedia traffic can be encoded at variable rates, i.e., multimedia traffic can occupy multiple levels of bandwidth depending on the current system load and the bandwidth available for transmission. In recent literature, numerous FRA schemes for wireless and mobile networks have been proposed which take advantage of these adaptive bandwidth requirements of the newer multimedia services.

A graceful bandwidth degradation strategy for both real-time and non-real-time multimedia traffic is outlined in [48]. To support increased traffic in the system, some of the ongoing calls are degraded to accommodate more new calls. The degradation strategy and call admission control in this scheme is governed by a cost function based on the net revenue generated by the system. This scheme balances the *positive* revenue generated by the increased carried traffic against the *negative* impact on system revenue by calls receiving degraded service in the system. In [47], the system bandwidth is divided among the different classes of traffic proportional to their arrival rates (*inter-class fairness*). The *intra-class fairness*, i.e., fairness among calls belonging to the same class, is achieved by fairly allocating bandwidth among users in the same class with the most degraded calls receiving the highest priority. This scheme has fixed resource partitions and may not perform well when there is a burst of new call arrivals belonging to an already overloaded class of traffic.

Another example of a FRA scheme is presented in [49] which is the basis of our proposed scheme – DARA. The dynamic resource allocation scheme of [49] modifies the bandwidth of both ongoing and incoming calls according to the current occupied bandwidth in the system. Fairness between the different classes of traffic is achieved by a *dynamic* prioritization scheme unlike the static priority schemes of previously discussed FRA schemes. But the above work performs poorly with regard to average QoS fluctuation metric, i.e., the calls in this system suffer from a high average bandwidth reallocation frequency during their lifetime. This frequent upgrade/degrade of calls involves consider-

able signaling overhead at the BS and also leads to an annoying user experience [55]. We address these issues in our proposed scheme, DARA and also extend it to handle the higher priority handoff calls.

We conclude the discussion on related work by mentioning two more schemes proposed in [52], HMax and HMin, which always allocate the maximum and minimum number of resources respectively to incoming calls. These schemes are also considered in our comparisons because they represent the upper and lower performance bounds for any FRA scheme [53].

Chapter 3

Preliminaries

3.1 Introduction

In this chapter, some of the key topics are presented which provide a better understanding of the proposed scheme, DARA, in subsequent chapters. In the first section, we introduce three important probability distributions — exponential, Poisson and Pareto distributions with emphasis on their unique properties which are useful for the modeling of different classes of traffic in a wireless and mobile system. Next, the handoff process is described in detail with information about the different types of handoffs, the handoff criteria and the various handoff schemes.

3.2 Important Probability Distributions

The performance evaluation of wireless and mobile networks is specified by various non-deterministic factors which help define the whole system. Some examples of such random parameters are the nature of call arrivals, the expected duration of calls and the movement of non-stationary users in the system [4]. For the efficient modeling and analysis of any wireless and mobile network, it is necessary to quantify or approximate

the aforementioned parameters into well-defined probability distributions. Since the underlying probability and statistics theory is too extensive to be covered by the scope of this thesis, we briefly define some of the probability distributions and their important properties that are used to define our system model and traffic model in subsequent chapters.

3.2.1 Exponential Distribution

The exponential distribution, due to its simplicity, is one of the most widely used distributions in the simulation of computer systems and networks. A random continuous time variable X is said to be exponentially distributed with parameter λ if its probability density function (pdf) is mathematically defined as:

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & \text{for } 0 \leq x < \infty. \end{cases} \quad (3.1)$$

Memory-less Property of the Exponential Distribution

The exponential distribution is the only continuous time random distribution that has no memory, i.e., the probability of an event happening in a specific time interval is the same regardless of the starting point of that time interval. For any two non-negative real numbers $t, \delta \geq 0$, consider the events $\{x > t + \delta\}$ and $\{x > t\}$. If $P\{x > t + \delta \mid x > t\}$ denotes the conditional probability of the event $\{x > t + \delta\}$ assuming that the event $\{x > t\}$ has already happened and $P\{x > t\}$ is probability of the event $\{x > t\}$, then the memory-less property states that:

$$P\{x > t + \delta \mid x > t\} = P\{x > \delta\}. \quad (3.2)$$

If we interpret x as the lifetime of an equipment [56], then (3.2) states that if the equipment has already been in use for time t , then the probability of it surviving for an additional time δ is the same as the probability of a brand new equipment surviving for the time δ . In other words, the equipment does not remember that it has already been used for time t , hence the “memory-less” property.

The exponential distribution is widely used to model the duration of calls and the channel holding time in wireless and mobile systems.

3.2.2 Poisson Distribution

The Poisson distribution is typically used to model random events occurring over a fixed interval of time. The Poisson distribution $P_\lambda(n, T)$ gives the probability for observing n discrete events in a time-interval of length T , when the mean number of events per unit time is λ , assuming that each event occurs *independently* of all others in the past or future:

$$P_\lambda(n, T) = \frac{e^{-\lambda T} (\lambda T)^n}{n!}. \quad (3.3)$$

The Poisson process is the widely accepted standard distribution that is used to approximate the nature of call arrivals in a wireless and mobile system.

Inter-arrival Times of a Poisson Process

Starting with an arbitrary time t , let T_1 denote the time until the first event occurrence in a Poisson process, then it can be shown [4] that T_1 has an exponential distribution with mean rate λ :

$$f_{T_1}(t) = \lambda e^{-\lambda t}. \quad (3.4)$$

Similarly, if we define T_2 as the time between the first and second events, T_3 as the time between second and third events and so on, then the inter-arrival times $T_1, T_2, T_3, \dots$ of a Poisson process are independent of each other and each of them follow the same exponential distribution with mean rate λ :

$$f_{T_2}(t) = f_{T_3}(t) = \dots = \lambda e^{-\lambda t}. \quad (3.5)$$

This property proves very convenient in simulation of Poisson events in computer programs by letting the inter-arrival times between the events follow an exponential distribution.

Merging Property

If we merge n Poisson processes with mean arrival rates λ_i , where $i = 1, 2, \dots, n$, then the result is a single Poisson process with mean arrival rate $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$ [4]. This property can be used to model the combined traffic load on a wireless system serving multiple classes of traffic, with each class being represented by its corresponding Poisson process.

Splitting Property

If a Poisson process with mean number of events per unit time λ is split into n processes so that the probability of assigning an event to the i th process is P_i ($i = 1, 2, \dots, n$) then the resulting i th process will also be a Poisson process with a mean rate of $P_i\lambda$ events per unit time [4].

3.2.3 Pareto Distribution

The Pareto distribution is one of the simplest examples from the family of heavy tailed distributions. A continuous random variable T which follows Pareto distribution has its pdf $f_T(t)$ and expected value $E[T]$ mathematically defined as:

$$f_T(t) = \frac{\alpha\beta}{(t + \beta)^{\alpha+1}} \quad \alpha, \beta > 0, t \geq 0 \quad (3.6)$$

$$E[T] = \frac{\alpha\beta}{(\alpha - 1)}, \quad (3.7)$$

where α is the shape parameter of the distribution and describes the ‘heaviness’ of the tail of the distribution while β is a scaling factor. When the condition $1 < \alpha < 2$ is true, the Pareto distribution has an infinite variance, a finite mean and exhibits self-similar behavior [57]. Because of this self-similar nature of the Pareto distribution, it is widely used to model the nature of call arrivals and duration of calls generated by data or non-voice traffic in a wireless and mobile network.

3.3 Handoff Process

The wireless and mobile networks are characterized by one unique factor, i.e., the users of these networks are *mobile*. To provide roaming access to these mobile users as they move across the different service regions, it is necessary to design an efficient handoff algorithm. For cellular networks, handoff (also sometimes known as hand-off or hand-over) can be defined as “the process of changing the channel (frequency, time slot, spreading code or a combination of them) associated with current connection while the call is in progress” [58]. A depiction of the handoff process in cellular networks, due to crossing of the cell boundary, by a mobile user is shown in Figure 3.1.

3.3.1 Types of Handoff

Handoffs can be broadly divided into two categories — hard handoffs and soft handoffs [4]. Hard handoffs are characterized by a *break before make* connection, i.e., the resources held by a Mobile Station (MS) with the current BS (Base Station) are released first before acquiring new resources from the next BS. A soft handoff is a *make before break* connection because in a soft handoff, it is possible for a MS to communicate simultaneously with two adjacent BSs while staying in the handoff region. The FDMA and TDMA systems employ hard handoff while soft handoffs are implemented in CDMA-based systems.

Handoff occurs as the MS moves from the service area of BS_1 to the service area of the adjacent BS_2

Figure 3.1: Example of Handoff in Cellular Networks

Handoffs are also sometimes classified as intra-cell and inter-cell handoffs [59]. Intra-cell handoff occurs when there is a channel change within the same cell while in an inter-cell handoff, the channel change occurs from one cell to another cell.

3.3.2 Handoff Criteria

To make a handoff decision, it is necessary to choose certain criteria which provide useful information about the call and their threshold values which determine the need for

handoff. Some of the criteria that are traditionally considered in a wireless and mobile network for handoff decisions are presented below [59].

Received Signal Strength (RSS)

The RSS of a MS provides a simple criterion — we can choose the BS which has the strongest signal available. But, in practical scenarios, if we consider only this parameter, it gives rise to unnecessary frequent handoffs due to rapid fluctuations in the signal strengths of adjacent BSs in the handoff area. It is recommended to use the RSS value along with other measurement parameters for a more efficient handoff process.

Carrier to Interference Ratio (CIR)

This is one of the very important criteria for handoff decisions and is widely used in mobile networks to initiate handoff. It determines the size of the cluster, cell shape and the frequency reuse distance in cellular networks. It also gives a measure of the bit error rate (BER) and the available signal strength at the MS.

Network Criteria

In these cases, handoff is done for reasons other than degradation of the channel or the loss of signal strength; network criteria are usually managed by a central network entity such as the Mobile Switching Center (MSC). For example, in a directional antenna based system, the channel of the MS might have to be changed when it moves from one sector to another sector

In usual practice, a combination of the above criteria is chosen to make a handoff decision instead of relying on just one parameter, resulting in a more effective handoff process.

3.3.3 Handoff Schemes

Once the handoff criteria have been chosen, it is necessary to designate the network entities that are responsible for measuring these criteria and making a handoff decision based on their threshold values. Numerous handoff schemes have been discussed in [60] outlining the decision-making process for handoffs.

Network Controlled Handoff (NCHO)

In this type of handoff control, the network (usually the MSC) makes a decision based on the measurements of the handoff criteria of the MSs at a number of BSs. Even though the delay associated with just the handoff process is around 100-200 ms, the overall delay can be in the range of 5-10s. This type of handoff scheme was first used in 1G analog cellular networks such as AMPS, TACS and NMT systems.

Mobile-Assisted Handoff (MAHO)

In a MAHO system, the responsibility for measuring the handoff criteria and making the handoff decision is assigned to the BS and the MSC, respectively. The delay in this case from the handoff initiation till the handoff execution is around 1s. Some examples of present cellular networks which implement MAHO are the GSM system and the IS-95 system.

Mobile-Controlled Handoff (MCHO)

In this case, the MS is the sole entity which measures the handoff criteria and makes a decision based on them — the MSC is not involved in the handoff process resulting in reduced burden on the MSC. Since the handoff process is implemented in the MS itself, the delay is usually smaller with a typical value of 0.1s. The MCHO type of decentralized fast handoff control is useful for micro-cellular systems (radius <1 km) where there is a high user density and frequent handoffs occur. An example of a MCHO based handoff control network is the standard for cordless phones in Europe — DECT (Digital European Cordless Telephone) [61].

Chapter 4

Dynamic Adaptive Resource Allocation (DARA) Scheme

4.1 Introduction

In this chapter, we present our proposed BAA (Bandwidth Allocation Algorithm) approach, namely Dynamic Adaptive Resource Allocation (DARA) scheme [50], which is specially suited to handle multiple classes of multimedia traffic in wireless and mobile networks. Our proposed scheme, DARA, is based on the concept that multimedia traffic has flexible bandwidth requirements and the system capacity can be increased if we *intelligently* degrade some ongoing calls to accept more calls in the system. This leads to a better system performance when compared to traditional resource allocation schemes such as CS (Complete Sharing) and FP (Fixed Partitioning) schemes which do not dynamically adjust the bandwidth of ongoing calls to increase system capacity.

In related work, a lot of research has been done in the area of resource management for multiple classes of traffic in wireless and mobile networks. Static resource allocation schemes, proposed in [44][54], implement intelligent partitioning of the system band-

width based on priority schemes or on the distinct QoS requirements of the different classes of traffic. Since these schemes, although more sophisticated than traditional schemes, do not benefit from the flexible bandwidth requirements of the multimedia services, their performance will always be inferior to FRA (flexible resource allocation) schemes. One such FRA scheme is presented in [49], which implements dynamic partitions and also adaptively modifies the bandwidth of ongoing multimedia calls to increase system capacity. But the drawback with this FRA scheme is that the calls in the system experience high QoS fluctuation because of the frequent upgrading and degrading of calls. This leads to an increased signaling overhead at the BS and an annoying user experience. In our proposed scheme, DARA, we improve upon the resource allocation scheme proposed in [49] by introducing a *dynamic* bandwidth borrow unit. This drastically reduces the QoS fluctuation experienced by a call during its lifetime while giving better system performance in terms of the blocking probabilities for both originating and handoff calls.

The rest of the chapter is organized as follows: we start with a description of the underlying system model and explain the various associated system level parameters. The traffic model is presented next and it defines the behavior of call arrivals for the different classes of traffic and channel occupancy by those calls in the system. After the system model definition, the bandwidth allocation rules and the call admission control policy of our scheme are specified in the next section. We then proceed to introduce two non-traditional QoS parameters – Average Degraded Bandwidth (ADB) and Bandwidth Reallocation Frequency (BRF) which are relevant for a comprehensive performance evaluation of our scheme.

4.2 System Model

Before we start with the description of our system model, we need to define certain parameters necessary to provide an overview of the system model.

4.2.1 System Parameters

- N : Total number of traffic channels or Bandwidth Units (BUs) available in a single cell.
- K : Number of classes of multimedia traffic.
- $b_{i\max}$: Maximum BUs required by a class i call, where $i \in \{1, 2, \dots, K\}$; $b_{i\max}$ can take only integer values.
- $b_{i\min}$: Minimum BUs required by a class i call; $b_{i\min}$ can also take only integer values.
- b_{icur} : Current BUs occupied by an ongoing class i call, valid values for b_{icur} are in the range $[b_{i\min}, b_{i\max}]$.
- Δb_i : Number of BUs that will be borrowed from an ongoing class i call — Δb_i is dynamic in our scheme and can take integer values in the range $[1, b_{i\max} - b_{i\min}]$.

- C_i : Allotted fair share BUs for calls belonging to class i traffic — suitable values for C_i can be chosen based on the range $[b_{i\min}, b_{i\max}]$ or the expected proportion of class i traffic.
- C_o : Originating bandwidth pool, defined as $C_o = \sum_{i=1}^K C_i$.
- C_h : Handoff bandwidth pool, channels exclusively reserved for the handoff calls of all traffic classes — the condition $C_o + C_h = N$ should always be satisfied.
- B_i : Number of BUs currently being used by all ongoing class i calls.
- U_i : Utilization index of class i calls, defined as $U_i = B_i / C_i$.

After an overview of the system parameters, we now proceed to describe our system model in detail. In our system model, we consider a wireless and mobile network with *homogeneous* cells, i.e., cells which belong to the same network experience identical arrival rates and service times for the different classes of multimedia traffic. Further elaboration of some of the system parameters and their importance in the system model is discussed below.

The total channel capacity of a single cell in a wireless and mobile network is taken to be N BUs where a BU is the basic measurement unit of the system capacity. For example, in a cellular system, a BU could be used to denote a frequency slot (for FDMA systems), a time slot (for TDMA systems), or a user specific code (for CDMA systems). To accommodate the different types of multimedia services available in today's wireless and

mobile networks, e.g., e-mail, video conferencing and TV-quality video streaming, we further assume that the incoming calls can be divided into K different classes of traffic based on their varied QoS requirements and bandwidth demands.

Whenever a call belonging to class $i \in \{1, 2, \dots, K\}$ traffic arrives in the system, it specifies the maximum bandwidth $b_{i\max}$ and the minimum bandwidth $b_{i\min}$ required by the call. Since multimedia traffic can be encoded at variable rates, the incoming call can be serviced by the system if it is allotted bandwidth in the range $[b_{i\min}, b_{i\min} + 1, \dots, b_{i\max}]$. Thus, the current bandwidth b_{icur} held by any ongoing class i call can hold valid values in the range $[b_{i\min}, b_{i\max}]$. For example, consider an incoming class i call with specified bandwidth demands of $b_{i\max} = 4$ and $b_{i\min} = 2$, then this call can be accepted into the system provided there are 2, 3 or 4 free BUs available in the system bandwidth. If at least the minimum bandwidth, $b_{i\min}$ (in this case 2), cannot be provided to the incoming class i call, it will be rejected. We take advantage of the flexible bandwidth requirements of these multimedia calls by degrading some existing calls to increase system capacity.

Each class of traffic further comprises of both originating and handoff calls. An originating call is a new call which is initially generated in the same cell. As mentioned in the previous chapter, handoff calls can be either inter-class or intra-class handoff calls. In our work, we focus on inter-cell handoff call requests which are generated when an ongoing call moves away from the range of one BS and into the service area of an adjacent BS.

Since the termination of an ongoing call is generally considered to be more disruptive to the user than the blockage of a new call, in our proposed scheme, handoff calls are given higher priority than originating calls. This enhanced QoS provisioning for handoff calls is implemented in our scheme by reserving some bandwidth C_h exclusively for the handoff calls. In addition, handoff calls can also access the originating bandwidth pool C_o along with the originating calls.

The sharing of the system bandwidth between the different classes of traffic is implemented by bandwidth partitions C_i where $i \in \{1, 2, \dots, K\}$. It is to be noted here that these are not fixed partitions; rather the value of C_i indicates the amount of bandwidth that should ideally be allocated to class i calls based on their respective QoS requirements. Dynamic prioritization of the different classes of traffic is done by monitoring their respective utilization indexes, $U_i = B_i / C_i$. Classes are prioritized according to their U_i values; classes which have higher U_i are given lower priority so when a class i becomes overloaded ($U_i > 1$), its priority automatically goes down and newer calls of the same class will be blocked from taking up excess bandwidth from other class calls.

The system model diagram for two classes of multimedia traffic is shown in Figure 4.1. As we can see handoff calls can borrow bandwidth across the whole system bandwidth while the originating calls cannot touch the handoff bandwidth pool C_h . The dotted lines indicate the bandwidth borrowing options that are available for the calls when they use up all the bandwidth in their corresponding bandwidth partitions.

λ_{oi} is the mean arrival rate of class i originating calls
 λ_{hi} is the mean arrival rate of class i handoff calls
 C_i is the shared bandwidth partition for class i originating calls
 C_h is the handoff bandwidth pool for both class 1 and class 2 handoff calls
 - - - indicates bandwidth borrowing across the different partitions

Figure 4.1: System Model Diagram

4.2.2 Traffic Model

In this section, we describe the traffic model assumed in our scheme for a single cell belonging to a wireless and mobile network.

(a). *Call Arrival Process*

We adopt two different approaches to model the arrival processes of originating and handoff calls belonging to class $i \in \{1, 2, \dots, K\}$. In the first approach, we adopt the traditional Poisson process (i.e. exponential inter-arrival times) to model the nature of call arrivals of originating and handoff calls belonging to class i with mean arrival rates λ_{oi}

and λ_{hi} respectively. The Poisson arrival process is used because the main underlying assumptions of Poisson process - ‘independence of calls’ and ‘infinite population’ hold true for the different classes of traffic in a wireless and mobile network [62][63]. The total arriving class i traffic in the system is a combined sum of the class i originating and handoff calls. Thus, in effect, the total class i traffic can be defined by a Poisson process with mean arrival rate of $\lambda_i = \lambda_{oi} + \lambda_{hi}$.

In another approach, the Pareto distribution is utilized to generate the inter-arrival times for originating and handoff calls belonging to class i with mean arrival rates of λ_{oi} and λ_{hi} respectively. This approach can be justified as realistic because the Pareto distribution is self-similar in nature [57] and effectively describes the long-range dependence present in multimedia and data traffic of a wireless and mobile network [63]. The arrival rate of class i traffic λ_i is a measure of the offered traffic load and is related to the mean of Pareto distribution as follows:

$$\frac{1}{\lambda_i} = \frac{\alpha\beta}{(\alpha-1)}, \quad (4.1)$$

where α is the shape parameter and β is the scale parameter of the Pareto distribution. The parameter α determines the ‘burstiness’ of traffic — the closer the value of α is to one, more bursty the traffic is. Once a suitable value of α is chosen to describe the traffic characteristics, the scale parameter β can be determined based on the desired traffic arrival rate λ_i by the relation shown in (4.1).

(b). *Call Holding Time*

The call holding time T_{ic} (random variable) for class $i \in \{1, 2, \dots, K\}$ calls (originating and handoff calls) is defined as the total time that is taken by the call for completion, i.e., the time taken from the generation of the call in the system until the termination of the call.

(c). *Cell Dwell Time of a Mobile User*

The cell dwell time T_{dwell} (random variable) is defined as the time duration from the time of call generation to the time when the user leaves the cell and moves to an adjacent cell. It is assumed that all users have the same mobility pattern; hence it is sufficient to define one common distribution of cell dwell time distribution for all the users. One simple mobility model, which describes the movement of mobile users in cellular system, is proposed in [64] based on the fluid flow model. The fluid flow model assumes that all the users are uniformly distributed throughout the system and are equally likely to move in any direction. In a two-dimensional mobility model, the average outgoing rate of mobile users in a cell, μ_{dwell} , is given by:

$$\mu_{dwell} = \frac{E[V]L}{\pi A}, \quad (4.2)$$

where $E[V]$ is the average of the moving speed V (random variable) of the mobile user, L is the length of the perimeter of the cell and A is the area of the cell.

(d). *Channel Holding Time*

The channel holding time T_i (random variable) for a class $i \in \{1, 2, \dots, K\}$ call is de-

defined as the amount of time for which the call occupies some bandwidth in the cell. It is intuitive that the channel holding time T_i is the minimum of the call holding time T_{ic} and the cell dwell time T_{dwell} . To generate the channel holding times for class i calls in our traffic model, we choose two different probability distributions — the exponential distribution and the Pareto distribution.

With the definition of our system model complete, we now proceed to outline the rules governing resource allocation in our proposed scheme.

4.3 DARA Bandwidth Allocation Algorithm

In our scheme, depending upon the type of calls (originating or handoff) and the class i of traffic, appropriate set of bandwidth allocation rules are processed. These rules can be conveniently implemented in a central entity of an infrastructure-based wireless and mobile network, e.g., the BS in a cellular network. All the different call arrival and departure events along with their associated processing rules are defined in the following subsections.

4.3.1 Originating Call Arrival Handling

If the utilization index U_i of class i is less than 1, i.e., $U_i \leq 1$ where $i \in \{1, 2, \dots, K\}$, the class is said to be *under-loaded*, or in other words, the class is using less than its allotted fair share bandwidth C_i . In this case, $b_{i_{\max}}$ bandwidth is assigned to the incoming call

either from the originating bandwidth pool C_o or by degrading ongoing calls of class j ($j \neq i, j \in \{1, 2, \dots, K\}$) with higher U_j values than U_i , i.e., $U_j > U_i$. If $b_{i\max}$ bandwidth is not available, $b_{i\min}$ bandwidth is allotted following the same procedure as above. If neither $b_{i\max}$ nor $b_{i\min}$ bandwidth is available, the incoming call is rejected. The reason behind trying to allot first $b_{i\max}$ bandwidth and then $b_{i\min}$ bandwidth is to ensure that originating calls belonging to under-loaded, i.e., higher priority classes are given the maximum possible bandwidth.

When the incoming originating class i call is *overloaded* ($U_i > 1$), the bandwidth allocation process is similar to that of *under-loaded* classes except only $b_{i\min}$ bandwidth is assigned if possible. The reason to assign only $b_{i\min}$ bandwidth instead of $b_{i\max}$ bandwidth to an originating call belonging to an overloaded class is to restrict future calls of an overloaded class from hogging the bandwidth; this would lead to a fair QoS performance among all classes. The detailed steps for originating call arrival handling are given in Algorithm I. The bandwidth upgrade/degrade process and the criteria for choosing ongoing calls for the upgrade/degrade procedure are explained in the later sections.

4.3.2. Handoff Call Arrival Handling

In our proposed scheme, handoff call arrivals for all classes are treated according to identical rules. Since handoff calls have the strictest QoS requirements, the system reserves some bandwidth C_h exclusively for the handoff calls. To further enhance QoS provisioning for handoff calls, only $b_{i\min}$ bandwidth is provided to all the incoming handoff calls.

In this manner, more number of handoff calls can be accommodated in the system.

Initially, if the required $b_{i\min}$ bandwidth is available in the handoff bandwidth pool C_h , the incoming handoff call is accepted. If $b_{i\min}$ bandwidth is not available in the handoff bandwidth pool C_h , $b_{i\min}$ bandwidth is made available either from the originating bandwidth pool C_o or by degrading ongoing calls. The detailed steps for handoff call arrival handling are presented in Algorithm II and the flowchart for both originating and handoff call arrival handling is shown in Figure 4.2.

Algorithm I. Originating Call Arrival Handling

-
- STEP 1. ORIGINATING CLASS i CALL ARRIVAL OCCURS.
 - STEP 2. IF CLASS i IS OVERLOADED, i.e., $U_i > 1$, GO TO STEP 8.
 - STEP 3. IF $b_{i\max}$ BANDWIDTH IS AVAILABLE IN THE ORIGINATING BANDWIDTH POOL C_o , ASSIGN $b_{i\max}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 4. IF $b_{i\min}$ BANDWIDTH IS AVAILABLE IN THE ORIGINATING BANDWIDTH POOL C_o , ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 5. IS $b_{i\max}$ BANDWIDTH AVAILABLE BY DEGRADING LOWER PRIORITY CLASS j ($U_j > U_i$) CALLS? IF YES, ASSIGN $b_{i\max}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 6. IS $b_{i\min}$ BANDWIDTH AVAILABLE BY DEGRADING LOWER PRIORITY CLASS j ($U_j > U_i$) CALLS? IF YES, ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 7. INCOMING CALL CANNOT BE ACCEPTED, REJECT IT, STOP.
 - STEP 8. IF $b_{i\min}$ BANDWIDTH IS AVAILABLE IN THE ORIGINATING BANDWIDTH POOL C_o , ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 9. IS $b_{i\min}$ BANDWIDTH AVAILABLE BY DEGRADING LOWER PRIORITY CLASS j ($U_j > U_i$) CALLS? IF YES, ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
 - STEP 10. INCOMING CALL CANNOT BE ACCEPTED, REJECT IT, STOP.
-

Algorithm II. Handoff Call Arrival Handling

- STEP 1. HANDOFF CLASS i CALL ARRIVAL OCCURS.
- STEP 2. IF $b_{i\min}$ BANDWIDTH IS AVAILABLE IN THE HANDOFF BANDWIDTH POOL C_h , ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
- STEP 3. IF $b_{i\min}$ BANDWIDTH IS AVAILABLE IN THE ORIGINATING BANDWIDTH POOL C_o , ASSIGN $b_{i\min}$ BANDWIDTH TO INCOMING CALL AND ACCEPT CALL, STOP.
- STEP 4. IS $b_{i\min}$ BANDWIDTH AVAILABLE BY DEGRADING ONGOING CALLS? IF YES, ASSIGN $b_{i\min}$ BUS TO INCOMING CALL AND ACCEPT CALL, STOP.
- STEP 5. INCOMING CALL CANNOT BE ACCEPTED, REJECT IT, STOP.
-

4.3.3 Call Departure Handling

If the departing call was initially generated in the same cell, the released bandwidth by that call is distributed back to the current ongoing calls which are receiving degraded service in the system. The bandwidth upgrade process is described in the later sections. If excess bandwidth is still left over after upgrading all the calls (possible under low traffic conditions), it is added to the originating bandwidth pool C_o .

If the departing call arrived in the system after being handed over from an adjacent cell, the released bandwidth is distributed as follows. If bandwidth was allotted from the handoff bandwidth pool C_h when the handoff call arrived, the released bandwidth is returned back to the handoff bandwidth pool C_h . If bandwidth for the departing call was obtained from the originating bandwidth pool C_o or by degrading ongoing calls, the released bandwidth by that call is distributed back to the current ongoing calls which are receiving degraded service in the system. The detailed algorithm and flowchart for call departure handling are shown in Algorithm III and Figure 4.3, respectively.

Figure 4.2: Flowchart for Call Arrival Handling

Algorithm III. Call Departure Handling

-
- STEP 1. CLASS i CALL DEPARTURE OCCURS.
- STEP 2. IF DEPARTING CALL WAS INITIALLY GENERATED IN THE SAME CELL, REDISTRIBUTE THE RELEASED BANDWIDTH b_{icur} AMONG THE ONGOING CALLS ACCORDING TO THE PREDEFINED UPGRADE PROCESS, STOP.
- STEP 3. IF DEPARTING CALL WAS HANDED OVER FROM AN ADJACENT CELL AND BANDWIDTH WAS ALLOTTED TO IT FROM THE HANDOFF BANDWIDTH POOL C_h , GIVE BACK THE RELEASED BANDWIDTH TO THE HANDOFF BANDWIDTH POOL C_h , STOP.
- STEP 4. IF DEPARTING CALL WAS HANDED OVER FROM AN ADJACENT CELL AND BANDWIDTH WAS ALLOTTED TO IT FROM THE ORIGINATING BANDWIDTH POOL C_o OR BY DEGRADING ONGOING CALLS, REDISTRIBUTE THE RELEASED BANDWIDTH b_{icur} AMONG THE ONGOING CALLS ACCORDING TO THE PREDEFINED UPGRADE PROCESS, STOP.
-

Figure 4.3: Flowchart for Call Departure Handling

4.3.4 Bandwidth Degrade Process

The following guidelines are used to borrow bandwidth from ongoing calls when there is not enough bandwidth available in the system to accommodate incoming calls:

- (a). Initially, ongoing calls belonging to the lowest priority class, i.e., the traffic class with the highest U_i are selected.
- (b). Calls across the same class i occupying the highest bandwidth levels are the first ones to be degraded.
- (c). While degrading a class i call, its bandwidth is reduced from b_{icur} to b_{imin} . This approach results in freeing up maximum bandwidth for incoming call (either originating call or handoff call) while affecting the least number of ongoing calls thereby reducing QoS fluctuation across the system.
- (d). Calls in a class i already occupying the minimum bandwidth b_{imin} cannot be degraded further

4.3.5 Bandwidth Upgrade Process

When a call departure occurs, the released bandwidth is redistributed in the system according to the following guidelines:

- (a). The calls belonging to the highest priority class, i.e., the traffic class with the lowest U_i are upgraded first.
- (b). Calls across the same class i occupying the lowest bandwidth levels are the first ones to be upgraded.

- (c). When an ongoing class i call is upgraded, its bandwidth is increased from b_{icur} to b_{imax} .
- (d). Calls belonging to class i already occupying the maximum bandwidth b_{imax} cannot be upgraded further.

4.4 QoS Metrics

In the performance evaluation of the DARA scheme, we take into account conventional QoS parameters such as the blocking probabilities of originating calls and handoff calls. Apart from evaluating the traditional performance parameters, we also need to measure the signaling overhead required at the BS for the bandwidth upgrade/degrade process and the proportion of calls receiving degraded service in the system. To accommodate these new features, we introduce two additional QoS metrics which reflect the overhead involved in implementing a dynamic resource allocation scheme as opposed to traditional static schemes such as CS and FP.

4.4.1 Average Degraded Bandwidth (ADB)

The parameter ADB is the average bandwidth held by a class i call in our proposed scheme. The value of ADB can vary from b_{imax} to b_{imin} as the amount of traffic in the system increases. This parameter gives an approximate measure of user satisfaction because it indicates the average level of degraded service that a user is subjected to in our system.

4.4.2 Bandwidth Reallocation Frequency (BRF)

The parameter BRF is defined as the number of times a call undergoes bandwidth reallocation during its lifetime. It reflects the average QoS fluctuation experienced by a user and it is also indicative of the amount of signaling overhead required at the BS to modify the bandwidth of ongoing calls.

In the next chapter, we present the performance analysis of our proposed scheme, DARA, which is accomplished by extensive computer simulations along with detailed discussions of the obtained results.

Chapter 5

Performance Analysis

5.1 Introduction

For the performance evaluation of our proposed scheme, DARA, we adopt an event-driven simulation model. The rest of the chapter is organized as follows:— the simulation model along with the simulation parameters is described in the first part and the simulation results are presented in the next part. The first section details the programming environment and lists down the assumed values for the different simulation parameters. In our experiments, we consider three classes of multimedia traffic with flexible bandwidth requirements comprising of both originating and handoff calls. The next section presents the extensive simulation results for various QoS parameters along with relevant discussions. For the purpose of comparison, we also simulate three other resource allocation schemes. The schemes chosen for performance comparison with our scheme are — HMax, HMin (presented in [52]) and the flexible resource allocation (FRA) scheme proposed in [49] (hereafter referred to as HFRA). As mentioned earlier, HMax and HMin are the boundary conditions for FRA schemes as they always allot the maximum and the minimum number of BUs respectively to all the incoming calls irrespective of their traffic class and current system bandwidth. In other words, HMax

and HMin provide the upper and lower performance bounds for FRA schemes. The HFRA scheme is a dynamic FRA scheme which modifies the bandwidth of both ongoing and incoming calls according to system bandwidth; it serves as the base scheme whose performance we aim to improve in our work. The performance comparison between the various schemes is carried out for traditional QoS parameters such as the blocking probabilities of originating calls and handoff calls. In addition, we also evaluate the performance based on two non-traditional QoS parameters — ADB (Average Degraded Bandwidth) and BRF (Bandwidth Reallocation Frequency).

5.2 Simulation Model

5.2.1 Programming Environment

The programming environment for our simulation experiments is the general purpose programming language, JAVA (build jdk1.5.0) [65], augmented with a discrete-event process oriented simulation package, SimJAVA [66]. SimJAVA is an API (Application Peripheral Interface) which extends the JAVA programming language to conveniently model and execute event-driven simulations. These unique features translate to ease of use for developers and considerable reduction in programming time when used for modeling the random call arrival/departure events in a wireless and mobile network.

5.2.2 Simulation Parameters

In our simulation model, we focus on a single cell of a mobile and wireless network serving three classes ($K = 3$) of multimedia traffic with handling for both originating and

handoff calls. We choose the number of classes of multimedia traffic to be three because the third generation (3G) cellular system identifies four main categories of expected traffic : (a). conversational class; (b). streaming class; (c). interactive class and (d). background class [41]. Since our proposed scheme, DARA, is specifically tailored for multimedia or non-voice traffic with flexible bandwidth requirements, we ignore the conversational class which consists of voice traffic. Additionally, multimedia services are generally divided into three broad groups based on their bandwidth requirements [67]:

1. *Light* multimedia traffic such as E-mails, web browsing and so on.
2. *Medium* multimedia traffic which contains services such as multimedia messaging, file transfer and music downloads.
3. *Heavy* multimedia traffic which needs a lot of bandwidth, such as TV-quality video streaming, video conferencing and broadband Internet.

With suitable selection of the maximum and the minimum bandwidth requirements for each of the three classes of traffic, all the above categories of multimedia traffic can be adequately represented.

Traditionally, the traffic in wireless and mobile systems is modeled as a Poisson arrival process with an exponentially distributed channel holding time. But, recent traffic measurements have shown that the traditional traffic model does not accurately describe the nature of data traffic generated by internet and multimedia services [68][69]. Instead, data traffic is self-similar in nature and is more realistically modeled by heavy-tailed distributions such as the Pareto distribution [57]. To accommodate these different types of distributions, our overall simulation setup is comprised of the following subsets:

1. *Set 1* — The inter-arrival times for originating and handoff calls belonging to class $i \in \{1, 2, \dots, K\}$ are assumed to follow the Pareto distribution with exponentially distributed channel holding times.
2. *Set 2* — The arrival processes for originating and handoff calls belonging to class i are assumed to be Poisson processes with exponentially distributed channel holding times.
3. *Set 3* — Both the inter-arrival times and channel holding times for originating and handoff calls belonging to class i are generated according to the Pareto distribution.
4. *Set 4* — The arrival processes for originating and handoff calls belonging to class i are assumed to be Poisson processes while the channel holding times are generated according to the Pareto distribution.
5. *Set 5* — The performance of our proposed scheme, DARA, is analyzed when subjected to different types of traffic by varying the ‘burstiness’ of call arrivals into the system.

Furthermore, in our simulations, we assume that the mean arrival rate of class i handoff calls is proportional to the mean arrival rate of class i originating calls. This assumption is reasonable because, in effect, class i handoff calls are generated by a fraction of class i originating calls leaving from one cell to another cell due to user mobility. The rest of the values for the various simulation parameters are listed down in Table 5.1.

Table 5.1: Simulation Parameters

<i>Parameter</i>	<i>Symbol</i>	<i>Value</i>
Total number of BUs in a single cell	N	120
Number of classes of multimedia traffic	K	3
Allotted fair share BUs for class 1 traffic	C_1	50
Allotted fair share BUs for class 2 traffic	C_2	30
Allotted fair share BUs for class 3 traffic	C_3	20
Originating bandwidth pool	$C_o = C_1 + C_2 + C_3$	100
Handoff bandwidth pool	C_h	20
(Maximum, Minimum) BUs required by a class 1 call	$(b_{1\max}, b_{1\min})$	(8, 4)
(Maximum, Minimum) BUs required by a class 2 call	$(b_{2\max}, b_{2\min})$	(6, 3)
(Maximum, Minimum) BUs required by a class 3 call	$(b_{3\max}, b_{3\min})$	(4, 2)
Ratio of class 1 and class 2 originating traffic arrival rates	$\lambda_{o1} : \lambda_{o2}$	1:1
Ratio of class i handoff and originating traffic arrival rates	$\lambda_{hi} : \lambda_{oi}$	1:9
Mean channel holding time for a class i call	$1/\mu_i$	120

5.3 Simulation Results and Discussions

In this section, we present the simulation results obtained by comparing the different resource allocation schemes — our proposed scheme DARA, the basic dynamic FRA scheme HFRA and the boundary FRA schemes, HMax and HMin. The system parameters that are chosen for the performance evaluation are the blocking probabilities

of originating calls and handoff calls, BRF (Bandwidth Reallocation Frequency) and ADB (Average Degraded Bandwidth).

5.3.1 Simulation Set 1 (SS1) Results

For the simulation set 1 (SS1), the inter-arrival times for all the three classes ($K = 3$) of traffic are generated according to the Pareto distribution with the shape parameter α_{arr} fixed at 1.50. The channel holding times for both originating and handoff calls belonging to class $i \in \{1, 2, \dots, K\}$ are exponentially distributed with mean channel holding time of $1/\mu_i = 120s$.

Figure 5.1: (SS1) Comparison of the Blocking Probability for Class 1 Traffic

Figure 5.2: (SS1) Comparison of the Blocking Probability for Class 2 Traffic

Figure 5.3: (SS1) Comparison of the Blocking Probability for Class 3 Traffic

Figure 5.1, Figure 5.2 and Figure 5.3 plot the blocking probabilities for calls belonging to class 1 traffic, class 2 traffic and class3 traffic respectively against their traffic arrival rates. The proposed DARA scheme offers much better performance than the HFRA scheme presented in [49] by having lower blocking probability over different values of traffic load for all the three classes of traffic. This is due to the fact that for an under-loaded class i originating call, the HFRA scheme tries to allocate only $b_{i\max}$ bandwidth either from the free system bandwidth or by degrading lower priority class calls. This is a

sub optimal implementation because if $b_{i_{\max}}$ bandwidth is not available, $b_{i_{\min}}$ bandwidth might still be available and can be allotted to the incoming call to increase the system capacity.

The performance of HMax and HMin strategies is also shown and as expected, both the DARA and the HFRA schemes offer performance in between these upper and lower bounds. In addition, our proposed DARA scheme provides better QoS guarantees for the higher priority handoff calls as can be seen by the DARA_HO plot in the figures.

Figure 5.4: (SS1) Comparison of the ADB for Class 1 Traffic

Figure 5.5: (SS1) Comparison of the ADB for Class 2 Traffic

Figure 5.6: (SS1) Comparison of the ADB for Class 3 Traffic

In Figure 5.4, Figure 5.5 and Figure 5.6, the ADB (Average Degraded Bandwidth) for calls belonging to class 1 traffic, class 2 traffic and class 3 traffic are plotted against their respective traffic arrival rates. From the figures, it is clear that both the DARA and the HFRA schemes perform a graceful degradation of the ADB from $b_{i\max}$ at low traffic loads to $b_{i\min}$ at heavy traffic loads. However, it should be noted that the HFRA scheme does perform marginally better than the proposed DARA scheme. This is because in the HFRA implementation, whenever bandwidth is released upon call departure from the system, it is re-distributed equally to maximum possible ongoing calls. On the other hand, in the DARA implementation, the released bandwidth is distributed to the minimum number of ongoing calls in order to minimize the overall BRF (Bandwidth Reallocation Frequency) experienced by ongoing calls in the system.

Figure 5.7: (SS1) Comparison of the BRF for Class 1 Traffic

Figure 5.8: (SS1) Comparison of the BRF for Class 2 Traffic

Figure 5.9: (SS1) Comparison of the BRF for Class 3 Traffic

As Figure 5.7, Figure 5.8 and Figure 5.9 show, the proposed scheme, DARA, significantly reduces the average BRF (Bandwidth Reallocation Frequency) experienced by a call when compared to the HFRA scheme for all the three different classes of traffic. This reduction in BRF translates to a higher user satisfaction due to the reduced QoS fluctuation experienced by the user and also helps to reduce the signaling overhead required to modify the bandwidth of existing calls. The reason for the low value of BRF can be attributed to the fact that Δb_i is dynamic in our scheme and can take integer values in the range $[1, b_{i\max} - b_{i\min}]$ as opposed to the constant value of 1 for Δb_i in the HFRA scheme. As a result, fewer calls are subjected to the upgrade/degrade procedure in DARA than in the compared HFRA scheme to redistribute/borrow the same number of BUs.

5.3.2 Simulation Set 2 (SS2) Results

In this simulation set, we adopt the traditional traffic model historically used to model the nature of call arrivals and channel holding times in a wireless and a mobile system. The call arrival processes for all the three classes of traffic are modeled as Poisson processes while the channel holding times for these calls are exponentially distributed with a mean value of $1/\mu_i = 120\text{s}$.

Figure 5.10: (SS2) Comparison of the Blocking Probability for Class 1 Traffic

Figure 5.11: (SS2) Comparison of the Blocking Probability for Class 2 Traffic

Figure 5.12: (SS2) Comparison of the Blocking Probability for Class 3 Traffic

Figure 5.10, Figure 5.11 and Figure 5.12 plot the blocking probabilities for calls belonging to class 1 traffic, class 2 traffic and class3 traffic against their respective traffic arrival rates. The blocking probabilities for all the different classes of traffic follow a similar trend as observed in the SS1 (simulation set 1) results. In this case also, the proposed scheme, DARA, outperforms the HFRA scheme with substantially lower blocking probabilities for all the classes of traffic.

Figure 5.13: (SS2) Comparison of the ADB for Class 1 Traffic

Figure 5.14: (SS2) Comparison of the ADB for Class 2 Traffic

Figure 5.15: (SS2) Comparison of the ADB for Class 3 Traffic

In Figure 5.13, Figure 5.14 and Figure 5.15, the ADB (Average Degraded Bandwidth) for calls belonging to class 1 traffic, class 2 traffic and class 3 traffic are plotted against their respective traffic arrival rates. Similar to the results obtained for the simulation set SS1, both the DARA and the HFRA schemes degrade the ADB held by a class i call from $b_{i\max}$ at low traffic loads to $b_{i\min}$ at heavy traffic loads with the HFRA scheme performing slightly better than the DARA scheme.

Figure 5.16: (SS2) Comparison of the BRF for Class 1 Traffic

Figure 5.17: (SS2) Comparison of the BRF for Class 2 Traffic

Figure 5.18: (SS2) Comparison of the BRF for Class 3 Traffic

Figure 5.16, Figure 5.17 and Figure 5.18 plot the average BRF (Bandwidth Reallocation Frequency) experienced by calls belonging to class 1 traffic, class 2 traffic and class 3 traffic respectively against their traffic arrival rates. The observed results are comparable to the results obtained for the simulation set (SS1), i.e., the proposed DARA scheme achieves a significant reduction in average BRF when compared to the HFRA scheme.

5.3.3 Simulation Set 3 (SS3) Results

For the simulation set 3 (SS3), both the inter-arrival times and the channel holding times for calls belonging to all the three classes of traffic follow Pareto distributions. To generate the inter-arrival times of all the three classes of traffic, the shape parameter for the arrival Pareto distribution α_{arr} is chosen to be 1.50. The channel holding times for both originating and handoff calls belonging to class $i \in \{1, 2, \dots, K\}$ are generated using a Pareto distribution with shape parameter α_{hold_time} equal to 1.35 and a mean channel holding time of $1/\mu_i = 120s$.

In the analysis of this simulation set, it was observed that the performance of the compared schemes with regard to chosen QoS metrics, namely — blocking probabilities of originating and handoff calls, BRF (Bandwidth Reallocation Frequency) and ADB (Average Degraded Bandwidth), are similar to the results obtained for simulations sets - SS1 and SS2. For the sake of completeness, the obtained simulation figures are included below (Figure 5.19 to Figure 5.27). To interpretation these results, the reader is directed to the detailed discussions presented for simulation set 1 (SS1) results in the preceding sections as they hold true for this set too.

Figure 5.19: (SS3) Comparison of the Blocking Probability for Class 1 Traffic

Figure 5.20: (SS3) Comparison of the Blocking Probability for Class 2 Traffic

Figure 5.21: (SS3) Comparison of the Blocking Probability for Class 3 Traffic

Figure 5.22: (SS3) Comparison of the ADB for Class 1 Traffic

Figure 5.23: (SS3) Comparison of the ADB for Class 2 Traffic

Figure 5.24: (SS3) Comparison of the ADB for Class 3 Traffic

Figure 5.25: (SS3) Comparison of the BRF for Class 1 Traffic

Figure 5.26: (SS3) Comparison of the BRF for Class 2 Traffic

Figure 5.27: (SS3) Comparison of the BRF for Class 3 Traffic

5.3.4 Simulation Set 4 (SS4) Results

The traffic model adopted for this simulation set is as follows: – the call arrival processes for all the three classes of traffic are modeled as Poisson processes while the channel holding times by these calls are generated according to a Pareto distribution with scale parameter α_{hold_time} equal to 1.35 and a mean channel holding time of $1/\mu_i = 120s$.

The simulation results obtained for the different QoS metrics, namely — blocking probabilities of originating and handoff calls, BRF (bandwidth reallocation frequency) and ADB (Average Degraded Bandwidth), are presented in Figure 5.28 through Figure 5.36. In this case also, the observed results are similar to the results obtained for simulations sets - SS1 and SS2. For further explanation of these results, please refer to discussions presented for simulation set 1 (SS1) results in the preceding sections as they hold true for this set too.

Figure 5.28: (SS4) Comparison of the Blocking Probability for Class 1 Traffic

Figure 5.29: (SS4) Comparison of the Blocking Probability for Class 2 Traffic

Figure 5.30: (SS4) Comparison of the Blocking Probability for Class 3 Traffic

Figure 5.31: (SS4) Comparison of the ADB for Class 1 Traffic

Figure 5.32: (SS4) Comparison of the ADB for Class 2 Traffic

Figure 5.33: (SS4) Comparison of the ADB for Class 3 Traffic

Figure 5.34: (SS4) Comparison of the BRF for Class 1 Traffic

Figure 5.35: (SS4) Comparison of the BRF for Class 2 Traffic

Figure 5.36: (SS4) Comparison of the BRF for Class 3 Traffic

5.3.5 Simulation Set 5 (SS5) Results

In this simulation set, we investigate the performance of the proposed scheme, DARA, when it is subjected to different levels of ‘burstiness’ in incoming traffic. For the traffic model, we utilize the Pareto distribution to generate call arrivals and the exponential distribution to model the channel holding times.

To vary the ‘burstiness’ in incoming traffic, we use three distinct values — 1.2, 1.5 and 1.8 for α_{arr} , the shape parameter of the Pareto distribution which describes the nature of call arrivals into the system. It is to be noted here that as the value of α_{arr} becomes closer to 1, the traffic becomes more ‘bursty’ in nature. In addition, results for the traditional traffic model, i.e., call arrivals generated by a Poisson process, are also included for comparison purposes. The channel holding times for all the three classes of traffic follow the exponential distribution with a mean channel holding time of $1/\mu_i = 120$ s. In the rest of the figures, the ‘alpha 1.2’, ‘alpha 1.5’ and ‘alpha 1.8’ plots denote the results obtained when the value of α_{arr} was chosen to be 1.2, 1.5 and 1.8 respectively while the ‘Poisson’ plot indicates the results obtained when the call arrivals were modeled as a Poisson process.

Figure 5.37: (SS5) Comparison of the Blocking Probability for Class 1 Traffic

Figure 5.38: (SS5) Comparison of the Blocking Probability for Class 2 Traffic

Figure 5.39: (SS5) Comparison of the Blocking Probability for Class 3 Traffic

Figure 5.37, Figure 5.38 and Figure 5.39 plot the blocking probabilities for originating calls belonging to class 1 traffic, class 2 traffic and class 3 traffic respectively against their traffic arrival rates. From the figures, it is clear that for all the three different classes of traffic, the blocking probabilities become higher as α_{arr} is decreased from 1.8 to 1.2. This is understandable because as α_{arr} becomes closer to 1, the traffic becomes more ‘bursty’ and it is quite possible that more calls are lost due to tighter arrivals to the system. Incidentally, the blocking probabilities of the “Poisson” plot are very close to the

‘alpha 1.8’ plots in the figures. This is indicative of the non-bursty nature of traffic when it is modeled as a Poisson process.

Figure 5.40: (SS5) Comparison of the ADB for Class 1 Traffic

Figure 5.41: (SS5) Comparison of the ADB for Class 2 Traffic

Figure 5.42: (SS5) Comparison of the ADB for Class 3 Traffic

From Figure 5.40, Figure 5.41 and Figure 5.42, it can be deduced that the ADB (Average Degraded Bandwidth) held by a class $i \in \{1, 2, \dots, K\}$ call is reduced as the value of α_{arr} is reduced from 1.8 to 1.2. This is also intuitive because as the call arrivals become more closely spaced (α_{arr} closer to 1), more number of calls tend to arrive together at the system. As a result, the total free bandwidth in the system has to be shared between an increased number of calls. Therefore the ADB, which is a measure of the average bandwidth held by a call, reduces as the value of α_{arr} moves closer to 1. Because of the non-bursty nature of a Poisson arrival process, when the call arrivals are modeled as a Poisson process (indicated by the ‘Poisson’ plot), the value of the ADB is higher than when the inter-arrival times are modeled as a Pareto distribution (indicated by the ‘alpha 1.2’, ‘alpha 1.5’ and ‘alpha 1.8’ plots).

The conclusions of our research and potential future work in related areas are discussed in the next chapter.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we have proposed and analyzed a resource management approach, namely Dynamic Adaptive Resource Allocation (DARA) scheme, which exploits the flexible bandwidth requirements of multimedia traffic to increase system capacity of wireless and mobile networks.

Initially, a brief discussion on the emerging multimedia services in current and future wireless systems and the associated resource management issues has been presented. To address these issues, numerous resource management techniques have been published in recent literatures which handle multiple classes of traffic. A review of some of these resource management approaches has been documented in this work highlighting their advantages and disadvantages. Our proposed scheme, DARA, is then described in detail along with the system model, the traffic model and the bandwidth allocation algorithm. In the performance analysis of our proposed scheme, the extensive simulation results indicated that the proposed scheme provided enhanced system capacity by allowing incoming calls to borrow bandwidth from ongoing calls at heavy traffic loads. Also, it has

been demonstrated that the proposed system improves QoS parameters such as the call blocking probabilities of originating calls and handoff calls for the different classes of traffic. In addition, by introducing a dynamic bandwidth borrow unit Δb_i , the QoS fluctuation or the BRF (Bandwidth Reallocation Frequency) experienced by a user was considerably reduced when compared to a similar FRA scheme [49].

6.2 Future Work

In this section, we introduce some of the prospective future issues in our research work. An improved priority assignment scheme could be implemented which assigns higher priority to calls expected to leave the system soonest either by call termination or by being handed over to another cell. This would further lower the signaling overhead at the BS and the average QoS fluctuation because calls which are anyway going to release bandwidth back to the system within a short period of time will not be subjected to the degradation procedure. Another challenging task would be to build an analytical model for our proposed scheme and to conduct system performance analysis in more detail with both analytical and simulation models. Some other potential interesting areas could be a dynamic partition for the handoff bandwidth pool C_h and intelligent selection of the bandwidth partitions C_i , where $i \in \{1, 2, \dots, K\}$, based on the analytical model.

Bibliography

- [1] P. Chaudhary, W. Mohr and S. Onoe, “The 3GPP Proposal for IMT-2000,” *IEEE Communications Magazine*, vol. 37, no. 12, pp 72-81, Dec. 1999.
- [2] M. Shafi, A. Hashimoto, M. Umehira, S. Ogose, and T. Murase, “Wireless communications in the twenty-first century: A perspective,” *Proceedings of the IEEE*, vol. 85, pp. 1622–1638, Oct. 1997.
- [3] Nobel Lectures, “Guglielmo Marconi – Biography,” Available: <http://nobelprize.org/physics/laureates/1909/marconi-bio.html>.
- [4] D. P. Agrawal and Q-A. Zeng, *Introduction to Wireless and Mobile Systems*, Thomson Publishing, USA, 2005.
- [5] P. Nicopolitidis, M. S. Obaidat, G. I. Papadimitriou and A. S. Pomportis, *Wireless Networks*, John Wiley & Sons, Chicester, England, 2003.
- [6] M. Zeng, A. Annamalai and V. K. Bhargava, “Recent Advances in Cellular Wireless Communications,” *IEEE Communications Magazine*, vol. 37, no. 9, pp. 128-138, Sep. 1999.
- [7] V. H. MacDonald, “AMPS: The Cellular Concept,” *BELL System Technical Journal*, vol. 58, no. 1, pp.15-49, Jan. 1979.
- [8] J. E. Padgett and C. G. Gunther, “Overview of Wireless Personal Communications,” *IEEE Communications Magazine*, vol. 33, no.1, pp. 28-41, Jan. 1995.

- [9] W. Johnston, "Europe's Future Mobile Telephony System," *IEEE Spectrum Magazine*, vol. 35, no. 10, pp. 49-53, Oct. 1998.
- [10] J. Uddenfeldt, "Digital Cellular – its Roots and its Future," *IEEE Proceedings*, vol. 86, no. 7, pp. 1319–1324, July 1998.
- [11] U. Black, *Second Generation Mobile & Wireless Networks*, Prentice Hall, New Jersey, USA, 1999.
- [12] M. Rahnema, "Overview of the GSM system and protocol architecture," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 92–100, Apr. 1993.
- [13] M. R. L. Hodges, "The GSM Radio Interface," *British Telecom Technology Journal*, vol. 8, no. 1, pp. 31-43, Jan. 1990.
- [14] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*, published by the authors, France, 1992.
- [15] C. Scholefield, "Evolving GSM Data Services," *Proceedings of the IEEE International Conference on Universal Personal Communications*, vol. 2, pp. 888-892, Oct. 1997.
- [16] GSM Technical Specification, "GSM 04.11, Point-to-Point (PP) Short Message Service (SMS) Support on Mobile Radio Interface," Version 5.3.0. ETSI, Feb. 1996.
- [17] GSM Technical Specification, "GSM 02.34, High Speed Circuit Switched Data (HSCSD), Stage I," Version 5.2.0. ETSI, July 1997.

- [18] GSM Technical Specification, "GSM 03.34, High Speed Circuit Switched Data (HSCSD), Stage 2+," Version 5.2.0. ETSI, May 1999.
- [19] GSM Technical Specification, "GSM 02.60 version 6.1.0, GPRS Service Description, Stage I," ETSI 1998.
- [20] GSM Technical Specification, "GSM 03.60 version 7.4.0, Digital cellular telecommunications system (Phase 2+), General Packet Radio Service (GPRS) Service Description - Stage 2," ETSI, 2000.
- [21] GSM Technical Specification, "GSM 03.64 version 7.1.0, Digital cellular telecommunications system (Phase 2+), Overall description of the GPRS radio interface, Stage 2," ETSI 1999
- [22] TIA/EIA IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread-Spectrum Cellular Systems," Telecommunications Industry Association, July 1993.
- [23] R. Padovani, "The application of spread spectrum to PCS has become a reality: Reverse Link Performance of IS-95 Based Cellular Systems," *IEEE Personal Communications*, vol. 1, no. 3, pp. 28-34, 3Q 1994.
- [24] TIA/EIA-95-B, "Mobile Station-Base Station Compatibility Standard for Wideband Spread Spectrum Cellular Systems," Telecommunications Industry Association, Mar. 1999.
- [25] Y-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architecture*, John Wiley & Sons, New York, USA, 2000.

- [26] TIA/EIA IS-54, "Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard," Telecommunications Industry Association, Apr. 1992.
- [27] TIA/EIA IS-136.1, "800 MHz TDMA Cellular - Radio Interface - Mobile Station - Base Station Compatibility - Digital Control Channel," Telecommunications Industry Association, Dec. 1994.
- [28] B. Sarikaya, "Packet Mode in Wireless Networks: Overview of Transition to Third Generation," *IEEE Communications Magazine*, vol. 38, no. 9, pp. 164-172, Sep. 2000.
- [29] R. Prasad, W. Mohr and W. Konhauser, *Third Generation Mobile Communication Systems*, Artech House Publishers, Boston, USA, 2000.
- [30] M. H. Callendar, "Future Public Land Mobile Telecommunication Systems," *Proceedings of the IEEE International Conference on Universal Personal Communications*, vol. 1, pp. 568-572, Oct. 1994.
- [31] M. H. Callendar, "International Mobile Telecommunications-2000 Standards Efforts of the ITU," *IEEE Personal Communications*, vol. 4, no. 4, pp. 6-7, Aug. 1997.
- [32] GSA Press Release, "GSA Says Number of Commercial 3G/WCDMA Networks Leaps to 100," Dec. 2005, Available: http://www.gsacom.com/news/gsa_189.php4
- [33] W. Mohr, "Development of Mobile Communications Systems Beyond Third Generation," *Wireless Personal Communications: An International Journal*, Kluwer Academic Publishers, vol. 17, nos. 2-3, pp. 191-207, June 2001.

- [34] J. Lilliberg and R. Prasad, "Research Challenges for 3G and Paving the Way for Emerging New Generations," *Wireless Personal Communications: An International Journal*, Kluwer Academic Publishers, vol. 17, nos. 2-3, pp. 355-362, June 2001.
- [35] U. Varshney and R. Jain, "Issues in Emerging 4G Wireless Networks," *IEEE Computer*, vol. 34, no. 6, pp. 94-96, June 2001.
- [36] Wideband Orthogonal Frequency division Multiplexing (W-OFDM), Wi-LAN Inc., version 1.0, 2000.
- [37] R. W. Chang and R. A. Gibby, "A Theoretical Study of Performance of an Orthogonal Multiplexing Data Transmission Scheme," *IEEE Transactions on Communications*, vol. 16, no. 4, pp. 529-540, Aug. 1968.
- [38] F. A. Cruz-Perez and L. Ortigoza-Guerrero, "Resource Sharing for Data Services (RSDS)," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 2, pp. 995-999, Oct. 2001.
- [39] P. C. Whale, "Challenges and trends in 3G devices," *Proceedings of the International Conference on 3G Mobile Communication Technologies*, vol. 494, pp. 173-177, June 2003.
- [40] J. Wang, *Broadband Wireless Communications: 3G, 4G and Wireless LAN*, Kluwer Academic Publishers, Norwell, USA, 2001.
- [41] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, Chicester, England, 2001.

- [42] L. Ortigoza-Guerrero, F. A. Cruz-Perez and H. Heredia-Ureta, "Call Level Performance Analysis for Multiservices Wireless Cellular Networks With Adaptive Resource Allocation Strategies," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 4, pp. 1455-1472, July 2005.
- [43] A. Dornan, *The Essential Guide to Wireless Communications Applications*, Prentice Hall, New York, USA, 2001.
- [44] L. Li, B. Li, B. Li and X-R. Cao, "Performance Analysis of Bandwidth Allocations for Multi-services Mobile Wireless Cellular Networks," *Proceedings of IEEE Wireless Communications and Networking Conference*, vol. 4, no. 1, pp. 1072-1077, Mar. 2003.
- [45] C-T. Chou and K. G. Shin, "Analysis of Adaptive Bandwidth Allocation in wireless Networks with Multilevel Degradable Quality of Service," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 5-17, Jan.-Mar. 2004.
- [46] T. S. Randhawa and R. H. S. Hardy, "Performance Analysis of Multi-service Cellular Networks with Mobile Users," *Proceedings of the IEEE Wireless Communications and Networking Conference*, vol. 3, no. 1, pp. 1390-1397, Sep. 2000.
- [47] Y. Xiao, C. L. P. Chen, and Y. Wang, "Fair Bandwidth Allocation for Multi-class of Adaptive Multimedia Services in Wireless/Mobile Networks," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3, pp. 2081–2085, May 2001.
- [48] S. Sen, J. Jawanda, K. Basu and S. Das, "Quality-of service Degradation Strategies in Multimedia Wireless Networks," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3, pp. 1884-1888, May 1998.

- [49] H. Chen, Q-A. Zeng and D. P. Agrawal, "Evaluation of a new Adaptive Resource Management Scheme for Multiclass Wireless and Mobile Networks," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 4, pp. 2187-2191, Oct. 2003.
- [50] P. K. Lakshmanan and Q-A. Zeng, "Dynamic Adaptive Resource Allocation Scheme for Multimedia Services in Wireless and Mobile Networks," *Proceedings of the IEEE World Telecommunications Symposium (WTS)*, Apr. 2006.
- [51] P. K. Lakshmanan, R. Lenagala, H. Liu and Q-A. Zeng, "An Adaptive Call Admission and Resource Management Scheme for Multi-class Wireless and Mobile Networks," *Proceedings of the International Conference on Wireless Networks (ICWN)*, pp. 311-316, June 2005.
- [52] J-Y. Jeng, C-W. Lin and Y-B. Lin, "Dynamic Scheduling for GSM Data Services," *IEICE Transactions on Communications*, vol. E80-B, no. 2, pp. 296-300, Feb. 1997.
- [53] F. A. Cruz-Perez and L. Ortigoza-Guerrero, "Equal Resource Sharing Allocation with QoS Differentiation for Conversational Services in Wireless Communication Networks," *IEE Proceedings – Communications*, vol. 150, no. 5, pp. 391-398, Oct. 2003.
- [54] Y. Wang, W. Wang, and S. Cheng, "Priority-based QoS Support for Multiservice Mobile Networks," *Proceedings of the IEEE Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 2059–2063, Sep. 2003.
- [55] S. Choi and K. Shin, "Location/mobility-dependent Bandwidth Adaptation in QoS-sensitive Cellular Networks," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3, pp. 1593–1597, Oct. 2001.

- [56] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, Tata McGraw-Hill Publishing, New Delhi, India, 2002.
- [57] J. Gordon, "Pareto Process as a Model of Self-similar Packet Traffic," *Proceedings of the IEEE Global Telecommunications Conference*, vol. 3, pp. 2232–2235, Nov. 1995.
- [58] Q-A. Zeng and D. P. Agrawal, *Handbook of Wireless Networks and Mobile Computing*, ch. Handoff in Wireless Mobile Networks, John Wiley & Sons, New York, USA, 2002.
- [59] E. A. Frech and C. L. Mesquida, "Cellular Models and Handoff Criteria," *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1, pp. 128-135, May 1989.
- [60] N. D. Tripathi, J. H. Reed and H. F. VanLandinoham, "Handoff in Cellular Systems," *IEEE Personal Communications*, vol. 5, no.6, pp. 26-37, Dec. 1998.
- [61] O. Grimlund and B. Gudmundson, "Handoff Strategies in Microcellular Systems," *Proceedings of the IEEE Vehicular Technology Conference*, pp. 505-510, May 1991.
- [62] Y-H. Bae, Y-T. Lee, H-S. Choi and W-J. Park, "Capacity Planning for IMT-2000 Using Long Range Dependent Traffic," *Proceedings of the IEEE Vehicular Technology Conference*, vol.2, pp. 909-913, Sep. 1999.
- [63] B.C. Sowden and K.W. Sowerby, "The Impact of Long-range Dependent Traffic in a CDMA System Supporting Real-time Services," *Proceedings of the IEEE Global Telecommunications Conference*, vol.6, pp. 3509-3513, Nov. 2001.

- [64] Q-A. Zeng, K. Mukumoto and A. Fukuda, "Performance Analysis of Mobile Cellular Radio System with Two-level Priority Reservation Handoff Procedure," *IEICE Transactions on Communications*, vol. E80-B, no. 4, pp. 598–607, 1997.
- [65] Java Programming Language, "Java 2 Platform Standard Edition (J2SE)," Available: <http://java.sun.com/j2se/>
- [66] Introduction to SimJAVA, "SimJAVA 2," Available: <http://www.icsa.informatics.ed.ac.uk/research/groups/hase/simjava/>
- [67] A. Krendzel, Y. Koucheryavy, J. Harju, S. Lopatin, "Method for Estimating Parameters of 3G Data Traffic," *Proceedings of the IEEE International Conference on Communications*, vol.7, pp. 4312- 4316, June 2004.
- [68] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, "On the Self-similar Nature of Ethernet Traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol.2, no.1, pp. 1-15, Feb. 1994.
- [69] V. Paxson and S. Floyd, "Wide Area Traffic: the Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol.3, no.3, pp.226-244, June 1995.