A MAC Protocol for IP-based CDMA Wireless Networks

By Simon Bonginkosi Mahlaba

University of KwaZulu-Natal

2005

Submitted in fulfillment of the academic requirement for the degree of Master of Science in Engineering in the School of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, 2005

Abstract

The evolution of the internet protocol (IP) to offer quality of service (QoS) makes it a suitable core network protocol for next generation networks (NGN). The QoS features incorporated to IP will enable future IP-based wireless networks to meet QoS requirements of various multimedia traffic. The Differentiated Service (Diffserv) architecture is a promising QoS technology due to its scalability which arises from traffic flow aggregates. For this reason, in this dissertation a network infrastructure based on DiffServ is assumed. This architecture provides assured service (AS) and premium service (PrS) classes in addition to best-effort service (BE). The medium access control (MAC) protocol is one of the important design issues in wireless networks. In a wireless network carrying multimedia traffic, the MAC protocol is required to provide simultaneous support for a wide variety of traffic types, support traffic with delay and jitter bounds, and assign bandwidth in an efficient and fair manner among traffic classes. Several MAC protocols capable of supporting multimedia services have been proposed in the literature, the majority of which were designed for wireless ATM (Asynchronous Transfer Mode). The focus of this dissertation is on time division multiple access and code division multiple access (TDMA/CDMA) based MAC protocols that support QoS in IP-based wireless networks.

This dissertation begins by giving a survey of wireless MAC protocols. The survey considers MAC protocols for centralised wireless networks and classifies them according to their multiple access technology and as well as their method of resource sharing. A novel TDMA/CDMA based MAC protocol incorporating techniques from existing protocols is then proposed. To provide the above-mentioned services, the bandwidth is partitioned amongst AS and PrS classes. The BE class utilizes the remaining bandwidth from the two classes because it does not have QoS requirements. The protocol employs a demand assignment (DA) scheme to support traffic from PrS and AS classes. BE traffic is supported by a random reservation access scheme with dual multiple access interference (MAI) admission thresholds. The performance of the protocol, i.e. the AS or PrS call blocking probability, and BE throughput are evaluated through Markov analytical models and Monte-Carlo simulations. Furthermore, the protocol is modified and incorporated into IEEE 802.16 broadband wireless access (BWA) network.

Preface

The research work presented in this dissertation was performed by Mr. Simon Mahlaba, under the supervision of Prof. Fambirai Takawira, at the Centre for Radio Access Technologies in the School of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Howard College. This work was partially supported by Alcatel SA, THRIP and Telkom South Africa as part of the Centres of Excellence programme at the University of KwaZulu-Natal.

The whole dissertation, unless specifically indicated to the contrary in the text, is the author's own work and has not been submitted in part or in whole to any other university for degree purpose.

Acknowledgements

Firstly, I would like to give special thanks to Professor Takawira for his invaluable help, guidance and patience as my supervisor over the last two years.

Secondly, a great thanks is owed to my family for their patience, encouragement and support, not only for the past two years, but throughout my education. In particular, I would like to give a special thanks to my grandparents who laid the foundation of my life and have always been a source of my inspiration.

Thanks are also owed to Telkom S.A for financing my tertiary education studies and the sponsors of the Centre for Radio Access Technologies namely Telkom S.A, Alcatel S.A, and THRIP.

Finally, thanks go to my friends and postgraduate colleagues for their assistance and making the past two years an enjoyable and memorable experience.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 3

Chapter 5

List of Acronyms

3GPP	: Third Generation Partnership Project
ACK	: Acknowledgement
ABR	: Available Bit Rate
ARIB	: Association for Radio Industries and Businesses
ARQ	: Automatic Repeat Request
ATM	: Asynchronous Transfer Mode
AS	: Assured Service
BE	: Best Effort
BEB	: Binary Exponential Backoff
BER	: Bit Error Rate
BS	: Base Station
BWA	: Broadband Wireless Access
CAC	: Call Admission Control
CBR	: Constant Bit Rate
CDMA	: Code Division Multiple Access
CLS	: Channel Load Sensing
CRA	: Collision Resolution Algorithm
CRC	: Cyclic Redundancy Check
CSMA	: Carrier Sense Multiple Access
DA	: Demand Assignment
DECT	: Digital European Cordless Telephony
DiffServ	: Differentiated Service
DSCP	: DiffServ Code Point
EDGE	: Enhanced Data Rate for Global Evolution
ETSI	: European Telecommunication Standards Institute
FDD	: Frequency division duplex
FDMA	: Frequency Division Multiple Access
FEC	: Forward Error Correction
FIFO	: First In First Out
FTP	: File Transfer Protocol
GSM	: Global System for Mobile Communication
IEEE	: Institute of Electrical and Electronics Engineers

IETF	: Internet Engineering Task Force
IP	: Internet Protocol
ITU	: International Telecommunication Union
ISMA	: Idle Sense Multiple Access (or Inhibit Sense Multiple Access)
LAN	: Local Area Network
MAC	: Medium Access Control
MAN	: Metropolitan Area Network
MAI	: Multiple Access Interference
MPEG	: Motion Picture Expect Group
MT	: Mobile Terminal
NACK	: Negative acknowledgement
nrtPS	: non real time Polling Service
OFDM	: Orthogonal Frequency Division Multiplex
OFDMA	: Orthogonal Frequency Division Multiple Access
OSI	: Open System Interconnection
PAN	: Personal Area Network
PDU	: Protocol Data Unit
PN	: Pseudo-random Noise
PRMA	: Packet Reservation Multiple Access
PrS	: Premium Service
QoS	: Quality of Service
REQ	: Request
RRA	: Random Reservation Access
rtPS	: real time Polling Service
S-ALOHA	: slotted ALOHA
SDU	: Service Data Unit
SIR	: Signal-to-Interference Ratio
SINR	: Signal-to-Interference and Noise Ratio
SS	: Subscriber Station
SS/ALOHA	: Spread-spectrum ALOHA
TDD	: Time division duplex
TDMA	: Time Division Multiple Access
UBR	: Unspecified Bit Rate
UGS	: Unsolicited Grant Service
UMTS	: Universal Mobile Telecommunications System
VBR	: Variable bit rate
Wi-Fi	: Wireless Fidelity Forum

WiMAX : Wireless Interoperability for Microwave Access Forum

Chapter 1

Introduction

1.1 Overview of Wireless Access Networks

Wireless access networks have become an important and convenient access technology to provide telecommunication services. The advantages offered by wireless access networks include support for mobility, easy and low cost deployment (especially in areas which lack existing copper infrastructure or have low tele-density) and solution for areas where cable theft is a problem. Wireless access networks include cellular mobile networks, wireless local area networks (WLANs), wireless personal area networks (WPANs), and wireless metropolitan area networks (WMANs). These systems are differentiated by their application, support for mobility, channel access methods, and coverage area.

1.1.1 Cellular Networks

Cellular networks are the broadest range wireless networks, providing global coverage. The first generation (1G) cellular systems were characterized by transmission of voice only, and had severe capacity limitations. An example of a 1G system is Advanced Mobile Phone system (AMPS) [5]. Developments in digital technology resulted in the emergence of the second generation (2G) cellular systems that support voice and low speed data. 2G cellular standards include global system for mobile communication (GSM), IS-95, and cdmaOne. Third generation (3G) cellular networks are characterized by providing multimedia service and achieving a maximum bit rate of 2 Mb/s. 3G systems were developed and approved under the International Mobile Telecommunications 2000 (IMT-2000) program formed by the International Telecommunication Union (ITU). There are five IMT-2000 radio interfaces. These are wideband CDMA, CDMA2000, CDMA TDD, TDMA single carrier (for EDGE), and FDMA/TDMA (for DECT). CDMA2000 is the evolution of cdmaOne, while EDGE is the evolution of the GSM system.

1.1.2 WLANs

WLANs were designed for short range communication in a corporate, public, or home environment. Today, WLANs are found in enterprise networks, homes and public hotspots such as schools, hotels, coffee shops, and airports. They allow a wireless terminal (e.g. notebook) to access the Internet and various other network resources via a nearby radio access point (AP). Most deployed WLANs are based on the IEEE 802.11a/b/g standards. The promotion and interoperability of 802.11b WLANs is ensured by the Wireless Widelity Forum (Wi-Fi). Other standards for WLANs are ETSI High Performance local area network 2 (HiperLAN/2) and ARIB high speed wireless access network (HiSWAN). WLANs operate in the 2.4 and 5 GHz bands. The 5 GHz WLANs such as IEEE 802.11a and HiperLAN/2 uses orthogonal frequency division multiplexing (OFDM) modulation and achieve data rate up to 54 Mb/s. IEEE 802.11b uses DS-CDMA and provides approximately 11 Mb/s in 2.4 GHz bands. The typical coverage area for a WLAN is in the range of 100-500 m.

1.1.3 WPANs

The objective of WPANs is to provide interconnection between devices (e.g. laptop computer, cell phone and palmtop) within a personal operating space. The coverage area for WPANs is less than 10m. The standard for WPANs is IEEE 802.15, which is also known as Bluetooth. Bluetooth operates in the 2.4 GHz band at 720 kb/s, using frequency hoping spread spectrum (FHSS) consisting of 79 channels.

1.1.4 WMANs

WMANs are designed to provide wireless last-mile broadband access. They offer broadband access alternatives to cable access networks such as, fiber optics links, coaxial systems using cable modems and digital subscriber line (DSL) links. The advantage of WMAN systems is the capacity to address a broad geographic area without the costly infra-structure development required in deploying cable links to individual sites. The standards defining WMANs are IEEE 802.16, ETSI High Performance Metropolitan Area Network (HiperMAN), and ETSI HiperACCESS. These standards share many functions, as they were developed in close co-operation between the standards organisations. Since WMANs have large coverage, they can serve as a backhaul to WLAN hotspots. The deployment of 802.16 systems is facilitated by the Worldwide Interoperability for Microwave Access Forum (WiMAX). The objective of WiMAX is to ensure interoperability between different vendor IEEE 802.16 systems. The IEEE 802.16a/REVd/e standards describe radio link interfaces that operate in the 10-66 GHz and 2-11

GHz frequency bands. The lower frequency bands support non line-of-sight propagation which is not possible at higher frequencies. There are three physical layer specifications based on single carrier format, 256-point fast Fourier transform (FFT) orthogonal frequency division multiplexing, and 2048-point FFT orthogonal frequency division multiple access (OFDMA). An IEEE 802.16 broadband wireless access (BWA) system can be configured as point-to-multipoint (PMP) or mesh network topology. In IEEE 802.16e standard, the 802.16a/REVd standards are enhanced to support subscriber station (SS) mobility. The expected offered data rate of IEEE 802.16 BWA systems is up to 70 Mbps per base station over 2 to 50km coverage area.

1.1.5 Fourth Generation (4G) Wireless Networks

The next generation wireless network, referred to as 4G is expected to be the convergence of different wireless access networks. The 4G wireless networks are characterized by the following features: supports a broad range of current and future multimedia services based on quality of service (QoS) requirements, IP-based core networks, global mobility and service portability, extend 3G capacity by an order of magnitude, and integrate wireless access networks such as cellular networks, WLANs, WMANs, and WPANs. With these features, users will have access to different services, increased coverage, the conveniences of single mobile terminal, one bill for all access costs and a reliable wireless access even with the failure or loss of one or more networks. Mobile terminals shall be equipped with multiple physical or software defined interface to allow users to seamlessly switch to different access technologies. The architectures, protocols, services, and wireless technologies that constitute 4G wireless networks are still under consideration and subject to great debate.

1.2 QoS in IP-based wireless networks

Quality of service is the collective effect of service performance parameters which determine the degree of satisfaction of a user of the service [82]. QoS support can occur at packet, transaction, user, and connection levels.

- **Packet-level QoS**: applies to packet-delay bounds, throughput, and error performance. Resource reservation, packet scheduling, and connection-admission control (CAC) as well as a proper medium access control (MAC) protocols are the key components of solutions to these QoS issues. Channel coding and power control schemes are also essential to error performance in wireless networks.

- **Transaction level QoS**: refers to the time it takes to complete a transaction and packet loss rate. Some transactions may have delay constraints, while others cannot tolerate any packet loss.

- **User level QoS**: depends on user mobility and application type. The new location may not support minimum QoS required, even with adaptive applications.

- **Connection-level QoS**: relates to connection establishment and management, which are very important, especially in dealing with user mobility. Connection level QoS parameters include the blocking probability of new calls and dropping probability of hand-offs.

The Internet protocol provides a universal network layer protocol for wireline packet networks and is expected to play the same role in wireless networks. IP provides a globally successful open infrastructure for creating and providing services and applications [73]. An *all*-IP wireless and wireline structure could make wireless networks more robust, scalable and cost effective. It will also enable the abundant applications and software technologies developed for wired IP networks to be used over wireless networks. However, the Internet and IP were originally designed to support best-effort (BE) service. Under such a scheme, packets are treated equally throughout the network. During high traffic loads, the network becomes congested, causing all packet delivery to slow down. If congestions become severe, packets are randomly dropped to ease the congestions. No distinction is made in terms of relative importance of any packet or of the delay requirements of the packet. The best-effort IP network is inadequate for multimedia services that demand QoS guarantees. Solutions for providing QoS in both wireless and wireline networks are required.

For QoS provisioning in IP networks, the Internet Engineering Task Force (IETF) proposed two QoS architectures, namely; integrated service (IntServ) [1] and differentiated service (DiffServ) [2]. The IntServ approach uses Resources Reservation Protocol (RSVP) to explicitly signal and dynamically allocate resources at each intermediate node along the path for each traffic flow. This model enables strong per-flow QoS guarantees, but it suffers from scalability problems. The Diffserv approach is the preferred QoS architecture, because it avoids the scalability and complexity problems of IntServ by dealing with aggregates of flows. Data packets entering a DiffServ edge router are classified, marked, and policed on a per-flow basis. The packets passing through the edge routers are marked as certain flow aggregates, each corresponding to per-hop behaviour (PHB). The PHB defines how packets should be forwarded or handled by the core routers. Each PHB is represented by a 6-bit value called a DiffServ Code Point (DSCP) in the IPv6 header. All packets with the same DSCP are referred to as a behavior aggregate, and they receive the same forwarding treatment. DiffServ classes can be created based on the DSCPs and PHBs. In order for a network user to receive differentiated services from its Internet service provider (ISP), it must have a service level agreement (SLA) with its ISP. SLA specifies the service classes supported and the amount of traffic allowed in each class. DiffServ currently defines two PHBs in addition to best-effort, the Expedited Forwarding [79] and Assured Forwarding [78] from which premium and assured service classes are derived, respectively.

- **Premium service (PrS)**: This service class is intended to provide low-loss, low-latency, low-jitter and guaranteed bandwidth service. PS is suitable for VoIP, videoconferencing, and for creating virtual leased lines for virtual private networks. For this service, the user traffic should not exceed the guaranteed bandwidth specified by the SLA; otherwise excess traffic will be dropped.

- **Assured service (AS)**: This service class is intended to support services that require a minimum bandwidth during network congestions, but not strict in terms of delays. AS packets can be classified as in-profile or out-of-profile depending on their conformance with the SLA. During network congestions the packets are dropped according to a random early detection (RED) with In and Out (RIO) queue management scheme [3]. RIO provides reliable service to in-profile packets, and drops out-profile packets aggressively.

- **Best-effort service**: This service class is intended to support services with no guarantees or QoS as in the current Internet.

In order to provide QoS in wireless domain, the wireless access points should support the DiffServ QoS architecture. The extension of DiffServ to the wireless domain is necessary to provide consistent end-to-end QoS behaviour. However, the characteristics of wireless channels make it difficult to provide QoS guarantees in wireless networks. Therefore, when designing wireless network protocols, these characteristics should be taken into consideration:

- Wireless channels are inherently unreliable and prone to location-dependent, time-varying, and bursty errors due to noise, multipath fading, shadowing and interference.

- The wireless link is a bottleneck of the wireline network in terms of bandwidth. This disparity is expected to hold in future even though rapid progress is being made for high-speed wireless transmission, due mainly to the physical limitation of the wireless media.

- Users tend to move around during a communication session causing hand-offs between adjacent cells. The current trend in cellular networks to reduce cell size in order to accommodate more mobile users in a given area will make it even more difficult to deal with the mobility-related problems [74].

There are general ideas that can be used to deal with problems related to these characteristics. Link layer protocols and MAC layer protocols can be used to combat channel errors and provide efficient channel utilization. Power control can be used to combat the radio wave attenuation; a transmitter can set the power of the transmitted radio wave such that it will be received with acceptable power levels.

1.3 Medium Access Control Protocol

The medium access control resides in the second layer of the OSI protocol stack. In wireless networks the protocol stack usually consists of three layers, as shown in Figure 1.1. The second layer (i.e. the data link layer) consists of two sub-layers: the medium access control (MAC) layer and the link layer. The MAC layer is responsible for invoking a procedure to control, distribute, and co-ordinate the use of the limited channel bandwidth among several users. The link layer is responsible for controlling the link between communicating users. The lowest layer in the model is the physical layer. It provides the physical medium for the information flow which it is responsible for activating, maintaining, and deactivating the physical circuit between the sender and the receiver. Depending on the type of the wireless network, the third layer is generally the network layer, providing transparent transfer of data between two users. IP is an example of a layer 3 protocol. Ensuring end-to-end QoS is a distributed function to be realised at various layers of the network protocol stack.

The MAC protocol is one of the important design issues in the emerging IP-based wireless networks. Due to the heterogeneous nature of multimedia services, the traditional voice/data based MAC protocols do not perform well under multimedia environments [77]. Future generation wireless networks require flexible MAC protocols that can efficiently accommodate multimedia traffic and ensure end-to-end QoS guarantees. The existing MAC protocols proposed for 3G and wireless ATM are capable of handling multimedia media traffic and offer QoS. To deal with various QoS requirements, scheduling algorithms are incorporated in to the MAC protocols. A scheduling algorithm usually allocates data slots according to the traffic characteristics and the current needs of each connection. A complex scheduling algorithm allocates slots precisely, but it needs a lot of computation and related information of the connection. This results in wastage of the limited wireless bandwidth and increasing system complexity. An ideal MAC protocol should provide a simple, fair and efficient mechanism to share the wireless channel.

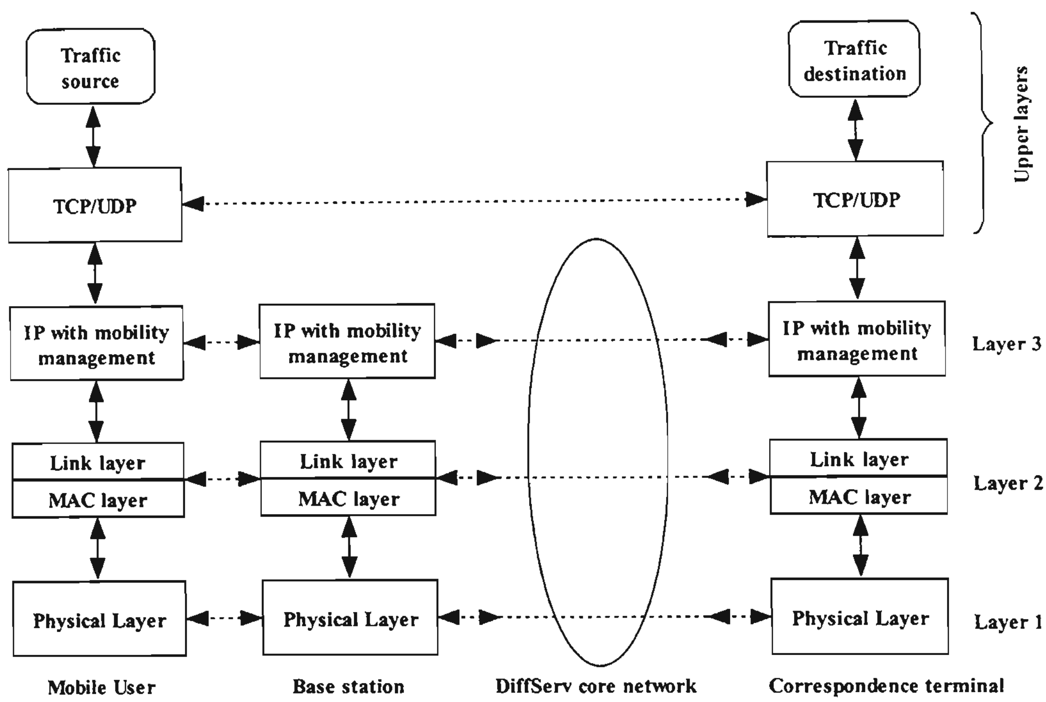


Fig. 1.1: The protocol stack architecture in 4G CDMA cellular networks [76]

MAC protocols can be implemented by employing a variety of multiple access technologies to divide available resources into accessible sections that may be shared amongst mobile terminals (MT). Three well known multiple access methods are frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA). Other multiple access methods are spatial division multiple access (SDMA) and orthogonal frequency division multiple access (OFDMA). In addition to the multiple access technology, the design of a MAC protocol is also influenced by the duplexing techniques as well as the network topology.

1.4 Code Division Multiple Access

CDMA is being used by three of the five air interfaces proposed for 3G cellular systems and is expected to be adopted for 4G wireless systems [75]. CDMA offers several advantages over other multiple access techniques such as high spectral efficiency, soft capacity, soft handoff, and increased system capacity. To support multi-rate transmissions with direct sequence code division multiple access (DS-CDMA) systems, many techniques have been proposed such as multi-modulation CDMA, multi-processing gain CDMA, and multi-code CDMA [74]. Multi-code (MC) CDMA is known to have some advantages over other methods, the multi-modulation CDMA degrades the performance for the users with high data rates, and the multi-processing gain CDMA is expected to cause problems to users with a very high source rate to have a too small processing gain to maintain good cross correlation among different user codes. MC-CDMA is expected to work well with multimedia traffic; when it integrates multimedia traffic, traffic streams with significantly different transmission rates can be easily integrated into a unified architecture with all the transmission channels having the same bandwidth and spread spectrum processing gain.

1.5 Motivation for Research

The 4G wireless networks are faced with many challenges such as interworking between heterogeneous access networks and QoS provision to multimedia services. At the network layer, DiffServ has emerged as an efficient and scalable solution to enable QoS in the current best-effort IP network. This research was motivated by the lack of work done on MAC layer protocols for IP-based wireless networks that employ DiffServ QoS architecture. The current research on DiffServ mainly focuses on the wireline network. Only limited work has been done on DiffServ over IP-based wireless networks, and most of the work focuses only on the transport and network layers, without consideration for the utilization of valuable resources in the lower layers (i.e. data link and physical). The focus of this dissertation is on the study of MAC protocols and the derivation and evaluation of a MAC protocol for IP-based CDMA wireless network supporting DiffServ IP traffic.

1.6 Outline of the Dissertation

The dissertation is organized into six chapters. Chapter 1, gives an overview of existing wireless networks, and introduces the concepts of QoS, MAC protocols and CDMA, as applied in wireless networks. The focus and the motivation for this dissertation research were also presented.

Chapter 2 is a literature survey on existing MAC protocols design for centralised wireless networks such as 2G, 3G, and wireless ATM. The chapter starts with a general overview in which MAC protocols are classified and their requirements when carrying multimedia traffic are presented. Thereafter, a discussion of some of the existing MAC protocols under each class of MAC protocols is presented starting with random access protocols, followed by guaranteed assignment protocols, after which follows random reservation protocols and ending with demand assignment protocols.

In chapter 3, we present our MAC protocol integrating DiffServ QoS architecture and CDMA for wireless IP networks, called Wireless DiffServ IP over CDMA (WDIP/CDMA). The proposed MAC protocol is based on a random reservation protocol with channel load sensing scheme, and demand assignment schemes for supporting BE service, AS and PrS respectively. The traffic models corresponding to each service type and the operation of the systems are described. The simulation model is then presented, along with the environment for the proposed MAC scheme. The simulation, which is time driven was developed using the C++ Builder 4 software package. The performance of the scheme is then presented and evaluated.

In chapter 4, Markov analyses of AS and PrS traffic which solve for the system state distributions, are performed for the WDIP/CDMA protocol. The BE traffic is also analysed by a Markov analysis assuming an infinite population model with Poisson arrival process. The analytical results are presented and are verified by comparing them to the simulation results.

In chapter 5, the MAC protocol and physical layer of the IEEE 802.16 standard for WMANs are discussed in detail. The shortcomings of the MAC protocols for the IEEE 802.16 based broadband wireless access networks are highlighted. To address these issues, techniques used in our WDIP/CDMA protocol are proposed to improve performance of the protocol. The performance of the modified MAC protocol is evaluated through simulations.

Chapter 6 concludes this dissertation with a summary of the research findings, including important concepts and techniques behind this research effort.

1.7 Original Contribution in this Dissertation

The original contributions of this dissertation include:

1. The derivation of a novel MAC Protocol, called WDIP/CDMA, for DiffServ over CDMA IP-based wireless networks.

2. Derivation of analytical methods to predict the performance of the proposed MAC protocol.

3. Modification of the MAC protocol for IEEE 802.16 broadband wireless access networks

Parts of the work presented in this dissertation have been presented by the author at the following local conferences:

1. S. B. Mahlaba and F. Takawira, "Performance analysis of a MAC protocol for wireless IP/CDMA networks", proceedings of SATNAC, Spier, September 2004.

2. S. B. Mahlaba and F. Takawira, "A MAC Protocol for Wireless IP Networks", proceedings of SATNAC, George, September 2003.

Chapter 2

Medium Access Control (MAC) Protocols for Wireless Networks

2.1 Introduction

This chapter is a literature survey of selected MAC protocols employed in wireless access networks. Since 1970 MAC protocols have attracted a lot of researchers' attention. They define the strategy by which a user gets access to the network. The ultimate goal is to provide a reliable and dynamic schedule by which a common channel is distributed among network users.

In Section 2.2, different categories of MAC protocols are presented, sections 2.3 to 2.6 discuss MAC protocols for centralized wireless networks falling under some of the categories in section 2.2.

2.2 General overview of MAC protocols

Wireless MAC protocols can be broadly classified into two categories according to the type of network architecture for which they are designed, these categories are: distributed and centralised [34]. The extension of wireline to wireless domain requires a centralised network architecture with a BS acting as the interface. Since this dissertation focus on extending IP to wireless domain, we focus on the centralised MAC protocols. Centralised MAC protocols can be classified according to their multiple access technology and as well as their method of resource sharing [4]. The multiple access scheme of a MAC protocol establishes a method of dividing resources into accessible sections. Multiple access schemes include FDMA, TDMA, and CDMA, as well as hybrids of such techniques. The resource sharing methods include:

1. **Random Access**: In random access schemes, users contend for access to the medium by transmitting as soon as packets are available to send. When two or more users decide to transmit at the same time, a collision occurs. Collisions prohibit the successful separation of an individual user's signal and they are a degrading factor in random access protocol performance. Random access is suitable for bursty traffic but is not desirable for delay sensitive traffic.

2. **Guaranteed Access**: In guaranteed access schemes, collisions are eliminated by using a scheduling scheme such as polling. The base station (BS) polls each user in the network in a round robin fashion and the user sends data in response to the poll. These schemes can achieve high channel utilization when many users are active with data to transmit. However, when the user being polled has nothing to transmit, the bandwidth can be wasted.

3. **Fixed Assignment**: In dedicated assignment schemes, each user is assigned a predetermined and fixed allocation of resources, regardless of the user's need to transmit. These schemes offer desirable features such as fixed delay and guarantee that once transmission is in progress, no collisions can occur. They are appropriate for continuous traffic, but are wasteful for bursty traffic. Fixed assignment MAC protocols can be classified into FDMA, TDMA, and CDMA.

4. **Hybrid Access**: Hybrid protocols bridge the gap between statistical access with random access protocols and deterministic access in the guaranteed protocols by blending the best qualities of both types of the protocols. Based on the scheduling and reservation policies imposed by the BS, these protocols can be further classified into two classes: random reservation access (RRA) protocols and demand assignment (DA) protocols. In RRA protocols, the BS has implicit rules of reserving upstream bandwidth. For example, a success in contention results in periodic reservation of an uplink slot. Demand assignment scheme assign resources according to mobile terminal requests. They are suitable for variable-rate traffic and the hybrid conditions of multimedia traffic.

The typical requirements of MAC protocols carrying multimedia traffic are that they must [4]:

- Provide simultaneous support for a wide variety of traffic types (e.g. e-mail, rt-video)
- Support traffic that requires delay and jitter bounds
- Assign bandwidth resources in an efficient manner (between different classes, on demand)
- Support both fair and prioritised access to resources

A MAC protocol can employ either time division duplex (TDD) or frequency division duplex (FDD) to separate uplink and downlink traffic. TDD refers to multiplexing in transmission and reception in different time periods in the same frequency bands. FDD uses different frequency bands for uplink and downlink transmission. A majority of current generation wireless MAC protocols use FDD as it allows easy analyses of uplink and downlink separately. However, TDD has advantage over FDD when there is asymmetric traffic between the uplink and downlink. TDD offers greater flexibility than FDD since the number of timeslots dedicated to up-link/downlink transmission can vary as a function of the service demand. The disadvantage of TDD is the time delay caused by switching between uplink and downlink and variety of interference conditions (e.g. inter-cell interference). Many current researchers are focusing on reducing the interference related to TDD [80].

2.3 Random Access Protocols

For these types of protocols, no regulations or guarantees are provided in the network, so mobile terminals can access the channel freely and randomly. They provide high efficiency while maintaining simplicity. The major disadvantage of random access protocols is packet collision, where a collision occurs when two or more mobile terminals choose the same slot to transmit their data.

2.3.1 Narrow-band Random Access Protocols

2.3.1.1 ALOHA

ALOHA is a fundamental protocol for most modern channel contention protocols. It was developed at the University of Hawaii in 1970 for packets radio networks [46]. ALOHA involves a single channel being shared by a population of terminals in a totally asynchronous and uncoordinated fashion. When a user has generated a new packet, it is transmitted immediately. After transmission, the user waits for an acknowledgement from the receiver. If no acknowledgement is received within a predefined period, the transmitted packet is assumed to have collided or been lost. The user then enters a collision resolution state, in which it waits for a random time before retransmitting the packet. The ALOHA access protocol provides completely free access to the radio channel, and can be used on any network topology. The maximum throughput for this protocol is 18 percent.

2.3.1.2 Slotted ALOHA

The slotted-ALOHA (S-ALOHA) protocol [48] is an improvement on the ALOHA protocol. In the S-ALOHA protocol, time is divided into fixed length slots that are equal to the size of one packet. Users are synchronized to transmit packets at the beginning of a time slot. This approach halves the vulnerability of the transmission to collision consequently doubling the system efficiency. S-ALOHA is similar to ALOHA in that the user waits for an acknowledgement from the receiver. If the acknowledgement is not received after a predefined period, then the user assumes that a collision has occurred. The user then backs off for a random number of slots before retransmitting the packet. The ALOHA protocol is widely used as the random access method for the more complex protocols.

2.3.1.3 Collision Resolution Algorithms (CRA)

The CRA define rules followed by terminals to resolve collisions. To avoid retransmitted packets from causing collisions again, users wait a random time before retransmitting. Random waiting time ensures that packet retransmission is started at different times. The random waiting time is commonly referred to as the backoff time and the terminals that are waiting to retransmit are commonly referred to as being backlogged. Some examples of common CRA schemes are:

1. **P-persistence and Non-persistence:** The CRA used in ALOHA and S-ALOHA are based on non-persistence and p-persistence respectively. In non-persistence, the backoff time is explicitly chosen in a random fashion according to some distribution function, the most common being the uniform and truncated negative exponential distributions [5]. In p-persistence, instead of explicitly choosing a backoff time period, the terminal attempts to retransmit in the next slot with probability p, or delay the retransmission until the following slot with probability $1-p$. This process is repeated until the packet is successfully transmitted. The parameter p should be chosen to be small so as to ensure that the number of retransmitted packets that a transmitted reference packet finds across the channel from one collision to the next is virtually independent [5].

2. **Binary Exponential Backoff**: The binary exponential backoff (BEB) is one of the most commonly used collision resolution algorithms. The protocol is easy to implement, does not require many hardware resources and can work on top of the slotted-ALOHA protocol [42]. BEB increases the backoff time of collided terminals to relieve the traffic. When a transmission from a user collides, the user backs off. If a collision occurs again for the same packet, the backoff time is doubled. The backoff time of the user doubles each time the same packet collides.

3. **Tree Algorithm**: Tree algorithms can usually provide highly efficient collision resolution in a time slotted channel and much research has been conducted in this particular area. The IEEE 802.14 standard for hybrid fibre coaxial cable TV networks (HFC-CATV) uses a highly optimized ternary tree algorithm for collision resolution [42]. The multi-accessing tree protocol was one of the first tree algorithms proposed by

Capetanakis [70] in the 1970s. It has been shown that Capetanikis' tree algorithm has an improved maximum performance of 0.43 and is stable for all input rates less than 0.43, as opposed to 0.37 for S-ALOHA [5]. This scheme involves subdividing the collided users into two groups (sub-branches) and resolving the collisions within each branch. Successive collisions create new branches, and the process repeats until all collisions have been resolved and there is no chance of them occurring again.

2.3.1.4 Carrier Sense Multiple Access (CSMA)

The CSMA protocol [49] attempts to minimise collisions by co-ordinating channel access through a channel sensing scheme. A user that has a packet to transmit senses the channel for a set duration before transmitting. If the channel is sensed to be busy, the sensing user refrains from transmitting and reschedules according to one of several strategies. If the channel is sensed idle, the user transmits its packet immediately. Collision of packets only occurs if two or more packets were transmitted within the propagation time of one another. Since the CSMA protocol cannot totally avoid collisions because of propagation delays, a Collision Detection (CD) scheme was incorporated into CSMA. In the resulting protocol, called CSMA/CD [50], a user monitors the channel to see if what is on the channel agrees with the packet being transmitted. If not, a collision is assumed and the user aborts transmission. The user may send a noisy impulse to inform all users, which in turn will prevent others from transmitting. Collision detection improves system performance by preventing a user from wasting time in transmitting the remainder of a corrupted packet. There are variations of CSMA and CSMA/CD basic strategy, as outlined below:

1. **Non-persistent CSMA**: If the channel is idle, send. If the channel is busy, wait a random time and try again.

2. **P-persistent CSMA**: P-persistent CSMA is used for slotted channels. If the channel is idle, send with probability p and defer until the next slot with probability $1-p$ if the channel is busy. This is repeated until the packet is successfully transmitted, or until another user is sensed to have begun transmitting, in which case wait a random time and try again.

3. **1-persistent CSMA**: 1-persistent CSMA is a special case of p-persistent CSMA in which $p=1$. A ready user senses the channel, if the channel is idle the user transmits the packet. If the channel is busy the user waits until the channel becomes idle and then transmits the packet with probability one.

2.3.1.5 Inhibit Sense Multiple Access (ISMA)

In wireless networks, carrier sensing is very difficult due to the hidden terminal, exposed terminal and capture effects [34] and such problems are performance degrading factors in CSMA protocol. Geographical obstruction can cause some terminals to be unaware of the transmitting state of other terminals in the network. Consider two transmitting terminals, A and B, which may be within range or line of sight of a receiving terminal C or BS, but be out of range of each other or have no direct line of sight of one another. In this case, neither A nor B will be able to detect whether the other is using the channel and may attempt to simultaneously transmit to C, resulting in conflict at C's receiver. In ISMA [52], [53] these problems are avoided by having a central BS that is within range and line of sight of all terminals in the network that performs carrier sensing and collision detection. The BS uses a separate out of band signalling channel to inform users of the state of the shared inbound channel by sending a busy or idle signal (tone). There are several versions of the ISMA protocol found in literature, such as Idle-signal Casting Multiple Access (ICMA) [54], ISMA [51] and Busy-tone Multiple Access (BTMA) [55]. In most distributed wireless networks, collision avoidance schemes such as those in [63], [64] and [65] are used, since ISMA is not always feasible in such topology.

2.3.2 CDMA Random Access Protocols

2.3.2.1 Spread-Spectrum ALOHA (SS/ALOHA)

In spread spectrum ALOHA (SS/ALOHA), users follow the same access algorithms and CRAs as in ALOHA. In pure ALOHA systems it is assumed that when a single packet is transmitted in a given time period it is always received correctly. If there are two or more transmissions then all the packets are destroyed. The difference is that a spread-spectrum system permits the simultaneous transmission (with some error rate) of two or more user's data. The number of transmissions is limited by the multiple access interference (MAI), which does not affect all users equally. Some of the transmissions may be successful, while others may not. The packet error probability, which is almost zero when few users are transmitting, approaches one as the number of simultaneous transmissions tends to infinity. A detailed survey and brief performance analysis of SS/ALOHA is found in [5].

2.3.2.2 Channel Load Sensing (CLS)

In CDMA based random access systems channel sensing is performed by measuring the number of transmitting users (or channel load) in the channel. Based on this value, steps can be made to control channel access such that multiple access interference (MAI) remains within acceptable levels. Measurement of channel load in CDMA is complicated, compared to narrowband CSMA where the sensing mechanism can have only two states (idle or busy). There are two main techniques for measuring the number of transmissions in a channel [5]:

- **Estimation of channel load using power levels**: In this scheme the total power received from all terminals is measured and its value divided by the expected power received from each user. This scheme requires all users to be received with near equal power. Therefore, it is only valid for centralised CDMA systems that implement near-perfect power control. However power level estimations are generally not required in such systems, because the BS is usually capable of estimating the number of transmitting users from the fact that it takes part in all communications and will know how many users it is dealing with. In distributed systems this form of channel load estimation is extremely inaccurate because of non-equal power levels from different terminals due to different distances between terminals and that each user is responsible for measuring the channel load itself, consequently each user will arrive at a different estimation for the channel load.

- **Measurement of channel load through multiple receivers**: A more accurate channel load measurement can be obtained by using multiple receivers at each terminal to despread all signals in the channel. Each receiver is tuned to one of the PN codes in the code population. In distributed systems, fully connected schemes where power control is impossible to implement, this technique provides good results. The main disadvantage is that each terminal needs to have at least as many receivers as the number of terminals in the population. This disadvantage is eliminated in centralised systems, since only the BS needs to implements multiple receivers.

Intuitively, channel load sensing schemes are most useful in asynchronous systems and slotted systems in which packets are transmitted over multiple slots [5]. In slotted-SS/ALOHA where packet length is equal to the slot size, the channel load is completely independent from slot to slot. A packet that enters the channel in slot t will not be in the system in slot $t+1$. It is not effective to measure the channel load in slot t and use this value to make admission decisions in slot $t+1$. In contrast to this intuition, a channel load sensing scheme where the packet length is equal to slot duration is studied in [45] for frequency-hopping slotted random access protocol. The value of the channel load can be used to control channel access and usage in two ways as follows [5]:

- **Channel load regulation through blocking**: This scheme attempts to avoid excessive MAI by regulating the channel load around a safe operating level. The concept is analogous to the blocking mechanism used in CSMA. If the channel load is less than a certain threshold then access is allowed, otherwise access is denied until the ongoing transmission falls below the threshold. Packets that are denied access follow a rescheduling algorithm in the same way as CSMA.

- **Overload Detection**: Overload detection is analogous to collision detection in CSMA/CD. Overloads in the CDMA channel may be sensed and action taken to abort packets. The assumption can be made that once the channel load exceeds a certain threshold, all packets involved have high a probability of being corrupted. Aborting these packets may then result in improved system performance.

In [6], Judge and Takawira studied the improvement in system performance gained by implementing both overload detection and blocking scheme. The model in [6] was found to give superior performance over protocols which employ collision detection only or blocking only (e.g., [44]). The effect of packet length adaptation is studied in [43] as the way to improve performance of CLS/CDMA systems in highly correlated fading channels. It is shown that packet-length adaptation can provide a significant improvement of the system performance with efficient utilization of limited radio and battery resources.

2.4 Guaranteed Assignment Protocols

As already mentioned, guaranteed assignment protocols are based on polling algorithms. Polling is defined as a control handshake that is similar to the handshake used in the collision avoidance ad hoc network, and is initiated by the BS using a small packet that carries a message to a specific MT. Once the MT receives this packet, it responds to the BS according to the protocols it uses. The BS polls each MT in the network for data, one after the other, in a round robin fashion. For these types of protocols, there exists no collision hence they can achieve high utilization when many MTs are accessing the channel. However, bandwidth can be wasted when the polled MT has nothing to transmit.

2.4.1 Zhang's Protocol

In this protocol the BS polls each MT in the network for transmission request in a round robin fashion [56]. The protocol operation is illustrated in Fig 2.1. If the polled terminal has outstanding packets to transmit, it responds with a transmission request packet. If its queue is empty, it transmits a *KEEP ALIVE* message to inform the BS that it is still in the network. This poll-request handshake ensures a good communication channel between the BS and the MT. After polling all the terminals in the network, the BS starts polling the terminals that had responded with the request packet for data. Zhang proposed that all terminals be polled once every time T, where T is the coherence time of the channel, the logic being that after T seconds the channel is likely to have changed sufficiently to affect the data transmission and hence channel needs to be re-sampled. The protocol is very simple, but it guarantees that all MTs are polled and all transmissions are free of collision.

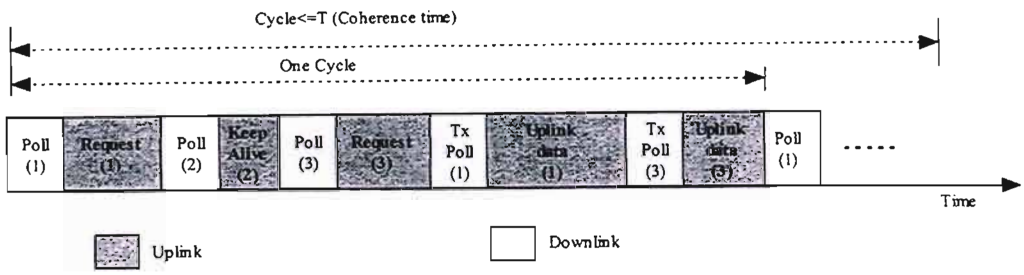


Fig. 2.1: Zhang's proposal [34]

2.3.2 Disposable Token MAC Protocol (DTMP)

The DTMP [57] modifies Zhang's cyclic format from a poll-request-poll-data, to just a poll-data cycle. The operation of the protocol is shown in Fig. 2.2. In DTMP, the data transmissions for both the uplink and the downlink are followed by a single poll. There are two types of poll in the protocol: the normal poll and the data poll. If the BS has no data for the polled MT, it sends a normal poll and if the polled MT has no data to transmit, it remains silent. After sending the normal poll, the BS waits for the data from the polled MT for a short period. If no data is transmitted from the polled terminal in that period, the BS polls the next MT. However, if the polled MT has data to transmit, it transmits its data immediately after being polled and the BS replies with an acknowledgement (ACK) after the data packet has been successfully received.

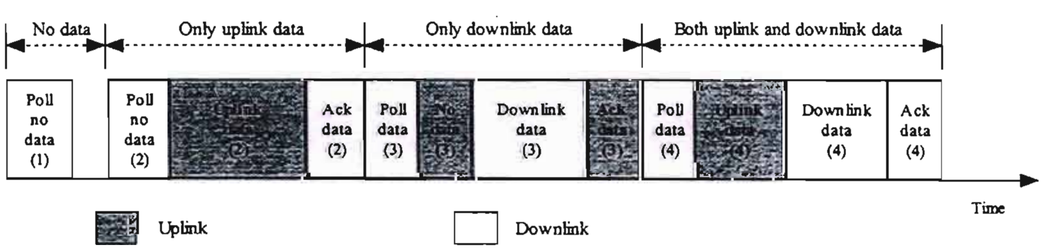


Fig. 2.2: Disposable token MAC protocol format [34]

A data poll is used when the BS has data for the polled MT. If the polled MT does not have any data for the BS, it transmits a "no data" message to the BS upon being polled. If the polled MT has data to send, it sends its data when it is polled. After receiving the "no data" message or the uplink data, the BS transmits its data to the polled MT. If the downlink data transmission is successful, the polled MT sends an ACK.

2.4.3 Acampora's Proposal

This protocol was proposed by Acampora *et al* [41] for smart antenna systems. The protocol operation is divided into three phases: polling phase, request phase and data phase. The frame structure of the protocol is shown in Fig. 2.3. Each MT in the network is assigned a unique codeword. In the polling phase, the BS identifies all active MTs by polling each of them with its codeword. The MT remains silent if it has no packet to send. An active terminal with data to transmit echoes the codeword back. Upon receiving the echo, the BS then broadcasts the codeword back to the network so that every terminal knows the number and order of the active terminals. In the request phase all the active terminals send their requests in order to the BS. The BS will then poll the terminals that have sent requests in the data phase.

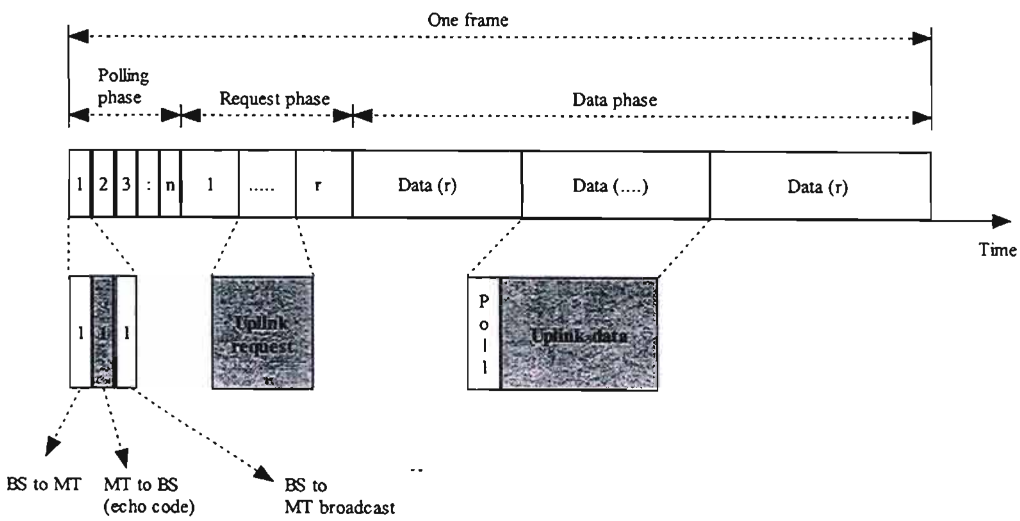


Fig. 2.3: Acampora's proposal frame structure [34]

2.5 Random Reservation (RRA) Protocols

The random reservation protocols attempt to combine the flexibility of random access with the guarantees of polling access. At the same time, the protocols must remain simple. Every RRA protocol consists of two components: random access and reservation. All MTs that have data to transmit use a random access protocol to make their first transmission. S-ALOHA is the most widely used random access protocol because it does not need state information. When an MT manages to successfully transmit the first packet, it follows a reservation policy (described by the BS) to reserve the uplink channel.

2.5.1 TDMA based RRA Protocols

2.5.1.1 Packet Reservation Multiple Access (PRMA)

PRMA [47] was designed for integrating voice and data on short range radio channels. The PRMA protocol is organized around time frames with duration matched to the periodic rate of voice traffic. Each frame is divided into time slots which are recognized as "reserved" or "available" according to feedback messages from the BS. Terminals with messages to transmit contend for available slots using the S-ALOHA protocol. Upon successfully transmitting in a slot, a terminal with voice traffic or long periodic message reserves that slot for uncontested use in subsequent frames until completion. If not successful, the terminal retransmits the first packet with probability q in subsequent unreserved slots. Data terminals are not allowed to reserve slots in the frame and they have lower priority than voice in terms of transmission permission probability. For example in Fig 2.4, a successful voice transmission occurs in slot 4 of frame K and this slot is reserved in the next frame. Similarly, in slot 6 a data terminal contends successfully but no reservation is made in frame $K+1$.

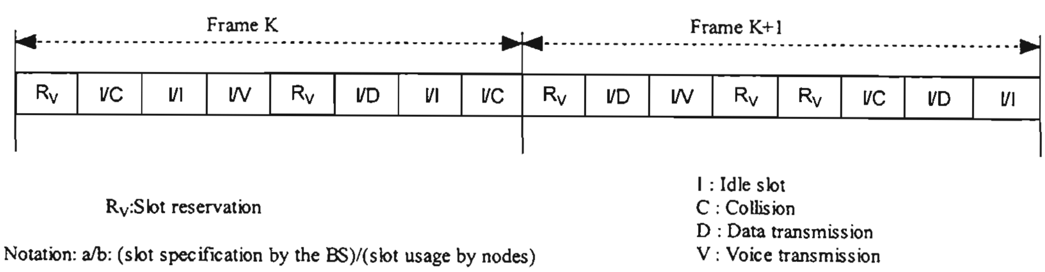


Fig. 2.4: Uplink frame structure of PRMA and its operation [34]

In [39], the classic PRMA is extended to accommodate rt-VBR video in addition to voice and data traffic. Each frame consists of 150 time slots and it is assumed that one video user and large number of voice and data users contends for these 150 slots. A fixed number of slots M out of 150 are reserved at the beginning of every frame to transmit some of the video packets arriving during the frame interval. The rest of the video packets contend with voice and data packets for the remaining time slots as in the normal PRMA. Several other protocols have been proposed for enhancing the performance of PRMA [69]; some of them are Aggressive (APRMA) [67], Data Steal over Voice (PRMA-DSV) [66], Mini-packets (MPRMA) [66], Centralised (C-PRMA), and PRMA/Dynamic Allocation (PRMA/DA) [68]. All of these enhancement PRMA protocols deal with improving channel efficiency and providing some kind of fairness for data applications. A comprehensive analysis of PRMA can be found in [47], [58] and [72].

2.5.2 CDMA based RRA Protocols

2.5.2.1 Multidimensional PRMA with Prioritised Bayesian Broadcast (MD PRMA BB)

MD PRMA BB is a proposed MAC strategy for the uplink channel of the UMTS terrestrial radio access (UTRA) [9]. This protocol is the derivative of the classic PRMA adapted for hybrid TDMA/CDMA schemes and multimedia traffic. In conventional PRMA, time slots are grouped into frames and resources are allocated based on packet spurts. In MD PRMA BB, slots are further defined in an additional dimension such as code or frequency, and each time slot supports up to eight spreading codes (sub-slots). As shown in Fig. 2.5, the frame consists of contention slots (C slots) and information slots (I slots) as indicated by the BS.

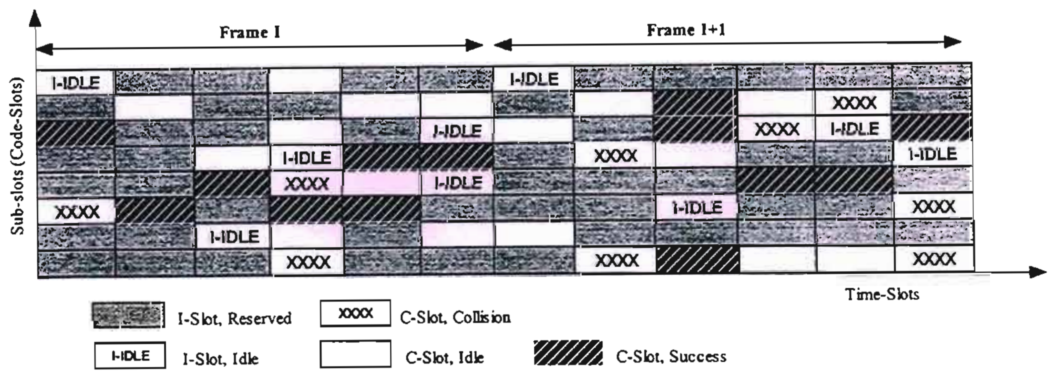


Fig. 2.5: MD PRMA frame structure [4]

When a packet spurt arrives at the mobile terminal (MT), it switches from an idle state to the contention state and tries to obtain a permission to send a packet on the next available C slot by performing a Bernoulli experiment with probability p. If successful, the MT transmits the first packets of its spurts in the contention slot. If the first packet of the spurt is successfully transmitted, the BS responds with an ACK which implies the reservation of the same slot. The C slot becomes an I slot in the subsequent frame and is reserved for that MT for the duration of the spurt in the case of voice or for certain number of frames in the case of data. If the Bernoulli experiment fails or the packet collided, the contention process is repeated in the slot-by-slot basis. It is assumed that ACKs will be received before the next transmission slot in the frame.

The transmission probabilities, p, of terminals in a contention slot depend on the type of service, available channels and the estimated number of backlogged terminals currently in the system. They are calculated according to protocol stability and best delay-throughput performance. The Bayesian broadcast method originally used to stabilise S-ALOHA networks is used to calculate transmission probabilities. These probabilities are calculated on a slot-by-slot basis. Prioritisation is used in MD PRMA BB to calculate the probabilities for contention slots, however to discriminate between the QoS of different service classes multiple transmission probabilities per slot are used, all derived from the original probability.

One of main drawbacks of this protocol is that it is not able to support high-bit-rate data service or real time services such as video, which requires multiple sub-slot allocations in the same frame. Another problem of MD PRMA BB is that it allows services with different BER requirements to share the same time slot. Thus the capacity of the slots will be variable and limited by the most demanding service.

2.5.2.2 Voice and Data integration in Dual-Threshold CDMA (DT/CDMA) MAC protocol

DT/CDMA is a MAC protocol proposed by Judge [5] for voice and data integration in CDMA packet radio networks. This protocol is based on a combined ISMA/CDMA [53] protocol that works in conjunction with two multiple access thresholds. The multiple access thresholds are used to regulate channel access such that the MAI capacity of the CDMA system is not exceeded. The system is time slotted and both voice and data traffic are assumed to use the same bit rate and spreading techniques. Two multiple access thresholds, $K^v_{\max}$ and $K^d_{\max}$, are defined as the maximum number of allowed simultaneous transmitting voice and data users, respectively, such that the expected packet error probabilities ($P^{pkt}_{Err}(k)$) are below pre-specified values,

$$\begin{aligned} P^{pkt}_{Err}(k) &\le P^v_{\max} \qquad && for\ k \le K^v_{\max} \\ P^{pkt}_{Err}(k) &\le P^d_{\max} \qquad && for\ k \le K^d_{\max} \end{aligned} \tag{2.1}$$

where $P^v_{\max}$ and $P^d_{\max}$ are the maximum acceptable packet error probabilities for voice and data respectively, and k is the number of transmitting voice and data users. Voice traffic admission is given unconditional priority over data traffic. This results in data traffic utilising the remaining capacity. The threshold value $K^d_{\max}$ for data is effectively the maximum allowable data users when the level of voice v is zero. If $v > 0$ voice users are transmitting, then the maximum number of allowed data users is give by

$$\beta = \max(0, K^d_{\max} - v) \tag{2.2}$$

Since v is a time varying process, β is expected to be time varying as well. To reduce the effects of collision due to data traffic in the system a second threshold, defined as α, which is less than β is implemented. With respect to these three thresholds, three signalling tones broadcasted by the BS are defined and referred to as follows:

- Voice Collision (**VC**) signal (tone); emitted by the BS if the number of voice users exceeds $K^{v}_{\max}$.
- Data Collision (**DC**) signal (tone); emitted by the BS if the number of data users exceeds β.
- Data Blocking (**DB**) signal (tone); emitted by the BS if the number of data users exceeds α.

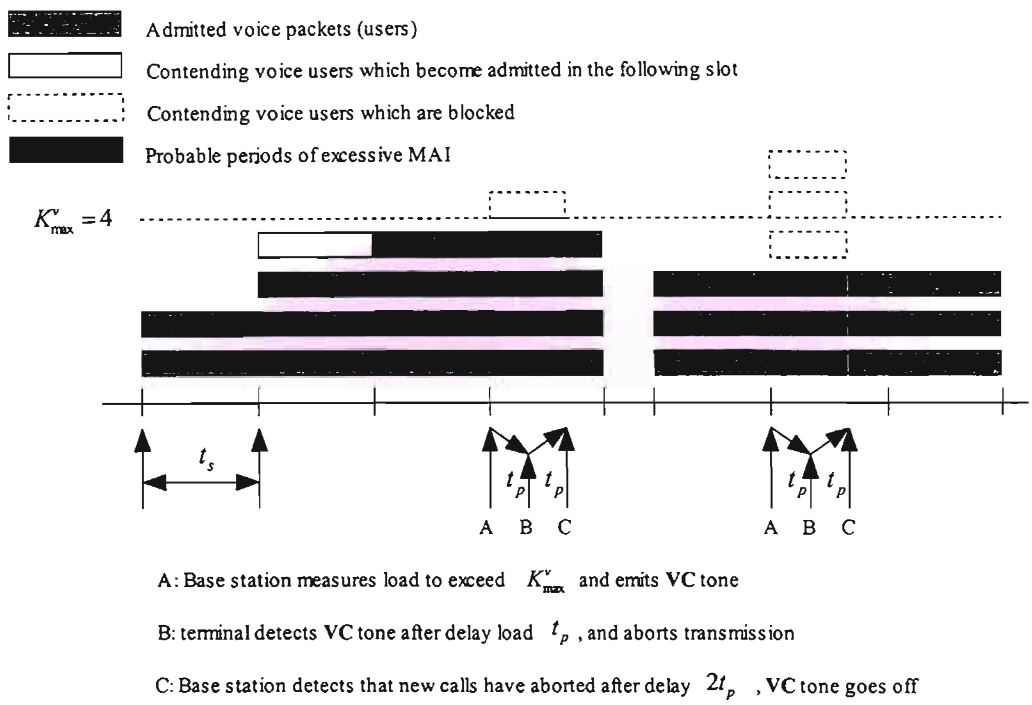


Fig. 2.6: Illustration of the voice call admission policy [5]

The MAC protocol that each terminal (voice or data) in the network adheres to is based on these thresholds and their associated signalling tones. Fig. 2.6 illustrates the admission procedure for voice traffic.

All new voice calls are accepted if the number of voice calls, v, in progress does not exceed $K^{v}_{\max}$. As soon as v exceeds $K^{v}_{\max}$, the BS emits the **VC** tone to block contending users. Upon detecting the **VC** tone, all contending voice users abort transmission and are considered blocked. If the **VC** is not detected, the contending voice users become admitted in the following slot. Once the voice user has been admitted, it effectively secures that channel for the duration of its call and is never dropped. Voice packets that are corrupted due to MAI are not retransmitted and are discarded.

The admission procedure for data traffic is illustrated in Fig. 2.7. Data terminals are admitted to the system if no **DB** and **DC** tones were sensed before and after (re)transmission of the first packet of the data message. If the **DB** tone is present, the terminals defer transmission (the messages are *blocked*). If the **DC** tone is detected, contending terminals abort transmission. Once users become admitted, they secure the channel for the duration of the message. Packets corrupted due to MAI are grouped together to form a new message. An ARQ feedback scheme is implemented for retransmission purposes. Aborted, blocked or corrupted message are retransmitted after a random time period, which is assumed to be geometrically distributed.

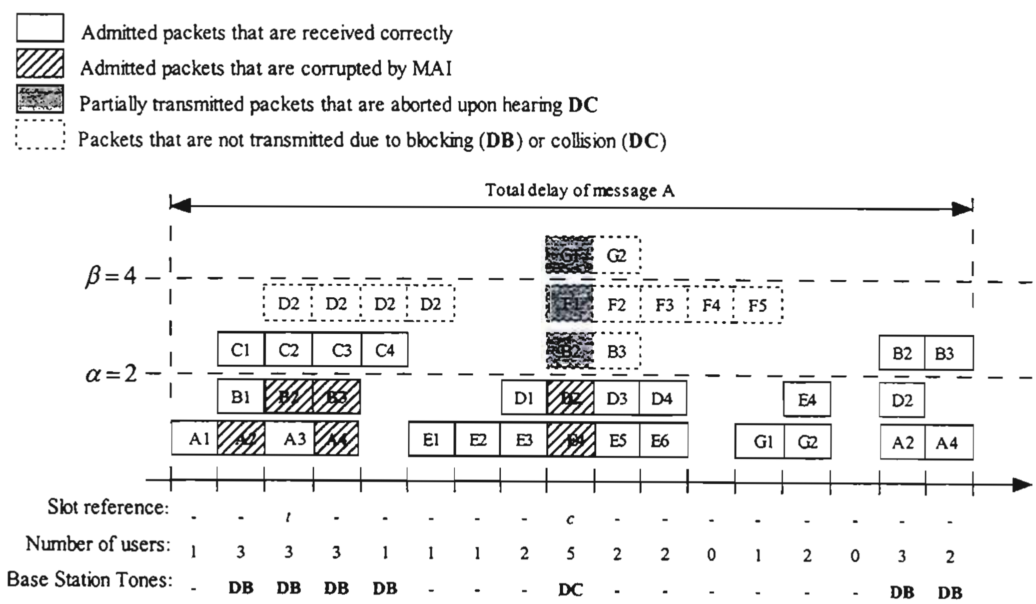


Fig. 2.7: Illustration of data admission policy [5]

2.6 Demand Assignment (DA) Protocols

DA protocols combine both random access and guaranteed access protocols, and allocate bandwidth to nodes according to their quality of service (QoS) requirement. Most of these protocols are designed for networks such as wireless ATM networks and wireless voice/data communication networks, where they are required to provide QoS guaranteed to multimedia traffic.

There are three phases to DA protocols: request, scheduling, and data transmission. In the request phase, a random access scheme is normally used by the MTs to send their QoS requirements to the BS. Once the requests have been gathered, the BS deploys a scheduling protocol to assign data slots to the MTs that have requested it. The notification of the assignment of the uplink data slots is broadcasted through the downlink stream, and the data for the MT from the BS is usually transmitted after notification. The last phase of the protocol is the data transmission phase, where terminals transmit their data without any collision in the data slots assigned by the BS. The disadvantage of DA protocols is the additional overhead and delay caused by the reservation process which can degrade performance. The scheduling algorithms incorporated to these protocols need a lot of computations and related information of the connection. This results in wastage of the limited wireless bandwidth and increases the system complexity.

2.6.1 TDMA based DA Protocols

2.6.1.1 Distributed-Queuing Request Update Multiple Access (DQRUMA)

In DQRUMA [40], Karol *et al* considered a time-slotted system in which a request access (RA) and packet transmission (Xmt) channels are formed on a slot-by-slot basis. The uplink and downlink transmission are organised according to FDD. The uplink consists of the RA and Xmt channels. The RA channel is used to send transmission requests (Xmt_Req) from the MTs to the BS. The transmission request includes mobile's *b*-bits *access ID* which it was assigned at call setup or after handoff. The downlink consists of an ACK, Transmission-Permission (Xmt_Perm) and packet transmission channels. The ACK channel is used by the BS for acknowledging the reception of MT's transmission requests. The Xmt_Perm channel carries a transmission grant for the MT that is allowed to use the next uplink slot. Fig. 2.8 shows the basic frame structure of the DQRUMA protocol.

The channel model considers the MTs to be in one of three states: empty, request and wait-to - transmit. When a MT has a packet to transmit, it sends an Xmt_Req to the BS on the RA channel using a random access protocol. The author considered two methods to randomly access the RA channel, these are: Dynamic Access Channel Slotted ALOHA with Harmonic Backoff algorithm and Dynamic Access Channel Binary Stack algorithm. Several results for normalised delay versus. throughput were given for both cases, without a final choice being made as to which would be implemented. After successfully receiving the Xmt_Req packet, the BS updates the corresponding entry in a request table, which has entry for each MT in the cell and sends an ACK to the relevant MT.

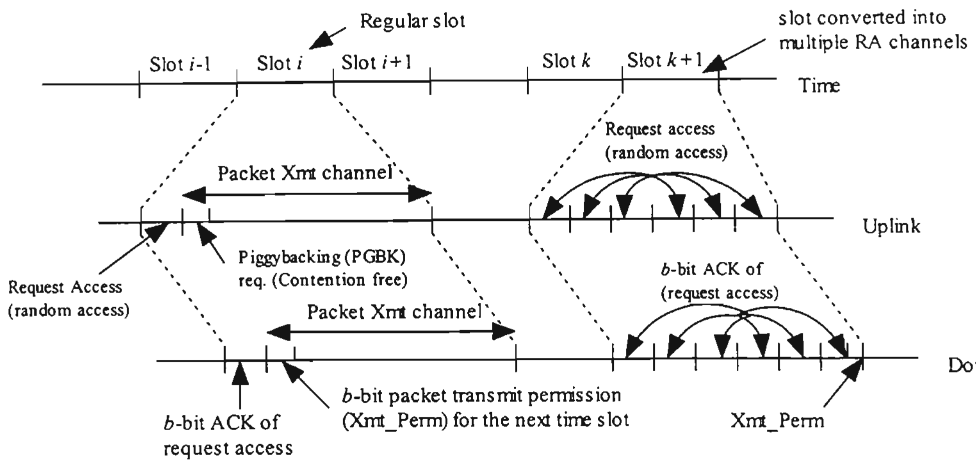


Fig. 2.8: DQRUMA timing diagram [37]

After the reception of the positive ACK to its Xmt_Req, the MT switches to wait-to-transmit state and keeps listening to the downlink Xmt_Perm channel, waiting for permission to transmit from the BS. The distributed queuing aspect of the protocol is that the MTs queue their packets for transmission and are served in a round robin fashion. When the BS has determined that there is sufficient capacity to accept data, it sends a transmission permission to the relevant MT through the downlink Xmt_Perm channel. After detecting its own *b*-bit ID, the MT will then transmit a packet in the next slot and switch to the empty state or wait-to-transmit (if it has more packets). If the BS becomes aware that there are many MTs in the cell making access contentions, it may convert an entire slot into a series of RA channels, and the corresponding downlink slot a series of ACK channels. Each time a MT transmits an ATM packet it includes a piggyback message if it has more packets to transmit.

2.6.1.2 Dynamic TDMA with Time-Division Duplex Protocol (DTDMA/TDD)

DTDMA/TDD was proposed by Raychaudhuri *et al.* for a prototype wireless ATM network (WATMnet) capable of providing integrated multimedia service to mobile terminals [62]. The protocol is based on TDMA/TDD with fixed-length frame. The downlink subframe consists of a single time division multiplex (TDM) burst which is divided into two parts. The first part contains bandwidth reservation (B-R) control and B-R feedback (ACK), while the second part is used for data transmission from the BS to MTs.

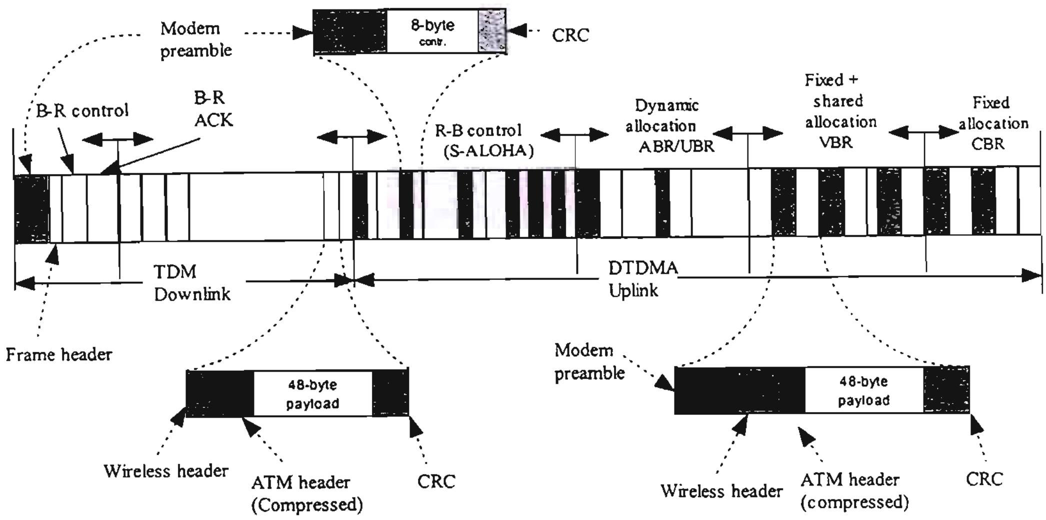


Fig. 2.9: Dynamic TDMA/TDD MAC frame format [4]

The uplink subframe is divided into four sections for reservation bandwidth (R-B) control, followed by ABR/UBR, VBR, and CBR slots respectively. Each section consists of an integer number of time slots. The boundary between these sections and uplink and downlink is variable according to the traffic experienced by the network. Fig. 2.9 shows the DTDMA/TDD frame format. When a MT has packets to transmit, it sends a request in the R-B control slots using S-ALOHA. At the beginning of the next frame, the BS transmits slot allocation information along with ACKs and other control information. For CBR traffic, slot allocation is done once during call establishment. A fixed allocation of slots is assigned according to user requests. When CBR slots are no longer available, arriving CBR calls are blocked. For VBR traffic, the allocation is accomplished on a fixed shared basis, with some slots assigned for the duration of an active period, plus some extra slot(s) assigned according to a usage parameter control (UPC) based statistical multiplexing algorithm. Arriving VBR calls are also blocked when VBR slots are no longer available. For the case of ABR/UBR traffic, slot allocation is performed on a burst-by-burst basis via dynamic reservation of ABR/UBR slots and unused CBR and VBR slots. ABR/UBR calls are always accepted subject to appropriate rate flow control.

2.6.1.3 Dynamic Slot Assignment (DSA++)

Dynamic Slot assignment (DSA++) is based on FDD and TDMA structure with a variable frame length [59]. The frame length varies from 8 to 15 slots and is known as a signalling burst. Each slot is the size of an ATM cell. Fig. 2.10 shows the frame structure of the protocol.

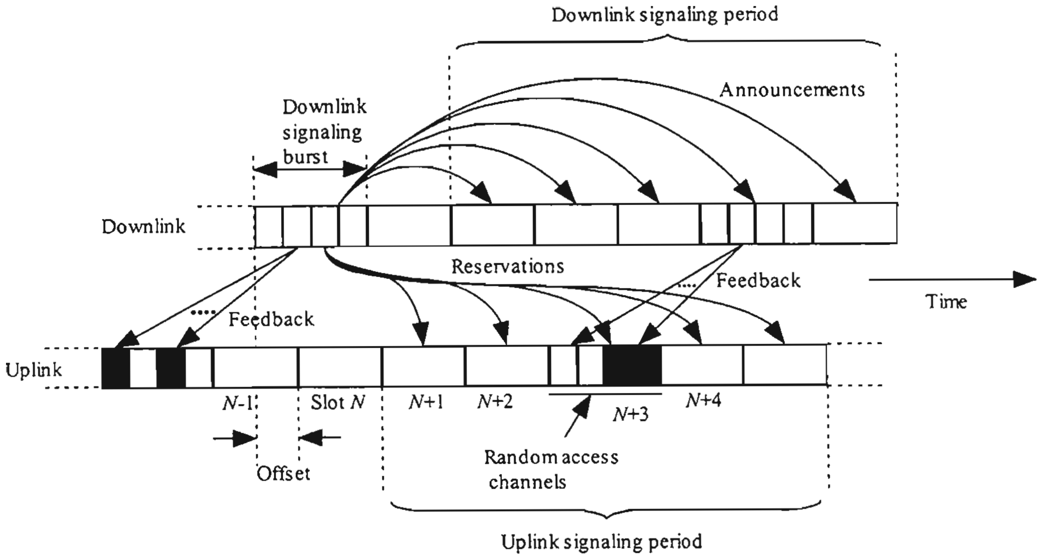


Fig. 2.10: DSA++ protocol [37]

The downlink frame starts with a downlink signalling burst, which specifies the length of the frame. The downlink signalling burst contains the following information:

- A reservation message for each UL slot of the signalling period.
- An announcement message for each downlink slot of the signalling period
- A feedback message for each random access slot of the previous signalling period
- Additional signalling messages such as collision resolution and paging channel.

Each downlink signalling period has the same length as the uplink signalling period. There is an offset between the starting points of each period to compensate for round trip propagation delay. After the initial registration procedure, the BS allocates transmission capacity to the MTs on a slot-by-slot basis. The allocated transmission capacity is determined by priority calculation for each MT being served. The priority is determined according to a set of dynamic parameters (DPs) which includes the number of waiting ATM packets and their due times. The DPs are included in the header of each ATM packet being transmitted by the each MT. The BS may ask a MT to update the DPs by either polling or random access using mini-slots. For this purpose, the BS uses an algorithm which calculates the number of mini-slots that must be available in the next frame according to the following parameters:

- Number of MTs in contention
- Probability of a new packet arrival at each MT in contention mode since the last transmission of their DPs
- Throughput of random access procedure

The priorities assigned to ATM classes of services are: CBR>VBR>ABR>UBR. For the CBR and VBR classes, which are delay-sensitive, a factor called relative urgency is used to decide which MT will transmit or receive in the next signalling period. The advantage of this protocol is that the downlink signalling burst releases all other slots in the next signalling period, which means that a mobile power control algorithm could be implemented if a need arises.

2.6.1.4 Mobile Access Scheme based on Contention and Reservation for ATM (MASCARA)

MASCARA [61] is the MAC protocol designed for the MAGIC Wireless ATM Network Demonstrator project. MASCARA uses variable length time frames as the DSA++ protocol. The multiplexing of uplink and downlink traffic is based on TDD. As shown in Fig. 2.11, the MASCARA time frame is divided into three periods: broadcast, reserved, and contention, which are further subdivided into time slots. Each of the three periods has a variable length, depending on traffic load to be carried in the wireless channel.

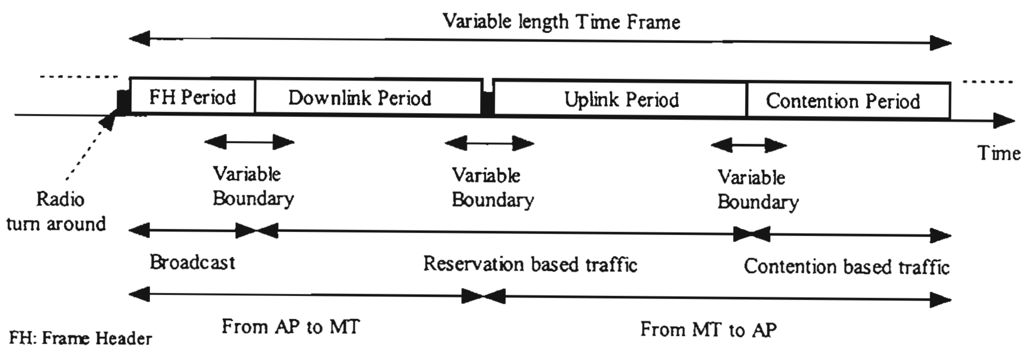


Fig. 2.11: General frame structure of MASCARA [61]

The broadcast period is used to notify all MTs of the structure of the current time frame and the scheduled uplink transmissions, and to acknowledge the requests from the previous frame. The reserved period consists of a downlink period in which the AP or BS transmits the downlink data and the uplink period where the MTs transmit packets in the order scheduled by the BS. The contention period is used by MTs to send new reservation requests to the BS. MTs with established connection may use piggybacking to request more bandwidth. All packets that are transmitted through the contention period use the slotted-ALOHA access protocol. After the request reception, the BS makes uplink data slot assignments based on a leaky bucket token scheme called Prioritized Regulated Allocation Delay-oriented Scheduling (PRADOS) [60].

2.6.2. CDMA Demand Assignment Protocols

2.6.2.1 Wireless Multimedia Access Control with BER Scheduling (WISPER)

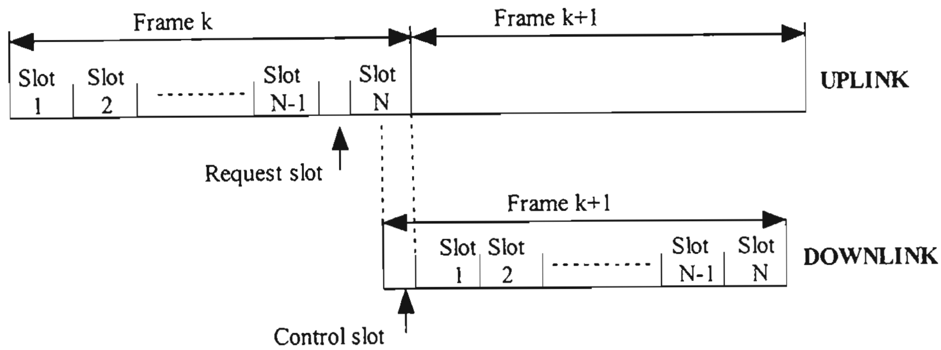
The WISPER [10] protocol is based on MC-CDMA, and uses FDD to organise the uplink and the downlink. For both uplink and downlink, time is divided into frames, and each frame is further divided into time slots as shown in Fig. 2.12a. The frame length is chosen to coincide with the packet arrival rate of the most abundant traffic class such as voice. The uplink consists of one request slot and several packets slots that can carry any traffic class. The request slot can be used by new MTs to place admission requests to be admitted to the wireless network, or by registered MTs to place transmission requests. The request slot is placed in a position that will allow the BS enough time to process the request and assign slots before the next downlink frame. In order for the MTs to have transmission information on time, the downlink control slot precedes the start of an uplink frame. The downlink control slot is used by the BS to acknowledge MT requests and provide slots allocations in the next uplink frame.

The BS is assumed to allocate a unique primary PN code to each mobile terminal admitted in the network. When a mobile wants to be admitted to the network, it randomly selects a primary PN code from a pool (managed by the BS) of codes. A mobile terminal, n, admitted to the system can derive m different spreading codes from the primary PN code and transmit at higher rate than the basic rate. The different spreading codes are derived from the primary PN codes as follows: when C_n^{PN} is the primary PN code for user n, the different spreading codes $\{C_n^{(i)}, i = 1,2,....., M\}$ are obtained by

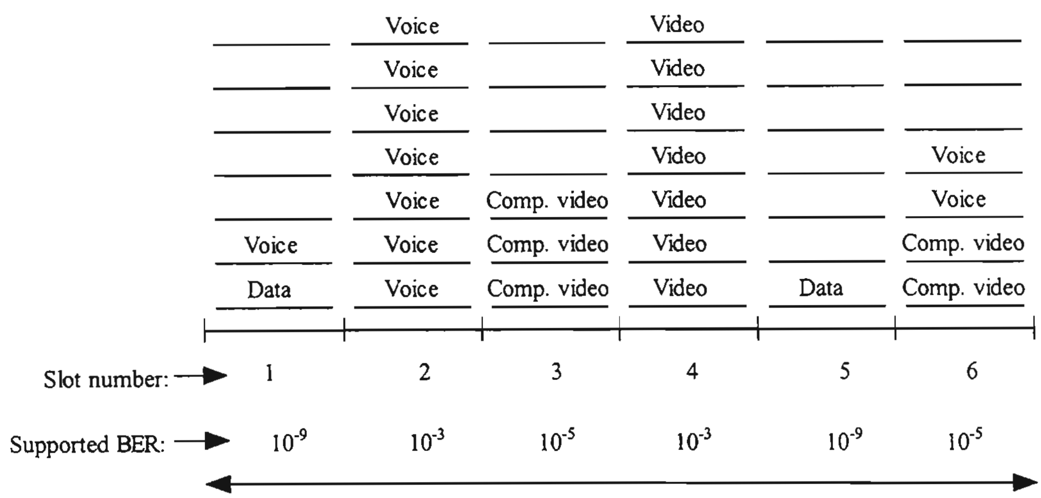
$$C_n^{(i)} = C_n^{PN} \times D_i, \quad D_i \perp D_j, \quad i \neq j \tag{2.3}$$

where D_i's are from a set of orthogonal codes, such that $C_n^{(i)} \perp C_n^j$, $i \neq j$ is guaranteed.

The MTs are assumed to generate packets in batches, where all packets in a batch have the same time-out specifications. Whenever a MT has new packets ready for transmission it sends a transmission request to the BS indicating the number of packets in the new batch as well as their time-out value. The transmission request can be sent on the request slot or piggybacked onto another packet being transmitted. Once a request has been received, the BS uses a data structure or tables to keep track of the batch associated with the request. This information is kept until the packets have been successfully received or they are timed out.



a) Uplink and downlink channels: a timing diagram



b) An example of frame structure and slot assignment

Fig. 2.12: Frame structure in WISPER [4]

WISPER designates slots that can support certain BERs and schedules packet transmissions in these slots in such a way that the bandwidth can be utilised efficiently. To use the available bandwidth efficiently, packets that have equal or similar maximum BER specifications are transmitted in the same slot. As a result, the capacity of a time slot is determined by the packet with most stringent BER requirements. Fig. 2.12b shows an example of slot assignment for multimedia traffic. A packet scheduler is used to assign packets to slots and apart from setting BER also ensures that time-out values are not exceeded. When there are more packets to be transmitted than can be accommodated in the frame the packet prioritizer function of scheduler is used.

WISPER claims to offer improved performance over CDMA protocols that use single BER thresholds. Packet losses are minimised by a packet prioritization scheme that determines packet transmission order by considering remaining time before timeout.

2.6.2.2 An Uplink CDMA System Architecture with Diverse QoS Guarantees for Heterogeneous Traffic

In [15] a MAC protocol based on the principles of the Distributed-Queuing Request Update Multiple Access (DQRUMA) [40] and utilizing MC-CDMA as WISPER is considered. The protocol support two traffic classes, namely class I and class II in the same slot. Class I carries connection orientated voice and video traffic, while class II carries both delay sensitive and delay tolerant loss-free data (e.g. e-mail, remote login). Class II utilizes bandwidth remaining from Class I on a best effort basis. The uplink and downlink frame structures for the protocol are shown in Fig. 2.13. The uplink frame consists of a mini-slot for ACK/NAK for each downlink packet, a contention-based transmission request mini-slot, a piggyback rate-request mini-slot, and a packet-transmission slot. The downlink frame consists of a mini-slot for ACK/NAK for each uplink packet, a mini-slot for result announcements of contention-based transmission requests, a mini-slot for uplink packet transmission permissions, and a slot for downlink packet transmission.

In this protocol MTs are assumed to carry a combination of traffic and their maximum rates are dependant on their traffic mix. Whenever a MT has new traffic to send and is not scheduled for transmission in the next frame it requests code-channels via a BS orientated transmission request. The access method used in the contention slot is based on slotted ALOHA with harmonic backoff in case of request failures. MTs that have established connections and are scheduled for transmission, uses piggyback mini-slots to request code-channels from the BS. The BS permits transmission for class II traffic according to a round-robin scheduling policy. For class I connections an admission test is performed such that the target SIR ratio is met. This protocol offers BER guarantees for class I traffic through power control and FEC techniques. Class II traffic use FEC and selective repeat ARQ scheme, with power control to obtain an optimum SIR target that maximizes aggregate throughput.

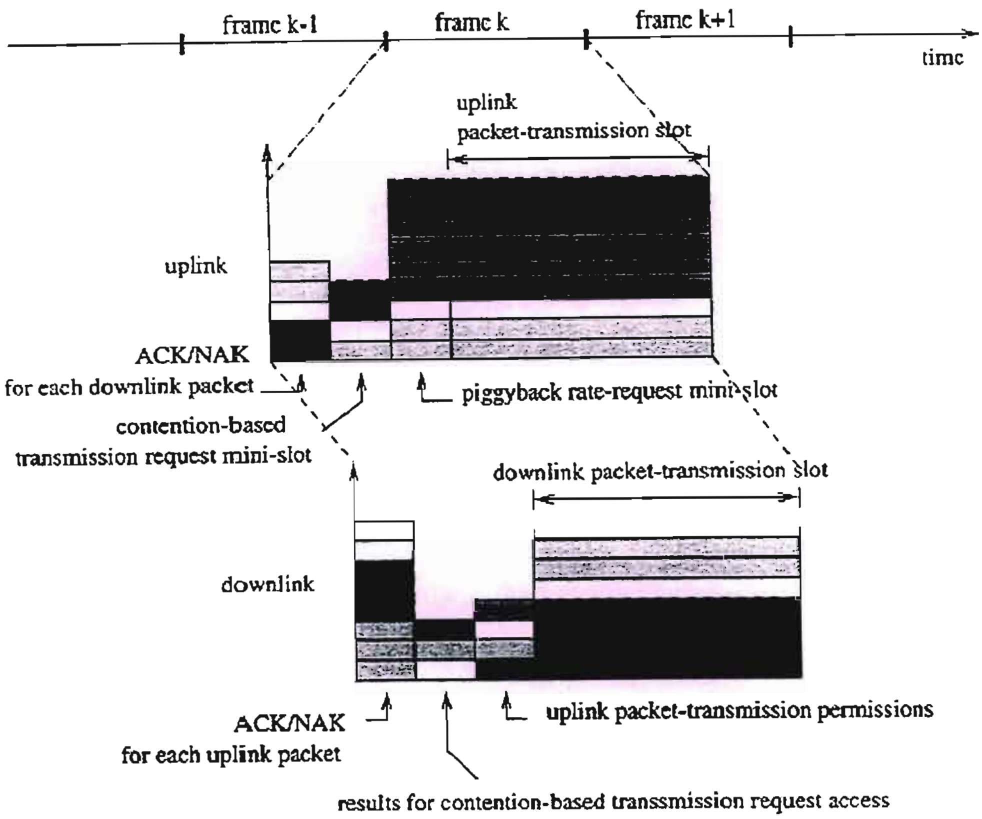


Fig. 2.13: Uplink and downlink frame structure [15]

2.6.2.3 IP QoS Delivery in a Broadband Wireless Local Loop

In [14] the authors proposed a MAC protocol and physical layer for IntServ IP-based broadband fixed wireless access (FWA) architecture supporting differentiated QoS traffic. The physical layer is based on OFDM-CDMA with FDD technique. A symbol interval is used to transmit data symbols belonging to K users and as result the soft capacity nature of CDMA is not utilised. The MAC protocol supports two service classes: guaranteed bandwidth (GB) and best effort (BE). The GB class carries traffic with guaranteed bandwidth requirements and delay sensitive multimedia traffic such as voice and video. The BE classes accommodate the existing Internet traffic and utilizes the bandwidth remaining from GB.

The frame structure of the protocol is depicted in Fig. 2.14. A frame consists of N time slots, each supporting up to K orthogonal codes for simultaneous transmission. A radio terminal (RT) may transmit on several slot-code pair without restrictions. Bandwidth requests are transmitted in the uplink mini-slots and the downlink acknowledgement channel informs the RTs of the number of time-code pairs and positions. Each RT is assigned its own Request-Acknowledge slot-code pair.

A generalised processor sharing (GPS) scheduler which shares the capacity among competing users according to their actual and predefined weights is considered. The users' weights in the protocol are related to the RTs traffic descriptor (TD) and are passed down to the MAC layer from higher layers. The scheduling operation is divided in two phases.

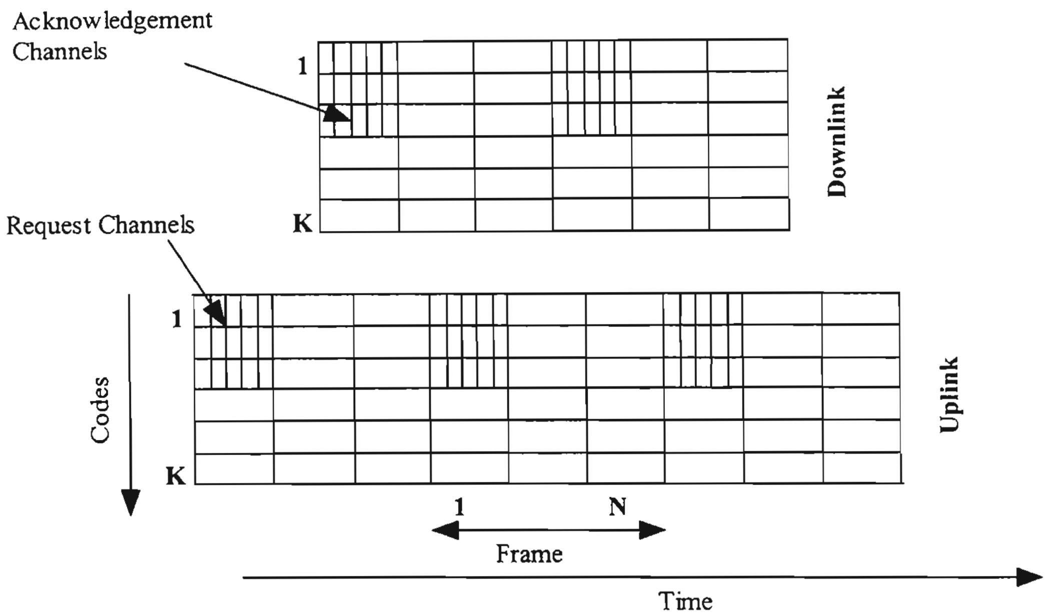


Fig. 2.14: MAC frames structure of [14]

In the first phase the overall radio link capacity is shared among RTs according to their overall request and weights. In the second phase each RT shares the bandwidth it obtained among the competing GB and, if possible the BE traffic.

A useful lemma is given relating to the aggregation of source with common delay bounds. The authors show that if L packet flows with different TDs yet the same delay bounds are multiplexed, common delay bounds can be met provided the output capacity of the FIFO multiplexer is equal to that required by a GPS scheduler. Traffic is characterised using a dual leaky bucket specification and an admission algorithm formed accordingly.

2.6.2.4 Wide-Band TD-CDMA MAC with Minimum-Power Allocation

In [35], Wang proposed a wide-band time-division/code-division (TD-CDMA) MAC protocol with minimum-power and rate and BER-scheduling. The protocol was developed to satisfy

3GPP resource management and MAC protocol specifications for TD-CDMA networks. This protocol considers TDD technique for multiplexing uplink and downlink. Fig. 2.15 shows the frame structure of the protocol. Due to the asymmetric nature of uplink and downlink traffic load, multiple switching points exist between the uplink and downlink. These switching points are determined by a dynamic channel allocation algorithm. MC CDMA operation is assumed for both uplink and downlink. Three transport channels, namely dedicated channel (DCH), random access channel (RACH) and broadcast channel (BCH) are defined. The DCHs are used for packet transmissions, while RACH are for sending mobile terminal requests to the BS. The BCH is used by the BS to send feedback of resource allocation from the BS to MT.

The MTs are assumed to generate packets in batches, where all packets in a batch have the same time-out specifications as in WISPER. When a mobile terminal with real-time traffic wants to establish a connection, it enters an *admission state* and sends an admission request in the RACH by spreading its signal with a randomly selected primary PN code. Upon receiving an admission request the BS invokes an effective bandwidth call admission control (CAC) algorithm to check whether enough bandwidth is available in the system. If bandwidth is available, the request is accepted, and the primary PN code is reserved for the mobile terminal. The MT then enters a *transmission state*. Otherwise the admission request is blocked, and the MT will try to send the request after a random backoff time.

Whenever a MT in transmission state or with non real-time traffic has new packets ready for transmission it first sends a transmission request to the BS indicating the number of packets in the new batch as well as their time-out value. The transmission request can be sent in RACH or piggybacked in DCH. Once the BS has collected all transmission requests from MTs, it uses a packet scheduling scheme to determine how the packets of multimedia services are accommodated to each time slot. The packet scheduler is a joint rate- and BER-scheduling scheme based on minimum power allocation for each code channel. The rate scheduler reserves a time slot with more available code-channels for a MT with more packets waiting for transmission. These schemes maximise slot capacity by minimising mutual interference between code-channels due to different MTs in a time slot. The BER scheduler attempts to accommodate packets with different BER requirements in different time slots as in WISPER. Based on the permission feedback from the BS, MTs transmit packets in the specified time slots with allocated power levels. ARQ is applied for non real-time loss-sensitive service. The protocol is shown to perform better than WISPER.

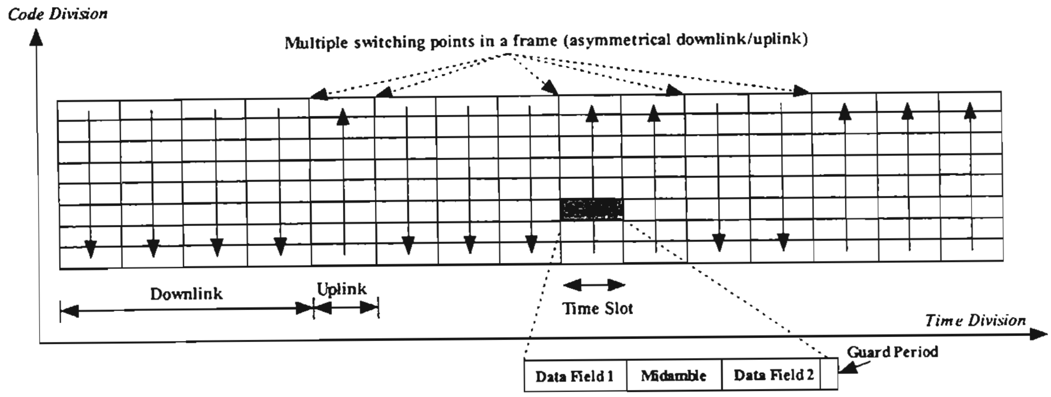


Fig. 2.15: The structure of wide-band TD-CDMA frame [35]

2.6.2.5 Wireless ATM over CDMA with Multi-Class BER guarantees (WAC/MB) protocol

The WAC/MB protocol was proposed by Majoor for wireless ATM [7]. The protocol supports CBR, VBR, and ABR traffic and offers multi-class BER guarantees through appropriated power assignment. The uplink transmission is considered. Time is divided into fixed length frames which are further divided into four sections, as depicted in Fig. 2.16. The first section is a mini-slot for CBR and VBR reservation requests. The following three sections carry padded ATM cells for CBR, VBR, and ABR traffic, respectively.

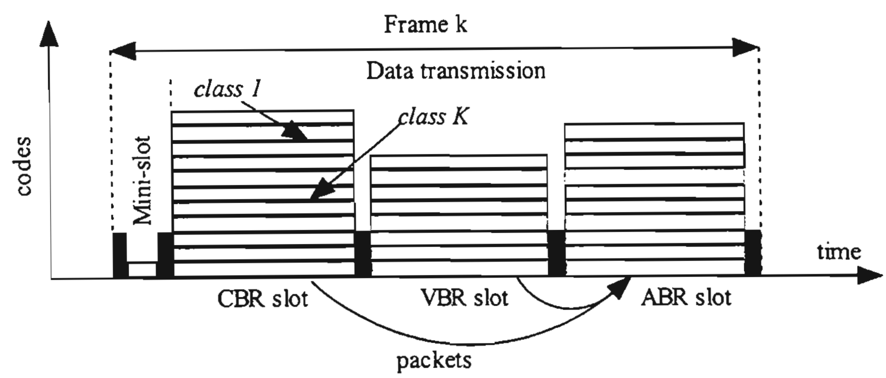


Fig. 2.16: Uplink frame structure for WAC/MB [7]

A CBR or VBR MT who wishes to initiate a session selects a spreading code from a finite pool at random and transmits a reservation packet in the mini-slot. A connection admission algorithm is performed at the BS to determine whether the connection may be accommodated in the relevant data section without violating QoS of admitted terminals. Mobiles with successful reservation requests are informed via the downlink channel and their allocated bandwidth and QoS requirements are guaranteed for the duration of the session. A mobile that was not allocated resources by the admission algorithm will remain in the contention state until either the termination of its call or reservation is achieved.

MTs with ABR traffic do not make reservations, but transmit their data in the ABR section in connection with other users in a SS/ALOHA fashion. CBR and VBR mobile terminals in the connection state are allowed to transmit their packets in the ABR section in conjunction with regular ABR traffic in order to reduce system cell losses. Since VBR source varies its data rate from frame to frame, a piggyback scheme is used to inform the BS of the required data rate in the next frame. The number of packets that a VBR mobile may transmit in the VBR section is limited. Excess packets are transmited in the ABR section. To accommodate mobile terminals transmitting at higher data rates, MC CDMA is used. BER QoS guarantees are achieved by assigning higher powers to users with more stringent BER requirements. All users with the same BER will transmit such that they are received at the same power, and consequently receive the same interference. The disadvantage of this protocol is that ABR traffic is not allowed to use the CBR and VBR sections when they have low traffic loads, and as a result bandwidth is under utilised.

2.7 Summary

In this chapter, the literature survey on existing MAC protocols for centralised wireless networks was presented. The protocols were classified into fixed assignment schemes, random access protocols, guaranteed assignment schemes, random reservation protocols and demand assignment protocols. For random access protocols, narrowband schemes were discussed; namely ALOHA, S-ALOHA, CSMA, and ISMA, followed by CDMA based schemes: SS/ALOHA and CLS. In guaranteed assignment protocols, three protocols were discussed, i.e., Zhang's proposal, DTMP, and Acampora's proposal. This was followed by the discussion of random reservation access protocols: PRMA, MD PRMA BB and DT/CDMA. Finally, selected demand assignment protocols were presented, and they were classified into TDMA and CDMA. The TDMA based DA protocols discussed are DQRUMA, DTDMA/TDD, DSA++, and MASCARA. CDMA based DA protocols presentation included, WISPER, and Uplink CDMA protocol in [15], IP QoS protocol in [14], Wide-Band TD-CDMA MAC protocol in [35] and WAC/MB.

Chapter 3

The Wireless DiffServ IP over CDMA (WDIP/CDMA) MAC Protocol

3.1 Introduction

In this chapter we present our proposed MAC protocol for carrying DiffServ IP traffic over a centralised CDMA wireless network. The proposed protocol which is called Wireless DiffServ IP over CDMA (WDIP/CDMA) is based on demand assignment (DA) and random reservation access (RRA) classes of MAC protocols. The DA scheme is used to support traffic from assured service (AS) and premium service (PrS), since they require QoS guarantees. The Best effort (BE) traffic is supported by the RRA scheme and is served using remaining capacity. By using RRA for BE traffic, scheduling complexity and congestions in the DAs request channels can be reduced. Thus, the overheads and access delays which are the drawback in DA protocol could be reduced as well. The RRA part of the WDIP/CDMA protocol is based on the data-only DT/CDMA MAC protocol proposed and analyzed by Judge in [6]. We have extended the ideas of this protocol to a frame structured system. The advantage of the frame structure is that the packets in each time slot in the frame will not be equally affected by the MAI. Furthermore, the packets in the multi time slots frame are small compared to an equivalent frame with a single slot. When smaller packets are used packet loss due to burst errors when the channel is in fade can be minimised.

A detailed description of the MAC protocol is presented in section 3.2. In section 3.3 the traffic models for the MAC protocol are described. This is followed by the discussion of the system capacity in section 3.4. Finally, in section 3.5 simulation results are given and discussed.

3.2 Protocol description

3.2.1 Frame Structure

Time is slotted and organised into fixed length TDMA frames as shown in Fig. 3.1. TDD is used to multiplex downlink and uplink transmissions. In the uplink, each frame is divided into a mini-slot and L_U packet slots; in the downlink, each frame is divided into a control slot, and L_D packet slots. In each type of slot, CDMA multiplexing is used with a fixed spreading gain. Up to

Q_{CS} spreading codes for packet transmission (1 packet = 1 code-channel) called code-channels can be supported in a time slot. The code-channels in each time slot are divided into AS and PrS section using a movable boundary scheme. Movable boundary techniques [21] are used to provide efficient bandwidth utilisation in TDMA protocols that divide a frame into sections for integration of different traffic types. The maximum number of code-channels Q_{CS} in a slot is determined according to the BER requirements of the admitted traffic. The mini-slot is used by AS and PrS mobile terminals (MTs) to send transmission requests to the Base Station (BS). The downlink control slot is used by the BS to acknowledge (ACK) successful requests and provide information about bandwidth allocations in the current frame.

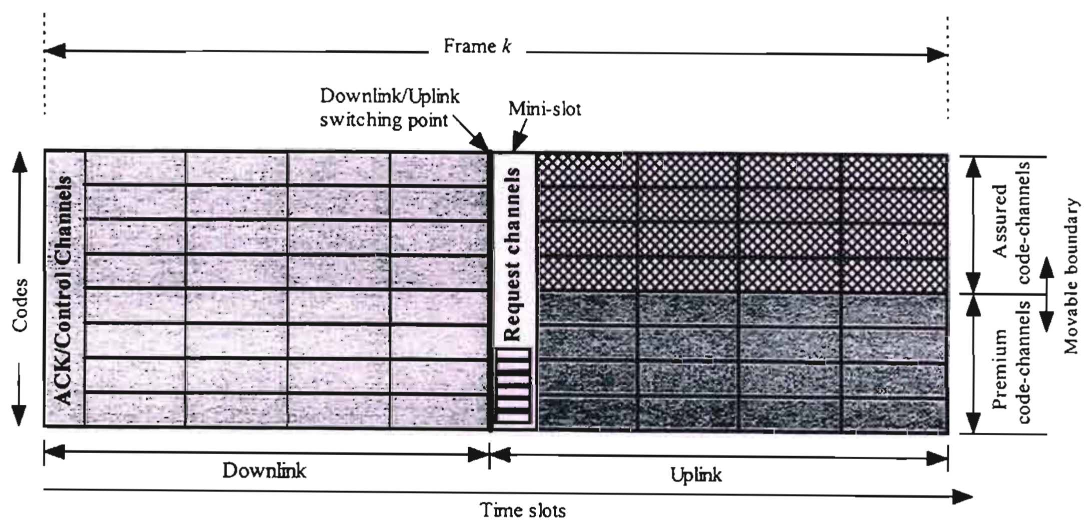


Fig. 3.1: Frame Structure

AS and PrS MTs transmitting at different data rates are accommodated by using multi-code CDMA. Thus one MT may transmit several packets in a frame by modulating data packets by different PN codes. MTs with BE traffic always make use of one code-channel per slot per frame.

3.2.2 Bandwidth allocation strategy

A maximum guaranteed bandwidth (number of code-channels) of $P_{\max}^{ps}$ code-channels/frame is allocated to premium service. To meet QoS requirements of assured service, a minimum guaranteed bandwidth of $A_{\min}^{as}$ code-channels/frame is reserved. Since BE service has no QoS requirements, no bandwidth is reserved for it, and consequently no call admission control is required. The sum of the bandwidth allocated to AS and PrS gives the total system capacity C. Therefore, BE service is served with the remaining bandwidth from AS and PrS classes. In addition to the minimum bandwidth, AS can utilize unused PrS bandwidth.

Each AS or PrS traffic flow will be assigned a guaranteed bandwidth according to service level agreement specifications. Since MTs with bursty (VBR) traffic vary their data rate from frame to frame, their allocated guaranteed bandwidth will not always be utilized. To achieve high bandwidth utilization, this bandwidth can be filled with BE traffic during the periods when it is not required.

3.2.3 Call Admission Control for AS and PrS

The function of a call admission control (CAC) is ensuring that the system capacity is not exceeded and that QoS for admitted users is not violated. Before a new call is admitted, the CAC is invoked to determine if the call can be accepted or rejected based on available bandwidth. From the bandwidth allocated per traffic class, we can determine the CAC. For PrS, a new call is admitted if after its admission the following CAC criterion is satisfied.

$$\sum_{i=1}^{R_{ps}} P_i \leq P_{\max}^{ps} \tag{3.1}$$

where P_i is the guaranteed number of code-channels assigned to the i^{th} premium service user and R_{ps} is the number of admitted premium users including the new user.

For AS, a new call is admitted if after its admission the following CAC criterion is satisfied.

$$\sum_{i=1}^{T_{as}} A_i \leq \left\{ C - \sum_{i=1}^{R_{ps}} P_i \right\} \tag{3.2}$$

where A_i is the minimum guaranteed code-channels for the i^{th} assured service user and T_{as} is the number of admitted assured service users including the new user. If PrS capacity is fully used, the CAC for AS is given by

$$\sum_{i=1}^{T_{as}} A_i \leq A_{\min}^{as} \tag{3.3}$$

3.2.4 MAC procedure for AS and PrS

When a PrS mobile terminal wants to transmit packets, it randomly selects a primary spreading PN code from a finite code pool and transmits its request packet in the mini-slot using the slotted SS/ALOHA protocol. The request contains information such as DSCP, BER and transmission rate requirements. Upon receiving the request the BS invokes a CAC based on equation (3.1) and then decides whether to admit or reject the call. If (3.1) is satisfied even after the user is admitted, the BS replies with an acknowledgement for the request and relevant feedback information. The feedback information specifies the number of code-channels, time slots, and transmission power level for each time slots. Once the user is admitted it reserves the allocated code-channels for the duration of its call and hence cannot be dropped. Otherwise, the request is blocked. The MT will then try to send the request after a random back-off time. MTs already admitted use a piggybacking scheme to inform the BS of the required bandwidth in the next frame. This allows the BS to efficiently determine the bandwidth available for BE traffic. Due to the bursty nature of some of the traffic, users may generate traffic in excess of their guaranteed bandwidth. In this protocol, the excess traffic will be transmitted as best effort or dropped immediately. The excess traffic is immediately dropped because premium service is delay sensitive and its packets cannot be buffered for a long time.

An AS terminal will follow the same procedures as the PrS one, except that now the BS will use a CAC based on equation (3.2). The admitted AS terminals are categorised into permanently admitted (PA) and temporarily admitted (TA) groups. Permanently admitted AS terminals are those that are allocated a portion of the bandwidth specifically assigned to AS. Once admitted, a PA terminal reserves the allocated code-channels for the duration of its call and cannot be dropped. Since assured service does not have strict delay requirements, excess traffic can be buffered instead of being dropped and transmitted when bandwidth become available. Temporarily admitted AS terminals are those that are allocated unused PrS bandwidth. They have unconditional priority over new AS requests when the AS bandwidth become available. When a PA user terminates its call, the BS reallocates the bandwidth that remains available to a TA terminal. If the PrS bandwidth allocated to a TA terminal is required for a new PrS call and immediate reallocation cannot be performed, the TA terminal is forced to terminate in order to achieve the PrS terminal's QoS guarantees. This phenomenon is known as 'break' [27].

With respect to the break phenomenon, we define a signalling message referred to as the "*TA breaks*" (**TB**) signal (or tone). The **TB** tone will be broadcasted by the BS to notify temporary admitted AS users about a break. Upon detecting the **TB** tone, TA users abort transmission and

PA users transmitting excess traffic reduce their data rates by buffering the extra packets with a probability given by

$$\gamma = \frac{B_t^a}{B^p} \tag{3.4}$$

where B_t^a is the bandwidth used by TA assured users and B^p is the bandwidth required by new PrS users requesting admission. We assume that the probability γ is included in the **TB** tone.

3.2.5 MAC procedure for BE service

For this MAC procedure, we implement a MAI threshold (denoted by Q_{CS}) which represents the maximum allowable number of users per time slot. The MAI threshold of a time slot is time variant depending on BER requirements of packets being transmitted. This threshold should be chosen such that QoS guarantees for all admitted users are not violated. The expected packet error probability must satisfy the following criteria:

$$P_{Err}^{pkt}(a+b+p) \leq P_{\max}^{BE} \qquad for\ (a+b+p) \leq Q_{CS} \tag{3.5}$$

where $P_{\max}^{BE}$ is the maximum acceptable bit error rate for BE traffic, $P_{Err}^{pkt}(a+b+p)$ is the bit error rate given the total number of simultaneous transmissions $(a+b+p)$ in a time slot, a is the number of packets transmitted by AS users, b is the number of transmitting BE users, and p is the number of packets transmitted by PrS users. Based on equation (3.5), the maximum number of channels available to BE service is given by

$$\beta = \max\ (0, Q_{CS} - (a+p)) \tag{3.6}$$

The BS is able to determine the capacity available, β, for best effort in every time slot of the next frame based on the transmission updates and requests it receives from AS and PrS terminals for every frame.

In order to avoid the threshold from being crossed frequently during contention and currently admitted user's packets corrupted, we defined another threshold α below β such that $0 \leq \alpha \leq \beta$. With respect to β and α, we define the following signalling tones to notify BE users about the state of a time slot in the previous and the next frame:

- **BSB**: **B**est effort **S**ervice **B**locking tone is emitted by the BS to indicate which time slots in the next frame are blocked or have the BE traffic load exceeding α .
- **BSC**: **B**est effort **S**ervice **C**ollision (Overload) tone is emitted by the BS to indicate which time slots in the previous frame have the BE traffic load exceeding β .

The BS broadcasts these tones in an out-of-band downlink-signalling channel on a frame-by-frame basis. Based on these thresholds and the associated signalling tones, we can describe the MAC protocol that each BE terminal in the network adhere to as follows:

1. A BE terminal that has a message ready to transmit in the current frame, first select a PN code and a time slot in which to transmit randomly. Then it listens to the downlink-signalling channel for the presence of the **BSB** tone in the selected time slot. If the **BSB** tone is present, the packet is blocked and the terminal (re)enters to the retransmission routine.
2. If the **BSB** tone is not present, the terminal initiates transmission immediately at the start of the chosen time slot and known as contending. During the next frame, the terminal listens to the downlink-signalling channel for the presence of the **BSC** tone. Upon detecting the **BSC** tone, all contending users immediately abort transmission and (re)enter backlog retransmission routine (i.e. step 4). If no collision has occurred, the users continue transmitting using the initially acquired code-channel.
3. Admitted BE terminals are similar to temporary admitted AS terminals. During high traffic loads, they are forced to abort message transmission to open channel capacity to AS or PrS terminals.
4. Messages that were aborted, blocked (**BSB** tone) or dropped by the **BSC** tone are retransmitted in full after a random back-off delay. This delay is assumed to be geometrically distributed with parameter p_R and mean duration $R = p_R^{-1}$.
5. Terminals are notified via the **ARQ** system or **ACK** slot whether a transmitted message was successfully received or not. The corrupt packets in a message are combined to form a new message and then retransmitted. The new message is transmitted after the transmission of the actual message is completed.

3.3 Traffic Model

The premium and assured service traffic in the network is assumed to be generated by a finite population N_{PS} and N_{AS} of premium and assured terminals, respectively. Each terminal can be in one of two states, namely transmission state (TRA), or silence state (SIL). When a mobile terminal has no calls to transmit, it is in SIL state otherwise it is in transmitting state with a call in progress. This model is derived from the ON/OFF process well-known for characterizing voice traffic [15].

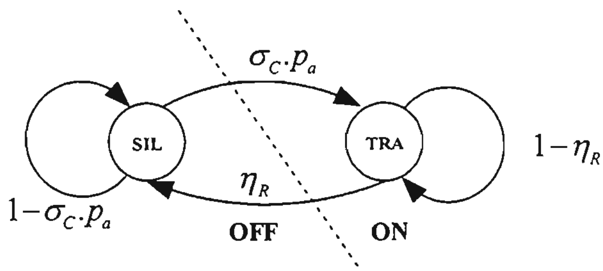


Fig. 3.2: Traffic model for premium service terminal

For PrS, the traffic model with Markov transition probabilities between states is illustrated in Fig. 3.2. The probability that a call is generated in a frame is σ_C, and η_R is the probability that a call ends in a frame. The probability of a user being admitted, p_a, is a function of the bandwidth used by the number of terminals in the transmitting state.

The traffic model for AS is shown in Fig.3.3. The transmitting state is further divided into temporarily (TM) and permanent (PM) states. Terminals in the TM state transmit on unused PrS bandwidth. When a call terminates from the TM state, it is converted to PM state.

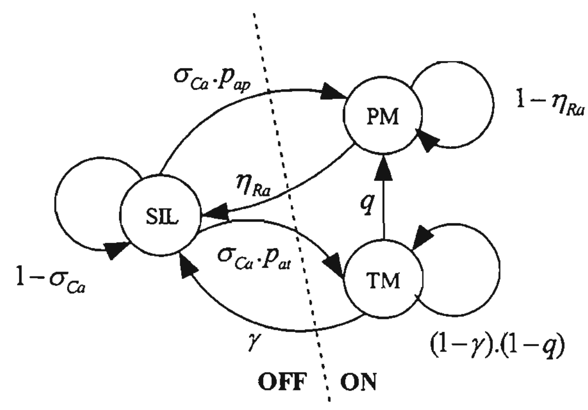


Fig. 3.3: Traffic model for assured service terminal

The probability that a call is generated in a frame is σ_{Ca}, and η_{Ra} is the probability that a call ends in a frame. The probability that a temporarily admitted AS terminal is forced to terminate is γ, while q is the probability that the terminal is converted to the permanent state. The probability that an AS terminal is admitted to the PM (TM) state is given by p_{ap} (p_{at}) and is a function of the bandwidth used by the terminals already admitted.

In the transmitting state, a premium service or assured service terminal with VBR traffic may be assumed to take one of several states. This model is based on [4], where it is used to model VBR video traffic. Each state is statistically independent and corresponds to a constant bit rate for an exponentially distributed transmission time, with a mean call holding time equal to 160 ms. The bit rate values for the states are obtained from a truncated exponential distribution. This distribution is defined with minimum and maximum bit rate values.

The BE traffic is assumed to be generated by an infinite population of identical BE terminals $N_{BE} = \infty$. A BE terminal can be in one of the three states shown in Fig. 3.4. A BE terminal in SIL becomes active by entering the backlog (BK) state or transmission (TRA) state with probability p_o. Terminals in BK state attempt retransmission in the current frame with probability p_R. New messages are generated only by BE terminals in the SIL state. The probability that a message is successfully transmitted is given by σ_s, and σ_{cp} the probability that a transmitting terminal returns to the BK state due to one or more corrupted packets that need to be retransmitted. The success probability, p_{fs} on which the success of the first packet of a BE message depends, is a function of the bandwidth used by BE, AS, and PrS users in the relevant time slot.

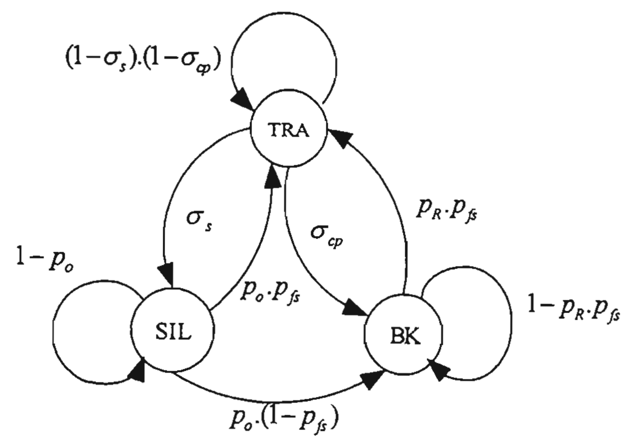


Fig. 3.4: Traffic model for best effort service

The average length of BE message is assumed to be in the order of the state holding time for AS or PrS VBR traffic. Therefore, the unused bandwidth for AS or PrS when transmitting at data rates less than their guaranteed data rate can be effectively used by BE traffic.

3.4 System Capacity

Unlike TDMA systems, the capacity of a CDMA system is a function of some maximum tolerable bit error rate due to MAI, i.e. the channel capacity is interference limited and time variant. In TDMA/CDMA the capacity is difficult to estimate because of capacity variations in each time slot of the frame. Power control algorithms are required to reduce MAI in each time slot and maximise slot capacity. With power control each mobile terminal adjusts its own transmit power to ensure an adequate BER QoS or SINR at the BS. In CDMA/TDMA systems using conventional power control schemes it is assumed that signals from all users are received at equal power levels [4]. In equal received power level systems, the capacity of a time slot is limited by the traffic type with stringent BER requirements. This results in under utilization of the system resources. In [4] this problem is solved by scheduling traffic with different BER requirements at different time slots, thus maximizing capacity. In [35] the slot capacity is maximized by using a minimum power allocation algorithm in which the transmit power for each code channel is minimised. The slot capacity is maximised by accommodating packets from the BER classes in the same time slot.

In our system we assumed a maximum fixed slot capacity (for simplicity), set by the least stringent BER service capacity. The system capacity is then given by [7],

$$C = N_{Slot} \times \left(1 + \frac{3.G}{SINR_{\max}}\right) \tag{3.7}$$

where N_{Slot} is the total number of uplink time slots in the frame, G is the processing gain and $SINR_{\max}$ is the target SINR of the least stringent BER traffic type. Users with stringent BER requirements and with higher data rate are assigned higher received powers which are multiples of the least stringent BER service to achieve target SINR as in [33]. The maximum slot capacity decreases with more stringent BER requirements. We define the received power levels for BER class with least stringent BER requirements as the minimum power, $Pw_{\min}$ and the corresponding slot capacity is denoted by $Q_{cs}^{\max}$. If a packet can tolerate $(Q_{cs}^{\max}-1)$ other simultaneously transmitted packets with power $Pw_{\min}$, it can tolerate other $(Q_{cs}^{\max}-1)/p_i$ transmitted packets with power $p_i Pw_{\min}$. Therefore, a packet with stringent BER requirements can be visualised as consuming p_i code-channels for its transmission.

For a system which employs retransmission schemes, like the BE service MAC protocol, Judge showed in [6] that the optimum MAI capacity is set such that the BER is approximately equal to the inverse of the packet length. Since the BE service MAC is random in nature, the blocking threshold α needs to be set at an appropriate value below the optimum MAI capacity, to ensure that BER requirements of admitted AS and PrS terminals are not violated.

3.5 Simulation Results

3.5.1 Protocol Parameters

The parameters which were used for the simulation of the proposed protocol are provided in table 3.1. The call holding times for PrS and AS and guaranteed data rates are assumed to be equal. Two BER classes are considered for both PrS and AS terminals. The maximum slot threshold β for BE traffic is equal to the slot capacity $Q_{cs}^{\max}$.

Table 3.1: Simulation parameters

Parameter	Value
Chip Rate	20 MHz
Spreading Gain	62.5
Basic rate	32 kbps
PrS guaranteed data rates (packets/frame)	1,2,3,4,5,6,7,8
AS guaranteed data rates (packets/frame)	1,2,3,4,5,6,7,8
BER Classes	10^{-3}, 10^{-6}
Frame duration	20 ms
Packet size	640 bits
Average call holding time	3 min
Average BE message length	160ms
Data time slots/uplink frame	4
Code channels/time slot	20
Capacity (code channels)	80
Premium service population (N_{PS})	100
Assured service population (N_{AS})	100
Best effort service population (N_{BE})	∞

3.5.2 PrS Results

The performance measures of interest for the premium service subsystem are the average number of admitted premium service calls and the average premium service call blocking probability for a given offered load and the average premium service call blocking probability for a given maximum guaranteed PrS capacity and offered load. In Fig. 3.5, the plot for number of premium service calls versus the offered call load is shown. The level of PrS traffic is low for low call arrival rates as one would expect. As the offered traffic load increases the traffic level increases as well. It can be seen from the curves that when high capacity is reserved for PrS, more calls will be admitted. Since the terminals are allowed to transmit using more than one code-channel, fewer calls are admitted than it would be when one code channel was allowed per user. As the traffic load is increased further, the number of calls is expected to approach the system capacity. However, this means that only the users with smaller data rates will be admitted to the system. When the users are transmitting at an equal average data rate (or using equal number of code-channel), as it will be shown in chapter 4, the system capacity limit would have been reached at the call arrival rate equal to 0.009.

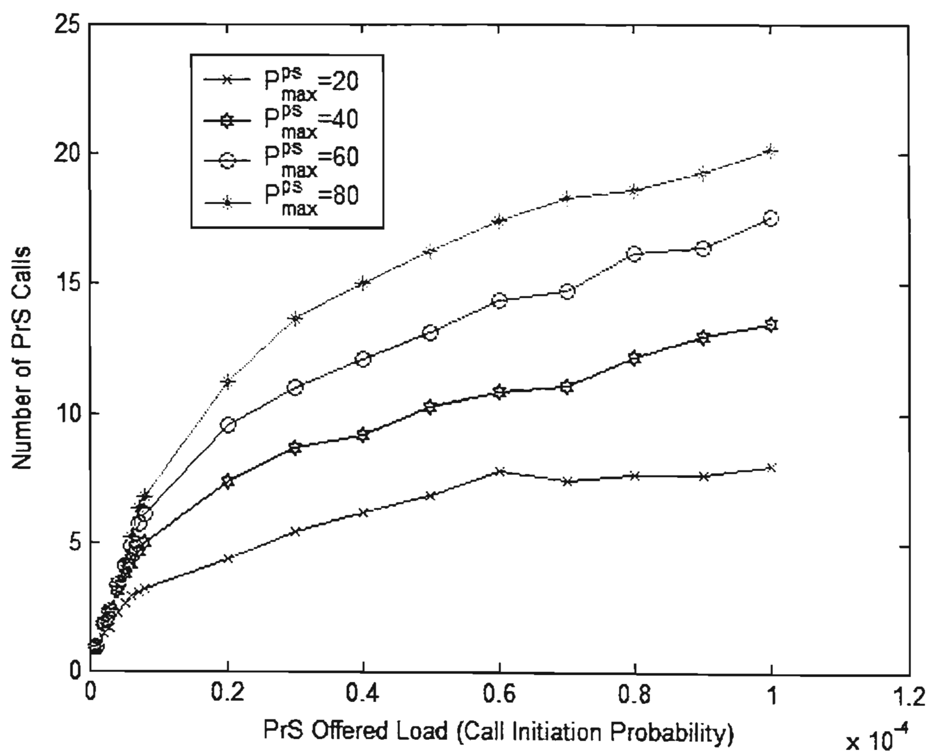


Fig. 3.5: Number of premium service calls versus the offered call load for various guaranteed Premium service capacities

The premium service call blocking probability due to the CAC procedure versus the offered PrS load is plotted in Fig. 3.6. As the call arrival rate increases, the expected call blocking probability increases. To maintain the admitted calls below the allowable channel limit, the blocking probability will be high for high traffic loads. From the graphs it is also be noticed that when the maximum capacity available for premium service is minimized, the blocking probability increases.

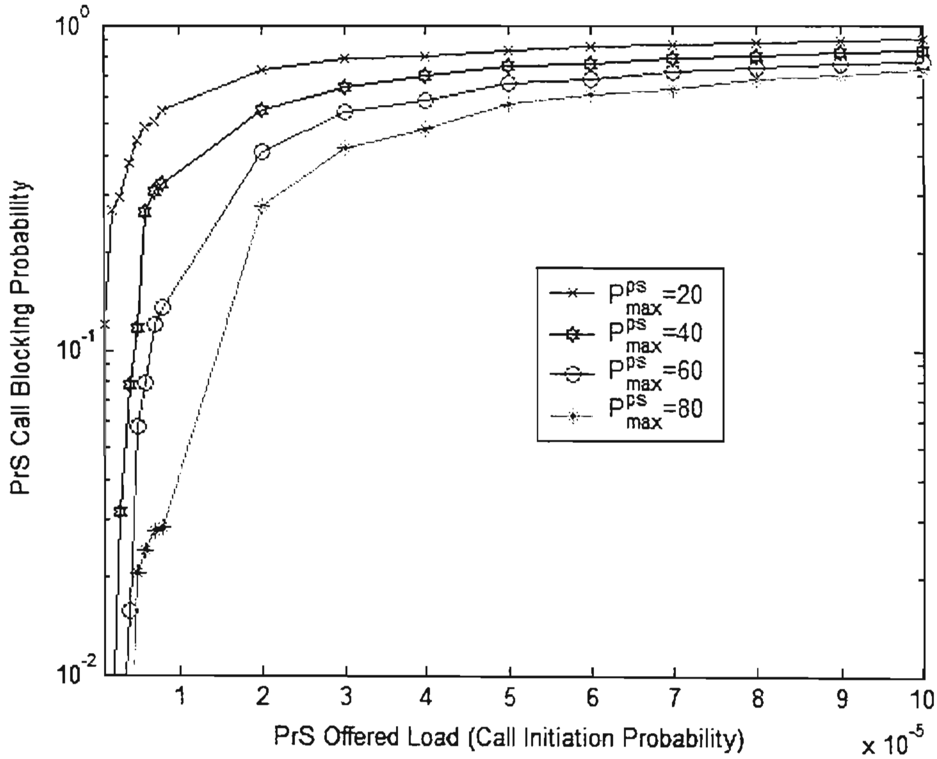


Fig. 3.6: Premium service call blocking probability versus the offered call load

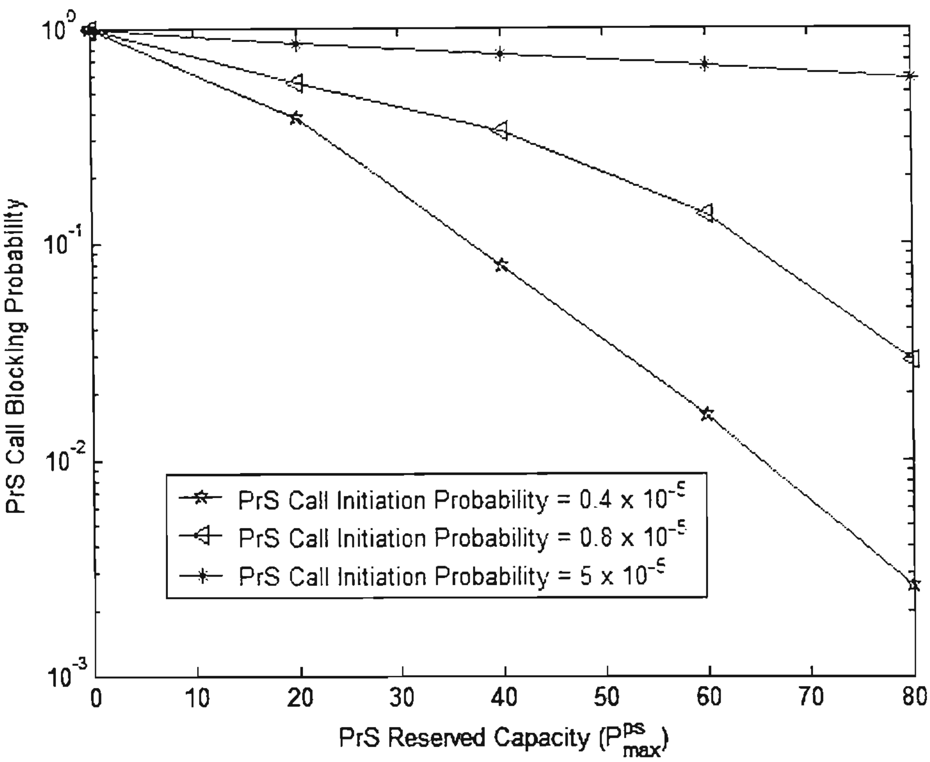


Fig. 3.7: Premium service call blocking probability versus the maximum reserved PrS capacity

Fig. 3.7 illustrates the premium service call blocking probability versus the guaranteed capacity for PrS (i.e. P^{ps}_{max}) considering three call arrival rates. As the maximum guaranteed capacity increases, more calls are admitted and therefore the call blocking probability is low. For low PrS call arrival rate the blocking probability is low as well.

3.5.3 AS Results

The performance measures for assured service are similar to those of premium service. Since AS partially depends on PrS, the number of AS calls and the AS call blocking probability are determined at a given value for PrS arrival rate. The guaranteed minimum capacity reserved for AS service considered is $A^{as}_{min} = 40$. The same results will be obtained if any value of AS guaranteed minimum capacity was used, because when more capacity is allocated to PrS, less capacity is available to be temporarily used by AS and visa versa.

Fig. 3.8 shows the number of admitted AS calls versus the offered AS load for various PrS arrival rates. The number of admitted AS calls increases as the offered AS load increases. At very high offered load, the number of admitted AS calls are expected to approach the system capacity and only users with very low data rate should be admitted. For lower premium service loads, more capacity is available for assured service; therefore there will be a higher number of admitted AS calls. This is also reflected in Fig. 3.9 which illustrates the expected call blocking probability for AS shown for various PrS arrival rates. For higher traffic loads, the probability is high in order to limit the number of users that get admitted. As can be seen, the blocking probability is low for low PrS arrival rates, since more capacity is available for assured service.

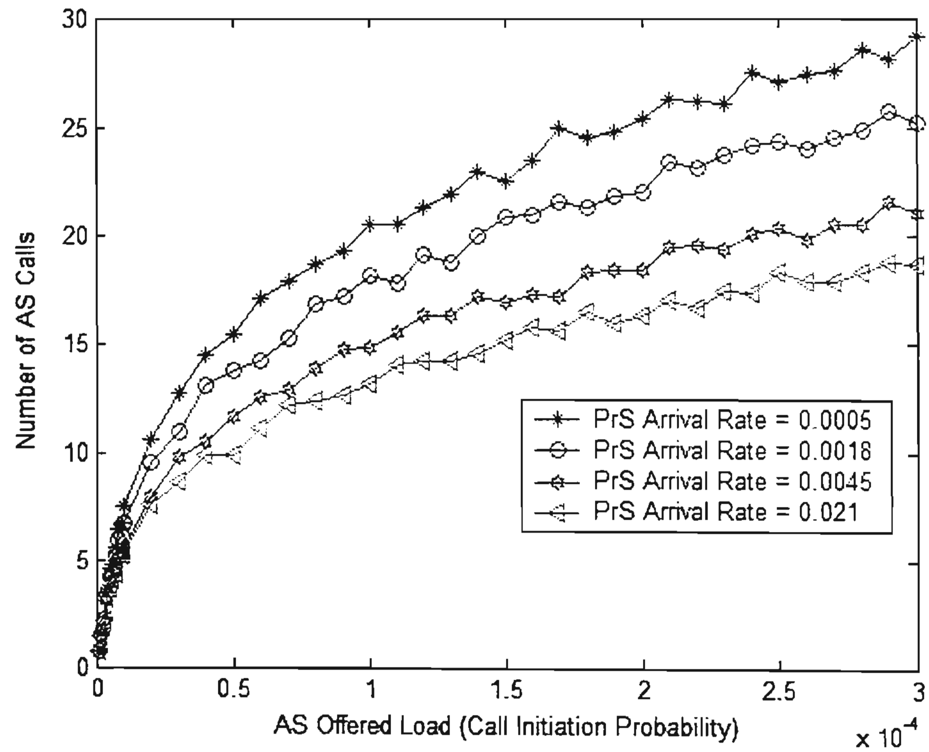


Fig. 3.8: Number of assured service calls versus the Assured service offered call load for various Premium service call arrival rates

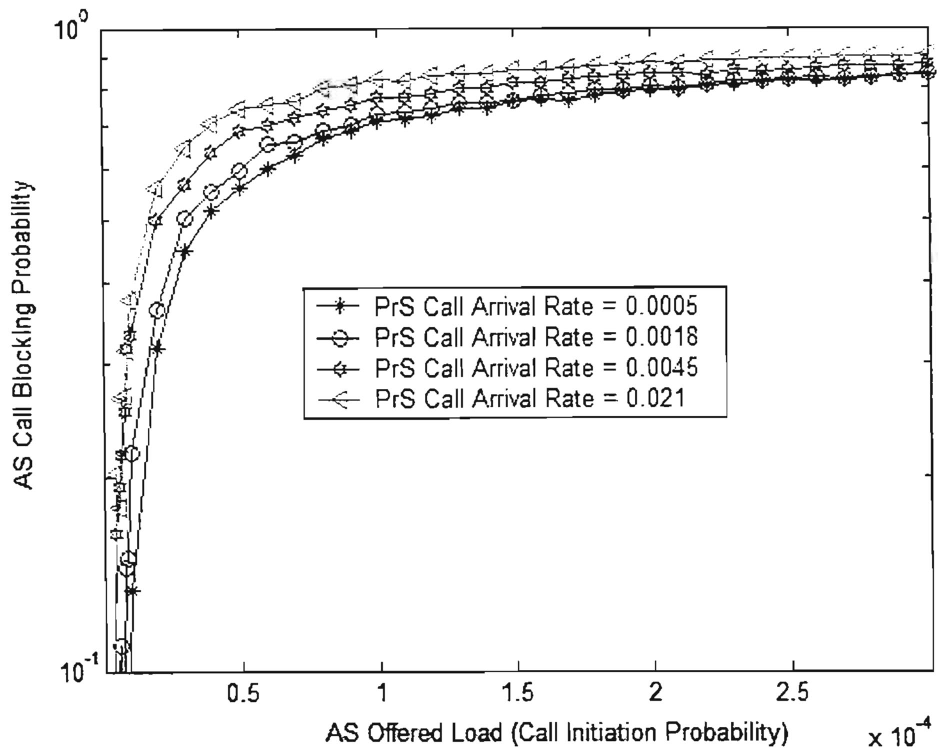


Fig. 3.9: Assured service call blocking probability versus the Assured service offered call load for various Premium service call arrival rates

3.5.4 BE Results

The BE performance is determined under various combined PrS and AS average call arrival rates. For each PrS and AS call arrival rate, the BE throughput and BE message blocking probability is determined. The PrS and AS call arrival rate determines the average capacity available for BE traffic. The blocking threshold α is set to be equal to the maximum slot threshold β.

Fig 3.10 illustrates the BE message blocking probability as a function of the offered BE load for four PrS and AS call arrival rates. As expected, the message blocking probability increases as the offered BE load increases. The dependence of BE on the remaining PrS and AS capacity is clearly illustrated. As the PrS and AS call arrival rates increases, fewer messages can be admitted due to small remaining capacity. This results in high message blocking probabilities.

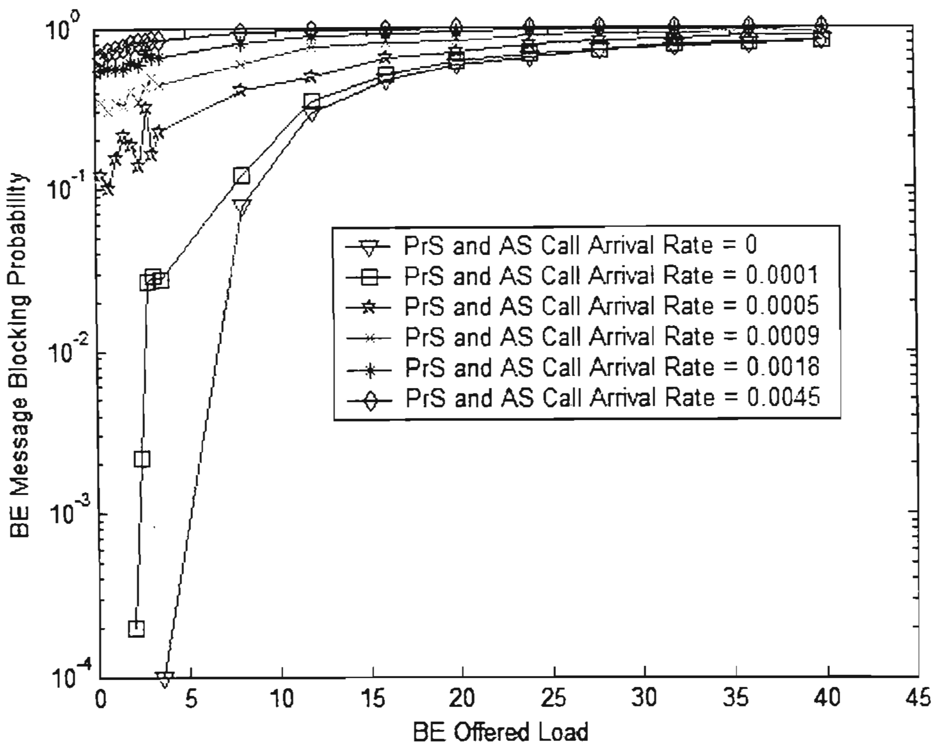


Fig. 3.10: Best effort services message blocking probability versus the Best effort service offered load for various Premium and Assured service arrival rates

In Fig. 3.11, the average BE throughput is shown versus the offered BE load for four PrS and AS call arrival rates. The throughput increases as the offered load increases. However, as the offered load approaches infinity, the throughput approaches zero. It is expected that the performance resembles that of SS/ALOHA, since the MAC protocol uses a random access scheme derived from SS/ALOHA. When the call arrival rate is zero, no PrS and AS calls are admitted, therefore the whole capacity is available for BE and the throughput is high. For small PrS and AS call arrival rates, a large portion of the capacity is still available for BE, thus the throughput is still high. As the call arrival rates for PrS and AS increase, the capacity available for BE decrease as a result the throughput decreases as well. When the offered load is high and the available capacity is small, the chances of the MAI threshold being exceeded are high. Therefore a lot of packets will be corrupted by the MAI and the throughput decreases as the graphs shows.

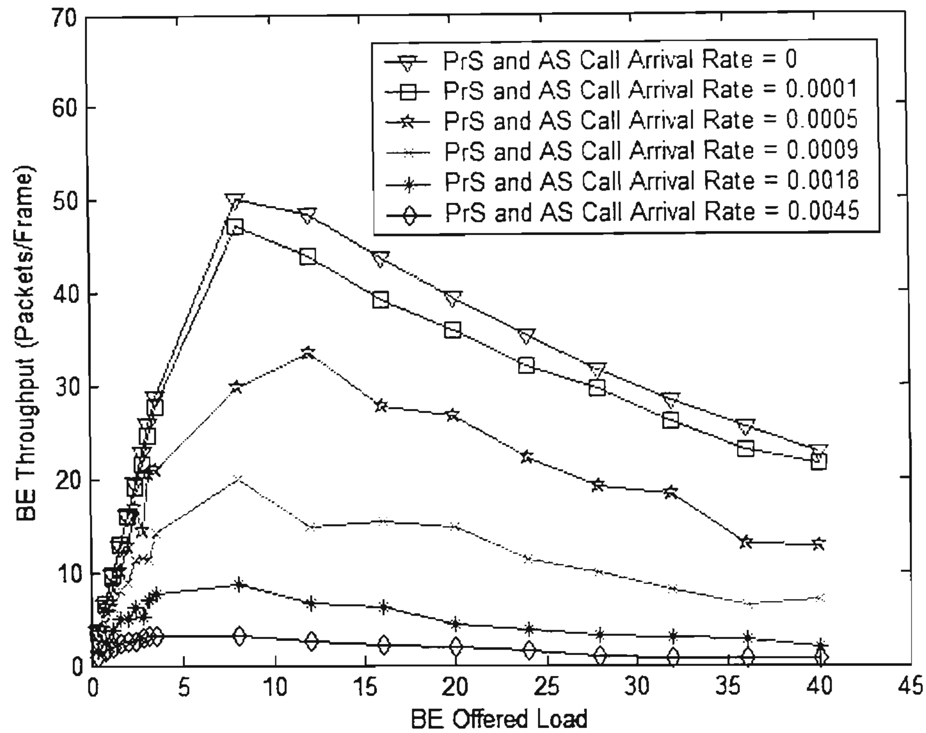


Fig. 3.11: Best effort service throughput versus the Best effort service offered load for various Premium and Assured service rates

3.6 Summary

This chapter described our proposed MAC protocol designed for DiffServ IP-based wireless networks utilizing a CDMA air interface. The protocol uses a DA scheme to deal with PrS and AS traffic. By implementing the DA scheme PrS and AS users are able to communicate their QoS requirements to the BS. The BS will ensure that once admitted the QoS requirements of the users are not violated. To provide efficient bandwidth utilization, the remaining capacity is used to serve BE traffic using a RRA scheme for BE traffic.

The performance of the protocol was determined through simulations in terms of the number of admitted PrS (or AS) calls, call (or message) blocking probability, and BE throughput. The results showed how AS and BE traffic depend on PrS and AS and PrS offered loads, respectively. It was observed that as the offered load increases, the number of AS calls and BE throughput decreases because of the reduction of the remaining capacity which causes aggressive blocking of new calls (or messages).

Chapter 4

Performance Analysis of the WDIP/CDMA MAC protocol

4.1 Introduction

In this Chapter we consider the analysis of the proposed WDIP/CDMA MAC protocol using discrete-time Markov chains, also called Markov analysis. Markov analyses are the commonly used analytical models for MAC protocols (e.g. [5], [7], [21], [22], [23], [26] and [27]). They permit a user in a communication system to occupy discrete states and permit transitions between these states to take place only at discrete times. The stationary distributions of users in various states are used to derive statistics of interest such as message throughput and delay. The analytical models for the proposed protocol are derived from [5] and [27].

In Chapter 3 it was indicated that PrS and AS mobile terminals have different guaranteed data rate requirements which were considered in the simulations. The inclusion of the data rates in the Markov analysis results in a complicated system with too many states. Except [7], no other work was found in literature that considers users with different data rate requirements. Most analyses presented in literature are mainly for voice/data MAC protocols. A voice or data terminal uses one channel for transmission; hence the analyses are simplified since the states are simply specified in terms of the number of terminals (e.g. idle or transmitting). Although the rates of MTs were considered in [7], it was finally assumed that MTs transmit at their average rate in order to simplify the analysis. Due to the difficulties encountered when trying to incorporate the data rate information in the analysis, the assumption that users transmit at their average data rate had been made. The state diagrams for individual PrS, AS, and BE terminals described in Chapter 3 are used for the Markov analysis.

In section 4.2 the analysis of the PrS MAC scheme is presented. This is followed by section 4.3 where the analysis of AS which partially depends on PrS is presented. The analysis of BE is presented in section 4.4.

4.2 PrS Analysis

This PrS analysis is based on [27]. Since the PrS MAC scheme is completely independent of the AS and BE traffic, the analysis is greatly simplified. The PrS subsystem can be fully described by two state variables $\{R_{PS}, S_{PS}\}$, namely the number of PrS terminals in reservation state (i.e. transmission state) and the number of PrS terminals in silence states. Since the number of MTs is finite, the system can be described by the number of terminals in reservation state and the number of terminals in silence state is then given by $S_{PS} = N_{PS} - R_{PS}$. The evolution of the system is modelled as one dimensional (1-D) discrete-time Markov chain with the embedded Markov points at the beginning of the frame. We assume that the stationary distribution of the system exists, and is denoted as:

$$\bar{\pi} = \pi(r_{ps}) = \left\{\Pr(R_{ps} = r_{ps})\right\} \tag{4.1}$$

In order to find the stationary distribution, the transition probability matrix P_{ij}^{S} from reservation state $R_{PS} = i$ in frame t to $R_{PS} = j$ in frame $t+1$ is required. For the state to be j, the following events must occur at beginning of frame $t+1$.

- Among the i terminals in reservation state, k terminals leave the reservation state and enter the silence state, where k is a random variable within the range $i - j \le k \le i$.

- Among the $N_{PS} - i$ PrS terminals in silence state, $j-(i-k)$ terminals leave the silence state to enter the reservation state by generating new premium service calls. New calls are admitted by a CAC scheme to the reservation state if capacity is available.

The probabilities for these events are given by $B(i,k,\eta_R)$, $B(N_{PS}-i, j-(i-k), \sigma_C)$, where $B(m,n,p)$ indicates the Binomial distribution:

$$B(m,n,p) = \binom{m}{n} p^{n} (1-p)^{m-n} \tag{4.2}$$

Let P_{ij}^{R} be the steady transition probability of moving from state i to state j given that the capacity threshold is $PS_{\max}$. This matrix can be shown to be:

$$P_{ij}^{R}=\left\{P\left(R_{SP}(t+1)=j\mid R_{SP}(t)=i\right)\right\}$$
$$=\begin{cases}\sum\limits_{k=\max(0,i-j)}^{i} B\left(N_{PS}-i,j-(i-k),\sigma_C\right)B\left(i,k,\eta_R\right) & for\ j<PS_{\max}\\ \sum\limits_{j_1=PS_{\max}}^{N_{PS}}\sum\limits_{k=0}^{i} B\left(N_{PS}-i,j_1-(i-k),\sigma_C\right)B\left(i,k,\eta_R\right) & for\ i\le j=PS_{\max}\end{cases} \tag{4.3}$$

where $PS_{\max}$ is the maximum number of users that can be admitted to the system given that they all transmit at an average data rate. Hence, the following criteria must be satisfied for the CAC:

$$\Pr\left\{\sum_{j=1}^{PS_{\max}^{j}}\overline{P_j}>P_{\max}^{ps}\right\}=0 \tag{4.6}$$

where $\overline{P_j}$ is the average number of code-channels (corresponding to the average data rate) reserved for each admitted user and $P_{\max}^{ps}$ is the total number of code channel capacity reserved for PrS.

The transition probability size is given by $(PS_{\max}+1)\times(PS_{\max}+1)$. Given the transition probability, the stationary distribution for the reservation state is obtained by solving for

$$\sum\pi(r_{sp})=1 \text{ and } \pi(r_{sp})=\pi(r_{sp}).P_{ij}^{R} \tag{4.7}$$

Once the steady stationary distribution has been obtained, we can calculate the expected number of terminals in reservation state per frame, i.e., the number of PrS calls, and the blocking probability. The number of PrS calls in the system is defined as

$$\text{Number of PrS Calls}=\sum_{r_{ps}=0}^{PS_{\max}} r_{ps}.\pi(r_{ps}) \tag{4.8}$$

The blocking probability is defined as follows:

$$P_{PS}=\frac{E[B_{R_t}]}{E[R_t]} \tag{4.9}$$

where $E[B_{R_t}]$ is the expected number of blocked premium service calls in a frame, and $E[R_t]$ is the average number of premium service calls arriving in a frame. $E[R_t]$ is determined as follows:

$$E[R_t] = \sum_{r_{ps}=0}^{PS_{max}} \sum_{z=0}^{N_{PS}-r_{ps}} z.B\left(N_{PS} - r_{ps}, z, \sigma_C\right).\pi(r_{ps}) \tag{4.10}$$

The expected number of blocked calls $E[B_{R_t}]$ is obtained as follows:

$$E[B_{R_t}] = \sum_{r_{ps}=0}^{PS_{max}} \sum_{b=1}^{N_{PS}-PS_{max}} \sum_{k=0}^{r_{ps}} b.B(N_{PS} - r_{ps}, PS_{\max} - r_{ps} + k + b, \sigma_C) \\ .B(r_{ps}, k, \eta_R).\pi(r_{ps}) \tag{4.11}$$

4.3 AS analysis

The AS analysis is dependent on PrS, since AS can utilize the unused PrS capacity in addition to its minimum reserved capacity. This analysis follows from the PrS analysis and [27]. The difference between this analysis and the abovementioned analyses is apparent from the AS state diagram, i.e., the consideration of the temporary and permanent state. The AS subsystem can be fully described by three state variables $\{T_{AS}, P_{AS}, S_{AS}\}$, namely the number of temporary admitted AS terminals, number of permanently admitted terminals, and the number of assured terminals in silence state. The number of MTs in the system is finite, therefore the number of terminals in silence state will be $S_{AS} = N_{AS} - T_{AS} - P_{AS}$ and the total number of transmitting AS terminals will be $AS = T_{AS} + P_{AS}$. The stationary distribution for AS subsystem states is denoted as:

$$\pi(t_{as}, p_{as}, r_{ps}) = \left\{P(T_{AS} = t_{as}, P_{AS} = p_{as}, R_{PS} = r_{ps})\right\} \tag{4.12}$$

Since the number of transmitting PrS terminals in reservation state is independent of the AS processes, the stationary distribution of system states is simplified as follows:

$$\pi(t_{as}, p_{as}, r_{ps}) = \pi(t_{as}, p_{as} \mid r_{ps}).\pi(r_{ps}) \tag{4.13}$$

The stationary distribution $\pi(t_{as}, p_{as}, | r_{ps})$ is obtained by evaluating a transition probability matrix denoted by $P_{nm|ij}^{AS} = \left\{P\left(t_{as}^{t+1} = n, p_{as}^{t+1} = m \mid t_{as}^{t} = i, p_{as}^{t} = j_{ps}\right)\right\}$, conditioned on the number of PrS users $R_{PS} = r_{ps}$ in reservation state. The variables (t_{as}^{t}, p_{as}^{t}) denote the number of terminals in temporary and permanent states in the current frame, and $(t_{as}^{t+1}, p_{as}^{t+1})$ in the next frame. The transition probability matrix for the AS call process is determined as follows:

Firstly, consider $m < AS_{\max}$, when all AS terminals from the temporary state and new calls are admitted, and consequently $n = 0$. $AS_{\max}$ is the guaranteed maximum number of AS calls that is allowed to satisfy the following CAC criteria:

$$\Pr\left\{\sum_{j=1}^{AS_{\max}} \overline{A_j} > A_{\max}^{as}\right\} = 0 \tag{4.14}$$

where $\overline{A_j}$ is the average guaranteed code-channels reserved for each admitted user and $A_{\max}^{as}$ is the maximum guaranteed code-channel capacity reserved for AS. Then,

$$P_{nm|ij}^{AS}(r_{ps}) = \sum_{l=0}^{i} \sum_{k=0}^{\min(j, N_{AS}-m)} B(N_{AS} - i - j, m + k - (i + l), \sigma_{ca}) .B(j, k, \eta_{Ra}).\Phi(i, l) \tag{4.15}$$

where k is a dummy variable representing the number of calls terminating in frame $t+1$, and $\Phi(i,l)$ is the probability that among the i terminals in temporary state at frame t, l terminals are converted from temporary state to permanent state in frame $t+1$. In this case $l = i$ and $\Phi(i,l) = 1$. It follows that $P_{nm|ij}^{AS}(r_{ps})$ can simply be obtained as:

$$P_{nm|ij}^{AS}(r_{ps}) = \sum_{k=0}^{\min(j, N_{AS}-m)} B(N_{AS} - i - j, m + k - (i + j), \sigma_{ca}).B(j, k, \eta_{Ra}) \tag{4.16}$$

For the case, $j \le m = AS_{\max}$ and $n < PS_{\max} - r_{ps}$, some of the AS calls in temporary state during frame t are converted to permanent state and all new calls are admitted to temporary state. Considering the permanent state, the total number of call arrivals from silence and temporary states in frame $t+1$ is $AS_{\max} - j + k$. Let l denotes the number of calls from the temporary state which are converted to permanent state, then the number of call arrivals from silence state to permanent state is given by

$$a_p = AS_{\max} + k - j - l \tag{4.17}$$

Since the temporary admitted users have priority over new calls, the maximum number of calls from temporary state converted to permanent state is $l = AS_{\max} + k - j$. If $i > l$, only l out of i users in temporary state at frame t are converted to permanent state. For $i \le l$ all terminals in temporary state are converted to permanent state. Given l, the number of call arrivals from the silence to the temporary state is

$$a_t = n + l - i \tag{4.18}$$

The total number of call arrivals from the silence state is

$$a_p + a_t = n + AS_{\max} + k - (i + j) \tag{4.19}$$

Then,

$$P^{as}_{nm|ij}(r_{ps}) = \sum_{l=0}^{i} \sum_{k=0}^{\min(j, N_{AS}-(n-AS_{\max}))} B(N_{AS} - i - j, n + AS_{\max} + k - (i + j), \sigma_{ca}) .B(j, k, \eta_{Ra}).\Phi(i,l) \tag{4.20}$$

Since l is deterministic,

$$\Phi(i,l) = \begin{cases} 1 & \text{if } l = AS_{\max} + k - j \\ 0 & \text{Otherwize} \end{cases} \tag{4.21}$$

Therefore, for all legal values of l

$$P^{as}_{nm|ij}(r_{ps}) = \sum_{k=0}^{\min(j, N_{AS}-(n-AS_{\max}))} B(j, k, \eta_{Ra}) .B(N_{AS} - i - j, n + AS_{\max} + k - (i + j), \sigma_{ca}) \tag{4.22}$$

For $j \le m = AS_{\max}$ and $n = PS_{\max} - r_{ps}$, some of the new calls are blocked. Thus,

$$P^{as}_{nm|ij}(r_{ps}) = \sum_{n_1 = PS_{max} + AS_{max} - r_{ps}}^{N_{AS}} \sum_{k=0}^{\min(j, N_{AS} - (n - AS_{max}))} B(j,k,\eta_{Ra}) \\ .B(N_{AS} - i - j, n_1 + k - (i + j), \sigma_{cu}) \tag{4.23}$$

Due to the constraint that the reserved capacity for AS should be fully occupied for terminals to be temporary admitted on PrS capacity, the following cases for the transition probability will be illegal:

For $m < AS_{max}, n > 0$ or $j < AS_{max}, i > 0$

$$P^{as}_{nm|ij}(r_{ps}) = 0 \tag{4.24}$$

Removing the condition on r_{ps}, the average transition probability $P^{as}_{nm|ij}$ is obtained as follows

$$P^{as}_{nm|ij} = \sum_{r_{ps}=0}^{PS_{max}} P^{as}_{nm|ij}(r_{ps}).\pi(r_{ps}) \tag{4.25}$$

Given the transition probability, the stationary distribution of the temporary state and permanent state is obtained by solving for:

$$\sum \pi(t_{as}, p_{as}, r_{ps}) = 1 \text{ and } \pi(t_{as}, p_{as}, r_{ps}) = \pi(t_{as}, p_{as}, r_{ps}).P^{as}_{nm|ij} \tag{4.26}$$

Using the stationary distribution the following performance metrics are derived:

$$\text{Number of AS Calls} = \sum_{r_{ps}=0}^{PS_{max}} \sum_{t_{as}=0}^{PS_{max} - r_{ps}} \sum_{p_{as}=0}^{AS_{max}} (t_{as} + p_{as}).\pi(t_{as}, p_{as}, r_{ps}) \tag{4.27}$$

and the AS call blocking probability defined as follows:

$$P_{AS} = \frac{E[B_{AS}]}{E[R_{AS}]} \tag{4.28}$$

where $E[B_{AS}]$ is the expected number of blocked assured service calls in a frame, and $E[R_{AS}]$ is the average number of assured service calls arriving in a frame. Computation results showed that when the number of transmitting AS users are separated into temporary and permanent state the blocking probability is inaccurate. For this reason when determining, the call blocking probability, the temporary state and permanent state are combined into a single state for the number of transmitting AS. This is justifiable since new AS calls will only be blocked when both the reserved AS and PrS capacities are fully occupied. $E[R_{AS}]$ is determined as follows:

$$E[R_{AS}]=\sum_{r_{ps}=0}^{PS_{\max}}\sum_{w=0}^{PS_{\max}+AS_{\max}-r_{ps}}\sum_{x=0}^{N_{AS}-w}x.B\left(N_{AS}-w,x,\sigma_{ca}\right).\pi(w,r_{ps}) \tag{4.29}$$

where w is the combined state for permanent and temporary admitted AS users, i.e., the total number of transmitting AS users. $E[B_{AS}]$ is obtained as follows:

$$\begin{aligned}E[B_{AS}]=&\sum_{r_{ps}=0}^{PS_{\max}}\sum_{w=0}^{PS_{\max}+AS_{\max}-r_{ps}}\sum_{b=1}^{N_{AS}-(AS_{\max}+PS_{\max}-r_{ps})}\sum_{k=0}^{w}b\\&.B(N_{AS}-w,AS_{\max}+PS_{\max}-(r_{ps}+w)+k+b,\sigma_{ca})\\&.B(w,k,\eta_{Ra}).\pi(w,r_{ps})\end{aligned} \tag{4.30}$$

4.4 BE analysis

The Markov analysis of BE service follows the work performed by Judge [5] due to similar operation of the best effort and data admission schemes. Judge's analysis is applicable to a single slot frame. In this analysis his work is extended to a multiple slots frame structure. We consider Poisson traffic model because of its simplicity. The analysis based on a finite model was considered in [5], however if extended to the frame model, the complexity of the analysis increases due the unknown expected number of users per frame, unknown message length of retransmitted message, and unknown backlogged terminals. In the Poisson model these unknown parameters are absorbed into the Poisson arrival process. Because of this reason we can analyse each slot in the frame independently. The frame throughput will simply be obtained by summing the throughputs of the all the time slots.

4.4.1 BE system stationary distribution

To solve for the stationary distribution of BE system per time slot per frame, we first have to determine the one-step transition probability of the system. Let Y and $\prod_{ij}$ denote the system state and the stationary state transition probability in time slot s from state i in frame t to state j in frame $t+1$, respectively. Three cases are considered for determination of $\prod_{ij}$. The first case is when $i \leq \alpha$ and $j > i$, in which case all the new or retransmitted messages will be admitted. Given that there are i transmitting BE terminals at the end of frame t, the following two events must occur to be in state j at the end of frame $t+1$.

- Among the i terminals in transmitting state, k terminals terminate transmission with probability $B(i,k,\mu)$, where k is a random variable within the range $i-j \le k \le i$.
- Among the $N_{PS}-i$ BE terminals in backlog and silence states, $j-(i-k)$ terminals transit to the transmitting state. The probability of this event is given by

$$f_{\infty}(m) = \frac{e^{-\lambda.t_s}\left(\lambda.t_s\right)^m}{m!} \tag{4.31}$$

The second case is when $i \le \alpha$ and $j \le i$, in which there will be blocking of some messages. This case can result from the occurrence of events in case 1 or the event that $i-j$ terminals terminate normally with probability μ and that a large number of new terminals arrived exceeding the system capacity. The probability of the latter is given by $\{B(i,i-j,\mu).\Pr(m > \beta - j)\}$. Due to propagation delay, the new arrivals causing collisions in some of the time slots will be dropped in the next frame. Since the colliding users will be dropped immediately, the effect of collisions can be included in the blocking process to simplify the analysis. Therefore, the blocking process implies the blocking of new arrivals (messages) when no code-channels are available as well as the messages dropped due to excess MAI.

The third case is when $i \ge \alpha$ and $j \le i$, in which terminals are inhibited from transmitting in the specified time slot. The only possible event in this case is that of departing terminals after successful transmission of their messages. The probability of this event is given by $B(i,i-j,\mu)$.

Considering the above events, the one-state transition probability can be obtained as follows:

$$
\Pi_{ij} = \begin{cases} \sum\limits_{k=\max(0,i-j)}^{i} f_{\infty}(j-i+k).B(i,k,\mu), & i \le \alpha, j > i \\ \sum\limits_{k=\max(0,i-j)}^{i} f_{\infty}(j-i+k).B(i,k,\mu) \\ \qquad + B(i,i-j,\mu).\sum\limits_{m=\beta-j+1}^{\infty} f_{\infty}(m), & i \le \alpha, j \le i \\ B(i,i-j,\mu), & i > \alpha, j \le i \end{cases} \tag{4.32}
$$

The probability of the number of terminals in transmitting state is obtained by solving the simultaneous equations,

$$
\Pr(Y=j) = \sum_{i=0}^{\beta} \Pr(Y=i).\Pi_{ij} \text{ and } \sum_{i=0}^{\beta} \Pr(Y=i) = 1 \tag{4.33}
$$

The average number of transmitting users per slot can now be obtained and is given by:

$$
\bar{y} = \sum_{k=0}^{\beta} k.Y(k) \tag{4.34}
$$

4.4.2 Message blocking probability

The message blocking probability is determined by considering the probability of a reference message being blocked given that it sees $x = j$ transmission in its arrival slot. This message blocking probability is given by

$$
P_B(j) = \begin{cases} \sum\limits_{i=\alpha+1}^{\beta} \dfrac{\Pr(X=i).\Pi_{ij}}{\Pr(X=j)}, & j > \alpha \\ \sum\limits_{i=\alpha+1}^{\beta} \dfrac{\Pr(X=i).\Pi_{ij}}{\Pr(X=j)} \\ + \sum\limits_{i=j}^{\alpha} \left(\dfrac{\Pr(X=i).B(i,i-j,\mu)}{\Pr(X=j)} . \sum\limits_{m=l-j+1}^{\infty} f_{\infty}(m) \right), & j \le \alpha \end{cases} \tag{4.35}
$$

The average expected message blocking probability is then given by

$$P_B = \sum_{j=0}^{\beta} P_B(j) \tag{4.36}$$

4.4.3 Packet's success probability considering MAI

The successful transmission of a packet in any time slot is conditioned on the fact that the MAI is insufficient to cause bit errors in the packets. The effect of packets transmitted by AS and PrS terminals should be taken into consideration since they interfere with the reference packet. Given the variables r_{ps} and t_{as} as the number of transmitting PrS and AS terminals in a frame, respectively. The average number of packets transmitted per frame can be evaluated based on the number of packets that each admitted user transmits in a frame. Let k_λ be the number of AS and PrS terminals transmitting at data rate $\lambda \in \{1,2,...,R_{\max}\}$ per frame such that $r_{ps} + t_{as} = \sum_{\lambda=1}^{R_{\max}} k_\lambda$. Then, the cumulative probability distribution for number of packets transmitted by AS and PrS terminals per frame is given by

$$\Pr(\theta_{pk}^{f} = k_1 + 2k_2 + ... + R_{\max}.k_{R_{\max}} \mid r_{ps}, t_{as}) = \sum_{\forall k_1, k_2 ... k_{R_{\max}}} \frac{(r_{ps}+t_{as})!}{k_1!k_2!k_3!....k_{R_{\max}}!} p_1^{k_1} p_2^{k_2} p_3^{k_3} ....p_{R_{\max}}^{k_{R_{\max}}} \tag{4.37}$$

where $p_\lambda = \{p_1, p_2, ..., p_{R_{\max}}\}$ is the probability that a PrS or AS terminal takes on λ for the current frame. We assume that packets transmission for VBR AS and PrS terminals is modelled by an autoregressive function [7]. The average number of packets transmitted by PrS and AS terminals per frame is given by

$$\overline{\theta_{pk}^{f}} = \sum_{j=1}^{(r_{ps}+t_{as}).R_{\max}} j.\Pr(\theta_{pk}^{f} = j \mid r_{ps}, t_{as}) \tag{4.38}$$

In order to determine the average number of packets transmitted per slot, an assumption is made that packets are transmitted in any time slot of the frame with equal probability. The average number of packets transmitted by AS and PrS terminals per time slot is given by

$$\overline{l_{as}^{ps}} = \sum_{x=0}^{\overline{\theta_{pk}^{f}}} x.B(\overline{\theta_{pk}^{f}}, x, \varepsilon_s) \tag{4.39}$$

where $\varepsilon_s = \dfrac{1}{N_S}$ is the probability of choosing to transmit in any time slot of the frame.

Now that the numbers of packets transmitted per slot is known, we can define the packet success probability as follows:

$$P_{pk}^{succ}(\overline{l_{as}^{ps}} + j + 1) = [1 - P_{Err}^{bit}(\overline{l_{as}^{ps}} + j + 1)] \tag{4.40}$$

where $P_{Err}^{bit}(\overline{l_{as}^{ps}} + j + 1)$ is the probability of bit error during the $\overline{l_{as}^{ps}} + j + 1$ spread spectrum transmission. The bit error probability, modelled as in [89] is denoted as

$$P_{Err}^{bit}(\overline{l_{as}^{ps}} + j + 1) = \frac{1}{2} erfc \sqrt{\frac{3.G}{2\left((\overline{l_{as}^{ps}} + j + 1) - 1\right)}} \tag{4.41}$$

4.4.4 Message success probability

The success probability of the first packet of a message is computed as follows:

$$R_1(j) = X(j).[1 - P_B(j)].P_{pk}^{succ}(\overline{l_{as}^{ps}} + j + 1) \tag{4.42}$$

The success probability of the entire message is obtained by solving the following equation recursively

$$R_n(j) = \sum_{i=0}^{\beta-1} R_{n-1}(i).\prod_{ij}^{ref}, \qquad n > 1 \tag{4.43}$$

where $\prod_{ij}^{ref}$ denotes the transition probability matrix as observed by the reference packet. The transition probability matrix is as follows:

$$\Pi_{ij}^{ref} = \begin{cases} \sum_{k=\max(0,i-j)}^{i} f_{\infty}(j-i+k).B(i,k,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j+1) & i \le \alpha, j > i \\ \sum_{k=\max(0,i-j)}^{i} f_{\infty}(j-i+k).B(i,k,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j+1) & \\ +B(i,i-j,\mu).\sum_{m=\beta-j+1}^{\infty} f_{\infty}(m).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j+m+1), & i \le \alpha, j \le i \\ B(i,i-j,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j+1), & i > \alpha, j \le i \end{cases} \tag{4.44}$$

The success probability of the reference packet is incorporated into Π_{ij}^{ref} to model the blocking of the reference packets due to MAI.

The probability of success of a message containing L packets is defined by

$$R_L = \sum_{j=0}^{\beta-1} R_L(j) \tag{4.45}$$

The total average message success probability considering all possible message lengths is given by

$$P_{SC} = \sum_{l=1}^{\infty} R_l.L(l) \tag{4.46}$$

4.4.5 Message and Packet Throughput

The message throughput per slot conditioned on $\overline{l_{as}^{ps}}$ is given as follows

$$S_M^s\left(\overline{l_{as}^{ps}}\right) = G.P_{SC} \tag{4.47}$$

where G is the average offered load per slot from the silent state and the backlog state denoted by $G = \lambda.t_s$.

The offered load per frame is given by

$$G_f = \sum_{s=1}^{N_S} \lambda . t_S \tag{4.48}$$

The expected message throughput per frame is given by

$$S_M^{Fr} = \sum_{s=1}^{N_S} S_M^s (\overline{l_{as}^{ps}}) \tag{4.49}$$

The expected packet throughput per slot is given by

$$S_P^s \left(\overline{l_{as}^{ps}}\right) = \sum_{j=0}^{\beta} \sum_{i=0}^{\beta} j.X(i).\Pi_{ij}^S \tag{4.50}$$

where Π_{ij}^S represents the state transition probability matrix as seen by the network (base station) and is computed as follows:

$$\Pi_{ij}^S = \begin{cases} \sum_{k=\max(0,i-j)}^{i} f_\infty(j-i+k).B(i,k,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j), & i \le \alpha, j > i \\ \sum_{k=\max(0,i-j)}^{i} f_\infty(j-i+k).B(i,k,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j) & \\ +B(i,i-j,\mu). \sum_{m=\beta-j+1}^{\infty} f_\infty(m).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j+m), & i \le \alpha, j \le i \\ B(i,i-j,\mu).P_{pk}^{succ}(\overline{l_{as}^{ps}}+j), & i > \alpha, j \le i \end{cases} \tag{4.51}$$

The expected packet throughput per frame is then given by

$$S_P^s = \sum_{s=1}^{N_S} S_P^s (\overline{l_{as}^{ps}}) \tag{4.52}$$

4.5 Analytical Results

In this section we provide the analytical results for the WDIP/CDMA MAC protocol. The accuracy of the analytical results is verified by simulations.

4.5.1 Protocol Parameters

The protocol parameters used in the analysis are as outlined in section 3.5.1 and table 3.1. Instead of the different data rate classes, an average data rate class that requires five code-channels is considered.

4.5.2 PrS Results

The analytical performance of PrS is determined in terms of the number of PrS calls in progress as well as the blocking probability of new PrS calls for a given offered PrS load. Fig. 4.1 shows the number of PrS calls versus the offered load for various maximum guaranteed capacities P_{max}^{ps}. The number of PrS calls increases as the offered load increases (by increasing the call initiation probability). For heavy premium traffic load, the number of PrS calls reaches saturation, since most new calls are blocked.

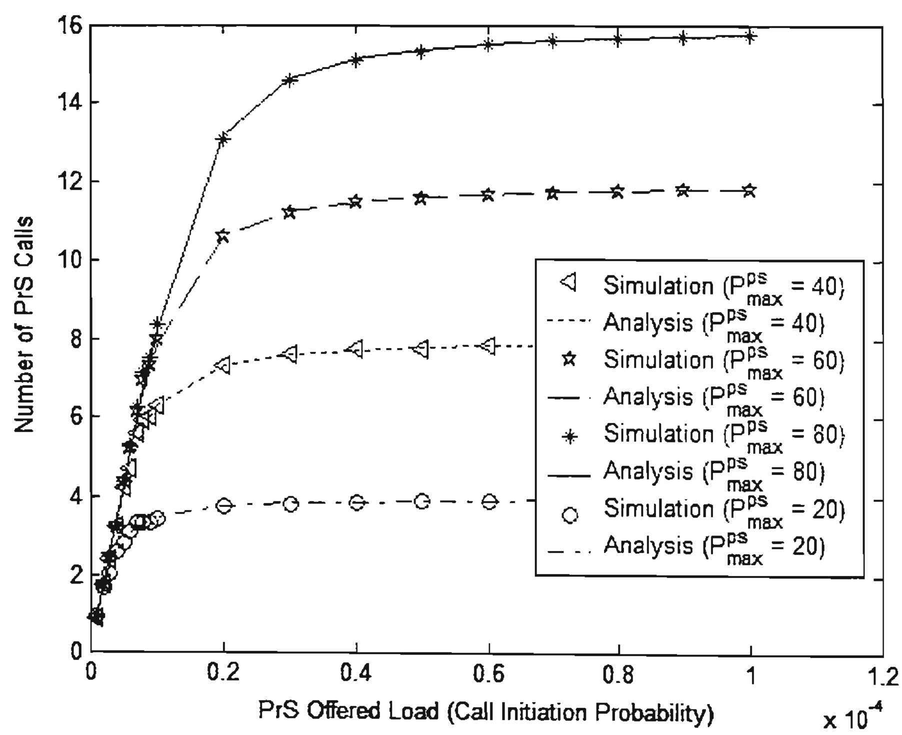


Fig. 4.1: Average number of premium service calls versus the offered premium service load for various system capacities.

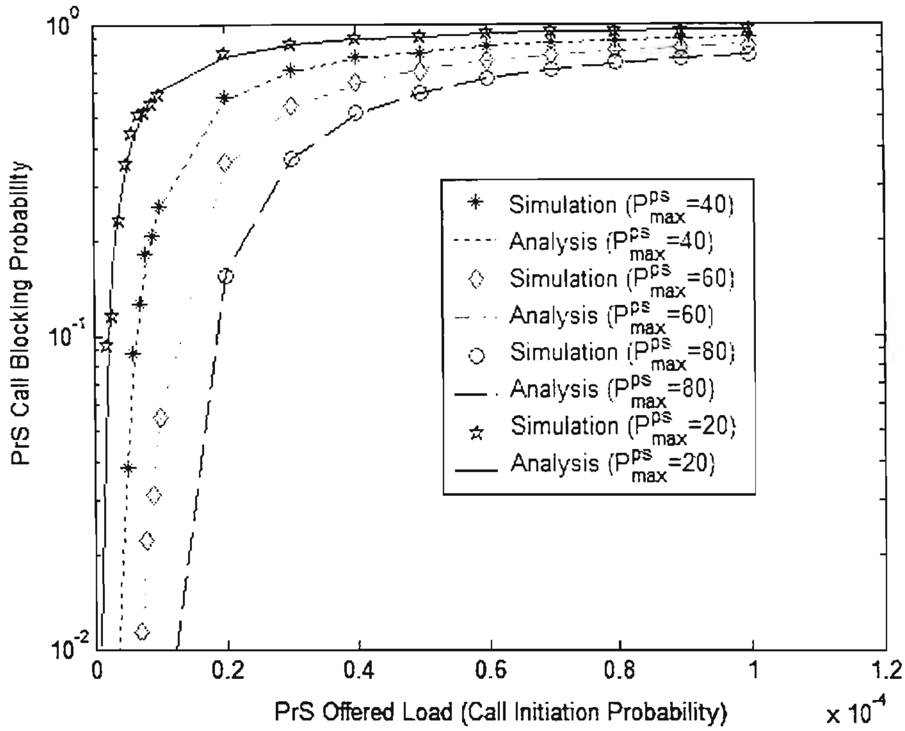


Fig. 4.2: Average premium service call blocking probability versus the offered premium service load for various system capacities.

In Fig. 4.2, the call blocking probability for new premium calls is shown versus call initiation probability. As the call initiation probability increases, the expected call blocking probability increases. The blocking probability decreases as the capacity reserved for PrS increases, which means that more calls can be accepted. This is further illustrated in Fig 4.3 where the PrS call blocking probability versus P^{ps}_{max} is plotted.

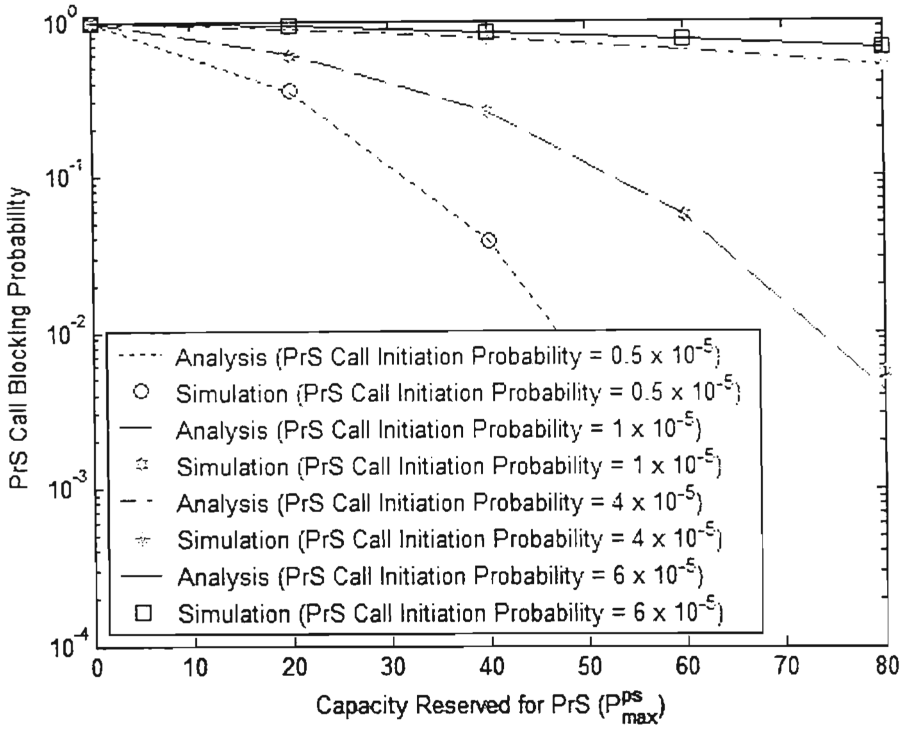


Fig. 4.3: Premium service call blocking probability versus the maximum reserved PrS capacity

4.5.3 AS Results

As in chapter 3, the performance measures (number of AS calls and AS call blocking probability) are determined under various conditions of PrS traffic level (i.e. PrS arrival rate) to illustrate the dependence of AS on PrS. Fig. 4.4 shows the number of AS calls versus the offered load for various values of premium service call initiation probability. It can be seen that for very low PrS traffic level, the number AS calls reach saturation very close to the system capacity. As in PrS, the saturation of the curve indicates that the system capacity limit has been reached, therefore new calls are blocked. For high PrS traffic level, we can see that AS users use the reserved AS capacity since there is no remaining capacity from PrS. In Fig. 4.5, the call blocking probability for new AS calls is shown versus call initiation probability for various values of premium service call initiation probability.

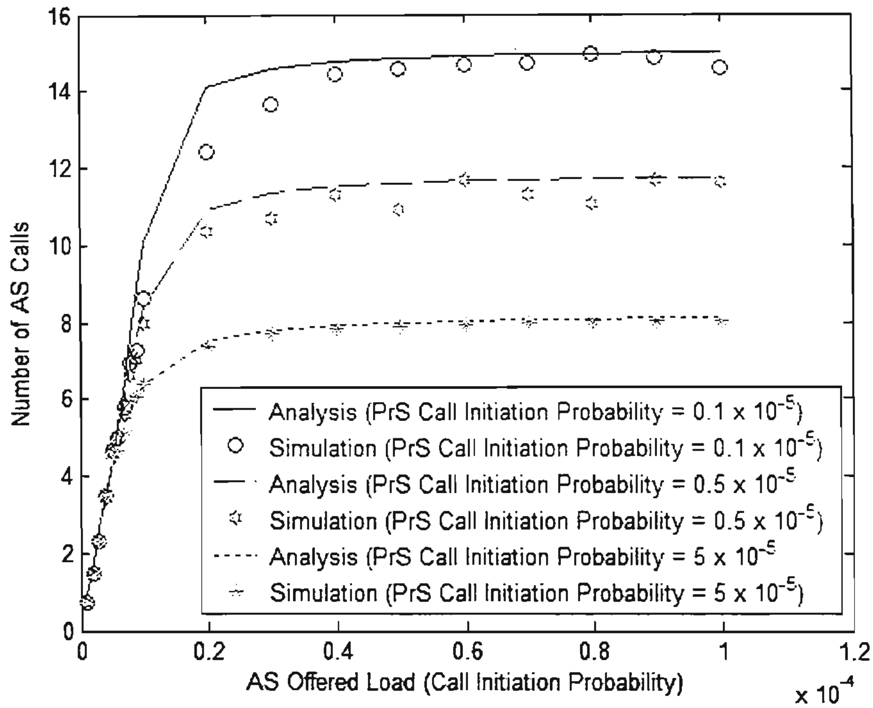


Fig. 4.4: Average number of AS calls versus the offered AS load for various values of premium service call initiation probability

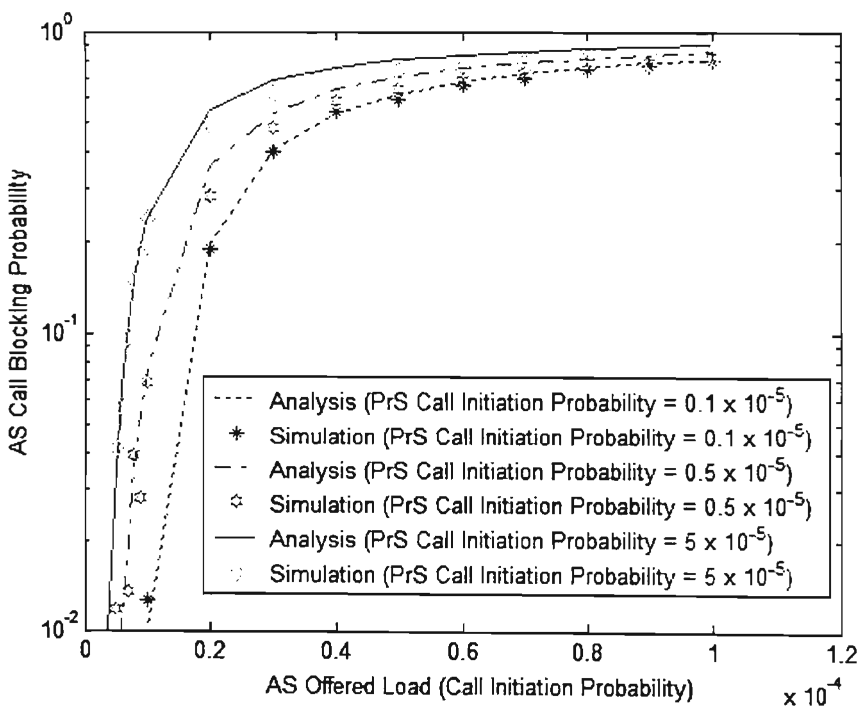


Fig. 4.5: Average AS call blocking probability versus the offered AS load for various values of PrS call initiation probability

4.5.4 BE Results

The analytical model cannot use PrS arrival rates as in chapter 3, therefore we choose to use the average number of packets that PrS and AS users transmit per slot to illustrate the effect of AS and PrS on BE traffic. Fig 4.6 shows the BE message throughput as a function of the average best effort offered load. In Fig 4.7, the BE packet throughput is shown as a function of the average best effort offered load.

Fig 4.8 illustrates the effect of varying the blocking threshold α on the performance of BE. We can see that the effect of α on the BE performance is load dependant. At low traffic loads, high values of α provide better results than low values of α. For high values of traffic loads we can see that low values of α give better results.

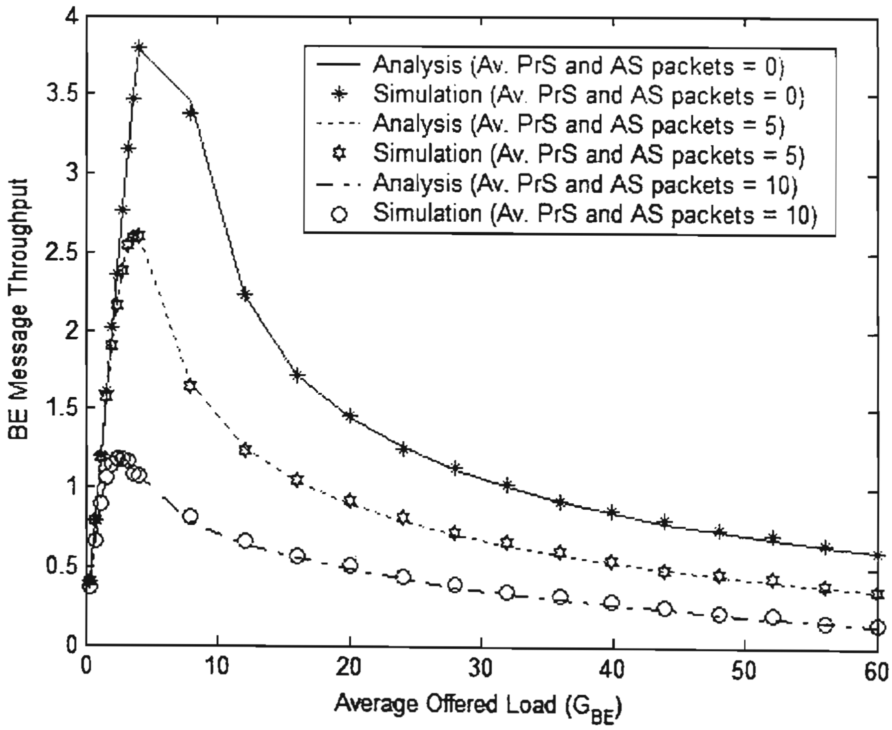


Fig. 4.6: Best effort service message throughput versus the offered Best effort service load for various average transmitted premium and assured service packets per slot

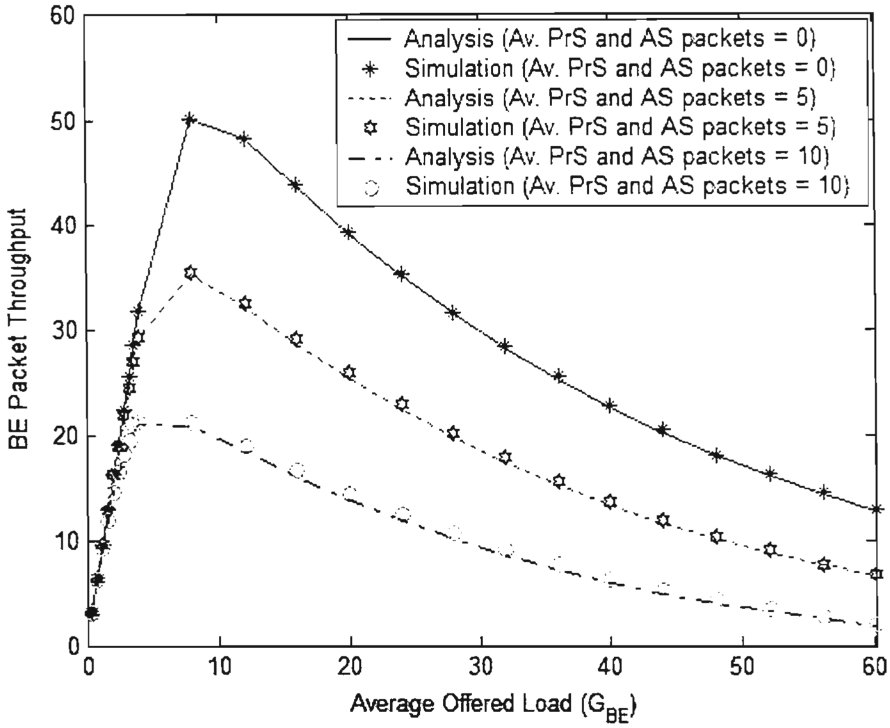


Fig. 4.7: Best effort service packet throughput versus the offered Best effort service load for various average transmitted premium and assured service packets per slot

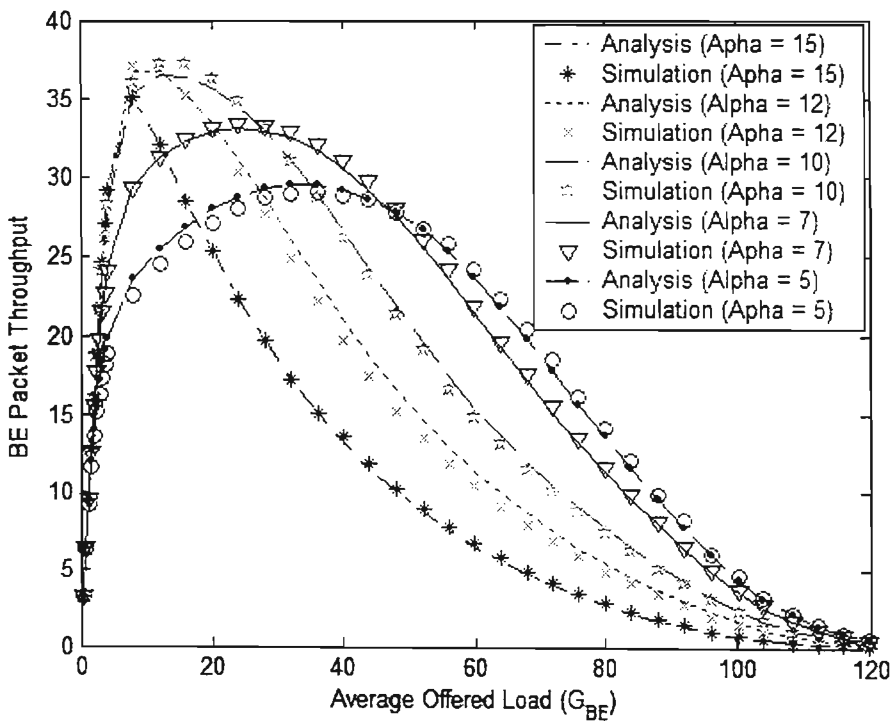


Fig. 4.8: The effect of varying the blocking threshold for average assured and premium packets per slot = 5

4.6 Summary

This chapter dealt with the Markov analysis of the proposed MAC protocol for DiffServ IP-based CDMA networks. We began by presenting the analysis for PrS which is followed by AS analysis. Due to fruitless efforts to incorporate users' data rates to the analysis, we assumed that all users transmit at the same average data rate. For both the PrS and AS subsystems, we computed the expected number of admitted calls and the expected call blocking probability. In the case of BE subsystem, we computed the expected BE state distribution, the mean BE message blocking probability and the expected packet and message throughputs.

In the last section of the chapter results obtained from the analytical model and simulations were presented. The close correlation between the analysis and simulation results validated the accuracy of the analysis. For AS and BE, we showed how they depend on the remaining PrS capacity. For AS when there is no remaining capacity from PrS, all AS users are served with the reserved AS capacity, as a result the blocking probability of new users will be high. In the case of BE, we showed that the performance of BE degrades as the number of AS and PrS packets (due to an increase in AS and PrS users) increase in the channel. This degradation is a result of the fact that the channel capacity for BE is reduced by the number of transmitting AS and PrS users, which also increase the level of MAI in the channel. The effect of varying the blocking threshold α was also illustrated. At low traffic loads, high values of α provide better results than low values of α. For high values of traffic loads, we can see that low values of α give better results.

Chapter 5

WDIP/CDMA MAC Protocol in IEEE 802.16 Broadband Wireless Access (BWA) systems

5.1 Introduction

In this chapter, we adapt ideas from our WDIP/CDMA MAC protocol to IEEE 802.16 BWA systems. The IEEE 802.16 standard provides signalling mechanisms and specifies a MAC layer protocol to create a framework for designing QoS architecture. It does not specify the scheduling algorithm, call admission control and traffic integration schemes to provide fairness,and efficient bandwidth allocation. In order to provide different QoS guarantees to various applications, while still achieving high system utilization such factors should be integrated into the MAC protocol. Much research, e.g., [84] and [85] focus on developing the scheduling algorithms, without considering how to integrate them to MAC protocol and the complexity they introduce into the system.

Since the IEEE 802.16 BWA systems are based on OFDM (or OFDMA), the air interface of the proposed MAC protocol is changed to OFDM. The problems associated with CDMA such as MAI are eliminated in the modified protocols. The main idea that we incorporate to the IEEE MAC protocol is the use of boundary techniques to divide the channel resources into sections which are reserved for particular types of service. Boundary techniques allow for efficient integration of different types of service traffic in a shared medium. The boundaries can be moved to optimise the system performance. In each section, a different MAC protocol (e.g. demand assignment or random access) and scheduling algorithm is applied according to the service assigned to that section. The scheduling algorithm and call admission control are simplified, since the traffic is separated and can be dealt with independently. Furthermore, this simplifies the performance evaluation of the MAC protocol. Due to the complexity of the 802.16 MAC protocol, little has been done in terms of its performance evaluation.

An overview of the IEEE 802.16 based BWA system was presented in Chapter 1. In this dissertation we consider the WirelessMAN-OFDM Physical layer (PHY) air interface, the details of which are discussed in Section 5.2. In Section 5.3, we present the details of the MAC layer protocol and the QoS architecture. In Section 5.4, we describe the modified MAC protocol. Section 5.5, provides the simulation results of the modified MAC protocol.

5.2 WirelessMAN-OFDM Physical layer (PHY)

The WirelessMAN-OFDM PHY is based on OFDM modulation and designed for non line of sight (NLOS) operation in the 2-11 GHz frequency bands. Systems operating in 2-11 GHz frequency and NLOS conditions suffer from severe multi-path propagation. OFDM functions by transmitting multiple modulated sub-carriers in parallel. The spectra of the individual sub-carriers are permitted to overlap, but orthogonal to each other. OFDM signal is carried out by using 256-point inverse fast Fourier transform (IFFT) and fast Fourier transform (FFT) at the transmitter and receiver, respectively. Fig. 5.1 shows the time and frequency domain structure of an OFDM symbol.

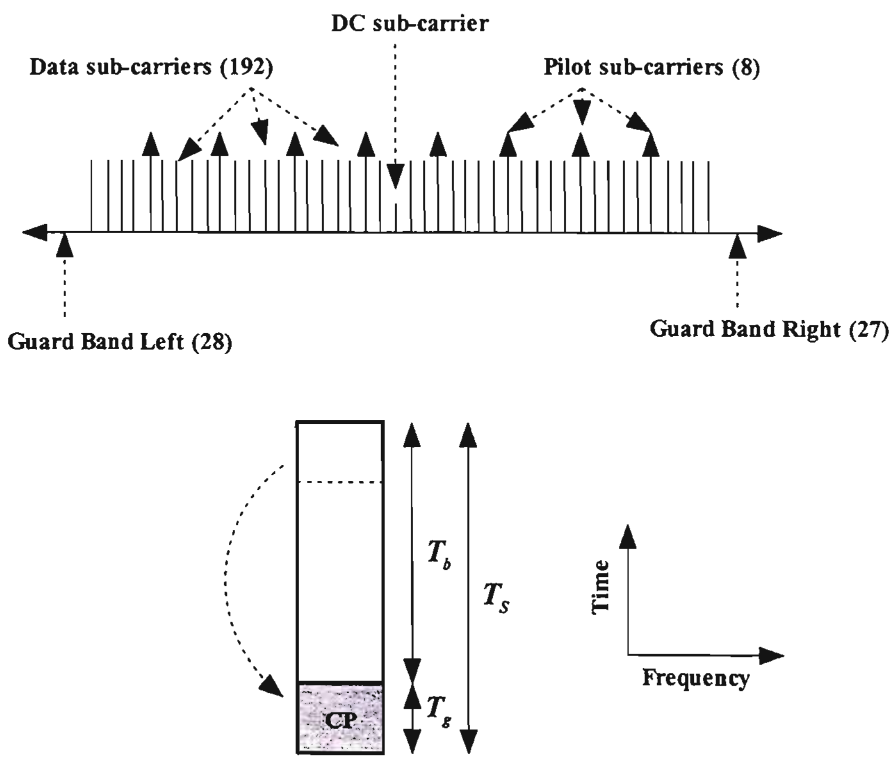


Fig. 5.1: OFDM Symbol structure in frequency and time domain

As shown in Fig. 5.1, an OFDM symbol in frequency domain is made from sub-carriers, namely data sub-carrier, pilot sub-carriers and null sub-carriers. Data sub-carriers are used for data transmission; pilot sub-carriers are used for channel estimation, time and frequency offset estimations; and the null sub-carriers carry no transmission. The null sub-carriers are used for guard band, non-active sub-carriers and DC sub-carriers. The time duration of the OFDM symbol is referred to as the useful symbol time T_b. The useful symbol time is preceded by a cyclic prefix (CP) which is a periodic copy of the last block of useful symbol time. The CP allows the multi-path delay spreads to fall under the guard time, thus eliminating the inter-symbol interference.

The standard defines as mandatory, a combined variable-rate Read-Solomon (RS)/Convolutional Coding (CC) scheme, supporting code rates of 1/2, 2/3, 3/4, and 5/6. Variable-rate Block Turbo Code (BTC) and Convolutional Turbo Code (CTC) are also optionally supported. The standard supports multiple modulation levels, including Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (QAM) and 64-QAM. Finally, the PHY supports (optionally) transmit diversity in the Downlink (DL) using Space Time Coding (STC) and Adaptive Antenna Systems (AAS) with Spatial Division Multiple Access (SDMA).

The OFDM PHY layer supports TDD and FDD operations. FDD SSs may support half-duplex FDD which may be less expensive, since it does not simultaneously transmit and receive. In the license exempt bands, the duplexing technique is TDD. The data transmission is based on frames, which are further divided into time slots. The frame durations supported are 2.5 ms, 4 ms, 5 ms, 8 ms, 10 ms, 12.5 ms, and 20 ms.

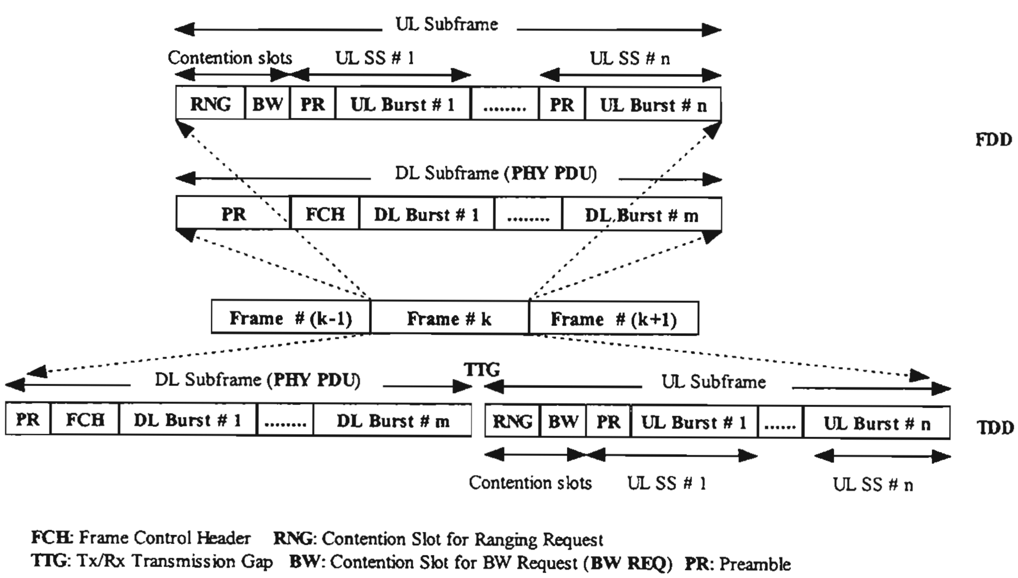


Fig. 5.2: OFDM PHY frame structure

Fig. 5.2 illustrates the frame structure employed by the OFDM PHY. The frame consists of a downlink subframe and an uplink subframe. The downlink subframe starts with a long preamble which is used for synchronization. The preamble is followed by the frame control header (FCH) burst. The FCH contains information that specifies burst profile (modulation/coding and FEC changes) and length of downlink data bursts. The FCH burst is modulated by ½ BPSK and utilises one OFDM symbol. The first burst of the downlink contains MAC control messages, e.g. DL-MAP and UL-MAP. Burst #1 may use a more efficient modulation/coding scheme supported by all SSs and the BS. The downlink burst transmitted to each SS uses a negotiated burst profile. The bursts are transmitted in order of decreasing robustness to allow SSs to receive their data before being presented with burst profile that could cause them to lose synchronisation with the downlink. The downlink subframe may consist of an optional space time coding (STC) zone where all downlink burst are STC encoded. The STC zone starts from a preamble and STC encoded FCH-STC burst. FCH-STC burst is modulated by ½ BPSK.

The uplink subframe consists of uplink contention slots used by SSs for initial ranging[1] and bandwidth requests. The contention slots consist of a number of transmission opportunities (TO) which are scheduled by the BS. Every initial ranging request (RNG-REQ) message starts with a long preamble. The frame supports two types of contention based bandwidth request regions. A region is a two dimensional allocation of a group of contiguous subchannels in a group of contiguous OFDM symbols. Bandwidth request (BW) can be sent on a REQ Region-Full or REQ Region-Focused. In REQ Region-Full the SS sends the bandwidth request header in full contention transmission. Each TO utilises a short preamble over one OFDM symbol. The uplink subcarriers can be grouped to form subchannels. There are 16 subchannels that can be supported. When subchannelization is employed, the RS encoder is bypassed and each TO shall consist of subchannel preamble, number of subchannels and OFDM symbols. The REQ Region-Focused consists of two phases. In the first phase, the SS sends a request signal to BS utilizing a contention code modulated on a Contention Channel consisting of 4 subcarriers over 2 OFDM symbols. During the second phase, the SS responds with a bandwidth request header in region allocated by the BS.

Unlike the downlink subframe, the uplink subframe conveys uplink bursts from different SSs. Uplink bursts modulation/coding is specific to the source SS. Each uplink burst starts with a short or subchannelisation preamble. AAS alert slots may be provided in the uplink subframe to support AAS SSs initial ranging. Each AAS new entry request shall be started with a long preamble.

---

[1] Ranging is a collection of processes by which the SS and BS maintain the quality of the RF communication link between them

In addition to the contention slots, the uplink subframe shall also consist of periodic ranging and non-contention BW request transmission intervals. The BS schedules these transmission intervals according to SS requirements or timer interrupts. Uplink midamble may be used to enhance SSs performance. If used, uplink midambles will be identical to the preamble used by the uplink burst.

5.3 IEEE 802.16 MAC protocol

The MAC protocol is capable of supporting multiple PHY specifications optimized for the frequency band of operations. Different types of traffic, including IP and ATM are supported. In PMP mode the wireless link operates with a central BS capable of handling multiple independent sectors simultaneously. The downlink is broadcast with data to subscriber stations (SSs) multiplexed in time division multiplexing fashion. The uplink is shared by SSs in a TDMA or OFDMA fashion. The MAC is connection-orientated; the applications first establish a connection to the BS. All services, including connectionless services are mapped to a connection. This allows bandwidth requests for connections to be associated with predefined service flows parameters such as QoS.

Connections are identified with a 16-bit connection identifier (CID) located in the MAC header fields. When a SS joins the network through a registration process, two management connections in the downlink and uplink directions are established between the BS and the SS. Another management connection can be optionally generated. These three connections reflect that there are three different QoS requirements used by different management traffic between the BS and SS. The first management connection is the basic connection which is used to transfer short and time-critical messages. The primary management connection is used for transmission of longer and more delay-tolerant messages such connection setup. The secondary management connection is used to transfer delay tolerant standard based management messages such as Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP) and Simple Network Management Protocol (SNMP).

Each SS has a 48-bit universal MAC address, which serves mainly as an equipment identifier since primary addresses used during operation are the CIDs. The data unit exchange between the BS and SS is called the MAC protocol data unit (PDU). The MAC PDU consists of a generic MAC header, variable length payload, and an option cyclic redundancy check (CRC). There are two types of MAC header defined in the IEEE 802.16 standard, and they are distinguished by a header type field. The generic MAC header is used at the beginning of each MAC PDU containing MAC management messages. The bandwidth request header is used for bandwidth request. The bandwidth request MAC PDU contains no payload. The CRC is used for error detection for PDUs for all ARQ-enabled connections. The implementation of CRC is mandatory for OFDM and OFDMA physical layers.

In addition to the MAC headers, four types of subheaders are defined. The grant management subheader is used by a SS to convey bandwidth management to the BS. The Fragmentation subheader contains information that indicates the presence and orientation of service data units (SDU) fragments in the payload. The packing subheader is used to indicate the packing of multiple SDUs into a single MAC PDU. The ARQ feedback payload (or subheader) is used to carry negative or positive acknowledge signalling information.

5.3.1 Connection setup between SS and BS

IEEE 802.16 uses the concept of service flows to define unidirectional transport of packets on either downlink or uplink. Service flows are characterized by a set of QoS parameters such as latency and jitter. To most efficiently utilize network resources such as bandwidth and memory, 802.16 adopts a two phase activation model in which resources assigned to a particular admitted service flow may actually be committed until the flow is activated. Each admitted or active flow is mapped to a MAC connection with a unique CID.

Service flows are preprovisioned, and setup of the service flows is initiated by the BS during SS initialization. However, service flows can be dynamically established by either the BS or the SS. The SS typically initiates service flows only if there is a dynamically signalled connection, such as switched virtual connection (SVC) from ATM network. The establishment of a service flows is performed via a three-way handshaking protocol in which the request for service flow establishment is responded to and the response acknowledged. Dynamic service establishment supports dynamic services in which service flow parameters are renegotiated.

5.3.2 Classes of services

There are four classes of services supported in the 802.16 standards, namely the Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS) and Best Effort (BE). Each uplink connection is mapped to one of the classes of uplink services. Each class of uplink service is associated with a set of QoS parameters which quantify aspects of its behaviour.

The UGS is designed to support real-time data streams consisting of fixed-size data packet transmitted at periodic intervals, such as T1/E1 and Voice over IP without silence suppression.

The BS regularly schedules grants of the size negotiated at connection setup to eliminate the overheads and latency of bandwidth requests in order to meet the delay and delay jitter requirements of the underlying service. No explicit requests from the SS are required. The SSs use a poll me bit from the grant management subheader to report that the transmission queue is backlogged due to factors such as lost grants.

The rtPS is designed to support data streams consisting of variable-sized data packets that are transmitted at fixed intervals, such as moving pictures experts group (MPEG) video. The service offers real-time, periodic, unicast request opportunities, which meet the flow's real-time needs and allow the SS to specify the size of the desired grant. This service requires more request overhead than UGS, but supports variable grant sizes for optimum data transport efficiency.

The nrtPS is designed to support delay-tolerant data streams consisting of variable-sized data packets for which a minimum data rate is required, such as FTP. In addition to unicast request opportunities, SSs with nrtPS are allowed to use contention request opportunities.

The BE service is designed to support data streams for which no minimum service level is required and therefore may be handled on a space-available basis. The SS sends requests for bandwidth using contention opportunities.

5.3.3 Bandwidth allocations and requests schemes

Bandwidth is allocated according to the SS's bandwidth requirements and QoS level assigned to the SS connections during network entry and initialization. The BS can use a polling (unicast or multicast) process to allocate bandwidth to the SSs specifically for the purpose of making bandwidth requests. These allocations may be to individual SSs or to groups of SSs. Bandwidth grants shall be scheduled by the BS in response to bandwidth requests. UGS, bandwidth grants are periodically scheduled as negotiated at connection setup. The request/grant mechanism is a self-correcting protocol rather than an acknowledged protocol. Therefore SSs shall periodically use aggregate bandwidth requests. Bandwidth requests are always per connection while grants are either per connection (GPC) or per subscriber station (GPSS). In GPSS, the BS grants bandwidth to the SS. The SS may re-distribute bandwidth among its connections maintaining QoS and service-level agreements. It is suitable for many connections per terminal and allows more sophisticated reaction to QoS with low overheads. For GPC, the BS grants bandwidth to a connection. GPC is more suitable for few users per SS. However, GPC results in high overheads but allows simpler SS.

A request may come as a stand alone bandwidth request header or may come as piggybacked requests. Bandwidth requests can be incremental or aggregate, as indicated in the MAC header. The piggybacked requests used by non-UGS service shall always be incremental. A Poll-Me bit in the grant management subheader in MAC packet of a UGS connection may be set to indicate to the BS that they need to be polled to request bandwidth for non-UGS connections.

5.3.4 Contention based bandwidth requests

Contention based bandwidth requests is used by BE and nrtPS services. The OFDM PHY layer supports two types of contention based bandwidth request mechanisms. Bandwidth requests can be sent on a REQ Region-Full or REQ Region-Focused. In REQ Region-Full the SS send the bandwidth request header in full contention transmission. The REQ Region-Focused consists of two phases. In the first phase, the SS sends a request signal to BS utilizing a contention code modulated on a contention channel consisting of four sub-carriers. During the second phase, the SS responds with a bandwidth request header in region allocated by the BS.

The OFDMA PHY also supports two types of contention based bandwidth requests mechanisms. In the first mechanism, the SS simply sends (in contention transmission) a bandwidth request header to the BS. The second mechanism is based on CDMA. The OFDMA PHY specifies ranging subchannels and subset of ranging codes that shall be used for contention based bandwidth requests. The SS requests bandwidth by selecting, with equal probability, a ranging code from the code subset allocated to bandwidth requests. This ranging code shall be modulated onto the ranging subchannel and transmitted during the appropriate uplink allocation.

5.4 Modified 802.16 MAC protocol

The IEEE 802.16 MAC protocol defines the signaling mechanism for information exchange between BS and SS connection setup, bandwidth request, and UL-MAP, and QoS architecture with four classes of service (i.e. UGS, rtPS, nrtPS and BE). The following points are open for discussion or under the responsibility of the implementers; the design of scheduling algorithms, call admission control, and traffic integration schemes to provide fairness and efficient bandwidth allocation. In order to provide a complete solution for QoS provision, these issues must be integrated to the MAC protocol.

To address the above issues, we propose to incorporate ideals from our WDIP/CDMA protocol described in Chapter 3 to complete the missing parts in the IEEE 802.16 MAC protocol. Fig. 5.3 illustrates the uplink frame structure of the modified IEEE 802.16 MAC protocol. In the OFDM domain, the subchannels are divided into compartments for UGS, rtPS, and nrtPS using the movable boundary concept. A call admission control implemented in the BS is used to accept new connections to the corresponding service compartment.

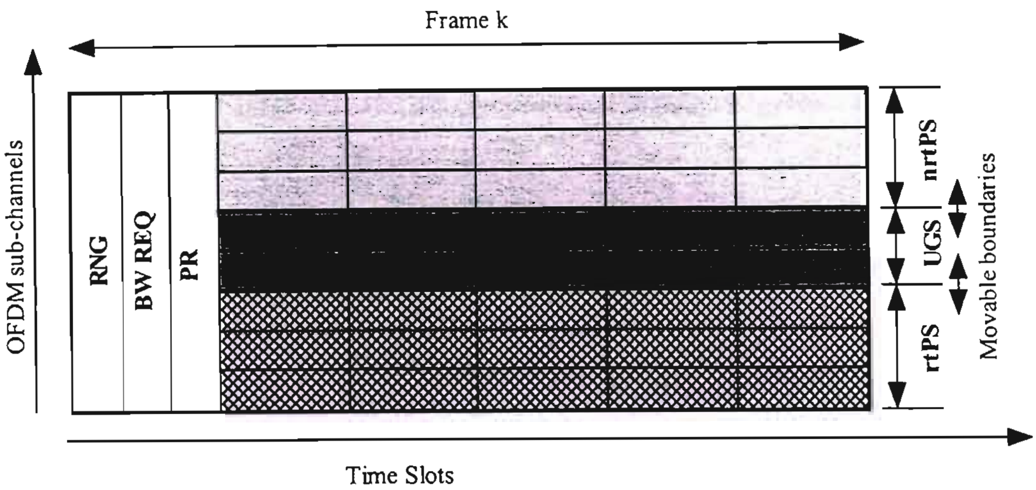


Fig. 5.3: Uplink frame structure

For UGS and rtPS, the CACs are similar to the CAC for premium service. The difference is that the system resources are now expressed in terms of OFDM subchannels and time slots. Once a UGS connection is admitted and its service flow activated, the BS will grant bandwidth to corresponding SS as described in Section 5.3.2. A new UGS connection is blocked when channels are no longer available in the UGS compartment. Like UGS, a new rtPS connection is blocked when no channels are available in the rtPS compartment. Since the SS may exceed the traffic parameters negotiated during connection setup and overload the network, the BS implements a traffic policing function to ensure that the connections conform to the negotiated traffic parameters.

The bandwidth allocation for nrtPS traffic is based on the reserved nrtPS channels and unused UGS and rtPS channels. Therefore, new nrtPS connections are blocked when the total system capacity is used. This service is treated the same as assured service (see Chapter 3). For BE service the MAC protocol for DiffServ BE service described in Chapter 3 is adopted. In this MAC protocol, collision is associated with two or more SSs transmitting on the same OFDM subchannel. The BS determines the unused subchannels in each time slot and broadcasts them using the DL-MAP and UL-MAP control messages. When no subchannels in a particular time slots are available to be used by SSs with BE traffic, the BS broadcasts a tone indicating blocking in these time slots.

5.5 Simulation Results

In this section we evaluate the performance of the modified MAC protocol through simulations. The performance measures considered are: the number of connections, call blocking probability, and throughput (only for BE traffic). The apparent difference between the results in this chapter and chapter 3 is that BER is not considered as the modified MAC protocol is based on OFDM. Otherwise, the systems behaviour is similar to that in chapter three.

5.5.1 Protocol Parameters

The following parameters shown in Table 5.1 were used in the simulation.

Table 5.1: Protocol parameters

Parameters	**Values**
Channel bandwidth	20 MHz
Total channel data rate	75 Mbit/s
UGS guaranteed data rates (packets/frame)	1,2,3,4,5,6
rtPS minimum data rates (packets/frame)	1,2,3,4,5,6
nrtPS minimum data rates (packets/frame)	1,2,3,4,5,6
Frame duration	20 ms
Average call holding time	3 min
Average BE message length	160ms
Uplink data time slots/frame	4
OFDM sub-channels/time slot	16
OFDM sub-channel data rate	448 kbit/s
Capacity (sub-channels)	64
UGS population (N_{UG})	80
rtPS population (N_{RT})	80
nrtPS population (N_{NRT})	80
Best effort service population (N_{BE})	∞

5.5.2 UGS, rtPS and nrtPS Results

The total system capacity was divided into 31.25, 50, and 18.75 percentage of capacity among UGS, rtPS and nrtPS traffic, respectively. Fig. 5.4 shows the performance of UGS, rtPS and nrtPS in terms of the number of connections versus the offered load. Fig. 5.5 shows the call blocking probability versus the offered load.

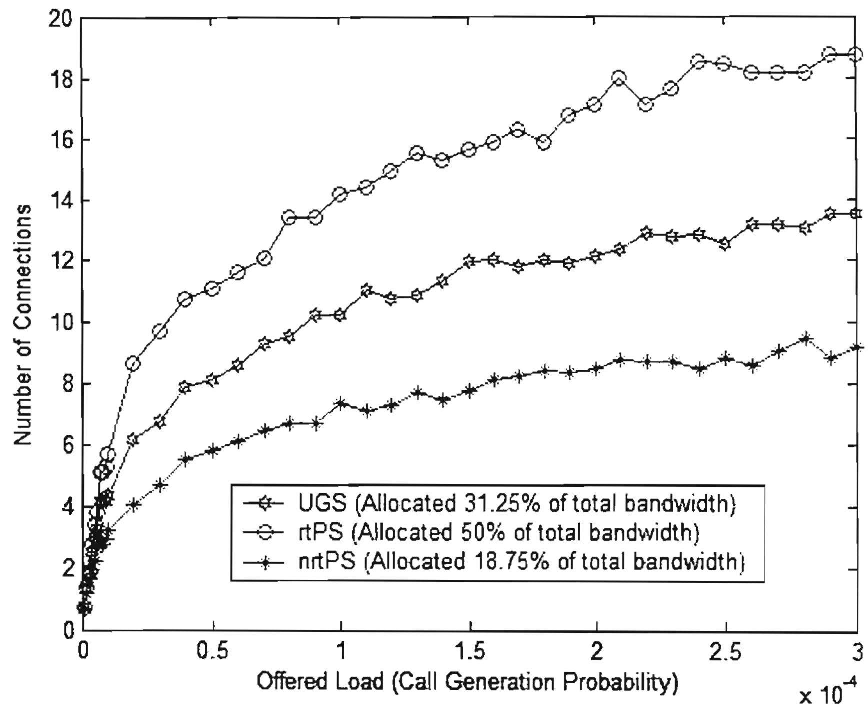


Fig. 5.4: Number of connections versus the offered load

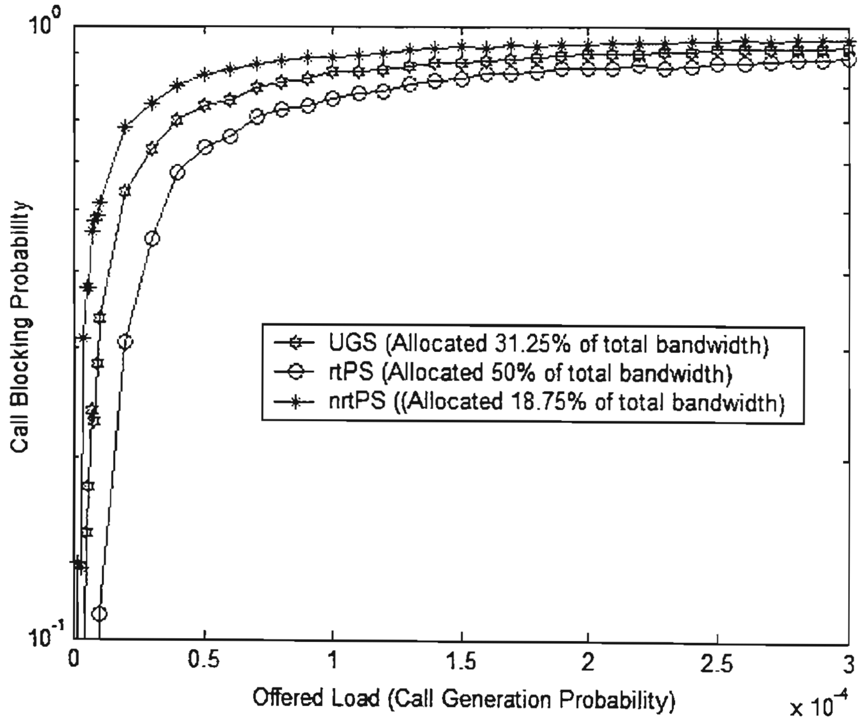


Fig. 5.5: Average call blocking probability versus the offered load

5.5.3 nrtPS Results

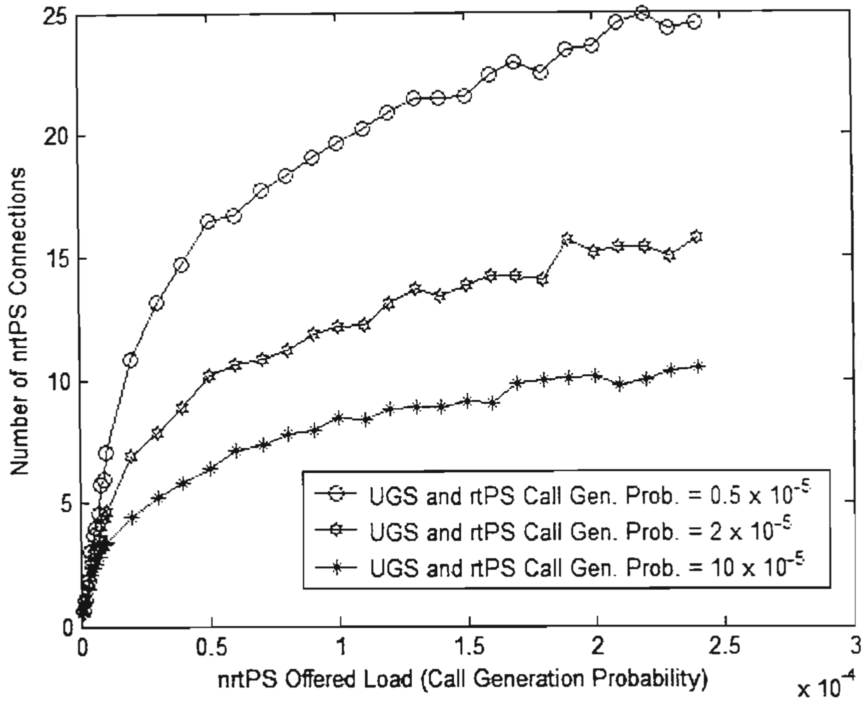


Fig. 5.6: Number of nrtPS connections versus the offered nrtPS load

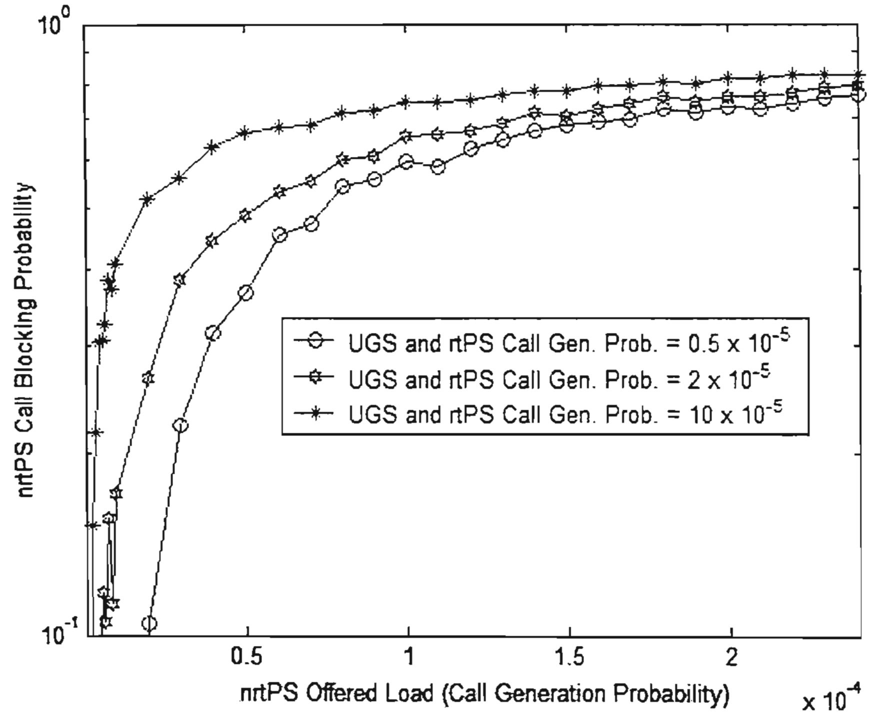


Fig. 5.7: Average nrtPS call blocking probability versus the nrtPS load

Fig. 5.6 shows the number of nrtPS connections versus the offered load for various UGS and rtPS call generation probabilities. In Fig. 5.7, the call blocking probability is shown versus the offered load for various UGS and rtPS call generation probabilities. The performance of nrtPS is considered under various UGS and rtPS call generation probabilities to illustrate how it is affected by UGS and rtPS traffic. At high UGS and rtPS traffic load, nrtPS relies on its reserved capacity. Therefore, the number of active nrtPS connections is low and the blocking probability is high.

5.5.4 BE Results

Fig. 5.8 shows the BE packet throughput versus the offered BE load for four combined UGS, rtPS, and nrtPS call arrival rates. The different call arrival rates illustrate the dependence of BE service on the capacity from the other services. For low BE traffic loads (<15), the throughput increases as the offered load increases. For traffic loads greater than 15, the throughput decreases as the offered load increases. The degradation of the system at high offered traffic load is attributed to the increase of collisions due to message choosing to the same subchannels to transmit. In Fig. 5.9, the BE message throughput versus the BE offered load is illustrated for various combined UGS, rtPS, and nrtPS call arrival rates.

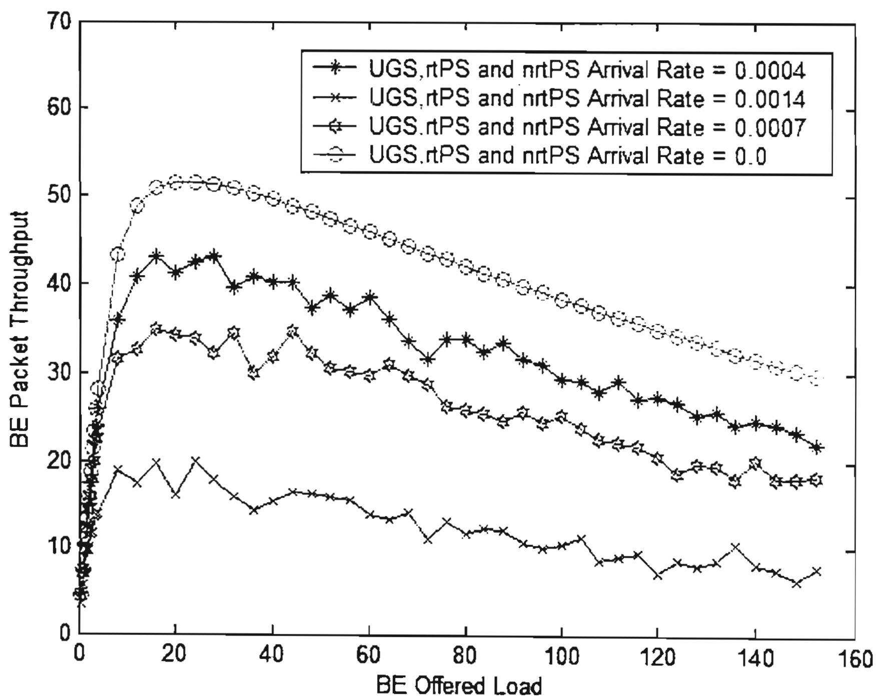


Fig. 5.8: Best effort service packet throughput versus the Best effort service offered load for various UGS, rtPS and nrtPS call arrival rates

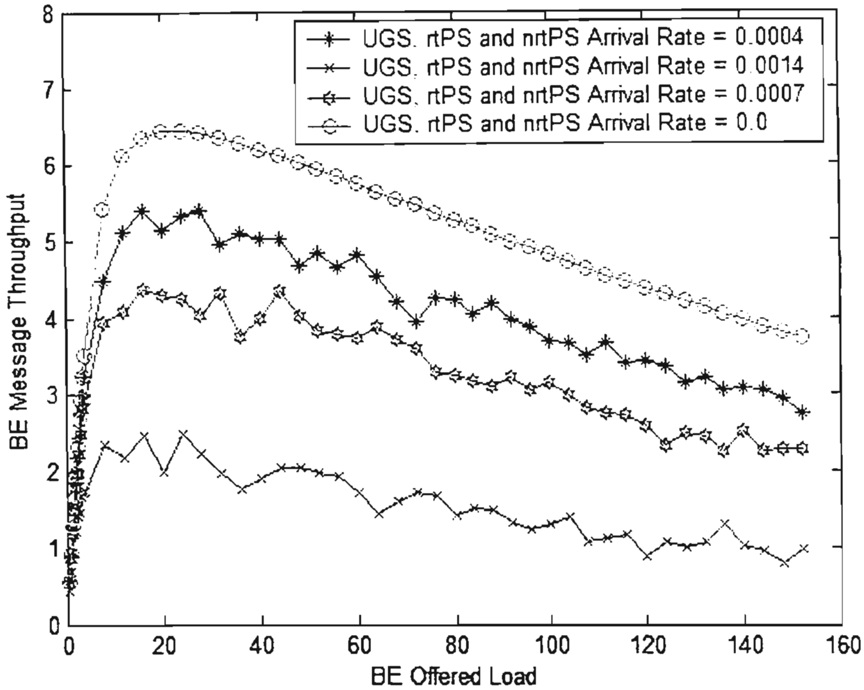


Fig. 5.9: Best effort service message throughput versus the Best effort service offered load for various UGS, rtPS and nrtPS call arrival rates

5.6 Summary

In this chapter, the MAC protocol proposed in chapter 3 was modified to adapt to OFDM environment and its features incorporated to the MAC protocol for IEEE 802.16 BWA systems. An overview of the physical layer and MAC layer of the IEEE standard was first given. The features incorporated to the IEEE MAC protocol include the partitioning of the system capacity using boundaries to provide simple service integration and efficient bandwidth utilization by allowing nrtPS and BE to use remaining capacity. The admission policy for the UGS and rtPS and nrtPS IEEE 802.16 service classes are similar PrS and AS. BE is treated as in chapter 3, however collisions are associated with two or more SSs transmitting on the same OFDM subchannel.

Simulation results were presented to illustrate the performance of the modified MAC protocol. The difference between the simulation results in chapter 3 and this chapter is the non consideration of BER due to MAI as the system is based on OFDM. The performance behaviour is similar to chapter 3. For BE, the degradation of the performance is due to high message collisions when the traffic load is high as in S/ALOHA.

Chapter 6

Conclusions

This dissertation started by giving an overview of wireless networks and their corresponding standards. This was followed by the discussion of QoS provisioning in IP-based wireless networks. DiffServ emerged as the efficient and scalable QoS mechanism that needs to be extended to the wireless domain. The roles of the MAC protocol in wireless networks and in the QoS issue were highlighted. A brief background of CDMA, the motivation of the research and the dissertation outline were then given.

Chapter 2 contains the literature survey of MAC protocols for centralized wireless networks. The protocols were first classified into four categories namely; fixed assignment, random access, guaranteed assignment, and the hybrid access protocols. The hybrid access protocols were further classified into random reservation protocols and demand assignment protocols. The requirements of a MAC protocol carrying multimedia traffic as it would be in the 4G wireless networks were described. Under each category of MAC protocols a few examples of proposed protocols were given. For random access, the protocols discussed are ALOHA, SS/ALOHA, CLS, S-ALOHA, CSMA and ISMA. In guaranteed assignment protocols, the Zhang's proposal, DTMP, and Acampora's proposal were discussed. For random reservation protocols, the discussed protocols are PRMA, MD PRMA BB and DT/CDMA. The protocols which falls under demand assignment schemes looked at include, DQRUMA, DTDMA/TDD, DSA++, MASCARA, WISPER, Uplink CDMA protocol, IP QoS protocol, Wide-Band TD-CDMA MAC protocol and WAC/MB.

In Chapter 3, our proposed MAC protocol which supports the heterogeneous DiffServ IP traffic and provides QoS guarantees in wireless CDMA networks was presented in detail. The protocol is a hybrid of DA and RRA classes of MAC protocols. The DA supports AS and PrS traffic, while the RRA protocol which is based on DT/CDMA supports BE traffic. The DA scheme is most suitable for AS and PrS traffic, since their QoS requirement must be declared upfront for the CAC. The total system capacity is divided amongst AS and PrS. When the bandwidth allocated to PrS is not used, AS is allowed to use it. However, PrS is only allowed to use its reserved capacity. Therefore, when the network is dimensioned, the reserved capacity for PrS should be maximized such that its QoS requirements are always met. BE users contend for the remaining capacity from both AS and PrS. Each AS or PrS is allocated a maximum guaranteed capacity which is negotiated at call setup. The excess traffic that a user can generate is transmitted as BE. The proposed protocol was then evaluated through time driven computer simulations. The simulations provide a more detailed performance of the protocol which is not possible in statistical analysis due to unavoidable simplifying assumptions. The performance metrics considered for AS and PrS are call blocking probability as well as the number of admitted calls. For BE, the message blocking probability and packet throughput was considered.

Chapter 4 focused on the Markov analysis of the proposed MAC protocol. To simplify the analysis, AS and PrS users were assumed to transmit at the same average data rate. For PrS, a two state system was analysed, but due to finite MT population the states reduced to one. For AS a three state system which reduced to two due to finite MT population was analysed for the determination of the number of admitted AS calls. In order to accurately determine the call blocking probability of AS, two (permanent state and temporary state) of the three states were combined to form one state of transmitting AS users. In the BE analysis one state of transmitting BE users was analysed since all other states (silence and backlogged) are absorbed in the infinite BE MT population. For both the PrS and AS subsystems, the expected number of admitted calls and the expected call blocking probability were computed. In the case of BE subsystem, the expected BE state distribution, the mean BE message blocking probability and the expected packet and message throughputs were computed. The analytical results were presented and validated with simulation results. For BE service, it was shown how the blocking threshold below the maximum MAI threshold affect the performance of BE. A dynamic blocking threshold which varies according to the offered BE load can improve the performance.

In Chapter 5, we modified our proposed MAC protocols for implementation in IEEE 802.16 BWA systems and adaptation to the OFDM environment. This chapter began by giving a comprehensive overview of the physical and MAC layers of the IEEE 802.16 standard. This was followed by the proposed modifications of the proposed MAC protocols. The objective was to provide the QoS features such as CAC, traffic integration and efficient bandwidth allocation schemes which are not specified in IEEE 802.16 MAC protocol. The modified MAC protocol was evaluated through time driven simulations. The different between the simulation results in chapter 3 and chapter 5 is the fact that BER due to MAI is eliminated in chapter 5. Collisions are the degradation factor in the performance of BE. These collisions are associated with two or more users choosing the same OFDM subchannel for transmission rather than channel overloads as in CDMA based systems.

References

[1] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet: an Overview" *RFC 1633*, June 1994.

[2] S. Blake, D. Black, M. Carlson, E. Davis, Z. Wang and W. Weiss, "An Architecture for Differentiated Services", *RFC 2475*, Dec. 1998.

[3] X. Xiao and L. M. Ni, "Internet QoS: A Big Picture", *IEEE Network*, March/April 1999.

[4] I. F. Akyildiz and J. McNair, "Medium Access Control Protocols for Multimedia Traffic in Wireless Networks", *IEEE Network*, July/Aug. 1999.

[5] G. Judge, "Protocols for Voice/Data Integration in CDMA Packet Radio Network", PhD Thesis, University of Natal, 1999.

[6] G. Judge and F. Takawira, "Spread-sprectrum CDMA packet radio MAC protocol using channel overload detection and blocking", *Wireless Networks*, vol. 6 no. 6, pp 467-479, Jan. 2000.

[7] R. J. Majoor, "A MAC Protocol for Wireless Networks with QoS Guarantees", PhD Thesis, University of Natal, 2002.

[8] R. Majoor and F. Takawira, "Markov Analysis of a Joint WATM/CDMA MAC Protocol," *IEEE AFRICON, Cape Town, South Africa*, pp 269-274, Sept. 1999.

[9] A. E. Brand and A. H. Aghvami, "Multidimensional PRMA with Prioritized Bayesian Broadcast- A MAC Strategy for Multiservice Traffic over UMTS", *IEEE Trans. Veh. Technol*, vol. 47, No. 4, Nov. 1998.

[10] I. F. Akyildiz, "A Slotted Protocol with BER Scheduling for Multimedia Networks", *IEEE/ACM Trans. Networking*, vol. 7, No. 2, April 1999

[11] X. Wang, "Providing Wireless Internet Access in TDMA/TDD Wireless LAN through a new wireless Packet Fair Queuing Scheme", *proceedings of IEEE VTC Fall 2001*, Atlantic City, NJ, Oct. 2001.

[12] E. Hossain and V. K. Bhargava, "A centralized TDMA-Based Scheme for Fair Bandwidth Allocation in Wireless IP Networks", *IEEE J. Select. Areas Commun.*, vol 19, No. 11, Nov. 2001.

[13] J.Y. Khan, "A RLC/MAC protocol architecture for a wireless IP network", *Personal, Indoor and Mobile Radio Communications*, 2002, The 13th IEEE International Symposium on, Vol. 3, 2002.

[14] A. Baiocchi, F. Couma, and S. Bolognes, "IP QoS Delivery in a Broadband Wireless local Loop: MAC Protocol Definition and Performance Evaluation", *IEEE J. Select. Areas Commun.*, vol 18, No. 9, September 2000.

[15] S. Choi and K. G. Shin, "An Uplink CDMA System Architecture with Diverse QoS Guarantees for Heterogeneous Traffic", *IEEE/ACM Trans. Networking*, vol. 7, No. 5, Oct. 1999

[16] X. Wang, "A new Scheduling Scheme for the Wideband TD-CDMA MAC Protocol", in proceedings of IEEE ICC 2001, Helsinki, Finland, June 2001.

[17] B. Bae and D. Cho, "Performance Analysis of an Integrated Voice/Data CDMA System with Dynamic Admission/Access Control", *proceedings of IEEE VTC'2001 spring*,

[18] I. Koo, S. Bahng and K. Kim, "QoS-Sensitive Admission Policy for Non-real-Time Data Packets in Voice/Data Integrated CDMA Systems" *proceedings of GLOBECOM'01 Conference*, vol. 4, pp 2611-2616, 2001.

[19] W. Mohr and S. Onoe, "The 3GPP Proposal for IMT-2000", *IEEE Commun. Mag.*, Dec. 1999, pp 72-81.

[20] X. Wang, "An FDD Wideband CDMA MAC Protocol for Wireless Multimedia Networks", *IEEE INFOCOM 2003*.

[21] J. E. Wieselthier and A. Ephremides, "Fixed- and Movable-Boundary Channel-Access Schemes for Integrated Voice/Data Wireless Networks", *IEEE Trans. Commun.*, vol. 43, no. 1, pp. 64-74, Jan. 1995.

[22] X. Qiu and V. O. K. Li, "A Unified Performance Model for Reservation-Type Multiple-Access Schemes," *IEEE Trans. Veh. Technol.*, vol. 47, no. 1, Feb. 1998.

[23] M. Soroushnejad and E. Generanoits, "Multiple-Access Strategies for an Integrated Voice/Data CDMA Packet Radio Networks", *IEEE Trans. Commun.*, vol. 43, no. 2, pp.934-945, Feb. 1995.

[24] H. van Roosmalen, J. Nilhof and R. Prasad, "Performance analysis of a Hybrid CDMA/ISMA Protocol for Indoor Wireless Computer Communications", *IEEE J. Select. Areas Commun.*, vol. 12, no. 5, pp. 909-916, June. 1994.

[25] F. L. and V. Grassi, "Markov analysis of the PRMA protocol for local wireless networks", *Wireless Networks*, vol. 4, no.4, pp. 297-306, Jan 1998.

[26] W. Yue and Y. Matsumoto, *Performance Analysis of Multiple Traffic on Wireless Communication Networks*, Kluwer Academic Publishers, London, 2002.

[27] M. Ma and E. Gunawan, "Single-code and multi-code DS/CDMA MAC protocols for video/voice integration", *International Journal of Computer Systems and Engineering*, vol. 17, no. 6, Nov. 2002.

[28] C. J. Chang and C. H. Wu, "Slot allocation for an integrated voice/data TDMA mobile radio system with finite population of buffered users", *IEEE Trans. Veh. Technol..*, vol. 43, no. 1, pp. 21-26, 1994.

[29] R. Fantacci and A. Naldi, "Performance evaluation of a CDMA protocol for voice and data integration in personal communication networks", *IEEE Trans. Veh. Technol..*, vol. 49, no. 2, pp. 307-320, Mar. 2000.

[30] G. Benelli, R. Fantacci, G. Giambene and C. Ortolan, "Performance Analysis of a PRMA Protocol Suitable for Voice and Data Transmissions in Low Earth Orbit Mobile Satellite Systems", *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp.156-168, Jan. 1995.

[31] M. Chatterjee and S. K. Das, "A Hybrid MAC Protocol for Multimedia Traffic in Wireless Networks", *The Proceedings of the IEEE International Conference on Networks*, 2000.

[32] D Zhao, X. Shen and J. W. Mark, "Radio Resource Management for Cellular CDMA Systems Supporting Heterogeneous Services", *IEEE Trans. Mobile Computing*, vol. 2, No 2, June 2003.

[33] V. Huang and W. Zhuang, "Optimal Resource Management in Packet-Switching TDD CDMA Systems", *IEEE Personal Communications*, Dec 2000.

[34] Chandraa, A, Gummalla, V, and Limb, JO (2000). "Wireless Medium Access Control Protocols," *IEEE Communications Surveys*, Second Quarter, 2000.

[35] X. Wang, "Wide-Band TD-CDMA MAC with Minimum-Power Allocation and Rate- and BER-Scheduling for Wireless Multimedia Networks", *IEEE Trans. Networking,* Vol. 12, no. 1, pp.103-116, Feb 2004.

[36] S. K Kang, "An Adaptive Channel Access Control Methods for CDMA/PRMA", *IEICE Trans. Commun.*, Vol. E86-B, no. 1, pp.3-15, Jan 2003.

[37] J. Sánchez, R. Martinez, and M. W. Marcellin, "A Survey of MAC Protocols Proposed for Wireless ATM", *IEEE Network*, pp. 52-62, Nov/Dec 1997.

[38] J. Y. Park, S. Y. Nam, and D. K. Sung, "Performance Analysis of an Enhanced DQRUMA/MC-CDMA Protocol for Voice Traffic", *IEEE Trans. Veh. Technol.*, Vol. 52, no. 2, pp. 440-434, March 2003.

[39] S. Elnoubi, and A. M. Alsayh, "A Packet Reservation Multiple Access (PRMA)-Based Algorithm for Multimedia Wireless System", *IEEE Trans. Veh. Technol.*, Vol. 53, no. 1, pp. 215-222, Jan 2004.

[40] M. J. Karol, Z. Liu, and K. Y. Eng, "An Efficient Demand-Assignment Multiple Access Protocol for Wireless Packet (ATM) Networks," *Wireless Networks, ACM-Press*, Baltzer Science Publishers, vol. 1.3, pp. 267-279, 1995.

[41] A. S. Acampora, and S.V. Krishnamurthy, "A New Adaptive MAC Layer Protocol for Wireless ATM Networks in Harsh Fading and Interference Environments," *Proc. ICUPC '97*, vol. 2, Oct. 1997, pp. 410-415.

[42] E. C. Kao, "A Dynamic Random Channel Reservation for MAC Protocol in Multimedia Wireless Networks", MSc Dissertation, Department of Computer Science, University of Canterbury, 2001

[43] V. V. Phan, S. G. Glisic, and D. D. Luong, "Packet-Length Adaptive CLSP/DS-CDMA: Performance in Burst-Error Correlated Fading Channels", *IEEE Trans. Wireless Commun.*, Vol. 3, No. 1, pp 147-157, Jan. 2004

[44] K. Toshimitsu, T. Yamazato, M. Katayama, and A. Ogawa, "A novel spread slotted ALOHA system with CLSP," *IEEE Select. Areas Commun.*, Vol. 12, pp. 665–672, May 1994.

[45] K. Sakakibara, T. Shimoi and J. Yamakita, "Analysis of frequency-hopping slotted ALOHA systems with channel load sensing in finite population model," *ECTI Transactions on Electrical Engineering, Electronics and Communications*, Vol.1, No.1, pp. 49-55, Aug. 2003.

[46] N. Abramson, "The Aloha system - Another Alternative for Computer Communications", *AFIPS Conference Proceedings and Fall joint Computer Conference,* Vol. 37, pp. 281-285, 1970.

[47] D. J. Goodman, R. Valenzuela, K. Gayliard, and B. Ramurthy, "Packet Reservation Multiple Access for Local Wireless Communications", *IEEE Transactions on Communications*, Vol. 37, pp. 885-890, 1989.

[48] L.G. Roberts, "Aloha Packet System with and without Slots and Capture", *ACM SIGCOMM, Computer Communication Review*, Vol 5, No. 2, Apr. 1975.

[49] L. Kleinrock and F. Tobagi, "Packet Switching in Radio Channels: Part I -- Carrier Sense Multiple Access Modes and Their Throughput-delay Characteristics", TRANSCOM COM-23(12):1400-1416, Dec.1975.

[50] F. A. Tobagi and V. B. Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection", *Computer Networks*, Vol. 4, No. 7, pp. 245-259, Oct.-Nov. 1980.

[51] G. Wu, K. Mukumoto, and A. Fukuda, "An Integrated Voice and Data Transmission System with Idle Signal Multiple Access – Dynamic Analysis", ICICE *Trans. Commun.*, Vol. E76-B, No. 11, Nov.1993.

[52] K. J. Zdunek and D. R. Ucci, "Throughput of Non-persistent Inhibit Sense Multiple Access with Capture", Electron Letter, Vol. 25, pp. 30-32, Jan. 1989.

[53] R. Prasad, "Performance analysis of mobile packet radio networks in real channels with inhibit-sense multiple access", *IEE Proceedings-I, Vol. 138*, No. 5, pp. 458-464, Oct. 1991.

[54] J. H. Lee and C. K. Un, "Performance analysis of Non-persistent Idle-signal Casting Multiple Access with Collision-Detection (ICMA/CD) Protocol", *IEE Proceedings-Communications*, Vol. 143, No.3, pp. 162-166, June 1996.

[55] F. A. Tobagi and l. Kleinrock, "Packet Switching in Radio Channels: Part II – The Hidden Terminal Problem in Carrier Sense Multiple Access Modes and the Busy-tone Solution" *IEEE Trans. Communication*, COM-23, pp. 1417-1433, 1975.

[56] Z. Zhang and A. Acampora, "Performance of a Modified Polling Strategy for Broadband Wireless Lans in a Harsh Fading Environment", Proceeding of IEEE GLOBECOM '91, pp. 1141-1146, 1991.

[57] R. J. Haines and A. H. Aghvami, "Indoor Radio Environment Considerations in Selecting a Media Access Control Protocol for Wideband Radio Data Communications", IEEE International Conference on Communications '93, Vol. 2, pp. 990-994, 1993.

[58] S. Nanda, D. J. Goodman, and U Timor, "Performance of PRMA: A packet Voice Protocol for Cellular Systems", *IEEE Trans. Veh. Technol.,* Vol. 40, No. 3, pp. 584-598, Aug. 1991.

[59] D. Petras, "Medium Access Control Protocol for Wireless, Transparent ATM access", *IEEE Wireless Communication Systems Symposium*, Long Island, NY, Nov. 1995.

[60] G. Colombo, L. Lenzini, E. Mingozzi, B. Cornaglia, and R. Santaniello, "Performance Evaluation of PRADOS: A Scheduling Algorithm for Traffic Integration in a Wireless ATM Network", *Proceedings of the fifth annual ACM/IEEE International conference on Mobile computing and networking (MobiCom'99)*, pp. 143-150, Aug. 1999.

[61] N. Passas, L. Merakos, D. Skyrianoglou, F. Bauchot, G. Marmigère, S. Decrauzat, "MAC protocol and traffic scheduling for wireless ATM networks", *Mobile Networks and Applications*, Vol. 3, no. 3, pp. 275-292, Sept. 1998.

[62] D. Raychaudhuri, L. J. French, R. J. Siracusa, S. K. Biswas, R. Yuan, P. Narasimhan, and C. A. Johnston, "WATMnet: A Prototype Wireless ATM System for Multimedia Personal Communication", *IEEE J. Select. Areas Commun.*, Vol. 15, No. 1, pp. 83-95, Jan. 1997.

[63] C. L. Fullmer and J. J. Garcia-Luna-Aceves, "Floor Acquisition Multiple Access (FAMA) for Packet-radio Networks", Proc. ACM SIGCOMM 95.

[64] P. Karn, "MACA – A New Channel Access Method for Packet Radio", *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, pp. 134-140, 1990.

[65] B. Crow et al., "IEEE 802.11: Wireless Local Area Networks", *IEEE Commun. Mag.*, Vol. 35, no. 9, pp. 116-126, Sept. 1997.

[66] L. S. Koh and M. T. Liu, "A Wireless Multiple Access Control Protocol for Voice-Data Integration," *Proc. Int'l. Conf. Parallel and Dist. Sys.*, Tokyo, Japan, pp. 206-13, June 1996,

[67] F. Khan and D. Zeghlache, "Analysis of Aggressive Reservation Multiple Access Schemes for Wireless PCS," *Proc. ICC '96*, Dallas, TX, pp. 1750-55, June 1996.

[68] J. G. Kim and I. Widjaja, "PRMA/DA: A New Media Access Control Protocol for Wireless ATM," *Proc. ICC '96*, Dallas, TX, pp. 240-44, June 1996.

[69] O. kubbar and H. T. Mouftah, "Multiple Access Control Protocols for Wireless ATM: Problems Definition and Design Objectives", *IEEE Commun. Mag.*, pp. 93-99, Nov. 1997.

[70] J. Capetenakis, "Generalized TDMA: The Multi-Accessing Tree Protocol", *IEEE Transactions on Communications COM-27*, no. 11, pp. 1476-1484, Oct. 1979.

[71] T. Madsen and R. Prasad, "Hybrid ISMA/FDMA Protocol for Wireless Packet Transmission", *IEEE PIMRC 2002*, Lisbon, Sept. 2002.

[72] H. Qi and R. Wyrwas, "Performance analysis of joint voice-data PRMA over PRMA random packet error channel" *IEEE Trans. Veh. Technol.*, Vol. 45, pp. 332-345, May 1996.

[73] P. Agrawal, T. Zhang, C. J. Sreenan, and J. Chen, "All-IP Wireless Networks", *IEEE J. Select. Areas Commun.*, Vol. 22, no. 4, pp. 613-616, May 2004.

[74] S. Choi, "QoS Guarantees in Wireless/Mobile Networks" Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, University of Michigan, 1999.

[75] L. Xu, X. Shen, and J. W. Mark, "Dynamic Bandwidth Allocation with Fair Scheduling for WCDMA Systems", IEEE *Commun. Mag.*, pp. 26-32, April 2002.

[76] H. Jiang, and W. Zhuang, "Quality-of-Service Provisioning in Future 4G CDMA Cellular Networks", IEEE *Commun. Mag.*, pp. 48-54, April 2004.

[77] V. Huang, and W. Zhuang, "QoS-Oriented Access Control for 4G Mobile Multimedia CDMA Communications", IEEE *Commun. Mag.*, pp. 118-125, March 2002.

[78] J. Heinanen , F Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group," RFC 2597, June 1999.

[79] V. Jacobson, et. al., "An Expedited Forwarding PHB group," IETF RFC 2598, June 1999.

[80] H. Haas and S. McLaughlin, "A Dynamic Channel Assignment Algorithm for a Hybrid TDMA/CDMA-TDD Interface Using the Novel TS-Opposing Technique", *IEEE J. Select. Areas Commun.,* vol 19, No. 10, October 2001.

[81] F. Hao, E. W Zegura, and M. H. Ammar, "QoS Routing for Anycast Communications: Motivation and an Architecture for DiffServ Networks", *IEEE Commun. Mag.*, vol. 40 No. 6, pp. 48-65, June 2002.

[82] ITU-T. Recommendation E.800: *Terms and definitions related to quality of service and network performance including dependability*, Aug. 1994.

[83] V. Upkar and J. Radhika, "Issues in Emerging 4G Wireless Networks", IEEE Computer, pp.94-96, June 2001.

[84] K. Wongthavarawat and A. Ganz, "IEEE 802.16 Based Last Miles Broadband Wireless Military Network with Quality of Service Support", *MILCOM 2003 - IEEE Military Communications Conference,* Vol. 22, No. 1, pp. 779-784, Oct. 2003.

[85] G. Chu, D. Wang, and S. Mei, "A QoS Achitecture for the MAC protocol of IEEE 802.16 BWA systems", *IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions*, Vol. 1, pp. 435- 439, July 2002.

[86] C. Erlung, R. B. Marks, K. L. Stanwood, and S. Wang, "IEEE Standard 802.16: A Technical Overview of the WirelessMAN^{TM} Air Interface for Broadband Wireless Access," *IEEE Commun. Mag.,* pp. 98-107, June 2002.

[87] IEEE P802.16-REVd/D5-2004, "Draft IEEE Standard for Local and Metropolitan Area Networks-Part 16: Air Interface for Fixed Broadband Wireless Systems", May 2004.

[88] G. Nair, *et al*, "IEEE Medium Access Control and Service Provisioning", Intel Technology Journal, Vol. 8, No. 3, Aug. 2004.

[89] R.K. Morrow and J.S. Lehnert, "Bit-to-bit error dependence in slotted DS/SSMA packet systems with random signature sequences", *IEEE Trans. Commun.* Vol. 37, No. 10, pp 1052 - 1061, Oct. 1989.