Performance optimization of CSMA/CA based ad hoc networks

Jaya Shankar Pathmasuntharam

2006

Jaya, S. P. (2006). Performance optimization of CSMA/CA based ad hoc networks. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/35744>

<https://doi.org/10.32657/10356/35744>

Performance Optimization of CSMA/CA Based Ad Hoc Networks

Jaya Shankar Pathmasuntharam

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2006

Abstract

Ad hoc wireless networks introduce a number of challenges for protocol designers. Stations using such networks are constrained by factors such as limited bandwidth, link errors, and collisions. Changes are needed at various levels of the protocol stack, but most importantly at the Medium Access Control (MAC) layer. The medium access mechanism in multi-hop wireless networks should minimize collisions, and take care of the hidden and exposed node problems. It should also be flexible enough to allow all nodes to maximize the utilization of available bandwidth. Due to the broadcast nature of wireless transmission, when two hosts are communicating, all other hosts within the range of the two hosts must defer their communication in order to avoid collision. This results in significant throughput degradation as the number of active hosts increases. Studies have shown that wireless networks provide dismal capacity in the presence of many contending nodes in a given area.

In the recent past, research has been carried out to improve the throughput of wireless networks and various approaches have been proposed. Some of the areas of enhancements include power control, antenna directivity, multi-channel capability, concurrent transmission, multi-rate, etc. In this thesis, the focus of my work is on the design of new and innovative methods to enhance the throughput and delay performance of ad hoc networks using CSMA/CA-based MAC protocols. In the later chapters, we highlight three new MAC layer schemes and a link layer hop-by-hop flow control scheme. The first MAC scheme is a multi-channel MAC protocol that is capable of enhancing the throughput of CSMA/CA protocols in a multi-channel

environment and exhibits good channel scaling properties. To optimize the performance of this MAC protocol, we introduce a channel assignment scheme that can be used along with this multi-channel MAC protocol. The second MAC scheme is a concurrent transmission scheme that is able to improve the throughput of a single channel ad hoc network by a factor of 2. The third MAC scheme is a multi-rate MAC protocol, which exhibits improved throughput and delay performances when compared to existing protocols. Finally, we introduce a unique hop-by-hop flow control framework, which has low per-flow processing complexity and improves the overall throughput and energy efficiency of wireless ad hoc networks.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisors, Associate Professors Amitabha Das and Anil Kumar Gupta for their trust, excellent guidance, wisdom and patience throughout my entire doctoral programme. The directions shown during the course of my studies have been tremendously constructive. I am truly fortunate to have received excellent guidance from them in many aspects of my studies and they also became excellent role models for me in terms of work ethics. I would also like to thank Professor Prasant Mohapatra for his valuable guidance and insightful comments during the last phase of my doctoral programme. I am grateful to Professor Lye Kin Mun for encouraging me to pursue my doctoral studies.

Thanks are also due to my colleagues of Institute for Infocomm Research for their help in one way or another. I would also like to thank my wife and parents for their love, support and understanding during the course of my research. This work would not have been possible without them. Lastly, I would like to thank the Institute for Infocomm Research, Singapore, for generously sponsoring this research work.

Contents

Abstract 1

Acknowledgements 3

Contents 4

List of Figures 9

List of Tables 14

Chapter 1. Introduction 15

 1.1 Research motivation..... 17

 1.2 Research Approach 22

 1.3 Main contributions of this thesis..... 24

 1.4 Organization of this thesis 26

Chapter 2. Primary Channel Assignment based Multi-channel MAC Protocol 27

 2.1 Problems in Multi-channel MAC designs..... 28

 2.1.1 Hidden terminal problem 28

 2.1.2 Busy Receiver 29

 2.1.3 Broadcast messages 29

 2.2 Related Work 29

 2.3 The PCAM protocol..... 33

 2.4 Performance Evaluation 38

 2.4.1 Wireless LAN 40

 2.4.2 Chain 44

 2.4.3 Multi-hop lattice..... 46

 2.4.4 Multi-hop with mobility..... 48

2.5	Conclusions.....	50
Chapter 3. Channel Assignment for Nullifying the Critically Exposed Node		
Problem in Ad Hoc Wireless Networks		51
3.1	Wireless protocol design issues	52
3.2	Exposed node problem.....	54
3.3	Experiments to study the critically exposed node problem	57
3.3.1	Single channel MAC.....	57
3.3.2	Experiments with channel assignment and the PCAM protocol.....	59
3.4	Analysis of Channel assignment for eliminating critically exposed node problem	64
3.4.1	Design issues.....	64
3.4.2	Channel assignment and guard zones	65
3.5	Channel assignment strategy.....	68
3.6	Clustering and Channel assignment.....	69
3.7	Channel assignment algorithm.....	74
3.8	Distributed heuristics-based algorithm	75
3.9	Experimental results.....	80
3.10	Related Work	82
3.11	Conclusion	84
Chapter 4. SCENT – A MAC Protocol for Concurrent Transmission in Wireless		
Ad-hoc Networks		85
4.1	Preliminaries	85
4.1.1	Packet capture and conditions for parallel transmission.....	86
4.1.2	Location information	89
4.1.3	Busy Tone	90

4.2	Related Work	91
4.3	The SCENT Protocol	94
4.3.1	Message sequence	95
4.3.2	Finite state machine	98
4.3.3	Design of Secondary contention window	102
4.3.4	Enabling tertiary pairs	103
4.3.5	Deterministic vs. non-deterministic scheduling	104
4.4	Determination of maximum number of communication pairs in the transmission range	105
4.5	Performance Evaluation	111
4.5.1	Simulation Model	111
4.6	Simulation results	113
4.7	Conclusions	120
Chapter 5.	An Efficient Multi-rate Relaying (EMR) MAC Protocol for Wireless Ad hoc Networks	121
5.1	Related Work	122
5.2	Preliminaries	124
5.2.1	SIR condition	125
5.2.2	Packet Rates and Overheads	125
5.2.3	Broadcast messages and routing	126
5.3	Efficient multi-rate relaying MAC (EMR)	127
5.3.1	Relay selection	127
5.3.2	Relay discovery and selection	130
5.3.3	Message Sequence	131

5.4	Analytical Model to Studying Throughput Performance Bounds of Localized Links	133
5.4.1	Error Probability Model	135
5.4.2	Enhancement to Bianchi's Markov Chain Model	139
5.4.3	Computing Throughput	141
5.5	Performance Evaluation	145
5.6	Simulation Results	151
5.7	Conclusion	154
Chapter 6.	A Flow Control Framework for Improving Throughput and Energy Efficiency in CSMA/CA based Wireless Multihop Networks	156
6.1	Related Works	158
6.2	Factors affecting throughput in multihop transmission	161
6.2.1	The critically exposed node problem	161
6.2.2	Location dependency and uneven carrier sensing resulting in congestion	162
6.2.3	False NAV	163
6.2.4	Frozen MAC State	165
6.3	Severity of the Congestion, False NAV and Frozen MAC state Problem in Multihop networks	166
6.3.1	Metrics	166
6.3.2	Experiments	166
6.4	A flow control framework	173
6.5	Enhanced CSMA/CA MAC protocol	174
6.5.1	Solving the false NAV problem	175
6.5.2	Modifications to the 4-way handshake	175

6.6 Hop-by-hop Hybrid window and rate based flow control 176

6.6.1 Flow ID representation and management 177

6.6.2 Bandwidth-Delay Product and flow control design..... 179

6.6.3 Static hop-by-hop window control..... 183

6.6.4 Rate control 184

6.7 Simulation Results 190

6.8 Conclusion 198

Chapter 7. Conclusions and Future Work 200

7.1 Summary 200

7.2 Future Work 203

References..... 206

List of Figures

Figure 0-1. Radio irregularity	22
Figure 0-2. Degree of Irregularity (DOI)	23
Figure 2-1. Hidden terminal problem	28
Figure 2-2. Channel assignment in PCAM	34
Figure 2-3. Channel assignment and transceiver arrangement for a 6-node chain setup	35
Figure 2-4. Two-Ray propagation and path loss.....	37
Figure 2-5. WLAN with low node density	40
Figure 2-6. WLAN with medium node density	40
Figure 2-7. WLAN with flows in opposite directions	43
Figure 2-8. Scalability with respect to available channels.....	43
Figure 2-9. Chain setup.....	44
Figure 2-10. End-to-end Throughput in Chain setup.....	45
Figure 2-11. Multi-hop lattice.....	46
Figure 2-12. Throughput performance in ad hoc setup.....	47
Figure 2-13. Aggregate throughput performance in a mobile ad hoc setup	48
Figure 3-1. Regions in CSMA/CA based protocols.....	54
Figure 3-2. Throughput at R1	57
Figure 3-3. Throughput at R4	57
Figure 3-4. Throughput fairness for R1/R4	58
Figure 3-5. Chain setup at 200 m between neighboring nodes.....	60
Figure 3-6. Channel assignment and transceiver arrangement in PCAM.....	61

Figure 3-7. Critically exposed interference in a Grid setup	64
Figure 3-8. Critically exposed interference boundary	66
Figure 3-9. Clustering schemes.....	69
Figure 3-10. Approximation to ideal clustering and ranges	70
Figure 3-11. Clustering for boundary nodes	72
Figure 3-12. Example channel assignment with $N_{cluster} = 7$	73
Figure 3-13. Initialization phase	76
Figure 3-14. Main distributed channel assignment algorithm	77
Figure 3-15. Channel assignment request function.....	78
Figure 3-16. Maximum channels vs. number of clusters.....	81
Figure 3-17. Distribution of channel assignment in a 10000 clusters setup	81
Figure 4-1. Condition for parallel transmissions	86
Figure 4-2. Node placement highlighting the problem in MACA-P	93
Figure 4-3. Parallel transmissions within transmission range of node A	94
Figure 4-4. SCENT protocol message sequence for Primary, Secondary and Tertiary pairs.....	97
Figure 4-5. FSM for primary sender	98
Figure 4-6. FSM for primary recipient and secondary pair	99
Figure 4-7. Preventing collisions from overlapping regions.....	101
Figure 4-8. Primary pair and best case location of secondary senders and recipients.	107
Figure 4-9. Best case location of the first secondary pair	107
Figure 4-10. Best case location of the second secondary recipient	107
Figure 4-11. Best case location of the second secondary sender	108
Figure 4-12. Best case location of the third secondary pair.....	108
Figure 4-13. Aggregate throughput for long-range pairs.....	113

Figure 4-14. Aggregate throughput for medium-range pairs.....	114
Figure 4-15. Aggregate throughput for short-range pairs.....	114
Figure 4-16. Simulation setup with 8 pairs of links (20m, 80m, 140m and 200m).....	115
Figure 4-17. Throughput of various transmission ranges for 802.11 (20m, 80m, 140m and 200m)	115
Figure 4-18. Throughput of various transmission ranges for 802.11 (20m, 80m, 140m and 180m)	116
Figure 4-19. Throughput contributed by Primary contention in SCENT	116
Figure 4-20. Throughput contributed by Secondary contention in SCENT	117
Figure 4-21. Total throughput using SCENT	118
Figure 4-22. Aggregate Throughput of SCENT and 802.11 in a multihop scenario...	119
Figure 5-1. Path selection for various schemes. Path selection with Shortest Path (SPR) Algorithm = $A \leftrightarrow C \leftrightarrow E$. Path selection with MTM, MARO, and EMR = $A \leftrightarrow B$ $\leftrightarrow C \leftrightarrow D \leftrightarrow E$	123
Figure 5-2. Transmission Ranges and Regions.....	127
Figure 5-3. Relaying Nodes	128
Figure 5-4. EMR MAC message sequence.....	131
Figure 5-5. PLCP Protocol Data Unit (PPDU) format	134
Figure 5-6. SNR vs. BER.....	137
Figure 5-7. CSMA/CA - 1 Mbps links under various blocking probability	145
Figure 5-8. CSMA/CA - 11 Mbps links under various blocking probabilities.....	146
Figure 5-9. Position of nodes used in analytical model	148
Figure 5-10. MTM with high density	149
Figure 5-11. EMR with high density	149
Figure 5-12. MTM with low density and added noise of 3dBm.....	150

Figure 5-13. EMR with low density and added noise of 3dBm.....	150
Figure 5-14. Throughput performance in a chain setup.....	151
Figure 5-15. Delay performance in a chain setup	151
Figure 5-16. Aggregate throughput in the presence of mobility.....	153
Figure 6-1. Chain setup with flow traveling from node 1 to node 9.....	161
Figure 6-2. Uneven carrier sensing due to node location	162
Figure 6-3. End-to-end throughput for chain setup.....	167
Figure 6-4. Number of received and dropped DATA frame for chain setup (packet inter-arrival time = 0.005 seconds)	168
Figure 6-5. Number of received and dropped DATA frame for chain setup (packet inter-arrival time = 0.057 seconds)	168
Figure 6-6. Transmission cost for chain setup	169
Figure 6-7. Percentage of various receiver states in 6-node chain setup when packet inter-arrival at node 1 is 0.005 sec.	170
Figure 6-8. Grid setup and high fan-in chain setup.....	171
Figure 6-9. Flow control framework.....	173
Figure 6-10. (a) Wired 8-hop link. (b) Spiral wireless multihop links	180
Figure 6-11. Receive function.....	186
Figure 6-12. Transmit function	189
Figure 6-13. Cleanup function	189
Figure 6-14. Directed 5-chain links fan-in.....	191
Figure 6-15. Throughput performance – chain setup.....	191
Figure 6-16. Transmission cost – chain setup.....	193
Figure 6-17. Throughput performance – directed 5 chain links fan-in	194
Figure 6-18. Transmission cost – directed 5-chain links fan-in.....	194

Figure 6-19. Aggregate throughput – mobile ad hoc network setup 197

List of Tables

Table 2-1. Distance vs. received power 37

Table 2-2. MMAC protocol parameters..... 39

Table 2-3. Aggregate Throughput vs. Packet size 41

Table 2-4. Throughput and Delay performances in Chain setup 45

Table 2-5. Average end-to-end delay (in seconds) 48

Table 3-1. Channel Assignment and Throughput in PCAM..... 60

Table 3-2. Notations..... 75

Table 4-1. Number of Interfering node vs. possible Primary sender-receiver distance
..... 109

Table 5-1. Typical 802.11 system Parameters 125

Table 5-2. Effective throughput and priority level for various relay combination..... 129

Table 5-3. Orinoco Wireless Card Parameters..... 136

Table 5-4. Path loss related parameters 138

Table 5-5. IEEE 802.11b Parameters..... 145

Table 5-6. Aggregate End-to-End Throughput for MTM and EMR scheme 148

Table 6-1. Performance results for Grid setup 171

Table 6-2. Performance results for High Fan-in chain setup 172

Table 6-3. Notations..... 186

Table 6-4. Summary of schemes..... 191

Table 6-5. Performance of Flow Control Framework using Grid Setup 196

Table 6-6. Performance of Flow Control Framework using High Fan-in chain Setup
..... 196

Introduction

Ad hoc networks research, with previously assumed names such as packet radio or multi-hop networks research, can be traced back to 1972 when US Department of Defense (DoD) sponsored a project known as Packet Radio Network (PRNET) [20]. Since then, the area of research has expanded tremendously partly due to the fact that it has enormous commercial and military potentials. Ad hoc networks can be set up anywhere and anytime because they eliminate the complexities of infrastructure setup and central administration.

An ad hoc network is a collection of communications devices that wish to communicate but have no fixed infrastructure available. They operate in a self-organized and decentralized manner and message communication takes place via multiple hops. A packet is sent to its target node via a set of intermediate nodes that act as routers. Individual nodes are responsible for dynamically discovering neighboring nodes, which they can directly communicate with. A significant feature of ad hoc networks is that rapid changes in connectivity and link characteristics may be introduced due to node mobility and interference. Ad hoc networks are not specific to a particular application area and could include packet radio networks, sensor networks, personal communication systems, rooftop networks, and wireless local area networks.

While ad hoc networking is a multi-layer problem, our main focus in this research work revolves around the medium access control (MAC) layer issues. Since the wireless medium is a broadcast medium, multiple devices can access the medium at the same time. Simultaneous transmissions can result in garbled information at any receiver. Besides minimizing collisions, the MAC layer needs to fairly allocate resource, and efficiently transport data over the shared wireless links in the presence of rapid changes and hidden or exposed terminals.

The performance of a wireless system is greatly influenced by the MAC protocol used. Imperfection in the MAC layer introduces additional throughput losses, which can be easily propagated to the upper layer resulting in degradation of the system performance. Due to these reasons, an efficient MAC protocol, which moderates access to the shared medium by defining specific access rules, is necessary in an ad hoc network.

Wireless MAC protocols have been studied extensively since the 1970s and can be logically divided into two classes namely, distributed and centralized [11]. Distributed protocols are normally contention-based [9,23,24,45,49,85] and implemented using a random access technique. Access to the channel is not time synchronized and nodes use a random access protocol to contend for the channel when they have packets to send. The centralized protocols [7,106] are normally time synchronized and packets are transmitted in fixed time slots. These time-slotted protocols normally require a centralized node to distribute transmission schedules to nodes within the vicinity. Due to the fixed time slots and predetermined schedules, these protocols can achieve contention free transmissions and result in higher throughputs. However, due to time

varying properties of the network such as node mobility, traffic characteristics, etc., there is high complexity in scheduling the slots efficiently. Due to the simplicity, adaptability, and robustness of the distributed MAC protocols, they offer a more attractive solution in such a dynamic environment. The MAC protocols that we explore in this research work, fall under the category of distributed MAC schemes. More specifically, we will focus on the throughput analysis as well as the introduction of new and innovative methods to enhance the throughput and delay performance of contention-based MAC protocols.

1.1 Research motivation

One of the main concerns in the design of ad hoc networks for military and commercial applications is the scaling law associated with throughput capacity of ad hoc networks. Designers raised the question on whether an ad hoc network with large number of nodes was feasible. In the landmark paper by Gupta and Kumar [28], the authors proved through geometry, random graphs and convex analysis that per-node throughput values of $O(1/\sqrt{n})$ bits/sec and $O(1/\sqrt{n \log n})$ bits/sec are attainable for arbitrary and random networks respectively. In the arbitrary case, the nodes are optimally placed in a disk of unit area, traffic patterns are optimally assigned and transmission ranges are optimally chosen. In the random case, nodes are uniformly distributed in a topology area and source and destinations nodes are randomly chosen. The authors also showed that this theoretical scaling is true for two different models; the protocol model and the physical model. These two models are used to describe when a transmission is received successfully by its intended recipient over a one hop transmission. The protocol model considers all the nodes locations in the vicinity of a recipient and uses a “guard zone” factor to calculate which transmissions will be successful. The model is simpler

compared to the physical model and allows for imprecision in the achieved range of transmission. The physical model takes into consideration the physical layer characteristics and studies the power level of the subset of nodes transmitting at the same time instant over a certain sub-channel. This model assumes that a minimum signal to noise ratio (SNR) is necessary for successful receptions. The earlier value of $O(1/\sqrt{n})$ bits/sec is in some sense the upper bound value for a random ad hoc network if suitable protocols for routing, scheduling and medium access control can be designed. Their findings however are based on the assumption that nodes are static and the transmission and reception are omni-directional.

The main reasons why the overall capacity follows the aforementioned values are that the receiving nodes have to obey a basic signal to noise ratio (SNR) for proper frame reception and as the number of hops increases, the intermediate nodes spend a larger fraction of their capacity to relay traffic from other nodes rather than their own.

Previous work by Li *et al.* [55] studied how these reasons affect the capacity of the network. The authors examined interactions of the 802.11 MAC protocol and how forwarding by intermediate node affects the capacity of the network for several simple configurations and different traffic patterns. In their study, they also showed that the 802.11 MAC instead of a global scheduling scheme¹ does not seem to affect the asymptotic bound derived by Gupta and Kumar [28].

The studies by Gupta and Kumar [27,28] showed that the capacity of ad hoc networks is somewhat disappointing and paved the way for many other studies and

¹ Gupta and Kumar [28] demonstrated the existence of a global scheduling scheme achieving $O(1/\sqrt{n \log n})$ bits/sec for a uniform random network with random traffic pattern.

improvements to reach or even surpass this theoretical bound. The initial assumption that mobility could degrade capacity was demystified by Grossglauer and Tse [26], who demonstrated that mobility can achieve higher rates asymptotically as the number of nodes increases. In their paper, they claimed that per-node throughput capacity of $O(1)$ can be achieved. However, there is a caveat to this claim. The attained long-term throughput is averaged over the time-scale of node mobility and hence, delay of that order will be incurred. Practically, the result is of little use to most applications since most of them have real-time delay requirements. In this work, the per-session throughput is studied for applications with loose delay constraints, where the topology changes over the time-scale of packet delivery.

Yi *et al.* [101] investigated the capacity of ad hoc wireless networks using different combinations of directional antennas. In their work, arbitrary networks and random networks with static nodes similar to Gupta and Kumar's [28] model were used. The capacity of networks using directional antennas was shown to improve by a factor of $2\pi/\sqrt{\alpha\beta}$, where α and β are the beamwidths of transmitting and receiving directional antennas, respectively. If the beamwidths of transmitting and receiving antennas are decreased asymptotically as fast as $1/\sqrt{n}$, the throughput capacity will keep constant with the increase of the number of nodes in the network. Practically, this has some implication on the physical size of the directional antenna because a pencil beamwidth is difficult to achieve with small antenna structures. In conclusion, there is still a limit to which an ad hoc network can be scaled when this approach is used.

In a study to verify the theoretical bound, Gupta *et al.* [27] carried out experimental study to check their scaling law attainable in theory using actual hardware based on

IEEE 802.11 MAC protocol. The results obtained showed that the per-node throughput declines like $O(1/n^{1.68})$ bits/sec, which is considerably worse than the best-case theoretical limit of $O(1/\sqrt{n})$ bits/sec and even worse than the purely co-located case where one expects the value of $O(1/n)$ bits/sec based on simple time-division multiplexing. Their results suggest that the current hardware and supporting suite of protocols at the link and network layers are far from optimal. In particular, they concluded that a much more efficient medium access protocol would be required. With this main challenge in mind, we felt that there is still room for developing innovative ideas to improve the existing MAC and link layer protocols for ad hoc network, which could eventually improve the overall throughput capacity of ad hoc networks.

A multi-channel approach [38,63,64,82,93,94] is one of the methods that can be used to improve the throughput of ad hoc networks. However, most MAC protocols have been designed for use in a single channel environment. The existing single channel MAC protocols cannot be readily adapted for use in a multi-channel environment. The existing multi-channel MAC protocols have poor channel scaling properties due to control channel saturation. This has motivated us to design a multi-channel MAC protocol [70] that is able to effectively increase the capacity of ad hoc networks when multiple channels are available.

While the multi-channel MAC protocol that we developed allows efficient use of multiple channels, the optimal performance can only be obtained if a proper channel assignment strategy exists to sufficiently reuse the channels. The main idea of this work is to ensure that hidden terminal interference in whatever form is reduced. We therefore

introduce a distributed channel assignment strategy and describe the algorithm used [68].

While it is easier to have parallel transmissions in a TDMA based MAC protocol, CSMA/CA based protocols pose a greater challenge when parallel transmissions are used. One of the main reasons why this is difficult is that transmissions are not synchronized and packet lengths vary from one node to another. This has motivated us to introduce a novel synchronized parallel transmission MAC protocol that is able to increase the throughput of ad hoc networks [103].

Another well-explored approach of increasing throughput is by using multi-rate transmissions [5,32,79,80] to increase throughput in ad hoc networks. Since most systems today are designed with multi-rate capability, the use of shorter but effective transmission rates can benefit the system in terms of power and capacity. We present a MAC protocol [69] that uses the standard shortest path routing scheme together with a localized secondary MAC layer forwarding scheme to forward packets in a more efficient manner when compared to existing multi-rate techniques [5,80].

In multi-hop wireless networks where a random access MAC scheme such as CSMA/CA is used, nodes greedily compete in a distributed manner and are unaware of the interference they cause to other surrounding nodes. In these networks, excessive interference at a receiver or a potential forwarding node causes severe blocking and reduction in throughput. In addition, the unbalanced interference experienced at a particular node can force the node to consume more time receiving packets rather than sending them, resulting in dropped packets due to buffer overflow. This motivated us to

develop a flow control framework for regulating the transmission and improving the overall throughput of multi-hop wireless networks based on CSMA/CA MAC protocol. The framework prevents congestion, reduces packet loss and is attractive because per-flow information at each node is kept to a minimum.

1.2 Research Approach

Before we detail the main contributions of this thesis, we will discuss some of the important assumptions that we have adopted. The two major challenges in carrying out research in wireless ad hoc networks are how to model multi-hop communication and how to model mobility. The theoretical work available for modeling wireless ad hoc networks is still in its infancy. There have been better efforts in modeling wireless communication in a single hop situation [10]. Most people in the networking research community have always studied protocols with the assumptions of a static wireless network and with the assumption that radio propagation is a uniform disk model.

Figure 0-1. Radio irregularity

Irregular radio propagation is actually a common phenomenon in wireless ad hoc networks. A typical radio transmission and interference range of a node could resemble the diagram shown in Figure 0-1. Radio irregularity is a common phenomenon, which arises from multiple factors, such as variance in radio frequency (RF) sending power, reflection, scattering and differences in path losses depending on the direction of

propagation. Zhou, *et al.* [105] have shown through empirical tests that the variance in received signal strength is largely random.

Figure 0-2. Degree of Irregularity (DOI)

Due to irregular radio ranges and variations in packet loss in different directions, the layers above the physical layer in the protocol stack will be affected. Recently, Zhou, *et al.* [105] have developed a model known as Radio Irregularity Model (RIM) to represent the irregularity of the transmission and carrier sensing ranges. Their model was based on the Degree of Irregularity (DOI) model described in [29]. The example of different DOI values and their effect on radio propagation signal is shown in Figure 0-2 above. The DOI model defines an upper bound and lower bound on signal propagation for successful communication to occur. Zhou, *et al.* [105] have used their RIM model to study the impact of radio irregularity on MAC and routing protocols. Perhaps one of the most important finding of their work is that radio irregularity has little impact on the relative performance of MAC protocols but has a significant impact on the relative performance of routing protocols.

While new models are emerging, few researchers have actually pursued the study of networking protocols by using this approach, because of the complexity and cost of performance evaluations on a simulator escalates, when the network scales up to

hundred or more nodes. Moreover, repeatable results are extremely hard to obtain from random time varying parameters.

Due to the complexity in carrying out the analysis and the fact that radio irregularity has lesser impact on the relative performance of MAC protocols, we have chosen to pursue our analysis using a more simplistic model based on a uniform disk model. Moreover, in this thesis, our primary aim is to develop enhancement to MAC protocols based on CSMA/CA MAC protocol and compare them against the original MAC IEEE 802.11 MAC protocol or alternative schemes presented by others. In all of our studies, we have decoupled the impact of the routing protocol inefficiencies from the MAC layer to accurately study the performance of the MAC protocols.

1.3 Main contributions of this thesis

The main contributions of this thesis are as follows:

- **A multi-channel MAC protocol for enhancing the throughput of wireless ad hoc networks:** We present a new MAC protocol based on the assumption that there are multiple non-overlapping channels available in the system. The protocol eradicates multi-channel hidden terminal issues, scales well with the channel availability and avoids the saturation of dedicated control channel typically encountered in other schemes.
- **A channel assignment scheme to mitigate the critically exposed node problem² in wireless ad hoc networks:** Based on the multi-channel MAC protocol, we conducted some experiments and explored methods to mitigate the

² The critically exposed node problem is a special case of the exposed node problem, which has a significant impact on the throughput and it is defined and explained in section 3.2.

critically exposed node problem. We proposed a distributed heuristics-based channel assignment scheme and conducted simulations to find the minimum number of channels that are required.

- **A synchronized and concurrent transmission scheme for improving throughput in CSMA/CA based wireless ad hoc networks:** We propose a new MAC protocol that makes use of neighboring node's location information to enable parallel transmissions.
- **A multi-rate MAC protocol for improving the capacity of ad hoc networks:** We propose a new multi-rate MAC protocol that is able to relay packets in a more efficient manner compared to existing multi-rate schemes. We developed a simple theoretical model to evaluate our scheme and an existing multi-rate MAC scheme and carried out simulations to verify the throughput and delay performance of the MAC protocol.
- **A flow control framework for improving throughput and energy efficiency in CSMA/CA based wireless ad hoc networks:** We propose a novel hop-by-hop flow control scheme for regulating the transmission and improving the overall throughput of multi-hop wireless networks. We designed a cross layer framework that utilizes the 4-way MAC protocol handshake for efficient flow control signaling and evaluated the theoretical flow control parameters that should be used in a CSMA/CA based ad hoc networks. We also developed a technique to reduce per-flow information storage at intermediate nodes.

1.4 Organization of this thesis

The remainder of this thesis is organized as follows. Chapter 2 describes the primary channel assignment-based multi-channel MAC protocol. The focus is on a thorough study of existing multi-channel MAC protocols, the inherent problems that exist in multi-channel MAC protocol designs and finally the design of our scheme coupled with extensive simulations. Chapter 3 describes the inherent throughput capacity problem that exists in wireless ad hoc networks due to critically exposed node interference. We then describe a channel assignment scheme that can be coupled with the MAC protocol described in Chapter 2 to solve the critically exposed node problem. Chapter 4 describes the parallel transmission MAC scheme for enhancing the throughput in a wireless ad hoc network. Chapter 5 presents a multi-rate MAC scheme for forwarding packets in a more efficient manner. We enhance an existing analytical model, which is commonly used for performance evaluation of CSMA/CA in a WLAN setup, to study the performance of our MAC protocol. Chapter 6 describes a flow control framework for improving throughput and energy efficiency of wireless ad hoc networks based on the CSMA/CA MAC protocol. We conclude this thesis with a summary and a discussion of future research ideas in Chapter 7.

Chapter 2. Primary Channel Assignment based Multi-channel MAC Protocol

The basic IEEE 802.11 MAC protocol operating on a single-transceiver device was not designed to capitalize on the multiple non-overlapping wireless channels. In this chapter, we propose a new MAC protocol scheme known as the “Primary Channel Assignment based MAC” (PCAM) for increasing the per-node capacity and overall throughput of system through the use of multiple channels. Although a number of schemes for multi-channel MAC exist, most of the methods that have been proposed so far require either a frequency or time separated dedicated control channel which suffers from saturation at heavy loads.

In contrast, the PCAM scheme does not require any dedicated control channel or complex synchronization. It however requires 3 half-duplex transceivers per node, which can be reduced to 2 if special transceivers are used. In this chapter, we compare the performance of PCAM protocol with two other protocols, Dynamic Channel Assignment (DCA) [93] and Multi-Channel MAC (MMAC) [82].

2.1 Problems in Multi-channel MAC designs

2.1.1 Hidden terminal problem

Figure 2-1. Hidden terminal problem

Hidden terminal problems are usually present in most single channel MAC schemes and the use of Ready to Send (RTS) and Clear to Send (CTS) does not eliminate them completely. Only schemes such as those proposed by Deng, *et al.* [18] is known to eliminate hidden terminal problems completely. In designing a multi-channel scheme, the hidden terminal problem has to be addressed carefully. Wu, *et al* [93] gives an example of how hidden terminal issues arise when a single-channel 802.11 design is adapted to a multi-channel scenario. Another example problem that applies to designs that employ channel switching without prior reservation in the newly switched channel is depicted in Figure 2-1. Node A is communicating to node B in a common channel. Node C switches to this channel around time T1 where it misses the CTS sent out by B. Using its carrier sense mechanism, Node C will assume the channel to be idle since it did not receive the Network Allocation Vector (NAV) sent out by Node A prior to the switch. Node C, which is a hidden node to Node A, will also physically sense an idle channel immediately after switching to the new channel. If Node C proceeds to transmit to some other node after sensing an idle channel, the signal could collide at B.

2.1.2 Busy Receiver

Another issue that needs to be taken into consideration is the availability of the receiver when a transmitter switches to presumed receiver's channel. In a design where the receiver is tuned to another channel, while this switching occurs, it will not be able to listen to broadcast or RTS signals.

2.1.3 Broadcast messages

Wireless ad-hoc network protocols are dependent on the broadcast nature of wireless medium to operate effectively. Route discovery, route maintenance, address resolution are some of the messages that depend on broadcast messages. Due to mobility, congestion and unpredictable wireless channel in wireless ad-hoc networks, nodes usually lose connectivity and have to update their routes frequently. In designing a multi-channel scheme, we have to take into consideration the need for broadcast messages. It is inefficient for a multi-channel scheme to have nodes switch to another channel and miss out on the broadcast messages. In schemes that use a separate dedicated control channel, nodes will be required to use this channel to broadcast to others. We propose a multi-channel MAC protocol that addresses all these problems.

2.2 Related Work

There have been many research efforts on multi-channel MAC protocols for wireless ad-hoc networks in the past few years covering contention based MAC schemes [38,63,64,82,86,89,93,94]. The use of multiple channels by a single node in contention-based schemes in a sense is not new and was introduced by Kleinrock and Tobagi [49], with the primary aim of combating hidden terminal issues. Similarly, Deng, *et al.* [18]

extended the busy tone idea by using 2 separate tones to eliminate hidden terminal problems. Gain in throughput is achieved by effectively eliminating the hidden terminals transmissions. The drawback of these schemes is that the former requires 2 transceivers while the latter requires 3 transceivers. One of the main issue that is addressed in all multi-channel protocol design is the method by which channel assignment is performed. Nasipuri, *et al.* [63] studied the performance of multi-channel CSMA protocol in a fixed bandwidth system that is divided into several channels. The main goal of study is to show that multi channel CSMA can reduce collision significantly in wireless networks. The scheme distinguishes itself from the rest by using carrier sensing mechanism to perform channel assignment as opposed to other schemes that use broadcast scheduling or via additional messaging. The so called “soft channel reservation” gives preference to channels that was successfully used last. The scheme basically uses the same back off principle as IEEE 802.11 and in addition senses the channel for an extended period before selecting the channel. The study covers 2 types of receivers; a single user receiver and a multi-user receiver. The scheme however does not address the hidden terminal problems highlighted in Figure 2-1. The scheme also does not address the problem of busy receiver for single user receiver or in a special case the deadlock issue discussed by Wu, *et al.* [93].

Jain, *et al.* [38] proposed a CSMA-based MAC called receiver-based channel selection (RBCS) that uses short reservation frames over a dedicated control channel. The scheme makes use of the receiver-side channel state to select the best channel at the sender that reduces collision at the receiver. One control channel and N data channel is used in the scheme and RTS/CTS exchanges only takes place in the control channel. The use of separate control channels also eliminates the interference between data and control packets. The study is based on a slightly different motivation whereby a fixed

bandwidth system, is broken up into multiple channels to reduce collision and enable more concurrent transmissions and thus improving the performance. To address nodes that might switch to another channels and miss out on control signals, the scheme also assumes a multi-user receiver that is capable of receiving in different channels at the same time. One of the major drawbacks of this scheme is the presence of a fixed bandwidth control channel that might cause saturation when the number of nodes increases. The only obvious method for sending broadcast message to all nodes via the control channel also means that the control channel is further loaded (or crowded).

Tzamaloukas and Garcia [89] introduced a MAC protocol for frequency hopping systems called RICH-DP which is a receiver-initiated collision-avoidance scheme that does not require carrier sensing or assignment of unique codes. The motivation for such a protocol is to improve the capacity by reducing collision and hidden terminal issue while making use of the available frequency hopping bands (channel). The scheme requires the nodes in the systems to use a common frequency hop sequence. A dwell period is defined within each hopping sequence for the exchange of collision avoidance signals between receiver and sender. Nodes that are successful in exchanging the control signals will remain in the same channel and continue transmitting data while the rest of the nodes will continue on hopping in search for an available channel. The major drawback of this scheme is the lack of a clear method for sending broadcast messages to all neighboring nodes, which is seen as an essential function in ad-hoc networks. Since nodes switch to other channels in some predefined sequence, the sender will not be able to communicate its broadcast message.

Wu, *et al.* [93] proposed Dynamic Channel Assignment (DCA) that uses 2 transceivers and divides the overall bandwidth into one control channel and N data channels. The

scheme utilizes two lists; a “current usage list” (CUL) maintained by the receiver and a “free channel list” (FCL) maintained by the sender. The main idea of this scheme is to match the channel that is free at the receiver end with the channels suggested by the sender. The scheme uses a reservation control frame (RES) in addition to RTS and CTS in the control channel to complete the channel assignment while DATA and ACK signals are sent in data channels only. The major drawback of this scheme and that proposed in [38] is the limitation in bandwidth utilization due to control channel saturation. Wu *et al.* [93] points out that the number of channels is limited to no more than $L_d/3L_c$ whereby L_c is the length of the control frames (CTS+RTS+RES) and L_d is the length of the data frames. The scheme favours data messages that are larger. Although the load of broadcast signals in ad-hoc networks was not studied, this would further reduce the capacity since broadcast signals will have to be sent on the control channel. Also pointed out in [82] is the waste of channel utilization when the number of channels in a system is low. The scheme suggest the use of 2 transceivers; one fixed to the control channel and another free to switch to other data channels.

One of the most recent scheme proposed by So and Vaidya [82] is based on the use of power saving mechanism (PSM) concepts adopted from 802.11 for the design of multi-channel scheme. The scheme in essence also utilizes a dedicated control channel but is time multiplexed rather than channel separated. Periodically transmitted beacons are used to define a channel negotiation (ATIM) window and a data transmission interval. The ATIM window in one of the channels is used as a dedicated control channel where all nodes have to listen and contend for the available channels. Once reservation is successful, the nodes switch to the selected channel and sends data. The scheme also proposes the use of RTS, CTS and RES in the control window for negotiation of the data channels. The scheme proposed has some advantages because it uses only a single

half-duplex transceiver and reuses techniques designed for 802.11 protocol without much modification. The major drawback of the scheme is the need for synchronising beacons which might be difficult to implement in a wireless ad-hoc network environment. Other drawback of this scheme is the waste of the bandwidth during the period defined by the ATIM window in other channels. If a fixed ATIM window period is used, it might also cause control channel saturation.

Compared to the previous works, our proposed scheme does not require any separate control channel or any complex synchronization.

2.3 The PCAM protocol

Our main goal is to design a MAC protocol that removes the constraints of control channel saturation, complex time synchronization and at the same time to improve the maximum utilization of channel bandwidth. In the proposed scheme, three half-duplex transceivers are used by a single node. Two transceivers, namely the primary and the secondary transceivers, are mainly used for data transmission purposes. A third transceiver is used for transmitting and receiving broadcast messages. In some special cases, the third transceiver can also be used for data transmission. In PCAM, the primary transceiver is assigned a certain channel known as the “primary channel”, which will be made known to the neighboring nodes via a channel assignment protocol. Since the focus of this chapter is not on the channel assignment protocol, for simplicity of the study, we assume the existence of a distributed protocol, which can be used for channel assignment. The channel assignment scheme will be covered in the next chapter. In PCAM, the primary transceiver with an assigned “primary channel” serves as a means for other nodes to contact this node. Ideally, “primary channel” assignments should not be the same for nodes that are placed two hops away from each other but

this is impossible in most situations where the number of nodes can far exceed the number of available channels.

Figure 2-2. Channel assignment in PCAM

The secondary transceiver is used mainly for sending data and is not assigned any fixed channel. Figure 2-2 illustrates the possible communication scenarios and the channel assignments used by the transceivers. We assume channel 1 is assigned as the broadcast channel in the example scenarios above. When two nodes have different primary channel (P_{ch}) assignments (Figure 2-2a), the sender will use its secondary transceiver to communicate to the primary transceiver of the receiver by switching to the receiver's primary channel. In the example above, the sender switches its secondary transceiver's channel (T_{ch}) to channel 4. When both nodes have the same primary channel, $P_{ch} = 3$ (Figure 2-2b), nodes will make use of their primary transceivers to communicate with each other. In such a scenario, the secondary transceivers of both nodes can be used to communicate to other nodes in other channels. In Figure 2-2c, the receiver uses the broadcast channel (channel 1) as the primary channel. In such a case, the tertiary transceiver at the receiver is disabled (indicated as $T_{ch} = \#$) and the sender uses the tertiary transceiver to communicate with primary transceiver at the receiver.

Figure 2-3. Channel assignment and transceiver arrangement for a 6-node chain setup

To give a better illustration of the PCAM protocol, a chain setup with 6 nodes is shown in Figure 2-3. In this particular setup, the tertiary transceivers are assigned channel 1. The primary channel for node 1 is channel 2, node 2 is channel 3, node 3 is channel 4, node 3 is channel 2, etc. When node 1 wants to communicate to node 2, it switches its secondary transceiver to channel 3. If any of the nodes wants to broadcast a message, it uses its tertiary transceiver, which is set at channel 1 to send out the broadcast message.

Since the primary channels are pre-assigned, the scheme does not require any dedicated control channel to negotiate channels dynamically. Unlike other schemes where channel assignments occur regularly and require nodes to rendezvous in a common channel, nodes in this scheme are always contactable via their primary channel. The scheme has no restrictions on the total number of channels that can be utilized in the system as compared to those schemes that require a dedicated control channel. In schemes that require a dedicated control channel, the limited bandwidth of the control channel normally prevent nodes from exchanging control messages freely to reserve the available channel. In such cases, the control channel saturation limits the utilization of the multiple channels.

In our design, if a node wants to broadcast to all neighboring node, it would be infeasible for the secondary transceiver to switch to all channels to broadcast. Such an operation will be too time consuming and incurs delays that scales up with the number of channels used in the system. For this reason, a tertiary transceiver is used for

broadcasting purposes. A fixed common channel is assigned in the system and all tertiary transceivers are tuned to this common channel. Having a third transceiver just for implementing broadcast messages has its advantages since we can separate management signals from data signals. Routing messages should take precedence over data transmission since each data packet is dependent on the state of the routing table. By separating the management and data messages by channel usage and ensuring less contention in the broadcast channel, we can ensure some form of priority for management messages. In our proposed scheme, if one channel is dedicated solely for broadcast purposes, it may be underutilized, especially when the number of nodes in the vicinity is small. Because of this, we propose to share this channel for data transmission purposes as well. When nodes are assigned the same primary channel as the broadcast channel, the primary transceiver will handle broadcast messages and incoming data as well. The third transceiver could be switched off or used for transmitting data on another channel.

In the proposed scheme, route discovery is implemented using the broadcast channel. Nodes send out, for example, AODV Request [96] on the broadcast channel and receive replies on the same channel. Similar to route discovery process, primary channel discovery can be achieved by piggybacking the information in the routing packet or implementing a separate set of query-response messages. Using for example the AODV Reply [96] messages, the downstream sender can append its primary channel to this message to notify the upstream node.

As pointed out in Section 2.1, nodes can potentially become hidden terminals in the new channel and disrupt the ongoing conversation in the new channel. So and Vaidya [82] proposed that nodes could wait a complete maximum data transmission cycle

before transmitting when they switch to another channel. Such a method could introduce more delays since nodes could potentially wait for one maximum MAC protocol data unit even if they sense the channel to be idle. Wireless receivers are designed to exploit the packet capture phenomenon whereby the stronger signals are recovered even in the presence of interfering signals. The power capture method [4] typically used in wireless hardware allows capture to occur when the stronger packet arrives almost simultaneously, regardless of arrival time but within a stipulated interval T_c known as capture time. There are two stages required for successful reception of a frame by a radio transceiver. Initially, the frame must be detected properly and subsequently captured by the receiver.

Figure 2-4. Two-Ray propagation and path loss

Distance (m)	Power (dBm)
200	-90.49755
250	-94.37395
300	-97.54798
350	-100.21907

Table 2-1. Distance vs. received power

To solve the problem caused by interfering hidden terminals, we introduce a new value for transmission range and capture threshold. Typical receivers are able to perform capture when the stronger signal is 4-10 dB higher than the weaker unwanted signal

[52]. Figure 2-4 illustrates the use of this new threshold to reduce the hidden terminal problem. Based on a typical 802.11 system, the power levels at various distance is shown in Table 2-1. By assuming the node positions shown in Figure 2-1, a capture threshold of 10 dB and using the values shown in Table 2-1, we can explain the hidden terminal problem. If node C is at 300 m from B and B is 250 m from A, the power ratio of the packets received at B, P_A/P_C , will be less than 10 dB, thus causing the packet received at B to be discarded. To prevent this phenomenon and increase the chance of capturing, we reduce the effective transmission range threshold of the system from Tx_{dist} to Tx'_{dist} by increasing the received power threshold from PTx to PTx' while keeping the same transmission power. In ns-2 [65], the typical 802.11 transmission range, carrier sense range and capture threshold values used in the model are 250 m, 550 m and 10 dB, respectively. To solve the hidden terminal problem, we propose to reduce the transmission range to 200 m and capture threshold to 8 dB while maintaining the same carrier sense range. Based on Figure 2-1, if node B is able to communicate to node A, then nodes A and B have to be within 200 m of each other. Node C, however, has to be more than 550 m from node A to sense the channel as idle at time T1. Given the worst-case scenario, where node A is 200 m away from B and node C is 350 m away from B, the power ratio of the packets received at B, P_A/P_C , will be more than 8 dB, thus ensuring successful reception of packet at node B.

2.4 Performance Evaluation

In this section, we present the performance evaluation of the proposed PCAM protocol. The protocol is implemented in ns-2 [65] with CMU wireless extension [97] and is compared with the DCA [93] and MMAC [82] protocols on the same simulation platform. The channel bandwidth is assumed to be 1 Mbps. In our PCAM

implementation, the capture threshold is reduced to 8 dB while the effective transmission range is reduced to 200 m to handle the hidden terminal problem. The carrier sense range is set to 550 m for all the three protocols, namely PCAM, DCA and MMAC. The transmission range for the DCA and MMAC protocols is set to 250 m.

Beacon Interval	100 ms
ATIM window size	20 ms
ATIM RES	304 bits
ATIM	304 bits
ATIM ACK	304 bits

Table 2-2. MMAC protocol parameters

Unless stated otherwise, in all our experiments, the packet size is fixed at 512 bytes and 4 channels are used. All channels are assumed to be noiseless and error free. Thus, packet losses are due to collision. The MMAC parameters such as beacon interval, ATIM window size, etc. have the same values as set in [82] and are shown in Table 2-2. For all simulations, the packet arrival rate varies to simulate the various offered loads. The simulated time for all simulations is set to 200 seconds and the routing protocol used is AODV [96]. The following sections discuss the various experiments carried out, assuming stationary nodes. In most experiments, we vary the offered load and observe the aggregate throughput from the simulation results. Aggregate throughput is defined as the sum of the throughputs delivered to the all the active users in a system.

2.4.1 Wireless LAN

Figure 2-5. WLAN with low node density

Figure 2-6. WLAN with medium node density

The first setup is a Wireless LAN configuration whereby all nodes are within the transmission range of each other (co-located). Two scenarios to simulate low and medium contention are used. The number of nodes in each scenario is set to 4 and 16, respectively, while the number of flows is half the number of nodes. For PCAM protocol, channels are evenly assigned among the nodes with some of the nodes having to use the broadcast channel as their primary channel (as shown in Figure 2-2

Reference source not found.c). Figure 2-5 shows the throughput performance of the 3 protocols at various offered loads for a 4-node case. When the network load is low, all protocols have the same performances. This happens because, all packets that are injected into the network is serviced (transmitted). At higher offered load, where packets arrival rate is greater than 300 packets-per-second, the MAC protocol reaches a saturation point whereby, any increase in offered load will not cause anymore increase in channel utilization. In the low offered load case, there is a throughput gain in PCAM of about 8.5% with respect to DCA. MMAC performs the worst in this scenario, i.e., when the number of flows is small. Under a co-located scenario, there are no hidden node or exposed node complications. The performance of the MAC protocols is dependent on the timing sequence of the message exchanges, channel acquisition time and the backoff timer due to contention of nodes. Therefore, under such a scenario, we do not expect significant improvement of PCAM over the DCA and MMAC protocols. Figure 2-6 shows the performance of the protocols under medium level of contention (16 nodes) and the situation where the number of nodes exceeds the number of channels. Under heavy offered load conditions, where packet arrival exceeds 300 packets-per-seconds, the aggregate throughput of PCAM protocol shows 25.5% gain over the MMAC protocol. DCA performs slightly worse than MMAC under heavy offered load condition because the control channel in DCA encounters more contention and therefore allows less successful data exchange.

Protocol	Data Payload (bytes)		
	64	512	1024
PCAM	1034.81	2758.54	3022.48
MMAC	886.03	2176.65	2338.54
DCA	355.68	2007.96	2398.83

Table 2-3. Aggregate Throughput vs. Packet size

The same 16-node experiment was repeated by varying the packet size for various packet lengths while keeping the packet arrival rate at 200 packets/sec. Table 2-3Error! Reference source not found. shows the aggregate throughput of the 3 protocols. When the packet size is small, DCA performs dismally due to the saturation of the control channel. This happens because the control channel in DCA is only able to admit a fixed number of data packets, which is governed by the control channel bandwidth and the control frame size. The number of packets that can be admitted is also linked to the number of optimal channels that can be utilized by DCA. Wu *et al.* [93] have pointed out that the number of channels is limited to not more than $L_d/3L_c$ whereby L_c is the length of the control frames (CTS+RTS+RES) and L_d is the length of the data frames. DCA however performs slightly better than MMAC when the packet size is large. In all cases PCAM performs better than the rest and throughput is higher than DCA by a significant 26% when the packet size is high. PCAM naturally outperforms these two other protocols due to the omission of control channel in the design.

Figure 2-7. WLAN with flows in opposite directions

We also studied a special case of a 4-node setup with 2 pairs of nodes having flows in opposite directions. The aggregate throughput of the PCAM protocol (Figure 2-7) shows 114% gain compared to the DCA protocol. Such throughput gain is possible due to the fact that PCAM allows multiple transceivers to be utilized simultaneously for data transmission, effectively doubling the per-node capacity in this instance.

In another experiment, we used 36 nodes with 18 flows and studied the throughput by varying the number of available channels between 8, 12 and 16. In Figure 2-8, “DCA-8” indicates DCA protocol with 8 channels.

Figure 2-8. Scalability with respect to available channels

Figure 2-8 shows that the throughput of PCAM scales almost linearly as the number of channels increase. In all three protocols, the aggregate throughput reaches a maximum saturation point when offered load for each node is increased. DCA does not show any throughput improvement with increasing number of channels in the example above

because of control channel saturation. As discussed previously, this happens when the number of channels exceed the value of $L_d/3L_c$. In fact, once saturation of control channel occurs in DCA, the aggregate throughput decreases slightly as the number of channels increases. This is due to the fact that, as the number of channels increase, the available channel in the “free channel list” increases causing the nodes in the vicinity to believe that there are many freely available channel. This causes more collision in the control channel as more senders try to contend for the free channels. As a result, the number of successful reservations decreases. The same problem occurs for MMAC since MMAC also maintains a similar list known as the “Preferable Channel List”. For the optimal ATIM window size suggested in [82], MMAC shows saturation of control window around 12 channels. Figure 2-8 shows that there is higher performance with MMAC-12 when compared to MMAC-8 but the throughput performance in MMAC-16 diminishes when compared to MMAC-12. Overall, MMAC shows better channel scaling compared to DCA.

2.4.2 Chain

Figure 2-9. Chain setup

The second setup is a chain of nodes as shown in Figure 2-9 above. Each node is placed at a distance of 200 m from their neighboring nodes to ensure that they are within communication range of each other. The aim of this experiment is to study the end-to-

end delay and throughput in a multi-hop network. Traffic of various loads is sent from node 1 to node 6. The primary channel for node 1 is 1, node 2 is 2, node 3 is 3, node 4 is 4, node 5 is 1 and node 6 is 2. Channel 1 is also assigned as the broadcast channel. Figure 2-10 shows a throughput gain of 434% for PCAM over the DCA protocol while MMAC performs dismally in the chain configuration.

	MMAC	DCA	PCAM	1-channel 802.11
Throughput (kbps)	33.550	132.986	707.489	53.115
Delay (sec)	1.3676	0.8871	0.3699	0.4577

Table 2-4. Throughput and Delay performances in Chain setup

Figure 2-10. End-to-end Throughput in Chain setup

Table 2-4 also shows the average delay for the 3 protocols when packet arrival rate is fixed at 200 packets/sec. The results show that PCAM has very low delay compared to the rest of the protocols. MMAC performs the worst since only 20% of the channel dedicated for exchange of control signals can be used for broadcasting purposes. Routing messages from all nodes have to wait for this ATIM window period and contend with other control signals to setup and maintain the routes. Because of hidden

and exposed terminals, the ATIM window is further subjected to contention from neighboring nodes. In the case of DCA, the throughput and delay are slightly better because the entire bandwidth of the dedicated control channel can be used for broadcast messages. However, the dedicated control channel has to be shared between control signals (RTS, CTS, RES) and broadcast messages. The control channel in DCA is also subjected to contention from exposed and hidden nodes.

2.4.3 Multi-hop lattice

Figure 2-11. Multi-hop lattice

Figure 2-12. Throughput performance in ad hoc setup

The third setup is meant for studying the performance of the protocols in a multi-hop wireless network as shown in Figure 2-11 above. One hundred (100) nodes were placed in a 10 by 10 grid where the distance between adjacent nodes is 140 m. We chose a distance of 140 m in the setup above to ensure that each node is able to communicate to neighboring nodes that are placed in the vertical, horizontal and diagonal axis's. Seventeen (17) pairs of source and destination nodes were selected randomly and throughput and delay were studied for various offered loads. For the PCAM protocol, the four (4) primary channels were interleaved from nodes 1-100. The experimental result was averaged over 12 simulation runs. Figure 2-12**Error! Reference source not found.** shows the aggregate throughput of the three protocols.

The PCAM protocol outperforms the MMAC protocol with regard to the aggregate throughput. Indeed, PCAM leads to an improvement on aggregate throughput in the order of 140 % compared to MMAC under heavy load condition. The performance of MMAC degrades in a multi-hop scenario due to difficulty in sending and receiving the broadcast messages during the ATIM window. As a result, there is significant latency in route setup and frequent route breaks occur because of loss of AODV Hello messages during route maintenance. In MMAC, flows with lower hop count are more successful in getting their messages across while flows with higher hop counts are starved.

The throughput shown in Figure 2-12**Error! Reference source not found.** for MMAC is mostly contributed by the flows with lower hop count. The throughput of DCA is low because of two reasons. The first is the control channel saturation effect and the second reason is the increased contention between routing messages and data messages. Table

5-2 shows the end-to-end delay observed when the packet arrival rate is set to 200 packets/sec for all flows.

DCA	MMAC	PCAM
1.1815	0.3629	0.5805

Table 2-5. Average end-to-end delay (in seconds)

The results show that MMAC has the lowest end-to-end delay when compared to the other two protocols. This happens because the packets, which are successfully delivered to the destination, are those that have smaller end-to-end hop count. In DCA and PCAM, flows with higher hop count are also successful. PCAM shows significantly lower delay compared to DCA.

2.4.4 Multi-hop with mobility

Figure 2-13. Aggregate throughput performance in a mobile ad hoc setup

The fourth setup is meant for studying the performance of the protocols in the presence of mobility. Fifty (50) nodes were placed randomly in a 1500 m by 1500 m area and allowed to move randomly using the Random Way Point mobility model [43]. The speed of nodes was increased from 5 m/s to 20 m/s and a pause time of 5 seconds was used in all setup. Constant bit rate (CBR) connections with a rate of 20 packets per second between sender and receiver pairs were started at random time intervals. Each simulation lasted for 200 seconds and the results were averaged over 12 simulation runs. An average 35 connections were used in each simulation run. Four channels were used in this simulation. For the PCAM protocol, the four (4) primary channels were interleaved from nodes 1-50. Figure 2-13 shows the aggregate throughput of the three protocols.

Consistent with the lattice scenario, the PCAM protocol outperforms the MMAC and DCA protocols with regard to the aggregate throughput. PCAM shows an average improvement in the order of 110% compared to DCA protocol for the aggregate throughput. The performance of MMAC is worse than DCA for all the different speeds that were tested. In all three protocols, the aggregate throughput decreases with the increase in speed because there are more occurrences of route breakages when the speed increases. The performance of MMAC suffers in a mobility scenario due to difficulty in sending and receiving the broadcast messages during the ATIM window. Similar to the chain setup, there is significant latency in route setup and frequent route breakages due to the inability of receiving AODV Hello messages. In MMAC, flows with lower hop counts are more successful in getting their messages across while flows with higher hop counts are starved. The throughput of DCA is also low because of control channel saturation effect and the increased contention between routing messages and data messages.

2.5 Conclusions

This chapter describes a new multi-channel MAC protocol scheme. We have also compared the performance of the proposed protocol with two other recent multi-channel MAC protocols, namely MMAC and DCA protocols. The results show that the PCAM protocol has the best aggregate throughput and end-to-end delay when compared to other protocols. The performance results highlight the importance of the broadcast mechanism in wireless networks and its impact on the throughput and delay for multi-channel MAC protocols. We have shown that by isolating the broadcast messages from the data messages, throughput and delay can be improved tremendously. We have also shown that if a channel is dedicated for the control purposes, it results in severe performance degradation of the multi-channel environment and the failure of the protocol in taking advantage of the increasing number of channels.

Chapter 3. Channel Assignment for Nullifying the Critically Exposed Node Problem in Ad Hoc Wireless Networks

Interference or collisions can be classified into two types. The first is direct (primary) interference, where a pair of nodes that are within transmission range, tries to transmit to each other simultaneously. The second is the hidden terminal (secondary) interference, where nodes not in the transmission range of each other but at 2-hops away, transmit to the same receiving station intentionally or non-intentionally. In the latter case, two senders actually try to communicate to two different receivers but cause interference at one of the receivers. The primary interference can be avoided by using random access with some testable conditions such as carrier sensing, whereas the hidden terminal problem is solved by using additional signaling such as busy tones in DBTMA [18] or channel assignment in [8,34,58]. In protocols such as IEEE 802.11, the hidden terminal problem is not totally eliminated but reduced by using a combination of RTS-CTS signaling and carrier sensing. While hidden terminal has been shown to affect throughput, there are also other factors that cause performance degradation of wireless networks. Particularly, CSMA/CA protocols such as 802.11 suffer from other problems such as exposed node problem and inefficiency of binary exponential backoff mechanism. In this chapter, we provide an extensive study on a special case of exposed node problem, which we term as the “*critically exposed node problem*.” While exposed node problem has been discussed in great detail in previous

works [9,81,99], there have been few solutions to address this problem. To our knowledge, there has been no solution to address the “*critically exposed node problem*.” We suggest a couple solutions to address the critically exposed node problem, which is based on regulating channel access rate and channel assignment. We then elaborate on the latter solution based on channel assignments using a cluster-based approach and multi-channel MAC. The question of how many channels are required to eliminate the critically exposed node in a regular shaped cluster will be investigated in this chapter. We also propose a distributed sub-optimal heuristics-based algorithm for channel assignment, applicable for general topologies.

3.1 Wireless protocol design issues

Before discussing the critically exposed node problem, we discuss some important design considerations in CSMA/CA based wireless systems. In all wireless systems, due to radio propagation characteristics, the average power received at a receiver, P_r , from a transmitter at a distance d is given by [78]

$$P_r = P_t \left(\frac{\lambda}{4\pi d} \right)^n g_t g_r \quad (\text{Eq. 3-1})$$

where P_t is the transmitter power, λ is the carrier wavelength, n is the path loss exponent, g_t and g_r are antenna gains at the transmitter and receiver, respectively. The value of n is at least equal to 2 and depends on the physical environment. In a typical wireless system, the physical hardware imposes a certain Signal-to-Interference (SIR) Ratio condition in order for the transmission to be successful. We adopt the physical model as described in Gupta and Kumar [28] to study the successful transmission in the wireless network. The SIR condition can be stated as follows. Let X_i represent the location of node i and Z be the subset of nodes transmitting on the same channel within

the vicinity at the same time instant. Assuming that the transmission power P_t is equal for all nodes, for successful reception to occur at node j , when an intended transmitter i , $i \in Z$, is transmitting, the following condition [28] has to be satisfied.

$$\frac{\frac{P_t}{|X_i - X_j|^\alpha}}{N_o + \sum_{k \in Z; k \neq i; k \neq j} \frac{P_t}{|X_k - X_j|^\alpha}} \geq \beta \quad (\text{Eq. 3-2})$$

whereby, N_o is the ambient noise level and β is the minimum SIR necessary. In a CSMA/CA protocol, this physical layer constraint given by inequality (Eq. 3-2) is normally mapped into the design of the MAC protocol. A certain guard zone in the form of a carrier sensing range is built into the MAC layer protocol such that upon sensing a carrier, nodes intending to transmit will backoff and prevent interference. The guard zone design is normally based on a single interfering node and can be described as follows. This is similar to the protocol model used by Gupta and Kumar [28]. Using the same notation as in inequality (Eq. 3-2), for successful reception to occur at node j , when an intended transmitter i and another nearby transmitter k are transmitting simultaneously, the ratio of guard zone to transmission zone, δ , is given as:

$$\frac{|X_k - X_j|}{|X_i - X_j|} \geq \delta = \frac{CS_range}{Tx_range} \quad (\text{Eq. 3-3})$$

where CS_range is the carrier sensing range and Tx_range is the transmission range. We normalize Tx_range to make it 1 to simplify our analysis in this chapter. While the mapping of inequality (Eq. 3-2) to (Eq. 3-3) is based on a single-node interference, in reality there could be interference from multiple nodes transmitting simultaneously outside the guard zone, causing failure of reception at the main node. Based on

inequality (Eq. 3-3), since P_t , which translates to the Tx_range , is normally fixed, designing for more interference would result in higher δ or CS_range . This would lead to more nodes being exposed and hence channel underutilization. The value of δ is normally chosen based on a tradeoff between exposed node interference and probability of reception failure due to multiple interferences, and it is also based on the transmission rate. In a typical system, δ is greater than 2 because of radio propagation properties. A typical 1 Mbps 802.11 system such as Lucent's WaveLAN uses δ and β as 2.2 and 10dB, respectively. A 11 Mbps 802.11b system can have a δ value of around 4.47 [5].

3.2 Exposed node problem

Figure 3-1. Regions in CSMA/CA based protocols

Most wireless MAC protocols are designed to address the harmful effects of hidden terminal interference. While there have been many past research efforts on reducing hidden terminal problems [8,18,34,58,88] in wireless MAC protocols, there have been very few research efforts addressing the exposed terminal problem in CSMA/CA-based MAC protocols. In this section, we epitomize the various exposed node problems affecting throughput in CSMA/CA-based wireless ad hoc networks and explore some

of the known works addressing these issues. We then classify the most critical exposed node problem and try to address it in later sections. Figure 3-1 shows the various regions in a CSMA/CA-based MAC protocol when S1 and R1 are the sender and the receiver, respectively. In general, any node that falls in region II, III, IV and VI when S1 is transmitting is an exposed node. In 802.11, nodes in region II and III are exposed mainly due to the feedback from the virtual carrier sense (NAV) mechanism, whereas nodes in region IV and VI are exposed mainly due to the feedback from the physical carrier sense mechanism. When R1 is transmitting CTS and ACK messages, all nodes in the region II, III, IV and V will be exposed.

Bharghavan *et al.* [9] have previously suggested that exposed nodes could transmit in parallel. Recently, Shukla, *et al.* [81] proposed an enhancement for 802.11 whereby nodes could actively listen to sender's RTS and DATA messages and deduce that they are in the exposed region of the main transmission. The exposed node may proceed to transmit its DATA packet in parallel to the main sender. This scheme, however, works for nodes residing in region III only and is restricted to transmission flowing out of region III to region IV (e.g. S2 to R2) or VI and not in the reverse direction (e.g. S3 to R3). Besides being restricted to region III, this scheme can only work if the exposed node in region III does not cause degradation to signal-to-interference ratio at receiver R1 as stated in inequality (Eq. 3-2). This further reduces the potential parallel senders that can benefit from this scheme

While Shukla, *et al.* [81] tries to improve the throughput of current 802.11 systems by promoting parallel transmissions of exposed nodes in regions II and III, there is a more serious kind of exposed node problem that causes severe degradation in the capacity of wireless ad hoc networks. In Figure 3-1, when node S1 is communicating to node R1,

node S4 can initiate a transmission by sending an RTS because it senses the channel as idle. However, a receiver such as R4, which could be residing in region IV or VI will not reply since it senses the channel as busy. The location dependent interference causes disparity in the carrier sensing state between the sender, S4, and receiver, R4. Due to this disparity and the use of binary exponential backoff (BEB) mechanism, a CSMA/CA scheme can experience severe degradation in multi-hop capacity. In an example CSMA/CA scheme such as 802.11, when a node such as S4 sends an RTS to R4 but fails to receive a CTS, it will double its contention window. Because node S4 senses an idle channel, it will retry the RTS until the maximum retry limit is reached before dropping the RTS and declaring the route as broken. In 802.11, the maximum retry limit is 7. The route discovery process is then tried again before sending out a new RTS. We term the node such as R4 in flow S4-R4 as the “*critically exposed node*”, when S1 is the sender. In the rest of this chapter, we shall refer to flow S4-R4 as the “*critically exposed pair*.”

3.3 Experiments to study the critically exposed node problem

3.3.1 Single channel MAC

Figure 3-2. Throughput at R1

Figure 3-3. Throughput at R4

Figure 3-4. Throughput fairness for R1/R4

We conducted a simple experiment using NS-2 [65] to study the critically exposed node problem using a two-flow setup. Based on Figure 3-1, we only used flows S1-R1 and S4-R4 in the setup. We used a 1 Mbps 802.11 MAC protocol in the test setup. We varied the packet inter-arrival time and data packet size to study the throughput of the two flows and RTS drop rate. We used the same arrival time and packet size values for the two flows.

Figure 3-2 and Figure 3-3 **Error! Reference source not found.** shows the throughput of packets received at R1 and R4, respectively. Figure 3-4 shows the ratio of throughput for R1/R4. Figure 3-4 shows that at higher packet inter-arrival times, the two flows can still achieve equal fairness (ratio=1) whereas at lower inter-arrival times, the fairness drops tremendously. The unfairness also increases as the packet size increases. An interesting observation from Figure 3-2 and Figure 3-3 show that at certain ranges of packet inter-arrival times and data packet sizes, the critically exposed pair S4-R4 can still achieve high throughput and equal fairness. A careful study also

shows that the RTS drop rate has a linear relation to the unfairness in the two flows. The maximum unfairness between the throughput at R1 and R4 can be as high as 18.5 times. To address this type of exposed node problem, we propose the following two solutions:

Proposal 1: Find the optimal packet scheduling for ad hoc forwarding. In other words, we should regulate the greedy senders by controlling their channel access rate. Figure 3-2 and Figure 3-3 suggest that this method could be possible if we limit S1's sending rate based on its packet size.

Proposal 2: Assign different channels for the main flow and the critically exposed pair.

We will not focus on proposal 1 in this chapter and the work is left for future investigation. However, we will address the critically exposed node problem by using channel assignment strategies.

3.3.2 Experiments with channel assignment and the PCAM protocol

We use the PCAM protocol, which was described in chapter 2 to conduct a few experiments to show the effect of critically exposed node problem and normal contention, on throughput of multi-hop systems. We also show how channel assignment using a multi-channel MAC protocol can be used to curb the critically exposed node effects.

Figure 3-5. Chain setup at 200 m between neighboring nodes

Experiment	Channel assignment	End-to-end Throughput (kbps)
1	1,1,1,1,1,1	48.2204
2	4,4,4,4,4,4	118.4870
3	2,3,4,2,3,4	49.1568
4	3,4,2,3,4,2	49.1568
5	2,3,4,5,2,3	705.7725
6	1,2,3,4,1,2	705.9000
7	4,1,2,3,4,1	705.8150
8	3,2,3,2,3,2	265.6170
9	2,3,4,3,2,3	264.4250
10	2,3,4,3,2,5	357.4189
11	3,2,3,4,3,2	357.3763
12	3,3,2,2,4,4	327.2438
13	3,4,4,5,5,3	327.7971
14	4,4,4,3,3,3	205.6074
15	3,4,4,4,3,3	215.8643

Table 3-1. Channel Assignment and Throughput in PCAM

We use a chain setup with 6 nodes (Figure 3-5), where the distances between two neighboring nodes are 200 meters. For the remaining part of this section, we refer to the flow from node 1 to 2 as flow 1, node 2 to 3 as flow 2, etc. In PCAM, the transmission range, shown as a solid circle is set to 200 m while the carrier sense range is set to 550 m. The packet inter-arrival time and data packet size were fixed at 0.005 sec and 512 bytes, respectively, in this experiment. UDP packets were streamed from node 1 to node 6 and AODV was used as the routing protocol. Each transceiver has a transmission bandwidth of 1 Mbps. Based on Figure 3-5, in a single channel system, critically exposed pairs 1-2 and 2-3 will experience the interference due to pairs 4-5 and

5-6, respectively. The aggregate end-to-end throughputs for various channel assignments are shown in Table 3-1. In all experiments, the different channels indicate non-overlapping channels and channel 1 is assigned as the broadcast channel. The channel assignment and transmission pattern for the transceivers in the PCAM-based nodes using experiment 3 as an example is shown in Figure 3-6. In this particular setup, the tertiary transceivers are assigned channel 1. In cases where primary transceiver uses channel 1, the tertiary transceiver is disabled.

Figure 3-6. Channel assignment and transceiver arrangement in PCAM

In experiment 1, all primary transceivers are assigned with the broadcast channel and therefore the tertiary transceivers are disabled. The throughput of the system is comparable to 802.11 running on a single channel. Flows 1 and 2 experience frequent route breaks due to the critically exposed node effect. All flows also experience normal contention. In experiment 2, we assign all primary transceivers with same channel but different from the broadcast channel. All flows experience normal contention as well as the critically exposed interference. We notice that even when flows 1 and 2 experience critically exposed interference, by separating the routing messages to the broadcast channel, the routes can be repaired faster, therefore ensuring higher throughput compared to experiment 1.

In experiment 3, three different channels excluding the broadcast channel are used but the channels are repeated 3 hops away. Intuitively we would expect this setup to perform better than experiment 1 and 2 because more channels are used, but

surprisingly the performance does not improve. The flows do not experience any normal contention hence nodes forward packets faster. However, when compared to experiment 2, flows 1 and 2 experience higher critically exposed interference due to flows 4 and 5, since the forwarding is faster. This explains why the throughput is about the same as in experiment 1. Experiment 4 is the same as experiment 3 except the channels assignment are shifted to the left. The throughput remains the same for the same reasons.

Experiment 5 shows that, by assigning different channels to the primary transceivers from the broadcast channel and with sufficient physical separation of channels, we ensure that there are no critically exposed interferences or normal contention. We notice that the maximum end-to-end throughput, which is equivalent to the maximum single-hop throughput, is achievable after taking into consideration the overheads such as RTS, CTS, ACK, etc. [55].

In experiment 6, four channels including the broadcast channel are assigned with sufficient physical separation, achieving the maximum line throughput similar to experiment 5. The throughput remains the same for experiment 7 when the channels assignment is shifted to the right. This shows that, as in experiment 6, although the broadcast channel transmission is used from node 4 to 5, the frequency of the routing messages is not high enough to cause interference to flow 1. This argument can be supported by the inter-arrival time versus fairness curve shown in Figure 3-4.

In experiment 8, two non-broadcasting channels are interleaved such that flows using the same channel are in the carrier sense range of the main sender. The critically exposed interference is present especially when flow 1 is in the forward direction while flow 3 is in the reverse direction (node 4 is sending CTS and ACK on channel 2) but is

not significant enough to cause throughput degradation. However, flows 2-hop away are using the same channels therefore resulting in normal contention (normal exposed node). Flows 1, 3 and 5 contend with each other for channel 2 while flows 2 and 4 contend for channel 3. Due to the normal contention, the end-to-end throughput is slightly more than $\frac{1}{3}$ the maximum throughput. However, the throughput is still better than experiments 2 and 3. In experiment 9, the channels are assigned in a manner that interference of only flows 1, 3 and 5 are present. The throughput achieved is the same as in experiment 8. This shows that contention in flows 1, 3 and 5 is the limiting factor for throughput in experiment 8. In experiment 10, we ensure that only flows 1 and 3 are contending for the same channel. The throughput is about half of the maximum end-to-end throughput. This experiment clearly shows that, while there is possible occurrence of critically exposed interference problem due to flow 1 in forward direction and flow 3 in reverse direction, the degradation is not significant. Similarly, in experiment 11, the throughput is half the maximum throughput because only flows 2 and 4 are contending with each other.

In experiments 12 and 13, we ensure that there are no critically exposed node problem and no repetition of the same primary channel assignments 2-hops away. The throughput is almost half the maximum throughput because 2 adjacent flows experience normal contention. The throughput is, however, lesser than that obtained for experiments 10 and 11 because the two flows that are contending for the same channel are sharing a common node. In experiments 14 and 15, flows 1-hop and 2-hop away are contending for the same channel. The throughputs are obviously worse than that obtained for experiment 8 but much better than those with the critically exposed node scenarios.

From the experiments above, we can conclude that critically exposed node problem causes severe degradation of throughput. The throughput drop caused by normal contention between communicating pairs within the transmission and carrier sense ranges, which are using the same channel is not as severe as that resulting from the critically exposed node problem. If the end-to-end flow experiences normal contention in two adjacent hops as in experiment 12, the drop in throughput is by a factor of 2.16, while the pure critically exposed interference as in experiment 3 causes a drop in throughput by a factor of 14.

3.4 Analysis of Channel assignment for eliminating critically exposed node problem

3.4.1 Design issues

Figure 3-7. Critically exposed interference in a Grid setup

From the study of throughput versus channel assignment in Section 3.3.2 , we showed that it is imperative to assign the critically exposed pair with a different channel from the main sender. We make two key observations in this section that will help us design the channel assignment scheme.

Observation 1: In an wireless ad hoc network, there can be many critically exposed nodes due to a single main transmission. Figure 3-7 shows a simple grid setup with nodes placed 200 m away from adjacent nodes. A single flow F1, transmitted from node S1, can results in 20 critically exposed pairs. Since pairs communicate on the same channel, when the reversal of transmission for flow F1 occurs (during CTS, ACK), the interference now shifts to another set of 20 critically exposed pairs.

Observation 2: In a random topology, it is difficult to pinpoint the exact location of the exposed pair due to the transmission of a particular main sender, unless some form of location information is available. Based on the typical transmission and carrier sense range of IEEE 802.11, the critically exposed node will be placed outside the transmission range but within the carrier sense range of the main sender. Because of this, the node is not able to receive any data from the main sender and determine the exact identification of the main interfering node. In a dense network topology, it is safe to assume that nodes several hops away are not in the critically exposed range of a main flow.

3.4.2 Channel assignment and guard zones

Based on observation 1 above, due to pairwise usage of the same channel, it does not matter if the MAC scheme is based on transmitter oriented or receiver oriented channel assignment scheme [8]. There is a minimum and maximum bound to the critically exposed interference due to a pairwise communication. Since, we intend to study the minimum number of channels required to eliminate the critically exposed node problem, we make the following proposition to impose the lower limit on the channels for any wireless ad hoc network.

Proposition 1: To eliminate the critically exposed interference, the maximum number of adjacent hops, h_{tx_max} , for a chain of nodes using the same channel should be less than δ , where each hop is within the maximum transmission distance.

Proof: Critically exposed interference will exist if any hop within the group of hops using the shared channel intersects the carrier sense range boundary at a normalized distance δ of any node using the same channel. Now, suppose that $h_{tx_max} > \delta$. Then under a chain configuration, the maximum number of consecutive member nodes using the same channel is $h_{tx_max} + 1$. Suppose these nodes are numbered as 1, 2, ..., $h_{tx_max} + 1$. Since each hop is of maximum transmission distance or the normalized distance 1, the distance between the two furthest members of the group of nodes sharing the common channel is h_{tx_max} which is greater than δ . Therefore, there must be one hop that intersects the boundary defined by the distance δ from node 1, causing critically exposed interference.

On the other hand, if $h_{tx_max} < \delta$, then any two nodes within the group sharing the common channel must remain within the distance δ from each other, avoiding critically exposed interference. This proves the proposition.

Figure 3-8. Critically exposed interference boundary

Obviously, critically exposed interference will not exist if the number of consecutive hops using a shared channel is less than h_{tx_max} , but this would mean that other channels have to be assigned to the remaining hops. Although this would increase throughput due to less direct contention, the number of channels required will increase. Proposition 1 also suggests a circular clustering approach for nodes using the same channel, limited by cluster diameter of h_{tx_max} hops. Figure 3-8 illustrates this concept. The dotted circles show the carrier sense or the critically exposed interference boundary due to each transmitter (1, 2, 3 and 4). In this example, a cluster using the same channel and with diameter of 2 hops is formed by nodes 1, 2, 3 and 4. Figure 3-8 shows that the individual carrier sense boundaries which cause critically exposed interference lie between two main boundaries, with radiuses $\delta + 1$ and $\delta - 1$ from the cluster center (node 2). While the inner boundary determines the value of h_{tx_max} , the outer boundary imposes the minimum co-channel reuse distance³, γ . This distance is defined by the distance between the centers of two clusters sharing the same channel and each with a diameter of h_{tx_max} hops.

Assuming that we use the clustering approach stated above, we make the following proposition:

Proposition 2: To eliminate the critically exposed interference, the minimum distance, γ , by which two clusters of nodes using the same channel must be separated is given by:

$$\gamma = \delta + h_{tx_max} \quad (\text{Eq. 3-4})$$

where, each cluster has a diameter of h_{tx_max} hops, and the distance is measured between the cluster centers.

³ The term co-channel reuse distance is widely used in literature discussing channel assignment and reuse for cellular wireless networks [41,78] and is also used here to mean the same.

Proof: Suppose we refer to the two clusters sharing the same channel as cluster A and cluster B. To eliminate critically exposed interference, the minimum distance between any node a in cluster A and any node b in cluster B must be δ . Obviously, if a and b are the nearest inter-cluster nodes, then they must be at the edge of their respective clusters. Since the clusters are of a maximum diameter h_{tx_max} , the distance of each of these nodes from their respective cluster centers must be $h_{tx_max} / 2$. It follows that, the minimum distance between the cluster centers,

$$\gamma = \delta + h_{tx_max} \cdot$$

3.5 Channel assignment strategy

Notice that in Figure 3-8, the nodes 1, 2, 3 and 4 can communicate to each other using the same channel. However if node 3 tries to communicate to 5 using the same channel, this would cause critically exposed interference problem. For node 3 to communicate to node 5, a different channel would be required. However, to achieve such a communication, where nodes such as 3 can communicate on two or more channels, we require a multi-channel MAC protocol such as PCAM. For the remaining of this chapter, we will study the channel assignment strategy based on PCAM as the MAC protocol.

Section 3.4 suggests that, for keeping a minimum channel assignment while eliminating critically exposed node problem, channel assignment should be done in a cluster-based approach. To maximize the channel reuse and minimize the number of channels, the ideal clustering algorithm should cluster the nodes and ensure equal cluster diameter of h_{tx_max} hops. Based on this observation, we propose to adopt the clustering algorithms such as that mentioned in [14,25], which is normally used for routing, to be used for

channel assignment. However, we do not preclude the use of non-cluster based routing protocols such as AODV [96], DSDV [73], etc., are also applicable with the proposed channel assignment strategy. However, we will still require a cluster-based algorithm to perform channel assignment. The ideal condition would be to combine the clustering algorithms for routing with the channel assignment scheme. For the remainder of this chapter, we study the channel assignment strategy for a cluster-based wireless ad hoc network.

3.6 Clustering and Channel assignment

Figure 3-9. Clustering schemes

We will study the channel assignment using the two clustering techniques, overlapping [14,25] (Figure 3-9b) and non-overlapping [56] (Figure 3-9a) with the assumption that cluster size can have diameter ≥ 2 hops. We then show that the overlapping clustering technique is more optimal for our requirements. For analysis, we first study the channel assignment using the overlapping clusters as shown in Figure 3-9b. We now work out the minimum number of channels required for this type of clustering.

Figure 3-10. Approximation to ideal clustering and ranges

In an ideal case, the overlapping clusters can be approximated to a 2-dimensional hexagonal cell layout as shown in Figure 3-10. To simplify our analysis, we assume the existence of a clustering algorithm that is capable of forming the clusters in a consistent manner such that each central-point of the cluster is separated from its neighbors by an equal distance throughout the network. Notice that, we refer to the center of the cluster as the central-point rather than the clusterhead, because in our clustering problem, there is possibility of having odd number of hops h_{tx_max} , defining the diameter of the cluster. The boundary nodes are placed on the edges of the hexagon and they connect two clusters together. In a special case, the boundary node can be placed at the point where three clusters overlap (Figure 3-9b) and can connect three different adjacent clusters. We will consider the channel assignment for this special case later. For a fully connected and optimally spaced out clusters, the maximum possible separation between two adjacent clusters is given as $2 \times R_m \times \sin(\pi/3)$, where R_m is the maximum possible radius of the cluster. Since the co-channel reuse distance, γ , is dependent on $h_{tx_max}/2$ (see section 3.4), which is the radius of that cluster, the number of channels required might increase as h_{tx_max} decreases. This happens because more clusters might be placed within the radius defined by the co-channel reuse distance, γ . The channel

assignment problem in a 2-dimensional hexagonal cell layout resembles the channel assignment problem in a typical cellular network. We can assign a single channel for each cluster shown above if the diameter of each cluster is h_{tx_max} . In channel assignment schemes for cellular networks, the minimum number of channels N , required to serve the entire area is related to the reuse distance σ and is given as [41]:

$$N = \left(\frac{1}{3}\right)(\sigma)^2 \quad (\text{Eq. 3-5})$$

The reuse distance σ in cellular network is defined as D/R_a , where D is the distance between the center of two cells where channels are reused and R_a is the radius of the cell. Note that the radius R_a of a cell is in fact the radius of the circle circumscribing the corresponding hexagon. Since these circles are overlapping, the distance D will not be an integral multiple of R_a in general.

In (Eq. 3-5), N can only assume integer values $\{3, 4, 7, 9 \dots \text{etc.}\}$ as generally represented by the series $U = (i+j)^2 - ij$, where i and j can assume any integer numbers [54]. Let $N(k)$ represent the series U :

$$N(k) = (i + j)^2 - ij \quad (\text{Eq. 3-6})$$

where $k = \{0, 1, 2, 3, \text{etc.}\}$, such that $N(0)$, $N(1)$, etc. represent the series U in ascending order. Mapping our ad hoc clustering problem into the cellular problem, we use R_a as the cluster radius and D as the distance between center points of two clusters. Based on Figure 3-10, for a cluster to reuse the channel assigned in the central cluster A, the cluster center must be outside the circle with radius γ and centered on cluster A. However, to find the minimum number of channels, $N_{cluster}$, required in the overlapping cluster-based wireless ad-hoc network, we cannot directly use (Eq. 3-5) as unlike D , γ

is not the exact distance between two cluster centers, but only a lower bound on the distance. Therefore to determine the value of $N_{cluster}$ we can use one of two equivalent approaches. The first method is to list the values of the ratio D/R_a in ascending order, and choose the lowest value exceeding γ , and determine the corresponding N using (Eq. 3-5). Alternatively, based on equation (Eq. 3-5), we define N_a as:

$$N_a = \left(\frac{1}{3}\right)(\gamma)^2 \quad (\text{Eq. 3-7})$$

where N_a is not necessarily an integer. To find $N_{cluster}$, we use equation (Eq. 3-7) and find $N_{cluster} = N(k+1)$ such that

$$N(k) < N_a < N(k+1) \quad (\text{Eq. 3-8})$$

Figure 3-11. Clustering for boundary nodes

The minimum number of channels used in the channel assignment also depends on the MAC protocol used. As discussed earlier, border nodes can communicate to two or three adjacent clusters depending on the placement of the border nodes and the clustering algorithm used. If border nodes can communicate to other clusters by switching effectively and maintaining the states and connectivity in an effective manner, then the minimum number of channels required is $N_{cluster}$. This is not possible when we are using standard single channel MAC protocols. If PCAM is used as the

MAC protocol, $N_{cluster} + 1$ channels will be required if border nodes are restricted to linking two clusters. The additional channel is use for broadcast purposes. If border nodes can link three clusters, then using PCAM, we require $2 \times N_{cluster} + 1$ channels. The additional $N_{cluster}$ channels required for the border nodes is justified as follows.

In Figure 3-11, the shaded area covers three edges that are adjacent to each other and suggests that, a similar cluster size to the main clusters can be used for the three adjacent edges. Since the same cluster size is used, this does not violate proposition 1 and these border nodes can be assigned the same channel. Translating this shaded area to a similar 2-dimensional hexagonal network, as shown in Figure 3-11, the minimum number of channels N_{border} required for the border nodes = $N_{cluster}$ because the same co-channel reuse constraints applies.

Figure 3-12. Example channel assignment with $N_{cluster} = 7$

Figure 3-12 illustrates an example channel of assignment using a typical 1 Mbps 802.11 system, where $\delta = 2.2$. Then $h_{tx_max} = 2$, $R_a = 1$ and $\gamma = \delta + h_{tx_max} = 4.2$. Using equation (Eq. 3-7), we get $N_a = 5.88$. From (Eq. 3-8), we select $N(k+1)$ as 7 for cluster channel

assignment. Using PCAM as the MAC protocol, and assuming border nodes link only two clusters, we need a total of 8 channels for this example.

Notice that for the overlapping cluster algorithm, we require at least $N_{cluster}$ in the ideal case. For a non-overlapping cluster algorithm (Figure 3-9a), the border nodes need to communicate across the cluster boundary. Because of this, the boundary nodes need to use different channels from those assigned to the clusters. If the distance between the two boundary nodes, shown in Figure 3-9a, is less than h_{tx_max} , then the spatial reuse for this channel is not optimized. Therefore, the overlapping cluster with boundary nodes linking 2 clusters is the preferred choice.

3.7 Channel assignment algorithm

Notice that the channel assignment problem with the knowledge of the location information is trivial. Based on their location, nodes can calculate their pre-determined channel assignment for the given coordinates, thus avoiding the critically exposed node problem. However, in the case where location information is not available, we require an algorithm that is computationally fast to assign the channels. Assigning channels in an optimal manner for hidden terminal interference avoidance in TOCA [58] and ROCA [8] based networks has been shown to be an NP-complete problem [8]. In this section, we will present a distributed sub-optimal heuristics-based algorithm for channel assignment and show the performance of the algorithm using a regular cluster-based topology shown in Figure 3-12. Although a centralized algorithm would probably result in lesser number of channels compared to the distributed algorithm, the centralized algorithm is not preferred since it works best when the topology is known and rarely changes. However, in a wireless ad hoc network, this condition is not true as frequent mobility and power constraints cause the topology and routing states to

change. Moreover, the centralized scheme has a single point of failure, and the convergence time also increases as the network size increases. Thus a good choice for channel assignment would be a distributed algorithm that can react fast to changes in the network.

3.8 Distributed heuristics-based algorithm

For convenience, the notation and terminology used to describe the algorithm are summarized in Table 3-2.

i	The clusterhead which runs the channel assignment algorithm
N_i^{ce}	The set of clusterheads whose clusters are in the critically exposed range of clusterhead i
N_i^δ	The set of clusterheads whose cluster are just outside the critically exposed range of clusterhead i and can potentially reuse the channel used by clusterhead i
$T_{request}$	The timeout interval for confirming the channel assignment request
T_{init}	The timeout interval for collecting neighboring clusters' channel assignment
$i.reuseHop$	The minimum number of hops from clusterhead i to the clusterhead whose cluster is just outside the critically exposed range
$i.degree(k)$	The number of channel reuse occurrences for clusters just outside the critically exposed range using channel k
$i.chan$	The channel selected by clusterhead i
$i.nodeID$	Node identification (ID) of clusterhead i
$i.defChan$	Default channel for clusterhead i
$Pkt.ReqNeighCh$	Packet for requesting neighboring clusters' channel assignment
$Pkt.RcvNeighCh$	Reply message from neighboring clusters indicating channel assignment
$Pkt.ReqChAsg(i)$	Channel assignment request message sent by node i
$Pkt.ChAsgConfirm$	Confirm message for channel assignment
$hops(i,j)$	Number of hops between node i and node j
$AvailableChanID$	Set of channels available to the system
$PrefChan$	Set of channels preferred for assignment
$PrefChan(n).cnt$	Number of times channel n occurs in the preferred list
$UnusableChan$	Set of channels that cannot be reused due to conflict

Table 3-2. Notations

We make several assumptions in the process of designing this algorithm. In Section 3.4.1 , observation 2 showed that it is difficult to pinpoint the exact boundary of the critically exposed interference in a sparsely connected network. However for a dense

network, we can safely assume that nodes that are several hops away are out of the critically exposed range. Based on Table 3-2, we let $i.\text{reuseHop}$ describe the minimum number of hops from a central clusterhead i to the clusterhead j whose cluster is just outside the critically exposed range. In an evenly spaced out cluster topology as shown in Figure 3-12, the same value of $i.\text{reuseHop}$ applies to all clusterheads. We assume that each node has a unique identifier and a clustering algorithm is executed prior to running the channel assignment algorithm. The algorithm can be triggered periodically or whenever there is a change in cluster formation.

```

/* Initialize */
Init(i)
{
1       $N_i^{ce} = \emptyset$ ;
2       $N_i^{\delta} = \emptyset$ ;
3       $i.\text{chan} = i.\text{defChan}$ ;
4      /* Request neighbor channel assignment */
5      Broadcast Pkt.ReqNeighCh up to distance of  $i.\text{reuseHop}$ ;
6      while( $T_{init} > 0$ ) {
7          Receive Pkt.RcvNeighCh from clusterhead  $j$ 
8              if ( $\text{hops}(i, j) < i.\text{reuseHop}$ )
9                   $N_i^{ce} \leftarrow j$ ;
10             if ( $\text{hops}(i, j) = i.\text{reuseHop}$ )
11                  $N_i^{\delta} \leftarrow j$ ;
12         }
} /* End of Init */

```

Figure 3-13. Initialization phase

```

/* Distributed Assignment */
doDistributedAssignment(i)
{
1       AvailableChanID  $\leftarrow \{1, 2, 3, \dots, n\}$ 
2       UnusableChan =  $\emptyset$ ;
3       PrefChan =  $\emptyset$ ;
4       Assigned = 0;
5       for ( $\forall j \in N_i^{ce}$ ) {
6           UnusableChan  $\leftarrow j.chan$ ; /* mark as unusable */
7       }
8       for ( $\forall j \in N_i^{\delta}$ ) {
9           if ( $j.chan \notin UnusableChan$ )
10              PreferredChan  $\leftarrow j.chan$ ;
11              /* count channel usage frequency */
12              increment PrefChan( $j.chan$ ).cnt;
13          }
14      }
15      PreferredChan  $\leftarrow \text{SortDegree}(\text{PreferredChan}, \text{PrefChan}(n).\text{cnt})$ ;
16      while ( $\neg \text{Assigned}$ ) {
17          for ( $k \in \text{PreferredChan}$ ) {
18              i.chan = k;
19              /* Request channel assignment */
20              if (doChanAssign (i.chan) = True) {
21                  Assigned = True;
22                  end;
23              }
24              else
25                  UnusableChan  $\leftarrow i.chan$ ;
26          }
27          for ( $l \in \text{AvailableChanID}$  and  $l \notin \text{UnusableChan}$ ) {
28              i.chan = 1; /* select lowest ID first */
29              /* Request channel assignment */
30              if (doChanAssign (i.chan) = True) {
31                  Assigned = True;
32                  end;
33              }
34              else
35                  UnusableChan  $\leftarrow i.chan$ ;
36          }
37      } /* End of while */
} /* End of doDistributedAssignment */

```

Figure 3-14. Main distributed channel assignment algorithm

```

/* Channel Assignment Request */
doChanAssign (i.chan)
{
1   Send broadcast Pkt.ReqChAsg(i) with i.degree (i.chan) to distance of i.reuseHop;
2   while( $T_{request} > 0$ ) {
3       Receive Pkt.ReqChAsg(j) from node j
4       if (i.chan == j.chan and hops(i,j) < i.reuseHop) {
5           /* ResolveConflict */
6           if (i.degree (i.chan) < j.degree (i.chan))
7               return false;
8           else if (i.degree (i.chan) == j.degree (i.chan)) {
9               if (i.nodeID < j.nodeID)
10                  return false;
11           }
12       }
13   }
14   if ( $T_{request} < 0$ )
15       Send Pkt.ChAsgConfirm;
} /* End of doChanAssign */

```

Figure 3-15. Channel assignment request function

The algorithm for distributed heuristics-based channel assignment is described in Figure 3-13 to Figure 3-15 using C-style pseudo code. Figure 3-13 describes the initialization phase, before a cluster starts assigning a channel to itself. To execute the channel assignment algorithm, the clusterhead needs to find out the channel assignment of its neighboring clusters up to a distance of $i.reuseHop$. When a cluster is newly formed, a clusterhead might not have any information on the neighboring clusters' channel assignment. One option is to request this partial information from its neighboring clusters. However, this information could be stale depending on how frequently the channel assignment broadcast is performed.

To get more accurate information, the cluster first assigns a default channel, $i.defChan$ to itself and starts communicating to its neighbors. It sends out a broadcast message, $Pkt.ReqNeighCh$ requesting for neighboring clusters' channel assignment up to a distance of $i.reuseHop$. To reduce flooding caused by this message, only clusterheads and border nodes will participate in the re-broadcasting of this message. These nodes will decrement the hop count in the $Pkt.ReqNeighCh$ message and help propagate this broadcast request until the hop-count reaches zero. Duplicate $Pkt.ReqNeighCh$

messages from the same clusterhead with higher hop-count are dropped. When the Pkt.ReqNeighCh message is sent out, a timer is set. Line 6, in Figure 3-13 shows that the clusterhead i waits for a timeout interval T_{init} to collect the reply messages, Pkt.RcvNeighCh. The reply messages are unicast messages. The channel assignment from clusterheads which are at distance $i.reuseHop$ are then kept in a table defined by N_i^δ , while the assignment from clusterheads which are less than distance $i.reuseHop$ are kept in a table defined by N_i^{ce} .

Figure 3-14 describes the main part of the distributed algorithm executed by clusterhead i . The node first initializes the *AvailableChanID* list with a set of valid channels in the system starting from the lowest ID. We used channel 1 as the lowest ID in this example. Lines 5-7 ensure that the channels that are already assigned within the critically exposed region are not used. The clusterhead i then proceeds to check the channels that are used by clusters just outside the critically exposed region. To enable efficient reuse of the channels and to reduce the total number of channels required, we search for the channels that are not assigned in the critically exposed region but used by the maximum number of clusterheads at distance $i.reuseHop$. We then build a preferred channel list for cluster i from the channels that do not appear in set N_i^{ce} but appears in set N_i^δ . We then build a channel list preferred for cluster i from the remaining channels in set N_i^δ . We sort this list in a descending order based on the frequency of channel reuse. In case of a tie in reuse frequency, the channel with the lowest ID is listed first. Lines 17-26 test the channels in the preferred list starting from the highest occurring channel for the assignment for cluster i .

The actual request for channel assignment is carried out by the procedure in Figure 3-15. The clusterhead i sends out a broadcast message, $\text{Pkt.ReqChAsg}(i)$ requesting information on channels being used by other clusters. A timer is set with period $T_{request}$ to wait for the responses. The clusterhead i then waits for timeout period $T_{request}$ to check if any other neighboring clusterheads, with hop count less than $i.\text{reuseHop}$ have requested for the same channel assignment by sending out a similar request. If the channel assignments requested by both clusterheads are the same, then lines 6-11, in Figure 3-15 ensure that the clusterhead with the higher occurrence of this channel reuse at $i.\text{reuseHop}$ succeeds. In cases where the degree of channel occurrence is the same, the node identification is used to resolve the tie. The node that succeeds will assign the channel after the timeout period of $T_{request}$ by sending out a Pkt.ChAsgConfirm message to neighboring clusterheads (Figure 3-15, line 15). The node that fails in the conflict resolution will drop this channel selection and try the next channel in the preferred list (Figure 3-14, lines 17-26). Once the preferred list is exhausted, the algorithm then selects the other untested channel from the *AvailableChanID* list in the system starting from the lowest ID channel (Figure 3-14, lines 27-28). By selecting the priority of channel assignment based on degree of occurrence and then followed by lowest ID, the algorithm ensures that the number of channels required remains low.

3.9 Experimental results

We implemented a simulator to evaluate the performance of the distributed channel assignment algorithm. Our objective is to study the minimum number of channels required to eliminate the critically exposed node problem using this sub-optimal heuristics-based algorithm. We used the regular shaped overlapping cluster-based topology shown in Figure 3-12 to test the algorithm. Based on the parameters selected

in the same example in Section 3.6 , the minimum channel assignment, $N_{cluster}$, was calculated as 7. Based on Figure 3-12, the cluster whose members are just outside the critically exposed range has a clusterhead with distance 6-hops away from the central cluster. Hence, we set $i.reuseHop = 6$.

The algorithm was tested for 5 different cluster densities ranging from 100-10000 clusters. The maximum number of channels required in all the different densities is recorded in Figure 3-16. Each density was tested with 100 simulations. Figure 3-16 shows that the maximum number of channels reaches up to 14 channels as the density grows.

Figure 3-16. Maximum channels vs. number of clusters

Figure 3-17. Distribution of channel assignment in a 10000 clusters setup

Figure 3-17 shows the average distribution of channel usage in a 10000-clusters setup. The points plotted are an average of 100 simulations. The last two channels, channels 13 and 14 are rarely assigned in the 10000-clusters experiment and these two channels have only 47 and 2 assignments, respectively.

3.10 Related Work

In general there has been extensive research work on channel assignment in cellular networks [47,62] and multi-hop packet radio networks [8,34,58,76]. Katzela, *et al.* [47] have provided a comprehensive survey on work related to dynamic, fixed and hybrid channel assignment schemes for cellular networks. The work by Robert *et al.* [62] provides another comprehensive study on frequency assignment problem in cellular networks. The channel assignment problem in cellular network poses a different but easier problem compared to that for a wireless ad hoc network due to several reasons. In a cellular network, there is a notion of high-powered fixed base station, which has the necessary processing power to process the channel assignment algorithm. The base stations are also normally connected by fixed line systems therefore allowing channel information to be communicated fast. Communication is in a star like fashion with reference to the base station. Whereas in a wireless ad hoc network, the nodes are usually identical and they are usually mobile. Communication can be arbitrary in any direction. There has been no prior work on using channel assignment to address exposed terminal problems in wireless ad hoc networks.

Makansi [58] introduced a Transmitter-Oriented Code Assignment (TOCA) scheme for multi-hop packet radio. In TOCA scheme, receivers are dynamically tunable to receive the waveform of differing codes from neighbors. The scheme is difficult if the degree of connectivity is high since receivers have to scan for all transmitter codes. The main

idea is to reduce the interference caused by hidden terminal in a network using random access protocol by proposing that codes two hops away differ. The work was based on a TDMA channel but can be used for a code or frequency based wireless system. The number of codes, N_c required for a graph with maximum connectivity Δ is given by $N_c \geq \Delta$. A simple non-ideal algorithm for assigning codes was proposed and simulation shows a $2 \times \Delta$ bound. The work by Bertossi, *et al.* [8] was an extension of [58]. Bertossi, *et al.* [8] shows that finding the minimal assignment of codes for suppression of hidden terminal is NP-complete and thus computationally intractable. The work also introduces a fast but sub-optimal centralized and distributed heuristic algorithms based on a naive greedy graph coloring heuristics approach. In the process of assigning codes, the algorithm uses two-hop information and does node ordering based on node ID. Hu [34] introduces the concept of pairwise code-assignment (POCA) scheme where codes are assigned to edges of the connectivity graph in a multi-hop network. The idea of code assignment is to increase spatial reuse by reducing collision due to co-channel interferences. Simulation based on some sparse topologies has shown that the number of codes required is less than that of the TOCA scheme while maintaining similar throughput. Hu [34] also introduces a two-phase algorithm for assignment and reassignment of codes to transmitter and receivers when there are clashes in code assignment at certain boundaries of the network. Garcia *et al.* [76] presents a distributed code assignment algorithm for a dynamic, multi-hop wireless network that can be applied to transmitter oriented (TOCA) and receiver oriented code assignment (ROCA) systems. The algorithm assigns a code to each node in a way that no interference occurs after the algorithm converges, provided that the number of codes available for assignment is at least $\Delta(\Delta+1) + 2$, where Δ is the maximum number of one

hop neighbors any node can have. The assignment of code at a node is similar to [8,58] where code assigned is to be different from that used two hops away.

3.11 Conclusion

In this chapter, we identified a new type of exposed node problem, known as the critically exposed node problem, which causes severe throughput degradation in wireless ad hoc networks as we have shown through simulation. We also proposed two possible methods; via scheduling and channel assignment to solve this problem with further elaboration of the later scheme. We also analyzed the proper method of channel assignment and worked out the minimal channel assignment required to eliminate the critically exposed node problem. We suggested the use of certain clustering approach to achieve the minimal assignment. Under the conditions discussed in Section 3.6 , and assuming the existence of a centralized algorithm and suitable MAC protocol such as PCAM, 8 non-overlapping channels would be required. However, under a non-ideal distributed heuristics-based algorithm, we will require an additional 6 channels.

Chapter 4. SCENT – A MAC Protocol for Concurrent Transmission in Wireless Ad-hoc Networks

In this chapter we propose a new distributed CSMA/CA-based MAC scheme, known as SCENT, which uses a unique synchronization technique to increase the per-node capacity in wireless ad hoc networks. Unlike power control schemes, we assume that the transmission power and hence maximum transmission range of a node is a fixed pre-determined parameter set by the hardware vendor. The protocol exploits the use of location information, power capture and a modified message sequence. To our best knowledge, the scheme proposed by Acharya *et al.* [2,3], known as MACA-P is the only other concurrent synchronization technique proposed so far, but this scheme differs considerably from SCENT. While MACA-P tries to synchronize transmissions in an area encompassing 2-hop neighbors, in SCENT we only synchronize transmissions in the transmission range of a main sender. While the state of 2-hop neighborhood nodes are less deterministic and difficult to synchronize, in SCENT we have better control over the single-hop transmissions.

4.1 Preliminaries

In a wireless network, the shared medium is subjected to contention from many surrounding nodes. The performance of a wireless system is greatly influenced by the MAC scheme and the modulation technique used. We discuss some of the conditions

that have to be satisfied when reusing the wireless channel for simultaneous transmission and a couple of techniques that are useful for our protocol.

4.1.1 Packet capture and conditions for parallel transmission

In a wireless multiple access environment, capture has been shown to improve the throughput of random access protocols such as slotted ALOHA [107]. Exploiting power capture [4] in wireless hardware allows successful reception of a packet in the presence of, typically weaker, contending transmissions. Current hardware allows capture to occur when the stronger packet arrives almost simultaneously, regardless of arrival time but within a stipulated interval T_c known as capture time.

Figure 4-1. Condition for parallel transmissions

We adopt the physical model used by Gupta and Kumar [28] to describe the necessary conditions for parallel transmissions in this section. We use the following notations to describe the protocol. Let Rx_Thresh and CS_Thresh denote the minimum receiving power necessary for successful reception and sensing range power threshold to indicate an idle channel, respectively. Let $Capture_Thresh$ denote the minimum power ratio required by the receiver to “latch on” to the stronger signal and successfully receive the packet. The conditions for power capture can be generalized as follows. Let $i_k \in V$ and $j_k \in W$ be the k^{th} sender and receiver pair, and P_{ij} the power received at node j from the

transmission of node i . Then for successful reception at any receiver j_k , the following condition must be satisfied [28]:

$$\forall k : i_k \in V, j_k \in W; \frac{P_{i_k j_k}}{\sum_{n \neq k} P_{i_n j_k}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-1})$$

If a stronger signal packet arrives after the interval T_c , the hardware will not resynchronize with this new packet even if it exceeds the capture threshold. This implies that, if parallel transmissions are to be admitted within the transmission range of a node, the packet transmission has to be synchronized within interval T_c .

We use Figure 4-1a to explain the converse case. Tx_A , Rx_B and P_{AB} denote the transmitter A, receiver B and the signal power from A to B respectively. Assuming that the transmission is not synchronized and Tx_A transmits to Rx_B first. Rx_D would pick up Tx_A 's packet after locking on to its packet training sequence although it is not meant for Rx_D . If Tx_C starts transmitting after Rx_D has locked on, the signal received at Rx_D will be corrupted since the newly arriving signal has higher power than the initial signal from node Tx_A . In SCENT, we prevent this phenomenon from occurring. In a CSMA/CA scheme there is also a reversal of transmit-receive role for transceivers during the RTS/CTS/DATA/ACK messaging. We notice that there is also a location dependent packet capture requirement when this occurs. We use Figure 4-1a to illustrate the additional conditions that have to be satisfied. For power capture to successfully occur at both receivers D and B, assuming the signals are sent simultaneously (within interval T_c), the following two conditions should be satisfied.

$$\frac{P_{CD}}{P_{AD}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-2})$$

$$\frac{P_{AB}}{P_{CB}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-3})$$

When the transmissions are reversed for both pairs, A-B and C-D, we also require the following two conditions to be satisfied.

$$\frac{P_{DC}}{P_{BC}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-4})$$

$$\frac{P_{BA}}{P_{DA}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-5})$$

We term conditions (Eq. 4-2) to (Eq. 4-5) collectively as “*S-R capture power reciprocity*” or sender-recipient capture power reciprocity. In SCENT, since we want to synchronize parallel transmissions within the transmission range of node A (Figure 4-1a), we are interested in studying the combined power generated by transmitters within this region. Let $i_k \in V$ and $j_k \in W$ be the k^{th} sender and receiver pair within the transmission range of node A, $l \in Z$ be the possible transmitter outside node A’s carrier sense range and P_{ij} and P_{lj} are the power received at node j from the transmission of nodes i and l respectively. We want the following equation to hold for any receiver j :

$$\forall k : i_k \in V, j_k \in W \quad \text{and} \quad \forall l : l \in Z ; \quad \frac{P_{i_k j_k}}{\sum_{n \neq k} P_{i_n j_k} + \sum_l P_{l j_k}} \geq \text{Capture_Thresh} \quad (\text{Eq. 4-6})$$

At this point, it is also worth pointing out that the transmission pattern in Figure 4-1b is also possible as long as equation (Eq. 4-6) is satisfied. However, from simple geometry we can work out that the distance of secondary pair (D-C) promoted by such an arrangement is less favorable than the orientation shown in Figure 4-1a. In summary,

there are several conditions that have to be satisfied to ensure successful parallel transmission in a local neighborhood. These include:

- Equation (Eq. 4-6) has to be satisfied at the receiver for the entire duration of packet reception.
- Local parallel transmissions should occur simultaneously within the interval T_c while satisfying equation (Eq. 4-6).
- During the reversal of transmit/receive role in RTS-CTS-DATA-ACK sequence, the “S-R capture power reciprocity” has to be satisfied for all parallel pairs.

4.1.2 Location information

Location information has been proposed for use in routing protocol [50] and for enhancements of spatial reuse of CSMA based MAC protocols [61,95]. In [61,95], the received signal strength (RSS) is used to estimate the range of the neighboring nodes. In SCENT, we assume that location information is also obtainable from the wireless system. There are several ways in which location can be obtained in a wireless network. The Global Positioning System (GPS) has been suggested as a location information device in wireless ad-hoc networks [50] but is suitable mainly for outdoor use. Recently there have been advances in using Ultra-Wide-Band (UWB) techniques to obtain accurate location information [53]. Patwari *et al.* [72] have shown that peer-to-peer relative location system can provide accurate location estimation using RSS as a ranging method when locations of a few nodes in the network are known. Their technique is based on maximum-likelihood method to estimate the location of other devices based on the few known locations.

In SCENT, we only require a node to know accurate relative positioning of nodes within a radius of twice its transmission range.

The SCENT protocol relies on location updates that are broadcast to nodes 2 hops away and will naturally incur some overhead since the same channel is used for data transmission as well as broadcasting the location update. The signaling overhead for location update in SCENT is similar to those encountered in ad hoc clustering protocols where some form of topology information (node ID, node degree) is also propagated several hops away periodically. In both cases, mobility affects the reliability of the data that was broadcast previously to neighboring nodes. In highly mobile scenarios, we do not recommend the use of SCENT protocol since frequent location updates would be required to maintain the topology information.

The frequency of location update in SCENT is a design parameter and should be set based on the mobility of nodes in the wireless ad hoc network. In a static or less mobile environment where the rate of transmission is much greater than the rate of change in location, the location updates can be less frequent. In SCENT, nodes can operate using either 802.11 4-way handshake scheme or the concurrent transmission scheme. In a highly mobile environment, nodes can choose to operate in the normal 802.11 mode.

4.1.3 Busy Tone

In [18], Deng and Haas pointed out inefficiencies of MAC schemes using RTS/CTS-based collision avoidance technique and showed that busy tone methods could improve channel utilization by a factor of 2. RTS/CTS based techniques basically suffer from hidden terminal and exposed terminal problems and suffer from high number of collisions when the traffic load is heavy. In SCENT, we make use of the busy tone

concept to control the interfering signals in a more deterministic manner. However, unlike [18,95], we only use a single busy tone to realize our protocol.

4.2 Related Work

In this section, we will focus only on the directly relevant scheme, i.e. the local concurrent synchronization scheme [2,3].

Acharya *et al.* [2,3] proposed MACA-P for concurrent transmissions in multi-hop wireless networks. The aim of the protocol is to improve the overall network utilization by exploiting spatial diversity to increase the number of concurrent transmissions in a wireless ad-hoc network. The scheme enhances 802.11 DCF by allowing parallel transmission for situations where two neighboring nodes are either both receivers or both transmitters but not a combination of receiver and transmitter. MACA-P divides the RTS-CTS-DATA-ACK message exchange into two distinct schedules, T_{data} and T_{ack} . These schedules, which are carried in the RTS and CTS messages, essentially inform the neighbors when the reversal of role from transmit to receive and vice-versa will occur. The protocol introduces the concept of a “master pair” or “primary pair”, who invites other “slave” or “secondary” pairs to transmit simultaneously. It utilizes a control phase interval to allow multiple sender-receiver pairs to contend and synchronize their data transfer with the master pair. Each node also maintains the state of neighbors by overhearing the RTS/CTS exchange and forming Network Allocation Vector (NAV) entries consisting of the neighbor’s MAC address and neighbor’s state (sender, recipient, idle). Based on the entries, a node can deduce whether it can communicate to a certain node.

One main problem in MACA-P is that, a sender-recipient pair can schedule a data transmission only if there is at most one master transmission in the sender's neighborhood or at most one master reception in the recipient's neighborhood, but not both. This degrades the MACA-P performance since it is likely that in a wireless ad-hoc network, sources 2-hops away might setup a master transmission. This causes the nodes in the 1-hop neighborhood to frequently receive 2 master transmissions. Simulation results in [3] confirm this anomaly in a grid setup when the aggregate throughput gain over 802.11 is marginally higher. The SCENT protocol will not suffer from same problem because the concurrent transmissions are confined to a 1-hop range of a primary sender. Due to the transmission and carrier sense ranges, it is unlikely (barring circumstances due to abnormal physical channel conditions) for a secondary pair to pick up invitations from two primary senders.

Another major drawback of MACA-P design is the use of a large and fixed control phase interval, typically around 512-640 bytes or $\sim 16 \times \text{RTS}$ interval, to admit other parallel transmissions. The long interval introduces additional delay and might degrade the capacity when there are no transmissions in the neighborhood to utilize the opportunity for concurrent transmissions. Another problem that arises due to the large control phase is the possible initiation of other master schedules in the carrier sense range of the initial master pair during the silent period of the control phase. This eventually leads to the reception of more than one master transmission by potential pairs in the 1-hop neighborhood. MACA-P introduces preferential triggering where backoff intervals are halved for secondary pairs in the hope that this would help cascade several transmissions but this does not offer much improvement in throughput either. In SCENT, the use of busy tone prevents nodes in the carrier sense range of a primary sender from initiating a new primary transmission. The problems of large

control interval overhead and multiple neighboring masters encountered in MACA-P are naturally avoided in SCENT.

Another key problem that was overlooked by the MACA-P design is the position of the secondary sender in relation to the primary receiver. We use the Figure 4-2 to illustrate this problem. Let node B be the recipient of the primary pair. Node C which is a potential secondary sender is just outside B's transmission range, will not receive the NAV from B. Based on MACA-P design, C will not have an entry in its NAV table indicating any receivers nearby, therefore it proceeds to transmit. However, based on equation (Eq. 4-6), the short distance between A and C causes capture threshold to be violated at node B and D.

Figure 4-2. Node placement highlighting the problem in MACA-P

Simulation results in [3] using concentric ring topology with outer senders also demonstrate the effect of collisions due to CTS and RTS at hidden terminals rendering the performance of the protocol worse than 802.11. To prevent this problem from occurring, the authors suggest the use of a special enhancement to the receiver such that it could capture a stronger signal at any point in the packet reception interval even when the receiver had already “locked on” to another packet. This could potentially be a problem in the hardware design since any temporary high power noise could trigger

the hardware to resynchronize and corrupt a normal reception. However, even if this enhancement works, it still does not solve the problem highlighted in Figure 4-2.

4.3 The SCENT Protocol

The main idea in SCENT is to enable parallel transmissions in the neighborhood of a sender-recipient pair (A-B) as shown in Figure 4-3. We want to enable parallel transmissions for pairs such as C-D and E-F if they satisfy the power conditions mentioned in equation (Eq. 4-6). To achieve this goal, we make use of location information and synchronized transmissions. In SCENT, data and ACK frames are sent out synchronously for pairs A-B, E-F and C-D to satisfy the power capture principles. In our design, we assume that the capture time interval is large enough to compensate for inaccuracy in clock synchronization between transmitters and maximum propagation delay between transmitter and receiver. Using a 1Mbps 802.11 system as an example, typical values for capture time and maximum propagation delay are $4\mu s$ [2] and $2\mu s$, respectively. We assume location information can be obtained using techniques mentioned in Section 4.1.2 . The location information is propagated to 2-hop neighborhood nodes using broadcasting techniques.

Figure 4-3. Parallel transmissions within transmission range of node A

In SCENT, we utilize a single narrow bandwidth busy tone channel and a common data channel used for sending control and data messages. The setup is similar to BTMA protocol [88], whereby only a single busy tone from the primary sender (node A in Figure 4-3) is used. We also assume that the channel characteristics (path loss, shadowing, multi-path, etc.) experienced in both channels are the same. Therefore the area covered by the data signal and the busy tone is the same. Similar to [3], we introduce the concept of a “primary pair” whose transmission is used to invite other pairs to transmit concurrently. In Figure 4-3, the pair A-B is known as the primary pair. In SCENT, the primary pair uses the same binary exponential backoff scheme used in IEEE 802.11 DCF. Since nodes maintain location information, it can decide to admit other parallel transmissions based on this knowledge. As shown in Section 4.4 later, if typical 1 Mbps 802.11 system parameters are used, a primary pair transmission length of less than 157.75m is required to admit other neighboring pairs. In our implementation, we set this threshold to 140 m. Also due to additional delay incurred in sending out extra control signals, primary pairs will only invite other pairs if its data is sufficiently large.

4.3.1 Message sequence

In SCENT, the primary pair uses a modified version of the RTS-CTS-DATA-ACK message sequence, which is typically used in IEEE 802.11 DCF scheme. We instead use the RTS-CTS-DS-DATA-ACK sequence proposed in MACAW [9]. In addition to notifying the successful reception of CTS packet to exposed nodes, in SCENT, DS serves as a token to invite nodes in the transmission range of the primary sender to contend for parallel transmission. The information carried in DS includes the identification of the primary sender, primary recipient and the duration of the DATA

packet. Based on IEEE 802.11 standard, the length of DS is typically designed as 44 bytes including physical layer header. We utilize a second contention window (S_CW) shown in Figure 4-4 to invite a “secondary pair” to transmit concurrently with the primary pair. Unlike the large and fix control phase used in [3], we design a small secondary contention window to allow only a single pair to successfully contend and send an RTS in this window. In SCENT, the secondary window consists of 8 contention slots plus an interval equivalent to an RTS packet. This amounts to 64 bytes in a 1Mbps system. If a primary pair chooses not to invite other pairs, it will not use the DS and the secondary window and reverts to the RTS-CTS-DATA-ACK sequence. In all cases, the secondary pair always uses the RTS-CTS-DATA-ACK message sequence. Later in Section 4.4 , we will show that by keeping the distance of the parallel pairs reasonably long enough, we are able to admit another four (4) pairs in addition to the primary pair. Besides a single secondary pair, other pairs known as the “tertiary pairs” can be admitted. Unlike other message sequence, the tertiary pairs will use a DATA-ACK message sequence. In SCENT, we will limit the maximum number of tertiary pairs to 2.

Based on the information carried in DS, the secondary sender will evaluate its position and its intended receiver’s position with respect to the primary pair to see if it satisfies condition (Eq. 4-6). The data size of the secondary pair plays an important part in the decision to contend during the secondary window. The data size with the closest value to the primary pair’s data size is favored else the utilization is not maximized. The secondary sender will also base its decision to contend on the range of transmission. In our design, longer-range transmissions are favored over shorter-range transmissions. Admitting shorter-range transmissions is also possible in our design. However, we design it in this manner to reduce the number of hops in a multi-hop transmission.

The message sequence starts when the primary sender sends out a busy tone. The range covered by the busy tone is equivalent to the carrier sense distance of the data channel. In SCENT, carrier sensing is also performed by the primary sender in the busy tone channel. While primary nodes can choose not to enable parallel transmission, the busy tone is still sent out to ensure compatibility with neighboring transmissions. Figure 4-4 is used to illustrate the message sequence for the pairs shown in Figure 4-3.

Figure 4-4. SCENT protocol message sequence for Primary, Secondary and Tertiary pairs

Node A sends out a busy tone, BT_A, in the busy tone channel as well as an RTS in the data channel when it senses no busy tone in the busy tone channel. Node B replies with a CTS and in return, node A sends out a DS message inviting other nodes to contend in the secondary window. While there is a precondition for carrier sensing for primary sender before transmitting an RTS, the primary recipient, secondary pair and tertiary pair need not perform carrier sensing on the busy tone channel. Instead, we use 2 timers to determine the channel states correctly. In Figure 4-4, node C succeeds in contending in the secondary window. It manages to send an RTS to node D, which replies with a CTS. Since the secondary contention window is a fixed interval, the data transmission in primary and secondary can be synchronized. While the nodes are contending in the

secondary window, primary nodes in the carrier sense range of node A will not start a new transaction. This is due to the constant busy tone signal that prohibits primary nodes in this region to start transmitting. Nodes E and F are the tertiary pair in the above scenario. The DATA frame in nodes A, C and F are sent out simultaneously. Notice that the secondary DATA frame and tertiary DATA frame might be smaller than the primary DATA frame. Finally, the ACK frames are sent out at the same time even if the DATA frames are of different lengths.

4.3.2 Finite state machine

Figure 4-5. FSM for primary sender

Figure 4-6. FSM for primary recipient and secondary pair

Figure 4-5 and Figure 4-6 depict the finite state machine (FSM) of the SCENT protocol. In the state transitions shown above, the statements in “[]” are the condition whereas the statements in “{ }” are the actions. Figure 4-5 illustrates the state transition of the primary sender. A node that has no packets to send stays in the IDLE state. When it has packets to send it will transition to the PRIM_CONTENTEND state and set its backoff timer. When the backoff timer expires, the primary sender performs a carrier sense in the busy tone channel. If it senses a clear channel, it will proceed to transmit RTS and a busy tone. It then waits for the CTS frame in the PRIM_R_CTS state. If the

CTS does not arrive or the frame is corrupted, the primary sender will cancel its busy tone and return to IDLE state. If it receives the CTS, it will transition to PRIM_S_DS and send out a DS frame. It will then send data in PRIM_S_DATA state and wait for ACK in PRIM_R_ACK state. Once the ACK is received or the ACK timer expires, the primary sender will return to IDLE state and switch off its busy tone.

Figure 4-6 illustrates the full state transition of the primary recipient and the secondary sender. A primary recipient or a secondary node that does not receive any busy tone remains in the IDLE state. If a busy tone is detected, the nodes start a “receive-RTS” timer and transition to the PRIM_WF_RTS state to wait for a primary RTS. The receive-RTS timer defines a starting and ending time delimiter, which defines when the reception of the primary RTS is expected in the presence of differences in clock synchronization time between nodes and channel propagation delay. If a primary RTS intended for this node is received within this interval, it is then marked as the primary recipient and the node transitions to PRIM_S_CTS state to reply a CTS. If it does not receive a primary RTS intended for this node or receive the RTS after the receive-RTS timer expires, it will mark itself as a potential secondary pair. The node will then start a new timer known as “DS-timer”. Similar to receive-RTS timer, the DS-timer also defines a starting and ending delimiter to handle the reception of a DS frame. If a DS frame is received in this interval, it will then transition to the SEC_USER state. In SEC_USER state, the secondary sender will then transition to the SEC_CONTEND STATE to contend for the parallel transmission. If the potential secondary nodes do not receive a DS in the interval defined by DS-timer, it will transition to the BLOCK state. If the secondary sender is not successful in the secondary contention, it will also transition to the BLOCK state. If the secondary sender’s backoff counter reaches zero and the channel is idle, it will send out an RTS frame. It will then wait for a secondary

CTS frame. If the frame arrives correctly, it will then proceed to send the DATA frame. If the RTS frame is not received or corrupted, it will transition to the BLOCK state. Once the DATA frame is sent, the secondary sender waits for the ACK frame. If the ACK frame is received correctly, it transitions to the IDLE state. For the primary recipient, after the PRIM_S_CTS state, the node will transition to PRIM_R_DATA state to wait for the DATA frame. If the DATA frame does not arrive or a timeout occurs, the node will transition to a BLOCK state if the busy tone is still active. If the primary sender turns off the busy tone while the node is in PRIM_R_DATA or BLOCK state, the node will transition to IDLE state.

Figure 4-7. Preventing collisions from overlapping regions

The use of the receive-RTS timer prevents hidden terminal problems associated with the primary pairs, whereas the combination of busy tone and DS-timer prevents collisions when nodes in overlapping regions receive DS token from different directions. Figure 4-7 illustrates this problem. If nodes F and G receive a DS token from node E while nodes D and C receive a DS token from node A, they might proceed to contend for the secondary transmission. Due to their location and power capture condition described in (Eq. 4-6), collision could happen for pair F-G and C-D. To prevent this, we use the DS-timer to ensure that only one of the pairs in the overlapping region succeeds in capturing the opportunity to transmit. In the example shown in Figure 4-7, If busy tones initiated by nodes A and E are sent out at different time slots,

nodes in the overlapping region will expect DS frames according to the first busy tone only. Therefore one of the pair will be disabled from using the DS token. Collisions could however arise if the busy tone is sent out simultaneously.

4.3.3 Design of Secondary contention window

There is possibility of collision in the secondary window given the small contention window size and possibility of large number of contending nodes in the transmission range. If collision happens, then the secondary and tertiary transmissions will also fail. To increase the chance of success, we use a different backoff scheme for the secondary window instead of the binary backoff scheme. While the binary backoff scheme has an equal probability distribution curve for all nodes, we use a truncated, increasing geometric distribution similar to that proposed by Jamieson *et al.* [42]. The probability distribution is given by the following expression

$$P_r = \frac{(1-\alpha)\alpha^{CW}}{1-\alpha^{CW}} \bullet \alpha^{-r} \quad \text{for } r = 1, \dots, CW. \quad (\text{Eq. 4-7})$$

where $0 < \alpha < 1$ is a distribution parameter, r is the slot position and CW is the maximum number of slots in the contention window. While the motivation in [42] is to ensure the success of only a single report in a sensor network given a fixed CW period, and possibly N number of unknown senders, where $N \gg CW$, the situation in SCENT is similar in the sense that only one successful transmission should occur given a variable and large number of contending nodes. However instead of using a fixed α as in [42], we vary α to differentiate the pairs having different transmission distances and data packet sizes. Assuming Y is the maximum possible length of the secondary pair

given the location of primary pair and L is the length of the primary data, then we calculate α as

$$\alpha = \frac{Y_0 W_0}{Y} + \frac{L_0 W_1}{L} \quad (\text{Eq. 4-8})$$

where L_0 is the data length of the contending secondary node, Y_0 is the distance between the secondary sender and secondary recipient. W_0 and W_1 are constants used to control the preference of the two weights.

4.3.4 Enabling tertiary pairs

An alternative method for allowing concurrent pairs in the transmission range is to schedule different set of pairs that obey the condition set forth in Section 4.1.1 rather than using a secondary window. However, by using such a scheme, there is possibility that this fixed set of nodes might not have any data to send due to the time-varying density of flows. In addition to this, the scheduling of new and old nodes entering the system increases the complexity. For this reason, we use a secondary window to ensure that at least a single other pair gets to transmit in parallel. The contention scheme also promotes fairness among possible secondary transmission. To increase the capacity of the system, we propose a simple query response protocol that can be used to admit additional 2 tertiary pairs. A tertiary sender will listen to the channel for the identity of the primary pair and the secondary pair. Based on this combination, it will decide if it can transmit along with these two pairs. It will then send a broadcast query message (TERT_REQ) to primary sender to request for parallel transmission. The primary sender will then decide if the tertiary pair can be admitted. Since there is a possibility of receiving multiple queries from potential tertiary pairs, the primary sender will schedule a primary-secondary-tertiary combination for a certain interval and send out a

broadcast response (TERT_RES). Once this interval expires, tertiary senders will have to re-request for new primary–secondary–tertiary schedules.

4.3.5 Deterministic vs. non-deterministic scheduling

To schedule the concurrent transmissions, a non-deterministic scheduling approach (contention based) as opposed to a deterministic scheduling approach was chosen for reason of protocol simplicity and reduced overheads. The scheduling of concurrent transmissions shares the same complexity of scheduling slots in TDMA-based wireless ad hoc networks. In the deterministic approach, timetables or schedules for individual nodes or links have to be setup prior to transmission. Distributed scheduling in wireless ad hoc networks can be done in a topology-dependent [17] manner or a topology-transparent [15,16] manner. In a topology-dependent scheme, nodes have to exchange schedules dynamically and resolve time slot requests. The problem of optimally scheduling the transmissions is proven to be an NP-hard problem in graph theory [19]. Due to this, suboptimal solutions using randomized approach or heuristics are used for simplifying the scheduling problem. While the topology-dependent scheme suffers from higher communication overhead, the topology-transparent scheme reduces the overhead incurred by the conflict-free schedule maintenance. However, there are two problems associated with this scheme. Firstly, this scheme only works in a time division multiplex system, where all nodes are time synchronized. Secondly, collisions are still possible since nodes do not share their slot allocation function and will continue to contend in their allocated slot until they successfully transmits [83].

The SCENT protocol is more suited for wireless ad hoc networks, which are static or with slow mobility. Broadcasting of location update is assumed to be done less

frequently than those required for topology-dependent distributed scheduling. The topology-transparent scheduling is not suitable due to the reasons mentioned above.

4.4 Determination of maximum number of communication pairs in the transmission range

In multi-hop transmissions, longer-range transmissions are necessary to ensure lower delay performances. Therefore, it is essential that we maximize the range while maximizing the throughput of the system. In this section, we determine the maximum number of pairs that could possibly transmit in parallel, given the condition that we also want to maximize the range of the pairs. We also want to find the minimum required distance of the primary pair necessary for invitation of other concurrent pairs. This problem is actually a constraint satisfaction problem but we use simple geometrical method to derive the maximum number of pairs by optimally placing the nodes in the transmission radius. We assume pairs are equidistant to simplify the analysis.

We denote the maximum transmission distance by Rx_Dist and carrier sensing distance by CS_Dist . Note that Rx_Dist is the maximum possible transmission distance where the receiver would receive the message successfully. Let Rx_Thresh be the received power at distance Rx_Dist . The channel propagation model in a typical wireless system can be modeled as follows. In the absence of any reflection or multi-path, radio wave propagation can be modeled using the free space propagation model. Using this model, the received signal power in free space at distance d from the transmitter is given by the following equation [78].

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} = P_1(d) \quad (\text{Eq. 4-9})$$

P_t denotes the transmitted signal power, G_t is the antenna gain of the transmitter and G_r is the antenna gain of the receiver. L is the system loss, and λ is the wavelength. While equation (Eq. 4-9) describes the path loss accurately for shorter distances, for a more realistic profile at longer distances, Rappaport [78] has shown that the two ray ground propagation, given by equation (Eq. 4-10), models the path loss more accurately. The cutoff distance d_c in which the two different equations are applied is known as the Fresnel breakpoint. Propagation starts out with a d^2 falloff rate and then transitions to a d^4 falloff rate at greater ranges.

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L} = P_2(d) \quad (\text{Eq. 4-10})$$

h_t and h_r are the heights of the transmitter and receiver antennas, respectively. Note that λ , G_r , G_t , h_r , h_t and L are constants in normal situations. Therefore, we have

$$P_r(d) = \begin{cases} P_1(d), & \text{if } d < d_c \\ P_2(d), & \text{if } d \geq d_c \end{cases} \quad (\text{Eq. 4-11})$$

We assume a disk area for transmission range to simplify the calculation. We assume capture threshold of 8dB in our calculation. Note that because of equation (Eq. 4-6), it is impossible to include other parallel transmission in the transmission region when the transmission distance for the primary pair is above a certain threshold, which we term as *Conccurency_ThreshDist*. We assume the power contributed by the term $\sum_l P_{ljk}$ in equation (Eq. 4-6) is negligible to simplify our analysis.

Figure 4-8. Primary pair and best case location of secondary senders and recipients

Figure 4-9. Best case location of the first secondary pair

**Figure 4-10. Best case location of the second secondary
recipient**

Figure 4-11. Best case location of the second secondary sender

Figure 4-12. Best case location of the third secondary pair

We now use Figure 4-8 to Figure 4-12 to determine the maximum number of equidistant parallel pairs that can possibly coexist by physically placing the nodes in an optimal way. In Figure 4-8, node S1, R1, S2 are primary sender, primary recipient and secondary sender, respectively. Nodes R2, R3, R4 are possible secondary recipients. We fix the position of node S1 to analyze the position of other nodes. Based on the *S-R capture power reciprocity*, we study the capture power received at S1 when node R1, R2, R3, etc. are transmitters. The distance of the secondary recipients R2, R3 and R4 governs the maximum distance for S1-R1.

We vary the distance of the primary recipient, R1 and possible secondary senders (R2, R3, R4) to find the optimal point such that

$$\frac{P_{S1-R1}}{P_{R2-S1} + P_{R3-S1} + P_{R4-S1}} = Capture_Thresh \tag{Eq. 4-12}$$

Maximum distance y for S1-R1, also denoted as *Conccurency_ThreshDist*, occurs when there is only a single secondary recipient at the periphery of circle with radius r centered at S1.

Number of interfering node (x)	Distance S1-R1 (m)
1	157.75
2	132.65
3	119.86
4	111.54
5	105.45
6	100.8
7	96.95
8	93.8

Table 4-1. Number of Interfering node vs. possible Primary sender-receiver distance

Before we calculate y , we list in Table 4-1 the possible distances of y given x number of interfering nodes in the periphery of this circle. We use typical values captured from an actual wireless transceiver such as the 914MHz Lucent WaveLAN DSSS radio interface [65]. Antenna gains, G_t and G_r are set to 1, while antenna height h_t and h_r are set to 1.5m. System loss $L=1$. P_t is set to around 0.2818 W, while Rx_Thresh is 3.652e-10 W [65]. This gives us a d_c of around 86.15m. The transmission range, Rx_Dist works out to be 250 m whereas the CS_Dist is 550 m. Substituting these values into (Eq. 4-11) and (Eq. 4-6), we have the *Conccurency_ThreshDist* = 157.75m.

Based on Figure 4-9, and under the condition that recipients should be placed at the periphery of circle radius r centered at S1, and S2 has to be $\geq r$ away from R1, the optimal placement of S2 to enable the most number of parallel pairs is at the

intersection of the 2 circles with radius r centered at S1 and R1. Under this optimal placement, the distance between S2 and R2 should be at a distance y . Notice that, if node S2 is placed closer to R2, the distance between S1 and R1 has to be less than distance y as shown in Figure 4-9 due to the *S-R capture power reciprocity*. We would also have to keep distance between S2 and R2 not greater than y because this would also violate the *S-R capture power reciprocity*. We now work out the location of R3, which must be placed at distance r away from S1 and must also be at a distance $\geq r$ from S2. The optimal placement of R3 is on the intersection of circles with radius r centered at S2 and S1 (Figure 4-10). We now determine the placement of S3. The possible location of S3 can be somewhere in the shaded area shown in Figure 4-11, but must be at a distance y from R3 and distance $\geq r$ from R2.

The optimal placement of S3 is shown in Figure 4-11, where it is placed at a distance y from R3. Notice that if S3 is placed in other locations, then R4 shown in Figure 4-12 will have to be further away from R3, resulting in less optimal placement. Repeating this exercise, we find that the optimal placement for equal distance pairs has to be interleaved in a S-R-R-S-S-R fashion and placed on the periphery of a circle of radius r centered at S1, where S denotes a sender and R denotes the recipient. To calculate the maximum number of possible additional pairs, we calculate the angle θ defined by points A', S1 and B' (Figure 4-9), which is equal to $360^\circ - 2\cos^{-1}(y/2r)$. The largest angle possible for θ occurs when there is a single interfering node or when $y = 157.75\text{m}$. Notice that as $y \rightarrow 0$, $\theta \rightarrow 180^\circ$. Since $\angle \text{S2-S1-R3} = 60^\circ$, we can show that based on the upper and lower bounds of angle θ , there can be at most 5 additional equidistant pairs including the primary pair placed in the transmission range of the primary sender. Based on Table 4-1 and given that there are four (4) recipients on the

boundary, the distance y is equal to 111.54m. Notice that this value is the upper bound on the number of equidistant parallel transmission pairs based on the maximum transmission range constraint. If all the pairs in the neighborhood have short ranges, we can admit many more parallel transmission pairs subjected to capture power stated in equation (Eq. 4-6). Since our objective is to admit as many long-range transmissions as possible, we have obtained the value that corresponds to the maximum range constraint. While the ideal positions shown in Figure 4-12 are not practical in real systems, we set a lower minimum threshold of 140 m to admit other pairs and allow only a maximum of 4 concurrent transmissions, inclusive of the primary pair to occur within the transmission range.

4.5 Performance Evaluation

In this section, we investigate the performance of SCENT compared to the IEEE 802.11 protocol [36]. The SCENT protocol was implemented in ns-2 [65] with CMU wireless extensions and is compared with 802.11 MAC on the same simulation platform.

4.5.1 Simulation Model

For the simulation, we used 1Mbps for channel bit rate. The transmission range, carrier sensing range and capture threshold are set to 250 m, 550 m and 8dB, respectively. The tertiary broadcast request (TERT_REQ) and response message (TERT_RES) are implemented as 48 and 96 bytes messages, respectively. The schedule for a single primary-secondary-tertiary combination has a timeout interval of 1 second. The DS message has the same length as an RTS message. For the first simulation topology, we placed 30 nodes (15 pairs) randomly in a 388×388 m² flat area, such that all nodes are

at least within the carrier sense range of each other. The purpose of this setup is to study the effect of nodes within the sensing range only and performance of nodes that are placed randomly. The 15 pairs of nodes have fixed transmission ranges but random direction. The distances between the pairs are varied from 20 m to 150 m for different scenarios. The packet size was fixed at 1024 bytes and the packet arrival rate was varied from 1 to 32 packets/sec. We ran these simulations with average 20 runs with each run lasting 30 seconds. The aggregate throughputs for both protocols were recorded. In the second scenario, we placed 8 different pairs of 4 different ranges (20 m, 80 m, 140 m, 200 m) in circular area with radius 250 m. The area used in this simulation is almost same as the first scenario but in this setup, the pairs are purposely arranged in a special manner. The pairs are placed in a manner that would allow all pairs (except for the 200 m pair) of equal range to transmit in parallel. The throughput performance for both SCENT and 802.11 were collected. In the third scenario, we placed 73 pairs (146 nodes), each with transmission distance of 80 m and random direction within a $1040 \times 1040 \text{ m}^2$ flat area to simulate a multi-hop scenario. The packet size was fixed at 1024 bytes and the packet arrival rate was varied from 1 to 32 packets/sec. Simulation was conducted with average 20 runs, with each run lasting 20 seconds. In our performance evaluation, we omit the study of the SCENT protocol in a mobile environment because we strongly believe that this protocol is not suitable for such an environment. In a mobile environment, constant movement of the nodes affects the node location. The process of triangulation can also be more complex due to imperfect channel propagation and Doppler Effect [41]. The incorrect location information would necessitate more frequent location updates, which leads to more control overheads and throughput degradation.

4.6 Simulation results

Figure 4-13. Aggregate throughput for long-range pairs

In the first simulation, we compared the performance of SCENT to 802.11 for long-range transmissions (150 m). Since DS message is only sent when the average distance between the primary sender and primary receiver is less than 140 m, in this setup, the SCENT protocol uses the same message exchange as 802.11. The performance of SCENT shown in Figure 4-13 shows a slight improvement (~8%) over 802.11 even when there is no parallel transmission. The reason for the improvement is due to the effectiveness of the busy tone used to prevent the hidden terminal problem, which is still possible in this setup. In 802.11, nodes in the carrier sense range can possibly transmit during the RTS/CTS phase as pointed out in [44] causing collisions in the main pair.

Figure 4-14. Aggregate throughput for medium-range pairs

Figure 4-15. Aggregate throughput for short-range pairs

In Figure 4-14, the same experiment is repeated for pairs with 80 m range. The results show improvement in throughput for SCENT by a factor of 2 compared to 802.11. In Figure 4-15, the range of the pairs was reduced to 20 m. The graph shows an increase of throughput by a factor of ~2.17. When we analyzed the trace file, we noticed that most of the transmissions in the 80 m scenario were primary and secondary transmissions. There were very few tertiary transmissions. In the 20 m case, more tertiary transmissions were admitted when compared to the 80 m scenario. This

explains why the throughput for the 20 m scenario, shown in Figure 4-15 is slightly higher than the 80 m scenario.

Figure 4-16. Simulation setup with 8 pairs of links (20m, 80m, 140m and 200m)

Figure 4-17. Throughput of various transmission ranges for 802.11 (20m, 80m, 140m and 200m)

Figure 4-18. Throughput of various transmission ranges for 802.11 (20m, 80m, 140m and 180m)

Figure 4-19. Throughput contributed by Primary contention in SCENT

In the second setup, our objective was to study the throughput fairness of the SCENT protocol in the presence of transmissions with different ranges. Figure 4-16 shows the simulation setup with location of the nodes and their transmission directions clearly indicated. Figure 4-17 shows the throughput achieved for different transmission ranges

when the 802.11 protocol and the setup in Figure 4-16 was used. The results (Figures. 4-16 to 4-18) show that under lower packet arrival rates, where collision is minimal, the throughputs for all ranges are equal. Notice that a significant drop in throughput is observed starting from a link length of 160 m and becomes more pronounced as the link length increases to 200 m. However, this is expected because, when there are two contending links of different lengths, the longer link will always suffer from Signal-to-Noise Ratio (SNR) conditions. Power capture is likely to succeed for the shorter links because the relative distance of an interfering source during a contention is greater. Because the longer links fail more often, they also go into more frequent backoff when compared to the shorter links. This causes further delay in accessing the channel. However, besides the link distance, the relative locations of the sender and receiver with respect to other flows also affects the throughput.

Using the setup as shown in Figure 4-16, only primary and secondary transmissions are possible for the SCENT protocol. The tertiary transmissions are not possible due to the placement of nodes.

Figure 4-20. Throughput contributed by Secondary contention in SCENT

Figure 4-21. Total throughput using SCENT

We measure the throughput contributed by the primary contention and the secondary contention in Figure 4-19 and Figure 4-20, respectively. For the 200 m pairs, the primary contention is the only component that contributes to the throughput since no parallel transmissions are allowed. Since the primary contention makes use of binary exponential backoff algorithm, the throughput performance of SCENT and 802.11 appears to be the same for high packet arrival rates. Under this circumstance, we state that the SCENT protocol has the same fairness as 802.11. Under low arrival rates, where the channel is not saturated, Figure 4-19 shows higher throughput for the 200 m pairs when compared to the other ranges. However, the throughputs appear to be lower for the shorter-range transmissions because the packets are also sent out using the secondary transmissions. The cumulative throughput as shown in Figure 4-21 confirms this case for the lower arrival rates. In SCENT, we have maintained the same fairness as 802.11 when primary contention is concerned. However, we have used the secondary contention to admit more parallel transmissions to increase the overall throughput of the system. Although it is possible to increase the throughput of the

longer-range transmissions by manipulating the binary exponential backoff scheme, this will reduce the chance of admitting more parallel pairs and hence lower the overall throughput.

Figure 4-22. Aggregate Throughput of SCENT and 802.11 in a multihop scenario.

In the third setup, our objective was to study the aggregate throughput of the SCENT protocol in a multi-hop scenario. Figure 4-22 shows the aggregate throughput performance of both SCENT and 802.11 protocols. When packet arrival rate is high, the performance improvement of the SCENT protocol over 802.11 is by a factor of 1.6. This value is lower than the factor of 2 observed in the scenario where nodes were collocated in the carrier sense range, which is shown in Figure 4-14. The main reason for this is the efficiency of primary pairs engaging the secondary pairs in the random layout. When the trace file is carefully studied, we can observe that not all primary transmissions are able to engage secondary pairs. Very few tertiary pairs were scheduled in this scenario. Nevertheless, this scenario shows that SCENT can still promote concurrent transmission in multi-hop scenarios and achieve reasonable throughput improvements.

4.7 Conclusions

There have been a few attempts to design MAC protocols that could increase the spatial reuse in the presence of shorter-range transmissions. Most of these approaches are based on power control and allow asynchronous transmission. In this chapter, we have pointed out some of the necessary conditions that have to be satisfied for enabling parallel transmissions within the transmission range of a sender. We have also proposed a new distributed MAC protocol scheme that allows parallel transmissions in a wireless network. By working within the constraints of power capture effect and using location information, we have shown that by synchronizing transmission in the transmission range of a primary sender, we can achieve more than twofold increase in throughput.

Chapter 5. An Efficient Multi-rate Relaying (EMR) MAC Protocol for Wireless Ad hoc Networks

Current wireless systems, such as 802.11b, 802.11a and HiperLAN2, utilize multiple transmission rates in order to accommodate a wide range of channel conditions. Higher rates can be used in the system when the signal-to-noise ratio (SNR) is sufficiently high. The use of multiple rates presents a significantly more complex challenge to ad hoc routing protocols than the traditional single-rate model. The hop count routing metric, which is traditionally used in single rate networks, tends to select short paths composed of maximum length links. In a multi-rate network, these long distance links operate at the slowest available rate, thus achieving low effective throughput and reduced reliability due to the low signal levels.

Intermediate nodes that are situated along the transmission path of a long distance link always remain silent during the ongoing transmission. This gives rise to the interesting possibility that these nodes can actually offer a higher throughput relaying service to the original long distance link by using higher transmission rates.

In this chapter, we will explore an enhanced CSMA/CA MAC protocol design known as Efficient Multi-rate Relaying (EMR) that relies on multi-rate capability and a relaying scheme to improve the throughput of wireless systems. Unlike other multi-rate aware routing schemes such as the one proposed by Awerbuch, *et al.* [5], our scheme

does not require any modification to the primary routing protocol. We use the traditional shortest path routing to form the main route and subsequently use a secondary local messaging technique to transform the overall route to a multi-rate route. We also propose a modification to the standard 4-way handshake of the CSMA/CA MAC protocol such that the contention and queuing is minimized when compared to the other schemes such as MTM [5] and the multi-rate routing protocol proposed by Seok, *et al.* [80]. Throughout this chapter, we use the IEEE 802.11b MAC protocol to illustrate the EMR MAC protocol enhancements although it is possible to implement the protocol by using other multi-rate systems.

5.1 Related Work

Past research efforts on multi-rate wireless protocols can be classified into two main areas. The first area is concerned with designing rate adaptation schemes such as ARF [74], OAR [32] and RBAR [79] to optimally adapt to the prevailing condition of the system, channel, etc. The second area of multi-rate research is concerned with improving routability and overall throughput in wireless ad hoc networks by capitalizing on the higher capacity routes [5,80]. The MAC protocol described in this chapter can be classified under the second category.

Awerbuch *et al.* [5] proposed an improved metric for route selection in a multi-rate wireless ad hoc network known as medium time metric (MTM), which can easily be used to replace the standard shortest path metric used widely in ad hoc routing protocols. The medium time metric is actually proportional to the time it takes to transmit a packet on a given link. As a result of using this metric, paths that have higher effective throughput instead of the minimum hop paths that have long but slow links are selected.

Seok, *et al.* [80] proposed a multi-rate aware sub-layer (MAS), which resides in between layer 2 and layer 3 of the OSI stack to improve resource utilization and minimize power consumption. The main idea behind this work is to route packets via intermediate nodes where higher data rates are available. The MAS keeps track of two-hop neighbourhood state information such as SNR and data rate of the adjacent links. This state information is either propagated proactively via HELLO messages sent by all nodes or reactively via listening to RTS-CTS-DATA-ACK message exchanges between potential relay and a receiver. With a view to minimize header overheads, MAS computes the effective path throughput and selects the relay path only if it provides improvement over the original path. The packet is passed from one hop to another using the same RTS-CTS-DATA-ACK sequence. The MAS layer also adds another 5-byte header field to help route the packets via the intermediate nodes. A time-to-live (TTL) and an IP address field are included in this header for routing purposes. This could cause some overhead but is insignificant if high data rate links are used. For the remaining of this chapter, we refer to the scheme proposed in [80] as MARO.

Figure 5-1. Path selection for various schemes. Path selection with Shortest Path (SPR) Algorithm = A ↔ C ↔ E. Path selection with MTM, MARO, and EMR = A ↔ B ↔ C ↔ D ↔ E

Figure 5-1 illustrates the path used by the various schemes discussed above. Shortest Path Algorithm selects the route A – C – E while MTM selects the route A – B – C – D – E. The MARO and our proposed scheme, EMR, uses the traditional shortest path algorithm to select route A – C – E, but uses a secondary routing process to change the route to A – B – C – D – E. One major drawback of the MARO and the MTM scheme

is the increased number of contenders in the system. The capacity of wireless ad hoc networks is known to be influenced by direct contention from nodes in the transmission range and transmissions from hidden nodes [55]. In these two schemes, the relaying nodes need to apply the standard RTS-CTS contention and binary exponential backoff process. As a result of this, more collisions will result in additional backoffs. Another significant drawback of the MARO and MTM scheme when compared to the shortest path scheme is the increase in queuing delay due to increase in hop count. Since packets are received and transmitted by using the same 4-way handshake, the relaying nodes now have to queue the packets for delivery to the next hop. Collisions and queuing delay in the MARO and MTM scheme could cause higher end-to-end delay when compared to the shortest path route.

Though the final path selected by our scheme (EMR) can resemble the path found by MARO or MTM, our scheme differs from MARO and MTM in a number of ways. First of all, EMR uses the shortest path method to select the basic route. Subsequently, an efficient local scheme is used for relay selection and forwarding thereby avoiding the problem of higher contention and larger queuing delay from which both MARO and MTM suffer. The contention in EMR is kept to about half of that introduced in the MTM scheme and we eliminate packet queuing at the intermediate nodes. The contention is halved for EMR because half of the nodes in the end-to-end link consist of relay nodes, which do not need to contend for the channel.

5.2 Preliminaries

Before discussing the efficient multi-rate relaying MAC scheme, we discuss some important issues that we need to consider when designing the multi-rate capable CSMA/CA-based MAC protocol.

5.2.1 SIR condition

Rate(Mbps)	Maximum transmission range (m)	Maximum throughput (Mbps)	δ
11.0	122	4.55	4.46
5.5	163	3.17	3.36
2.0	206	1.54	2.66
1.0	250	0.85	2.24
Carrier Sense	550	-	-

Table 5-1. Typical 802.11system Parameters

The original IEEE 802.11 protocol [36] was designed with a single base rate of typically 1 Mbps. With the multi-rate enhancement, transmission can take place at different rates depending on the channel quality. Higher data rates are achieved by using more efficient modulation schemes. However, with these more efficient schemes, the effective transmission range drops since the SIR requirements are higher. The ratio of carrier sensing range to transmission range, δ , that defines the guard zone varies depending on the modulation scheme used. Typically the ratio increases with the increase in transmission rate in a multi-rate MAC protocol. Typical 802.11b system parameters for different transmission rates are shown in Table 5-1 [5]. While the maximum transmission range drops for higher data rates, the carrier sense threshold remains the same for all transmission rates. This ensures a consistent guard zone is present for nodes receiving signals at different transmission rates.

5.2.2 Packet Rates and Overheads

In CSMA/CA protocols such as 802.11, the short RTS and CTS frames are used to reserve the channel and prevent collision due to hidden terminal problems. In 802.11b, even when the chosen transmission rate for data is higher, all RTS/CTS frames must be sent at the base rate, to ensure that all stations in the vicinity are able to receive the

messages error free. In addition to this, the physical layer header for all frames including DATA and ACK frames are also modulated using the base rate. This is to ensure that the receiver is able to lock-on to the incoming message and synchronize the packet reception. Due to these reasons, the overheads due to base-rate transmissions are the same irrespective of the transmission rate used.

5.2.3 Broadcast messages and routing

The use of broadcast messages is an integral part of the route discovery process in wireless ad hoc networks. Current 802.11-based protocols are designed to transmit broadcast messages using the base rate. The reason for using base rate is to ensure the reliability in packet reception by nodes in the vicinity. In a wireless ad hoc network, the use of base rate broadcast messages is advantageous when used with the shortest path algorithm since the path found is normally the least hop-count path. However, as pointed out, this path is not necessarily the highest capacity path when multi-rate links are available. The use of multiple rates presents a significantly more complex challenge to ad hoc routing protocols than the traditional single rate model. For example, the use of base rate broadcast messages with a new multi-rate aware metric such as MTM is not a trivial issue. Since broadcast messages are sent at base rate, the neighbouring node which receives the route request has to infer the highest data rate possible from the received signal strength of the base rate broadcast message and reply to the route request. Alternatively, one could send out broadcast messages at different rates but the time to discover the final solution to the optimal route will be much longer. In our design we keep the broadcast messages at the same base rate to ensure reliability but we use an innovative way to form the higher capacity multi-rate routes.

5.3 Efficient multi-rate relaying MAC (EMR)

The main idea in the EMR MAC design is the use of the multi-rate capability to forward packets at higher rates while keeping latency and queuing delay low. In the EMR scheme, the primary routing protocol such as AODV, DSDV, etc. can be used to form the main route from sender to destination. Naturally, this main route has the least hop count with mostly long-range transmissions. A secondary protocol that operates locally between two consecutive nodes along the main path is used to select the relays. These relays, when combined with the nodes that form the main route, eventually form a multi-rate route. The following sub-sections discuss the MAC protocol design.

5.3.1 Relay selection

Figure 5-2. Transmission Ranges and Regions

In EMR, we want to select the path that gives us the best effective throughput. There are two ways in which we can discover the relays. The first method is to estimate distances between relay and sender or receiver, by passively listening to the RTS and CTS frames transmitted at the base rate. The physical layer of the radio transceiver in 802.11 is capable of measuring the received signal strength, and based on equation (Eq.

3-1), the signal strength measurement can be used to make an estimation of the distance between any two nodes if the received packet is error free.

The second method is based on periodically broadcasting special messages meant for discovering relays. Although a system such as 802.11b transmits broadcast messages using the base rate, the transceiver should be capable of sending out broadcast messages at higher rates. These special messages are sent out by each node at different transmission rates. Nodes that receive these messages maintain a table indicating the nodes that they are able to listen to. However, this method is not favourable since the number of broadcast messages grows with the density of nodes. Moreover, frequent movement in wireless ad hoc networks could result in stale values and slow reaction to changes. For this reason, we select the signal strength measurement method for relay selection since the selection can be achieved with fewer packets and the ranging information can be obtained in a timelier manner.

Figure 5-3. Relaying Nodes

Figure 5-2 above shows the transmission range, carrier sense range and regions of a typical IEEE 802.11 system. We assume that nodes S1 and R1 can only communicate to each other using the base rate and this link is one of the hops discovered by the primary routing protocol. In Figure 5-2 above, nodes in region I are the only nodes that

receive both the RTS and CTS frames sent out by the sender S1 and the receiver R1, respectively. Nodes in regions II and III receive either the CTS or the RTS frames. To help relay the packets, we need to discover nodes that are situated in region I and with favourable distances from sender and receiver to provide the highest effective throughput. Consider the relay positions shown in Figure 5-3. Each of the relays contributes to different effective throughput when packets are sent via the relays.

Path	Link 1 capacity (Mbps)	Link 2 capacity (Mbps)	Effective Throughput (Mbps)	Priority
S1-R1	11	-	4.6885	10
S1-R1	5.5	-	3.2601	9
S1-F4-R1	11	11	2.9063	8
S1-F2-R1	11	5.5	2.2853	7
S1-F3-R1	5.5	11	2.2853	7
S1-F5-R1	5.5	5.5	1.8833	6
S1-R1	2	-	1.5777	5
S1-F1-R1	2	11	1.3078	4
S1-F6-R1	5.5	2	1.1653	3
S1-R1	1	-	0.8712	2
S1-F7-R1	2	2	0.8436	1
Forced S1-R1	Any	-	-	11

Packet Size = 1500 bytes

Table 5-2. Effective throughput and priority level for various relay combination

Based on the EMR scheme and data packet size of 1500 bytes, the effective throughput assuming no loss in the wireless channel for various possible combinations of sender-relay-receiver link are shown in Table 5-2. The effective throughput values were collected from simulations results and were obtained by measuring the packets received at the receiver (R1) when various combination of sender (S1), receiver (R1) and relay (F1) nodes were used. The packet from node S1 is sent as fast as possible. The effective throughput of various sender-relay-receiver and sender-receiver combinations is mapped to the priority values shown in Table 5-2. Note that all the columns in Table 5-2, except the first one, contain generic values that are independent of the nodes and

commonly available to every node. Only the first column, which shows the corresponding paths for the S1-R1 node pair, contains information specific to that node pair and is not a necessary component of the priority table. With the exception of the last option, denoted by “Forced S1-R1”, a high priority value indicates a higher preference because a higher effective throughput is achievable. In our design, it is possible for the primary routing protocol to discover shorter links even before the secondary scheme is used. This could especially happen for the last or the first link of the entire end-to-end route. For such links, a relay might not be necessary if the direct S1-R1 link rate is more favourable. For such links, we assume that the sender S1 and the receiver R1 are able to select the best link rate based on the estimated signal strength value.

5.3.2 Relay discovery and selection

In EMR, we encode the current link priority value into the RTS and CTS frames. The priority value, which is a 4-bit value, can be encoded into the unused fields⁴ of the Frame Control Field in the RTS and CTS frames. In EMR, the multi-rate relaying path is not formed immediately after the shortest path route is discovered but after a relay selection process is used. We refer to Figure 5-3 to illustrate the selection process. If the initial capacity between nodes S1 and R1 is 1 Mbps, node S1 will set its priority value to 2 according to Table 5-2. It will send out an RTS frame to node R1 using the base rate. Node R1 will reply with the CTS frame with the priority value similar to S1. The relaying nodes that are in region I (Figure 5-2) will be able to receive both the RTS and CTS frames. These relays will calculate the received signal strength for the RTS and CTS frames, deduce the effective link capacity and compute the appropriate

⁴ In ad hoc networks, we assume fields such as DS and WEP are unused.

priority value based on Table 5-2. If the priority value computed by the relay is higher than the current priority value, it will then send out a relay request broadcast packet, RELAY_RQST, to the sender S1 for the purpose of serving as a relay. The RELAY_RQST message consists of the relay MAC address and the path priority value. Sender S1 might possibly receive several RELAY_RQST messages from the other relays in the vicinity. Node S1 will select the best possible relaying node that has the highest priority value and send out a relay respond message, RELAY_RES, to the selected relay. The RELAY_RES message contains the selected relay MAC address and the agreed priority value. This relay will then be used for the subsequent data transmission and the new priority value will be encoded into the subsequent RTS frames sent out by node S1. In case of persistent data transmission failures due to loss of connectivity to a selected relay, node S1 can reset its priority value in the RTS frame to a low value and indicate a fallback to the normal 4-way handshake. This will trigger relays that can offer higher throughput links to start sending RELAY_RQST packets again. The process of selection repeats itself. In Table 5-2 above, the last entry labelled as “Forced S1-R1” is used if the sender S1 wants to use the direct route from S1 to R1, despite having a low link rate. The corresponding priority value encoded in the RTS and CTS packets will prevent relays in the vicinity from sending out RELAY_RQST packet.

5.3.3 Message Sequence

Figure 5-4. EMR MAC message sequence

The EMR MAC protocol message sequence is shown in Figure 5-4 above. We use F to represent one of the nodes F1–F7. Physical location of nodes S1, R1 and F are depicted in Figure 5-3 above. In Figure 5-4, we assume that nodes S1 and R1 are reachable to each other only by using the base rate of 1 Mbps. However, node F is capable of communicating with nodes S1 and R1 at data rates r_2 and r_1 , respectively. Using the relay selection scheme mentioned above, node F is selected. Node S1 sends out an RTS frame using the base rate and an encoded priority value. The use of a priority value greater than 1, explicitly informs the relay node and the receiver that this new message sequence will be used. The sender S1's and receiver R1's MAC addresses are encoded as usual in the RTS frame. Node R1 also replies with a CTS using the base rate. Based on the pre-agreed transmission rate supplied by relay F, nodes S1 and R1 will calculate the appropriate timeout period required for this message sequence and encodes the value in the Network Allocation Vector (NAV). Once S1 receives the CTS, it will transmit the DATA frame to the relay using rate r_2 . The destination address field in the DATA frame sent out by node S1 is replaced with the relay MAC address. If the relay node receives the data successfully, it will now adjust the transmission rate to r_1 and re-send the DATA frame to receiver R1. The destination MAC address field is encoded with receiver R1's address. Once node R1 receives the DATA frame correctly, it sends the ACK in the reverse direction using rate r_1 . The ACK frame is then relayed by node F to S1 using rate r_2 . If the ACK frame is received by S1 correctly, the transmission is completed. The scheme essentially uses an efficient multi-rate path with the same RTS-CTS overheads of the single rate scheme.

5.4 Analytical Model to Studying Throughput Performance

Bounds of Localized Links

In the recent past, many accurate models for CSMA/CA based protocols for WLAN scenarios [10,87] have been developed. While such models are suitable for co-located scenarios, they cannot be directly used for wireless ad hoc networks. Due to the complexity of wireless ad hoc networks where interference varies with time and location, there are very few accurate mathematical models to analyse the performance of a truly distributed wireless ad hoc network. The model suggested in [91] is by far the first attempt to model the throughput in wireless ad hoc networks by taking into consideration the interference from the neighbouring nodes. However the model makes certain simplifying assumptions such as density of interfering nodes and arrival rates are based on Poisson distribution. Consequently, the model does not entirely correspond to simulation results.

We extend the model proposed in [10] and limit the analysis of the throughput to the links shown in Figure 5-3 above. This approach is still valid since the links considered are co-located and are within range of each other when the base rate contention with RTS is used. We assume that the links formed by nodes S1, F and R1 in Figure 5-2 form a subset of a longer end-to-end link. The main objective of this performance analysis is to study the throughput performance bounds of a set of co-located links for the MTM and EMR protocols when the links are subjected to packet error, contention and hidden terminal interference. While it is difficult to quantify the saturation throughput of links because of time varying interference in wireless ad hoc networks, we simplify the analysis by fixing the level of blocking due to hidden terminal interference and channel noise to study the steady state throughput performances. In

contrast to this localized throughput analysis, in Section 5.6 , we use a packet level simulator to study the end-to-end delay and throughput performances.

For all cases, we assume that the contending nodes reside in region I as shown in Figure 5-2. To compute the throughput bounds of the two different schemes, we extend the model proposed in [10] by considering the packet error probability and blocking probability due to hidden terminal interference.

The PLCP Protocol Data Unit (PPDU) format of the data packet in IEEE 802.11b is shown in Figure 5-5. In our analysis, we assume that long PLCP header is used for all physical transmission rates. Based on Figure 5-5, the PLCP header and preamble are always transmitted at base rate whereas the MPDU can be transmitted at different rates.

Figure 5-5. PLCP Protocol Data Unit (PPDU) format

To simplify the packet error probability calculation, we model the system in the presence of Additive White Gaussian Noise (AWGN) in the wireless channel. For the subsequent performance study, we assume that a constant amount of noise is generated from the surrounding nodes and all nodes are affected in the same manner.

In a CSMA/CA based WLAN system, it is uncommon for nodes to have different channel sensing state because all nodes are usually in the transmission or carrier sensing range of each other. The model proposed in [10] only considers such a

situation. However, in a wireless ad hoc network this is not the case. In a multi-hop network, it is common for senders to sense an idle channel state while receivers sense a busy channel state. Such a phenomenon causes severe throughput degradation in wireless ad hoc networks [99] when the receivers do not reply to the RTS packet and the senders keep doubling their contention window. Notice that this phenomenon has also been classified as hidden terminal [91] or a form of exposed terminal [99]. It is essential that we take into account this phenomenon, while studying the throughput of the two MAC protocols. For the remaining part of this chapter, we quantify this parameter as the blocking probability, p_b . Similar to the channel noise assumption, we assume that nodes in region I of Figure 5-2 are subjected to the same blocking probability.

To simplify the model, we assume that perfect collision avoidance applies to the three nodes shown in Figure 5-2 since they are within transmission range of each other when RTS and CTS frames are used.

5.4.1 Error Probability Model

In this chapter, we consider the bit error probabilities of the different modulation schemes under AWGN and the modulation schemes that are used in IEEE 802.11b. The bit error probability, P_{be} , for BPSK and QPSK [75] is given as

$$P_{be} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \quad (\text{Eq. 5-1})$$

The bit error probability for CCK modulation which is a variation of M-ary Bi-orthogonal modulation, can be expressed as

$$P_{be} = 1 - \int_{-\sqrt{2E_b/N_0}}^{\infty} \left(\frac{1}{\sqrt{2\pi}} \cdot \int_{-(v+\sqrt{2E_b/N_0})}^{v+\sqrt{2E_b/N_0}} e^{\left(\frac{-x^2}{2}\right)} dx \right)^{\frac{M}{2}-1} \cdot e^{\left(\frac{-v^2}{2}\right)} dv$$

(Eq. 5-2)

where M is 4 and 8 for data rates 5.5 Mbps and 11 Mbps, respectively. The SNR values for a particular modulation scheme can be obtained from the following equation.

$$SNR = \frac{E_b}{N_0} \cdot \frac{R_b}{B_t}$$

(Eq. 5-3)

In (Eq. 5-3), E_b , R_b , B_t and N_0 , are the signal energy per bit, the maximum bit rate of the modulation scheme, the system bandwidth and the noise per bit, respectively. To compare the model with actual empirical values, we normalize the SNR curves obtained from the mathematical equations to fit the published values of the Orinoco wireless transceiver [75]. The receiver sensitivity and the minimum SNR required for various transmission rates are shown in Table 5-3. The SNR vs. BER curves for the various modulation schemes are shown in Figure 5-6.

Rate (Mbps)	Receiver Sensitivity	Minimum SNR requirement
1	-94 dBm	4 dB
2	-91 dBm	7 dB
5.5	-87 dBm	11 dB
11	-82 dBm	16 dB

Table 5-3. Orinoco Wireless Card Parameters

Figure 5-6. SNR vs. BER

The noise floor was computed by calculating the thermal noise using the well-known formula

$$N_t = kTB_t \quad (\text{Eq. 5-4})$$

where k = Boltzman constant (1.38×10^{-23} Joules/Kelvin), T = system temperature and B_t is the channel bandwidth (Hz). The system temperature is assumed to be 290K. We add a typical noise figure, N_f of 7 dB to account for the losses in the receiver.

To calculate the path loss at a distance d , given by $P_l(d)$, we use the log-distance path loss model given by the following expression

$$P_l(d) = P_l(d_0) + 10 \cdot n \cdot \log_{10} \left(\frac{d}{d_0} \right) + X_\sigma \quad (\text{Eq. 5-5})$$

where $P_l(d_0)$ is the median path loss at the reference distance of 1 meter, n is the path loss coefficient and X_σ is the shadow fading loss. To simplify the analysis, we ignore the shadow fading loss. We let the propagation model consist of two parts: 1) free-space propagation for distances less than the Friis cutoff distance, d_{friss} and 2) two-ray

ground propagation for distances greater than d_{friss} . The path loss exponent, n , for the two propagation models are 2 and 4, respectively. We use the standard Orinoco Wireless transceiver parameters [75], which are also used in NS-2. Consequently, the path loss in dB is given as

$$P_l(d) = 31.66 + 20 \cdot \log(d) \quad \text{if } d < d_{friss} \quad (\text{Eq. 5-6})$$

$$P_l(d) = 70.365 + 40 \cdot \log\left(\frac{d}{d_{friss}}\right) \quad \text{if } d > d_{friss} \quad (\text{Eq. 5-7})$$

Given the path loss, the receiver sensitivity, P_{rx} can be expressed in terms of the transmission power, P_{tx} as

$$P_{rx} = P_{tx} + G_{tx} + G_{rx} - P_l(d) - L_{fade} \quad (\text{Eq. 5-8})$$

where G_{tx} , G_{rx} and L_{fade} are the transmitter antenna gain, receiver antenna gain and fade margin, respectively. In addition, the receiver sensitivity, P_{rx} , can be expressed in terms of SNR by

$$P_{rx} = N_t + N_f + SNR \quad (\text{Eq. 5-9})$$

From (Eq. 5-9), we are able to use the SNR value to compute the BER for different distances of d . Table 5-4 describes the parameters used in the analysis.

PARAMETERS	
G_{tx}	0 dB
G_{rx}	0 dB
P_{tx}	24.5 dBm
L_{fade}	30 dB
d_{friss}	86.1424 m

Table 5-4. Path loss related parameters

5.4.2 Enhancement to Bianchi's Markov Chain Model

We now discuss the enhancements to the model proposed in [10] with additional consideration for frame error probability and blocking probability. For convenience, we reuse the notations found in [10]. In [10], the conditional collision probability p , which we refer to as p_c in this paper, was the only factor that affected the transitions in the Markov Chain Model. We reuse the same Markov Chain Model proposed in [10] but assume that a new conditional probability, p_f , which we term as the conditional probability of failure, causes the states to transition in the same fashion. We note that this failure event consists of three independent but not necessarily mutually exclusive events, i.e. collision, blocking due to hidden terminal interference and frame error due to noise. Since any one of these events can cause the binary backoff window to double, we state the conditional probability of failure as

$$p_f = p_c + p_e + p_b - p_c p_e - p_c p_b - p_b p_e + p_b p_e p_c$$

(Eq. 5-10)

where p_c is the conditional probability of collision, p_b is the conditional blocking probability and p_e is the conditional probability of packet error due to channel noise.

From [10], the probability τ that a station transmits in a randomly chosen slot time is given by

$$\tau = \frac{1 - p_f^{m+1}}{1 - p_f} b_{0,0}$$

(Eq. 5-11)

where $b_{0,0}$ is given as

$$b_{0,0} = \frac{2(1-2p_f)(1-p_f)}{(1-2p_f)(W+1) + pW(1-(2p_f)^m)} \quad (\text{Eq. 5-12})$$

where m is the maximum backoff stage, W is the minimum window size. These values are set to 5 and 32, respectively. Instead of p used in [10], we express p_c as

$$p_c = 1 - (1 - \tau)^{n-1} \quad (\text{Eq. 5-13})$$

where n is number of nodes directly contending with each other within the transmission range.

In the subsequent expressions, we assume that a packet of size L is transmitted using the PHY mode r , where $r = 1, 2, 3, 4$ represents the PHY rates of 1, 2, 5.5 and 11 Mbps, respectively. Since the conditional probability of packet error, p_e , is independent of τ , for the MTM scheme, we can expressed p_e as

$$p_e = 1 - P_{succ}^r(L) = 1 - (1 - P_{e_rts}^1)(1 - P_{e_cts}^1)(1 - P_{e_data}^r(L))(1 - P_{e_ack}^r) \quad (\text{Eq. 5-14})$$

where $P_{succ}^r(L)$ is the probability of successful transmission of a L -byte packet; $P_{e_rts}^1$, $P_{e_cts}^1$, $P_{e_data}^r(L)$ and $P_{e_ack}^r$ are the error probabilities for the RTS, CTS, DATA and ACK frames, respectively. Notice that RTS and CTS are always transmitted at base rate. For the EMR scheme, p_e can be expressed as

$$\begin{aligned} p_e &= 1 - P_{succ}^r(L) \\ &= 1 - (1 - P_{e_rts}^1)(1 - P_{e_cts}^1) \cdot (1 - P_{e_data}^{r1}(L))(1 - P_{e_data}^{r2}(L))(1 - P_{e_ack}^{r1})(1 - P_{e_ack}^{r2}) \end{aligned} \quad (\text{Eq. 5-15})$$

where $r1$ and $r2$ signify the rates used by the relay in Figure 5-3 above. Based on the PPDU format of the data packet shown in Figure 5-5, we can express

$$P_{e_data}^r(L) = 1 - (1 - P_e^1(24)) \cdot (1 - P_e^r(28 + L)) \quad (\text{Eq. 5-16})$$

where $P_e^1(24)$ is the probability of error due to the transmission of the PLCP header using PHY mode 1 and $P_e^r(28 + L)$ is the probability of error for the MPDU when mode r is used. The frame error probability $P_e^r(L)$, for different modes and distances, d , can be computed by using the BER value, P_b^r , obtained from the following equation:

$$P_e^r(L) = 1 - (1 - P_b^r)^{8L} \quad (\text{Eq. 5-17})$$

The value, $8L$, in equation (Eq. 5-17) represents the number of bits in the frame. As mentioned earlier, we will assume a constant blocking probability in this analysis. Since blocking is an independent event, we can easily substitute p_b and p_e into equation (Eq. 5-10). Using numerical techniques, the unknowns p_f and τ can be determined by using the equations (Eq. 5-10), (Eq. 5-12) and (Eq. 5-13).

5.4.3 Computing Throughput

Similar to [10], we consider the saturation throughput in this analysis. The single link saturation throughput, S_{link} , for a MAC scheme, can be expressed as

$$S_{link}^x = \frac{\text{E[payload information transmitted in a slot time]}}{\text{E[length of slot time]}} \quad (\text{Eq. 5-18})$$

In the MTM scheme, we analyse two adjacent links as a single end-to-end link, whereas in EMR, we consider the two-hop links via the relay as a single end-to-end link. Since the two adjacent links are within contention range of each other, the end-to-end throughput for the MTM scheme, S_{MTM} , for nodes S1 to R1 can be approximated as

$$S_{MTM} = \frac{S_{link1}^{MTM} \cdot S_{link2}^{MTM}}{S_{link1}^{MTM} + S_{link2}^{MTM}} \quad (\text{Eq. 5-19})$$

where S_{link1}^{MTM} is the maximum throughput of the first link assuming that only the first link, denoted by link S1-R, exist in region I of Figure 5-2. Similarly S_{link2}^{MTM} is computed by assuming that only the second link exists in region I of Figure 5-2. We modify the expression obtained from [10] by adding the packet error and blocking events. S_{link1}^{MTM} can be expressed as:

$$S_{link1}^{MTM} = \frac{P_s E[P]}{P_i \sigma + P_s T_s^{MTM} + P_c T_c + P_b T_c + P_{e1} T_{e_rts} + P_{e2} T_{e_cts} + P_{e3} T_{e_data}^{MTM} + P_{e4} T_{e_ack}^{MTM}} \quad (\text{Eq. 5-20})$$

In equation (Eq. 5-20), we assume that collisions are due to two or more frames colliding, therefore resulting in indiscernible information whereas frame error happens when there is only a single transmission and packet is received properly but with wrong frame check sequence (FCS). Frames with errors will normally trigger the EIFS backoff. The fractional error probabilities P_{e1} , P_{e2} , P_{e3} and P_{e4} are used to express the probability of frame error at different stages of the entire RTS-CTS-DATA-ACK message exchange for the MTM scheme. The values, P_i , P_s , P_c and P_b denote the probability of idle, probability of success, probability of collision and probability of blocking during the interval of the slot time, respectively. The term $E[P]$, denotes the average payload transmitted in a slot time. The values, σ , T_s^{MTM} , T_c , represent the

average time of an empty slot time, average time of a successful message exchange and average time due to collision, respectively. Whereas the terms T_{e_rts} , T_{e_cts} , $T_{e_data}^{MTM}$, $T_{e_ack}^{MTM}$, denote the average time incurred at the various stages of frame error during the entire message exchange. The various probability expressions and parameters are expressed as follows.

$$\begin{aligned}
 P_s &= n\tau(1-\tau)^{n-1} \cdot (1-p_e) \cdot (1-p_b) \\
 P_{str} &= n\tau(1-\tau)^{n-1} \\
 P_i &= (1-\tau)^n \\
 P_c &= 1 - n\tau(1-\tau)^{n-1} - (1-\tau)^n \\
 P_b &= P_{str} \cdot p_b \\
 P_{e1} &= P_{str} \cdot P_{e_rts}^1 \\
 P_{e2} &= P_{str} \cdot (1-P_{e_rts}^1) \cdot P_{e_cts}^1 \\
 P_{e3} &= P_{str} \cdot (1-P_{e_rts}^1) \cdot (1-P_{e_cts}^1) \cdot P_{e_data}^r \\
 P_{e4} &= P_{str} \cdot (1-P_{e_rts}^1) \cdot (1-P_{e_cts}^1) \cdot (1-P_{e_data}^r) \cdot P_{e_ack}^r
 \end{aligned}
 \tag{Eq. 5-21}$$

$$\begin{aligned}
 T_c &= RTS^1 + DIFS + \delta \\
 T_{e_rts} &= RTS^1 + \delta + EIFS \\
 T_{e_cts} &= RTS^1 + SIFS + \delta + CTS^1 + \delta + EIFS \\
 T_{e_data}^{MTM} &= RTS^1 + SIFS + \delta + CTS^1 + SIFS + \delta + DATA^r + \delta + EIFS \\
 T_{e_ack}^{MTM} &= T_{e_data}^{MTM} + SIFS + ACK^r + \delta \\
 T_s^{MTM} &= RTS^1 + SIFS + \delta + CTS^1 + SIFS + \delta \\
 &\quad + DATA^r + SIFS + \delta + ACK^r + DIFS + \delta
 \end{aligned}
 \tag{Eq. 5-22}$$

The term P_{str} in (Eq. 5-21), denotes the probability that only a single transmission occurs. The $DATA^r$ and ACK^r terms shown in equations (Eq. 5-22) include the physical layer headers and the superscript r represents the transmission mode used. The throughput for the EMR scheme can be expressed by equation (Eq. 5-23). Similar to (Eq. 5-20), the fractional error probabilities P_{e1} , P_{e2} , P_{e5} , P_{e6} , P_{e7} , P_{e8} , express the frame

error at different stages of the entire RTS-CTS-DATA-DATA-ACK-ACK message exchange for the EMR scheme. T_s^{EMR} represents the average time of a successful EMR message exchange. Whereas the terms T_{e_rts} , T_{e_cts} , $T_{e_data1}^{EMR}$, $T_{e_data2}^{EMR}$, $T_{e_ack1}^{EMR}$, $T_{e_ack2}^{EMR}$, denote the average time incurred at the various stages of frame error during the EMR message exchange. The various probability expressions and parameters for the EMR scheme are expressed in equations (Eq. 5-24), (Eq. 5-25) and (Eq. 5-26). Rates $r1$ and $r2$ in equations (Eq. 5-25) and (Eq. 5-26) signify the first and second link rates, respectively.

$$S_{link}^{EMR} = \frac{P_s E[P]}{P_i \sigma + P_s T_s^{EMR} + P_c T_c + P_b T_c + P_{e1} T_{e_rts} + P_{e2} T_{e_cts} + P_{e5} T_{e_data1}^{EMR} + P_{e6} T_{e_data2}^{EMR} + P_{e7} T_{e_ack1}^{EMR} + P_{e8} T_{e_ack2}^{EMR}}$$

(Eq. 5-23)

$$\begin{aligned} P_{e5} &= P_{str} \cdot (1 - P_{e_rts}^1) \cdot (1 - P_{e_cts}^1) \cdot P_{e_data}^{r1} \\ P_{e6} &= P_{str} \cdot (1 - P_{e_rts}^1) \cdot (1 - P_{e_cts}^1) \cdot (1 - P_{e_data}^{r1}) \cdot P_{e_data}^{r2} \\ P_{e7} &= P_{str} \cdot (1 - P_{e_rts}^1) \cdot (1 - P_{e_cts}^1) \cdot (1 - P_{e_data}^{r1}) \cdot (1 - P_{e_data}^{r2}) \cdot P_{e_ack}^{r2} \\ P_{e8} &= P_{str} \cdot (1 - P_{e_rts}^1) \cdot (1 - P_{e_cts}^1) \cdot (1 - P_{e_data}^{r1}) \cdot \\ &\quad (1 - P_{e_data}^{r2}) \cdot (1 - P_{e_ack}^{r2}) \cdot P_{e_ack}^{r1} \end{aligned}$$

(Eq. 5-24)

$$\begin{aligned} T_{e_data1}^{EMR} &= RTS^1 + SIFS + \delta + CTS^1 + SIFS + \delta + DATA^{r1} + \delta + EIFS \\ T_{e_data2}^{EMR} &= T_{e_data1}^{EMR} + SIFS + DATA^{r2} + \delta \\ T_{e_ack1}^{EMR} &= T_{e_data2}^{EMR} + SIFS + ACK^{r2} + \delta \\ T_{e_ack2}^{EMR} &= T_{e_ack1}^{EMR} + SIFS + ACK^{r1} + \delta \end{aligned}$$

(Eq. 5-25)

$$T_s^{EMR} = RTS^1 + SIFS + \delta + CTS^1 + SIFS + \delta + DATA^1 + SIFS + \delta + DATA^2 + SIFS + \delta + ACK^{r1} + DIFS + \delta + ACK^{r2} + DIFS + \delta$$

(Eq. 5-26)

5.5 Performance Evaluation

PARAMETERS	
RTS	352 bits
CTS	304 bits
Data payload	1500 bytes
ACK	304 bits
Propogation delay (δ)	2 μ s
Slot time(σ)	20 μ s
SIFS	10 μ s
DIFS	50 μ s
EIFS	364 μ s

Table 5-5. IEEE 802.11b Parameters

Figure 5-7. CSMA/CA - 1 Mbps links under various blocking probability

Figure 5-8. CSMA/CA - 11 Mbps links under various blocking probabilities

The results in this section have been obtained by using the standard parameters shown in Table 5-5 and a data packet size of 1500 bytes. Before we discuss the end-to-end throughput of the two schemes, we first show the effect of blocking probability on throughput of a wireless multi-hop link. Figure 5-7 shows the effect of throughput on a single-hop link under varying degree of blocking probabilities and number of contending nodes. The increase in node density causes more contention and collisions, which then affects the degree of exponential backoff. The objective of these experiments is to study how blocking probability along with the 802.11 binary exponential backoff mechanism affects the throughput of the system. Figure 5-7 was obtained when all links in the vicinity are using 1 Mbps transmission rate. Although in practical situations, studying fifty (50) nodes within a single hop link is reasonable, we increase the number of contending nodes to two thousand (2000) for the sake of observing the asymptotic trend in throughput for larger number of nodes.

Figure 5-7 and Figure 5-8 reveal some interesting properties on the achievable capacity of the links in a wireless ad hoc network under various blocking probabilities and are of fundamental theoretical importance to this study. In Figure 5-7, it is interesting to note that under zero blocking probability (no hidden node), which is the case simulated or

studied in previous WLAN cases [10,87], the system throughput decreases very slowly when contention increases. Therefore, when we increase the node density from 200 to 400 nodes, the throughput remains almost the same. Under higher blocking probability, the throughput of the system is low when the number of nodes in the system is low (e.g. a single node). However, the throughput reaches a maximum limit when $n \sim 100$ nodes before deteriorating. The reason for lower throughputs in the low node density region is that nodes spend most of their time in the backoff stage. Around the peak throughput region, the number of nodes is optimal. When the node density increases, more collisions occur, thus resulting in throughput degradation.

A similar performance evaluation carried out on a set of 11 Mbps links (Figure 5-8) show steeper throughput degradation when compared to the 1 Mbps link. Therefore, to compare the performance of the two protocols under varying degree of contention, it is essential that we take into consideration the system operating regions highlighted in Figure 5-7 and Figure 5-8.

Based on Figure 5-7 and Figure 5-8, we will compare the two MAC schemes under low, medium and heavy node density settings but under a constant blocking probability factor of 0.5. We use a total of m nodes in the system to analyse the two protocols. We also use the term n to represent the number of contending nodes in each system. Note that, while the total number of nodes m , in the two systems is the same; the number of contending nodes, n , in MTM is always indicated as double that of EMR for equal comparison. This is because MTM involves double the number of nodes for carrying out the channel contention (RTS-CTS and binary backoff) as compared to EMR.

Figure 5-9. Position of nodes used in analytical model

Figure 5-9 shows the position of the nodes used to evaluate the performance of the two MAC schemes, based on the analytical model. All contending nodes reside in the circular region shown in Figure 5-9 and sender S1 and receiver R1 are placed on coordinates $(-110, 0)$ and $(110, 0)$ on the x and y axis, respectively. Because the distance between nodes S1 and R1 is 220 m, only the base rate of 1 Mbps can be used to communicate directly. Since all nodes in this circular region fall in region I of Figure 5-2, the model described in Section 5.4 is applicable. The position of relay F is varied within the circular region shown in Figure 5-9 and the throughput is evaluated and shown in Table 5-6 and Figure 5-10 to 13. Table 5-6 shows the maximum and minimum aggregate end-to-end throughput of the two schemes under various node densities, no added noise interference and a fixed blocking probability of 0.5.

m	MTM n	EMR n	MTM – Aggregate Throughput (Mbps)		EMR – Aggregate Throughput (Mbps)	
			Max	Min	Max	Min
3	2	1	1.57	0.74	1.86	0.73
300	200	100	1.87	0.87	2.61	0.98
600	400	200	1.68	0.82	2.51	0.96

Table 5-6. Aggregate End-to-End Throughput for MTM and EMR scheme

Based on Table 5-6, under low density ($m = 3$ nodes), the throughput of EMR scheme only exceeds that of MTM scheme by 18.5%. Under medium contention, when the number of contending nodes in MTM and EMR are 200 and 100, respectively, the throughput of the EMR scheme shows a significant improvement of 39.6% over the MTM scheme.

Figure 5-10. MTM with high density

Figure 5-11. EMR with high density

Figure 5-10 and Figure 5-11, show the aggregate throughput when the number of contending nodes in MTM and EMR are 400 and 200, respectively. The axes labelled

as ‘X’ and ‘Y’ are the X and Y coordinates, respectively. The vertical axis is the aggregate throughput of the links from sender S1 to receiver R1 when node F is placed on various X and Y coordinate within the circular region shown in Figure 5-9. The throughput of EMR scheme exceeds MTM by 49.4%.

Figure 5-12. MTM with low density and added noise of 3dBm

Figure 5-13. EMR with low density and added noise of 3dBm

Figure 5-12 and Figure 5-13 show the throughput of the two schemes under a uniform noise interference of 3 dBm from the surroundings. The throughput of the two systems

drop but the improvement of EMR over MTM is only 13.1% as compared to the 18.5% improvement shown in Table 5-6 for the same number of nodes.

5.6 Simulation Results

Figure 5-14. Throughput performance in a chain setup

Figure 5-15. Delay performance in a chain setup

In this section, we present the simulation results that were obtained by using NS-2. Both the MAC protocols, MTM and EMR were implemented in NS-2. The objective of the simulation is to study the performance of the base rate IEEE 802.11, EMR and MTM MAC protocols in a simple multi-hop chain network and a more complex mobile ad hoc network scenario. In the multi-hop chain setup, each adjacent node is placed 100 m apart in a straight line and the number of nodes in the chain is varied from 3 to 11 nodes. In both setups, we used a packet size of 1500 bytes. Figure 5-14 describes the throughput of MTM, EMR and the standard 1 Mbps single-rate IEEE 802.11 DCF MAC. In Figure 5-14, the 'nodes' in the horizontal axis, signifies the number of nodes in the multi-hop chain setup. We notice an average 19.74% improvement of throughput for the EMR scheme over the MTM scheme. Figure 5-15 describes the delay performance of the three schemes. The EMR scheme exhibits the lowest delay when compared to the other schemes. For the low hop count (2 hops) path, MTM has 19.74% higher delay compared to EMR. When the hop count is high (10 hops), the difference in delay can increase to as much as 51.41% of EMR. This is primarily due to the higher rate of collisions encountered in the MTM case when compared to EMR. Clearly, the delay performance of the EMR scheme is far better than the 1 Mbps base rate and the MTM scheme.

Figure 5-16. Aggregate throughput in the presence of mobility

In the mobile ad hoc network scenario, fifty (50) nodes were placed randomly in a 1500 m by 1500 m area and allowed to move randomly using the Random Way Point mobility model [43]. The speed of nodes was increased from 5 m/s to 20 m/s and a pause time of 5 seconds was used in all setups. Constant bit rate (CBR) connections with a rate of 20 packets per second were initiated between randomly selected sender and receiver pairs at random time intervals. Each simulation lasted for 200 seconds and the results were averaged over 12 simulation runs. An average 35 connections were used in each simulation run. Figure 5-16 shows the aggregate throughput of the three protocols.

In general the difference for the aggregate throughput of the two multi-rate protocols, EMR and MTM, over the base rate is not as high for this scenario when compared to the static multi-hop chain setup. In the low mobility case (5 m/s), the EMR protocol still performs reasonably better than the MTM protocol. The improvement of aggregate

throughput recorded for EMR over MTM and the base rate 802.11 for this scenario is 14% and 43%, respectively. As the speed of the node increases to 10 m/s all three protocols suffer the reduction in aggregate throughput. In this scenario, the number of route breakages increases slightly and the AODV routing protocol tries to repair the routes. For this speed, most of the larger hop-count links are still repairable. The three protocols register a higher throughput for the 15 m/s case when compared to the 10m/s case. In the 15 m/s case, the longer hop-count links suffer from constant breaks in routes. The improvement of throughput for this speed as shown in Figure 5-16 is mainly contributed by shorter hop-count links. In the higher mobility case, with node speed of 20 m/s, the performance of the multi-rate over the base rate system is less significant. In the higher mobility cases, the EMR scheme suffers throughput degradation because of the frequent change of relay node locations. The constant broadcast of relay request and relay response messages also incurs some bandwidth overheads.

5.7 Conclusion

In this chapter, we introduced a new multi-rate relaying MAC protocol known as EMR, which is able to utilize the traditional shortest path routing algorithm and a localized secondary forwarding scheme to improve the throughput of a multi-hop network. In addition, we enhanced an existing analytical model proposed by Bianchi [10], which is primarily used for WLAN to study the performance of CSMA/CA based protocol in a wireless ad hoc network. Our objective in performing this analysis is to study the saturation throughput of a set of localized links, which are subjected to hidden terminal interference and channel noise. The end-to-end throughput and delay performance of the EMR scheme have been carefully studied using a packet level simulator. The EMR

scheme is shown to perform much better than the previously proposed multi-rate schemes. The analytical model shows that the throughput gains of our protocol can reach up to 49.4% when compared to the current state-of-the-art multi-rate routing scheme for the high-density case.

Chapter 6. A Flow Control Framework for Improving Throughput and Energy Efficiency in CSMA/CA based Wireless Multihop Networks

In this chapter, we carefully study some of the main contributing factors of throughput degradation in CSMA/CA based MAC protocol. In addition, we also analyze the congestion problem, and show its relationship to the inefficiency of the MAC scheme. Thus the primary focus of this work is to develop a framework for flow control for CSMA/CA based multihop wireless networks, which can improve the throughput performance while conserving communication energy.

Flow and congestion control have been well researched in the context of wired networks [46,51,59,67]. Over the past few years, they have been receiving wide attention in the context of wireless ad hoc networks [12,84,102,103], and in sensor networks [35,90]. Unlike the wired access medium, the wireless access medium, which is typified by hidden terminal problems, broadcast transmissions, and random access methods, has significant contributions to the congestion problem. In a typical wireless multihop network, each node has a single transceiver, which is used to compete with other nodes in the shared medium to access the channel. Besides sending its own packets, each node also serves to forward other transiting packets. If a transceiver is captured more often by other nodes for receiving incoming transit packets rather than

being allowed to send packets out, then congestion at that node will occur. In the wired medium, such a phenomenon does not exist since each link operates independently of others. Congestion, which leads to dropped packets at the downstream nodes, is obviously detrimental to the energy efficiency of a wireless ad hoc or sensor network because of the wasted energy used to forward packets along multiple hops.

In this chapter, we examine and design several techniques that span different layers of the traditional protocol stack to address the performance of the CSMA/CA MAC based multihop network. These include a novel hop-by-hop, hybrid rate and window based flow control scheme and an enhanced CSMA/CA MAC protocol. Collectively, we term the entire work as a flow control framework. The choice of a hop-by-hop flow control scheme over an end-to-end scheme is straightforward since feedback is faster resulting in shorter response time to congestion. In addition, it is easier to design a loss less scheme that can complement the modest energy level of the nodes.

As a precursor to the design of the proposed framework, we identify the factors that affect (some of which are new) the throughput of CSMA/CA-based multihop wireless networks. The impacts of these factors are quantified experimentally. An enhanced CSMA/CA-based scheme is developed to mitigate the problems identified. In proposing the flow control scheme, we pursue a fundamentally different approach, which we refer to as a hop-by-hop, hybrid window and rate-based flow control scheme. As the name suggests, the rate control and window mechanism is implemented on a per-hop basis. Previous hop-by-hop protocols use either a rate control [59] or a window scheme [51] but not both in combination. While previous hop-by-hop schemes suffered from per-flow management and rate measurement complexity [51,59], we make certain

simplifications to the rate control and window design to make our design feasible for multihop wireless networks. In total, the scheme prevents congestion and reduces packet loss. While this work is not specifically about IEEE 802.11, most of our experimental results and enhancements are based on the IEEE 802.11 DCF MAC protocol. However, the suggested framework is applicable to a general CSMA/CA MAC protocol.

6.1 Related Works

Flow control in wired networks have been proposed and studied for several decades with end-to-end flow/congestion control schemes dominating the earlier designs [37,60,77]. Some of these designs have been adopted into the well accepted TCP protocol suite [37]. Due to low complexity in design, scalability and ease of deployment, end-to-end flow control schemes such as TCP have dominated the wired networks.

However, as pointed out in [6,21,31,57,84,98,103], end-to-end flow control such as TCP fails to support high utilization of bandwidth in multihop wireless networks. The authors in [31] have specifically investigated the effects of mobility on TCP performance in mobile wireless ad hoc networks. They further proposed an explicit link failure notification (ELFN) to help TCP differentiate between congestion and link failure losses. The authors in [21] have also closely studied the interactions between TCP and the underlying MAC protocol such as 802.11. Their study reveals the optimum window size that allows TCP to achieve the highest spatial reuse and also a stabilized window size that TCP converges to but results in reduced throughput. Unlike [6,21,31,57,103], which proposed enhancements to TCP, ATP [94] is an entirely new transport protocol, which is not a variant of TCP. ATP, which is still an end-to-end

flow control scheme, introduces an assisted congestion control scheme, which uses intermediate nodes to mark the packets that travel towards the receiver. The receiver nodes will further perform an exponential averaging and send rate feedback to the sender. Rate control, is then performed at the sender.

Hop-by-hop flow control started to emerge in the early 90's [12,51,59]. Hop-by-hop flow is usually favored because it provides faster feedback response to congestion. However, the benefits of hop-by-hop schemes normally come at the expense of additional explicit notification messages [12,59] to notify the upstream nodes about the congestion and maintenance of per-flow information at each node [12,51,59]. Recently, hop-by-hop flow control schemes have been suggested for use in wireless ad hoc networks [12,102] and sensor networks [35,90].

Hop-by-hop schemes typically use either a rate control or a dynamic window to implement the flow control. The scheme proposed by [59] is a rate-based flow control implemented on a per-hop basis. Nodes typically measure the rate and buffer occupancy of flows passing through it and explicitly pass this information to upstream nodes. Each node is responsible for throttling the flows at the onset of congestion. In [35,12], the schemes defer as the explicit feedback is sent to the main source to control the source rate. These schemes can suffer from scalability issues because of the expensive resources required to maintain per-flow information and compute the actual rate of the flows. The scheme proposed by [51], which is known as the credit-based flow control is a typical dynamic window scheme implemented on a per-hop basis. A downstream receiver monitors queue lengths of each flow and determines the number of packets or "credit" an upstream sender can transmit on a link. The sender transmits only as many packets as allowed by the credit. The scheme essentially monitors the

outgoing links, determines the link round trip times and allocates the buffer evenly to flows passing through the node. Periodic feedback is required to indicate the available credits to the upstream nodes. Similar to [59], this scheme requires per-flow management and dedicated buffers, which is not scalable in wireless ad hoc or sensor networks. In addition, the schemes proposed in [12,51,59] do not promote spatial spreading of packets belonging to the same flow, which could result in self-contention and reduced throughput at the MAC layer.

The schemes proposed in [35,90] are examples of recent efforts in designing lightweight flow control schemes specially targeted for sensor networks. This scheme monitors the buffer occupancy level and uses a binary value to indicate congestion. The binary congestion value, which is coded in the data frame, is then broadcasted to upstream (child) node to stop the transmission of frames. The upstream (child) node, which hears this binary congestion value, will similarly broadcast this value upstream till it reaches the source. In addition to this simple hop-by-hop flow control, the scheme in [35] implements a per-hop rate control which controls the rate of upstream nodes. The sending rate is computed based on actively listening and computing the rate of the transit packets forwarded by the downstream nodes. In total, the scheme resembles a rate-based, hop-by-hop flow control scheme. The disadvantage of this technique is that if per-flow management is not considered, then the congestion bit is essentially propagated to all source nodes, which are serviced by the congested parent node. This could restrict some of the upstream nodes from forwarding packets in other directions if other paths exist. The scheme in [90] on the other hand implements a hop-by-hop flow control scheme with source rate control. In [90], congestion is detected by listening to the channel activity. Also, the congestion is signaled to the upstream nodes. Similar to

[35], this scheme is suitable for sensor networks where most data originate from sources and mostly travel in a single direction towards a sink.

In contrast to the schemes described in [35,90], our flow control scheme is targeted towards a more general multihop wireless network which typically have flows moving in arbitrary directions. As such, the distinction of flows is still required. The hop-by-hop flow control scheme that we described in this chapter differs significantly from other hop-by-hop schemes mentioned above. In our scheme, we introduced a hybrid per-hop static window and rate control scheme to dynamically adjust the flow rate.

6.2 Factors affecting throughput in multihop transmission

In this section, we diagnose and discuss some of the issues that affect throughput in a multihop wireless network when CSMA/CA MAC protocol is used. The effect of exposed node problem on throughput has been discussed in great detail in [55]. In [68], we have shown via simulation that throughput in CSMA/CA based multihop networks can be further degraded by the “critically exposed node problem”. We briefly reiterate this problem and further reveal three more reasons, which significantly cause the overall throughput in multihop networks to degrade.

6.2.1 The critically exposed node problem

Figure 6-1. Chain setup with flow traveling from node 1 to node 9

In Figure 6-1, when node 7 is communicating to node 8, node 4 can initiate a transmission by sending a request to send (RTS) frame because it senses the channel as idle. However, a receiver such as node 5, will not reply since it senses the channel as busy. The location dependent interference causes disparity in the carrier sensing state between the sender (node 4) and the receiver (node 5). Due to this disparity and the use of binary exponential backoff (BEB) mechanism, node 4 will double its contention window and retry with another RTS frame. If node 4 continuously does not receive a response and the maximum retry limit is reached, the data frame will be dropped at node 4. Node 5 is termed as the “critically exposed node”, when node 7 is transmitting. Node 4 is essentially blocked from forwarding the packet.

6.2.2 Location dependency and uneven carrier sensing resulting in congestion

In a multihop wireless network, the location of nodes and the direction of transmission affect the throughput of a particular flow. We use Figure 6-1 and Figure 6-2 to illustrate the effect of uneven carrier sensing.

Figure 6-2. Uneven carrier sensing due to node location

In Figure 6-1, data packets are transmitted from node 1 to node 9 via the intermediate nodes. Figure 6-1 also shows the transmission and carrier sense ranges of node 7, when node 7 is transmitting. The intended destination of node 7 is node 8. The timelines in Figure 6-2 describe the reception of the signals at nodes 6 and 5, respectively, when

node 7 is engaged in a 4-way, RTS-CTS-DATA-ACK handshake with node 8. The shaded frames in Figure 6-2 represent the carrier sense range frames received at nodes 5 and 6. When a node is within the carrier sense region but outside the transmission range of a particular sender, the received frame will be treated as noise. Therefore the Network Allocation Vector (NAV) cannot be received correctly and the node can only depend on the physical carrier sense instead of the virtual carrier sense (NAV) to perform backoff. In the setup shown in Figure 6-1, node 5 is outside the carrier sense region of node 8. Therefore node 5 experiences longer idle time when compared to node 6, especially during the CTS and ACK exchanges from node 8. Due to this uneven idle sensing and unequal termination times (denoted as T_0 and T_1), node 5 will start counting down its backoff counter faster than node 6. This causes node 5 to always capture the channel faster than node 6. If the offered load from node 5 is high, and this phenomenon persists, then node 6 will spend most of its time being engaged in receiving packets rather than sending them. When this occurs, the queue at node 6 will soon be exhausted and congestion will prevail. The same phenomenon will occur anywhere along the chain if the offered load is high and the idle sensing period of the upstream node is greater than the downstream node. In later sections, it will be apparent that one of the techniques that we propose for improving throughput is designed on the basis of balancing the time spent by a node for receiving and transmitting packets.

6.2.3 False NAV

The use of virtual carrier sense or Network Allocation Vectors (NAV) was made popular in the IEEE 802.11 [36] protocol and continues to be adopted in newer protocols such as IEEE 802.15.4 [33]. Both physical and virtual carrier-sense functions are used to determine the state of the medium. In [36], the standard states that if either

function detects a busy channel, the medium shall be considered busy; otherwise it shall be considered idle. The virtual carrier sense mechanism is achieved by distributing reservation information announcing the impending use of the medium. The exchange of RTS and CTS frames prior to the actual DATA frame is one means of distributing this medium reservation information. The RTS and CTS frames contain a “Duration” field that defines the period of time that the medium is to be reserved to transmit the actual DATA frame and the returning ACK frame. All nodes within the reception range of either sender’s RTS or receiver’s CTS shall obey the medium reservation.

The wrong use of NAVs in CSMA/CA MAC protocols can be detrimental to the throughput of multihop wireless networks. We use the chain setup as shown in Figure 6-1 to illustrate the problem that currently exists. Consider the situation when node 7 is in the midst of transmitting a DATA frame to node 8, and node 4 tries to send an RTS frame to node 5. Node 5 does not reply with a CTS frame because it is critically exposed by node 7. However, the RTS sent out by node 4 might be received by node 3. In 802.11, when a sender sends an RTS packet, it sets its NAV and informs the surrounding nodes to backoff for the entire duration of the 4-way handshake message sequence. As such, the channel will still remain reserved by the sender who had initially sent the RTS packet. If the data packet is large, the backoff period imposed by NAV can be larger than the average backoff interval chosen by the binary backoff algorithm at any backoff stage. In Figure 6-1, when node 2 tries to communicate to node 3, even when both nodes senses an idle channel, node 3 will not reply due to the initial NAV value sent out by node 4. Repeated RTS retries from node 2 will result in dropped packets at node 2 when the maximum retry limit is reached. Effectively, this “false NAV” indication from node 4 results in a larger exclusion area and prevents certain nodes from using the channel effectively. In some cases, this scenario could

have a domino effect, thereby, impacting drastically on the throughput of multihop wireless networks.

6.2.4 Frozen MAC State

Under perfect collision avoidance, a returned CTS frame should be free from collision since the RTS sender would have reserved the channel around it. However, due to hidden terminal problem, there is a possibility for frames such as ACKs to collide with the reception of the CTS frame. We use Figure 6-1 to illustrate this problem. Assuming that in the first instance, only nodes 4 and 5 are engaged in a 4-way handshake and node 4 is in the midst of sending a DATA frame to node 5. Midway through this transmission, node 7 transmits an RTS frame to node 8. Meanwhile, when node 7 is transmitting its RTS frame, node 5 completes the reception of the DATA frame and sends out an ACK frame regardless of the channel sensing state. During this period, node 8 receives the RTS frame properly and sends out a CTS frame because it senses the channel as idle. However, due to node 7's proximity to node 5, it receives the ACK frame from node 5 and the CTS frame collides at node 7. Node 7 will cease to send the DATA packet but the MAC state in node 8 anticipates the reception of a DATA frame. However, it will remain frozen in this state and will not respond to other incoming RTS frames, including those from the original sender, until the initial DATA frame interval has timed out. Such a scenario also causes blocking and results in unnecessary RTS retries.

6.3 Severity of the Congestion, False NAV and Frozen MAC state Problem in Multihop networks

In this section, we quantify the impact of the causes for performance limitations outlined in the previous section.

6.3.1 Metrics

Throughout this chapter, we use a couple of measures to evaluate the performance of the original and newly proposed scheme: *end-to-end throughput*, (S), and *transmission cost*, (ψ). The *transmission cost*, (ψ), measures the amount of bits expended by the nodes in the system to transmit a single data bit from the source to the destination. The proposed metric is motivated by the metric proposed in [104]. However, this definition is modified to cater for a multihop scenario and the overheads incurred by all the 4-way handshake frames. Therefore, the *transmission cost*, ψ , which gives us some indication of the energy wasted by the system is defined as the ratio of all transmitted frames in the system over the data frames received at the receiver.

6.3.2 Experiments

Using a series of experiments carried out by using NS-2 simulator, we demonstrate the problem of congestion, false NAV and frozen MAC state in multihop transmissions. All the simulations in this chapter are based on the 802.11 Distributed Coordination Function (DCF) protocol with transmission rate of 1 Mbps. The transmission and carrier sense ranges are 250 m and 550 m, respectively. A packet capture model which only accepts packets that arrive within a short capture interval, t_c , of $4\mu\text{s}$ and ignores other received packets even with higher power signal is implemented into the original

802.11 DCF scheme. Unless stated otherwise, static routing is used in all the experiments to omit the effect of routing layer inefficiencies and accurately study the effect of MAC protocol alone on the performance of the multihop network. In addition, a fixed packet length of 1500 bytes is used in all experiments.

Figure 6-3. End-to-end throughput for chain setup

The first experiment was carried out by using a chain setup as shown in Figure 6-1. However, we reduce the number of nodes in the chain to 6 for this experiment. The data packet is sent from node 1 to node 6. The end-to-end throughput is recorded in Figure 6-3. The packet inter-arrival time at the source node, which is labeled as node 1 in Figure 6-1, is varied from 1 to 0.0001 seconds. The throughput reaches a peak of 213.1 kbps when the packet inter-arrival time at node 1 is 0.057 seconds.

Figure 6-4. Number of received and dropped DATA frame for chain setup (packet inter-arrival time = 0.005 seconds)

Figure 6-5. Number of received and dropped DATA frame for chain setup (packet inter-arrival time = 0.057 seconds)

This optimal offered load depends on the data frame size and the simulated value of 0.057 seconds packet inter-arrival time can also be proven analytically by considering the timing sequence of the MAC protocol, the average backoff period when MAC layer congestion window is minimal and the timing sequence of the MAC frames. In Figure 6-3, when the optimal offered load is used, the highest throughput at packet inter-

arrival time of 0.057 seconds was also shown for a chain setup as reported by [55]. However, when the packet arrival is faster than this value, the end-to-end throughput saturates at 86.94kbps. In this scenario, when the packet inter-arrival time drops below 0.057 seconds, congestion will persist. Figure 6-4 and Figure 6-5 show the number of data frames received and dropped at various nodes along the chain for two packet inter-arrival times; 0.005 seconds and 0.057 seconds, respectively. In the first case (Figure 6-4), there are many dropped data packets at nodes 3 and 2 due to buffer overflow. We also notice that data packets are being dropped at node 1 and node 2 due to MAC layer maximum transmission retry limit. This phenomenon is attributed to the critically exposed interference and false NAV indication as discussed in Sections 6.2.1 and 6.2.3 above. The congestion phenomenon at nodes 2 and 3, which causes dropped packets from the buffer when the offered load exceeds a certain threshold confirms our observation, which is discussed earlier in Section 6.2.2 .

Figure 6-6. Transmission cost for chain setup

In Figure 6-5, there are no dropped DATA frames since the packets are scheduled by using an optimal packet inter-arrival time (0.057 seconds). In such a scenario, the

deviation of the end-to-end throughput, which is 213.1 kbps, from the maximum link throughput (1 Mbps) is attributed to the exposed node effect and other overheads.

Figure 6-6 shows the transmission cost of the plain 802.11 MAC protocol when the chain setup is used. When the packet arrival is slower than the critical value of 0.057 seconds, the transmission cost is optimal because only a single RTS frame is consumed for every DATA frame. Under such cases, the transmission cost is recorded at 5.57. When the packet arrival is faster than the critical value, the transmission cost starts to increase due to a variety of reasons discussed in Section 6.2 above. Under the heavy offered load cases, the difference in transmission cost can be as high as 210% from the optimal case.

Figure 6-7. Percentage of various receiver states in 6-node chain setup when packet inter-arrival at node 1 is 0.005 sec.

Figure 6-7 shows the overall receiver responses when an RTS frame is transmitted by one of the nodes in the 6-node chain setup. The packet inter-arrival time at node 1 is set to 0.005 seconds to simulate the saturated case. The graph show that for 55.08% of the RTS frames sent to a receiver, the receiver sees an idle channel and responds with a

CTS frame. A significantly high percentage (23.02%) of the RTS frames sent to the receivers are critically exposed and the receivers will not respond with CTS frames. False NAV and frozen MAC states amount to about 13.18% and 7.46% of the failed CTS responses, respectively. The remaining problems, which are insignificant compared to the other problems, are due to direct collision when the receiver transmits at the same instant as the sender (1.16%) and two transmission range signals colliding at the receiver (0.10%).

Figure 6-8. Grid setup and high fan-in chain setup

Grid Size	Distance between rows or columns	Transmission cost (ψ)	Aggregate Throughput (kbps)
6 by 6	200 m	11.35	253.536
11 by 11	100 m	13.11	219.184
21 by 21	50 m	15.56	182.4

Table 6-1. Performance results for Grid setup

The above experiment only reveals the throughput and energy performance under a single end-to-end flow scenario with low node density. To demonstrate the throughput

performance under multiple flow scenarios, we simulated the original 802.11 MAC in a grid topology depicted in Figure 6-8a where node density and channel contention are higher. The size of the grid is fixed at 1000 m by 1000 m, but we vary the number of nodes and distance between the rows and columns to increase the density of nodes. The flows are transmitted from the nodes in the upper row to the lowest row and nodes from the first column to the last column. In all three grid sizes shown in Table 6-1, the number of hops to reach the destination is 5 in all end-to-end flows. In addition to the grid setup, we simulated a high fan-in scenario as shown in Figure 6-8b, where many single-hop links fan-in into a chain of nodes. The latter scenario depicts a possible scenario in sensor networks, whereby an event is sensed by several nodes at the same instance and all nodes try to report back to a sink.. In both setups, we set the packet inter-arrival time of the source nodes to 0.005 seconds to simulate the condition where nodes transmit as fast as possible.

No of Sources	Transmission cost (ψ)	Aggregate Throughput (kbps)
1	11.48	86.94
10	31.21	31.31
20	31.75	30.70
40	30.52	31.92

Table 6-2. Performance results for High Fan-in chain setup

The results obtained in Table 6-1 are as expected since throughput and energy efficiency degrade as the node density and contention increase. The transmission cost is still high considering that each flow needs to travel only 5 hops to reach the destination. Table 6-2 records the performance evaluation of the topology shown in Figure 6-8b. The end-to-end throughput drops drastically to around 31 kbps when the number of flows going into the chain of nodes increases. The transmission cost also degrades by

548% compared to the ideal case ($\psi = 5.57$) when the number of flows is increased from 1 to 10. The experiments above demonstrate that under heavy traffic conditions and high fan-in situations, the throughput of wireless multihop networks based on CSMA/CA MAC protocol can easily degrade.

In most wireless networks, the MAC protocols and flow control schemes operate independently of each other. There have been very few research efforts on the study of an effective cross layer flow control design that increases the throughput of wireless multihop networks [103]. The MAC layer is normally isolated from the upper layers and an event such as buffer overflow is transparent to the MAC layer. Through the experiments above, we have examined some inherent MAC protocol flaws that cause unnecessary packet drops. In the later sections, we propose enhancements to the MAC protocol and introduce a novel hop-by-hop flow control scheme to improve the performance of the CSMA/CA based multihop wireless network.

6.4 A flow control framework

Figure 6-9. Flow control framework

The main objective of this flow control framework is to maximize throughput and energy efficiency of the CSMA/CA based multihop wireless network. The primary goal of the techniques that we introduce in this chapter is tuned towards improving the throughput in multihop wireless networks. Since the framework is targeted for wireless ad hoc and sensor networks, we adopt the techniques and algorithms that are efficient in terms of computations and transmissions, and thereby energy efficient. The overall flow control framework is shown in Figure 6-9 and consists of two major components; the enhanced CSMA/CA MAC and the hop-by-hop flow control. The two major components are explained in detail in the following sections.

We adopt a per-flow packet scheduler that tries to achieve reasonable fairness but with low processing complexity. We use the Distributed Round Robin (DRR) scheme to implement the packet scheduling scheme because it has $O(1)$ time packet processing complexity when compared to the other per-flow timestamp schedulers such as SCFQ, WF2Q, etc.

6.5 Enhanced CSMA/CA MAC protocol

The CSMA/CA MAC protocol block as shown in Figure 6-9, is modified such that the MAC layer is aware of flow control and buffer states. Unlike the traditional stack, our MAC layer is designed to act as an admission control point to prevent eventual packet drop due to buffer overflow. In this section, we discuss the MAC enhancements that are introduced to improve the throughput and rectify the MAC inefficiencies highlighted in Section 6.2 . Specifically, we discuss how the false NAV problem should be handled and discuss the additional modification to the CSMA/CA 4-way handshake to allow interoperation of the MAC protocol with the flow control scheme.

6.5.1 Solving the false NAV problem

To solve the false NAV problem, we recommend that nodes apply a selective reset on the received NAV. When a node receives a NAV from certain nodes, it is expected to physically sense some activity in the channel because it is likely to be within the transmission range of either one or both of the communicating sender-receiver nodes. To solve the false NAV problem, we propose to modify the DCF protocol such that nodes that are blocked with a NAV should still try to receive transmission range packet using the normal packet capture procedure [36]. If a valid RTS frame is received correctly, the node should cancel its NAV and proceed with the normal 4-way handshake. This operation is valid since the node would not have received a valid RTS frame if there was a valid 4-way handshake ongoing in the neighborhood of the node that received the NAV. We do not, however recommend that all nodes cancel their NAV by listening to the channel because resetting NAV in all nodes can result in increased interference.

6.5.2 Modifications to the 4-way handshake

To ensure interoperability between the MAC layer and the flow control module, we propose that the standard 4-way handshake in CSMA/CA is modified to include an RTS-NAK frame. In addition, the RTS frame is modified to carry a flow ID field, which is explained in Section 6.6.1. The modified 4-way handshake operates as follows. When a sending node sends an RTS frame, the receiver has the option to send an RTS-NAK frame or a CTS frame. Upon receiving the RTS, the MAC layer consults the flow control module, to check if the packet can be admitted into the node. The flow control examines the flow ID to check if a packet with the same ID exists at the receiver. If the flow violates the buffer occupancy rule, the MAC layer will reply with

an RTS-NAK frame to the sending node. The buffer occupancy rule is violated when a packet that belongs to the same flow exists in the receiving node or the total buffer limit has been reached. Accordingly, there are two different RTS-NAK types to represent these two events. The RTS-NAK frame is designed as small as possible to reduce the overhead incurred by this frame. 2 bits are used to represent the different RTS-NAK types. Another 14 bits are used to carry the flow rate information, which is explained in Section 6.6.4 . Using the same IEEE 802.11 standard, the RTS-NAK message is the same size as the CTS frame. If the flow control allows the data packet to be admitted, the 4-way handshake proceeds as per-normal.

6.6 Hop-by-hop Hybrid window and rate based flow control

In the previous section, we described the enhancement to the standard 4-way handshake with additional fields and an RTS-NAK frame for the purpose of compactly carrying flow control information. In this section we discuss the hop-by-hop flow control scheme. We refer the readers to [48] for a complete taxonomy on flow and congestion control techniques. Similar to the definition used in [48], we claim that flow control is used here as a means to solve the congestion problem. Due to scalability and ease of deployment, end-to-end flow control schemes such as TCP have dominated the wired networks. Hop-by-hop flow control is normally favored over the end-to-end control scheme since it provides faster feedback response to congestion. However, the benefits of hop-by-hop schemes normally come at the expense of additional explicit messages to notify the upstream nodes about the congestion. In a wireless multihop network, if special control messages are sent over multiple hops in the reverse direction to control the rate, it can result in self-contention, which results in a sluggish system.

Hop-by-hop flow control schemes have not been popular in wired networks because of its requirement to maintain per-flow information at each node. This is understandable since the number of flows passing through a core router in the Internet is enormous, making the management of per-flow information infeasible [40]. Unsurprisingly, end-to-end flow control schemes have dominated the Internet by far. In a multihop wireless network, the number of flows is presumably more manageable if we implement some form of admission control at the nodes to restrict the number of flows passing through a node. A case for hop-by-hop flow control for wireless multihop networks has been justified in [102]. Furthermore, depending on how a flow is classified, we can further reduce the complexity of per-flow information management. We present a lightweight flow ID management scheme that is beneficial to a computationally and energy challenged network such as wireless ad hoc and sensor networks.

6.6.1 Flow ID representation and management

To reap the benefits of the hop-by-hop flow control and per-flow packet scheduling schemes, the process of flow identification is compulsory. However, we need an efficient method, which incurs low communication and memory overheads to represent flows. In a typical IP system, the flow is defined in an end-to-end manner and the address information that identifies the flow is normally embedded in the data packet. Hence, the entire MAC layer 4-way handshake has to be completed before the system can identify the flow. In the standard network stack without cross-layer consideration, allowing the data packet to arrive at a node before inferring if a flow has violated the buffer occupancy limit, and finally discarding the data packet is wasteful in terms of channel resources. For this reason, we need an efficient way to notify the receiver of the impending data packet before actually sending the large data packet. The IP

addresses and port numbers, which are used to define the end-to-end flow, uses a total of 12 bytes. Encoding the 12 bytes into the MAC layer frames is one option but this makes the MAC frame size too large. Hence, we need a more compact flow ID representation scheme.

There are a few approaches that can be used. One could design a special field to maintain a globally unique flow ID but scalability of such a technique is difficult to achieve in a multihop wireless network. The other option is to design a flow ID scheme that is unique on a per-hop and flow direction basis. This scheme is much more attractive than the globally unique flow ID scheme because the same ID can be reused by the same node on other links. However, because the total number of flows, N , that can pass through a node in an interval can be much greater than the total buffer size, R , we need to maintain a reasonable amount of flow ID records. In addition, we require a flow ID management protocol that operates between any two nodes to assign a new ID or replace old per-hop flow IDs when a node runs out of IDs. To cater for a system that prevents frequent changes in per-hop flow IDs, we then require at least $\log_2 N$ bits to represent each flow and $O(N)$ memory records for each per-hop link.

To reduce the memory storage problem attributed to flow ID records and overcome the flow ID management problem, we propose an innovative method based on hashing to represent flow IDs. In our scheme, flow IDs are represented on a per-hop basis. The IDs are generated by using a standard hashing function [39] that takes the following triplets as inputs; the unique end-to-end source and destination addresses of the flow, and per-hop source MAC address. The per-hop MAC address is used as one of the inputs for the purpose of reducing hash code conflicts due to multi-path flows originating and terminating from the same source and destination but passing through

the same intermediate node. This simple but novel method obviates the need to have an additional flow ID assignment protocol.

Conflicts in the hashed codes can occur in our system if the hash code size is small. In our scheme, we maintain a table of the triplets mentioned above only for the packets that are present in the buffer of a node at any given instant. As such, the memory requirements for storage of per-flow ID is then reduced to $O(R)$, irrespective of the number of flows passing through a node.

In our experiments, we reserve a 10-bit field for the purpose of encoding the flow ID. The number of bits is a design parameter that has to take into consideration the statistical probability of hashing code conflicts at any given node in the system. The flow ID, which is compact, can then be encoded into the RTS frame for easy identification of the flows during the RTS-CTS phase. The benefit of this scheme is that the sender of the RTS frame is able to query the receiver's buffer occupancy state fast by using the same MAC layer collision avoidance frames. The receiver can then positively or negatively acknowledge the buffer occupancy state for the impending flow with a CTS or an RTS-NAK frame, respectively.

6.6.2 Bandwidth-Delay Product and flow control design

We propose a new hybrid hop-by-hop flow control scheme, which utilizes a combination of static window and rate control to regulate the flow of packets. The rate control is used on a per-hop basis rather than just at the originating node. The rationale of such a scheme will be further explained in this subsection.

Before we discuss the scheme and the parameter selection in detail, we study the *bandwidth-delay product*, BDP [48], of the typical wireless multihop network. The

bandwidth-delay product is an important parameter in flow control theory as it determines the optimal amount of packets that can be transmitted between endpoints in an uninterrupted manner. Most window based flow control schemes such as TCP’s AIMD [36] and traditional sliding window protocol in effect use the BDP to calculate the optimal window size, w , for transmission between endpoints. Assuming the bottleneck service rate of a node in the path is μ packets/seconds, and round-trip time is measured as RTT seconds, the *bandwidth delay product* is simply given by

$$BDP = RTT \times \mu$$

(Eq. 6-1)

If the flow control scheme is not controlled properly and w exceeds the BDP, then the number of packets buffered at a bottleneck node becomes $(w - RTT \cdot \mu)$. To reduce congestion and buffer build-up, ideally, we want to adjust w such that $w = RTT \cdot \mu$.

Figure 6-10. (a) Wired 8-hop link. (b) Spiral wireless multihop links

BDP in wired and wireless network has many fundamental differences. In a wired network, each intermediate node consists of more that one interface. Whereas, in a wireless multihop network, each node usually uses a single half-duplex transceiver. While the wired nodes can transmit in parallel using adjacent links, wireless nodes have

to contend spatially to transmit a single frame. Figure 6-10a illustrates a typical chain setup using wired networks whereas Figure 6-1 represents the chain setup for a wireless network. In Figure 6-10a, packet transmission can occur simultaneously in each wired link when packets are forwarded from node 1 to node 9. Assuming negligible acknowledgment packet transmission time and a fixed data packet size, the optimal window size can be worked out as 8 packets due to the 8 hops present in the topology. Whereas, in the wireless multihop network as shown in Figure 6-1, parallel transmissions is a function of the transmission and carrier sensing ranges of the MAC protocol. Denoting transmission range as d_{tx} and carrier sense range as d_{cs} , by geometry we can easily show that the maximum effective throughput, S_{chain_max} , of a straight chain is upper bounded by the following expression:

$$S_{chain_max} = \frac{\alpha \cdot S_{link_max}}{\left\lceil \frac{d_{cs}}{d_{tx}} \right\rceil + 1} \quad (\text{Eq. 6-2})$$

where α is simply a factor to account for the fixed overheads consumed by RTS, CTS and ACK frames, and other headers. S_{link_max} represents the maximum link capacity and the ceiling function, $\lceil \rceil$ rounds up the ratio between the carrier sense and transmission range. In a 1 Mbps IEEE 802.11b system, $S_{link_max} = 1$ Mbps and the ratio of d_{cs}/d_{tx} is typically 2.2 [5]. In a typical 11 Mbps system, the ratio of d_{cs}/d_{tx} is about 4.4 [5]. Using the typical 1 Mbps IEEE 802.11 ranges, it is shown in [55] that the optimal end-to-end throughput of the chain is governed by a factor of $1/4$. We simply refer to the factor, $\lceil d_{cs}/d_{tx} \rceil + 1$ from equation (Eq. 6-2) as the link reuse factor Q . Assuming a fixed packet size and ideal transmission scheduling, then it is easy to see that every link in a chain setup should be scheduled every Q slots, where each slot interval, T_{slot} is defined by the minimum 4-way handshake interval given as:

$$T_{slot} = T_{RTS} + T_{CTS} + T_{DATA} + T_{ACK} + 3sifs + difs + ave_cw \quad (\text{Eq. 6-3})$$

where the ave_cw parameter describe the average binary exponential backoff window slots used in the first stage of the binary exponential backoff process and T_{RTS} , T_{CTS} , etc. describes the time taken to transmit the RTS, CTS, etc. frames, respectively. In other words, in a chain setup using a 1 Mbps system with $d_{cs}/d_{tx}=2.2$, parallel transmissions can only be carried out at every 4 hops away assuming that each transmission is carried out at maximum transmission distance and the links are in a straight line as shown in Figure 6-1. However, if the links are much shorter and have arbitrary direction as shown in Figure 6-10b, then the effective throughput, S_{chain} , of the end-to-end chain will drop below that defined in (Eq. 6-2).

Unlike the wired case, we can show that by using equations (Eq. 6-1) and (Eq. 6-2), the maximum window size, w , necessary for optimal flow control in a multihop transmission is always lower than the number of hops, H , when $H \geq 2$. Assuming a negligible acknowledgement packet transmission time and wireless propagation delay, the upper bound on the window size, w_{u_bound} , for a particular flow that traverses H hops, where $H \geq Q$ can be approximated by

$$w_{u_bound} = \frac{H_{count_opt}}{\left\lceil \frac{d_{cs}}{d_{tx}} \right\rceil + 1} \quad (\text{Eq. 6-4})$$

where H_{count_opt} is the hop count for the end-to-end flow assuming the chain is in a straight line and distance between nodes are at maximum transmission distance. This result gives us some clue on how the per-hop static window and rate control should be implemented at each intermediate node. Intuitively, by observing equation (Eq. 6-4), we can conclude that it is unnecessary to allocate a total buffer space for a particular

flow along the chain of nodes where this flow passes through, which exceeds w_{u_bound} . These packets will anyway cause unnecessary backlog in the intermediate nodes and prevent other nodes from accessing the network. In addition, the additional packets originating from the same flow will result in self-contention during channel access.

6.6.3 Static hop-by-hop window control

The basic hop-by-hop flow control scheme that we adopt in this framework resembles a simple static window protocol of typical first generation flow control scheme [48]. We, however, reuse the CSMA/CA MAC protocol's collision avoidance frames (RTS, CTS), RTS-NAK and ACK frame as a means to implement the messaging protocol for the static window. In a multihop wireless network, it is difficult to achieve the maximum chain throughput, S_{chain_max} , since there will be other contending nodes that effectively reduce the throughput of the end-to-end link. In addition, a packet that experiences frequent blocking will have a longer average transmission period than that given in equation (Eq. 6-3) to send a packet successfully, due to multiple RTS retries and long binary exponential backoff periods.

Tuning the nodes to use the exact window size requires constant measurement and feedback of the round trip time or flow rate, which is too resource consuming for a wireless ad hoc or sensor network. For this reason and to keep the design simple, we simply use the upper bound of the window size given by equation (Eq. 6-4) to design the flow control scheme.

To simplify analysis and design of the flow control scheme, we assume that nodes transmit equal length packets and the queue length at each node is measured in terms of packets. The design and parameters can be easily converted for a variable packet length

system and a queue length measured in bytes. Averaging w_{u_bound} over the intermediate nodes, we note that each node should only hold a fraction of the packet. However, since the queue size is measured in packets, it is impractical to store a fraction of the packet. Therefore, we limit the static window size for each flow at each intermediate node to a single packet. We then rely on the rate control part of this hybrid hop-by-hop flow control scheme to spatially spread the packets. Since the per-hop queue used to serve a flow is limited to a fixed window size of one unit, we try to spread the packet distributed on the intermediate nodes such that they are placed at every Q^{th} link. By controlling the rate control part properly, we can effectively limit the equivalent flow control window size of a flow to w_{u_bound} along the links. The basic hop-by-hop flow control scheme with the static single-unit window alone will simply be referred to as the “Static H-b-H window” scheme.

6.6.4 Rate control

To realize the spreading of the packets or the rate control scheme, we implement a delay mechanism at each node. This rate control scheme cannot operate independently and needs to be supplemented with the basic static window scheme discussed in Section 6.6.3 above. This is necessary because the rate control scheme requires some form of binary feedback on congestion from the hop-by-hop static window scheme. The operation of the rate control scheme is captured in the algorithm described in Figure 6-11 to Figure 6-13, which is expressed using a C-style pseudo code. For convenience, the notation and terminology used to describe the algorithm are summarized in Table 6-3 below.

Unlike previous hop-by-hop flow control schemes, our scheme does not necessitate the storage of all per-flow information for packets that transit at each intermediate node.

However, we maintain flow information for packets that are currently residing at a node and for those that have exited a node, for a limited period. After this limited period, the flow record is flushed from the intermediate node's memory. However, like most of the rate control schemes, the historical flow rate at each node is crucial for optimally controlling the flow and reducing congestion. To retain the rate information of a flow and at the same time reduce the requirement of per-flow rate information storage at each node, we encode the flow rate on a given link, l , into the data packet that is forwarded to the next hop. In this way, when an upstream node tries to admit a packet originating from the same flow on link l and a packet still exist in the downstream node's buffer, the last used flow rate information for link l can be obtained from the downstream packet. The rate value is then encoded in the RTS-NAK frame and passed back to the upstream node. Since the rate information is coded in multiples of T_{slots} , the number of bits required to represent the rate can be compactly encoded in the RTS-NAK frame. The rate information can then be used to schedule the transmission of the data packet from the upstream node to the downstream node.

d	Destination node for flow f
s	Source node for flow f
Pkt_f^m	The m^{th} packet for flow f
$hops(i,j)$	Number of hops from node i to node j
Timer(T)	Timer function with count down period of T seconds
Q	System wide link reuse factor (as explained in Section 6.6.2)
T_{slot}	Average 4-way handshake period to transmit a data packet as given by equation (3)
$f.\text{record}$	Flow record for flow f
$f.T_{prev_delay}^{m-1}$	Previous delay time used by the $(m-1)^{\text{th}}$ packet of flow f . This is the time that must elapse for the m^{th} packet of flow f before it is scheduled for transmission after the $(m-1)^{\text{th}}$ packet has exited the node.
$Pkt_f^m.\text{delay}$	Current delay assigned to the m^{th} packet of flow f at a certain node in multiples of T_{slot} .
$Pkt_f^m.\text{base.delay}$	Base delay assigned to the m^{th} packet of flow f based on the remainder hops to destination
Schedule(Pkt, T)	A function to schedule the transmission of a packet, Pkt , after a delay of T seconds.

Table 6-3. Notations

```
Receive( $Pkt_f^m$ )      at node  $i$ 
{
1      /* initialize basic packet delay */
2      if ( $f.\text{record} \equiv \text{False}$ ) {
3          if ( $hops(i,d) \geq Q$ )
4               $Pkt_f^m.\text{base.delay} \leftarrow (Q-1) \times T_{slot}$ ;
5          else
6               $Pkt_f^m.\text{base.delay} \leftarrow (hops(i,d)-1) \times T_{slot}$ ;
7      }
8      if ( $f.\text{record} \equiv \text{False}$ ) {
9          Encode  $Pkt_f^m$  with  $Pkt_f^m.\text{base.delay}$ 
10          $Pkt_f^m.\text{delay} \leftarrow Pkt_f^m.\text{base.delay}$ ;
11         Schedule( $Pkt_f^m, 0$ );
12     }
13     else if ( $f.\text{record} \equiv \text{True}$  and Timer( $f.T_{prev\_delay}^{m-1}$ ) not expired) {
14         Encode  $Pkt_f^m$  with  $f.T_{prev\_delay}^{m-1}$ 
15          $Pkt_f^m.\text{delay} \leftarrow f.T_{prev\_delay}^{m-1}$ ;
16         Schedule( $Pkt_f^m$ , remainder of Timer( $f.T_{prev\_delay}^{m-1}$ ));
17     }
} /* End of Receive */
```

Figure 6-11. Receive function

Figure 6-11 describes the receiving function of the algorithm. A node, which receives a transit packet, first checks if a flow record for flow f , $f.record$, exists. The $f.record$ basically stores the time that must elapse, $f.T_{prev_delay}^{m-1}$, in between the transmission of packets $(m-1)$ and m of flow f , along a link l . The $f.record$ only exists at a node for a limited period given by $f.T_{prev_delay}^{m-1}$, after the $(m-1)^{th}$ packet has left the node. If a record does not exist, it then proceeds to schedule the packet for transmission immediately. Before transmitting, the delay in multiples of T_{slot} that corresponds to the rate in which flow f passes through link l is encoded into the data packet. In lines 3 to 6, the number of hops to the destination is checked to compute the base delay that should be used. The delaying technique that we design gives the receiver node and other contending nodes along the downstream path ample time to forward packets that belong to the same flow. Considering this, it is intuitive to schedule the packets for every Q^{th} slot when there is Q number of links ahead. However, when the remaining hop count from node i to destination node, d , given by $hops(i, d)$ is lower than Q , or an end-to-end flow has fewer than Q hops, then we need to regulate the rate control such that these flows are not unnecessarily penalized. For this reason, we apply the following delay rule at every intermediate node by considering the remainder hop-count to destination.

$$base.delay = \begin{cases} (Q-1) \cdot T_{slot} & \text{if } hops(i, d) \geq Q \\ (hops(i, d)-1) \cdot T_{slot} & \text{if } hops(i, d) < Q \end{cases} \quad (\text{Eq. 6-5})$$

In (Eq. 6-5), $base.delay$ denotes the base delay to be imposed for a flow at any given node. In Figure 6-9, the number of hops to destination is obtained from the routing layer. In lines 13 to 16 of Figure 6-11, the node waits for the timer, $Timer(f.T_{prev_delay}^{m-1})$, to expire if $f.record$ exists before scheduling the packet for transmission. If $f.record$

exists, the current delay period, which is stored in $f.record$, is encoded into the data packet before transmission.

The Schedule function in Figure 6-11 invokes the Transmit function shown in Figure 6-12. When an upstream node receives an RTS-NAK frame from a downstream node indicating that a packet of the same flow ID is still present at the downstream node, the upstream node will extract the delay information encoded in the RTS-NAK. This is described in line 2 of Figure 6-12. The upstream node will then select the highest delay value between the last used delay, $Pkt_f^m.delay$, and the delay information encoded in the RTS-NAK and additively increases the delay interval by a single T_{slot} . Notice that the last used delay value can be higher than the delay value encoded in the RTS-NAK due to multiple RTS-NAK replies for the same data packet. The upstream node will then restart the delay timer for transmitting the packet again.

If an upstream node receives an ACK frame as a result of successfully transmitting the m^{th} data packet, the upstream node will then setup the delay timer for the $(m+1)^{th}$ data packet. Lines 9 to 14 of Figure 6-12 describe this procedure. The upstream node will decrement the last used delay by T_{slot} if the delay period used is greater than the base delay given by equation (Eq. 6-5) and store it in $f.record$. Else, the upstream node will simply use the base delay given by equation (Eq. 6-5) and store this in $f.record$. The node will then set a timer based on the selected delay for the same flow f . Note that immediately after receiving the ACK, the upstream node will not have a packet with the same flow ID because the window size is a single unit but any packet that is received eventually will have to adhere to this delay. If the upstream node does not receive a packet of the same flow ID within this delay period, then this flow information is removed from the memory. This operation is described in Figure 6-13.

This operation reduces the memory storage complexity associated with storage of per-flow information. The reason for decrementing the delay by a single slot in line 11 of Figure 6-12 is to test if the sending rate can be increased. Line 11 also limits the rate to the achievable upper bound derived from equations (Eq. 6-2) and (Eq. 6-4).

```

Transmit( $Pkt_f^m$ )    at node  $i$ 
{
1      if (received an RTS-NAK) {
2          Extract  $Pkt_f^{m-1}.delay$  from RTS-NAK frame
3          if ( $Pkt_f^m.delay > Pkt_f^{m-1}.delay$ )
4               $Pkt_f^m.delay \leftarrow Pkt_f^m.delay + T_{slot}$ ;
5          else
6               $Pkt_f^m.delay \leftarrow Pkt_f^{m-1}.delay + T_{slot}$ ;
7          Encode  $Pkt_f^m$  with  $Pkt_f^m.delay$ 
8          Schedule( $Pkt_f^m$ , Timer( $Pkt_f^m.delay$ ));
9      }
10     else if (received an ACK) {
11          $f.T_{prev\_delay}^m \leftarrow Pkt_f^m.delay$ ;
12          $f.T_{prev\_delay}^m = \max[f.T_{prev\_delay}^m - T_{slot}, Pkt_f^m.base.delay]$ ;
13          $f.record \leftarrow \text{Store}(f.T_{prev\_delay}^m)$ ;
14         Start Timer( $f.T_{prev\_delay}^m$ );
15     }
16 } /* End of Transmit */

```

Figure 6-12. Transmit function

```

Cleanup () at node  $i$ 
{
1      if(Timer( $f.T_{prev\_delay}^m$ ) = 0 and  $Pkt_f^{m+1} \equiv 0$ )
2           $f.record = \text{False}$ ;
3  } /* End of Transmit */

```

Figure 6-13. Cleanup function

In this work, we have not suggested any special method to solve the critically exposed node problem; therefore some flows will still experience excessive blocking under normal circumstances due to unfavorable location of nodes. Some of the packets will also be dropped after experiencing the maximum retry limit. The simple additive delay increase method mentioned above will regulate the flows passing through such links

but it does not penalize them. As a result of this, the flow control framework proposed in this chapter still experiences dropped packet due to critically exposed interference. To reduce packet loss due to such interference, one could possibly modify the packet scheduler described in Section 6.4 to a weighted scheme that penalizes such flows.

In contrast to our per-hop rate control scheme, one could design a source node rate control mechanism [12] to regulate the flows in the same manner as described above but a rate control mechanism on a per-hop basis is more beneficial than a single rate control mechanism at the source. The reason for this is because a source rate control can only regulate the flow at the starting point of the flow. Contention and bottlenecks in the downstream nodes will quickly cause the packets to be clumped together at each node along the downstream path, further resulting in self-contention and reduced throughput.

6.7 Simulation Results

To demonstrate the effectiveness of our approach, we used NS-2 simulator to obtain simulation results. We modified the IEEE 802.11 DCF protocol to include the enhancements to the 4-way handshake and the solution to the false NAV problem. In addition, we developed additional clear channel assessment modes (CCA) to handle the proper carrier sensing states, which is absent in NS-2. All the performance results obtained in this chapter are based on CCA-2 mode [36]. We also implemented our flow control scheme and the MAC layer enhancements in a modular fashion such that different modules can be combined in different combinations to form different strategies. The various modular schemes that form the various strategies are summarized in Table 6-4. In addition, the metrics that were described in Section 6.3 are

used to evaluate the performance of the various strategies. All simulations were carried out by using a fixed data packet size of 1500 bytes.

Schemes	Remarks
NAV repair	The selective NAV repair scheme proposed in Section 6.5.1
Rate Control	The hop-by-hop rate control as described in Section 6.6.4
Static H-b-H Window	A static per-flow window of size 1 at each node
No Congestion Control or Original 802.11 MAC	The original stack with plain 802.11 DCF and FIFO queuing

Table 6-4. Summary of schemes

Figure 6-14. Directed 5-chain links fan-in

Figure 6-15. Throughput performance – chain setup

We evaluate the various strategies by using five distinct topologies. The first is a simple chain setup as shown in Figure 6-1, but with a single flow originating from node 1 and terminating at node 6. This simple setup has also been shown in Section 6.3 to suffer from congestion. In the second setup, we examine a directed 5-chain links fan-in situation as shown in Figure 6-14, where several flows traveling along different chain of nodes for some distance converge into a single chain of nodes. In Figure 6-14, the nodes denoted as S are the sources whereas the node denoted as D is the destination node. In the third setup, we show the performance of the high fan-in single-hop links that converge into a single chain as shown in Figure 6-8b. In the fourth setup, we evaluate the performance of our scheme under the grid setup with multiple cross flows, which are shown in Figure 6-8a and described in Section 6.3.2. Finally, we simulate a mobile scenario where fifty (50) nodes are allowed to move randomly in a 1500 m by 1500 m square area. Figure 6-15 shows the throughput performance of the various flow control strategies when the single-flow chain setup is used. When packet arrival is slow, the behavior of all the schemes is the same. However, when packet arrival is faster, there are significant differences in the throughput performances. Figure 6-15 shows that the combination of the static window and rate control is sufficient to ensure that the throughput is optimal and tallies with equation (Eq. 6-2). In this particular setup, using the rate control and static window negates the effect of false NAV; therefore the addition of NAV repair does not show any improvement. When all three schemes are used in combination, the throughput exceeds the “No Congestion Control” scheme by a factor of 2.47. When the static window is used alone, the throughput performance is increased by a factor of 1.77. Additional improvement to throughput is noticed when NAV repair is used with the hop-by-hop static window alone.

Figure 6-16. Transmission cost – chain setup

Figure 6-16 shows the transmission cost obtained by using the chain setup. The use of the three schemes in combination keeps the energy efficiency to the optimal value of 5.57. This corresponds to the transmission that occurs along 5 hops of the chain setup. In the “No Congestion Control” scheme, the transmission cost can exceed the optimal value by a factor of 2.4. This shows that significant amount of energy can be wasted when the flow is pumped into the chain as fast as possible and no congestion control is used. Figure 6-16 also shows that, significant saving of energy is possible even if the plain static window is used alone.

Figure 6-17. Throughput performance – directed 5 chain links fan-in

Figure 6-18. Transmission cost – directed 5-chain links fan-in

The throughput performance and transmission cost of the directed 5 chain links fan-in setup (as shown in Figure 6-14) is shown in Figure 6-17 and Figure 6-18. It is interesting to note that, in this setup, when the NAV repair is used with the plain 802.11 MAC, the throughput performance during faster packet arrival actually degrades when compared to the original 802.11 MAC. This occurs because the NAV repair employed at a node actually has the effect of improving the MAC layer service time used for receiving incoming frames. NAV repair also reduces the number of packets that are

dropped at a sender due to maximum retry limit because receivers are not blocked due to false NAV. However, this improvement in receiving time is achieved at the expense of reducing the service time available for transmitting frames. Therefore, when NAV repair is used without flow control, more packets can be sent by the source but are later dropped due to buffer overflow at certain nodes. The simulation traces for the 5 chain links fan-in setup used in Figure 6-14 confirm this observation. During faster packet arrival, the total packet dropped due to buffer overflow and maximum retry limit for the original MAC with NAV repair only, exceeds the original MAC by 17.6%. However, the packet drop due to buffer overflow alone, for the original MAC with NAV repair, exceeds the original MAC by 30%. The dropped packets due to buffer overflow for this particular setup occur mainly at the 2nd and 3rd node of each end-to-end flow. As a result of reduced service time for transmitting, fewer packets are successfully transmitted to the final destination.

When the NAV repair is used in combination with rate control and static window, it actually contributes to throughput improvements. Consistent with the chain setup, throughput and energy improvement can be significantly improved by just using the static window alone. However, the combination of the NAV repair, static window and rate control yields the best performance.

Table 6-5 shows the performance improvement of the complete flow control scheme when the grid setup as shown in Figure 6-8a is used and when sources transmit as fast as possible. The performance improvements for throughput and transmission cost are not as high when compared to the other two setups described previously but are still significant.

Table 6-6 shows the improvement in terms of throughput and energy efficiency when the high fan-in chain setup as shown in Fig 8b is used. The results collected in Table 6-6 are based on sources transmitting as fast as possible. Using the complete flow control framework, tremendous improvement in throughput and energy efficiency can be obtained when the number of sources is high. When the number of sources in the high fan-in chain setup is 10, the energy efficiency and throughput improves by factors of 5.14 and 5.58, respectively, when the complete flow control framework is used.

	No Congestion Control		Static Window + NAV repair + Rate control	
Grid Size	ψ	Throughput (kbps)	ψ	Throughput (kbps)
6 by 6	11.35	253.536	6.61	400.97
11 by 11	13.11	219.184	7.94	324.67
21 by 21	15.56	182.4	9.28	270.56

Table 6-5. Performance of Flow Control Framework using Grid Setup

	No Congestion Control		Static Window + NAV repair + Rate control	
No. of Sources	ψ	Throughput (kbps)	ψ	Throughput (kbps)
1	11.48	86.94	5.57	214. 01
10	31.21	31.31	6.07	174.89
20	31.75	30.70	6.13	172.67
40	30.52	31.92	6.26	170.24

Table 6-6. Performance of Flow Control Framework using High Fan-in chain Setup

Figure 6-19. Aggregate throughput – mobile ad hoc network setup

The throughput performance of the complete flow control scheme over the plain 802.11 MAC for the mobile ad hoc network scenario is shown in Figure 6-19 above. In this scenario, 35 constant bit rate (CBR) connections are made from the senders to receivers from the start of the simulation. Each CBR connection has a rate of 40 packets per second to ensure that the network is congested. The simulation period is 200 seconds and the results in Figure 6-19 are obtained by averaging the simulation over 12 runs. A Random Waypoint Mobility [43] model with a fixed pause period of five seconds was used for the mobility scenario. In addition, the Distributed Round Robin (DRR) scheduling is used for both schemes instead of the standard FIFO queue with drop tail. Figure 6-19 shows that the complete flow control scheme outperforms the plain 802.11 scheme by an average of 72% for the aggregate throughput. The aggregate throughput drops for both schemes with increase in node speed because of the more frequent breaks in links.

6.8 Conclusion

In this chapter, we have carefully studied the CSMA/CA MAC protocol based on 802.11 DCF to identify some of the main problems that cause congestion and degradation in throughput and energy efficiency. We have categorically identified the major problems such as “False NAV” and “Frozen MAC State” and proposed a solution to the “False NAV” problem. This chapter also presented a unique hybrid hop-by-hop flow control scheme, which uses per-hop static window and rate control to regulate congestion in multihop wireless networks. We proposed a flow control framework by coupling the hop-by-hop hybrid flow control and the false NAV enhancement. The hop-by-hop flow control scheme leverages on the CSMA/CA based MAC protocol by utilizing the MAC frames used for collision avoidance. The flow control scheme itself is a zero loss scheme since buffer occupancy is always monitored and fed back to the upstream nodes using the MAC layer’s collision avoidance frames. Unlike the traditional hop-by-hop flow control scheme proposed previously which suffers from per-flow maintenance of all flows passing through a node; our scheme is designed to reduce per-flow maintenance. We introduced a novel technique for storage and sharing of flow rate information. We encode the rate information of a link used by a particular packet into the packet and passed it to the downstream node. This rate information can then be extracted for eventual use by the upstream node. In addition to this, we introduced a novel technique for flow ID representation based on hashing. This simple but novel method eliminates the need for a complex flow ID assignment protocol and a large flow ID field to represent flows. The flow control scheme proposed in this chapter has several benefits when compared to an end-to-end window scheme such as TCP or a pure rate control hop-by-hop scheme. While the number of packets pumped into the system by a source can be the same in all these schemes, we

used a per-hop rate control scheme in combination with a static single packet window that tries to spatially distribute the packets along the links to avoid self-contention among flows originating from the same source. In TCP, clumping of packets belonging to the same flow can occur, resulting in congestion and buffer overflow. Our flow control framework demonstrates that throughput and energy efficiency can be improved tremendously in simple topologies and topologies that demonstrate high congestion.

Chapter 7. Conclusions and Future Work

We conclude this thesis with a summary of our contribution and directions for our future work.

7.1 Summary

In a wireless ad hoc network, mobility and location dependent interference cause frequent breaks in communication between nodes. The simplicity of CSMA/CA MAC protocol, which relies on a simple distributed contention scheme, makes it an attractive candidate for use in such a harsh wireless environment. However, the throughput capacity of such a MAC protocol has been shown to be miserable when packets have to travel multiple hops before reaching their destination. In this thesis, we proposed several MAC layer and link layer enhancements that can alleviate some of the problems that are common to the CSMA/CA-based MAC protocol.

In Chapter 2, we proposed a new multi-channel MAC protocol that can increase the throughput capacity of nodes in a wireless ad hoc network. More importantly, we have shown that the scheme does not suffer from the control channel saturation that plagues the existing multi-channel MAC protocol. The scheme uses a simple but effective method to circumvent hidden node problem associated with channel switching. We have shown that the new multi-channel MAC protocol performs better when compared

to existing multi-channel protocols such as MMAC [82] and DCA [93] in end-to-end throughput capacity, end-to-end delay and channel scalability. However, the performance improvement comes at the expense of a larger number of transceivers required for a single node. This MAC protocol would be most suitable for communication nodes that have less concern for power consumption such as wireless ad hoc networks connecting ships, airplanes and mobile ground vehicles.

The multi-channel MAC protocol that we developed, allowed us to study the effect of channel assignment on a distributed contention MAC protocol such as CSMA/CA. With the use of the multi-channel MAC protocol, we studied one of the main problems; the critically exposed node problem, which severely degrades multihop performances of CSMA/CA-based MAC protocol. The critically exposed node problem arises due to the combination of uneven carrier sensing state in the receiver and the sender, and also the use of the binary exponential backoff scheme. It is difficult to avoid this problem unless the receiver employs other forms of signaling. In Chapter 3, we explored the use of channel assignment to circumvent the critically exposed node problem. We have shown through simulation that based on ideal channel assignment in a chain setup, we are able to achieve up to 6 times improvement in end-to-end throughput compared to a normal single channel system. Based on the fundamental physical layer constraints of the CSMA/CA-based radio communications systems, which consist of carrier sensing and receiving power threshold, we made several propositions to completely eliminate the critically exposed node problem. We then detailed an ideal channel assignment strategy and non-ideal heuristics based channel assignment protocol that can be used to assign the channel in a distributed manner.

In Chapter 4, we explored the method of concurrent transmission scheme to improve spatial reuse and therefore improve throughput capacity of the overall wireless ad hoc network. The only prior work carried out in this area was performed by Acharya *et al.* [2] in a MAC protocol known as MACA-P. However, our work differs significantly from theirs due to a few key reasons. In MACA-P, a sender-recipient pair suffers from unwanted synchronization of neighborhood transmission schedules, which prohibits the pair to transmit. In our protocol, we confine the concurrent transmissions to a 1-hop range of a primary sender. We avoid unnecessary locking to neighboring transmissions and are able to achieve higher throughput performances. The other major difference lies in the unique way that we invite secondary pairs to transmit concurrently. Using this MAC protocol, we are able to achieve 60% improvement in throughput over the standard CSMA/CA MAC protocol for a multihop setup.

In Chapter 5, we explored the use of multi-rate systems to improve the throughput performance in CSMA/CA-based MAC protocol. Most sophisticated radio systems today are designed with several modulation techniques to capitalize on favorable channel conditions due to favorable distances between a transmitter and a receiver. However, very little work has been done based on this multirate feature to improve the end-to-end throughput for a multihop network based on CSMA/CA MAC protocol. The significant work done in this area is by Awerbuch *et al.* [5] and is known as MTM. However, this work is primarily a routing layer metric that can be coupled with existing routing protocol and has the objective of selecting effective routes based on favorable link rates. We managed to design a new multi-rate forwarding scheme that is able to perform up to 49% better than the existing multi-rate scheme such as MTM [5]. Our scheme does not change the existing routing layer protocol and operates in the MAC layer. We also developed a simple analytical model that is used to compare the

performance of our multi-rate scheme against the existing MTM [5] scheme and carried out simulations to verify the MAC protocol performances.

In Chapter 6, we performed extensive studies to gather all the factors that affect the throughput in multihop transmission for a CSMA/CA-based MAC protocol. As a result of this, we were able to reveal several other problems such as congestion due to location dependency and uneven carrier sensing, false NAV indication and frozen MAC state in addition to the critically exposed node problem that we discussed in Chapter 3. We then focused on rectifying the false NAV problem and designing a new link layer flow control protocol to cater for the congestion problem. This design is presented in a cross-layer approach and has very minor overheads because most of the flow control information is piggybacked in the CSMA/CA 4-way handshake. Our flow control method is unique because it combines the traditional window and rate-based technique and is implemented in a hop-by-hop basis. The entire flow control framework prevents congestion, reduces packet loss and is attractive because per-flow information at each node is kept to a minimum. The flow control scheme is augmented into the link layer and can work independently or in combination with other transport layer protocols such as TCP. In general network layouts, the flow control framework, which consists of the false NAV enhancement and the hop-by-hop link layer flow control, improves the throughput capacity by up to 50% when compared to the plain CSMA/CA MAC protocol.

7.2 Future Work

There are several areas that hold promising directions for future work. Some of these are extensions of our existing work while others are motivated by some real problems that are slowly starting to emerge.

Load-based Channel Assignment: In this thesis, we have described the use of the multi-channel MAC protocol to mitigate the critically exposed node problem and develop a heuristic-based algorithm for channel assignment. The heuristic approach for channel assignment is not efficient when compared to the ideal centralized channel assignment. Further work is required to reduce the number of non-overlapping channel assignments introduced by the proposed heuristic-based distributed channel assignment algorithm. In addition to this, the algorithm can consider the load of the flows or clusters before carrying out the channel assignment. The channel-borrowing concept as used in the cellular networks can also be explored to further improve the efficiency of channel utilization.

Location-aided MAC protocols: In Chapter 4, we proposed a MAC protocol that uses location and concurrent transmissions to improve the throughput performance of the CSMA/CA-based MAC protocol. However, the location information inferred from signal strength has significant errors especially if the system is used in severe multipath environment. UWB-based radio technology is already beginning to appear in most research labs and industry. One example standardization body that is actively pursuing the use of UWB is the Multi-Band OFDM Alliance (MBOA), which is primarily developing standards for a Wireless Personal Area Network (WPAN). The MAC protocol, which is being developed under the 802.15.3a group, is a hybrid TDMA-CSMA/CA system. It is also designed for ad hoc configurations. With UWB systems, the localization is accurate to a few centimeters and is an integral part of the radio systems. It would be interesting to use the location information in a cross-layer architecture to improve the performance of this new hybrid MAC protocol.

Improving the Fairness of the proposed flow control scheme: In Chapter 6, we identified various problems that can plague the CSMA/CA MAC protocol and we proposed a flow control framework to solve most of the critical issues. The critically exposed node problem was, however, not solved using the flow control framework. It would be ideal to integrate either our method or another method that can reduce or totally eliminate the critically exposed node problem, into this framework. In addition to this, the fairness issues related to scheduling of flows were not considered in the design. As a result of this, the scheme might run into problems where certain flows could possibly inhibit other flows. If the fairness issues are resolved, the framework could possibly be explored as the basis of developing a new end-to-end QoS framework.

Performance of TCP over the proposed flow control scheme: In Chapter 6, we studied the performance of our link layer hop-by-hop flow control scheme without considering any transport layer flow control schemes such as TCP. Some legacy applications have been designed for use with transport layer flow control protocols such as TCP. While we claimed that our flow control is generic enough to support any type of flows, there is a possibility that legacy applications will still rely on the transport layer flow control protocols to regulate application layer traffic. In such a scenario, it would be beneficial to study the impact of these transport layer protocol on our hop-by-hop flow control scheme and vice-versa.

References

- [1] Abramson N, "The ALOHA System - Another Alternative for Computer Communications," in *Proc. of the Fall Joint Computer Conference*, pp. 281-285, 1970.
- [2] Acharya A, Mishra A, Bansal S. "MACA-P : A MAC for Concurrent Transmissions in Multi-Hop Wireless Networks", IBM Research Report RC22528. 2002.
- [3] Acharya A, Mishra A, Bansal S. "MACA-P: A MAC for Concurrent Transmissions in Multi-Hop Wireless Networks", in *IEEE PerCom*, March 2003.
- [4] Arnbak JC, "Capacity of slotted aloha in Rayleigh fading channels", *IEEE J. Select. Areas Communications*, vol. SAC-5, pp. 261-269, 1987.
- [5] Awerbuch B, Holmer D, Rubens H, "High Throughput Route Selection in Multi-rate Ad Hoc Wireless Networks", *IFIP WONS*, January 2004.
- [6] Balakrishnan H, Padmanabhan V, Katz RH, "The effects of asymmetry on TCP performance," in *Proc. of ACM/IEEE MOBICOM*, Budapest, Hungary, Sep 1997.
- [7] Bao L, Garcia-Luna-Aceves JJ, "A New Approach to Channel Access Scheduling for Ad Hoc Networks," in *Proc. ACM Seventh Annual International Conference on Mobile Computing and networking*, (Rome, Italy), July 2001.
- [8] Bertossi AA, Bonuccelli MA, "Code assignment for hidden terminal interference avoidance in multihop packet radio networks," in *IEEE INFOCOM '92*, volume 2, pages 701-9, Florence, Italy, 1992.

- [9] Bharghavan V, Demers A, Shenker S, Zhang L, "MACAW: A Media Access Protocol for Wireless LANs," in *Proc. of ACM SIGCOMM '94*, 1994.
- [10] Bianchi G, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535-547, March 2000.
- [11] Chandra A, Gummalla V, Limb JO, "Wireless Medium Access Control Protocols", *IEEE Communications Surveys*, Second Quarter 2000.
- [12] Chen K, Nahrstedt K, Vaidya N, "The Utility of Explicit Rate-Based Flow Control in Mobile Ad Hoc Networks," in *Proc. of IEEE WCNC*, Atlanta, Georgia, USA, March 2004.
- [13] Chhaya HS, Gupta S, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol," *Wireless Networks* 3(3) (1997) 217–234.
- [14] Chiang CC, Wu HK, Liu W, Gerla M, "Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel," *Proceedings of IEEE SICON'97*, pp. 197--211, April 1997.
- [15] Chlamtac I, Farago A, "Making transmission schedules immune to topology changes in multi-hop packet radio networks." *IEEE/ACM Transactions on Networking*, 2(1):23–9, Feb. 1994.
- [16] Chlamtac I, Farago A, Zhang H. "Time-spread multiple-access (TSMA) protocols for multihop mobile radio networks." *IEEE/ACM Transactions on Networking*, 6(5):804–12, Dec. 1997.
- [17] Cidon I, Sidi M, "Distributed assignment algorithms for multihop packet radio networks." *IEEE Transactions on Computers*, 38(10):1353–61, Oct 1989.

- [18] Deng J, Haas ZJ, "Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio Networks," in *IEEE ICUPC'98, Florence, Italy*, October 5-9, 1998.
- [19] Ephremides A, Truong TV. "Scheduling broadcasts in multihop radio networks.," *IEEE Transactions on Communications*, 38(4):456–60, Apr. 1990.
- [20] Freebersyser J, Leiner B, "A DoD Perspective on Mobile Ad Hoc Networks," *Ad Hoc Networking*, ed. C. E. Perkins, Addison-Wesley, 2001, pp. 29–51.
- [21] Fu Z, Zerfos P, Luo H, Lu S, Zhang L, Gerla M, "The Impact of Multihop Wireless Channel on TCP Throughput and Loss", in *Proc. of IEEE INFOCOM*, San Francisco, USA, April 2003.
- [22] Fullmer CL, Garcia-Luna-Aceves JJ, "Solutions to Hidden Terminal Problems in Wireless Networks," in *Proc. of ACM SIGCOMM '97*, 1997.
- [23] Garcia-Luna-Aceves JJ, Fullmer CL, "Floor Acquisition Multiple Access (FAMA) in Single-channel Wireless Networks," *ACM/Baltzer Mobile Networks and Applications*, vol. 4, no. 3, pp. 157-174, 1999.
- [24] Garcia-Luna-Aceves JJ, Tzamaloukas A, "Receiver-initiated Collision Avoidance in Wireless Networks," *ACM Wireless Networks*, vol. 8, pp. 249-263, 2002.
- [25] Gerla M, Tsai JTC, "Multicluster, mobile, multimedia radio network," *ACM/Baltzer Journal of Wireless Networks*. vol. 1, (no. 3), p. 255-265, 1995.
- [26] Grossglauser M, Tse DC, "Mobility Increases the Capacity of Ad Hoc Wireless Networks", *IEEE/ACM Trans. On Networking*, Vol. 10, No. 4, pp. 477-486, Aug. 2002.
- [27] Gupta P, Gray R, Kumar PR, "An experimental scaling law for ad hoc networks". Technical Report, Univ. of Illinois at Urbana Champaign, May 2001.

- [28] Gupta P, Kumar PR, "The capacity of wireless networks, " in *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000.
- [29] He T, Huang C, Blum BM, Stankovic JA, Abdelzaher TF, "Range-Free Localization Schemes in Large Scale Sensor Networks", In *Proc. of ACM MOBICOM*, 2003.
- [30] Hill J, Szewczyk R, Woo A, Hollar S, Culler D, Pister K, "System architecture directions for network sensors," in *Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-IX)* Cambridge, MA, Nov. 2000.
- [31] Holland G, Vaidya N , "Analysis of TCP performance over mobile ad hoc networks." in *Proc. of MOBICOM*, 1999.
- [32] Holland G, Vaidya N, Bahl P, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Mobile Computing and Networking*, 2001, pp. 236-251.
- [33] <http://www.ieee802.org/15/>
- [34] Hu L, "Distributed Code Assignments for CDMA Packet Radio Networks," *IEEE/ACM Tran. on Networking*, Vol. 1, no. 6, pp. 668-677, Dec. 1993.
- [35] Hull B, Jamieson K, Balakrishnan H, "Mitigating Congestion in Wireless Sensor Networks", in *Proc. ACM SenSys*, Baltimore, MD, USA, Nov 2004,
- [36] IEEE Std 802.11b. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band, IEEE, 1999.
- [37] Jacobson V, "Congestion Avoidance and Control", in *Proc. of ACM SIGCOMM*, Stanford California, Aug 1988.

- [38] Jain N, Das S, "A Multichannel CSMA MAC Protocol with Receiver-Based Channel Selection for Multihop Wireless Networks," in *Proceedings of the 9th Int. Conf. on Computer Communications and Networks (IC3N)*, October 2001.
- [39] Jain R, "A Comparison of Hashing Schemes for Address Lookup in Computer Networks", Tech. Rpt.-593, Digital Equipment Corporation, February 1989.
- [40] Jain R, "Congestion control and traffic management in ATM networks: recent advances and a survey," *ATM Forum Contribution 95-0177*, 1995.
- [41] Jakes WC, Ed., *Microwave Mobile Communications*, IEEE Press, Reprinted, 1994.
- [42] Jamieson K, Balakrishnan H, Tay YC, "Sift: A MAC Protocol for Event-Driven Wireless Sensor Networks", MIT LCS technical report, available at <http://www.lcs.mit.edu/publications/pubs/ps/MIT-LCS-TR-894.ps>, 2003.
- [43] Johnson DB, Maltz DA, "Dynamic source routing in ad hoc wireless networks", In *Mobile Computing*, edited by Tomasz Imielinski and Hank Korth, chapter 5, pages 153–181. Kluwer Academic Publishers, 1996.
- [44] Jung E, Vaidya NH, "A Power Control MAC Protocol for Ad Hoc Networks", *Proc. ACM Mobicom 02*, pp. 36-47, 2002.
- [45] Karn P, "MACA - a New Channel Access Method for Packet Radio," in *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, pp. 134-140, ARRL, 1990.
- [46] Katabi D, Handley M, Rohrs C, "Internet Congestion Control for Future High Bandwidth-Delay Product Environments," in *Proc. of ACM SIGCOMM*, Pittsburgh, PA, August 2002.

- [47] Katzela I, Naghshineh M, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey." In *IEEE Personal Communications*, pages 10-31, June 1996.
- [48] Keshav S, *An Engineering Approach to Computer Networking*. Reading, MA: Addison Wesley, 1998.
- [49] Kleinrock L, Tobagi FA, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-access Modes and Their Throughput-delay Characteristics," *IEEE Trans. on Communications*, vol. 23, no. 12, pp. 1400-1416, 1975.
- [50] Ko YB, Vaidya N, "Location-aided routing (LAR) for mobile ad-hoc networks". In *ACM / IEEE Mobicom '98*, Oct. 1998.
- [51] Kung HT, Blackwell T, Chapman A, "Credit-based flow control for ATM networks: Credit update protocol, adaptive credit allocation and statistical multiplexing," in *Proc. of ACM SIGCOMM*, 1994, pp. 101-114.
- [52] LaMaire RO, Krishna A, Zorzi M, "On the Randomization of Transmitter Power Levels to Increase Throughput in Multiple Access Radio Systems," *Wireless Networks*, vol. 4, no. 3, pp. 263-77, Mar. 1998.
- [53] Lee JY, Scholtz RA, "Ranging in a Dense Multipath Environment Using an UWB Radio Link," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1677-1683, December 2002.
- [54] Lee WCY, *Mobile Cellular Communications Systems*, McGraw Hill, 1989.
- [55] Li J, Blake C, De Couto DS, Lee HI, Morris R, "Capacity of ad hoc wireless networks", in *Proceedings of ACM MOBICOM*, pages 61-69, July 2001.

- [56] Lin CR, Gerla M, "Adaptive Clustering for Mobile Wireless Networks ,," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 7, , pp. 1265-1275, Sept. 1997.
- [57] Liu J, Singh S, "ATCP: TCP for mobile ad hoc networks", in *Proc. of IEEE Journal on Selected Areas in Communications*, 19(7):1300–1315, 2001.
- [58] Makansi T, "Transmitted Oriented Code Assignment for Multihop Packet Radio," *IEEE Transactions on Communications*, Vol. 35, pp. 1379-1382, 1987.
- [59] Mishra PP, Kanakia H, "A hop by hop rate based congestion control scheme," in *Proc. of ACM SIGCOMM*, August 1992.
- [60] Mitra D, Seery JB, "Dynamic adaptive windows for high speed data networks: Theory and simulations", in *Proc. of ACM SIGCOMM*, Philadelphia, Pa., pp. 30-40, September 1990.
- [61] Monks J, Bharghavan V, Hwu W, "A power controlled multiple access protocol for wireless packet networks", In *Proc. IEEE Infocom*, pages 219–228, April 2001.
- [62] Murphey RA, Pardalos PM, Resende MGC, "Frequency assignment problems," *Handbook of combinatorial optimization*, Volume A. Kluwer Academic Publishers, 1999.
- [63] Nasipuri A, Das SR, "Multichannel CSMA with Signal Power-based Channel Selection for Multihop Wireless Networks," in *IEEE Vehicular Technology Conference (VTC)*, September 2000.
- [64] Nasipuri A, Zhuang J, Das SR, "A Multichannel CSMA MAC Protocol for Multihop Wireless Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, September 1999.

- [65] NS-2, <http://www.isi.edu/nsnam/ns/>
- [66] Orinoco wireless networks, <http://www.orinocowireless.com/>.
- [67] Ozveren C, Simcoe R, Varghese G, "Reliable and Efficient Hop-by-Hop Flow Control ", *IEEE Journal on Sel. Areas in Communications*, vol. 13, no. 4, May 1995, pp. 642-650.
- [68] Pathmasuntharam JS, Das A, Gupta AK, "Channel Assignment for Nullifying the Critically Exposed Node Problem in Ad Hoc Wireless Networks," in *IEEE SECON*, October 2004.
- [69] Pathmasuntharam JS, Das A, Gupta AK, "Efficient Multi-rate Relaying (EMR) MAC Protocol for Ad Hoc Networks" in *IEEE ICC*, May 2005, South Korea.
- [70] Pathmasuntharam JS, Das A, Gupta AK, "Primary Channel Assignment based MAC (PCAM) – A Multi-Channel MAC Protocol for Multi-Hop Wireless Networks," in *IEEE WCNC*, March 2004.
- [71] Pathmasuntharam JS, Das A, Gupta AK, "Synchronized and Concurrent Enabling of Neighborhood Transmission (SCENT) – A MAC Protocol for Concurrent Transmission in Wireless Ad-hoc Networks", in *Proc. of the International Journal on Wireless and Optical Communications (IJWOC)*, World Scientific Publishing, December 2005.
- [72] Patwari N, O'Dea RJ, Wang Y, "Relative location in wireless networks," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, (Rhodes, Greece), pp. 1149--1153, May 2001
- [73] Perkins C, Bhagwat P, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers" in *Proc. of the ACM SIGCOMM*, October 1994.

- [74] Prasad AR, Moelard H, "WaveLAN-II system design note 225: Enhanced data rate control," March 1999.
- [75] Proakis JG, Digital Communications, McGraw Hill, Fourth edition, 2001
- [76] Raju J, Garcia-Luna-Aceves JJ, "Distributed Assignment of Codes for Multihop Packet Radio Networks." In *Proceedings of IEEE MILCOM '97*, Monterey, California, November 1997.
- [77] Ramakrishnan KK, Jain R, "A Binary Feedback Scheme for Congestion Avoidance in Computer Networks", in *Proc. of ACM Trans. on Comp. Sys. Vol.* 8, no. 2, pp 158-181, May 1990.
- [78] Rappaport TS, Wireless communications, principles and practice. Prentice Hall, 1996.
- [79] Sadeghi B, Kanodia V, Sabharwal A, Knightly E, "Opportunistic Media Access for Multirate Ad Hoc Networks," in *Proceedings of ACM MOBICOM 2002*, Atlanta, GA, September, 2002.
- [80] Seok Y, Park J, Choi Y, "Multi-rate Aware Routing Protocol for Mobile Ad Hoc Networks," *IEEE VTC 2003-Spring*, Jeju, Korea, April, 2003.
- [81] Shukla D, Wadia LC, Iyer S, "Mitigating the exposed node problem in IEEE 802.11 adhoc networks," in *IEEE Intl Conference on Computer and Communication Networks (ICCCN)*, October 2003.
- [82] So J, Vaidya NH, "A Multi-channel MAC Protocol for Ad Hoc Wireless Networks", UIUC Technical report, 2003. Available: <http://www.crhc.uiuc.edu/~nhv/papers/jungmin-tech.ps>

- [83] Stankovic JA, Abdelzaher T, Lu C, Sha L, Hou J., "Real-Time Communication and Coordination in Embedded Sensor Networks", *Proceedings of the IEEE*, 91(7), 2003.
- [84] Sundaresan K, Anantharaman V, Hsieh HY, Sivakumar R, "ATP: A reliable transport protocol for ad-hoc networks.", in *Proc. of ACM MobiHoc*, Annapolis, MD, 2003.
- [85] Talucci F, Gerla M, "MACA-BI (MACA by Invitation): A Receiver Oriented Access Protocol for Wireless Multihop Networks," in *Proc. of PIMRC '97*, 1997.
- [86] Tang Z, Garcia-Luna-Aceves JJ, "Hop-Reservation Multiple Access (HRMA) for Ad-Hoc Networks," in *IEEE INFOCOM*, 1999.
- [87] Tay YC , Chua KC, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Networks*, v.7 n.2, p.159-171, March/April 2001.
- [88] Tobagi FA, Kleinrock L, "Packet Switching in Radio Channels: Part II - the Hidden Terminal Problem in Carrier Sense Multiple-access Modes and the Busy-tone Solution," *IEEE Trans. on Communications*, vol. 23, no. 12, pp. 1417-1433, 1975.
- [89] Tzamaloukas A, Garcia-Luna-Aceves JJ, "A Receiver-Initiated Collision-Avoidance Protocol for Multi-Channel Networks", in *Proceedings of INFOCOM* 2001.
- [90] Wan CY, Eisenman S, Campbell A, "CODA: Congestion Detection and Avoidance in Sensor Networks", in *Proc. ACM Sensys*, Los Angeles, CA, November 2003.
- [91] Wang Y, Garcia-Luna-Aceves JJ, "Performance of Collision Avoidance Protocols in Single-Channel Ad Hoc Networks," in *Proc. of IEEE Intl. Conf. on Network Protocols (ICNP '02)*, (Paris, France), Nov. 2002.

- [92] Wu H, Peng Y, Long K, Cheng S, Ma J, "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis And Enhancement," in *Proceedings of INFOCOM, 2002*.
- [93] Wu SL, Lin CY, Tseng YC, Sheu JP, "A New Multi-Channel MAC Protocol with On-Demand Channel Assignment for Multi-Hop Mobile Ad Hoc Networks," in *IEEE Wireless Communications and Networking Conference (WCNC), Chicago, IL, September 2000*.
- [94] Wu SL, Tseng YC, Lin CY, Sheu JP, "A Multi-Channel MAC Protocol with Power Control for Multi-Hop Mobile Ad Hoc Networks," *The Computer Journal*, vol. 45, pp. 101–110, 2002.
- [95] Wu SL, Tseng YC, Sheu JP, "Intelligent medium access control for mobile ad hoc networks with busy tones and power control", *IEEE Journal on Selected Areas in Communications*, 18(9):1647–1657, 2000.
- [96] www.ietf.org/rfc/rfc3561.txt
- [97] www.monarch.cs.rice.edu/cmu-ns.html
- [98] Xu S, Saadawi T, "Does the IEEE 802.11 MAC Protocol work well in multihop wireless ad hoc networks?," in *IEEE Communications Magazine*, June 2001.
- [99] Xu S, Saadawi T, "Revealing the Problems with 802.11 Medium Access Control Protocol in Multi-Hop Wireless Ad Hoc Networks," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, v.38 n.4, p.531-548, 15 March 2002.
- [100] Ye W, Heidemann J, Estrin D, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc. of IEEE INFOCOM, 2002*, pp. 1567–1576.

- [101] Yi S, Pei Y, Kalyanaraman S, "On the Capacity Improvement of Ad Hoc Wireless Networks Using Directional Antennas" in *Proc. of ACM MOBIHOC*, Jun 2003.
- [102] Yi Y, Shakkottai S, "Hop-by-hop Congestion Control over a Wireless Multi-hop Network", in *Proc. of IEEE INFOCOM*, Hong Kong, March, 2004.
- [103] Yu X, "Improving TCP Performance over Mobile Ad Hoc Networks by Exploiting Cross-Layer Information Awareness," in *Proc. of ACM MOBICOM*, Philadelphia, PA, USA, September 2004.
- [104] Zhao J, Govindan R, "Understanding Packet Delivery Performance in Dense Wireless Sensor Networks," in *Proc. of ACM Sensys*, Los Angeles, CA, November 2003.
- [105] Zhou G, He T, Stankovic JA, "Impact of Radio Irregularity on Wireless Sensor Networks." In The Second International Conference on Mobile Systems, Applications, and Services (MobiSys), June 2004.
- [106] Zhu C, Corson MS, "A Five-Phase Reservation Protocol (FPRP) for Mobile Ad Hoc Networks," in *Proc. of IEEE INFOCOM '98*, vol. 1, (San Francisco, CA, U.S.A.), pp. 322-31, 1998.
- [107] Zorzi M, Rao RR, "Capture and retransmission control in mobile radio," *IEEE J. Select. Areas Communications*, vol. 12, pp. 1289–1298, Oct. 1994.