"There are two things which cannot be attacked in front: ignorance and narrow-mindedness. They can only be shaken by the simple development of the contrary qualities. They will not bear discussion."

Lord Acton (1834-1902).

*To all I have known,
and will know....*

**DISTRIBUTED RESOURCE ALLOCATION FOR
DS-CDMA BASED MULTIMEDIA WIRELESS LANs.**

by

Sanjay Lal

**A thesis submitted in conformity with the requirements for the degree of
Master of Applied Science.
Graduate Department of Electrical and Computer Engineering
University of Toronto**

© Copyright by Sanjay Lal, 1997.

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-29411-0

Canada

Next generation multi-media WLANs are expected to support a wide variety of applications with varying service requirements. It is the role of the *resource management* scheme to map the service requirements of the various applications to network resources so that the QoS requirements of the various users are met. i.e. Given certain QoS constraints, what resources does the application require from the network such that the QoS requirements are met? Power control in DS-CDMA systems and more recently power/rate allocation in multi-rate DS-CDMA based networks is an open and interesting research area which has attracted a lot of attention. However, with a few exceptions, most researchers have emphasized centralized resource allocation algorithms for cellular systems where the base station keeps track of the requirements of the various users and is thus responsible for the management of network resources. Adhoc WLANs, on the other hand, are generally configured as peer-to-peer networks with no centralized hub or controller. Thus resource allocation has to be conducted in a distributed fashion. We address the issue of *distributed* resource management for multi-rate DS-CDMA based multi-media WLANs by:

- Presenting a distributed resource allocation protocol known as *Distributed Resource Negotiation Protocol (DRNP)* which is geared towards negotiating QoS guarantees through receiver based distributed power control and bandwidth rationing in ad-hoc DS-CDMA based multi-media WLANs.
- Investigating the performance of three resource allocation policies (minimum resource, maximum rate, maximum SIR) within the context of DRNP, in terms of network wide metrics such as blocking and *loss of QoS* rates. Surprisingly, the minimization of resources scheme which emphasizes conservation of transmission powers and yields the highest capacities in cellular (more generally centralized) networks, turned out to be the worst overall performer.

II. Acknowledgements

I would like to thank my supervisor Prof. E. Sousa for all his help and invaluable guidance through the course of this project. I am also grateful to Prof. Katzela and Prof. Banerjea for their input relating to protocol design and validation.

I am also very grateful to my family and friends, especially Sanjeev Verma and Diane Silva, for their support during this long and sometimes tiring endeavor of which they were forced participants.

I would also like to thank the ITRC for their financial support during the course of this project.

I. Abstract	i
II. Acknowledgements	ii
III. Table Of Contents	iii
IV. List of Figures	vi
V. List of Tables	vii
1 Introduction	1
1.1 Evolving Wireless Networks	1
1.2 Spread-Spectrum and Code Division Multiple Access	2
1.2.1 Direct Sequence Spread Spectrum	3
1.2.2 DS-CDMA As a Multiple Access Mechanism	4
1.2.3 Power Control In DS-CDMA Systems	5
1.3 Wireless LANs (WLANs)	5
1.3.1 Wireless LAN Topologies	7
1.3.2 Wireless LAN Standards	8
1.3.3 IEEE 802.11 WLAN Standard	8
1.3.3.1 Network Allocation Vector (NAV)	9
1.4 The Future of WINs: The Integration Issue	10
1.4.1 Where do Wireless LANs fit in?	11
1.5 QoS Parameters	11
1.6 Research Objectives	12
1.7 Thesis Organization	13
2 DS-CDMA Based Multi-Media WLANs	14
2.1 Preamble	14
2.2 Asynchronous Multi-Processing Gain DS-CDMA	14
2.3 Indoor Radio Propagation Modelling	15
2.3.1 Large vs. Small Scale Propagation Models	16
2.3.2 Log-Distance Path Loss Model	16
2.3.3 Small-Scale Fading	17
2.4 Packet Trains	19
2.5 WLAN Topology	22
2.6 Spread Spectrum Modelling	23
2.7 Dual Receiver Architecture	25
2.8 The Common control channel (CCCH)	26
2.9 QoS Parameters	26

3 Distributed Resource Negotiation Protocol	29
3.1 Preamble	29
3.2 Global vs. Incremental Resource Management	29
3.3 Notation	30
3.4 Resource Allocation List (RAL)	31
3.5 Control Message Formats	32
3.5.1 Ready To Send (RTS): Start Session	32
3.5.2 ESR: End Session Request	33
3.5.3 Clear To Send (CTS)	33
3.5.4 ESA: End Session ACK	33
3.5.5 Primary REJect (PREJ)	34
3.5.6 Secondary REJect (SREJ)	34
3.5.7 Update Maximum Sustainable Interference (UPD_MSI)	36
3.6 Functional Specification	37
3.6.1 The STATE Variable	38
3.6.2 The Dispatch Process	39
3.6.3 The Transmitter Process	41
3.6.4 The Receiver Process	41
3.6.5 The 3rd Party Process	41
3.6.6 DRNP Walk-through	41
3.6.6.1 Transmitter: Session Setup	42
3.6.6.2 Receiver: Session Setup	43
3.6.6.3 3rd Party: Session Setup	43
3.6.6.4 Transmitter: Session Setup Continued	45
3.6.7 Session Tear-Down	45
3.6.7.1 Transmitter: Session Tear-Down	46
3.6.7.2 Receiver: Session Tear-Down	46
3.6.7.3 3rd Party: Session Tear-Down	46
3.7 Validation of DRNP	46
3.7.1 Protocol Validation Using Spin & Promela	47
3.8 Message Loss in DRNP	48
3.8.1 Session Setup	49
3.8.2 Session End	49
3.8.3 Update MSI	50
3.8.4 Secondary Reject	50
4 Resource Allocation Policies	52
4.1 Preamble	52
4.2 Maximum Sustainable Interference	52
4.3 Four Terminal Network	53
4.3.1 Static Network	54
4.3.1.1 Minimum Resource Allocation	54

4.3.1.3 Maximum SIR Allocation	55
4.3.2 Dynamic K-Terminal Network	56
4.4 Message Loss & Resource Allocation Policies	56
4.4.1 RTS/CTS (Session Setup)	57
4.4.2 Update MSI (UPD_MSI)	57
4.4.3 SREJ	57
4.4.4 ESR/ESA (Session End)	57
5 Simulation Model & Results	59
5.1 Preamble	59
5.2 Simulation Model	59
5.2.1 Radio Propagation Model	59
5.2.1.1 Spread Spectrum Modelling	60
5.2.2 Network Model	61
5.2.3 Modelling Traffic Characteristics	61
5.2.4 Traffic Models	61
5.2.4.1 Packet Switched Traffic Models	62
5.2.5 Session Parameters	62
5.2.5.1 Traffic Patterns	63
5.3 Performance Metrics	63
5.3.1 Blocking Rates	63
5.3.2 Loss of QoS Probabilities (QoSLoss)	63
5.4 Simulation Modes	64
5.4.1 <i>Genie</i> Mode	64
5.4.2 Real Mode	64
5.5 Simulation Runs	64
5.5.1 Genie Mode Simulation Results	65
5.5.1.1 Blocking Rates	65
5.5.2 Real Mode Simulation Results	68
6 Summary & Future Work	71
6.1 The DRNP Protocol	71
6.2 Resource Allocation Policies	72
6.3 Future Work	72
7 Appendix A: DS-CDMA as a Multiple Access Mechanism	74
8 Appendix B: Flow Diagrams for DRNP	76
9 References	81

Figure 1: Categories of Wireless Information Networks	1
Figure 2: DS-SS Transmitter-Receiver Structure	3
Figure 3: Types of Wireless LANs	7
Figure 4: Carrier Sense Multiple Access with Collision Avoidance.	8
Figure 5: Bandwidth Reservation (NAV) in 802.11.	9
Figure 6: Typical 802.11 Packet vs. a Packet Train	20
Figure 7: Typical Ad-Hoc Multimedia WLAN	22
Figure 8: Dual Receiver Architecture	25
Figure 9: Out of date <i>RAL</i> leads to an incorrect value for π to be advertised.	34
Figure 10: SREJ and UPD_MSI messages issued by 3rd party Receivers	36
Figure 11: Top Level Process Architecture	37
Figure 12: Session Setup: Transmitter sends RTS_{ij}	42
Figure 13: QoS can be supported: Receiver sends CTS_{ji}	43
Figure 14: Receiver sends $PREJ_{ji}$	43
Figure 15: RTS_{ij} message received by l	43
Figure 16: CTS_{ji} received by l	44
Figure 17: Terminal l issues MSI Update for $\{m.l\}$	44
Figure 18: QoS violation, l sends $SREJ_{li}$	44
Figure 19: Loss of RTS at the receiver/ CTS message at the transmitter.	49
Figure 20: A four terminal WLAN	53
Figure 21: Blocking Rates (No BackOff)	65
Figure 22: Average Maximum Transmission Power vs. Network Load	66
Figure 23: Blocking Rates (BackOff Count = 3)	66
Figure 24: Avg. Maximum Transmission Power (BackOff Count = 3)	67
Figure 25: % Loss of QoS: Deaf Transmitter	68
Figure 26: QoS Loss Rates as a function of τ_{DATA} size.	70
Figure 27: Dispatch Process Algorithm	76
Figure 28: Transmitter Algorithm as executed by Terminal i	77
Figure 29: Transmitter Algorithm (Session Tear-Down)	78
Figure 30: Receiver Algorithm as executed by Terminal j	79
Figure 31: 3rd Party Algorithm as executed by Terminal l	80

V. LIST OF TABLES

Table I: Wireless LAN Technologies	6
Table II: Wireless Multimedia Communications	11
Table III: DRNP Messages and Processes	37
Table IV: The STATE Variable	38
Table V: Dispatch Process Behavior	39
Table VI: Radio Propagation Model Parameters	59
Table VII: Spread Spectrum Simulation Parameters	60
Table VIII: Session Parameters.	62

1.1 Evolving Wireless Networks

The rapid and accelerating evolution of communications from wired to wireless networks is leading to a wide variety of new services, systems and products. Despite the wide diversity of these developments, the wireless evolution has proceeded along two main paths, which can be distinguished as voice-oriented and data-oriented Wireless Information Networks (WINs) as shown in Figure 1 [1].

Figure 1: Categories of Wireless Information Networks

Local-area voice-oriented services are provided using inexpensive, low power terminals such as cordless phones and wireless PBXs. Wide-area voice-oriented services are typified by higher-powered analog and digital cellular telephone services. Current wireless data communications are provided either by a low-speed data service over wide areas or by high-speed Wireless LANs (WLANs) for local applications¹. Presently, mobile data services such as ARDIS provide length-limited wireless connections to portable users in metropolitan areas with data rates limited to 19.2 Kbps [2]. Mobile data services are used for transaction processing, such as credit card

¹ Cellular Digital Packet Data (CDPD) seems to be an exception to this clean cut distinction of voice versus data oriented services. CDPD is designed to use the idle channels in current generation AMPS systems to transmit data at low data rates.

primarily intended as extensions to wired LAN interfaces such as ethernet and token-ring and are designed to provide relatively high data rates (on the order of a few megabits) to a small number of users in a well defined geographical area².

1.2 Spread-Spectrum and Code Division Multiple Access³

The term *Spread Spectrum* (SS), defines a class of digital radio systems in which the occupied bandwidth is considerably greater than the information rate. The technique was initially proposed for military as its Low Probability of Intercept (LPI) and anti-jam properties made it an attractive choice for covert communications. Only more recently has the civilian telecommunications community come to consider using spread-spectrum systems primarily due to their interference rejection and anti-multi-path properties [3],[4],[5].

The operation of spreading the spectrum of an information bearing stream is accomplished by combining it with a complex waveform having the desired spectral characteristics. This waveform may be:

- A series of pulses of the carrier at different frequencies, in a pre-determined pattern. This is known as Frequency Hopping Spread-Spectrum (FHSS).
- A pseudo-random modulating binary waveform whose symbol (or chip) rate is a large multiple of the data rate of the original bit-stream. This is known as Direct Sequence Spread-Spectrum (DSSS).

The ratio of the transmitted bandwidth to the bandwidth of the information bearing signal is called the processing gain of the SS system.

² Typical indoor/campus areas such as offices, warehouses as well as university campuses etc. are well suited to WLAN operation.

³ A brief overview of CDMA is a necessary pre-requisite for the following sections dealing with Spread-Spectrum based WLANs.

$$G = \frac{W}{R} \quad (1)$$

where W is the transmission bandwidth and R is the bandwidth of the information bearing signal.

The term *Code-Division Multiple Access* (CDMA) is often used in reference to SS systems and refers to the possibility of transmitting several such signals in the same portion of the spectrum by using distinct pseudo-random codes for each one. This can be achieved with either FH or DS spread spectrum and provides an alternative to Frequency or Time Division Multiple Access (FDMA/TDMA) methods. However, the term CDMA has come to be used mainly in relation to DS spread spectrum systems, which have attracted more attention than FH systems. Therefore, the remainder of this section is devoted to CDMA based on direct-sequence techniques.

1.2.1 Direct Sequence Spread Spectrum

Figure 2: DS-SS Transmitter-Receiver Structure

A Direct Sequence Spread Spectrum (DS-SS) system spreads the baseband data signal by directly multiplying the baseband data pulses with a Pseudo-Noise (PN) sequence. Figure 2, shows a functional block diagram of a DS system with binary phase modulation. The received SS signal can be represented as:

$$S_{ss}(t) = \sqrt{\frac{2E_s}{T_s}} d(t) c(t) \cos(2\pi f_c t + \theta) \quad (2)$$

where $d(t)$ is the data sequence, $c(t)$ is the PN spreading code sequence, f_c is the carrier frequency and θ is the carrier phase angle at $t = 0$. The data waveform is a time sequence of non-overlapping rectangular pulses, each of which has an amplitude of +1 or -1. Each symbol in $d(t)$ represents a data symbol with duration T_s . Each pulse in $c(t)$ represents a chip, and is usually rectangular with duration T_c . Assuming code synchronization has been achieved at the receiver, the received signal passes through the wideband filter and is multiplied by the local replica of the PN code sequence $c(t)$. If $c(t) = \pm 1$, then $c^2(t) = 1$, and this multiplication yields the de-spread signal $s(t)$ given by:

$$s(t) = \sqrt{\frac{2E_s}{T_s}} d(t) \cos(2\pi f_c t + \theta) \quad (3)$$

at the input to the demodulator, which then extracts $d(t)$.

1.2.2 DS-CDMA As a Multiple Access Mechanism

In contrast to TDMA or FDMA systems, all signals in DS-CDMA system share the same frequency band at the same time. Multiple access capability is achieved by allocating different spreading sequences or codes to various users. The signal of each user is separated or *de-spread* from the others at the receiver by using a correlator synchronized with the associated code sequence. Signals which do not match the selected correlator *de-spreading code sequence* are not de-spread and contribute as Multiple Access Interference (MAI) to the system. A detailed overview of DS-CDMA as a multiple access mechanism is presented in Appendix A.

Due to its inherent interference rejection capability, CDMA allows for different systems to operate simultaneously in a multi-vendor environment. Moreover, in contrast to TDMA or FDMA systems which require explicit co-operation between entities sharing the channel, once the spreading codes have been assigned, there is no need for explicit co-ordination among

for considering CDMA for indoor WLAN applications. The multi-path fading resistance property stems from the fact that the delayed versions of the transmitted Pseudo-Noise (PN) signal have poor correlation with the original PN sequence, and hence appear as uncorrelated noise (similar to the MAI caused by other users) and are simply rejected by the receiver. In fact, in cases where the relative propagation delays of the various paths exceed a chip period, CDMA systems can take advantage of multi-path fading in the form of path diversity. This is accomplished by using a Rake receiver which combines the information obtained from several resolvable multi-path components [6].

1.2.3 Power Control In DS-CDMA Systems

The need for power control is the achilles' heel for DS-CDMA systems. If a CDMA receiver receives a low-power signal from an intended, but remote, transmitter, and a higher-power noise signal from a nearer transmitter, it has a hard time recovering the intended signal. Unless nearby interferers reduce their transmission powers, they can completely shut out distant transmitters. This is known as the *near-far problem*. The solution is to require each transmitter to vary its power in inverse proportion to its distance from the intended receiver. If the path from the transmitter to the receiver is symmetric, then a simple power management technique is for the transmitter to ensure that the sum of its transmission and reception power is constant. This automatically ensures that when a receiver is far away, the transmit power is increased. This is known as *open-loop* power control. In asymmetric channel systems (cellular), open loop power control schemes prove to be inadequate as the forward and reverse links undergo uncorrelated fading. As such, *closed-loop* power control, where the base station transmits explicit power control information to the various mobiles is used.

1.3 Wireless LANs (WLANs)

The release of the unlicensed Industrial, Scientific and Medical (ISM) bands in 1984 prompted a significant interest in the design of numerous WLAN systems. Although these bands can be used for any purpose, WLANs were one of the primary applications that were envisioned in the

with a minimal spreading factor of 10 specified for Direct Sequence (DS) and Frequency Hopping (FH) spread spectrum systems. By 1990, WLAN systems using Direct Sequence and Frequency Hopped Spread Spectrum in the ISM bands, licensed radio at 18-19 GHz and Directed-Beam Infra-Red and Diffused Infra-Red technology appeared on the market.

Table I, summarizes the main WLAN technologies and their characteristics [7]. Diffused IR systems provide moderate data rates and coverage. They are suitable for moderate-sized offices, short-distance battery operated applications, and environments with significant amounts of radio interference. Directed-beam IR offers higher speed with reasonable coverage for applications

Table I: Wireless LAN Technologies

Technology/Feature	RF	RF(DS/FH) (IEEE 802.11)	Diffused IR (IEEE 802.11)	Directed Beam IR
Data Rate (Mbits/sec)	15	2-6	1.0	10
Range (ft.)	40-130	100-800	70-200	80
Frequency (Ghz) /Wavelength (nm)	f = 18	f = 0.9, 2.4, 5.7	λ = 800-900	λ = 800-900
Radiated Power	25 mW	< 1	N.A	N.A

with fixed terminals. It is suitable for large file-transfers between servers and clients as well as large open offices with many fixed terminals. WLANs in the licensed cellular frequencies are suitable for high speed communications in large partitioned areas, such as large offices and some libraries. The spread-spectrum systems provide the largest coverage, and are suitable for penetration through building floors and walls. It is suitable for small-business applications where a few terminals are distributed over several floors of a building.

There are two fundamental types of topologies used in wireless networks. They are (1) the centralized or hub and spoke and (2) the peer-to-peer network. In the centralized configuration, there is a central hub that controls all communication between stations in the network. Cellular systems are excellent examples of such a topology where the base station acts as the hub relaying all data between the mobiles. There are two variations of the peer-to-peer network topology. In the fully connected version, every terminal has the functional capability to communicate directly with any other terminal. In some systems, where terminals may be distributed over a wide area, any terminal may be able to reach only a portion of the other terminals in the network due to signal blockage or transmitter power limitations. In this case, terminals act as packet forwarding agents between widely separated Source-Destination (SD) pairs. Such networks are called multi-hop networks. In general, WLANs fall under the centralized or fully connected peer-to-peer category although only the latter are considered in this dissertation.

Figure 3: Types of Wireless LANs

WLANs can further be classified as (1) *ad-hoc* or (2) *infrastructure*. Ad-hoc WLANs are formed

configuring and are characterized by short set-up times and the lack of access to any wireline/wireless infrastructure. Infrastructure WLANs include special terminals known as *access points* which are interconnected via a typically wired distribution system. Infrastructure WLANs provide access to the wireline network infrastructure as well as a network for wireless terminals encompassing a greater area than would be possible in the ad-hoc case [8].

1.3.2 Wireless LAN Standards

The IEEE 802.11 committee was formed in 1990 as an independent standards group within IEEE 802 organization, specifically geared towards establishing a WLAN standard [9],[10]. In the European Community (EC), HIPERLAN standards activities for the physical and Medium Access (MAC) layers were initiated under the direction of the European Telecommunications Standards Institute (ETSI). In 1993 the EC announced allocations in the 5.2 and 17.1 Ghz bands for *HIPERLAN* [11][12].

1.3.3 IEEE 802.11 WLAN Standard

The IEEE 802.11 committee is currently developing WLAN standards for operation at data transmission rates of 1 and 2 Mbits/sec for Direct-Sequence and Frequency Hopped spread spectrum systems in the 2.4-2.485 Ghz ISM band as well as 1 Mbit/sec for Diffused IR systems. The standard addresses the problem of mutual interference in the unlicensed ISM bands by providing five overlapping sub-bands for DSSS operation and three hopping patterns for FHSS operation.

A considerable part of the 802.11 standard is devoted to medium access protocols and is common to all three physical layers. The fundamental medium access mechanism within the 802.11 specification is a distributed co-ordination function known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). A station wishing to transmit under CSMA/CA rules operates as shown in Figure 4. The medium is sensed to determine if another terminal is transmitting. If not, the transmission is allowed to proceed. Alternatively, if the medium is sensed busy, the

Figure 4: Carrier Sense Multiple Access with Collision Avoidance.

terminal wishing to transmit must defer until the end of the current transmission. After deferral, the station selects a random time interval, *Data Inter-Frame Space* (DIFS), and then randomly selects a time slot within the back-off window. If no other station starts transmitting before the selected slot is reached, the terminal is allowed to proceed with its transmission. Collisions can only occur if two or more terminals choose the same time slot for transmission. If another station selects an earlier time slot, the terminal in question freezes its back-off counter, waits for the end of this transmission. At the end of the transmission, the countdown is resumed.

1.3.3.1 Network Allocation Vector (NAV)

Carrier sensing in wireless networks is unreliable. This leads to the classic *hidden terminal* scenario [13], where two transmitters not within range of each other transmit to a receiver that is reachable by both. This leads to a collision at the receiver. To make matters worse, unlike wireline networks, the collision cannot be detected by the transmitters until they time out due to the lack of an acknowledgement from the receiver.

The RTS/CTS mechanism is a form of Virtual Carrier Sense (VCS) and is intended to mitigate the effects of the hidden terminal problem and is used in 802.11 when relatively large amounts of data need to be transmitted⁴. After gaining access to the medium and before starting

⁴ Only when large amounts of data need to be transmitted can the overhead caused by the RTS/CTS exchange be justified.

Figure 5: Bandwidth Reservation (NAV) in 802.11.

transmission of the data, a short control packet i.e. *Ready To Sent (RTS)*, is sent to the intended receiver announcing the upcoming transmission. The receiver acknowledges the receipt of the RTS packet by a *Clear To Send (CTS)* packet, and thus indicating its readiness to receive the data. Both packets contain the projected length of the transmission and thus inform all stations within radio-range of both the transmitter and the receiver the length of time for which the medium will be in use [14]. This reservation of network resources by the transmitter and receiver is maintained by the other stations as the *Network Allocation Vector (NAV)*. Figure 5 indicates the *NAV* for stations that hear the RTS frame, while other stations may only receive the CTS frame, resulting in a different *NAV*.

1.4 The Future of WINs: The Integration Issue

The driving force behind the development of Third-Generation WINs seems to parallel that of ATM networks, i.e., they are both geared towards the integration of a variety of communication services traditionally offered in disparate networks into a single networking structure. In the short run this is expected to involve the merger of first and second generation cordless, cellular services, and other low-tier data services such as paging. The integration of high speed data services, such as those offered by WLANs, is a long term goal. This broad concept of

1.4.1 Where do Wireless LANs fit in?

With the allocation of unlicensed PCS bands intended for wireless local data communications, the future of WLANs within the context of PCS seems assured. Fueled by the explosive growth of portable computers in the last few years, researchers are contemplating new concepts such as ad-hoc networking, nomadic access and mobile computing, leading to the fusion of computers and communications in a ubiquitous computing environment [16]. Moreover, with the shift towards integrated multi-media networks such as ATM in the wireline arena, WLANs in their traditional roles as extensions to the wired infrastructure are expected to bring, in the form of multi-media WLANs, the revolutionary capabilities of wireline multimedia networks such as ATM, i.e., the ability to efficiently integrate disparate applications with varying service requirements into a single networking infrastructure, to the wireless arena.

1.5 QoS Parameters

Next generation multi-media WINs are expected to support a wide variety of applications with varying service requirements. These service requirements are typically expressed through application dependent parameters for three quantities: *bandwidth, loss and delay*. [13]. Typical bandwidth parameters are the minimum bandwidth and the sustained

Table II: Wireless Multimedia Communications

Media	Bit-Rate	Bursty	Loss Rate
Audio	4-64 Kb/s	medium	$< 10^{-2}$
Video	$n \times 64$ Kb/s	low	$< 10^{-4}$
Image	> 10 Kbytes	high	~ 0
Data	variable	high	~ 0

bandwidth. Typical loss parameters include packet loss probability and a bound on the maximum number of packets lost over every consecutive set of packets transmitted. Although not considered in this dissertation, delay is characterized by mean delay, the worst-case delay, the

requirements for the various application classes expected to be supported by multi-media WINs. For instance, voice communication can be classified as a low rate service that can tolerate relatively high loss rates due to the inherent redundancy of speech signals. In contrast, data traffic requires very low probability of loss/error to ensure the preservation of data integrity.

It is the role of the resource management scheme to map the service requirements of the various applications to network resources so that the QoS requirements of the various users are met. i.e. Given certain QoS constraints, what resources does the application require from the network such that the QoS requirements are met? The number of research papers dedicated to answering this question in the context of ATM networks is a testament to the complexity of the problem [17],[18],[19],[20].

1.6 Research Objectives

In multi-media WINs, the notion of resource management is inexorably tied to the Medium Access (MAC) mechanism. For instance, in TDMA based systems, resource management involves the allocation of time slots, while transmitted power is the core resource in DS-CDMA systems. Thus, the issue of resource management must be addressed within the context of a multiple access mechanism. To this effect, we propose multi-rate DS-CDMA as the multiple access mechanism of choice for next generation multi-media WLANs. We assume that the service requirements are expressed in terms of minimum bandwidth and maximum packet loss rates and address the issue of resource management in such networks in terms of power control as well as transmission bandwidth allocation.

Power control in DS-CDMA systems [21],[22] and more recently power/rate allocation in multi-rate DS-CDMA based networks [23],[24] is an open and interesting research area which has attracted a lot of attention. However, with a few exceptions [25], most researchers have emphasized centralized resource allocation algorithms for cellular systems where the base station keeps track of the requirements of the various users and is thus responsible for the

management of network resources. WLANs, on the other hand, are generally configured as peer-to-peer networks with no centralized hub or controller. Thus resource allocation has to be conducted in a distributed fashion. We address the issue of *distributed* resource management for multi-rate DS-CDMA based multi-media WLANs by:

- Presenting a distributed resource allocation protocol that builds on the RTS/CTS bandwidth reservation mechanism provided by IEEE 802.11, and provides Quality of Service (QoS) guarantees through distributed control of resources in DS-CDMA based multi-media WLANs.
- Investigating the performance of various resource allocation schemes in terms of network wide metrics such as overall throughput and blocking rates.

1.7 Thesis Organization

A model of the proposed ad-hoc multi-media WLAN system is presented in Chapter 2 with an emphasis on multi-rate DS-CDMA as a medium access mechanism. Chapter 3 details the formal specification and validation of the Distributed Resource Negotiation protocol geared towards providing QoS guarantees through distributed power control and bandwidth rationing in ad-hoc DS-CDMA based multi-media WLAN environments. Resource allocation schemes are introduced in Chapter 4 followed by a qualitative discussion of important network performance parameters. Chapter 5 outlines the simulation model used to simulate DRNP as well as the resource allocation schemes discussed in Chapter 4. The results are summarized in Chapter 6 followed by future directions of research.

2.1 Preamble

We now present our model for an adhoc multi-rate DS-CDMA based multi-media WLAN. A novel aspect of the architecture is a Common Receiver (C-R) based code protocol coupled with a dual receiver architecture. There are two classes of logical channels; (1) the Common Control Channel (CCCH) is responsible for the transmission of control messages and (2) the Data Channel(s) (DCH) used for transmitting data from one terminal to another. We assume that a session is defined as the transmission of a long series of sequenced *mini-packets*. The QoS requirements for each session are expressed through the maximum Packet Error Rate (PER) and minimum data rate requirement. As with all multi-rate DS-CDMA networks, the transmission power and data rates are the core resources to be managed. We also define several quantities specific to the resource negotiation protocol to be presented in the next chapter.

2.2 Asynchronous Multi-Processing Gain DS-CDMA

The integration of the myriad of services expected to be supported by multi-media WLANs requires a multiple access mechanism that can:

- fulfill the role of traditional Medium Access (MAC) protocols of co-ordinating access to the channel and
- provide the flexibility to support services with a wide range of QoS requirements traffic characteristics.

A MAC protocol that has recently attracted a great deal of attention, especially in the context of Variable Bit Rate (VBR) services, is *asynchronous multi-rate DS-CDMA* [26]. Generally, spreading in a DS-CDMA system can be performed synchronously based on short periodic spreading codes or asynchronously in connection with long-period PN sequences. Although synchronous spreading allows the design of orthogonal or near-orthogonal signals sets if the data

requires careful code assignment. This is because short codes can maintain good cross-correlation properties as long as the data symbol length is equal for all signals that interfere with each other. This is not the case in multi-media networks where services with quite different rates have to be offered simultaneously in the same frequency band. Thus the applicability of the synchronous spreading techniques is quite limited in an essentially distributed and asynchronous environment such as an ad-hoc WLAN. On the other hand, asynchronous DS spreading with long PN codes, has certain advantages in the context of the third-generation multi-media systems:

- The number of available codes is virtually unlimited.
- The flexibility, with respect to multiple bit rates and variable spreading factors, is extremely high.
- There is no need for synchronized transmissions in order to maintain low cross-correlations between the spreading codes.

Although several asynchronous multi-rate CDMA schemes have been proposed [27], the scheme characterized by all users, regardless of their transmission data rate, using the entire system bandwidth is considered in this dissertation. Specifically, consider a terminal i , that generates an information bit stream of rate R_i . The information bits are spread by a Pseudo-Noise (PN) code sequence with N_i chips per information symbol to obtain a transmission bandwidth of W . Thus the processing gain for terminal i is given by:

$$G_i = \frac{W}{R_i} \quad (4)$$

This scheme supports multiple data rates by varying the processing gain. Obviously, terminals transmitting a lower bit rates will have higher processing gains.

2.3 Indoor Radio Propagation Modelling

Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction and scattering. However, conditions are much more variable. For instance, signal levels are

indoor radio propagation is relatively new, with the first wave of research occurring with respect to the 800 Mhz band in the early 1980s [28] and more recently with the ISM bands [29].

2.3.1 Large vs. Small Scale Propagation Models

Propagation models focus on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location. Propagation models that predict the mean signal strength of an arbitrary transmitter-receiver (T-R) separation distance are called *large-scale* propagation models. On the other hand, models that characterize the rapid fluctuations of the received signal strength over very short distances (a few wavelengths) or short time durations (on the order of seconds) are called *small-scale* or *fading* models. Indoor channels can be characterized by a combination of Log-Distance large scale path loss and slow fading.

2.3.2 Log-Distance Path Loss Model

In general, indoor channels may be classified either as Line-Of-Sight (LOS) or Obstructed (OBS) with varying degrees of clutter [30]. The signal power (from each terminal's transmission) attenuates due to a combination of power-of-distance law and lognormal shadow fading:

$$g(d) = L_0 + 10 \cdot \alpha_0 \cdot \log\left(\frac{\lambda}{4\pi d_{ref}}\right) + 10 \cdot \alpha \cdot \log\left(\frac{d}{d_{ref}}\right) + X_\sigma \quad (5)$$

where:

L_0 = antenna gain/loss.

λ = wavelength

d_{ref} = reference distance from transmitter

d = transmitter-receiver distance

α_0 = path loss exponent below d_{ref} , usually within Line Of Sight (LOS).

α = path loss exponent above d_{ref} , usually OBS (Obstructed).

X_σ = represents the shadow fading component and is a log-normal random variable with a

2.3.3 Small-Scale Fading

While large scale propagation is dependent mainly on T-R separation, the type of fading experienced by a signal propagating through a radio channel depends on the nature of the transmitted signal with respect to the characteristics of the channel. Depending on the relation between the signal parameters (such as bandwidth, symbol period etc.) and the channel parameters (such as rms delay spread and Doppler spread), different transmitted signals can undergo different types of fading. While multi-path delay spread leads to time dispersion and either flat or frequency selective fading, Doppler spread leads to frequency dispersion and time selective fading [31]. It is important to note that the two propagation mechanisms are independent of one another.

If the channel possesses a constant-gain and linear phase response over a bandwidth that is smaller than the bandwidth of the transmitted signal, the transmitted signal experiences *frequency selective fading*. Specifically, signal undergoes frequency selective fading if:

$$B_s > B_c \quad (6)$$

and

$$T_s < \sigma_\tau \quad (7)$$

where: B_s is the transmitted signal bandwidth, B_c is the coherence bandwidth of the channel, T_s is the symbol period and σ_τ is the rms delay spread of the channel. The coherence bandwidth of a channel is a statistical measure of the range of frequencies over which the channel can be considered *flat*. i.e. a channel which passes all spectral components with approximately equal gain and linear phase. If the coherence bandwidth is defined as the bandwidth over which the frequency correlation is above 0.5, then the coherence bandwidth is approximately:

$$B_c \approx \frac{1}{5\sigma_t} \quad (8)$$

If the radio channel is relatively flat over a bandwidth which is greater than the bandwidth of the transmitted signal, then the received signal undergoes *flat fading*. Specifically:

$$B_s < B_C \quad (9)$$

and

$$T_s > \sigma_t \quad (10)$$

Due to the highly variable nature of indoor channels, RMS delay spreads, although generally smaller than in urban environments, generally vary between 10 ns to 1.5 μ s. Thus, in narrowband networks, the bandwidth of the transmitted signal is much smaller than the coherence bandwidth of the indoor channel. Thus such networks experience flat fading. However, in CDMA based networks the transmitted signal bandwidth is relative high (20 - 50 MHz). Thus, depending on the environment, a broadband spread spectrum signal may experience frequency selective fading. In actual practice, most indoor channels are modelled as flat fading channels and spatial diversity techniques (such as multiple antennas) are used to combat such fading.

Depending on how rapidly the transmitted baseband signal changes as compared to the rate of change of the channel, a channel may be classified either as a *fast fading* or *slow fading* channel. The time varying nature of a channel can be characterized by parameters such as the *Doppler spread* and *coherence time*. The Doppler spread B_D , is a measure of the spectral broadening caused by the time rate of change of the radio channel and is defined as the range of frequencies over which the received Doppler Spectrum is non zero. In general, if the baseband signal bandwidth is much greater than B_D , the effects of Doppler spread are negligible at the receiver. The Coherence Time T_c , is a statistical measure of the time duration over which the channel impulse response is essentially invariant. Although various definitions of T_c are proposed, a popular rule of thumb for modern digital communications is to define it as follows [32]:

$$T_C = \sqrt{\frac{9}{16\pi f_m^2}} = \frac{0.423}{f_m}$$

where f_m is the maximum Doppler shift.

In a *fast fading* channel, the channel impulse response changes rapidly within the symbol duration; i.e. the coherence time of the channel is smaller than the symbol period of the transmitted signal. This causes frequency dispersion (also called *time selective fading*) due to Doppler spreading. In the frequency domain, signal distortion due to fast fading increases with increasing Doppler Spread relative to the bandwidth of the transmitted signal. In a *slow fading channel*, the channel impulse response changes at a rate much slower than the transmitted baseband signal. In this case, the channel may be assumed to be static over one or several reciprocal bandwidth (symbol period) intervals.

Generally, in WLANs the transmitter and receiver can be assumed to be stationary with the variations in the received signal attributed to the movement of people or other objects (Doppler shift: 2-5 Hz). Moreover, in DS-CDMA based WLANs, the baseband bandwidth is typically much greater the Doppler spread. The channel can thus be classified as a *quasi-static* or very slow fading channel. The quasi-static nature of the channel is used to our advantage through the packet-train concept outlined below.

2.4 Packet Trains

In packet switched networks there is a tradeoff between packet sizes and protocol efficiency. Let t be the length of the overhead (headers, trailers etc.) in bits and d the length of the data field or payload. Further, let a be the length on an acknowledgement, p_d the probability of a data message being distorted or lost and p_a the probability of the same error for an acknowledgement. Let us first consider the case where $p_d = p_a = 0$. Due to the relatively high error rates in wireless channels, the transmission of one message/packet requires one data message and one acknowledgement message, i.e. a total of $d + t + a$ bits. The overhead is $t + a$, and the protocol

$$E = \frac{d}{d+t+a} \quad (12)$$

Therefore, in the absence of errors it is best to chose d as large as possible. Now consider the case where p_d and p_a are non-zero. The probability that neither the data message nor its acknowledgement are in error is $(1 - p_d)(1 - p_a)$. Therefore, the probability that the message must be re-transmitted is:

$$p_r = 1 - (1-p_d)(1-p_a) \quad (13)$$

The probability that it takes i subsequent transmissions to get the data message across, i.e. $i-1$ retransmissions and one successful transmission, is

$$p_i = (1-p_r)p_r^{i-1} \quad (14)$$

The expected number of transmissions per message, R , is given by: [33]

$$R = \sum_{i=0}^{\infty} i p_i = \frac{1}{1 - p_r} \quad (15)$$

The relative efficiency E_{err} , in the presence of errors is then:

$$E_{err} = \frac{d}{R(d+t+a)} \quad (16)$$

Thus, in the presence of errors, we have to strike a balance: If d is too small the overhead due to the headers etc. is comparatively large and thus efficiency is low. Alternatively, if d is large, the efficiency is approximately $1/R$. This implies that the majority of the overhead is incurred by the re-transmission of large messages.

We therefore propose the concept of a *packet train*. Instead of a single packet, a packet train consists of a large number of sequenced *mini-packets*, each suffixed with a CRC for error detection. Figure 6 illustrates this concept. At the receiver, the CRC's are used to detect errors

Figure 6: Typical 802.11 Packet vs. a Packet Train

within each of the mini-packets and a single Negative-Acknowledgement (NACK) is sent back to the transmitter with the sequence numbers of the mini-packets that were received incorrectly. The transmitter then selectively re-transmits the mini-packets that were incorrectly received⁵. As opposed to IEEE 802.11, once a session has been set-up, packet-trains do not require synchronization, source and destination addresses that must accompany individual packets. Moreover, when errors do occur, only the mini-packets that are damaged need to be retransmitted. Hence, packet-trains allow us to reduce overhead by transmitting a relatively large payload, while avoiding the overhead caused by the retransmission of large packets by isolating the damage to a small portion of the payload.

A *session* is defined as the transmission of a packet-train from the transmitter to the receiver.

⁵ The selective repeat ARQ is used only in cases where the ARQ is feasible. For instance, for voice applications the mini-packets that were received incorrectly may just be discarded.

negotiation protocol, a packet train should be as long as possible. Realistically, however, packet train sizes are constrained by two factors by the need for fairness in the network. Stations wishing to transmit should have equal access to network resources.

- Since the resource allocation is performed on a per session basis, the QoS or SIR must be more or less constant for the duration of the packet train. Hence, the size of a packet train is limited by the rate of change within the channel. From section 2.3.3, we know that typical indoor channels with stationary terminals are characterized by very slow fading. Thus, depending on the transmission data rate, the packet trains can be quite long. In the case of mobile terminals, the packet trains will have to be shorter to compensate for the variation in path loss. In the limiting case, packet trains sizes may be on the order of a packet. i.e. resources may have to be negotiated on a per packet basis.

2.5 WLAN Topology

A single-hop, ad-hoc, asynchronous DS-CDMA based multi-media WLAN with k terminals is considered. K is defined as the set of terminals in the network as follows:

$$K = \{1, 2, 3, \dots, k\} \quad (17)$$

Figure 7, illustrates a typical adhoc multi-media WLAN with desktop as well as portable workstations providing voice and data services. The terminals are assumed to be distributed randomly and their positions are fixed. It is assumed that diversity techniques, such as multiple antennas, are used to compensate for the flat-slow fading of the indoor channel. Thus the received signal strength is influenced mainly by path loss and shadow fading. Moreover, we assume that the shadow fading is time invariant. Using (5), we define a path loss matrix as follows:

$$H = h_{ij} = \{g(d_{ij}), i, j \in K, i \neq j\} \quad (18)$$

where d_{ij} is the distance from terminal i to terminal j .

Figure 7: Typical Ad-Hoc Multimedia WLAN

2.6 Spread Spectrum Modelling

The chip rate for all terminals is fixed and the total bandwidth, W , is used by all terminals. Additionally, we assume that all supported transmission bit rates $R_1, R_2, R_3, \dots, R_n$ are multiples of the lowest rate, R_1 , and that the bit rates are restricted to $R_1 \times 2^{(i-1)}$ bits/sec ($i = 1, 2, \dots, n$), so that the processing gain varies by a factor of 2. It is assumed that there are a set of spreading codes that may be used by the terminals:

$$C = \{c_1, c_2, \dots, c_n\}, n > k \quad (19)$$

A number of issues arise when considering the use of spreading codes in a distributed CDMA based environment. For instance, (1) Is each terminal assigned a unique receiving or transmitting code? (2) Is there a common code used by all terminals for transmitting as well as receiving. Given a set of spreading codes there must be a protocol which dictates how they are utilized.

- *Common Code*: With this approach, a single spreading code is used by all terminals. Addressing information can be placed at the beginning of a packet allowing a terminal to determine if the packet is addressed to it. Since all transmissions use the same spreading code, this approach is essentially like ALOHA. Indeed, it would be exactly like ALOHA if the terminals were sufficiently synchronized and propagation delays were small.
- In the *Receiver Based Code* case, a unique receiving code is allocated to each terminal. This code is essentially an address. A terminal with a packet to send uses the destination's code to modulate the data stream. In the receiving mode, a terminal constantly monitors its code. The drawback to this scheme is that it is vulnerable to the hidden terminal problem as two or more transmitters may transmit data to the receiver simultaneously. This may cause the destruction of the packets.
- In the *Transmitter Based Code* scheme is similar to the receiver code scheme, except that in this case, a unique transmitting code is assigned to each terminal. However, the fundamental issue with this scheme is that the receiver has no way of anticipating a transmission addressed to it.
- *Hybrid schemes* involve various combinations of the schemes outlined above and can be classified as either *Common-Transmitter (C-T)* or *Common Receiver (C-R)* based. With the C-T code protocol there is a common code that is used by all terminals for addressing purposes in addition to the unique transmitting code assigned to each terminal. The common code is monitored when a terminal is idle. If a receiving terminal detects its address, it then switches to monitoring the transmit code for the data packet. The C-T protocol addresses the main issue with transmitter protocols, as the common code can be used to get the intended receiver's attention. The C-R scheme is similar in that a common code is also used. But instead of a unique transmitting code, individual receiving codes are assigned to each terminal.

When using asynchronous DS-CDMA based on long random PN codes, the question of which

requirements of the network. For instance, one could simply use a transmitter-based protocol and have each terminal that is in receiving mode monitor all possible codes. For the system being presented, a dual receiver architecture with the hybrid *Common/Transmitter (C-T)* or *Common/Receiver (C-R)* code protocols is considered [34],[35]. Although prone to packet loss through collisions the common code is essential to the operation of the resource negotiation protocol presented in the next chapter. It allows *all* terminals in the network to track network activity and thus implicitly gain knowledge about the state of the system. This is imperative if QoS guarantees are to be sustained. Moreover, although the dual receiver architecture does add additional complexity to the terminal hardware, it does allow terminals to track network activity while receiving. This again limits the loss of critical state information.

2.7 Dual Receiver Architecture

Since each code in a CDMA system may be thought of as a channel, the common PN code forms a single channel. This is called the *Common control channel (CCCH)*, since it is used mainly for control messages and is shared by all terminals. Similarly, each transmitter code, C_i , forms a *Data Channel*, denoted DCH_i .

Figure 8: Dual Receiver Architecture

Each terminal is equipped with two receivers and a transmitter. One receiver is always synchronized to the common code while the second receiver is synchronized to the unique receiver code assigned to each terminal. A terminal cannot transmit and receive simultaneously⁶. A terminal that is not transmitting,

⁶ This is true of all single channel radio networks. When a terminal is transmitting, any received signal is comparatively much weaker than the transmitted signal. Hence, the received signal is essentially blocked out.

a terminal to receive control and data messages simultaneously.

Specifically, let $K' \subset K$, denote the set of terminals that are currently transmitting. Then, $K^c = (K')^c$, is the set of terminals monitoring the common control channel. Obviously, $K' \cap K^c = \emptyset$. Moreover, $K'' \subset K^c$, defines the set of terminals that are currently receiving data frames. A session between terminals, $i \in K'$, and $j \in K''$, is denoted by $\{i,j\}$. Also, $K^a = K' \cup K''$, is defined as the set of currently active terminals. Moreover, it is assumed that a transmitter can communicate with only a single receiver at a time and vice versa.

2.8 The Common control channel (CCCH)

The CCCH is assumed to have a slotted structure with each slot equal to the transmission time of a control message. All control messages are assumed to be of the same size. The data rate on the CCCH is fixed. All messages are transmitted on the CCCH with the same power. i.e. no power control is used. Moreover, it is assumed that control messages are transmitted with a high enough transmission power so that they can be received by all terminals $\notin K'$. Since the CCCH is a single channel, it is vulnerable to collisions. Due to the relatively light traffic on the CCCH, simple multiple access scheme such as slotted-ALOHA is proposed. Moreover, due to the critical nature of the control messages it is assumed that some form of Forward Error Correction (FEC) is used for messages on the CCCH.

2.9 QoS Parameters

The QoS requirements for a session are expressed through the maximum Packet Error Rate (PER) (loss parameter) which can be mapped into an equivalent E_b/N_0 (SIR) requirement [36]. The matrix Γ is defined as follows:

$$\Gamma = \{\gamma_{ij}\}, i,j \in K, i \neq j$$

where γ_{ij} represents the minimum SIR requirement for $\{i,j\}$. The transmission data rate requirement for $\{i,j\}$ represented by the matrix Ψ , which defined as:

where ψ_{ij} is the minimum transmission data rate for $\{i,j\}$.

2.10 Resource Variables

The maximum physical transmission power for a terminal is Θ . The transmitted power allocated for a session $\{i,j\}$ is represented by the matrix $\mathbf{P}$ is defined as follows:

$$\mathbf{P} = \{p_{ij}\}, i,j \in K, i \neq j \quad (22)$$

where p_{ij} represents the transmitted power from user i to user j . The Rate matrix $\mathbf{R}$ is defined similarly:

$$\mathbf{R} = \{r_{ij}\}, i,j \in K, i \neq j \quad (23)$$

where r_{ij} represents the allocated data rate for the session between users i and j . Background Additive White Gaussian Noise (AWGN) with one-sided power spectral density η_0 is assumed. Using the Gaussian approximation for the Multiple Access Interference (MAI) [37],[38], the interference matrix, $\mathbf{Z}$, is defined as:

$$\zeta_{ij} = \left\{ \sum_{\{l,m\} \neq \{i,j\}} p_{lm} h_{ij} + \eta_0 W \right\} \quad (24)$$

where ζ_{ij} is the interference experienced by $\{i,j\}$. The Signal to Interference Ratio (SIR) matrix $\mathbf{\Omega}$ can be defined as:

$$\mathbf{\Omega} = \omega_{ij}(p_{ij}, r_{ij}, \zeta_{ij}) = \left\{ \frac{W h_{ij} p_{ij}}{r_{ij} \zeta_{ij}} \right\} \quad (25)$$

where ω_{ij} is the SIR for the $\{i,j\}$ session.

Given that $\{i,j\}$ has been allocated a SIR of $\omega_{ij} > \gamma_{ij}$, the additional interference that can be sustained by $\{i,j\}$, is denoted by δ_{ij} . We thus have:

$$\frac{W}{r_{ij}} \frac{p_{ij} h_{ij}}{(\zeta_{ij} + \delta_{ij})} = \gamma_{ij} \quad (26)$$

Solving for δ_{ij} , the Maximum Sustainable Interference (MSI) matrix, Δ , is defined as:

$$\Delta = \delta_{ij} = \left\{ \frac{W p_{ij} h_{ij}}{r_{ij} \gamma_{ij}} - \zeta_{ij} \right\} \quad (27)$$

where δ_{ij} represents the maximum additional interference that the receiver, j , of $\{i,j\}$, given that the interference is ζ_{ij} and $\{i,j\}$ has been assigned a transmission power of p_{ij} and a data rate of r_{ij} .

Using the model outlined above, we now present a distributed resource allocation protocol known as *Distributed Resource Negotiation Protocol (DRNP)* that can be used to allocate resources on a per session basis.

3.1 Preamble

This chapter presents the specification for the *Distributed Resource Negotiation Protocol (DRNP)* geared towards providing QoS guarantees through receiver⁷ based distributed power control and bandwidth rationing in ad-hoc DS-CDMA based multi-media WLANs. Moreover, the protocol also fulfills the role of Call Admission Control (CAC) mechanism, blocking those users that violate the QoS constraints of the currently active sessions.

3.2 Global vs. Incremental Resource Management

Resource management can be performed on a *global* or *incremental* (per session) basis. The *Global* RM entails renegotiation of resources each time a session is activated or leaves the system. In the incremental scheme, resources are allocated only *once* per session. This is analogous to the *packing* problem, where certain boxes of various sizes are to be fit into a container. Here the volume of the container is the resource that needs to be allocated. In the global scheme, boxes that are already packed can be moved around to make room for incoming boxes, while in the incremental scheme, boxes once packed are static. Incoming boxes then have to fit around the boxes already there. For instance, let's assume that $\{i,j\}$ has been allocated certain resources. Another session $\{m,l\}$ now requests resources from the network. In the global case, the network will allocate resources to $\{m,l\}$ as well as re-allocate the resources allotted to $\{i,j\}$ so that some optimization criterion (such as the minimization of total transmission power, or maximization of total throughput) is met. However, in the incremental allocation case, the network allocates resources to $\{m,l\}$, while preserving the resources allotted to $\{i,j\}$.

Global RM has the advantage of being extremely flexible as it is able to re-evaluate resource

⁷ In CDMA systems, the rate of success for a transmission is determined primarily by the interference levels at the receiver. In a distributed, homogenous network, the receiver is the obvious choice to determine whether interference levels are low enough to support the QoS requested by the transmitter. Hence, the notion of a *receiver* based resource allocation protocol.

there is a price to be paid for this flexibility. Global RM schemes are highly computationally intensive and entail a lot of protocol overhead in a distributed environment as they require that the entire state of the system be known at each decision interval. This makes global RM schemes well suited to centralized implementations, such as in cellular systems, where the base station can monitor all activity and make re-allocation decisions as required. There is another fundamental limitation to global allocation protocols. They require that a terminal be able to receive resource allocation messages while it is transmitting. This requires some form of duplexing⁸ and is impossible in the single channel architecture being proposed. Although less efficient than global allocation schemes, incremental schemes are better suited to single channel based distributed implementation. *DRNP* is one such incremental resource allocation scheme.

However, incremental schemes do have the tendency of being unfair. For instance, continuing the example above, once $\{i,j\}$ is active, the $\{m,l\}$ session is essentially constrained by the resources allocated to $\{i,j\}$. In certain situations, this can have seriously degrade the performance of the network. For instance, if $\{i,j\}$ is allocated the minimum required transmission power, the $\{l,m\}$ session will not be able to transmit, as the MAI introduced by $\{m,l\}$ will violate the QoS for $\{i,j\}$. This issue is analyzed in detail in the next chapter which deals specifically with resource allocation policies.

3.3 Notation

For convenience, the notation introduced previously is summarized below:

K : set of all terminals in the WLAN.

K^c : set of terminals scanning the *CCCH*.

K' : set of terminals currently transmitting on the *CCCH* or *DCH*.

K'' : set of terminals currently receiving data on the *DCH*.

⁸ There are two forms of duplexing, Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD). Duplexing includes isolation of the send and receive channels either by separating them in frequency (FDD) or in time (TDD).

K_{ij} : set of 3rd party terminals with respect to $\{i,j\}$.

K_{ij}^c : set of 3rd party terminals with respect to $\{i,j\}$ that can track the CCCH.

γ_{ij} : minimum SIR requirement for $\{i,j\}$.

ψ_{ij} : minimum data rate requirement for $\{i,j\}$.

h_{ij} : Path loss between terminal i and terminal j .

p_{ij} : Transmission power allocated to $\{i,j\}$.

r_{ij} : Data Rate allocated to $\{i,j\}$.

ω_{ij} : SIR for $\{i,j\}$.

δ_{ij} : Maximum sustainable interference for $\{i,j\}$.

3.4 Resource Allocation List (RAL)

The *Resource Allocation List* is an extension of the *Network Allocation Vector (NAV)* used in 802.11 WLANs and the *Power Constraint List (PCL)* introduced by Whitehead [39]. Each terminal maintains a database which encodes its knowledge about other ongoing sessions in the network. For illustration purposes, let $\{i,j\}$ be a currently active session. The set of *3rd party* terminals with respect to $\{i,j\}$ can be defined as:

$$K_{ij} = K - \{i\} - \{j\} \quad (28)$$

Let K_{ij}^c be the set of 3rd party terminals with respect to $\{i,j\}$ that can track the CCCH. A terminal $l \in (K_{ij}^c)$, encodes information about the $\{i,j\}$ session in its database RAL_l , as a record containing the following fields:

- The source (i) and destination (j) addresses for $\{i,j\}$. This can be thought of the index or key field of the database. All entries are accessed through this field.
- The estimated path loss to the source h_{ii} , and destination h_{ij} .
- The Maximum Sustainable Interference (MSI) for $\{i,j\}$, δ_{ij} ⁹.

⁹ This information is advertised by $\{j\}$ and is explained in the next section.

RAL_l contains a similar record for every active session in the network. Using h_{ij} and δ_{ij} , l can determine its maximum transmission power such that the interference at j does not exceed δ_{ij} . This is given by:

$$\mu_{ij} = \min \left\{ \frac{\delta_{ij}}{h_{ij}}, \Theta \right\}, i \in K^t, j \in K^r, l \in K_{ij}^c \quad (29)$$

We now define π_l as the maximum transmission power for l , such that the MSI constraints for all active sessions are preserved as follows:

$$\pi_l = \min_{\{x,y\} \in RAL_l} \{ \mu_{ly} \} \quad (30)$$

3.5 Control Message Formats

Control messages are broadcast on CCCH and are used to setup/tear-down sessions etc. The following convention is used in the presentation of the control message formats:

- Physical layer entities such as synchronization and error detection/correction fields have been excluded. Only the data structures specific to the DRNP are presented.
- Each message is prefixed by the source and destination addresses.
- Each message contains a unique entry identifying the message type and is denoted by the respective message name in the presentation of message formats that follows.

3.5.1 Ready To Send (RTS): Start Session

The *RTS* message is used by a transmitter, $i \in K^c$, to initiate a session with another terminal, $j \notin RAL_j$. Its format is as follows:

¹⁰ The receiver j , estimates the duration of the transmission from i , by using the size parameter in the *RTS* message and the allocated data rate. This is explained in detail in the following sections.

where:

i : is the source address

j : is the source address

RTS : is the unique message identifier

σ_{ij} : is the size (in octets) of the packet train to be transmitted.

3.5.2 ESR: End Session Request

The ESR message is used by the transmitter, i , to signify the end of a session.

$$END_{ij} = \{ i, j, END \} \quad (32)$$

3.5.3 Clear To Send (CTS)

The CTS message is issued by the recipient, j , of the RTS message if it can support the QoS requested by j within the data rate and power constraints. Its format is as follows:

$$CTS_{ji} = \{ j, i, CTS, \{ \gamma_{ij}, \psi_{ij} \} p_{ij}, r_{ij}, \delta_{ij}, \tau_{ij} \} \quad (33)$$

where:

p_{ij} : is the allocated transmission power

r_{ij} : is the allocated data rate

τ_{ij} : is the duration of the session.

3.5.4 ESA: End Session ACK

The ESA_{ji} message is sent to the transmitter j , by the receiver i , in response to the ESR_{ij} message.

Its format is as follows:

$$ESA_{ji} = \{ j, i, ESA \} \quad (34)$$

The *Primary REject (PREJ)* is issued by the receiver, j , if it is unable to support the QoS requested by i . The format of the $PREJ_{ji}$ message is as follows:

$$PREJ_{ji} = \{j, i, PREJ, \{\gamma_{ij}, \psi_{ij}\}, \pi_{ij}\} \quad (35)$$

3.5.6 Secondary REject (SREJ)

Recall that a terminal cannot transmit or receive simultaneously. This implies that a terminal is essentially *deaf* while it is transmitting. Thus, even if no other form of message loss occurs, a terminal's RAL will not be updated while it is transmitting. An out of date RAL can cause a receiver to allocate resources to a session that would degrade the QoS of the currently active sessions in the network. This scenario is illustrated in Figure 9, where i has out of data information. When it tries to establish a session

Figure 9: Out of date *RAL* leads to an incorrect value for π to be advertised.

with j , an incorrect value for π_i will be advertised in RTS_{ij} . The receiver j , has no way of validating π_i and may allocate a transmission power that causes the SIR for $\{m, l\}$ to fall below γ_{ml} .

The *Secondary REject (SREJ)* mechanism allows 3rd party receivers¹¹ to interrupt the setup of sessions that violate their QoS guarantees. Continuing the scenario outlined above, the dual receiver architecture allows l to track the CCCH while it is receiving data. If the MAI introduced by $\{i, j\}$ causes the SIR for $\{m, l\}$ to drop below γ_{ml} , l issues a $SREJ_{li}$ message as follows:

It is imperative that the *SREJ* message reach i before it starts transmission. To help accomplish

¹¹ As apposed to *PREJ*, where the intended receiver denies access to the network.

this, i waits for a certain number of slots τ_{DATA} , before initiating transmission. The 3rd party terminals wishing to transmit a *SREJ* to i , randomly choose a slot for transmission of the *SREJ* message¹². However, the *SREJ* mechanism does suffer from certain fundamental limitations:

- A *SREJ* message may arrive after the timeout period. This is a drawback for all timeout based synchronization schemes and is unavoidable.
- The *SREJ* messages are themselves vulnerable to loss through collisions with other traffic on the *CCCH*. Standard collision resolutions mechanisms involving ARQ are not feasible in this case due to the delay overhead they cause.

In principle, the probability that a *SREJ* message is lost can be minimized by making τ_{DATA} as large as possible. Realistically, the gain in terms of QoS guarantees is far outweighed by the overhead through delay caused by excessively large values for τ_{DATA} . Ideally, the timeout period should be varied with network load, although this again might entail too much computational overhead, although it does form an interesting avenue for research. The effect of message loss involving *SREJ* is investigated in Section 3.8.

The *SREJ* mechanism introduces another trade-off by requiring that an active receiver transmit a *SREJ*. This entails that the incoming data stream be disrupted for the duration of the transmission. Although, the damage is limited to a few mini-packets, depending on the state of the ongoing session, transmitting a *SREJ* may cause more harm than good. This is especially true if the session in question is near completion. *DRNP* in its current form uses a hard decision metric to decide whether a *SREJ* message should be sent out. Nevertheless, it is clear from the above discussion that *DRNP*'s performance can be improved by:

¹² Note that there is an implicit assumption here that transmission delay is much greater than propagation delay. This is quite valid in WLAN environments as terminals are confined to a limited geographical area.

errant session must be taken into account. If the drop in SIR is minimal, it may be preferable to allow the session in question to proceed.

- The status of the current data transfer. If near completion, the transmission of the SREJ may cause more damage than the incoming session.

3.5.7 Update Maximum Sustainable Interference (UPD_MSI)

Figure 10: SREJ and UPD_MSI messages issued by 3rd party Receivers

The Update Maximum Sustainable Interference (UPD_MSI) message is broadcast by all 3rd party receivers, $l \in (K'_{ij})$ when the CTS_{ij} message is received. This message updates the interference margin, δ_{ml} , where $m \in K'$, due to the addition of $\{i,j\}$ to the network. It's format is as follows:

$$UPD_MSI_{tx} = \{x, l, \{m, l\}, \delta_{ml}\} \quad (37)$$

where x is a network specific broadcast address.

Since UPD_MSI messages are transmitted by one or more 3rd party receivers, they are

SREJ is applicable here as well. In fact, UPD_MSI and SREJ messages are extremely similar. i.e. they are both used to update 3rd party RALs. Thus, in its current incarnation, the time window, τ_{DATA} , is shared by both SREJ and UPD_MSI messages. This is illustrated in Figure 10 where 3rd party terminals issue either a UPD_MSI or SREJ depending on whether their SIR falls below the negotiated minimum.

3.6 Functional Specification

Figure 11: Top Level Process Architecture

Each terminal executes an instance of 4 concurrent processes (1) *dispatch* (2) *transmitter*, (3) *receiver*, (4) *3rd Party* that communicate through the *RAL* and a shared *STATE* variable. Figure 11 is a block diagram showing the highest level of process interaction. All messages are received by the dispatch process and are routed to the appropriate processes as outlined in Table III.

Message	Comment	Handler Process
<i>RTS</i>	Another terminal wants to initiate a session	Receiver Process
<i>CTS/PREJ</i>	The intended receiver has accepted/denied the session request previously sent.	Transmitter Process
<i>SREJ</i>	A 3rd party receiver's QoS is going to be adversely affected.	Transmitter Process
3rd Party Messages <i>RTS/CTS/PREJ/SREJ/UPD_MSI</i>	All messages concerning all other activity in the network.	3rd Party Process
<i>ESR</i>	The transmitter has finished transmitting the data and is initiating session tear-down	Receiver Process
<i>ESR</i>	Response from the receiver to the <i>ESA</i> message previously sent.	Receiver Process

3.6.1 The STATE Variable

The *STATE* variable is a shared memory segment used as a form of *Inter Process Communication* (IPC) by all the processes. At any given time, it contains the current status of the terminal. Table IV summarizes the possible values taken on by the *STATE* variable. A terminal is deemed *IDLE* if it is neither transmitting or receiving data. The transmitter process sets the value of *STATE* to *TX* when it is triggered by the arrival of a packet train for transmission. This informs the dispatch process to ignore session setup requests from other terminals. After transmitting a *RTS* message, the transmitter process sets the *STATE* variable to *WAIT_FOR_CTS_PREJ*. At this stage the terminal is waiting for a response from the intended receiver. On receiving a *CTS* message, the terminal waits for τ_{DATA} slots before initiating transmission of data. At this stage, *STATE* is set to *WAIT_FOR_SREJ* by the transmitter process. Alternatively, if a *RTS* message

Value	Comment	Set by
<i>IDLE</i>	The terminal is tracking the <i>CCH</i> and is $\notin (K' \cup K')$	Transmitter process. Receiver process.
<i>TX</i>	The terminal has a packet train for transmission.	Transmitter Process
<i>WAIT_FOR_CTS_PREJ</i>	The terminal has issued a <i>RTS</i> message and is waiting for a response from the intended receiver.	Transmitter Process
<i>WAIT_FOR_SREJ</i>	The terminal has received a <i>CTS</i> message from this receiver and is waiting for a <i>SREJ</i> message from one or more 3rd party receivers.	Transmitter Process
<i>WAIT_FOR_ESA</i>	The terminal has finished transmission and has issued an <i>ESA</i> message.	Transmitter Process
<i>RX</i>	The terminal has received a <i>RTS</i> message.	Receiver Process

is received by the terminal while it is *IDLE*, the receiver process sets *STATE* to *RX*. This serves a dual purpose. As with the *TX* case, this causes the dispatch process to ignore session requests from other terminals. Moreover, it also informs the transmitter process that the terminal is currently in receiver mode.

3.6.2 The Dispatch Process

The *dispatch* process is essentially a message router that receives messages from other terminals, and depending on the state of the terminal, routes them to the appropriate processes. The operation of the dispatch process is outlined in Table V. The process operates in an infinite loop

Message	STATE	Action
<i>RTS</i>	<i>IDLE</i>	<i>STATE = RX</i> Execute Receiver Process
<i>CTS/PREJ</i>	<i>WAIT_FOR_CTS_PREJ</i>	Deliver message to Transmitter Process
<i>SREJ</i>	<i>WAIT_FOR_SREJ</i>	Deliver message to Transmitter Process
<i>ESA</i>	<i>WAIT_FOR_ESA</i>	Deliver message to Transmitter Process
3rd Party Message (<i>RTS/CTS/PREJ/SREJ/UPD_MSI</i>)	N.A.	Deliver message to 3rd Party Process.

as follows:

1. If *RTS* message received and *STATE = IDLE* then *STATE = RX*. Execute the Receiver Process with the *RTS* message as a parameter.
2. If a *CTS* or *PREJ* message is received and *STATE = WAIT_FOR_CTS_PREJ*, then deliver message to Transmitter Process.
3. If a *SREJ* message is received and *STATE = WAIT_FOR_SREJ*, then deliver message to transmitter process.
4. If a *ESA* message is received and *STATE = WAIT_FOR_ESA*, then deliver message to the transmitter process.
5. All 3rd party messages are routed to the appropriate handler routine within the 3rd party process.
6. All other messages (error conditions) are ignored.

The dispatch process is designed to relay information to the other processes only when they

manageable as these processes do not have to deal with error conditions caused by the arrival of an unexpected message. A flow chart depicting the dispatch process algorithm is presented in Figure 27 (Appendix II).

3.6.3 The Transmitter Process

The transmitter process is triggered by the arrival of a packet train for transmission. If the terminal is currently idle, i.e. $STATE = IDLE$, the transmitter process initiates session setup by issuing a RTS message to the receiver. If the receiver responds with a CTS message, the transmitter waits for τ_{DATA} slots before assuming that the session setup has been successful. If a SREJ message is received within τ_{DATA} , the transmitter goes into backoff mode and attempts to setup a session again at a later time.

3.6.4 The Receiver Process

The receiver process is triggered by the arrival of a RTS message. It is essentially responsible for determining whether the QoS requested by the transmitter can be supported. If it can, it issues a CTS message containing the allocated transmission power and data rate to the transmitter. Otherwise a PREJ message is sent to the transmitter.

3.6.5 The 3rd Party Process

The 3rd party process is the all purpose handler for all 3rd party messages and is responsible for updating the RAL as described in Section 3.4. It is also responsible for implementing the timeout mechanism for all 3rd party sessions contained in a terminal's RAL. Thus if no ESA is received within the timeout period, the corresponding 3rd party session is deemed complete and is removed from the RAL. This is necessary as loss of ESA messages can result in resources being allocated even when the sessions using them have left the system.

3.6.6 DRNP Walk-through

DRNP has 3 main phases of operation: (1) Session Setup, (2) Data Transmission/Reception, and

involves negotiation for resources by the transmitter. Once resources have been allocated, the protocol enters the data transmission/reception phase. This involves modulation of the data stream by the transmitter's spreading code and transmission of the data at the negotiated power and data rate. This phase of the protocol is followed by the session tear-down phase which is also initiated by the transmitter.

We now present the sequence of events involved in setting up a session. For illustration purposes, we assume that terminal $i \in K^c$ is attempting to establish a session with terminal j . Terminals $l, m \in K_{ij}$ are two 3rd party terminals with respect to $\{i, j\}$.

3.6.6.1 Transmitter: Session Setup

This is the first stage of the protocol and is illustrated in Figure 12. The *STATE* variable is set to *TX*. If the intended receiver, j , is perceived to be idle by the transmitter, i , it sends a request for resources (RTS_{ij}), sets *STATE* = *WAIT_FOR_CTS_PREJ* and waits for a response from j . The process is repeated if (1) a response (CTS_{ji} or $PREJ_{ji}$) is not received within a timeout period or (2) If RAL_i is updated while it is waiting for a response. This phase of the protocol is detailed below:

Figure 12: Session Setup: Transmitter sends RTS_{ij}

1. Set *STATE* = *TX*. Check RAL_i . If $j \in RAL_i$ wait for τ_{BUSY} and *Goto* 1. If not, then send $RTS_{ij} = \{RTS, \{\gamma_{ij}, \psi_{ij}\}, \pi_{ij}\}$ and set *STATE* = *WAIT_FOR_CTS_PREJ*.
2. Set $\tau_{CTS/PREJ}$ and wait for CTS_{ji} or $PREJ_{ji}$.
3. If RAL_i is updated, set *STATE* = *TX* and *Goto* 1.
4. If CTS_{ji} or $PREJ_{ji}$ message not received within $\tau_{CTS/PREJ}$, set *STATE* = *TX* and *Goto* 1.

a session with terminal j .

3.6.6.2 Receiver: Session Setup

On receiving the RTS_{ij} message, j determines whether it can support the QoS requested by i . If it can, it sends a CTS_{ji} message containing the allocated transmission power, p_{ij} , and data rate, r_{ij} , and maximum sustainable interference, δ_{ij} (Figure 13).

Alternatively, if the session cannot be supported, a $PREJ_{ji}$ message is sent (Figure 14).

Figure 30 (Appendix II) illustrates the receiver algorithm as executed by j for both cases outlined above. This phase of the protocol is detailed below:

1. Receive RTS_{ij} , set $STATE = RX$.
2. Extract γ_{ij} , ψ_{ij} and π_i .
3. Estimate the path loss h_{ij} based on the received power of the RTS message.
4. Estimate z_{ij} .
5. If $\omega_{ij}(\pi_{ij}, \psi_{ij}) \geq \gamma_{ij}$, then send $CTS_{ji} = \{p_{ij}, r_{ij} \text{ and } \delta_{ij}\}$ and monitor C_i .
6. If $\omega_{ij}(\pi_{ij}, \psi_{ij}) < \gamma_{ij}$, then send $PREJ_{ji}$ message.

Figure 13: QoS can be supported: Receiver sends CTS_{ji}

3.6.6.3 3rd Party: Session Setup

3rd party terminals monitor the $RTS/CTS/PREJ$ and ESR/ESA handshakes between i and j and use the information to update their RAL .

On receiving a RTS_{ij} message, a 3rd party terminal, $l \in K^c$, estimates the path loss, h_{il} , and adds the $\{i,j\}$ entry to RAL_l . Depending on whether a CTS_{ji} or $PREJ_{ji}$ message is received and the status of l , one of several things may happen:

Alternatively, if a CTS_{ij} message is received, p_{ij} , r_{ij} and δ_{ij} and τ_{ij} are extracted and added to RAL_l . If l is idle (i.e. not receiving), it will use the information obtained from the RTS/CTS handshake to update its maximum transmission power, π_l (Figure 16). Additionally, l will start a timer for $\{i,j\}$ set to τ

Figure 14: Receiver sends $PREJ_{ji}$

If l is currently receiving data from terminal m , it will check whether the additional MAI introduced by $\{i,j\}$ will cause ω_{ml} to drop below γ_{ml} . If not, then an UPD_MSI_l message containing the modified δ_{ml} value is issued. This scenario is illustrated in Figure 17. (1) $\{m,l\}$ is assumed to be setup and is denoted by the bold arrow from m to l . (2) When the CTS_{ij} message is issued, (3) l broadcasts the new value for δ_{ml} .

However, if ω_{ml} will drop below γ_{ml} , a $SREJ_{li}$ message, containing δ_{ml} is sent to i (Figure 8, Figure 18).

A breakdown of the 3rd party algorithm is presented below:

1. Receive RTS_{ij} message. Extract the source (i) and destination (j) addresses. Estimate the path gain, h_{il} .
2. If a $PREJ_{ji}$ message is received, the corresponding entry is deleted from the 3rd party terminal's RAL.
3. If a CTS_{ji} message is received, extract p_{ij} , r_{ij} and δ_{ij} and τ_{ij} . Start τ_{ij} timer. Upon expiry delete entry for $\{i,j\}$ from RAL_l . Also estimate h_{ij} .
4. If $l \notin K'$, calculate μ_{ij} and update π_l .
5. If $(l \in K') \wedge (p_{ij}h_{il} \leq \delta_{ml})$ where $m \in K'$, then broadcast $UPD_MSI = \{\delta_{ml} = \delta_{ml} - p_{ij}h_{il}\}$.
6. If $(l \in K') \wedge (p_{ij}h_{il} > \delta_{ml})$ where $m \in K'$, then pick slot 1 to τ_{DATA} and send $SREJ_{li} = \{\delta_{ml}\}$.

Figure 31 illustrates the 3rd party algorithm $\forall l \in K_{ij}$ for all cases outlined above.

into backoff mode, and tries to establish $\{i,j\}$ again after τ_{RANDOM} . Alternatively, if a CTS_{ji} message is received, i extracts the transmission power and data rate information. It then waits for a pre-determined number of slots (τ_{DATA}) before transmitting on DCH_i . If one or more $SREJ$ messages are received, they are used to update π_i . After the expiration of the τ_{DATA} timer, the $\{i,j\}$ session is assumed to be setup correctly. The details of the above scenarios are outlined below.

5. If a $PREJ_{ji}$ message is received, then set $STATE = TX$, wait for τ_{RANDOM} . Goto 1.
6. If a CTS_{ji} message is received, extract p_{ij} and r_{ij} .
7. Set $STATE = WAIT_FOR_SREJ$ and Wait for τ_{DATA} slots.
8. If $SREJ_{li}$ message received within τ_{DATA} , extract source (m) and destination (l) addresses and δ_{ml} . Estimate h_{li} and calculate μ_{li} and update π_i . Goto 5.
9. If τ_{DATA} expires, transmit data using C_i , p_{ij} and at r_{ij} bits/sec.

3.6.7 Session Tear-Down

The session tear down phase is a lot simpler than session setup. Similar to the setup phase, this involves a two way handshake between the transmitter and the receiver. Continuing the example above, we now assume that i has finished transmitting the packet train to j and is initiating session tear-down.

Figure 15: RTS_{ij} message received by l .

Figure 16: CTS_{ji} received by l .

Figure 18: QoS violation, l sends $SREJ_{li}$

2. Set τ_{ESA} , $STATE = WAIT_FOR_ESA$ and wait for ESA_{ji} .
3. If ESA_{ji} not received within τ_{ESA} , *Goto* 1.

3.6.7.2 Receiver: Session Tear-Down

1. Receive ESR_{ij} message.
2. Send ERA_{ji} message.

3.6.7.3 3rd Party: Session Tear-Down

1. Receive ESR_{ij} message.
2. Receive ESA_{ji} message.
3. Delete $\{i,j\}$ entry from RAL_p .
4. Update π_p .

3.7 Validation of DRNP

Protocols must be able to deal appropriately with every feasible action and with every possible sequence of actions under all possible conditions. This entails an absence of deadlocks, livelocks and improper terminations. The process of verifying the correctness of a protocol is known as *protocol validation*. This is accomplished either through manual proofs based on the notion of state invariants or by using automated validation systems based on exhaustive reachability analysis. Each method has its advantages and disadvantages. Unlike most automated validation systems, the manual proofs method is based on system invariants and the inspection of state transitions. There are usually far fewer state transitions than reachable system states. However, finding the right invariants can be hard. Moreover, manual proofs can be tedious and are inevitably susceptible to human error.

Although simpler from the point of view of protocol verification automated validation systems suffer from the state space explosion problem. For protocols of a realistic size, the number of reachable system states is usually too large for purely exhaustive analyses. A reachability

Figure 17: Terminal l issues MSI Update for $\{m,l\}$

reachable from a given initial state. There are three main types of reachability analysis algorithms:

- The *full search* is the simplest, brute-force algorithm and performs the most thorough analysis, but can only analyze the smallest class of protocols, i.e. systems up to 10^5 states.
- The *controlled partial search* tries to optimize the quality of the reachability analysis specifically for those cases where a full search is infeasible. It attempts this by selecting an optimal fraction of the full state space that can be searched with given constraints of memory and time.
- *Random simulation* techniques are specifically targeted towards the validation of systems of a complexity that defeats even the controlled partial search. The system state space for these systems can be estimated to be so large that no partial search technique can make a sensible selection. The best possible search in these cases is a random, or biased random, walk of the state space.

3.7.1 Protocol Validation Using Spin & Promela

Spin is a protocol modelling tool developed at Bell Labs [40] that utilizes the reachability analysis techniques discussed earlier to simulate communication protocols. A model of DRNP was created using the Promela language and simulated using the *Spin* package. The model can be used for conformance testing as well protocol validation using lossy channels. Initially, due to the relative simplicity of DRNP, the full state-space search algorithm was used. But, the state space proved to be quite large¹³ and the model ended up exhausting all available memory on the workstation. We thus turned to controlled partial search algorithm and concentrated our efforts on three critical phases of the protocol. This proved to be computationally efficient as well as extremely productive. The model showed that the protocol had the possibility of reaching a *no progress* state in two out of the three critical scenarios. These are outlined below:

- The session setup phase of the protocol is susceptible to no progress conditions. This may

¹³ It was surprising to discover that even a relatively simple protocol like DRNP has a state space size greater than 10^{10} , especially when dealing with error conditions.

the protocol to get *stuck* in the session setup phase. This situation can be alleviated by using an exponential backoff scheme combined with a maximum backoff count when a RAL update occurs. If the backoff count is exceeded the session is forced to leave the system.

- Another critical stage of the protocol is the distribution of UPD_MSI messages by the 3rd party receivers after a successful session setup. These messages are particularly prone to loss due to collisions. It is important to guarantee that loss of the update messages does not cause a cascade of QoS violations. i.e. the system should return to normal operation within finite time of a UPD_MSI message loss.
- Finally, there is a danger that the loss of ESA messages during session tear-down may cause resources to be allocated even after the respective sessions have terminated. Over time this can cause a situation where a session advertises a value of zero for the maximum sustainable interference. If the ESA for this session is lost, other session requests will be incessantly blocked. To solve this problem, the session timeout mechanism is maintained by all 3rd party terminals. If a ESA is not received within the timeout period, the respective session is assumed to have left the system and the resources released.

3.8 Message Loss in DRNP

Protocol validation is concerned with a protocol's ability to recover from error conditions and is not meant to be a performance metric. In distributed message based systems, message loss is the most common error condition. In the case of wireless systems, message loss is most often caused by collisions on the shared channel. In our case this is the *CCCH*. We have shown through exhaustive reachability analysis that DRNP can recover from all error conditions. However, message loss does lead to performance degradation and in some cases loss of QoS guarantees. This is unavoidable due to the asynchronous nature of the protocol. In this section, we qualitatively analyze the effect of control message loss on the performance of the network. Although, this is highly resource allocation policy dependent, it is instructive to at least isolate the most troublesome scenarios. In DRNP message loss effects the 3rd party terminals as well

3.8.1 Session Setup

The session setup phase involves a two way handshake (*RTS/CTS*). The loss of either a *RTS* or *CTS* message is detected at the transmitter through a timeout mechanism ($\tau_{CTS/PREJ}$). Since an exponential backoff scheme is used, each loss of either message leads to exponentially increasing delays in session setup. This scenario is depicted in Figure 19.

Figure 19: Loss of *RTS* at the receiver/*CTS* message at the transmitter.

As far as the 3rd party terminals are concerned, even if they fail to receive the *RTS* message sent out by the transmitter, as long as they successfully receive the *CTS* message, they will be able to update their respective maximum transmission powers. However, if a 3rd party terminal is a 3rd party receiver and fails to receive the *RTS* message, it cannot determine the path loss from itself to the transmitter. This entails that the 3rd party terminal cannot determine if the session being setup will cause its SIR to drop below the negotiated minimum. In this case the effected 3rd party receivers will opt to issue a *SREJ* to delay session setup and update the offending transmitter's *RAL*. However, the tradeoffs regarding *SREJ* as discussed in Section 3.5.6, 3.6.6.3 apply here as well.

3.8.2 Session End

The Session End scenario also involves a two way handshake (*ESR/ESA*) and a timeout mechanism similar to the one used in session setup (τ_{ESA}) is used to detect and recover from the loss of either message. However, in the case of 3rd party terminals, loss of an *ESA* message may result in *ghost* entries in their respective *RALs* for sessions that have terminated. Depending on the resource allocation policy, these ghost entries may cause sessions to be rejected even if sufficient resources are available. The timeout mechanism described in 3.7.1, is designed to solve this possible deadlock problem. Here, a 3rd party terminal sets a timeout value that is

3.8.3 Update MSI

An *UPD_MSI* is issued by active 3rd party receivers when a new session is activated. These messages are particularly susceptible to collisions as they are issued by receivers just after a CTS message is issued. The loss of an MSI update entails that certain 3rd party terminals may not have the correct information in their RAL. Therefore, if and when they decide to initiate a session request, the RTS message will contain the incorrect value for π . The Secondary Reject mechanism is designed to minimize the possibility that a session advertising an incorrect value for π will be successfully setup.

3.8.4 Secondary Reject

As stated earlier, message loss in DRNP is caused by (1) collision on the CCCH and (2) The *SREJ* mechanism is the catch-all for the messages that are lost and is used by 3rd party receivers to update incorrect RALs. But it too is prone to loss through collisions. After a CTS message has been received, the t for a pre-determined period for any SREJ messages. Thus a 3rd party receiver has a certain window of time to transmit a SREJ if its QoS is adversely affected by the new session. SREJ messages may not be issued within the time frame or may be lost (if two or more terminals choose the same slot). In this case, a session that violates the QoS of an active session may be successfully setup and thus violate one of the tenants of DRNP. A qualitative effects of SREJ message along with the trade-offs involved is presented in Section 3.6.3¹⁴.

The presentation of DRNP thus far has focussed on negotiation of resources. Nothing has been said about how the receiver decides how much power and rate to allocate to a particular session. This isolation of negotiation protocol and resource allocation policy is a core design goal of DRNP. This allows us to investigate the performance of DRNP using a myriad of resource management schemes. Thus, as better policies are developed, they can be *plugged* into DRNP

¹⁴ The effect of the loss of MSI Update and SREJ messages on the performance of DRNP and various resource allocation schemes is investigated through simulations. The results are presented in Chapter 6.

4.1 Preamble

In the resource negotiation protocol presented in the previous chapter, it is up to the receiver to decide whether (1) It can support the QoS requested by the transmitter and (2) If the QoS for the session can be supported, what resources in terms of transmission power and data rate should it allocate to the session? The second issue is addressed in this chapter by considering the performance of several resource allocation schemes.

Performance analysis for spread-spectrum based networks is usually much more complex than similar narrowband networks due to the possibility of multiple simultaneous transmissions. As such, most analysis is carried out through simulation and even then some fundamental assumptions regarding the statistics of the MAI¹⁵ have to be made. Although, the performance of DRNP has been analyzed through simulation, it is instructive to qualitatively investigate the critical DRNP performance parameters.

4.2 Maximum Sustainable Interference

The Maximum Sustainable Interference (MSI) is a local parameter advertised by a session, that effects the global performance of the network. In general, the MSI for a session $\{i,j\}$, δ_{ij} , is given by:

$$\delta_{ij} = \frac{W p_{ij} h_{ij}}{r_{ij} \gamma_{ij}} - \sum_{\{m,l\} \neq \{i,j\}} p_{ml} h_{mj} - \eta_0 W \quad (38)$$

As can be seen from (38), δ_{ij} is a function of the SIR allocated for $\{i,j\}$ as well as the current interference levels at j . Moreover, the MSI advertised by a particular session determines the maximum transmission power of other terminals. Hence, the MSI reflects the fact that both transmission power and interference levels within the network effect the chances that a session setup will succeed.

¹⁵ The Gaussian Approximation is one of the most common assumptions made about the MAI.

terminals at a certain rate. Moreover, sessions that are successfully setup *never* leave the system. For illustration purposes let $\{i,j\}$ be the first session to be setup. This session advertises a non-zero value for δ_{ij} and since there are no other sessions, the network is essentially limited by the AWGN in the network. As other sessions are setup, the MAI at j (the second term in (38)) increases until:

$$\frac{Wp_{ij}h_{ij}}{r_{ij}\gamma_{ij}} = \sum_{\{m,l\} \neq \{i,j\}} p_{ml}h_{mj} + \eta_0 W \quad (39)$$

At this point, the MSI advertised by $\{i,j\}$ will be zero. Since no sessions ever leave the network, all other sessions arriving later will be blocked. Although unrealistic, the above scenario does illustrate strong correlation between the MSI parameter and global network performance. We now turn our attention to the relationship between MSI and resource allocation policies. In doing so we hope to gain insight into the effect that various resource allocation schemes have on DRNP performance.

4.3 Four Terminal Network

For illustration purposes, consider a four terminal network Figure 20. It is assumed that:

- All of the sessions in the four terminal network are of the same type. They all have a minimum SIR requirement of γ and minimum data rate requirement of ψ .
- Moreover, $K' = \{i\}, \{l\}$ and $K'' = \{j\}, \{m\}$. $\{i,j\}$ and $\{l,m\}$ are the only S-D pairs in the network.
- The system bandwidth is W and the Background Additive White Gaussian

Figure 20: A four terminal WLAN

4.3.1 Static Network

We consider a *static* network which is a snapshot of the system where:

- $\{i,j\}$ is an ongoing session and has been allocated a transmission power of p_{ij} and data rate r_{ij} by a resource allocation scheme.
- As part of the session setup process, terminal j has advertised a maximum tolerable interference level of δ_{ij} . This too is resource allocation policy dependent. As a result, terminals $l \in K_{ij}$ and $m \in K_{ij}$, have set their maximum transmission powers to π_l and π_m respectively.
- Terminal m wants to establish a session with terminal l .

4.3.1.1 Minimum Resource Allocation

The minimum power allocation scheme is one of the most obvious, and widely utilized resource allocation schemes (especially in centralized networks). This scheme emphasizes conservation of transmission power and only the minimum transmission power required to achieve γ_{ij} and ψ_{ij} is allocated and is, in general, given by:

$$p_{ij}^{\min} = \frac{\gamma_{ij} \psi_{ij} \left(\sum_{(m,l) \neq (i,j)} h_{mj} p_{ml} + \eta_0 W \right)}{W h_{ij}} \quad (40)$$

Since only the minimum required *SIR* is achieved at the receiver, the Maximum Sustainable Interference (MSI) advertised by $\{i,j\}$, δ_{ij} , will be zero. Thus, $\mu_{ml} = \pi_{ml} = 0$. As a result $\{m,l\}$ will not be successful for the duration of $\{i,j\}$. Except for the explicit QoS guarantee, the minimum power allocation scheme functions exactly like the RTS/CTS bandwidth reservation mechanism in 802.11. It essentially reserves network resources for the $\{i,j\}$ session and prevents any other sessions from being established.

Moreover, since only the minimum data rate is allocated, for a given payload size, the transmission time for $\{i,j\}$ is maximized. This implies that not only will the incoming $\{m,l\}$ session be blocked for the *duration* of $\{i,j\}$, but that the blocking duration will also be maximized. The above discussion implies that performance may be improved by either:

communicate while still satisfying the MSI constraint imposed by $\{i,j\}$.

4.3.1.2 Maximum Rate Allocation

The maximum rate allocation scheme allocates, if possible, higher data rates than the minimum required at the expense of additional MAI within the network. The Maximum Rate scheme is similar to the Minimum Power scheme in the sense that only the minimum required SIR is allocated to any session. Any extra transmitted power, over and above that required to support the minimum data rate and SIR is allocated to increase the data rate of the session which is given by:

$$r_{ij} = \frac{Wh_{ij}\pi_{ij}}{\gamma_{ij} \left(\sum_{\{m,l\} \neq \{i,j\}} h_{mj}P_{ml} + \eta_0 W \right)} \quad (41)$$

The effect of the $\{i,j\}$ session on the static network as a whole is also similar. Since, δ_{ij} is set to zero, then as in the previous case, the network will allow only one session to be active at a time.

However, since a higher data rate is allocated to the $\{i,j\}$ session, for a given session size, the duration of the $\{i,j\}$ transmission will be less than in the Minimum Power allocation case. Hence, the fact that this scheme introduces additional MAI into the network is inconsequential as it is the MAI advertised by $\{i,j\}$, δ_{ij} , that is the limiting factor. i.e. no other session will be successful as the transmission power of the other terminals is constrained to zero, irrespective of the MAI.

4.3.1.3 Maximum SIR Allocation

In the maximization of SIR allocation policy, any extra power ration is used to increase the QoS of the session. The SIR at the receiver, ω_{ij} , can be maximized by allocating a transmission power of π_{ij} and data rate, ψ_{ij} , and is then given by:

As opposed to the minimum power or maximum rate policy, δ_{ij} , and as a result π_m , is may be

$$\Psi_{ij} \left(\sum_{(m,l) \in (i,j)} n_{mj} P_{ml} + n_0 \nu \right)$$

non-zero. Thus, if terminals l and m are sufficiently separated from terminals i and j ¹⁶ respectively, then $\{m,l\}$ may be successfully setup. The maximum SIR allocation scheme embodies the essence of CDMA, (i.e. the possibility of multiple simultaneous transmissions), while preserving the QoS guarantees to each active session. Moreover, although ω_{ij} is always guaranteed to be greater than γ_{ij} , the introduction of $\{m,l\}$ into the network causes ω_{ij} to decrease. The variability in the QoS for currently active sessions is the price to be paid for the increased network throughput

4.3.2 Dynamic K-Terminal Network

The discussion presented previously in the context of the four terminal network can be readily extended to a general K-terminal where session requests arrive at a certain rate. Here too, network throughput is strongly correlated to the MSI advertised by each successful setup or equivalently the maximum transmission power allocated to an incoming session. Over the long term, network throughput is dependent on the average MSI advertised by each session. This *local* parameter has global ramifications as it determines the transmission powers for all sessions arriving within the duration of the session in question. We investigate the effect of average MSI on network throughput in detail in the simulations and the results presented in Chapter 5.

4.4 Message Loss & Resource Allocation Policies

We now continue our discussion in Section 3.8, relating to message loss in DRNP. As previously stated, the effect of control message loss on DRNP is dependent on the resource allocation policy. We now extend the arguments provided earlier in the context of the resource allocation policies presented earlier in this chapter.

¹⁶ Note that the success of a session setup is affected by both the maximum transmission power of the transmitter as well as the interference levels at the receiver.

minimum power and maximum rate schemes and the same arguments from 3.8.1 still apply. However 3rd party receivers are only possible with the maximum SIR scheme. In this case, if a CTS message is received without a corresponding RTS message, it is up to the SREJ mechanism to delay the setup of the offending session.

4.4.2 Update MSI (UPD_MSI)

UPD_MSI messages are critical in maintaining QoS guarantees in DRNP, but are only necessary when the maximum SIR policy is used. In the minimum power and maximum rate case, the MSI advertised by the CTS message during session setup is zero and is never updated.

4.4.3 SREJ

As with UPD_MSI messages, SREJ messages are expected to be issued by 3rd party receivers with higher frequency when the maximum SIR allocation scheme is used. When minimum power and maximum rate schemes are used, the only case where a SREJ will be issued is if one or more 3rd party terminals fail to receive the RTS/CTS exchange during session setup of a particular session. Even in this case, only a single SREJ needs to be issued (no UPD_MSI messages) and thus the possibility of it colliding with another message are extremely unlikely. Moreover, since the only likely candidates for collision with the SREJ message are session setup/teardown messages, QoS guarantees are still maintained as the incoming session setup is interrupted by the collision.

4.4.4 ESR/ESA (Session End)

Contrary to the discussion above, loss of session tear down encompass one scenario where the minimum power/maximum rate schemes may actually be worse than the maximum SIR scheme. Since the former advertise a MSI of zero, loss of the ESR/ESA messages by the 3rd party terminals essentially blocks off any incoming sessions, until the respective time-outs occur. Although the above scenario is also possible with the maximum SIR case, it is certainly less frequent.

can fluctuate as sessions arrive and leave the network. Moreover, due to the presence of multiple sessions, there is a higher risk of losing control messages due to collisions on the CCCH. In the case of maximum SIR resource allocation, this can lead to loss of QoS guarantees. The extent of this phenomenon is investigated in the next chapter through simulations.

5.1 Preamble

As stated earlier the complexity of analyzing the performance of CDMA networks is compounded by the possibility of multiple simultaneous transmissions. We use simulations as a valuable tool to gain insight into the performance of DRNP in a DS-CDMA WLAN environment. Specifically, session level simulations are used to:

- Verify the feasibility of DRNP in a realistic WLAN environment.
- Study the effect of message loss on the performance of DRNP.

5.2 Simulation Model

Modelling of wireless networks essentially falls into two categories; (1) simulating the physical radio channel which in our case also includes spread spectrum modelling and (2) Network level modelling which involves modelling network topology as well as traffic patterns and traffic characteristics.

5.2.1 Radio Propagation Model

The combined log-distance path and shadow fading model presented earlier is used to model large scale signal variations and is re-stated below:

$$g(d) = L_0 + 10 \cdot \alpha_0 \cdot \log\left(\frac{\lambda}{4\pi d_{ref}}\right) + 10 \cdot \alpha \cdot \log\left(\frac{d}{d_{ref}}\right) + X_\sigma \quad (43)$$

The values of the above parameters used in the simulation are presented in Table VI. We assume that small scale signal variations can be essentially ignored due to the slow fading quasi-stationary nature of the indoor channel. A carrier frequency of 2.4 GHz is chosen as most current WLANs operate in this unlicensed ISM band that offers a total bandwidth of approximately 80 MHz. Moreover, it is assumed that LOS is present for T-R separations that are less than 10 m. As such, a path loss exponent of 2 is chosen for T-R separations within 10 m. For T-R separations greater than 10 m, an obstructed channel is assumed with a path loss exponent of 3. In addition, the shadow fading component is represented by a lognormal random

Parameter	Value
L_0	-2 dB
f_c	2.4 GHz, $\lambda = 125$ mm
d_{ref}	10 m
α_0	2
α	3
X_σ	4 dB

variable with a standard deviation of 4 dB.

5.2.1.1 Spread Spectrum Modelling

The system bandwidth W , is set to 50 MHz. Since the maximum possible data rate is 1 Mbit/sec Kbits/sec the minimum processing gain is therefore 250. Under these circumstances the standard gaussian approximation is quite adequate. Moreover, it is important to note that DRNP is essentially independent of the model chosen. This is because, in an actual WLAN, DRNP will rely on empirical estimates of the MAI.

However, if more precise modelling of DRNP performance is required (especially in realistic WLAN environments), other approximations such as the improved gaussian approximation may be more appropriate¹⁷.

Table VII: Spread Spectrum Simulation Parameters

SS Parameter	Value
W	50 MHz
Max Data Rate	1 MHz
Max Tx Power (Θ)	1000.0

Also note that the physical constraint on the maximum transmission power, Θ , is set to an

¹⁷ The computational complexity of using the improved Gaussian approximation, especially in long simulations proved to be prohibitive.

5.2.2 Network Model

Terminals are assumed to be stationary and distributed uniformly in a two dimensional 50 m x 50 m grid. The radio propagation model presented in Section 5.2.1 is used to compute the path loss between terminals.

5.2.3 Modelling Traffic Characteristics

Modelling network traffic consists of:

- A *traffic pattern*, which defines the distribution of the source and destination of packets within the network.
- A *traffic model* which defines the statistics of the traffic arrival process.

5.2.4 Traffic Models

Traffic models summarize the expected behavior of a network application or an aggregate of applications and fall into two broad categories. Some models are obtained by detailed traffic measurements of thousands of connections over days or years. Others are chosen because they are amenable to mathematical analysis. Unfortunately, only a few models are both empirically obtained and mathematically tractable.

There are two broad categories of networks; (1) circuit switched or telephone networks and (2) packet switched networks. As such, traffic models can be functionally classified as Circuit Switched or packet switched models. The WLAN system being proposed is expected to carry a majority of data¹⁸ and some voice traffic in packetized form. Moreover, due to the fact that there is no clear consensus on how to model an aggregation of data and voice traffic we have opted to use one of the most common internet traffic models in the simulations.

¹⁸ More specifically Internet traffic. i.e. TCP/IP.

Most packet switched applications can be characterized with a combination of four parameters.

- The distribution of arrival times between application invocations
- The duration of a connection (if any)
- The number of bytes transferred during the connection
- The inter-arrival times of packets within a connection

Unfortunately, there is little consensus on models for these parameters. For instance, some studies indicate that the inter-arrival times between telnet and FTP sessions are best modelled by an exponential distribution [41], while others find a good fit with the *Weibull* distribution [42].

Since we are mainly interested in modelling traffic at the network layer, i.e. no notion of a connection, only the statistics for the sessions need to be modelled. Additionally, we are interested in investigating relative DRNP performance as a function of network load and not really concerned about making statements related to capacity. We thus model the session arrival process for the entire network as a Poisson random process with rate λ . Moreover, the size of a session is modelled by an exponential random variable with mean β . Using the above model, the average network load, N is given by:

$$N = \lambda \cdot \beta \tag{44}$$

For illustration purposes, it is more convenient to represent the network load as a dimensionless quantity. We thus normalize N by dividing it by the maximum allocated data rate. i.e. channel capacity.

5.2.5 Session Parameters

All sessions are of the same type. i.e. they have the same minimum SIR and data rate requirements. These parameters used are summarized in Table VIII. To ensure that simulation runs are comparable across resource allocation policies, the maximum possible data rate and SIR that can be allocated to a session are limited to a multiple (in this case 4) of the minimum data rate and SIR. Specifically, the maximum possible data rate and SIR are limited to 256 Kbits/sec

Parameter	Minimum Value	Max. Allowable
SIR (γ) (dB)	10.0	16.0
Data Rate (Kbits/sec)	64	256
Avg. Session Size (Kbits)	16	16

and 40 (16 dB) respectively.

5.2.5.1 Traffic Patterns

Traffic patterns dictate the distribution of messages/packets within a network. In LAN systems, every node usually needs to communicate with every other node in the network. In WLAN systems, the distribution of traffic is dependent on WLAN's topology. For instance, in infrastructure WLANs, which consist of one or more clusters of adhoc WLANs connected via a wired backbone, most messages are exchanged within a cluster. Since we are essentially modelling a single hop adhoc WLAN, which is topologically very similar to a wired LAN, the uniform traffic pattern seems most appropriate. i.e. each terminal receives and transmits a equal portion of the global network traffic.

5.3 Performance Metrics

5.3.1 Blocking Rates

The blocking rate is essentially a throughput measure and represents those sessions that have exceeded their backoff count and are deemed lost due to:

- The receiver not being available.
- Lack of sufficient resources to meet their QoS requirements.

5.3.2 Loss of QoS Probabilities (QoS Loss)

The QoS Loss parameter only applies to real mode simulations and denotes the percentage of sessions that are successfully setup *and* lose their QoS guarantees (their SIR falls below the

5.4 Simulation Modes

Using the above models, a custom network simulator (coded in C++) was used to analyze the performance of DRNP under a variety of scenarios. The simulations were performed in two modes, *genie* and *real*.

5.4.1 Genie Mode

The *Genie* mode simulations consider an idealized DRNP network where no message loss occurs. It this represents the best case scenario and is a useful gauge of performance degradation when message loss is considered.

5.4.2 Real Mode

Real mode simulations illustrate realistic scenarios that involve loss of control messages. It is important to note that even in *real* mode simulations we assume that:

- All terminals are within range of each other.
- A terminal that is not transmitting will receive all control messages except when there is a collision on the CCCH.

Specifically, the percentage of sessions that lose their QoS guarantees is investigated for the following scenarios:

- The *deaf transmitter* scenario where a transmitter is incapable of receiving any control messages. In this case, it has no knowledge of any new sessions while it is transmitting.
- Loss of UPD_MSI message through collisions. This scenario takes the deaf transmitter phenomenon as well as the fact that update MSI messages are prone to collisions into account.
- The SREJ mode of simulation demonstrates the efficiency of the SREJ mechanism towards limiting loss of QoS

5.5 Simulation Runs

Each simulation run consists of a randomly generated network. The session requests are

terminals. i.e. no loopback sessions are permitted. A run is deemed complete after 150,000 session arrivals. Each run is repeated 30 times with independently generated random networks and traffic patters. The required metrics along with their 95% confidence intervals were then computed and plotted.

5.5.1 Genie Mode Simulation Results

We now compare the blocking probabilities for the three resource allocation policies discussed in Chapter 4. i.e. Minimum Power (MP), Maximum Rate (MR) and Maximum SIR (MSIR). The effect of the backoff count parameter on global blocking probabilities is also investigated.

5.5.1.1 Blocking Rates

Figure 21 shows the blocking rates for the MP, MR and MSIR allocation schemes for a network where the backoff count is set to zero. i.e. an incoming session that cannot obtain resources is blocked and forced to leave the system. Contrary to results obtained in cellular multi-media CDMA systems where capacity is maximized by using the MP scheme, and as predicted by the qualitative discussion in Chapter 4, the MSIR scheme seems to be the best overall performer.

This can be attributed to the incremental nature of DRNP as opposed to the global RM schemes proposed for cellular architectures. The performance gap is especially evident in light to moderate load conditions. For instance, with a network load of 0.1 times channel capacity, blocking rate for the MP scheme is an order of magnitude greater than that for the MSIR scheme (almost 12 times as great). Although, the gap does narrow in extremely high load conditions, the MSIR scheme still allows 50% of the traffic through (almost full link capacity¹⁹), while the MP scheme causes almost 90% of the sessions to be blocked. The above discussion demonstrates that although the same resource allocation policies are applicable in centralized and distributed networks, their performance within the context of these networks is quite different.

¹⁹ This is achieved by having multiple active sessions each having being allocated the minimum required data rate; i.e. 64 Kbits/sec.

Figure 21: Blocking Rates (No BackOff)

The qualitative arguments relating to the MSI, and hence the Maximum Transmission power (MaxTx), being the limiting factor for network performance are confirmed in Figure 22 which shows the respective values for average maximum transmission powers for the three resource management schemes. The MaxTx values for all three schemes are quite similar for very light traffic conditions. This is to be expected as the network is idle most of the time and an incoming session will essentially experience no power constraints. But as the traffic load is increased, the MaxTx values for both the MP and MR schemes deteriorate rapidly. i.e. a large proportion of the incoming sessions find the network busy, resulting in their transmission powers being constrained to zero. The MSIR scheme seems to be most robust with maximum transmission power values close to the maximum limit for low and moderate traffic loads. The preceding discussion confirms the strong correlation between network performance and the MSI parameter.

The backoff count parameter dictates the maximum number of times a session can request

Figure 22: Average Maximum Transmission Power vs. Network Load

resources from network before being discarded. Higher values for this parameter can improve blocking rates at the expense of relatively larger buffer usage and delay. Figure 23 illustrates the blocking rates for the various RA policies with the maximum backoff count set to 3. As can be seen by comparing Figure 23 and Figure 21, the MSIR scheme seems to derive the most benefit from the non zero backoff count. For instance, at a network load of 2, the blocking rate was reduced from ~50% to slightly over 20%. The MP scheme is still an order of magnitude worse than the MR scheme. The performance of these schemes seems invariant to low values of the backoff count, especially for high levels of traffic.

Figure 24 shows that the MaxTx parameter is essentially independent of the backoff/retry policy used in the network. The improvements seen in the blocking rates are reflected in the comparatively higher valued for MaxTx.

Figure 23: Blocking Rates (BackOff Count = 3)

5.5.2 Real Mode Simulation Results²⁰

We now turn our attention to investigating the effect of message loss on DRNP's ability to maintain QoS guarantees. Unless stated otherwise, τ_{DATA} is set to 10 slots.

The *deaf* transmitter mode can be thought of as the best case real mode scenario. Here, the only messages lost are due to a transmitters inability to receive control messages. This means terminals, while transmitting, have no knowledge of sessions that arrive or leave. Thus when these transmitters issue RTS messages for future session setups, they advertise an out of date value for the maximum transmission power. Without the SREJ mechanism, this causes loss of QoS for already active sessions.

²⁰ Note that all simulations in real mode were performed with the backoff count set to zero.

Figure 24: Avg. Maximum Transmission Power (BackOff Count = 3)

As Figure 25 shows, under light load conditions the percentage of sessions that violate their QoS is limited to approximately 0.05%, but under extremely heavy load conditions, inspite of the heavy blocking probabilities (~50%), 5.6% of sessions end up losing their QoS guarantees. This is quite substantial. Ideally, we would like to limit QoS loss rates to 1-2%.

Including the effect of UPD_MSI message loss only makes matters worse. The SIR for slightly over 9% of the sessions fall below their negotiated minimum. However, this result does demonstrate DRNP's robustness, even under extreme traffic load conditions. This corroborates our results from the Spin validation model. i.e. there is no cascade effect in loss of QoS guarantees. DRNP is self-healing and can successfully recover from an incorrect state within a finite time frame.

The situation changes substantially with the introduction of the SREJ mechanism. Even though,

Figure 25: % Loss of QoS: Deaf Transmitter

SREJ messages are themselves lost due to collisions, even under extremely high traffic conditions, the loss of QoS is limited to $\sim 2.1\%$ which is almost a five fold improvement.

Figure 26 illustrates the QoS Loss rates for various sizes of the τ_{DATA} window. As expected, the performance improves as the window is increased. Even with very small values of τ_{DATA} , QoS Loss rates are limited to 10% for very high network load conditions. Although, unacceptably high, these results again demonstrate the effectiveness of the SREJ mechanism. As stated earlier, with a moderate value of τ_{DATA} (~ 10) QoS Loss rates are well within acceptable levels.

Figure 26: QoS Loss Rates as a function of τ_{DATA} size.

6 Summary & Future Work

6.1 The DRNP Protocol

DRNP was designed specifically to allow resources to be allocated in a multi-media environment. The protocol also fulfills the role of a Call Admission Control (CAC) mechanism, blocking those users that violate the QoS constraints of the currently active sessions. It is completely distributed and adaptive to changing network conditions. DRNP has been tested and validated using automated state space search techniques and found to be robust and fault-tolerant. In fact, for reasonable values of τ_{DATA} , even under severe traffic load scenarios, the loss of QoS guarantees is limited to around 2%.

manager. In fact, the notion of QoS guarantees can be totally removed. In this case, the network can function in *best effort* mode, attempting to allocate the best packet error and data rates to each session.

6.2 Resource Allocation Policies

We investigated the performance of three resource allocation policies (minimum resource, maximum rate, maximum SIR) within the context of DRNP, in terms of network wide metrics such as blocking and *loss of QoS* rates. Surprisingly, the minimization of resources scheme which emphasizes conservation of transmission powers and yields the highest capacities in cellular (more generally centralized) networks, turned out to be the worst overall performer. This can be attributed to the fact that DRNP is an *incremental* Resource Allocation (RA) protocol. Since resources are allocated on a per session basis, a session that has been allocated minimum SIR cannot sustain any additional interference without losing its QoS guarantees. This leads to a single session blocking all other sessions. The best overall performer was the maximum SIR RA policy which tries to maximize the SIR allocated to each session. Although this does increase the MAI within the network, it does allow other spatially dispersed terminals to setup sessions successfully, as each session is allocated a high QoS (SIR) that it needs. Thus we note that although the same resource allocation policies are applicable in centralized and distributed networks, their performance within the context of these networks is quite different.

6.3 Future Work

Future research can follow one of several avenues:

- The first involves retaining the DRNP framework and investigating the performance of other resource allocation policies. The RA schemes presented in this dissertation are adhoc, but an adaptive scheme combined with other performance metrics may yield more efficient allocation of resources. For instance, under extremely low traffic load conditions, the maximum SIR policies performance is not much better than the minimum power scheme. Thus in this case, the latter may suffice, especially if power conservation is paramount.

since issuing a SREJ requires that the reception of a current session be interrupted, a soft decision metric that takes the overall damage caused may yield better overall performance at the expense of minor QoS guarantee losses. A long term goal involves the extension of DRNP to multi-hop networks. Scalability issues arising from control channel traffic are especially important here.

- Lastly, other distributed resource allocation protocols can be investigated and benchmarked against DRNP.

Access Mechanism

Consider a DS-SS system with K multiple access users. Assume that each user has a PN sequence with N chips per data symbol period T such that $NT_c = T$. The transmitted signal of the k th user can be expressed as:

$$S_k(t) = \sqrt{\frac{2E_s}{T_s}} d_k(t) c_k(t) \cos(2\pi f_c t + \phi_k) \quad (45)$$

where $c_k(t)$ and $d_k(t)$ are respectively the PN sequence and data sequence for the k th user. The received signal consists of the sum of K different transmitted signals (one desired user and $K-1$ undesired users). Reception is accomplished by correlating the received signal with the appropriate signature sequence. The decision variable for the i th transmitted bit of user 1 is:

$$Z_i^{(1)} = \int_{(i-1)T+\tau_1}^{iT+\tau_1} r(t) c_i(t-\tau_1) \cos[2\pi f_c(t-\tau_1) + \phi_1] dt \quad (46)$$

where $r(t)$ is the received signal. If $d_{1,i} = 1$, then the bit will be received in error if $Z_i^{(1)} > 0$. The probability of error can be calculated as $\Pr [Z_i^{(1)} > 0 \mid m_{1,i} = -1]$. Since $r(t)$ is a linear combination of signals, can be rewritten as:

$$Z_i^{(1)} = I_1 + \sum_{k=2}^K I_k + \xi \quad (47)$$

where

$$I_1 = \int_0^T S_1(t) c_1(t) \cos(2\pi f_c t) dt = \sqrt{\frac{E_s T}{2}} \quad (48)$$

is the response of the receiver to the desired signal from user 1.

is a Gaussian random variable representing noise $n(t)$, with zero mean and variance:

and

$$E[\xi^2] = \frac{N_0 T}{4} \quad (50)$$

$$I_k = \int_0^T S_k(t - \tau_k) c_1(t) \cos(2\pi f_c t) dt \quad (51)$$

represents the Multiple Access Interference (MAI) from user k . Assuming that I_k is composed of the cumulative effects of N random chips from the k th interferer power the integration period T of one data bit, the *Central Limit Theorem* implies that the sum of these effects will tend towards a Gaussian Distribution. Since there are $K - 1$ users which serve as identically distributed interferers, the total MAI is:

$$I = \sum_{k=2}^K I_k \quad (52)$$

which may be approximated by a gaussian random variable [43].

Figure 28: Transmitter Algorithm as executed by Terminal i .

Figure 29: Transmitter Algorithm (Session Tear-Down)

Figure 30: Receiver Algorithm as executed by Terminal j .

Figure 31: 3rd Party Algorithm as executed by Terminal l .

1. Pahlavan, Kaveh and Levesque, Allen, *Wireless Information Networks*, John Wiley & Sons Inc., 1995.
2. Garg, V.K., Wilkes, J.E., *Wireless and Personal Communications Systems*, Prentice Hall PTR, New Jersey, 1996.
3. Viterbi, A.J., and Roberto Radovani, *Implications of Mobile Cellular CDMA*, IEEE Communications Magazine, pp. 38-41, Vol. 30, No. 12, December 1992.
4. Gilhousen, K.S., et al., *On the Capacity of a Cellular CDMA System*, IEEE Transactions on Vehicular Technology, Vol. 40, No.2, pp. 303-12, May 1991.
5. Pickholtz, R.L., Milstein, L.B., and Schilling, D., *Spread Spectrum for Mobile Communications*, IEEE Transactions on Vehicular Technology, Vol. 40, No. 2, pp. 313-322, May 1991.
6. Price, R., Green, P.E., *A Communication Technique for Multipath Channels*, Proceedings of the IRE, pp. 555-570, March 1958.
7. Pahlavan, K., *Wireless Local Area Networks (Chapter 11 in Personal Communications)*, Gardiner and West (Eds.), Artech House, Norwood, MA, 1994.
8. Diepstraten, Wim. et. al, *Distributed Foundation Wireless Medium Access Control*, IEEE 802.11 Proceedings (11-93/190), November 1993.
9. Cheah, J., *A Proposed Access Method for the WLAN Standard*, IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Boston, MA, pp. 145-148, October 1992.
10. Chu, T.S., Gans, M.J., *High Speed Infrared Local Wireless Communications*, IEEE Communications Magazine, Vol. 25, pp. 4-10, July 1987.
11. Wittenben, A., Liu, W., *The European Wireless LAN Standard HIPERLAN: Key Concepts and Testbed Results*, Proc. of IEEE 47th Vehicular Technology Conference, Vol. 3, pp. 1317-1321, May 1997.
12. Wilkinson, T.A., Phipps, T., Barton, S.K., *A Report on HIPERLAN Standardization*, International Journal of Wireless Information Networks, Vol. 2, pp. 99-120. March 1995.
13. Keshav, S., *An Engineering Approach to Computer Networking: ATM Networks, the Internet & The Telephone Network*, Addison Wesley Professional Computing Series, May 1997.

15. Cox, D.C., *Wireless Personal Communications: What Is It?*, IEEE Personal Communications Magazine, pp. 20-35, April 1995.
16. Pahlavan, K., *Trends in Local Wireless Data Networks*, Proc. of IEEE VTS 46th Vehicular Technology Conference, Vol. 1, pp. 21-25, 1996.
17. Elwalid, A., Mitra, D., and Wentworth, R.H., *A New Approach for Allocating Buffers and Bandwidth to Heterogenous, Regulated Traffic in an ATM Node*, IEEE Journal on Selected Areas in Communication, Vol. 12, No. 3, pp. 544-551, April 1994.
18. Floyd, S., and Jacobson, V., *Link Sharing and Resource Management Models for Packet Networks*, IEEE Transactions on Networking, Vol. 3, No. 4, August 1995.
19. Guerin, R., Ahmadi, H., and Nagshineh, M., *Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks*, IEEE Journal on Selected Areas in Communication, Vol. 9, No. 7, pp. 968-981, September 1991.
20. Guerin, R., and Gun, L., *A Unified Approach to Bandwidth Allocation and Access Control in Fast Packet Switched Networks*, Proceedings of IEEE INFOCOM '92.
21. Nettleton, R.W. and Alavi, H., *Power Control for a Spread Spectrum Cellular Mobile Radio System*, Proc. VTC'83, pp. 242-246, Vol. 1, 1983.
22. Grandhi, S.A., Vijayan, R., Goodman, D.J. and Zander, J., *Centralized Power Control in Cellular Radio Systems*, IEEE Transactions on Vehicular Technology, pp. 466-68, Vol. VT-42, No. 4, November 1993.
23. Sampath, A., Kumar, P.S., and Holtzman, J., *Power Control and Resource Management for a Multimedia CDMA Wireless System*, Proc. of IEEE PIMRC 95, Vol. 1, pp. 21-5.
24. Wong, Wilson Wai Shun, *Power Assignment Strategies and Traffic Control for Wireless Multimedia DS-CDMA Systems*, Masters Thesis, University Of Toronto, 1996.
25. Shum, Kenneth W., *Fuzzy Distributed Power Control in Cellular Radio Networks*, Proc. of PIMRC '95, pp. 51-55, Vol 1, 1995.
26. Aghvami, A.H., *Future CDMA Cellular Mobile Systems Supporting Multi-Service Option*, Proc. of the PIMRC '94, pp. 1276-79, September 1993.
27. Baier, A., *Multi-Rate DS-CDMA: A Promising Access Technique for Third Generation Mobile Radio Systems*, Proc. PIMRC '93, pp. 114-118, September 1993.
28. Cox, D.C., Murray, R.R., and Norris, A.W., *Measurements of 800 Mhz Radio Transmission into Buildings with Metallic Walls*, Bell Systems Technical Journal, Vol. 62, No. 9, pp. 2695-

29. Vannucci, Gerard J., Sagar Kulkarni A., Nasad, Ranjee, *Wideband Indoor Channel Measurements and BER Analysis of Frequency Selective Multi-path Channels and 2.4, 4.75, and 11.5 GHz*, IEEE Transactions on Communications, Vol. 44, No. 10, October 1996.

30. Rappaport, T., *Wireless Communications: Principles & Practice*, Prentice Hall PTR, New Jersey, 1996.

31. Sousa E., *Digital Mobile Communications (ECE 1543 Class Notes)*, Department of Electrical and Computer Engineering, University of Toronto, 1996.

32. Steele, R., *Mobile Radio Communications*, IEEE Press, 1992.

33. Holzmann, Gerard J., *Design and Validation of Computer Protocols*, Prentice Hall Software Series (New Jersey), 1991.

34. Sousa E., *A Spreading Code Protocol Utilizing an Echo From the Receiver*, Proc. Milcom 88, Vol.1, pp. 357-359.

35. Sousa, E., *Spreading Code Protocols for Distributed Spread-Spectrum Packet Radio Networks*, IEEE Transactions on Communications, Vol. 36, pp. 272-81, March 1988.

36. Georgiopoulos, M., *Packet Error Probabilities in Direct Sequence Spread-Spectrum Packet Radio Networks*, IEEE Transactions on Communications, pp. 1599-1606, Vol. 38, September 1990.

37. Lehnert, J.S. and Pursley, M.B., *Error Probabilities for Binary Direct-Sequence Spread Spectrum Communications with Random Signature Sequences*, IEEE Transactions on Communications, pp. 87-98, Vol. COM-35, No. 1, January 1987.

38. Sousa, E., *Interference Modelling in a Direct-Sequence Spread Spectrum Packet Radio Network*, IEEE Transactions on Communications, pp. 1475-1482, Vol. 38, September 1990.

39. Whitehead, James, *Distributed Packet Dynamic Resource Allocation (DRA) for Wireless Networks*, Proc. of the IEEE VTS 46th Vehicular Technology Conference, Vol. 1, pp. 111-5, 1996.

40. Holzmann, Gerard J., *Design and Validation of Computer Protocols*, Prentice Hall Software Series, 1991.

41. Paxson, V, and Floyd, S., *Wide-Area Traffic: The Failure of Poisson Modelling*, Proc. of ACM SIGCOMM 94, London, September 1994.

42. Feldmann, A., *On-line Call Admission for High Speed Networks*, PhD thesis issued by School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1995.

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved