Resource Management Schemes for Mobile Ad hoc Networks

Sridhar K. Nagaraja Rao

NATIONAL UNIVERSITY OF SINGAPORE

2007

Resource Management Schemes for Mobile Ad hoc Networks

Sridhar K. Nagaraja Rao

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

2007

Dedicated To

My parents, brother and my wife

Special Dedication To

DVG

Acknowledgements

First, I wish to thank my supervisor Dr. Chan Mun Choon for his excellent guidance throughout this work, for helping me to clear my thoughts and grasp problems from the right sides, and for the wonderful time we had working together. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped immensely completing this thesis work.

I would like to thank my previous advisors Dr. Lillykutty Jacob, and Dr. Rajeev Shorey. I am grateful to Dr. Jacob for the enthusiasm and inspiration, which was always there when I needed it. I thank Dr. Grabiel Ciobanu for introducing me to the field of Process Algebras, for providing vital information about writing, and for providing encouragement, advice and lots of good ideas. I thank Prof. Xie Ming for the technical discussions on the lifetime distribution models. This work has greatly benefitted by the comments from my internal examiners Dr. Pung Hung Keng and Dr. A. L. Ananda, many thanks to them. I am also thankful to Dr. A. L. Ananda for providing me with the opportunity and resources to work at CIRL.

Many thanks to all the colleagues and friends with whom I shared a lab, who helped me in reviewing papers, and who encouraged me throughout this work: Venky, Subbu, Sudhar, Auri, Ravi, Aseem, Rahul and Anand. I am grateful to all my lab mates: Hao Shuai, Eugene, XiuChao, Shao Tao, Bin Bin, MingZe, for numerous stimulating discussions on different topics in numerous meetings. It would be a long list to mention all the other friends I am indebted to. I gratefully thank all of them.

Special thanks goes to my wife Pallavi for putting up with my late hours, my spoiled weekends, my bad temper, but above all for taking lots of pain in reviewing my papers and thesis. Finally, I am immensely indebted to my parents Prema and Nagaraja Rao, and my brother Sripad for their love and support throughout my everlasting studies, and for the thirst for knowledge they infected me with.

Contents

Abstract

Mobile wireless ad hoc network (MANET) is a collection of mobile nodes dynamically forming a network without the use of any existing network infrastructure or centralized administration. The rapid growth in demand for mobile communication has led to intense research and development efforts towards a new generation of wireless ad hoc networks. It is desirable for such ad hoc wireless systems to support a wide range of services. Adaptive resource management schemes play a key role in next-generation ad hoc wireless systems for providing desired services.

In this work, we develop individual resource management schemes and a service differentiation solution combining the schemes for mobile ad hoc networks to achieve efficient utilization of scarce available channel bandwidth. The goal is to provide an improved network performance. The significance of this work arises from the need for efficient bandwidth management schemes to counter the ever-growing bandwidth demand and the scarcity of available spectrum. In addition, we found that the existing techniques, assumptions and approaches may not cater for all MANET needs and environments.

We develop mechanisms focusing on the challenges and the inherent aspects of mobile ad hoc networks. In particular, we focus on the features of ad hoc networks such as shared wire-

less medium, multihop, node mobility and time varying channel quality in developing routing (SHARC), admission control (iCAC) and packet scheduling schemes (CaSMA). We carried out detailed study on important inherent features such as node mobility and its effects on wireless link characteristics, interference and its effects on channel bandwidth measurements. For example, link lifetime, one of the characteristics of wireless link is analyzed following the approach used in reliability engineering studies. These studies helped us to develop metrics and devise mechanisms which are suitable for mobile ad hoc environments.

First, we develop a route computation mechanism termed as Stability and Hop-count based Approach for Route Computation (SHARC), which can be built into existing routing protocols, and which considers the link quality (represented as residual lifetime) as a metric, designed for ad hoc network environments. Link lifetime studies revealed that earlier assumptions such as, the longer the two nodes have remained as neighbors, the probability that the two nodes continue to remain as neighbors for longer time is high, does not apply to many mobility patterns. In some cases, the opposite may be true. Besides, link lifetime distribution models are different for different mobility patterns, and the exponential model (as considered by majority of previous works) is not a suitable fit for all the mobility patterns studied. Further, link failures are never random, and for majority of the mobility patterns link failures are similar to "wear-out" failures. In addition, it is difficult to have an accurate measure of the residual link lifetime, and heuristics-based estimation of link lifetimes perform considerably better (with average estimation errors ranging from 5 - 50 seconds) across various mobility patterns. Evaluation of SHARC that considers both stability and hop-count, shows that SHARC performs better than existing hop-based (DSR: 10% - 40%) and stability-based (ABR: 5% - 50%) routing mechanisms, and across various node mobilities

(Low Speed: 10% - 30%, High Speed: 10% - 45%).

Second, we develop a novel call admission control scheme termed as interference-based Call Admission Control (iCAC), which relies on the estimation of the positions of interfering nodes, and adheres to a fairness notion of equal-and-fair share. For position estimation, we exploit the wireless radio antenna states and noise measurements. We found that the estimation of position of interfering nodes helps in assessing the amount of available bandwidth for ad hoc environments. Performance evaluation of iCAC through simulation shows the following performance improvements: 50% more throughput, 30% less loss rate and 50% more calls admitted in comparison with existing schemes for single hop scenarios, and 30% to 50% decrease in average delay in comparison with IEEE 802.11 for multihop scenarios.

Third, we develop a packet scheduling scheme termed as Channel-aware Scheduling for MANETs (CaSMA), which considers end-to-end channel conditions in making the scheduling decisions. For efficient resource allocation, we found that it is advantageous to consider the end-to-end channel quality along with local channel quality while making the scheduling decisions. Combining both link lifetime and congestion level helps in modeling the end-to-end channel conditions effectively. Simulation results for CaSMA shows a 25% less packet loss, 30% - 40% less backlog and 50% increased TCP throughput in comparison with FIFO for estimation lifetime cases.

Finally, we combine above three schemes into single service differentiation solution, termed as UNIFIED. UNIFIED solution is developed to evaluate the combined performance, demonstrate the flexibility of the schemes and to have a comparative study with the existing service differentiation solutions. Performance evaluation of the combined service differentiation solution, UNIFIED, in comparison with an existing service differentiation architecture (SWAN) shows a 5% - 80% decrease in average delay and 25% increase in TCP throughput for varying real-time traffic. In addition, there is a 30% decrease in average delay and 5% - 15% increase in TCP throughput for various node mobilities.

Our findings show that it is important to develop mechanisms specifically for MANETs focusing mainly on the challenges and inherent features of MANETs. Such mechanisms, either used individually or combined into a resource management solution, perform better across various scenarios.

List of Figures

List of Tables

List of Abbreviations

AB	Available Bandwidth
ABR	Associativity Based Routing
AODV	Ad hoc On-demand Distance Vector
BSB	Bandwidth with Sensing as Busy
BSI	Bandwidth with Sensing as Idle
CAC	Call Admission Control
CaSMA	Channel aware Scheduling for Mobile Ad hoc Networks
CDF	Cumulative Distribution Function
CTS	Clear To Send
DSR	Dynamic Source Routing
EDF	Earliest Deadline First
iCAC	Interference based Call Admission Control
IEEE	Institute for Electrical and Electronics Engineers
IR	Interference Range
MAC	Medium Access Control
MANET	Mobile Ad hoc Network

MTTF	Mean Time To Failure
NS	Network Simulator
PDF	Probability Density Function
QoS	Quality of Service
QS	Queue Size
RLT	Residual Life Time
RMS	Rate Monotonic Scheduling
RPGM	Reference Point Group Mobility
RTS	Request To Send
RWP	Random Waypoint
SHARC	Stability and Hop-count based Approach for Route Computation
SNSDS	Stability and Neighbor State Dependent Scheduling
SSA	Signal Strength Adaptive
SWAN	Stateless Wireless Ad hoc Network
TR	Transmission Range
UNIFIED	Unique Features Influenced
WLAN	Wireless Local Area Network

Chapter 1

Introduction

This introductory chapter will provide the description of wireless mobile ad hoc networks, covering the features, advantages and history, followed by an overview of applications and technologies. Our motivation behind this work is described next, followed by a description of the problem addressed in this thesis, challenges involved, approach taken and significant contributions. We conclude this chapter by listing a few operational assumptions. In this thesis, we use the terms "mechanism" and "scheme" interchangeably.

1.1 Introduction to Mobile Ad Hoc Networks

There has been a tremendous advance in the development of small and smart devices, which users carry with them as they move around. Similar devices are also embedded in appliances and vehicles. Such devices can operate in a collaborative way, which drives the need for networking of such mobile devices without any support of infrastructure.

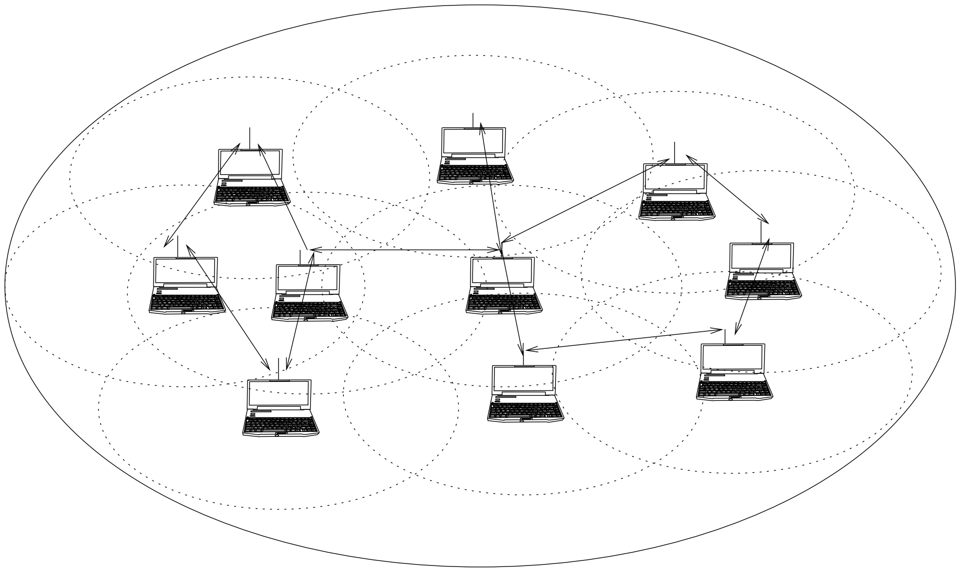


Figure 1.1: Ad hoc network

One such network of wireless and mobile devices is Mobile Ad hoc Networks (MANETs), shown in Figure 1.1. In Figure 1.1, the arrows indicate the communication links between the nodes, and the dotted circles indicate the transmission ranges of the nodes.

Mobile Ad Hoc Networks are defined as an autonomous system of mobile routers and associated hosts connected by wireless links [1]. The nodes are free to move randomly and organize themselves arbitrarily. Each node is equipped with a radio transmitter/receiver, which allows it to communicate with its neighboring nodes. These wireless radios, however, have limited transmission capabilities. Because of the limitation of transmission capabilities, not all nodes are within the range of each other. If a node wishes to communicate with a node outside its transmission range it has to take the help of other nodes by constructing a multihop route. Every node is capable of generating data, and carrying data for other nodes.

Typical characteristics of ad hoc networks include [2]: (1) Mobility - nodes are free to move in any random or well-defined paths (2) Multihop - path from source to destination can traverse through several nodes (3) Self-Organization - nodes must autonomously determine its own configuration (addressing and clustering) (4) Resources - both the available bandwidth and power are limited (5) Security - malicious nodes (intruders) may exist (6) Internet connectivity - might have to integrate with infrastructure standards (7) Scalability - network can grow from tens to thousands of nodes.

Inherent features of mobile ad hoc networks brings about various advantages. The basic concept that the network can be brought up or torn down in a short time provides a lot of flexibility. As ad hoc networks does not require any fixed infrastructure, they eliminate the infrastructure costs. This feature makes ad hoc networks economical compared to other networks. Existence of multi-hops provides larger coverage area, and results in increasing the scalability of the network. Further, ad hoc networks can extend the range of existing infrastructure based wireless and wired networks (WLANs and Internet) [3].

Brief History of Ad Hoc Networks

There have been lot of research and development in the field of ad hoc networks. The evolution of mobile ad hoc networks started with DARPA-sponsored PRNET (Packet Radio Networks) in 1970s to provide networking capabilities in a combat environment [4]. Around 1980s PRNET supported 138 nodes, and it used a flat distance vector routing. PRNET project was further enhanced and developed under the project called SURAN (Survivable Adaptive Radio Networks) program, which developed a packet-switched, infrastructure-less network for battlefield environment. This project ran from 1983 to 1992. SURAN was followed by Department of Defense (DoD) supported projects Global Mobile Information Systems (GloMo, 1995 - 2000) and Near Term Digital Radio (NTDR) [1]. These projects were developed to support higher number of nodes (400), and used two-level routing hierarchy.

In the earlier stages of growth, ad hoc networks used proprietary and single technology, and protocols used were technology specific. There was a strong need to develop IP based protocols for ad hoc networks. The main reasons for having an IP based solution were: hardware economics, standards based protocols, Internet connectivity, routing flexibility and future QoS support [5]. In this regard, a working group for mobile ad hoc networking was formed within the Internet Engineering Task Force (IETF). Spurred by the growing interest in ad hoc networking, various commercial standards were developed in late 90s. This includes IEEE 802.11 Physical and MAC protocols in 1995 [6], which influenced numerous applications to be developed for ad hoc networks. In the next part, we will focus on the various applications for ad hoc networks.

Applications

Ad Hoc networks are deployed in those places where building an infrastructure is difficult, due to constraints of cost and time. We have seen various advantages of mobile ad hoc networks in previous paragraphs. These advantages gave rise to initial applications such as battlefield and disaster recoveries. Figure 1.2 summarizes various class of applications of MANETs.

A popular class of applications are those that use autonomous agents such as unmanned ground vehicles (UGVs), unmanned underwater vehicles (UUVs) and unmanned airborne vehicles (UAVs) [2]. Ad hoc network involving these agents can be used for various purposes, such as intelligence, surveillance, damage assessment and search and rescue. An other recent application is home network, which includes communication between smart household appliances. Campus-wide communications is another growing application area of ad hoc net-

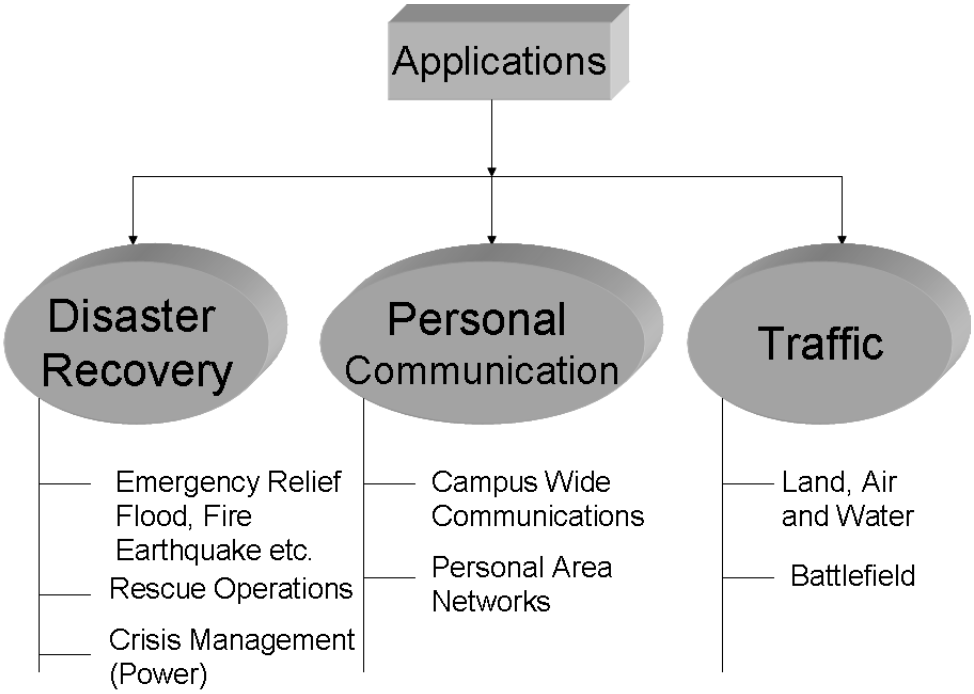


Figure 1.2: Ad hoc network applications

works. The term campus is used to refer to any place where people congregate for various reasons (work, study and entertainment). This can include technology parks, amusement parks, University campuses and shopping malls. Vehicular ad hoc networks is an upcoming application of ad hoc networks. This includes traffic control, hazard warning on roads and air traffic control.

Architectures and Technologies

Wireless networks can use different technologies. We highlight some of these technologies. *Bluetooth* is designed to meet low-power, low-cost and low-range goals. A technology, which was developed as a replacement for serial cable, Bluetooth can currently work with up to 7 devices in *piconet* (master-slave paradigm) [7]. Further, scalability is increased by connecting more than one piconets together to form a *scatternet*. *IEEE 802.11* is the most popular standard for WLANs [2]. Distributed Coordinated Function of IEEE 802.11 is proposed to support both ad hoc (infrastructure-less) WLAN and infrastructure WLANs. There is a category of broadband wireless ad hoc networks achieved by IEEE 802.16 recommendations. Typical ad hoc network deployed in this category is in the form of mesh networks (IEEE 802.16s) [8]. Some researchers, however, prefer to refer to ad hoc networks only for those networks where multihop exists. In this regard, they choose to exclude Bluetooth and infrastructure WLANs [9].

Resource Management in Ad Hoc Networks

The rapid growth in demand for mobile communication has led to intense research and development efforts towards a new generation of ad hoc networks. The new system must be able to provide quality-of-service (QoS), support a wide range of services and improve the system capacity. Efficient utilization of the scarce channel bandwidth for wireless communications is certainly one of the major challenges in MANET system design.

Important resource management functions include call admission control and scheduling. End-to-end routing also plays a major role as it complements resource management schemes to obtain various end-to-end information, and improves the efficiency of these schemes. Call admission control (CAC) is one method to manage radio resource in order to adapt to traffic and topology variations. CAC refers to the process of make a decision for new admission according to the amount of available resource versus users requirements, and effect on the existing calls imposed by new call. On the other hand, scheduling decides which flow among the set of backlogged flows within a node should get the chance to be transmitted over the network. The important features of any resource management solution in mobile ad hoc environments that we emphasize in our work can be broadly classified as: (1) accurate measure of available resource (2) fair allocation of available resource (3) efficient use of available resource.

In MANET environments, having an accurate measure of available bandwidth can be challenging due to the shared wireless medium feature. In addition, overheads involved in measurements of available bandwidth by mechanisms at network layer increases with node mobility and multiple hops. Once a measure of available bandwidth is obtained, designing a fair notion in wireless multihop environments is also a challenging problem. This fairness problem can be in two levels - fairness among set of competing nodes within a contending region and fairness among set of backlogged flows within a node. Finally, wireless, mobile and multihop features of MANET also hinders the efficient use of scarce channel bandwidth.

Hence, shared bandwidth among all the contending nodes, limited bandwidth availability, time varying nature, difficulty in estimating the available bandwidth, difficulty in reserving bandwidth, unable to hold multiple packets "back-to-back" in one transmission (sender has to contend for the channel again for the next transmission, which makes the delay (d) of sending out a packet over the wireless link tightly coupled with the link's bandwidth) define the dynamics of channel bandwidth and challenges involved in efficient resource management in MANETs.

1.2 Motivation

The use of wireless communications has become desirable if not unavoidable. One such communication system is infrastructure-less networks. This is an area that is rapidly evolving and has exciting possibilities for future research. We believe that in future, applications that require enhanced performance would be developed for MANETs. This is evident considering the amount of research that is carried out whose main focus is to improve the performance or provide guarantees. Further, it is important to consider the existing and foresee the possible operating environments of ad hoc networks. We can see that ad hoc network operates either as an independent network or as an extension of the Internet. In either case, it is expected to carry both multimedia and real-time traffic. This argument serves as a case for developing resource management schemes for ad hoc networks.

The three inherent characteristics of MANETs [2], as mentioned below, which also acts as design challenges in MANETs, further motivated us to develop resource management schemes for MANETs.

- Multi-hop exists, and flow on this multi-hop is affected by the frequent fluctuations of the channel quality *due to node mobility.*
- Wireless medium is shared, and even packets of the same stream contend for this media at adjacent nodes.
- Interference affects transmission at nodes beyond immediate neighbors.

A principal requirement of any resource management scheme is to make these challenges an important driving force. Catering for these challenges should not be an afterthought, but an integral part of the solution. Hence, the resource management schemes should achieve good performance by adapting to the inherent features of MANETs.

1.3 Problem Description and Approach

The problem addressed in this thesis is the *design of resource management schemes, focusing on the available channel bandwidth as resource, to improve the performance when multimedia applications are supported in MANETs.* Resource management problem can be seen as a subproblem of providing QoS. The resource management problem focuses on maximizing the system goodput, reducing the average delay and improving the fairness.

Radio environment, limited resources, lack of infrastructure and topology changes are the major hindrances in satisfying the resource and performance constraints in MANETs. Therefore, we believe that any resource management solution developed for MANETs should take into consideration the inherent features such as shared wireless medium, multihop and mobility. Challenges in developing resource management schemes can be explained by considering these mentioned inherent features.

In a shared wireless medium, transmission by one node will not just consume the bandwidth of that particular node, but also the bandwidth of other neighboring nodes. This problem is pronounced in the MANET environment, where multihop scenarios are present. Some of the important problems are: transmission of a flow at a node is interfered by transmission of same flow by neighboring nodes and available bandwidth measurement should consider the transmissions by all the interfering nodes.

Node mobility affects the network topology, which can result in frequent and dynamic changes. This implies that the multihop path between any source and destination also keeps changing with time. In addition, mobility in ad hoc networks also causes unpredictability in the quality of a wireless link between any two nodes. Finally, mobility makes the problem of achieving fairness (both among the set of flows within a node and among the set of nodes within a contention region) challenging.

Existence of multihop enforces any channel-aware mechanism to consider end-to-end channel quality information along with local channel quality information. Developing a consistent and suitable parameter to represent end-to-end channel qualities is also a challenging problem in MANETs.

The challenges described above poses new design requirements, and also requires solutions that are different from solutions developed for conventional wired/wireless infrastructure networks. For example, among the solutions proposed for Internet, there is a requirement of either maintenance of states or existence of end-to-end service architecture. Whereas, for MANET environments these solutions might be difficult to use due to the limitations within a node and the inherent features of MANET. Due to the decentralized nature of ad hoc networks, the quest for the distributed and adaptive solutions exacerbates the problem. Maintaining costly states will introduce a lot of overheads and at times might degrade the performance. In addition, dynamic topology changes also introduce challenges in the end-to-end service architecture. On the other hand, the mechanisms like admission control, queuing and scheduling, and policing that are used to realize the resource management in Internet can be incorporated in MANETs.

In infrastructure-less networks like MANETs, unlike Internet, focus is on the local mechanisms within a node. We focus on the minimum set of mechanisms that are required within a node to achieve the resource management goal. The set includes: (1) a routing protocol to find routes, may be with or without constraints (2) a mechanism to decide whether to allow a flow into the network or not (policing, admission control and constraint based routing). (3) a mechanism, which allocates the share of network bandwidth to different flows (queuing and scheduling). (4) a medium access control mechanism, which controls multiple access. In the remaining part of this section, we will describe the approach taken in our work in developing the resource management solution.

1.3.1 Approach

Our solution concentrates on the features unique to MANETs in designing the mechanisms for resource management, instead of porting the solutions designed from Internet. We identify a set of unique features (shared wireless medium, multihop and node mobility), to consider in our solution. We aim to consider following components: routing and admission control at Network layer and queuing mechanism at MAC layer. Figure 1.3 indicates the scope of the work. Apart from concentrating on individual mechanisms we also see how these mechanisms are inter-related, i.e., we study the inter dependence of the mechanisms.

Our motivation for including a route computation mechanism as part of resource management solution is that, in the context of MANETs, we believe that it is necessary to have reliable routes before carrying out actual resource management. Our routing mechanism considers the inherent feature of *dynamic topology changes due to mobility* while selecting the routes. We translate this feature by measuring the stability of the link. A link is stable if it endures for longer time than the other paths in a network. Path stability depends on the availability of all the links constituting that path. A link being available means the radio

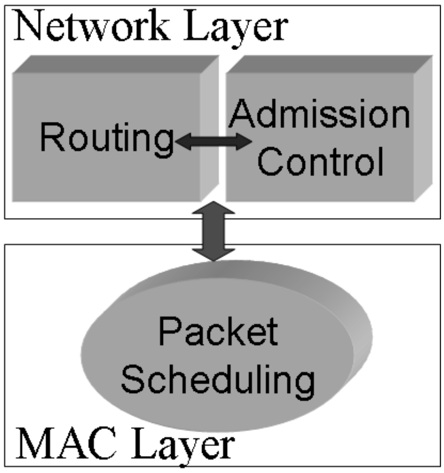


Figure 1.3: Mechanisms considered for resource management

quality of the link satisfies the minimal requirement for a successful transmission [9].

We understand that a stability-based routing proposal should be well supported by a detailed study on link quality variations. In this regard, we carry out a study on link lifetimes and attempt to associate a parametric statistical model to the lifetimes, which will help in understanding of various link/path quality aspects. We define link stability considering residual life-time of a link. Residual life-time of a link denotes the amount of time remaining for the link from the current age. We found that exact measurement of residual life-time is difficult. Therefore, we use the predicted value of the residual life-time. Various factors influence this prediction mechanism. Current age, environment in which the node is operating are factors that influence the residual life-time of the link. Prediction mechanism, however, is chosen after a detailed study of link lifetime distributions, and various prediction techniques.

We also argue that pure stability-based routing might not be helpful always just like pure hop-count based routing. Therefore, we propose a route computation mechanism (Stability and Hop-count based Approach Route Computation - SHARC) that considers both stability and hop-count, and which can be added to majority of routing protocols. We believe that the combination of both hop-count and stability would be appropriate for resource management. Our research, and works by other researchers have shown that pure stability-based mechanism might perform badly when it tends to choose long routes. Also, pure hop-count based mechanism might not consider stable routes when available. A major advantage of SHARC is that it can be included in majority of the available routing protocols. The details of our routing mechanism is explained in Chapter-2.

In admission control, a node has to decide whether it can admit a flow in the network, depending on its measure of channel capacity. It deals with provisioning of channel resource. Admission control is typically achieved by having a measure of available bandwidth, which can be measured by various techniques, and deciding whether the network can handle the new flow. Our approach for call admission is interference-based. We term our call admission control mechanism as Interference based call admission control (iCAC). Admission control mechanism proposed at the network layer considers the *Shared Wireless Medium* feature of MANETs.

In our approach, the bandwidth measurement is also accompanied by measurement of the interference (noise values), which capture the nature of shared wireless medium. The measurement of noise helps us in understanding the environment, which would be difficult by just bandwidth measurements. We use this information along with bandwidth measurements to first carry out position estimation of the interfering nodes. The position estimation information drives the admission control decision. These features highlight novelty in available resource measurements. End-to-end bandwidth measurement is incorporated to cater for multihop networks, which is achieved by enhancing the routing protocol. Our scheme is highly adaptive to the multihop and mobile environments, which is not present in majority of the existing proposals.

A scheduling algorithm determines which queued packet is to be processed next, and has a major impact on the performance of mobile ad hoc networks. Our packet scheduling mechanism selects packets, which have high probability of reaching the destination, and takes into account the cost of a link breaking by giving priority to flows that have a longer (normalized with path residual lifetime) backlog queue. We consider both *changing topology* and *shared wireless medium* features of MANETs in our scheduling mechanism. We term our scheduling mechanism as Channel-aware Scheduling for MANETs (CaSMA). We consider the end-to-end channel conditions, which is represented as path residual lifetime (RLT), in making the scheduling decision. This end-to-end consideration makes CaSMA channel-aware and increases the network performance. RLT is also combined with the workload at intermediate nodes, so that CaSMA is both channel-aware and congestion-aware. This combination attempts to approximate a global ideal scheduler that minimizes the backlog and provides a fair share of throughput. We have included a novel schedulable-list technique, which apart from providing better end-to-end co-ordination and approximation to an global ideal scheduler, also increases the goodput of the network.

Finally, to demonstrate the flexibility of the developed solutions, we combine the components (SHARC, iCAC and CaSMA) to achieve the service differentiation solution. We call our service differentiation solution as UNIque Features InfluencED (UNIFIED) solution for service differentiation in MANETs. We highlight the interaction between various layers of the network by providing a cross-layer design architecture.

Majority of performance evaluations of the proposed protocols are performed using simulations. While simulation studies have their limitations, Gerla et al. [10] have argued that "analytic models are practical only for small scope, microscopic tradeoffs. For complex studies, simulation is the only viable solution". We use simulators such as NS-2 [11] and GloMoSim [12] for our studies.

1.4 Contributions

The subject of this thesis is the resource management in wireless mobile ad hoc networks. This work on resource management leads us to develop individual mechanisms (route computation, admission control and packet scheduling) and a combined service differentiation solution. In this section, we describe our contributions considering all components and focusing on existing proposals and problems.

From our study on existing stability-based protocol and various temporal properties of wireless links we found that:

- Earlier assumptions such as, longer the two nodes have remained as neighbors, the probability that the two nodes continue to remain as neighbors for longer time is high, may not apply for all the mobility patterns. In some cases it may be the opposite.
- Link lifetime distribution models are different for different mobility patterns. A single distribution model for all mobility patterns is incorrect. In majority of earlier works researchers assume exponential model for link lifetimes. On the contrary we found that exponential model is not a suitable fit for all the common mobility patterns studied.

- Link failures are never random, and for majority of the mobility patterns link failures are similar to “wear-out” failures.
- It is difficult to have an accurate measure of the residual link lifetime, and heuristics based estimation of link lifetimes perform considerably better across various mobility patterns.
- Pure stability-based mechanism might perform badly when it tends to choose long routes. We propose a route computation mechanism (Stability and Hop-count based Approach Route Computation -SHARC) that considers both stability and hop-count, and which can be added to majority of routing protocols. Simulation results show that SHARC performs better than existing hop-based and stability-based routing mechanisms.

Contributions at the admission control scheme can be listed as below:

- We develop a novel scheme to estimate the position of interfering nodes. The scheme involves monitoring radio antenna states and measuring noise values.
- We consider the fairness notion (equal and fair share) in our admission control scheme, which is an important feature of resource management mechanism.
- Combining the position estimation and fairness features, we develop an available bandwidth estimation algorithm (iCAC).

From our study of channel-aware packet scheduling as resource allocation scheme we found that:

- For MANETs, it is important to consider the end-to-end channel quality along-with local channel quality while making the scheduling decisions.

- Combining both link lifetime and congestion level helps in modeling the end-to-end channel conditions.
- Performance evaluation show that packet scheduling mechanism based on channel conditions can prove advantageous even in mobile ad hoc environments (channel-aware schemes have proven advantageous in infrastructure WLANs).

Our findings show that it is important to develop mechanisms specifically for MANETs focusing mainly on the challenges and inherent features of MANETs. Such mechanisms either used individually or combined into a resource management solution perform better across various scenarios.

1.5 Network Model and Operational Assumptions

In this section, we describe the network model and various assumptions. This description helps the reader to understand the remaining chapters easily. We will consider a graphical modeling of ad hoc network. A graph, G is defined as set of vertices V and a set of edges E, and is denoted as $G = (V, E)$. We use set V to denote set of nodes and set E to denote set of links, and are assumed to be finite. Vertices i and j forms the end-nodes of a link l, and is denoted as $l_{i,j}$. If an edge (link) exists between two vertices (nodes), then the two vertices are termed as neighbors. Two edges (links) are considered adjacent if they have only one common end-node.

Every edge of a graph includes specific values termed as quality (Q) of an edge. Then,

$Q(l_{i,j})$ denotes the quality of the edge (link) $l_{i,j}$. This identifier is similar to that of "weights", and can be used for prioritizing. In ad hoc network, the communication between vertices is decided by this quality identifier, which may change over time depending on various conditions. Before proceeding further with the assumptions, we would like to first describe a few general terms that will be used in our work.

- Degree: The degree of a node i is the number of direct neighbors of that node in the network. If we consider $l_{i,j} = 1$ if vertices i and j are neighbors, and zero otherwise, then the degree d_i can be written as $\sum_{j=1}^{N} l_{i,j}$, where N is the total number of vertices.
- Path: A path between vertices i and j is said to exist if they are either direct neighbors ($l_{i,j}$ *exists*) or connected by only adjacent edges ($l_{i,k}, l_{k,l} \dots l_{n,j}$ *exists*).
- Hop-count: Hop-count specifies the number of edges on the path between two vertices i and j.
- Shortest Path: Shortest path between vertices i and j is the path, which has smallest number of hop-count among all the paths.
- Quality of a Path: Quality of a path is some mathematical formulation of the individual qualities of edges that forms the path. It can be additive, multiplicative, minimum, and maximum depending on the representation of the quality.

Protocol Stack

We assume the protocol stack (and corresponding responsibilities) for MANETs is similar to that of the 4-layer stack proposed for Internet: Transport, Network, MAC and Physical. One important difference is the power control (power level at which a packet on a hop is transmitted), which can be either addressed at the network layer or MAC layer.

Radio Technology

The radio technology used for ad hoc network can vary over a wide range of systems and standards. The suitable technology is typically based on the network size. Without providing the details on various available technologies, we would like to mention that we assume that the network uses Wireless Local Area Network (WLAN) technology IEEE 802.11a/b/g. The coverage area is limited to few hundreds of the meters. This communication range should suffice for majority of the applications mentioned in the preceding section.

Mobility Support

Support for mobility is an important advantage of wireless ad hoc networks. This support for mobility can be achieved either by Mobile IP or routing protocols [13]. Mobile IP [14] provides architectural solution for mobility support, which is suitable for nomadic users and not if the mobility is fast and topology changes are frequent. Whereas, routing protocols can be designed to cope with changes in network topology. Routing protocol approach is an ideal and widely accepted solution for ad hoc networks. Therefore, in our work we assume existence of a routing protocol, which provides mobility support.

Transmission Rate

In the previous sections, we mentioned the importance of cross-layer interactions in designing solution for MANETs. In this regard, we assume the following link transmission rate function [15] $R(t) = \Omega(Q_t, F)$, where $R(t)$ is the transmission rate at any time t, is a function of Q_t quality of the channel (single/multi hop) at time t and the feedback F from lower layer mechanisms (LLMs). In a mobile ad hoc network, the channel conditions vary for various reasons - mobility, congestion and interference. This varying conditions affect the transmission capabilities. Further, depending on the channel conditions, mechanisms proposed in our work such as admission control and scheduling provide feedback to the rate-control mechanism. This feedback can take different forms like, choosing a set of flows, deciding a rate for flow and blocking a set of flows. Therefore, our transmission rate function is dependent on both the channel condition and the feedback provided.

Other Assumptions

Our work is predominantly based on measuring and reacting to the link quality. Therefore, we assume a system where it is possible to obtain a timely feedback about the link quality. All the links in the network are bi-directional. We do not assume any additional error models during the packet reception. We also assume existence of a transport protocol, which may include flow control decisions. Our work does not cater to networks where there are malicious nodes, and we assume a network devoid of it and of only nodes who co-operate with each other.

Majority of our studies depend heavily on simulations. We use simulators such as NS-2 [11] and GloMoSim [12] for our studies. We add and modify protocols to suit our requirements

1.6 Thesis Organization

In this work, we develop three components: routing, admission control and packet scheduling, and a service differentiation solution. We organize the subsequent chapters in the same order. Chapter 2 describes the route computation component (SHARC). Chapter 2 emphasize on the importance of link stability based routing and carry out a detailed study of the temporal properties of wireless links. Chapter 3 explains the call admission control scheme (iCAC), where we concentrate on novel bandwidth measurement and fair allocation techniques. CaSMA, channel-aware scheduling mechanism, which stresses on the importance of considering end-to-end channel quality in packet scheduling is described in Chapter 4. Chapter 5 describes the service differentiation solution (UNIFIED), which is developed to demonstrate the flexibility of the three schemes, and which helps us to understand the combined performance of the schemes. Chapter 6 provide concluding remarks and propose few future directions.

Chapter 2

Routing

In this chapter, we focus on the protocol that underlies the establishment of paths using which the mobile nodes in ad hoc network can communicate with each other. There are various dimensions to the design domain of routing in ad hoc networks. We focus on one such dimension, termed as link stability-based routing. We perform a detailed link lifetime studies, and based on this study we introduce a route computation scheme, termed as Stability and Hop-count based Approach for Route Computation (SHARC).

2.1 Introduction

An ad hoc network is a network established by a collection of mobile nodes in a shared wireless media, by virtue of their proximity to each other. If all the wireless nodes in an ad hoc network are within the transmission range of each other (typically termed as fully connected), routing is not required. In practice, however, some of the wireless nodes are not within the transmission range of each other. Therefore, combined with restricted transmission range, node mobilities and lack of infrastructure, *multihop routing* is a challenging problem in MANETs.

Since the advent of packet radio networks, numerous routing protocols have been developed for ad hoc mobile networks [16–24]. Despite being designed for the same type of underlying network, the characteristic of each of these is distinct and the design principle varied. There have also been various works, which have done a comparative study among protocols [25–30]. Previous literature published by Royer et al. [31] reviews and presents important protocols of that time.

In mobile ad hoc network, each node if it volunteers to carry traffic for other nodes, participates in the formation of network topology. The concept is similar to the intermediate nodes/routers within the Internet, which cooperate to form multihop routing. This similarity has motivated many researchers to adapt existing routing protocols in Internet for use in ad hoc networks. In our work, we argue that apart from considering the functional similarity, researchers should also focus on the unique features that define ad hoc network. Therefore, we consider the intrinsic feature of mobile ad hoc networks such as *dynamic change in the topology* for designing the routing mechanism. Dynamic change in topology is a result of changes in the link stability either due to node mobility or due to congestion. *Dynamic change in topology* feature can be mapped to the link-stability metric of routing protocol. In this chapter, we focus on the stability-based mechanisms and propose a route computation mechanism, which combines both stability and hop-count features. We term this route computation mechanism as Stability and Hop-count based Approach for Route Computation (SHARC). The details of SHARC can be found in Section 2.5.

The link-stability metric that we consider in our work is residual link lifetime. Residual link lifetime can be described as follows. Let us consider a node n_1, with transmission range T. Consider a time t_1, where a node n_2 comes within the transmission range or n_1, then the link between the nodes n_1 and n_2 is said to be initiated. We call this time t_1 the *link initiation time*. Now, let us consider, at some point of time in future t_2 ($t_2 > t_1$), the node n_2 moves out of transmission range of n_1. Then the time t_2 is termed as link termination time. *Link lifetime is the difference between the link initiation time and link termination time* ($t_2 - t_1$). Residual link lifetime is the amount of time remaining in the link lifetime, at any given time t ($t_1 \leq t \leq t_2$), computed as ($t_2 - t$).

This chapter is organized as follows. In Section 2.2, we discuss the routing protocols classification and consider in detail the link stability based mechanisms. We also discuss existing studies on link and path lifetimes. In Section 2.3, we provide detailed study of link lifetime, which includes - collection of link lifetime data, association of statistical model, analysis of degradation process, and analysis and application of associated models. In addition, we describe the residual lifetime estimation process in Section 2.4. In Section 2.5, we describe our proposal of stability-based route computation mechanism (SHARC). We also provide a performance analysis of our approach, comparing it with other stability-based mechanisms. We conclude this chapter with a summary in Section 2.6. Detailed evaluation of the existing link stability based routing protocol (ABR), and brief description of the lifetime distribution models are provided as Appendix 2.A and Appendix 2.B, respectively, at the end of the chapter.

2.2 Related Work

In this section, we describe all the related works, which are classified into two subsections. First, we describe the various routing protocols proposed for MANETs. Second, we explain

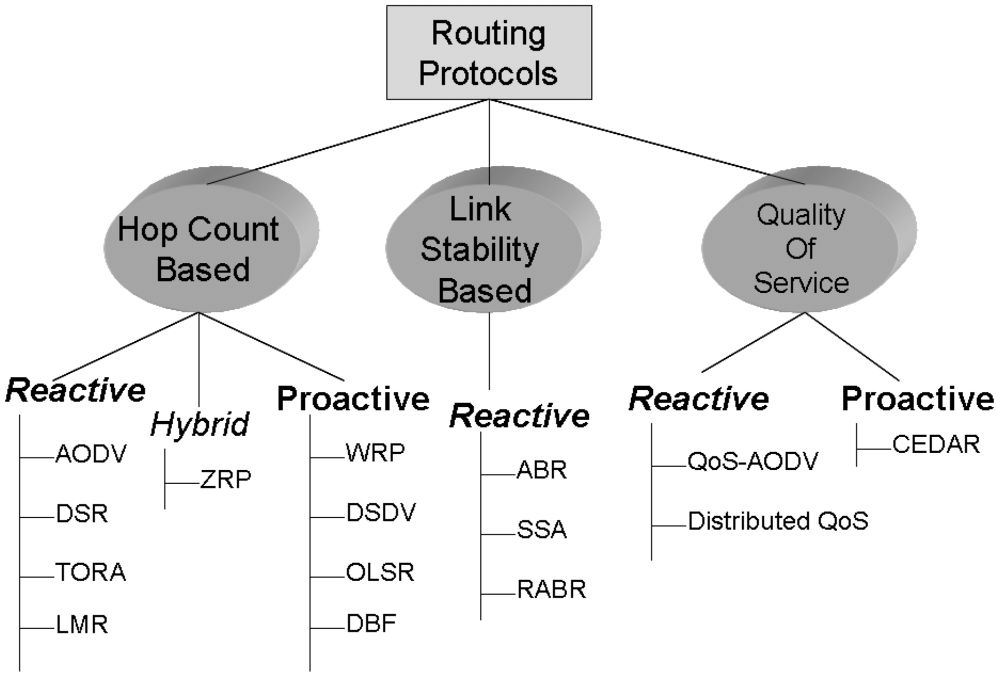


Figure 2.1: Routing protocol classification

works which study the impact of node mobility on the network performance. In the second part we also describe studies, which mainly focus on link or path lifetimes.

2.2.1 Routing Protocol Proposals

Typically, classification of routing protocols for MANETs is done by considering its route discovery philosophy. These protocols can be broadly classified as proactive and reactive. Proactive approach attempts to maintain consistent, up-to-date routing information from each node to every other node in the network. On-demand or reactive approach creates routes only when desired by the node. Studies have shown that routing protocols that build routes on-demand are practically more useful than those which are proactive in ad hoc wireless multihop environments [32, 33]. The major goal of on-demand protocol is to minimize control traffic overhead.

In our work we are more interested in the routing protocols, which track link quality and select higher-quality links over poor-quality ones. We want to emphasize on those protocols which considers some representation of wireless channel (whose condition vary significantly over time and space) as routing-metric. Therefore, we classify the routing protocols based on the routing metrics. Our classification is shown in Figure 2.1. Majority of the routing protocols are based on the hop-count metric. Hop count based metrics typically try to optimize the length of the route. There are both reactive and proactive protocols based on hop-count. Examples of reactive protocols based on hop-count include Dynamic Source Routing (DSR) [34] and Ad hoc On demand Distance Vector routing (AODV) [16]. Whereas, examples for proactive protocols based on hop-count include Destination Sequence Distance Vector (DSDV) [18] and Wireless Routing Protocol (WRP) [23]. There is another type based on optimizing the length, and which includes both reactive and proactive technique, called hybrid protocols. Zone Routing Protocol (ZRP) [22] belongs to the class of protocols, which take the hybrid approach.

The second class of routing protocols include those which consider metrics apart from hop-count [35], for example, stability-based protocols [20, 36]. Stability-based routing protocols select long-lived routes rather than short routes. There have been various proposals for this class of protocols, which are listed in the following paragraphs. As our approach is closely related to this class of protocols we end this section by explaining how our approach is different from earlier proposals of stability-based routing protocols.

Associative Based Routing (ABR) is probably the first protocol in this class of stability-based protocols for MANETs. ABR is based on the rule of associativity, which states that Mobile Host's (MH) association with its neighbor changes as it is moving, and its period of transit is defined by the associativity ticks. The movement is such that after a period of instability, there exists a period of stability, where the MH will spend some time within a wireless cell before it starts again [36]. The threshold where associativity transitions take place is defined by $A_{threshold}$. In simple terms, ABR is based on the idea that nodes which are neighbors for a threshold period are more likely to remain as neighbors for longer time, or less likely to move away. That is, the authors assume that after the threshold period, nodes move with similar speeds and directions and tend to stay together.

Signal Strength based Adaptive routing (SSA) [20] is a routing protocol, which finds routes based on signal stability and location stability. They distinguish links as strongly and weakly connected based on the average signal strength seen on that link by both the nodes, which form the link. Further, the location stability mechanism of SSA biases the routing protocol to choose a link which has lived for a longer time, which is similar to ABR. This location stability mechanism is considered only as a supplement to signal-strength measurements, and performance results were also not encouraging.

The protocol Route lifetime Assessment Based Routing (RABR) [37] is an extension to ABR, which assigns a threshold to the level of associativity, and based on this threshold, it chooses the routes. This protocol again suffers from the disadvantage of having to choose the optimal threshold values.

The third category of our classification is QoS Routing protocols. Routing protocol in general, and QoS routing in particular, is an essential component to realize complete QoS for MANETs. QoS routing informs a source node of the bandwidth availability (or any QoS metric) to the destination in the network. This helps in establishing QoS connection and the efficient support of realtime multimedia traffic. There have been many proposed solutions for QoS routing in MANETs [38–42]. Core-Extraction Distributed Ad hoc Routing (CEDAR) [38] algorithm is the proactive QoS routing for MANETs. CEDAR has three major components. Establishment of core is the first component. Core here refers to a self-organizing network of nodes which carries out majority of routing computations. The second component is the propagation of link states (high bandwidth availability) to core nodes. The final component is the route computation, which is carried out by core nodes. Proposal by lin et al. [40] is also a proactive based QoS routing for MANETs. They consider non-contention based MAC mechanisms like TDMA or CDMA.

Among the reactive routing protocols, QoS-AODV [41] extends the AODV mechanism to support QoS. This proposal is straight forward, in which every node checks for bandwidth availability before forwarding the request packets. Proposal by Chen et al. [39] also takes the idea of measuring the bandwidth from source to destination by probing, and then carries out the routing. They consider multipath routing and ticket probing mechanisms.

Though there are many proposals for routing in MANETs, the final selection of the protocol is purely based on the various factors like control message overhead, data throughput, delay and storage. In [43] the authors define the factors that should be considered in designing MANET routing protocol.

2.2.2 Path and Link Duration Studies

There have been numerous works, which study the impact of mobility on the performance of wireless ad hoc networks. An important part of these studies is the mobility models considered. A mobility model is defined as a set of rules that determine the movement of nodes within the network. The model typically encompasses the movement strategy and the degree of mobility. Movement strategy refers to a set of rules which decides to which target-position a node has to move, and also at which speed it has to move to that target. The degree of mobility denotes maximum speed of node and pause time. Pause time is the amount of time that a node waits between two consecutive movements. A variety of mobility models have been proposed for ad hoc networks [44].

In majority of the earlier works [45–49] , Link Change Rate (LCR) and Link Duration (LD) metrics are used to infer link stability. There are various algorithms, which use locally observable link statistics such as link duration or link change rate to trigger adaptivity in the routing protocol. The LCR [45] metric is defined as the number of communication links forming and breaking between nodes over a given time T. The LD metric describes the lifetime of communication links. Cho et al., [47] show that LCR is not suitable as a metric for link lifetime estimation as its relation with the route lifespan depends on the node density, which may not be uniform in many mobility models. The conclusion that LD is a good unified mobility metric is based on constant velocity (CV) model. This conclusion is mainly because of relation between LD and route lifespan, which they say is invariant of the mobility model used.

Lenders et al., [48] analyze the impact of human mobility on the link and route lifetime of mobile ad hoc networks. They analyze the data gathered from a real ad hoc network of 20 PDAs connected via 802.11b wireless interfaces. They found that the interruptions due to human mobility and collisions/interference have a completely different impact on the lifetime of links and routes. Authors also compared the empirical link lifetime with those obtained by statistical mobility models. The results show that the distribution of the random waypoint and the random reference point group mobility models are close to the empirical distribution.

One of the first studies concerning the analysis of path duration was by Bai et al., [46]. Based on experimental results obtained by simulations, they assume that the lifetime of a path with four or more hops can be approximated by an exponential distribution. The authors, however, do not consider the fit of any other standard distribution. Further, authors do not justify the selection of an exponential distribution with any mathematical validation. To cope with this shortcoming, Han et al., [50] basing their work on Palms theorem, state that, under some circumstances, the lifetime associated to those paths with a large number of hops converges to an exponential distribution. The previous works present a disadvantage as they provide a solution for the analysis of paths which is valid only for routes with a large number of hops. Therefore, their study could not be fully applied to usual ad hoc networks and practical MANET applications where the paths only consist of 1 to 4 hops. The importance of short paths is reinforced because majority of the existing protocols use the minimum hop-count as the metric to select the route in order to reduce the effects of the wireless retransmissions on the performance of the network. The popularity of the exponential fitting, however, has made it a common approximation in works like [51]. Most authors have analyzed path duration by means of empirical results. For instance, [47] have shown that the mean residual lifetime of routes depends on the number of hops as well as on the mean link duration. On the other hand, [52] analytically proves that the average lifetime of a path decreases with its length. An analytical study on this aspect is carried out by Tseng et al. They base the analysis of the route lifetime on a spatial discrete model [53]. This study simplifies a MANET into a cellular network composed of hexagonal cells to compute the path availability. In [54], authors formally describe the distribution function of path duration assuming that nodes move according to a CV model.

The work by Gerharz et al. [55] studies the characteristics of link duration in mobile wireless ad hoc network, and is closest to our work. They found that link durations vary with age and proposed techniques to measure residual lifetime. They then proceed to propose two metrics for selecting a stable link: highest average residual lifetime and highest 75% quantile. From their analysis they found that initially a link's average residual lifetime decreases with increasing current age, and after a threshold, where threshold corresponds to the modal value of link duration distribution, the residual lifetime increases with current age. In our simulations, however, we found that this may not be true in some cases.

Cheng et al., [56] also study the distribution of link lifetimes in ad hoc network. They focus mainly on the factors, which influence link lifetime. They consider the number of mobile nodes, node minimum speed and moving probability as dominating factors that influence link lifetime. From our experiments, we observed that average node density does not display any useful pattern, which can be exploited, across different mobility models. In [56], the authors also mention the possibility of considering route length along with route lifetimes though no algorithm was proposed.

In the "PATHS" analysis [57], the authors study the link durations and path durations. Detailed analysis was carried out by considering the effects of number of hops, maximum velocity and transmission range. In this study, the authors found that maximum velocity and number of hops have an inverse relation with path duration, whereas transmission range has a direct relation with path duration. The authors also mentioned that for higher mobile speeds, path durations can be approximated with exponential distributions. The simulation results on link and path distributions are used to develop a analytical model for path duration.

Yih-Chun Hu et al., [58] explore the cache strategies in DSR and propose some mobility metrics. They found that link-cache strategies are better than path-cache strategies. As one of the link-cache strategy, they propose technique to combine stability value of a link, which is dependent on the usage of the link, and hop-count. They found that this technique though performs better, but is not better than a static scheme of 5 seconds expiration. We will show that our scheme performed better than the best link-cache schemes.

We conducted a detailed performance study of a stability based routing protocol (ABR). This study compared hop count and stability based routing mechanisms, and showed that stability-based routing can be advantageous. Further, this study also showed that stability based routing can be advantageous if it is part of any resource management solution. For the comparative study we use AODV and ABR as hop-count and stability based routing protocols, respectively. We consider various parameters: throughput, delay, energy consumption, overhead. We also compare the performance of both AODV and ABR, along with service differentiation mechanism SWAN [59]. This study was useful in understanding the advantages and disadvantages of using stability based routing over hop count based routing, and helped us to develop SHARC. The details of the study can be found in the Appendix 2.A at the end of this chapter.

Existing proposal for stability-based routing includes assumptions, for example, in ABR the assumption is longer the existing lifetime longer the link will tend to exist, such assumptions may not be true in many cases. Such assumptions may be traced to lack of detailed study on link lifetimes. Further, purely stability-based routing, like ABR and SSA, have tendency to choose longer routes, which in some cases may prove disadvantageous. Finally, majority of the existing stability-based routing proposals involve a threshold value (associativity and signal-strength). It is difficult to have a single threshold value across different mobility patterns. Therefore, both the assumption and threshold value hinders the operation of routing mechanism when heterogeneous mobility patterns are considered.

Considering these shortcomings, we first begin with a detailed link lifetime study. Based on this lifetime study, we propose a route computation mechanism termed as SHARC - Stability and Hop-count based Approach for Route Computation.

2.3 Study of Link Lifetime

The approach we take in carrying out the link lifetime study is similar to the approach taken in the field of reliability engineering. In reliability studies, engineers study the probability that a system, (vehicle, machine, device) will perform its function for a specified period [60]. This study includes studying the lifetime of the entity considered, and various parameters that affect the lifetimes. We believe that a link between two nodes can also be one such entity, and we can take a similar approach in studying the link lifetime. We also study the process that affect the link lifetimes. Approach taken in these studies can be depicted as shown in Figure 2.2.

Traditionally, a statistical study on entity-lifetimes begins with collecting failure-time data (link lifetime data). This data is also referred to as failure data and the process is referred as "time-to-failure" measurements [61]. Along with failure data, degradation data is also collected whenever it is available. Degradation data is the measure of degradation (process that leads to failure) of the entity considered over a period [62]. Fortunately, for wireless links both the failure data and the degradation data can be obtained.

The next step in the process is to associate a parametric statistical model to describe a set of data or a process that generated the set of data. Reasons behind collecting lifetime data and associating statistical model are [60]: (1) studying the characteristics of the entity considered over a period (2) studying system stability and making estimations (durations) (3) studying the causes of failures and method to improve the reliability (4) comparing different environments under which the entities operate (5) checking the veracity of the performance claims.

Typical procedure in associating the statistical model involves two steps. First step uses the failure data and maximum likelihood (ML) approach, whereas, the second step uses the degradation data and carries out degradation analysis. The method of maximum likelihood is most popular method used for fitting statistical models to data [60]. Research [63] has also shown that under regular conditions, ML estimators are optimal when the samples are large.

From the point of view of statistical studies, ML estimation and degradation analysis can be categorized as enumerative study and analytic study, respectively [62]. Typically, enumerative study begins by collecting and carefully evaluating the samples, and further making an inference about the population from which the samples were collected. Whereas, analytic study answers questions about processes that generate samples over time. Together, they enhance the accuracy of lifetime distribution model estimation.

Once a statistical model is associated with the lifetime data, we carry out simple analysis of the model considered and also explore the possible applications of the associated model. The remaining part of this section is organized according to the link lifetime study process shown in Figure 2.2. Before we begin the description of the process, however, we first present the details about the considered mobility patterns and the simulation environment.

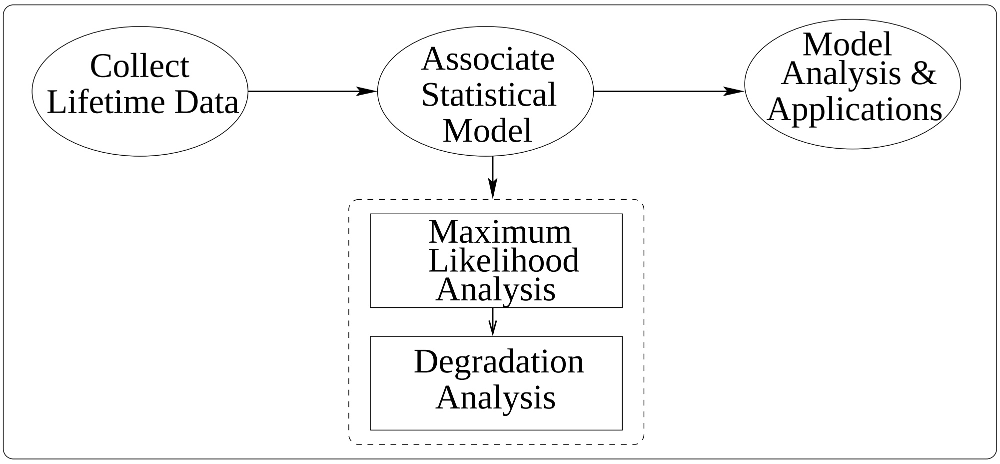


Figure 2.2: Link lifetime study process

Mobility Patterns

Node mobility is one of the most important characteristics of MANET. There have been various mobility models or patterns proposed for MANETs. These patterns try to capture most of the common mobility patterns, but few patterns capture realistic movements of nodes in MANETs. There have also been works that study various mobility models, and the performance of routing protocols across different mobility models. In our work, we use the following mobility patterns: random waypoint (RWP), Reference Point Group Mobility

(RPGM), freeway mobility and Manhattan mobility models [64]. We mainly concentrate on random mobility model (RWP) and group mobility model (RPGM) for associating statistical models. These four mobility models are chosen also considering the framework termed as IMPORTANT, and proposed in [46]. The IMPORTANT framework defined protocol independent metrics such as the average degree of spatial dependence, average degree of temporal dependence, average relative speed and geographic restrictions to capture the mobility characteristics. Mobility characteristics they include are spatial dependence, temporal dependence and geographic restrictions. With an extensive study, authors describe that it is important to make sure that the mobility models chosen (for simulation studies in MANETs) span all the mobility characteristics described in the framework. Further, they show that this set of mobility models (random waypoint, RPGM, freeway and Manhattan) satisfy these characteristics. Hence, along with the popularity and the simplicity of the models, we select these mobility models based on the recommendations of [46].

Simulation Environment

We use different mobility pattern generators for different mobility patterns. In addition, a single mobility pattern is generated using three different tools so that when we draw a conclusion, the probability of the conclusion being correct is high. We use the "setdest" tool, which comes along with the distribution of NS-2 [11], mobility generator obtained from the Toilers group [65] and the tool from the bonnmotion group [66] to generate random waypoint mobility pattern. Similarly, mobility generator obtained from the Toilers group [65], the tool from the Nile group at USC [67] and the tool from the bonnmotion [66] group is used to generate the group mobility (RPGM) patterns. For freeway and Manhattan mobility patterns, the generators from the Nile group at USC [67] and the bonnmotion [66], respectively, are used. For the group mobility models, we consider three cases. In the first case, termed as RPGM1, we have a single group with 50 nodes. In the second case, termed as RPGM2, we consider 5 groups with 10 nodes each. Finally, in RPGM3, there are 10 groups with 5 nodes each.

We have considered 50 nodes, with each node having a transmission range of 250 m. The simulation area is 1000 m x 1000 m. The simulation duration is for 1000 secs. To remove the effects due to traffic, we do not consider any traffic between the nodes. Therefore, link breaks are predominantly due to the mobility. For the study of link durations and residual lifetimes, only the speeds of 1 m/s and 10 m/s are considered. This simulation environment is used in all simulations in subsequent sections. Other specific simulation parameters, or any modifications would be mentioned in corresponding sections. All simulations are carried out for 1000 seconds, and we take the average of 5 to 7 runs unless stated otherwise.

2.3.1 Collection of Lifetime Data - Lifetime Duration Distribution

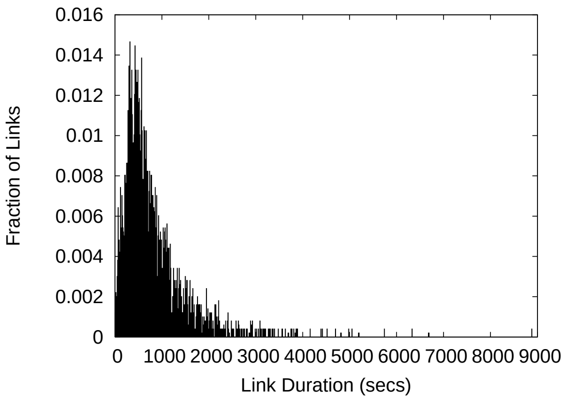
In this part, we collect the link lifetime duration data and plot the histogram of lifetimes. Further, with the same data we plot the cumulative distribution of residual lifetime.

Link lifetime duration is calculated as the duration of continuous connection time between a node and its neighbor. In order to remove any edge effect, a link duration is considered only when the link is broken before the end of the simulation. We look at the probability density function (PDF) of these durations using a bin size of 10 seconds.

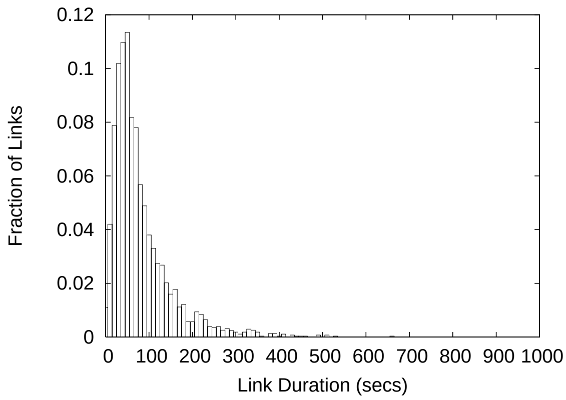
The results shown in this section can be categorized as follows. Random waypoint (high and low speeds), RPGM (with three classes)- RPGM1 (single group of 50 nodes), RPGM2 (5 groups of 10 nodes each) and RPGM3 (10 groups of 5 nodes each), heterogeneous - few nodes follow group mobility and few nodes follow random mobility (high and low speeds). For heterogeneous case we consider nodes with different speeds along with different mobility models. The speeds considered for all the mobility patterns are low speed (0.1 m/s - 1 m/s), varying speed (1 m/s - 10 m/s) and high speed (9.0 m/s - 10 m/s).

Lifetime Distribution: Random Waypoint

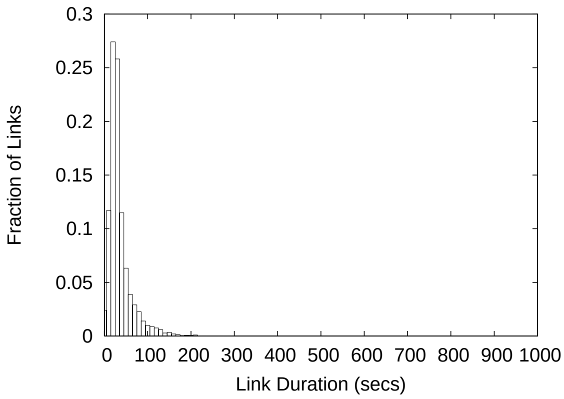
Figure 2.3 shows the plots for link lifetime distribution for random waypoint model. Figures 2.3(a), 2.3(b) and 2.3(c) shows the distribution for low speeds (maximum speed is 1 m/s), varying speeds (1 - 10 m/s) and high speeds (10 m/s), respectively. To remove the effects of short simulation time for low speeds (1 m/s), we conducted experiments for 9000 secs. We can see that all the plots exhibit similar behavior: unimodal and positively skewed. The modal values of link durations is a significant property in these plots. The modal values tend to decrease with increase in speeds. As the speed increases, the duration at which the 99 percentile value occurs also decreases. At a speed of 10 m/s, link durations above 500 secs are rare. Gerharz et al. [55] show in their link duration study that the histogram's peak (modal value) occurs roughly at the transit time of two mobile nodes crossing each other's transmission range. From our results we did not find this pattern in majority of the cases.



(a) RWP 1 m/s, 99% value: 3500 secs



(b) RWP 1-10 m/s, 99% value: 310 secs



(c) RWP 10 m/s, 99% value: 110 secs

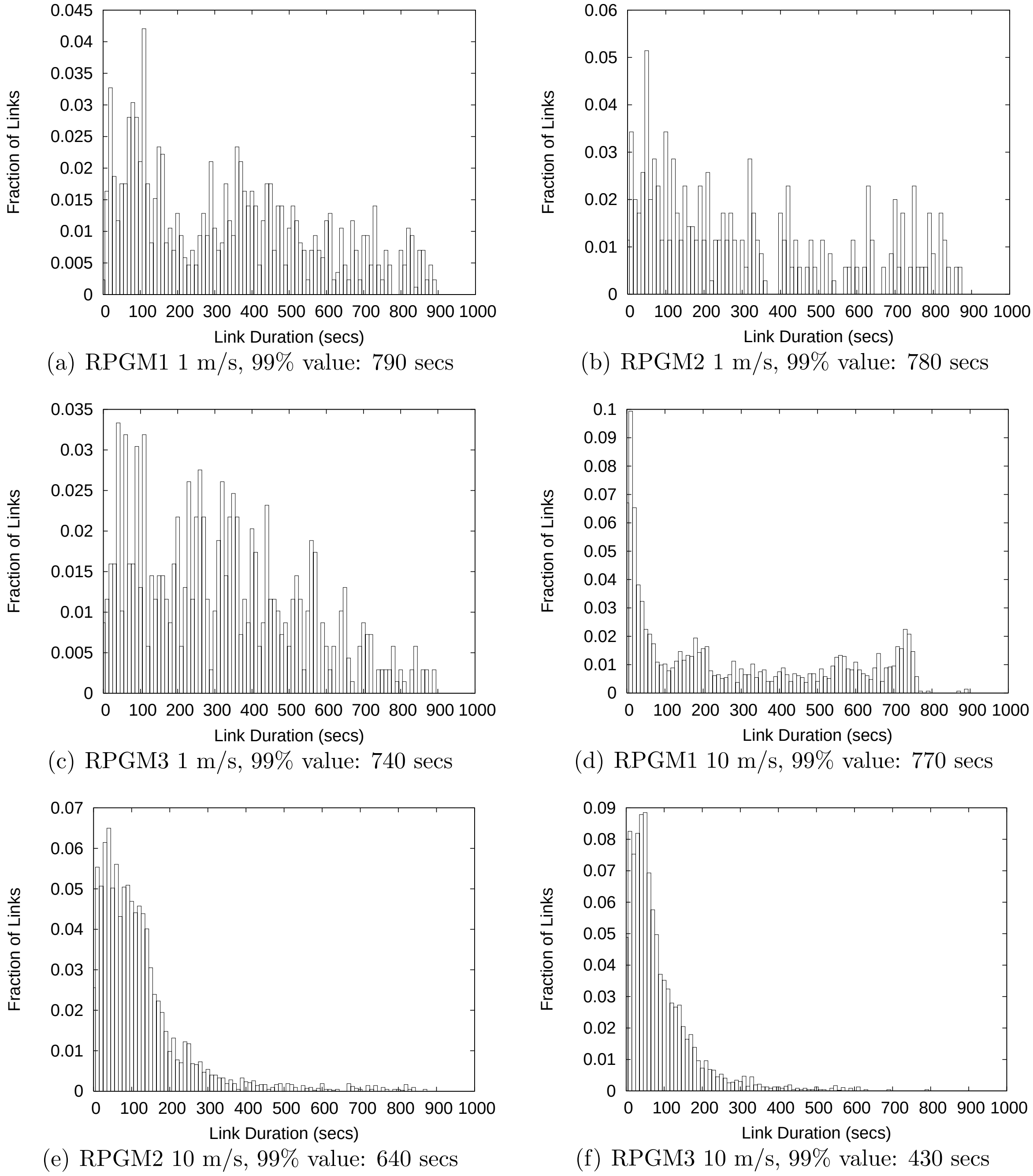
Figure 2.3: Random waypoint lifetime distributions

Lifetime Distribution: RPGM

Figure 2.4 shows the plots for link lifetime distribution for group mobility model. Figures 2.4(a) 2.4(b) 2.4(c) shows PDF for low speed (1 m/s) for RPGM1, RPGM2 and RPGM3, respectively. Similarly, Figures 2.4(d) 2.4(e) 2.4(f) shows PDF for high speed (10 m/s) for RPGM1, RPGM2 and RPGM3, respectively. The plots show that group mobility patterns have longer tails compared to random waypoint scenarios. Therefore, there are higher fraction of links with longer durations. This behavior is natural considering the properties of RPGM mobility pattern. The 99 percentile values are always greater than 400 secs. In fact, we also found that, even at the speed of 30 m/s, the 99 percentile values are above 300 secs. Regarding the modal values, from RPGM1 at low speed (1 m/s) the peak occurs at 100 secs, whereas for RPGM2 the modal value occurs at 70 secs. According to [55], it should have occurred at roughly 250 secs.

Lifetime Distribution: Manhattan and Freeway

Figure 2.5 shows the plots for link lifetime distribution for Manhattan and freeway mobility models. Figures 2.5(a) and 2.5(b) shows the distribution for Manhattan mobility with low speeds (maximum speed is 1 m/s) and high speeds (10 m/s), respectively. Similarly, Figures 2.5(c) and 2.5(d) shows the distribution for freeway mobility for low speeds (1 m/s) and high speeds (10 m/s), respectively. We can see that even for Manhattan mobility model all the plots exhibit similar behavior: unimodal and positively skewed. These plots have comparatively longer tail than random waypoint. Even in this category, similar to random waypoint, link durations above 500 secs are rare. Similarly, as the speed increases, the



(a) RPGM1 1 m/s, 99% value: 790 secs

(b) RPGM2 1 m/s, 99% value: 780 secs

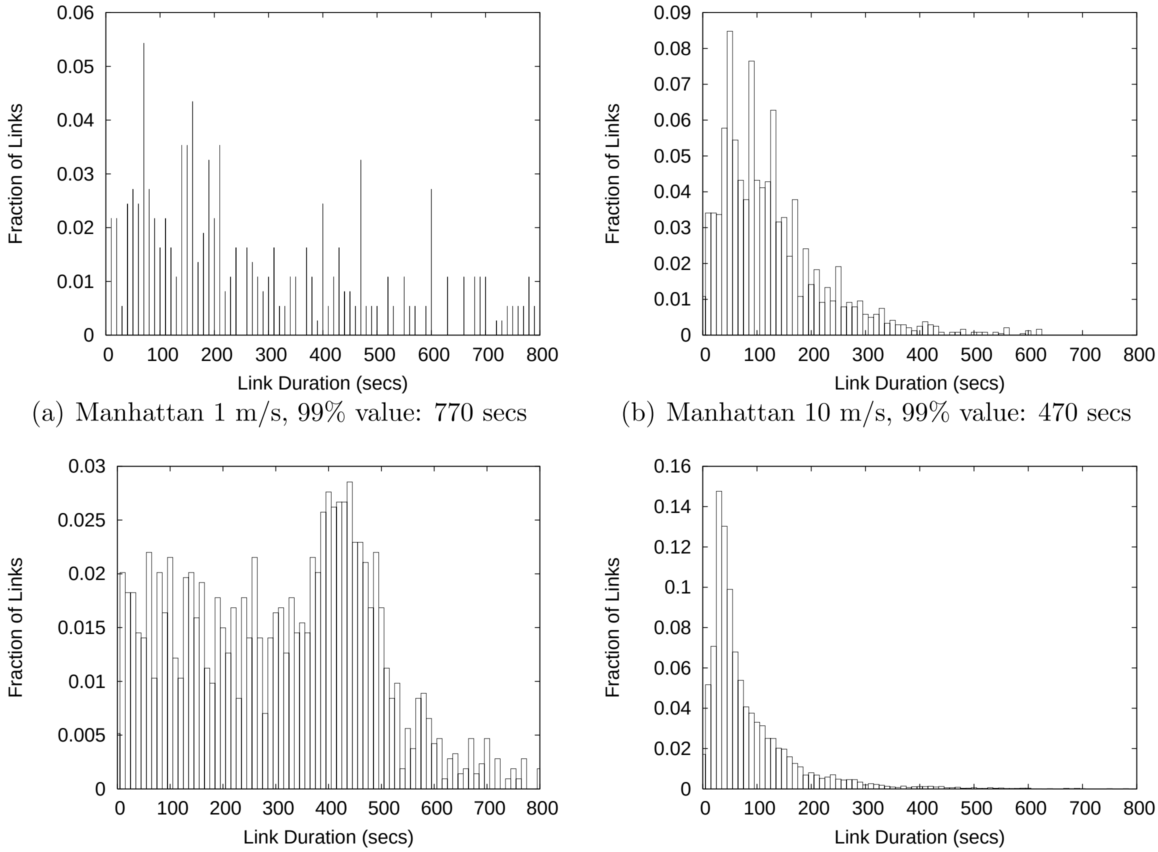
(c) RPGM3 1 m/s, 99% value: 740 secs

(d) RPGM1 10 m/s, 99% value: 770 secs

(e) RPGM2 10 m/s, 99% value: 640 secs

(f) RPGM3 10 m/s, 99% value: 430 secs

Figure 2.4: RPGM lifetime distributions



(a) Manhattan 1 m/s, 99% value: 770 secs

(b) Manhattan 10 m/s, 99% value: 470 secs

(c) Freeway 1 m/s, 99% value: 730 secs

(d) Freeway 10 m/s, 99% value: 420 secs

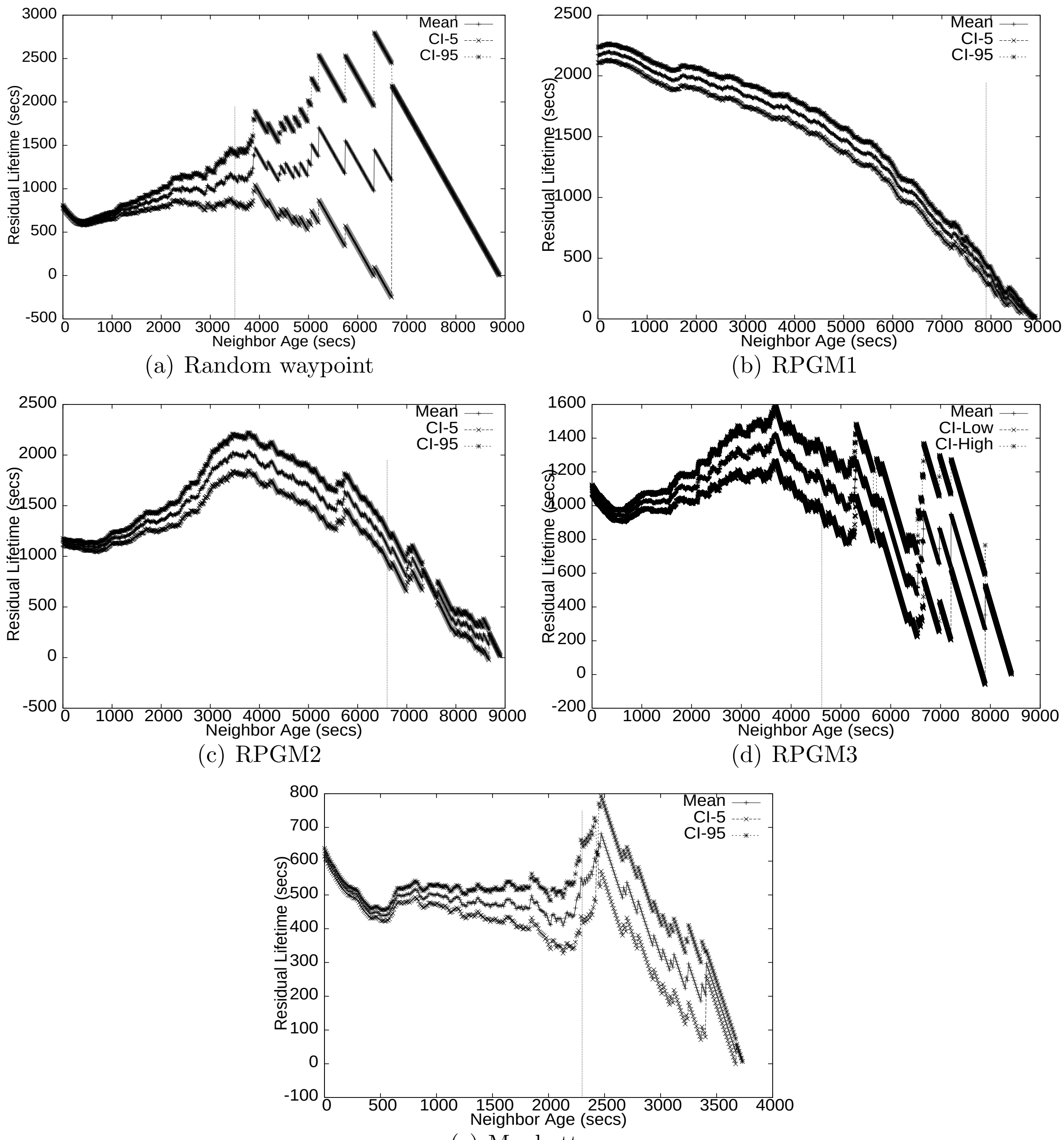
Figure 2.5: Manhattan and freeway lifetime distributions

duration at which the 99 percentile value occurs decreases.

Residual Lifetime

The collected link duration values are used to calculate the residual link lifetime. The residual lifetime value is computed as follows. Let l_i be the number of links with link duration i secs and R_a be the average residual link lifetime when the current link age is a.

$$R_a = (\sum_{i>a}(l_i * i)/\sum_{i>a} l_i) - a; \tag{2.1}$$

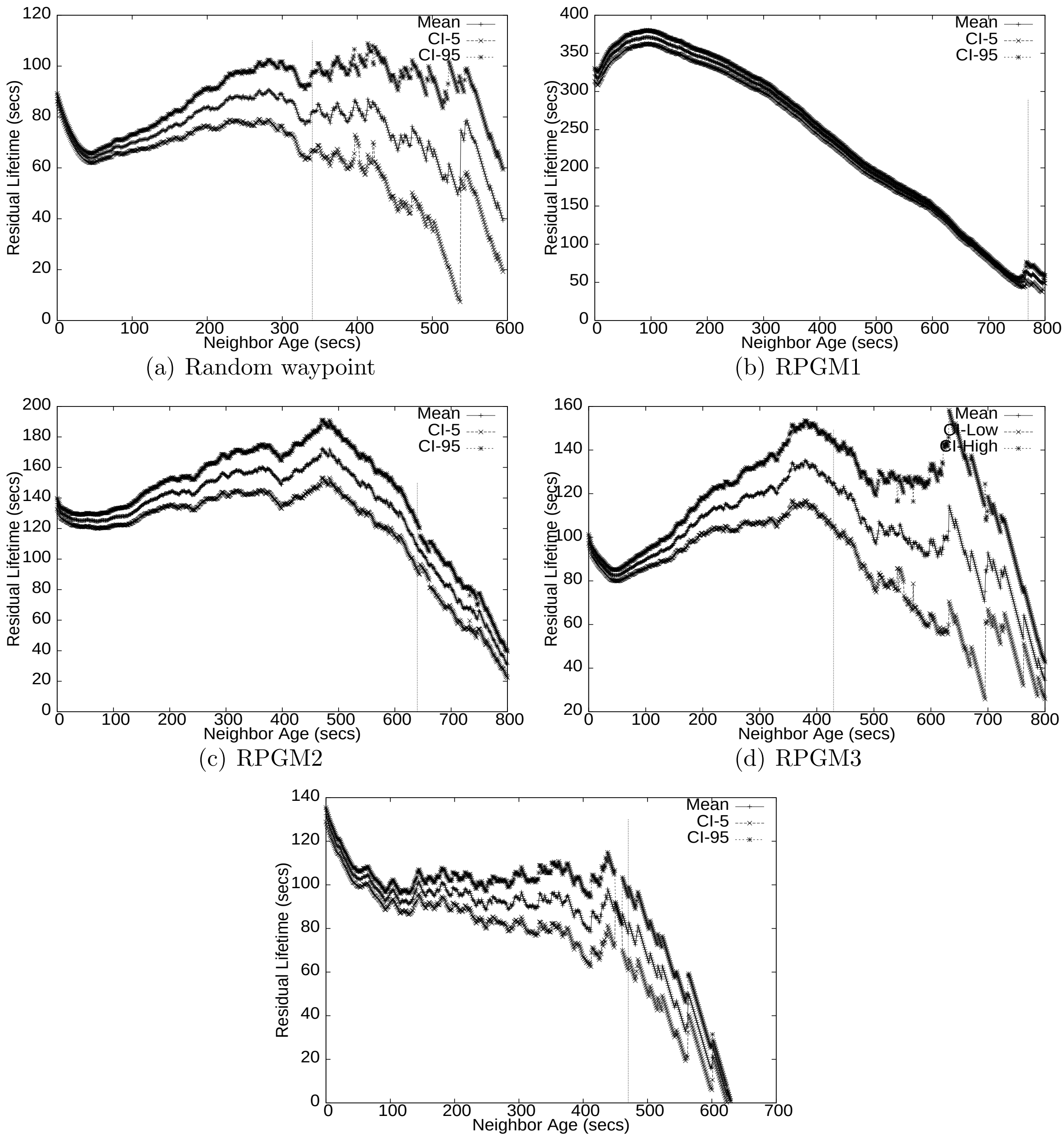


(a) Random waypoint
(b) RPGM1
(c) RPGM2
(d) RPGM3
(e) Manhattan

Figure 2.6: Residual lifetime, speed 1 m/s

In other words, the residual lifetime for a link of age a is the average lifetime of all links with durations above the age a, minus age a.

Figures 2.6 and 2.7 show the residual lifetime plots for speeds of 1 m/s and 10 m/s, respectively. We also show the 5% and 95% confidence interval values for each link age in



(a) Random waypoint
(b) RPGM1
(c) RPGM2
(d) RPGM3
(e) Manhattan

Figure 2.7: Residual lifetime, speed 10 m/s

the plots (labeled as CI-low and CI-High, respectively). Using results from the preceding section, we can also obtain link durations corresponding to the 99 percentile values. The obtained values are indicated in the Figures 2.6 and 2.7 as a vertical line. For example, in Figure 2.7(b), the vertical line is at 770 secs, which corresponds to Figure 2.4(d). In order to ensure that the simulation time is sufficiently long, the simulation duration for the 1 m/s and 10 m/s cases are 9000 secs and 900 secs, respectively.

From Figure 2.6, we see that for low speed (1 m/s), the residual lifetime decreases with an increase in age for the case of RPGM1. For the RPGM2 and random waypoint, the residual lifetime initially decreases, and then increases after some time. Finally, for the Manhattan model, the residual lifetime decreases and remains constant after some threshold. In all cases, for neighbors with sufficiently long lifetime, the residual lifetime decreases again. This final decrease, however, occurs for less than 1% of the links and is not considered important. At higher speeds (10 m/s and 30 m/s), the mobility models exhibit similar patterns.

Earlier work [55] has also noted a similar behavior where there is an initial decrease and then an increase. They do not consider the later decrease in link lifetime. At higher speeds, apart from RPGM1, all model confers with patterns of previous work: initial decrease and then further increase.

Based on the results obtained, we conclude that the heuristic of existing stability-based routing algorithms, for example Associativity Based Routing (ABR), of assuming that *older links are more stable* does not hold across a large spectrum of mobility speeds and models. In fact, in RPGM1 and Manhattan, the reverse is true. *Newer links are more stable.* In cases where the heuristic is correct, it is difficult to obtain a good estimation of the threshold when residual lifetime starts increasing, as the threshold depends on many factors including speed and mobility pattern. Even when this threshold is available, the likelihood of finding links with long residual lifetime may also be low, making the heuristic less useful for route selection.

In the succeeding section, we describe the next step after the collection of link lifetime data: the process of associating a parametric statistical model to the collected lifetime data.

2.3.2 Associating Parametric Statistical Model for the Lifetime Data

Statistical models helps in giving a definition of the target process or population. For many applications it will be useful to fit one or more parametric models to the data, mainly from the point of view of description, estimation and prediction [62]. The problem that we consider in this part can be described as follows. Suppose we have a random sample $X_1, X_2, \ldots, X_n$ of a parent random variable X, with distribution function F. Now it is required to decide that F is a member of one of a set of parametric families of distribution functions, say $F_1, F_2, \ldots, F_k$ [68]. In other words, we have to decide which of these k families best fits the sample.

Typical distribution models used for lifetimes are exponential, Weibull, lognormal and gamma [62]. The details about these models and descriptions of these models considering the failure rates and the failure modes are provided in the Appendix 2.B at the end of this chapter.

Selection of Model - Maximum Likelihood Analysis

In some circumstances, *physical considerations* alone can identify the appropriate family of distributions [69]. For example, when there is "no premium for waiting" and "lack of memory" property, the family of exponential distributions can be the ideal model. In practice, however, there may not exist any such considerations which would provide clue about the appropriate model to select. In such cases a choice of the model has to be made, and such a choice may be based upon mathematical analysis or on an understanding that a particular family is "rich" enough to include a good fit to the data [60]. There are few techniques, which help in making the choice. One useful approach is to choose two or more possible candidate parametric families of distributions, and then use the data to select the appropriate model. One such technique is maximum likelihood estimation [60], where for every alternative family under consideration it maximizes the likelihood over the parameter values, and selects the family that yields the largest maximum likelihood. In our work, we use minimum Kolmogorov distance method [70], which is similar to maximum likelihood method. The additional process involved here is a "Kolmogorov distance" between the specific candidate and the empirical distribution is determined, and a family is chosen which yields minimum distance. There have been numerous studies [69] in the field of reliability engineering, which uses Kolmogorov distance technique.

Many researchers [71] have suggested that replicated run experiments are the surest guide to the distribution of failure times. Therefore, we have performed 15 replicated experiments for every case. The distribution from the data is obtained using a curve-fitting software called Easy-Fit [72]. In this software, the selection procedure used is minimum Kolmogorov distance method.

When alternative families have different numbers of parameters, the appropriateness of the methods is unclear because the family with the greatest number of parameters would perhaps have an unfair advantage. In this regard, we have chosen all the families with equal number of parameters (number of parameters being 2).

Distribution Model for Random Waypoint

Considering the Figure 2.3, we can notice that many distributions - Weibull, gamma and lognormal can exhibit this behavior, depending on the parameters. For all the plots, considering Kolmogorov minimum distance technique, the best fit happens to be lognormal distribution.

We know that the shape of the lognormal is affected by the values of both μ (scale parameter) and σ^2 (shape parameter). The density is more spread for higher values of μ, whereas, it is more skewed (towards left) for higher values of σ^2. It was seen that for low speed, the parameters μ and σ (considering lognormal) were 6.3757 and 0.8181. Whereas, for high speeds the parameters were 4.163 and 0.7253. A small (less than 1) shape parameter indicates a narrow range of failure times and implies few early failures will occur. Also, large scale parameters implies a longer mean time to failure (MTTF) [62,68]. Further, exponential (which indicates random failures) is not a special case of lognormal distribution [62]. Hence, from these parameters we can conclude that the failure rates follow the wear-out type of failure (as it is neither infant mortality nor random failures). Random failures mean that failures are independent of time (failure modes are ageless) [60]. Whereas, for wear-out failures, the lifetimes are dependent on the current age, and the link wears out rather than experiencing a random breaks. From the plots for random waypoint, we can see that the lifetimes 0 - 400 (for low speeds) and 0 - 40 (for low speeds) are not the modal class. This also shows that the exponential distribution is a poor fit to the considered data.

Researchers [69] have also shown that with lognormal data and with lognormal and Weibull as alternatives (or vice versa), the probability of choosing the correct distribution (using minimum Kolmogorov distance) is closer to 1 with sample size greater than 100. This shows that, probability that the data generated by the experiments on random waypoint models are lognormal is very high [69].

Distribution Model for RPGM

Considering Kolmogorov minimum distance technique, the best fit happens to be Weibull for RPGM low speeds (1 m/s) and RPGM1(single group of 50 nodes) high speed (10 m/s). Whereas, the best fit is again lognormal for high speeds RPGM2 and RPGM3.

It was seen that Weibull distribution fits well for low speed RPGM mobility models. It was also seen that the β (shape parameter) value was around 1.5 to 2.5. This parameter gives clue about the failure mechanism, since different slopes (β's), imply different classes of failure modes. For Weibull distribution, if the shape parameter is less than 1 then the failure mode is infant mortality, if the parameter is equal to 1 it is random failure, and if it is greater than 1 the failure mode is wear-out [62, 68]. From the values of β obtained, we can conclude that *the failures are purely wear-out failures rather than random failures.*

We can also show that the best-fit model is indeed Weibull, by considering the alternatives. Earlier works [69] have shown that with exponential or Weibull data, and with exponential, gamma and Weibull as alternatives, selection of any of the three families can be regarded as "correct". This is true, however, when the shape parameter of the Weibull distribution is very high [69].

In RPGM cases, we have seen that the shape parameter (β) is greater than 1, and in such cases the probability of choosing the correct distribution is high. That is, with

Distribution	PDF	Estimated Cases
Exponential	$\lambda e^{-\lambda x}$	
Lognormal	$\frac{1}{\sqrt{2\pi}\sigma x}\ exp\ \left\{-\frac{(ln\ x-\mu)^2}{2\sigma^2}\right\}$	RWP, RPGM2/3 - High Speed
Weibull	$\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta}\ exp^{-\left(\frac{x}{\alpha}\right)^{\beta}}$	RPGM1, RPGM2/3 - Low Speed

Table 2.1: PDF and estimations of different distribution models

Weibull data, and with Weibull and exponential as alternatives, when the shape parameter of the parent Weibull distribution is higher than 1, then the probability of choosing correct distribution (using minimum Kolmogorov distance) is very high with the sample size greater than 100 [69].

Therefore, it is valid to assume that the probability, that the distributions chosen (Weibull for lower speeds RPGM and lognomal for higher speeds) is correct is very high (closer to 1). The Table 2.1 summarizes the distribution models, their density functions, and the best-fit cases (estimated).

We can also conclude that, for RPGM low speed, the link failures are linear with respect to time (Weibull). For random waypoint, the link failures are multiplicative and progressive (lognormal). Across all mobility models, link exhibits wear-out failures rather than random and infant-mortality failures.

In reliability engineering, it is important to consider the cause of failures (understanding the factors leading to failure) from the point of view of reducing the probability of a failure, thereby improving reliability [62]. In our work, we study the failure process mainly to support the claims we have made regarding the statistical models for link lifetime under different mobility models. The succeeding section makes an attempt to explain the reason behind the associated statistical model. We consider random waypoint mobility pattern as

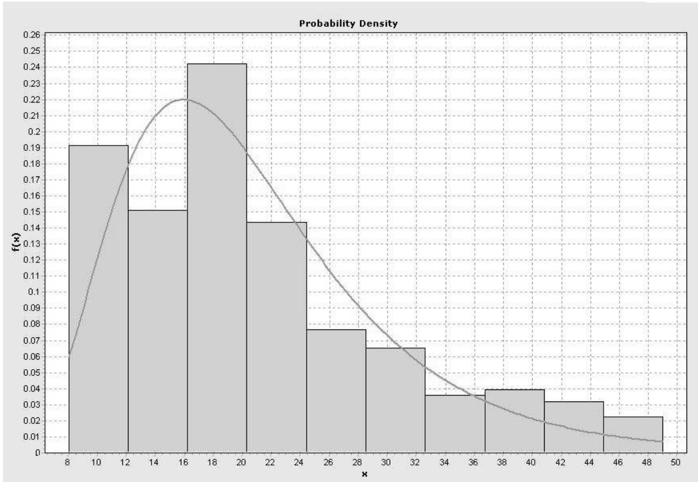


Figure 2.8: PDF of lifetimes considering 2 nodes

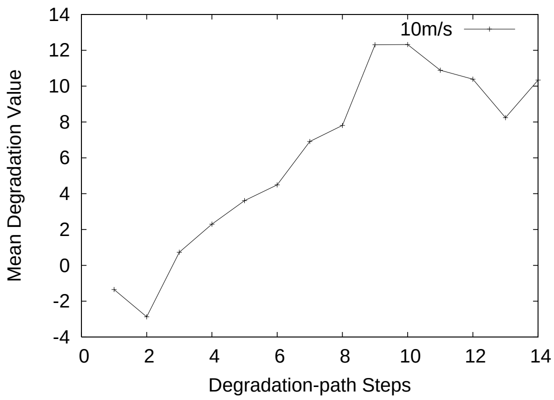


Figure 2.9: Aggregate degradation path

a case study, and similar analysis can be carried out for RPGM mobility pattern.

Link Degradation Analysis

In our work, the term degradation means how a node gradually moves out of other node's transmission range. Therefore, the degradation level is measured in terms of distance between the nodes (m). Another way to represent the degradation is by considering the received signal strength. It is known that, if we ignore any neighboring interfering nodes, and an area without any obstacles (as we do in this part of simulation study), even the received signal strength would predominantly depend on the distance between the nodes. Further, this approach is only to understand the degradation process, and we do not propose to incorporate this process in nodes and do not expect nodes to measure any degradation. This is because, in practice, if we do not assume the existence of GPS mechanisms, even the distance measurement would depend on measuring the signal strength. Therefore, to keep it simple, we directly measure the distance between the nodes and attempt to study the degradation process.

Direct observation of degradation mechanism allows direct modeling of the failure-causing mechanism [62]. Most failures can be traced to an underlying degradation process. Failure occurs when degradation crosses a threshold. Possible shapes for univariate degradation curves are: linear, concave and convex. The degradation rate ($\frac{dD(t)}{d(t)} = C$), is constant over time for linear degradation. Degradation level at time t, $D(t) = D(0) + C * t$ is linear in t. For concave, degradation rate decreases with time and the degradation level increases at a decreasing rate. For convex, degradation rate increases with time, and the degradation level increases at an increasing rate [73].

The degradation level, or true degradation path of a particular link (a function of time) is denoted by $D(t),\ t > 0$. In simulations, values of D(t) are sampled at discrete points in time, $(t_1, t_2, \ldots)$. Observed sample degradation path of link i at time t_j is

$$y_{ij} = D_{ij} + \epsilon_{ij}$$

where D_{ij} is the degradation path $D(t_{ij})$ for unit i at time t. ϵ_{ij} describe a combination of measurement and model error.

To understand the degradation process and collect the degradation data, we conducted a different experiment. Only 2 nodes, which are separated initially by 100 m are chosen. The area of the simulation is 1000 m x 1000 m, and other node parameters (like transmission range) remain same. The nodes randomly pick a destination and speed (maximum speed of 10 m/s, same as random waypoint model) and move towards that destination with the chosen speed. This process is considered only once. The simulation is considered for just 50 secs, and this is repeated for 1000 times.

We first plot the PDF of the link lifetimes, which is shown in Figure 2.8. We can see that this distribution is similar to what we obtained for random waypoint with 10 m/s. This is mainly because, nodes choose speed and destination, using the same method/process as used in random waypoint. The best-fit distribution for this again is lognormal (using the minimum Kolmogorov distance technique). The fit-curve is also shown in the Figure 2.8. This plot is obtained using the Easy-Fit [72] software package.

In addition, the process of nodes moving out of each other's transmission range the (link-degradation process) is also noted. Nodes take randomly 9 - 30 time-steps to move out of each others range. At each step we note the amount of degradation, where a single step is equal to 1 second. The Figure 2.9 shows the aggregate degradation path. The mean degradation value indicates the amount of degradation (distance in meters) that occurs at a particular step. The initial decrease in the plot can be attributed to the behavior that nodes tend to move towards each-other in the beginning, with high probability. Once the nodes begin to move away from each other, they continue to move away from each other until they go out of each other's transmission range (with high probability). We can see that till the 8th step the increase in mean degradation range from 1 - 2 m. Whereas, from 8th step to 9th step the increase in mean value is by 4 m. This can be attributed to the lognormal behavior. There is a decrease in mean degradation value after the 10th step, which can be explained as follows. In this simulation, we can classify the links based on how quickly they degrade. All the links which broke before 9th step falls under the class of quickly degrading links. When we compute the mean, we consider links where the corresponding nodes are still within each other's transmission range, and once the nodes move out of the each other's transmission range the links are not considered for the mean computations. As a result, after the 9th step, all the quickly-degraded links are not considered for the mean computation, and only the existing links are considered. Hence there is a decrease in the mean value. In the next section, we will model the degradation process analytically to justify the lognormal behavior of the link lifetimes under random waypoint mobility scenarios.

Analytical Evaluation of F(t)

If T is a random variable describing the failure time of a link, then the failure (probability distribution of failure time) can be written as:

$$Pr(T \leq t) = F(t) = Pr(D(t, \beta_1, \beta_2, \beta_n) \leq D_f)$$

D_f is the threshold degradation level (which is 250 m in our experiment).

Variability causes links between nodes to fail at different times. A degradation model should account for the important sources of variability in a failure process [62]. Considering the link degradation process, and the underlying reasons behind the link degradation, we can represent degradation path of a particular link as [73]:

$$D(t) = \beta_1 + \beta_2 * t$$

β_1 is the initial amount of degradation, and in our previous experiment the value of β_1 is fixed (100m). We know that the distance between two nodes, when nodes are moving, is largely dependent on the node speeds and directions. Therefore, we consider β_2 to be as the reciprocal of relative velocity. Further, we know that failure time is proportional to reciprocal of relative velocity times $D_f - \beta_1$. Let $v(n, t)$ and $\theta(n, t)$ be speed and direction of node n at time t, then magnitude of relative velocity is: $RV(i, j, t) = \sqrt{a^2 + b^2}$, where $a = v(i, t)cos\theta(i, t) - v(j, t)cos\theta(j, t)$ and $b = v(i, t)sin\theta(i, t) - v(j, t)sin\theta(j, t)$.

Therefore, $F(t)$ of link lifetime, in terms of $D(t)$ the link degradation can be written as [73]:

$$F(t; \beta_1, \beta_2) = Pr(D(t) > D_f) = Pr(\beta_1 + \beta_2 * t > D_f) = Pr(\beta_2 > \frac{D_f - \beta_1}{t})$$

Now, to show that F(t) is log-normal we have to show that right hand side (RHS) is lognormal, or β_2 has lognormal rate.

We found that the reciprocal of relative velocities between two nodes moving at random speeds follow lognormal distribution. This is obtained by using the Easy-Fit [72] software package and Kolmogorov minimum distance technique [69]. The cumulative distribution (CDF) of the relative velocities between two nodes are as shown in Figure 2.10. The initial

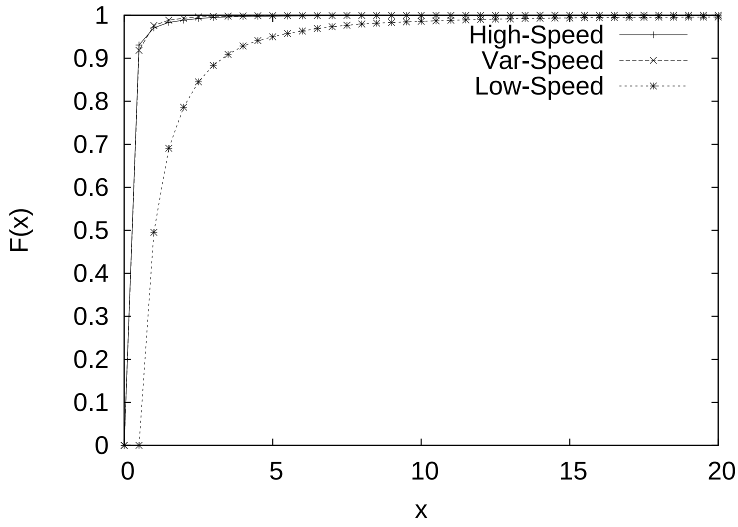


Figure 2.10: CDF of reciprocal of relative velocity

separation (β_1) is assumed to be 100 m. The three plots shown in the Figure 2.10 are for high (9-10 m/s), varying (1 - 10 m/s) and low (0.1 - 1 m/s) speeds.

Therefore, we can write β_2 of the previous equation as varying from link to link according to $LOGNORMAL(\mu, \sigma)$. This implies that:

$$Pr(\beta_2 \leq b) = \Phi\left[\frac{log(b) - \mu}{\sigma}\right]$$

where $\Phi(z)$, is the standard normal CDF and μ and σ are the mean and standard deviation of $log(\beta_2)$, respectively. Substituting this in previous equation,

$$F(t;\beta_1,\mu,\sigma) = 1 - \Phi\left[\frac{log(D_f - \beta_1) - log(t) - \mu}{\sigma}\right]$$
$$F(t;\beta_1,\mu,\sigma) = \Phi\left[\frac{log(t) - log(D_f - \beta_1) - \mu}{\sigma}\right], t > 0$$

This shows that T has lognormal distribution with parameters that depend on the basic path parameters. Hence, we have shown that the probability that the link lifetimes exhibiting lognormal behavior for random waypoint scenarios is high. Further, this lognormal behavior can be attributed to the relative velocities of the neighboring nodes moving at random speeds.

2.3.3 Model Analysis and Application

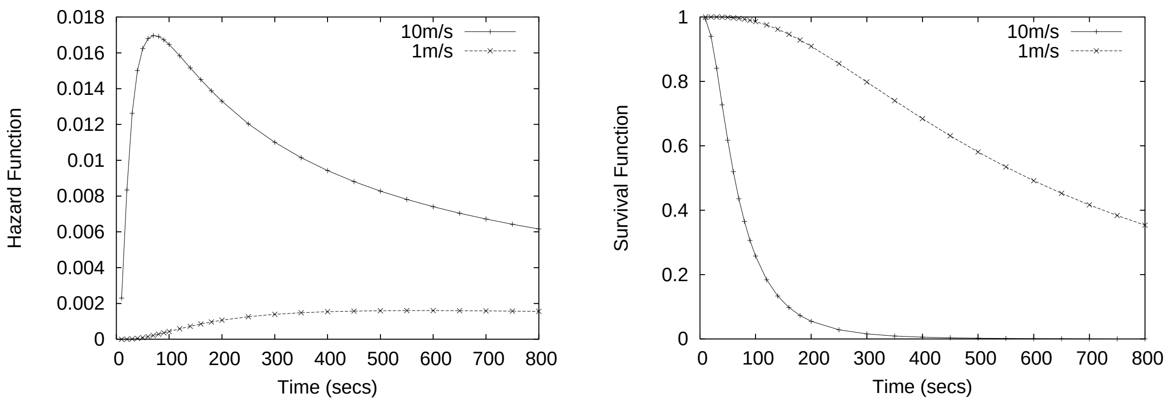
In the previous part of this section, we collected the link lifetime data and associated a parametric statistical to the data. In the remaining part we assume that the associated model to the data is correct and analyze some features, and explore a few applications of the model association.

Analysis

Analysis in reliability engineering typically aims to answer questions like: what is the probability that a link will sustain beyond some time t?, and what is the probability that a link will break in the next instant, given that it has survived to the time t. The answers to these questions are obtained by expressing the earlier distribution functions differently. Such different representations are survival and hazard functions.

We carry out the analysis considering the lifetime distribution model for random waypoint (lognormal). Apart from probability density function, given in Figure 2.3, typically probability distribution for failure time is also characterized by the hazard and survival functions. We consider the parameter values as given in the preceding section, and plot the hazard and survival functions as shown in Figures 2.11(a) and 2.11(b), respectively.

The hazard function (hazard rate, instantaneous failure rate) of the link is defined as the rate of change of the cumulative failure probability divided by the probability that the



(a) Hazard function - random waypoint (b) Survival function - random waypoint

Figure 2.11: Hazard and survival functions

link will not already be failed by time t [74]. That is:

$$\lambda = \frac{dF(t)/dt}{1 - F(t)} = \frac{f(t)}{1 - F(t)}$$

The hazard function of a lognormal process is defined by [61]

$$\lambda(t) = \frac{\phi(d)}{t\sigma\Phi(-d)}$$

where d is given as $\frac{\log_e t - \mu}{\sigma}$, ϕ is the probability density function of the standard normal distribution and Φ is the cumulative distribution function of the standard normal distribution.

The survival function ($R(t)$) is the probability that the time of failure is later than some specified time. It is also termed as reliability function [74]. Since a link either fails or survives, and one of these two mutually exclusive alternatives must occur, we have $R(t) = 1 - F(t), \quad F(t) = 1 - R(t)$. Therefore, the value $R(x)$ of the survival function at the point x gives the probability of survival beyond x.

Considering the Figure 2.11(a), the initial part the hazard function is concave for higher speeds, indicating the increase in failure rate (at a decreasing rate) as the time increases. For lower speeds, the hazard function still increases with time, but the increase is less.

After 100 secs for higher speed, however, the hazard function decreases with respect to time (approaching to 0 as $t \rightarrow \infty$). For higher speeds, such large durations are not of interest and lognormal model should be adequate enough to represent lifetimes.

From Figure 2.11(b) we can see that the survival function decreases as the time increases. We can also see that for higher speeds, the decrease in survival function is "considerable" compared to lower speeds. From this we can say that the network consisting of 50 nodes, in a 1 km x 1 km area, having transmission range of 250 m is fairly reliable if nodes are moving at low speeds, and considerably unreliable if they are moving at higher speeds.

In addition to being useful functions for reliability calculations, such analysis provides the information needed for troubleshooting, or classifying failure types. Similar to lognormal distribution (random waypoint and high speed group mobility), we plotted out reliability and hazard functions for Weibull distribution (low speed group mobility). We found that the change in reliability decreases slowly in the initial period and then decreases sharply as the characteristic life is approached. Whereas, failure rate increases (with lesser rate) with time.

Applications

Associating a statistical model to the lifetime data, and model-specific analysis has various applications. Obvious application of such a study is the understanding of the characteristics of the link over a period and across different mobility models. This understanding helps in studying the ad hoc network system stability. Such stability studies helps the network designer in making decisions on important design-parameters (number of nodes, maximum mobility speed). Further, study of the causes of failures helps to improve the reliability of the ad hoc system. Link lifetime study also plays an important role comparing different environments under which the wireless links exist. In addition, statistical model and model-specific analysis also helps in checking the veracity of the performance claims. Finally, statistical models are crucial in making estimations of various lifetime related temporal parameters. In the succeeding section, we will focus on the application of lifetime study in the estimation of residual link lifetime.

2.4 Residual Lifetime Estimation

In this section, we begin with describing the residual link lifetime estimation process. Next, we provide the lifetime estimation results considering RWP and RPGM mobility models. We also considered different combinations of RWP and RPGM, which are described below. Finally, we include the statistical model information from the previous section into the estimation process to reduce the estimation errors.

In this work, link lifetime estimation techniques are based on the heuristics. That is, nodes maintain a collection of link lifetimes, and based on this information it will estimate the lifetime for existing or future links. In this work, we use the term *history* to denote this collection of link lifetimes. There are both pros and cons with this approach. Some of the advantages are: if we assume that the mobility pattern of nodes remain similar, then greater the history present lesser would be the estimation errors. The approach is simple and feasible, and also the overhead of history maintenance is also not high (the amount of history does not exceed 1000 even for high mobility scenario). A disadvantage with history based estimation approach is in deciding how much history is useful, as higher is better is not always true. For example, when the node mobility pattern does not remain same, then estimation errors are bound to be high. Any estimation technique should counter this disadvantage. Another problem with this approach is non-existence or less history of link lifetimes present at the nodes, which can also cause increased estimation errors. The only way to cope up with the latter problem is to have a complete knowledge about link lifetime distribution. In this section, we will describe how our approach caters for both the disadvantages present in heuristic based link lifetime estimation technique.

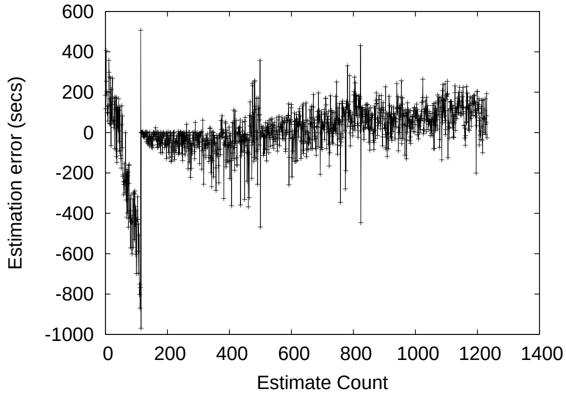
We study three estimation techniques, which are termed as: multi-node, single-node and average. In multi-node technique every node stores the link duration values of its neighbors. By collecting this information and aggregating them into bins of 10 s, each node makes an estimate of the residual lifetime distribution using all the samples collected in equation (2.1). Whereas, in single-node technique every node stores the link durations on a per-neighbor basis. That is, for every other node which is a neighbor, a separate bin of link lifetimes are maintained. Further, during the estimation process same equation (2.1) is used but for any neighbor, only its own history of lifetimes is used. In both the cases, where there does not exist any history, the estimate is chosen from a random distribution. The third technique is simple, it is the average value of single-node and multi-node estimation techniques.

Estimations for Random Waypoint Scenarios

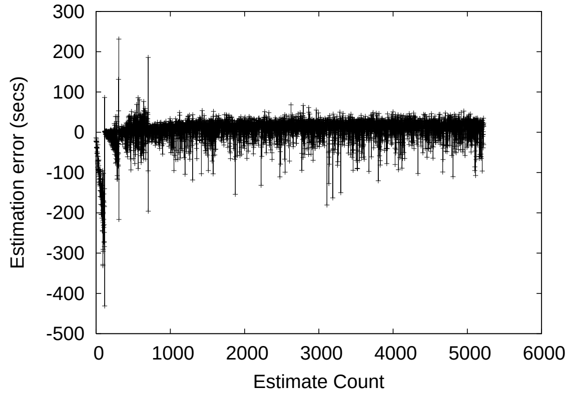
Figures 2.12(a) and 2.12(b) shows the estimation plots for random waypoint, multi-node technique, at low speeds (1 m/s) and high speeds (10 m/s), respectively. Whereas, Figures 2.12(d) and 2.12(e) are for single-node estimation technique. Figures 2.12(c) and 2.12(f) shows the estimation plots for random waypoint, with variable speeds for multi-node and single-node estimation techniques, respectively. The y-axis shows the estimation error in seconds. Positive values indicate that the estimations were high, whereas negative values indicate that the estimations were low. The x-axis shows the estimation count. Estimation count is chosen for x-axis instead of estimation time because multiple link breaks (in turn multiple estimation error values) can occur at any given time (or time interval), as we consider estimations from all the nodes in the network. Considering estimation time would require to consider average estimation error values. Further, this value (estimation count) also helps in understanding and classifying the errors based on the amount of history, which is an important factor in our study. We can notice from the plots that the results are classified into different sections separated by vertical bars. The first (left-most) section shows the result with no history, and the amount of history increases, as the plot moves towards right (1, 5 and greater). We can notice three aspects from the plot. First, for lower speeds, number of estimations having higher amount of history is lesser compared to higher speeds. Second, multi-node estimation technique has better estimations compared to single-node estimation technique. Finally, estimation errors decrease with increasing history.

Estimations for RPGM Scenarios

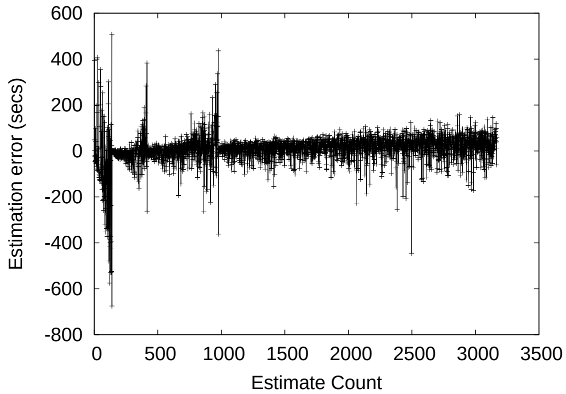
Figures 2.13(a) and 2.13(b) shows the estimation plots for RPGM, multi-node technique, and low speeds (1 m/s) and high speeds (10 m/s), respectively. Whereas, Figures 2.13(c) and 2.13(d) for single-node estimation technique. We only provide results for RPGM3 case. Similar to random waypoint plots, even for these plots we classify the estimation results. From the figures we can see that the estimation errors are similar for RPGM cases compared



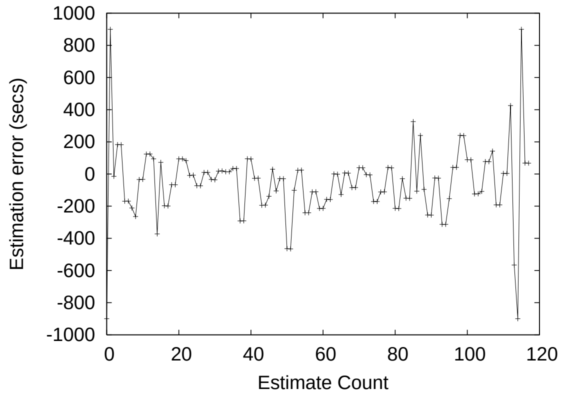
(a) Random waypoint, multi-node, low speed



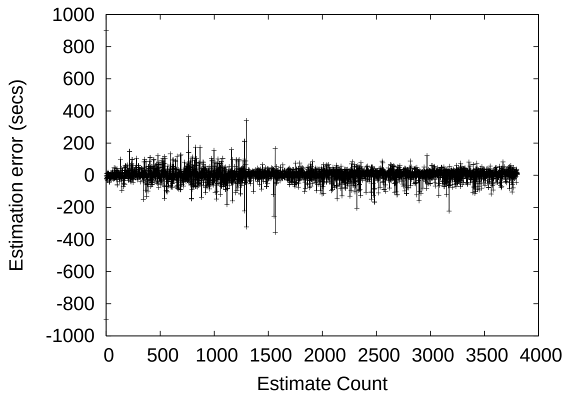
(b) Random waypoint, multi-node, high speed



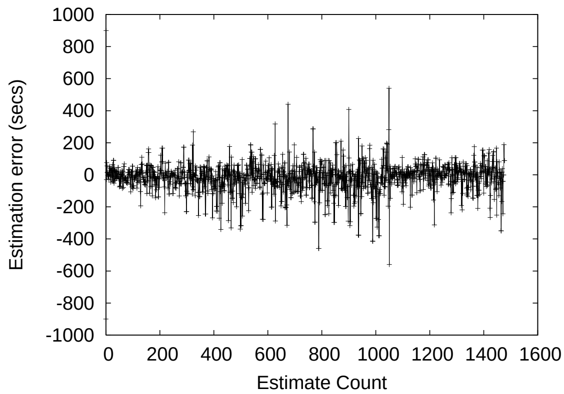
(c) Random waypoint, multi-node, variable speed



(d) Random waypoint, single-node, low speed

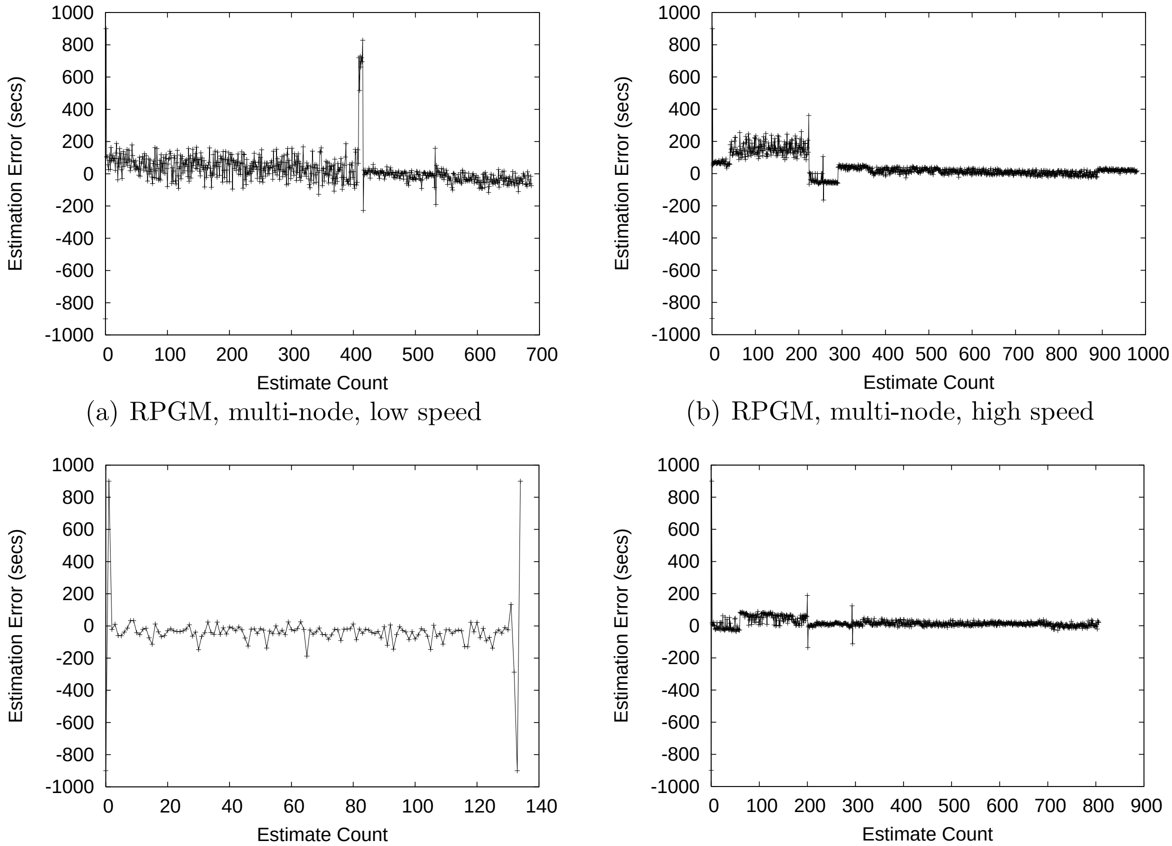


(e) Random waypoint, single-node, high speed



(f) Random waypoint, single-node, variable speed

Figure 2.12: Random waypoint residual lifetime estimations



(a) RPGM, multi-node, low speed

(b) RPGM, multi-node, high speed

(c) RPGM, single-node, low speed

(d) RPGM, single-node, high speed

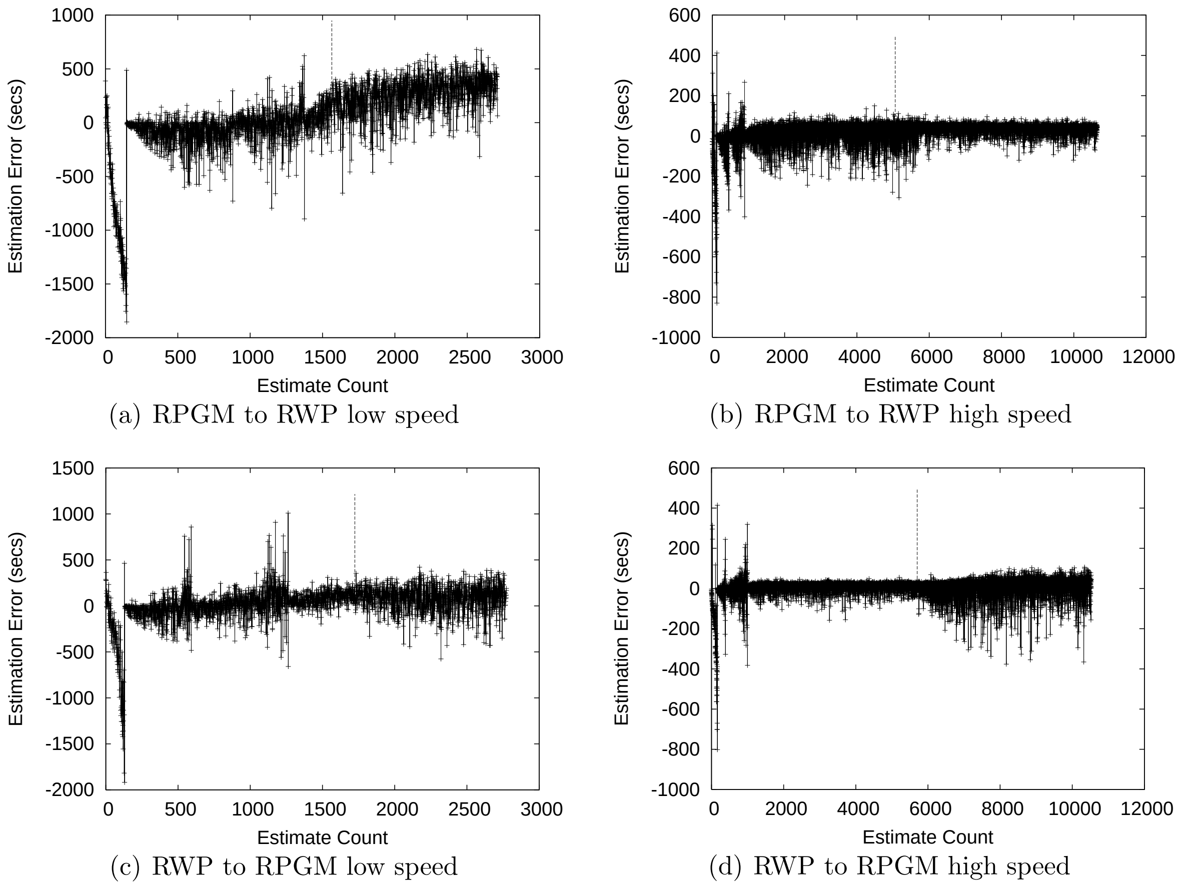
Figure 2.13: RPGM residual lifetime estimations

to random waypoint. In addition, for low speeds, there are few estimations using higher amount of history. We have not included results for RPGM1 and RPGM2 because there were less estimations for these cases.

Estimations Considering Scenarios with Transitions between Random and Group Mobility

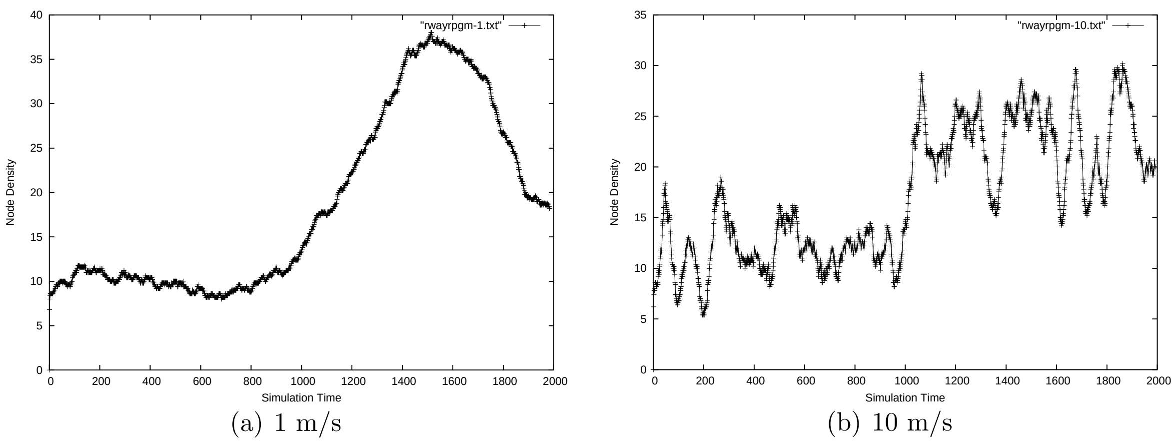
In the previous two subsections, we considered cases where mobility patterns remained the same throughout the simulation duration. In this subsection, we consider cases where the mobility patterns change from random waypoint to group mobility or vice versa. This transition occurs after 1000 secs of simulation period, where the total simulation period is for 2000 secs. Therefore, apart from the classification of estimations into different sections based on amount out history, we also indicate the transition with an impulse (dashed straight line). Figures 2.14(a) and 2.14(b) shows the estimation plots of group to random waypoint transition, with low speed and high speed, respectively. Whereas, Figures 2.14(c) and 2.14(d) show the estimation plots for random waypoint to group mobility transition. The disadvantage of heuristic based approach can be seen in these plots. For low speed RPGM to RWP transition (Figure 2.14(a)), we can see that there are large amount of over estimations present. That is, after the transition, all the estimations errors are due to over-estimation. This is because, after the transition from group to the random scenario, the history contains more number of long link lifetimes. This will affect the estimations in random waypoint scenario. Similar effect (opposite) can be seen in Figure 2.14(d), where there are a lot of high estimation errors after the transition period.

One technique to counter this affect is by having the node predict this transition period and lose some amount of history so that the estimation errors will get reduced. There are various ways for a node to predict the transitions. In this work, we propose a technique where nodes maintain the information about its neighbor densities. Neighbor densities give an indication about the mobility pattern. Figures 2.15 and 2.16 shows the variation of node densities with simulation duration, for RWP-to-RPGM and RPGM-to-RWP transitions, respectively. In each of these figures, there are two plots for low (1 m/s) and high (10 m/s) speeds. In these four Figures (2.15(a), 2.15(b), 2.16(a), 2.16(b)) we can see that after 1000 secs, there is a change (either increase or decrease) in the node-density values. This shows that node-densities can be used to estimate the transition process. It may not be straightforward for other combinations of mobility patterns.



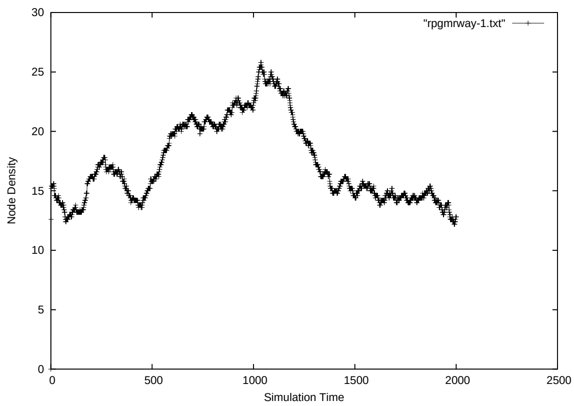
(a) RPGM to RWP low speed
(b) RPGM to RWP high speed
(c) RWP to RPGM low speed
(d) RWP to RPGM high speed

Figure 2.14: Residual lifetime estimations for transition cases (RWP-RPGM-RWP)

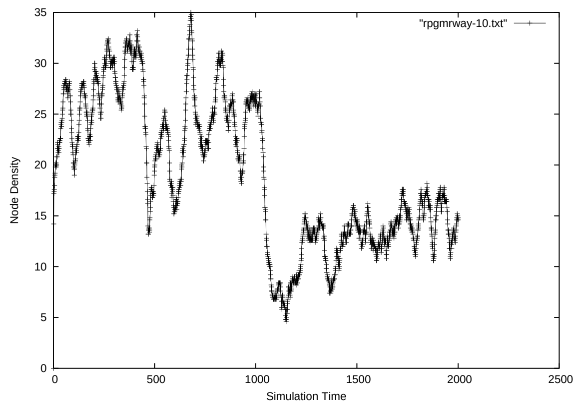


(a) 1 m/s
(b) 10 m/s

Figure 2.15: Random waypoint to RPGM - node density



(a) 1 m/s



(b) 10 m/s

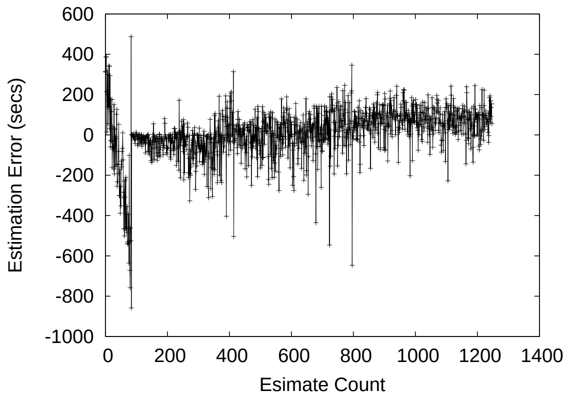
Figure 2.16: RPGM to random waypoint - node density

Estimations for Heterogeneous Scenarios

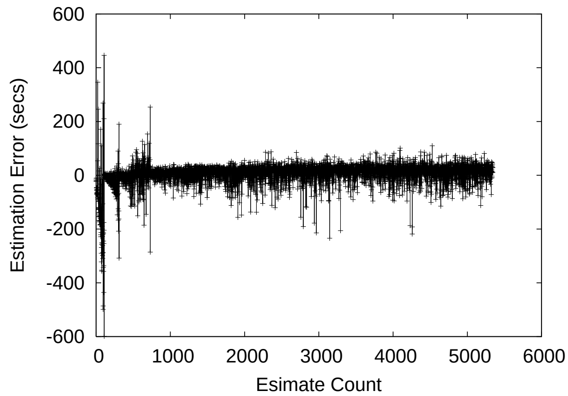
Finally, we consider cases where the mobility pattern is heterogeneous. That is, different nodes follow different mobility patterns. 50 nodes are divided into two sets of 25 nodes each. One set of nodes follows random waypoint and the other set follows group mobility. Figures 2.17(a) and 2.17(b) shows the estimation plots for mixed scenario with multi-node estimation technique, with low and high speeds, respectively. Whereas, Figures 2.17(c) and 2.17(d) shows the estimation plots for single-node estimation technique.

Further, Figures 2.17(e) and 2.17(f) shows the estimations for multi-node and single-node estimation techniques, for cases where along with heterogenous mobility patterns, nodes also have heterogenous speeds. That is, nodes following random waypoint mobility have higher speed, and nodes having group mobility has lower speeds.

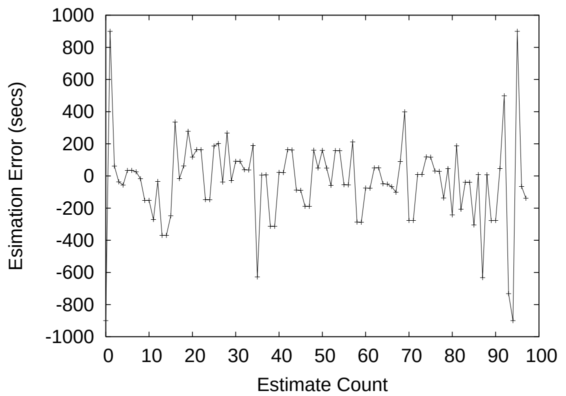
From the above plots we can see that the estimation errors, even for the heterogeneous cases, is fairly acceptable. Therefore, heterogenous cases may not be of a concern for heuristic based estimation technique.



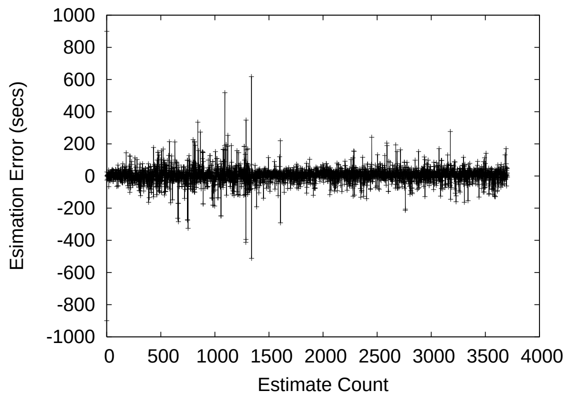
(a) Multi node, low speed



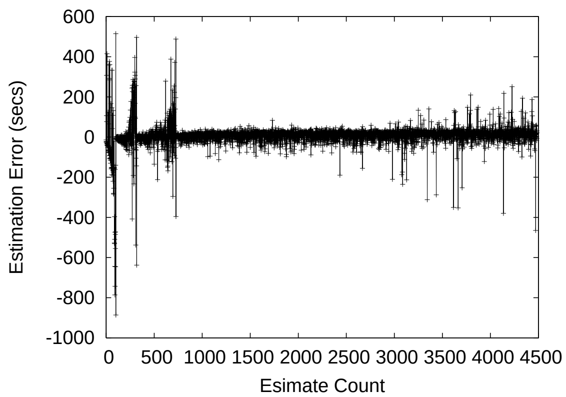
(b) Multi node, high speed



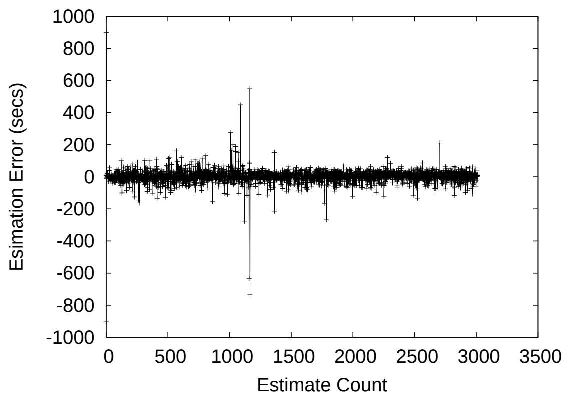
(c) Single node, low speed



(d) Single node, high speed



(e) Multi node, RWP high speed, RPGM low speed

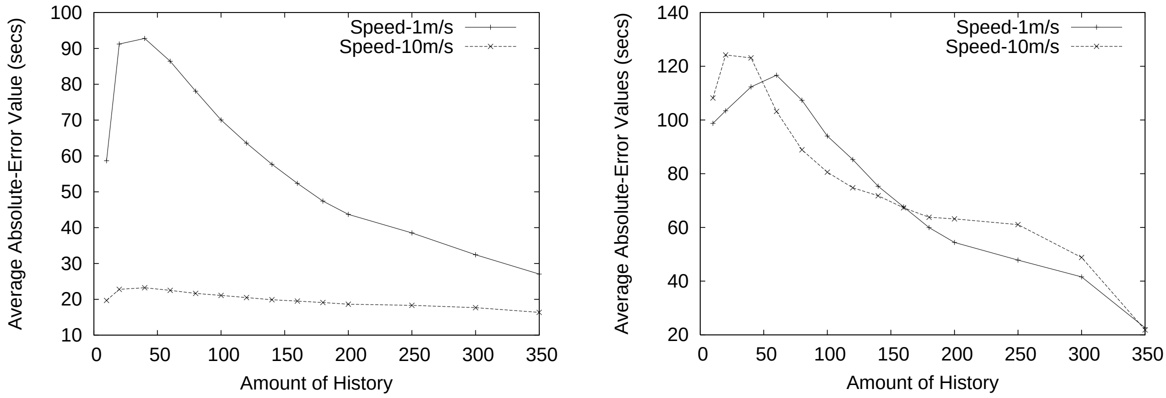


(f) Single node, RWP high speed, RPGM low speed

Figure 2.17: Residual lifetime estimations for heterogeneous cases

In summary, we found that the multi-node estimation technique provides better residual link lifetime estimation values, and can be used along with different mechanisms, for example, routing and scheduling. Figures 2.18(a) and 2.18(b) for random waypoint and RPGM, respectively, provide an information about the dependence of accuracy of estimation on the amount of history, and node mobility. We can see that for random waypoint, estimation errors are higher for high-speeds compared to low-speeds. Whereas, for RPGM there is less dependence on node speeds. Further, for both random waypoint and RPGM, the estimation accuracy increases with increasing speed.

We have also seen from the plots (Figure 2.18) that the estimation error decreases as the amount of history of link lifetimes increases. This advantage would motivate to maintain as many link-lifetimes as possible (as long as there is no change in mobility pattern). Excess storage, however, increases the memory requirements. Therefore, it is important either to design efficient data structures to reduce the storage space or to decide on an upper bound on the amount of history required. From the above simulations, we can find that it is difficult to decide on a single value for the upper bound (x) on the maximum history that needs to be maintained. This value x indicates that any number of collected link lifetimes greater than this x, may not improve the estimation errors. We found that, this value (x) varies across different mobility models, and within a single mobility pattern, this value varies across different speeds and node densities. We believe that a better approach would be to consider both the number of link lifetimes collected (x) and amount of time that has elapsed (α) since the lifetimes are collected. It is easier to develop a bound on this time duration α. Therefore, a node can choose a time-bound, say β, and discard all those values that are stored in the history before β when the number of lifetimes are greater than x. It would be



(a) Random waypoint amount of history versus error

(b) RPGM amount of history versus error

Figure 2.18: Amount of history versus estimation error values

part of future work to study the efficiency of this approach across different node mobility speeds, and mobility patterns.

2.4.1 Improving Estimation Process Using Distribution Information

There are also various ways in which we can exploit the link lifetime distribution information: network reliability studies, estimating lifetimes, understanding of failure rates. In this section, we will describe how we use the temporal properties study proposed in preceding sections to improve the link lifetime estimation technique. In the previous section, we concluded that lifetimes in random waypoint scenarios, with high probability, follow lognormal distribution, whereas group mobilities follow Weibull distribution. We will use this distribution information to estimate the lifetime values when the history does not exist (history here refers to collection of link-lifetimes by a node) to make an estimation. Further, we will also make use of node-density measurements to estimate the transitions from one mobility model to other mobility model (group to random to group). Once a node estimates the transition, it will give up all the collected lifetimes, and start collecting the lifetimes freshly.

In the plots as shown in Figure 2.19, we term the estimation technique which does not use the distribution as “Original”, and the one which uses the distribution information as “Enhanced”. We provide results for high-speed scenarios for three cases. Figures 2.19(a) and 2.19(b) show the “Original” and “Enhanced” estimation plots for high speed (10 m/s), and Figures 2.19(c) and 2.19(d) shows for variable speed (1-10 m/s). If we focus on the first part of the plot, which shows the estimation when there does not exist any history, we can notice that “Enhanced” version has lesser variations. Figures 2.19(e) and 2.19(f) shows the “Original” and “Enhanced” version estimations for the transition case (random waypoint to RPGM3), respectively. In these plots, we can see the improvements both at the initial part and at the later part. If we notice the last part of these plots (2.19(e) and 2.19(f)), i.e., after the transition, we can see that the estimations for “Enhanced” are shifted above, and also estimation errors are less for “Enhanced” compared to the “Original” version. Average error values for each plot is also provided for different sections of the plots. From these values, it is clear that the distribution information is effective when there does not exist any history. Non-existence of history also includes the cases when transition occurs, and node clears its collection of link lifetimes.

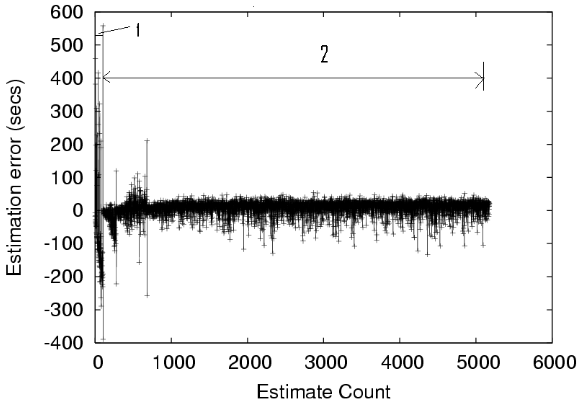
In addition, we conducted a comparative study, considering only the transition cases, with 2 other versions in which lifetimes are assigned following a “uniform” and “exponential” distributions when there does not exist any history. We have seen that the uniform and the exponential distributions were not good-fit models for either random or group mobility patterns. These versions do not give up the collected history after transition, as it does not carry out any estimations. Figure 2.20 shows the mean estimation error values for all the three cases. The versions, which uses the uniform and exponential distribution are termed as 'original (uniform)' and 'original (exponential)'. We can clearly see the improvements in the initial stage, and in the later stage when there are sufficient history information. The mean error value is halved using the Enhanced version.

In the following section, we propose a route computation mechanism SHARC, which uses the proposed link lifetime estimation technique.

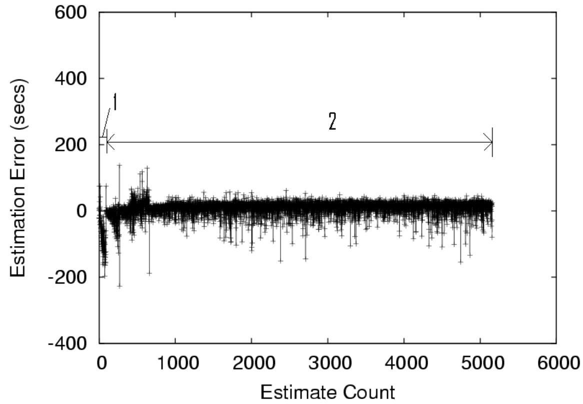
2.5 SHARC- Stability and Hop-count based Approach for Route Computation

In this section, we present our approach for MANET routing based on stability and hop-count, where the stability metric considered is the residual lifetime of a link. We view stability-based routing not as a separate routing protocol but as an enhancement to a hop-count based routing protocol (e.g. DSR or AODV), so that the expected residual lifetime as well as hop-count of a route are taken into account.

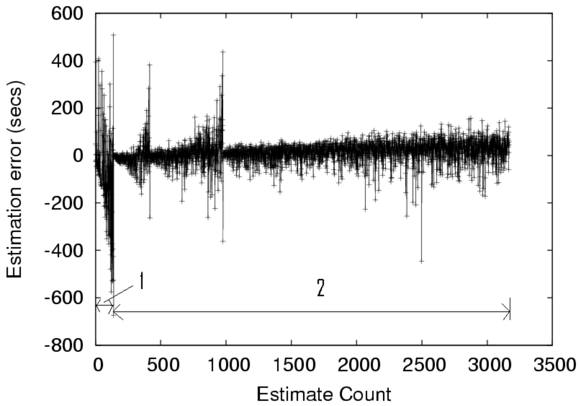
In this section, we first provide a building blocks view of the stability-based routing. This building blocks description is used to provide better understanding of the stability-based routing mechanics. In [46], Fan Bai et al. propose two frameworks: building blocks analysis, also termed as BRICS, and IMPORTANT. The IMPORTANT framework, which involves BRICS framework, aims to evaluate the impact of various mobility models on the performance of routing protocols. On the other hand, the goal of the BRICS framework is to identify the general building blocks of routing protocols. In BRICS, building blocks for a routing protocol include route setup and route maintenance. Route setup includes the



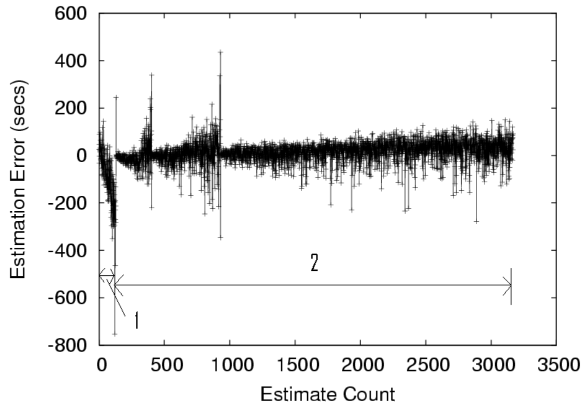
(a) RWP 10 m/s Original 1:140 2:17.827



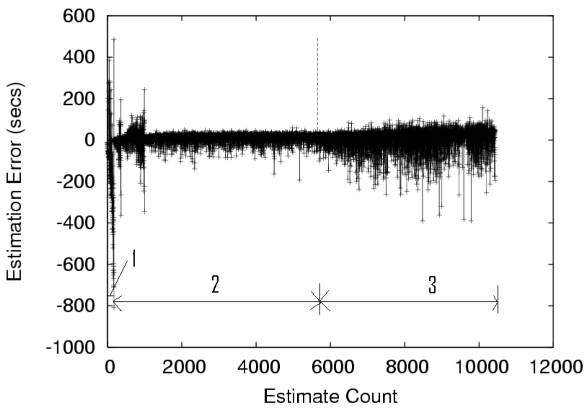
(b) RWP 10 m/s Enhanced 1:78.12 2:17.71



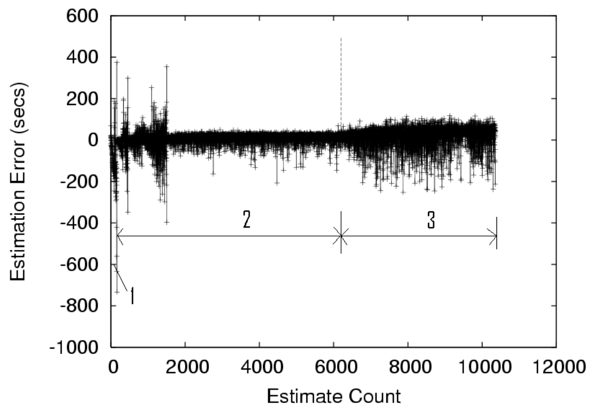
(c) RWP 1-10 m/s Original 1:152.725 2:35.25



(d) RWP 1-10 m/s Enhanced 1:101.21 2:35.19



(e) RWP-RPGM 10 m/s Original 1:263 2:21.5 3:44.15



(f) RWP-RPGM 10 m/s Enhanced 1:113 2:19.5 3:37.85

Figure 2.19: Residual lifetime estimations with and without distribution information

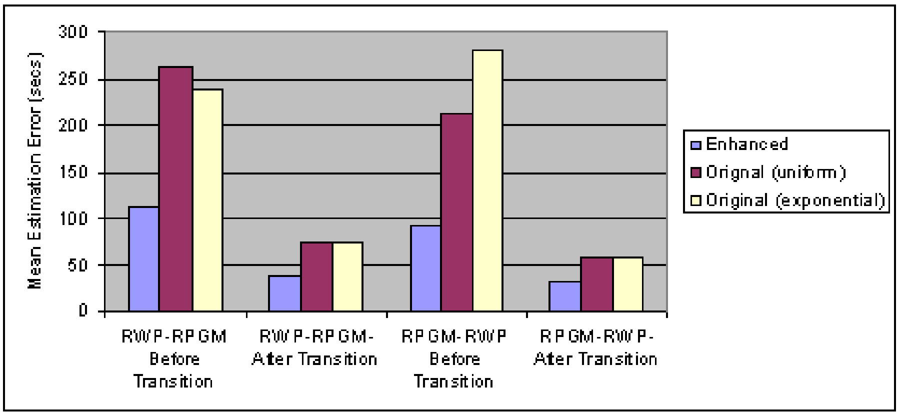


Figure 2.20: Comparison with other distributions

functionalities of flooding and caching, whereas route maintenance include error detection, notification and handling. We propose to add three additional blocks or functionality to the BRICS framework.

The three blocks we propose to add are: distribution of stability information, environment learning through neighbor management and route selection mechanism. Figure 2.21 shows the modified BRICS framework. These three additional blocks can be used to explain the components required in stability-based routing. When used in conjunction with the BRICS framework, these components can illustrate how stability-based routing can be integrated into other common routing protocols. Multipath support in the routing protocol is the only requirement for including these blocks.

The first block, distribution of stability information, is part of the flooding building block of BRICS. Here the stability information collected at the node is distributed along the route so that a node making routing decision can take this information into consideration.

The second block, environment learning through neighbor management mechanism, takes the form of exchanging hello/keep-alive messages. These messages are used to gather

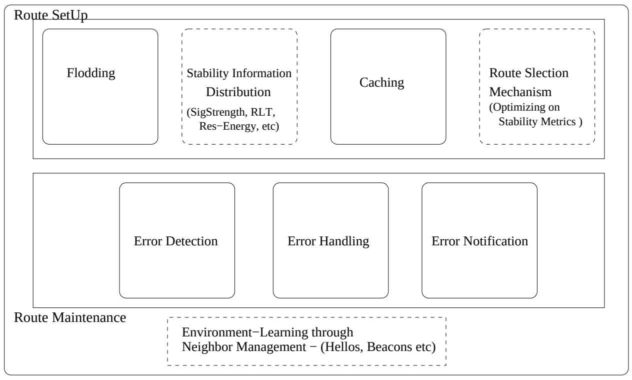


Figure 2.21: Modified BRICS framework

information about the environment. The environment includes number of neighbors, signal strength from a neighbor, neighbor lifetimes. This block also interacts with the first block (distribution of stability information) by providing it with residual lifetime information. This process involves an overhead in terms of both bandwidth and power consumption. At the MAC layer, neighbor state can be obtained during exchange of either control (RTS/CTS) or data messages. Whereas, at the network layer it is necessary to have a mechanism similar to “hello” protocol. We see in the succeeding chapters that the other proposed mechanisms (admission control and scheduling) also rely on the neighbor management mechanism. Therefore, the advantages obtained by neighbor management mechanism outweighs the overheads introduced.

The third block, route selection, selects a route which is most stable from a set of routes. This functionality can be included in the destination node, or source node or at every individual node, depending on the routing mechanism. For example, this selection mechanism is at the destination node for ABR, and if we enhance DSR with the proposed third block, this functionality is at every node.

In the remaining part of this section, we will describe our routing algorithm based on residual lifetime and hop-count, and conclude with the evaluation of SHARC.

Routing based on finding the minimum hop route has been used for a long time. In wireless network, the use of minimum hop route has several advantages, including simplicity, less interference, and lower consumption of network resources (bandwidth). Since minimum hop routing does not take into account link duration, shorter route may not be the best route (may be short-lived route). On the other hand, routing algorithms based on stability, like ABR, SSA and long lifetime routing (LLR), have also proven to be advantageous in some cases. Stability-based routing, however, may sometimes select longer routes, resulting in poorer performance caused by excessive node interference and wastage of network bandwidth. Although the hop-count and stability metrics may seem contradictory at times, it is possible to combine them in order to take advantage of the strengths of both.

We term our algorithm *SHARC* for *Stability and Hop-count based Algorithm for Route Computation*. In this work, we will consider an implementation of SHARC using the DSR routing protocol as the base routing protocol. The additions and modifications carried out on DSR are explained in the following paragraphs. We would like to emphasize that our approach can be applied to any hop-count based routing and DSR is chosen as a case study.

Implementation of Link Lifetime Estimation: In order to distribute stability information, the route-request packet of DSR is changed to carry residual lifetime information. Every node stores the link duration values of its neighbors. By collecting this information and aggregating them into bins of 10 s, each node maintains an estimate of the residual lifetime distribution using all the samples collected, and equation (2.1). During the initial period when the number of link duration samples collected is low, it is likely that a newer link will be chosen.

Every intermediate node on receiving the request packet includes the residual lifetime value in the route request message. The path data structure of DSR implementation is changed by associating every path with an additional stability value. This stability value of the path is the sum of all the residual lifetime divided by the length of the path. The cache structure is also enhanced to maintain the stability value along with the addresses of intermediate nodes. The route selection mechanism is incorporated in all the nodes to be compatible with DSR routing mechanism.

Typically, the stability value of a particular link is calculated based on the most recent (short-term) history, starting from the most recent link establishment. This is true in majority of the stability-based techniques (ABR and SSA). In our link estimator, we consider not just the most recent link behavior, but also the connectivity history of all neighboring nodes. This helps in having a better understanding of the environment in which the node operates, making the estimates accurate. An implicit assumption made is that the environment is homogeneous and nodes retain the same mobility patterns. Implications of heterogeneous mobility patterns are left as future work.

The amount of memory needed for link stability estimation depends on the length of the link lifetime history kept and can be easily bounded. The accuracy and effectiveness of the estimator will be investigated using simulation.

Route Selection Algorithm: The route selection mechanism assumes that routes are stored in the cache. *min_stability* is the current value of the stability available while searching the cache. It is initialized to -1. Similarly *min_length* is the current value of the hop-count available in the process of searching the cache. It is set to the maximum hop-count possible (configurable). *dest* refers to the destination node. Let the function *findRoute* finds the route from the cache matching the destination.

Algorithm 1 Route computation in SHARC

repeat
  route = findRoute in the Cache for dest
  if $route.length <= min\_length$ **then**
    if $route.length = min\_length$ **then**
      if $route.stability > finalRoute.stability$ **then**
        $finalRoute = route$;
        $min\_length = route.length$;
        $min\_stability = route.stability$;
      end if
    else if $route.length < min\_length$ **then**
      $finalRoute = route$;
      $min\_length = route.length$;
      $min\_stability = route.stability$;
    end if
  end if
until the end of the Cache

The route computation algorithm is as shown in **Algorithm** 1. The algorithm tries to find the most stable route among all shortest hop routes. The algorithm can be easily extended to the case where the most stable route among all routes with hop-count not more than N hops longer than the minimum hop route, where $N \geq 0$, are chosen.

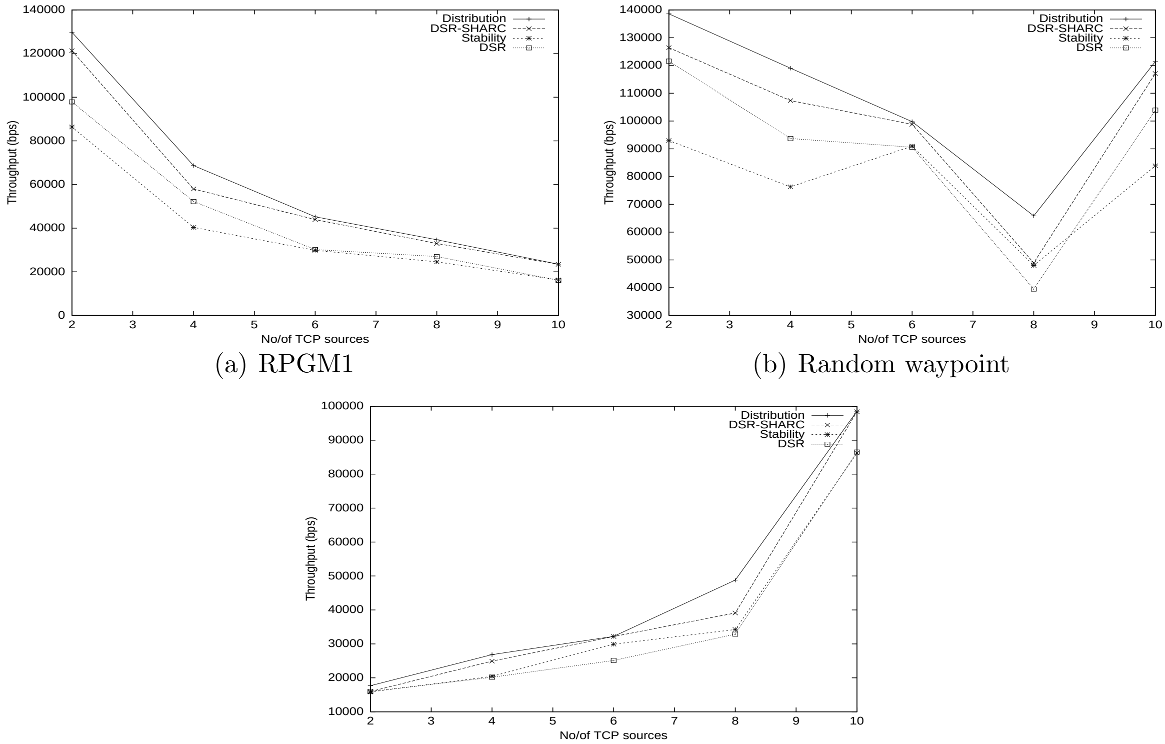
The idea behind SHARC can be explained as follows. SHARC attempts to find a route based on two objectives, path length and path stability. From Section 2.3.1, we know that link stability prediction is difficult and inexact. On the other hand, finding a shortest path is *precise*. In addition, there are often more than one shortest path. Hence, a good approach is to use the shortest path algorithm as the initial filter to narrow down the route selections and then use path stability, a less robust indicator, to choose the best route among the available routes.

2.5.1 Evaluation

In this section, we describe the performance evaluation of our route computation algorithm. We choose to use throughput of long-lived TCP traffic and response time of web-traffic as the performance metrics because we believe they can better capture the effects of link breakages. Many research works use CBR traffic and consider delay and packet delivery ratios. Since CBR traffic using UDP does not perform any congestion control and error recovery, we believe that the response times of short data transfers like web-traffic better reflect the impacts of link stability.

The simulation settings are similar to Section 2.3.1, except that we only consider RPGM1, random waypoint and Manhattan mobility models. For throughput measurements, we consider maximum speeds of 1 m/s and 10 m/s. The number of TCP sources is varied from 2 to 10. Plots for only 10 m/s is shown. For response time measurements, we consider maximum speed of 1 m/s and vary the number of web-client and web-server pairs from 2 to 10.

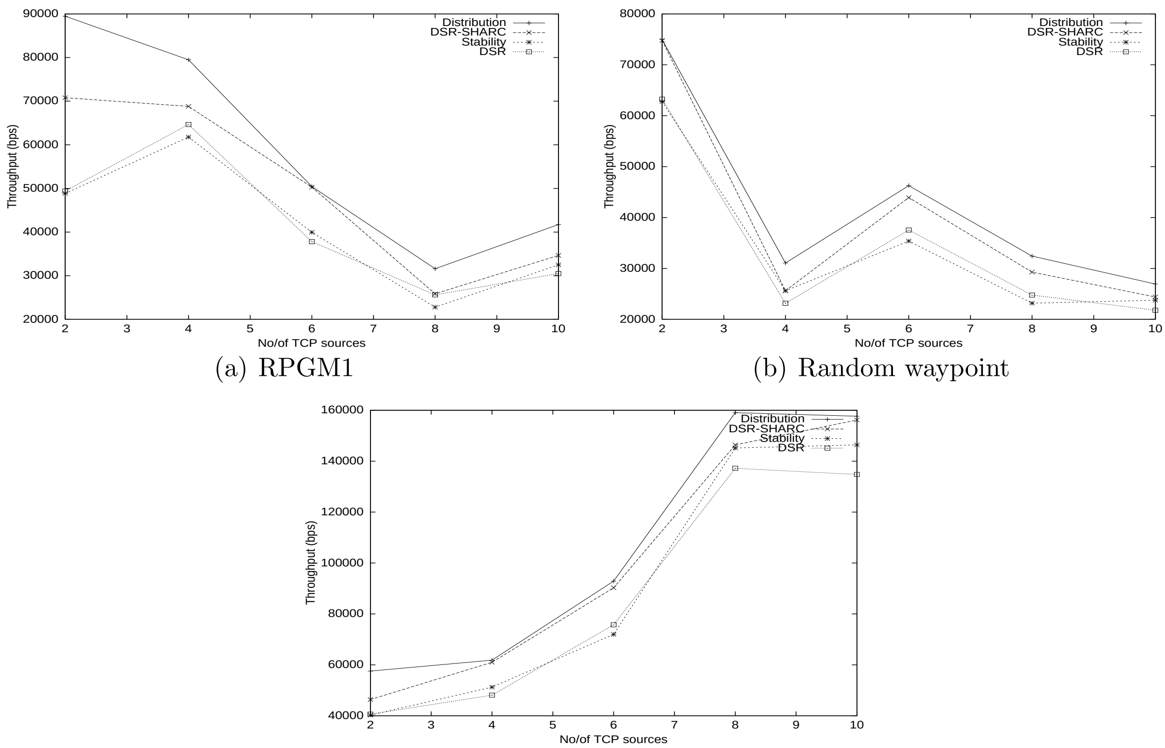
For our evaluations of SHARC, we compare SHARC which is implemented over DSR and labeled as DSR-SHARC, with three other algorithms. As we have combined both hop-count and stability into the route computation, we use as baseline the performance of *DSR*, which is hop-count based and termed as "DSR" in the plots. Second, an algorithm similar to DSR, and which uses the stability metric only is termed as *stability* in the plots. We have simulated DSR with the best of the link caching schemes [58].



(a) RPGM1 (b) Random waypoint

(c) Manhattan

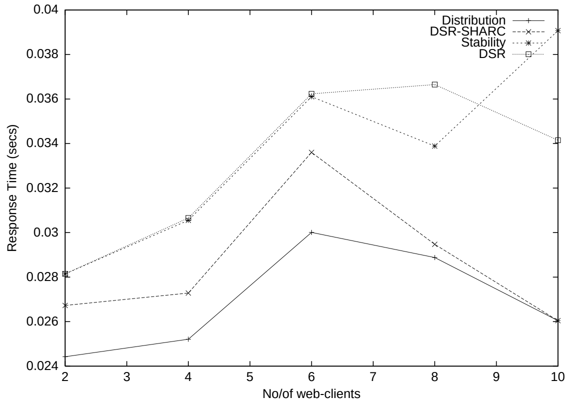
Figure 2.22: Throughput versus number of sources, 1 m/s



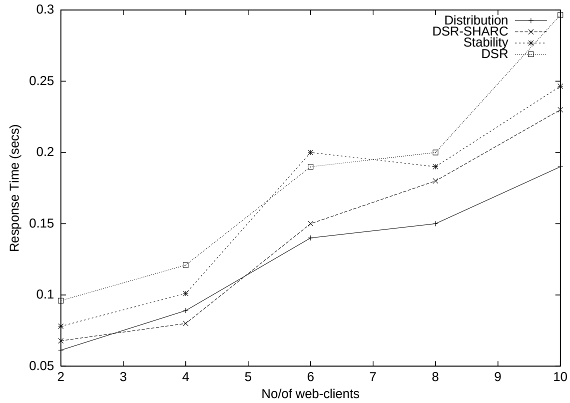
(a) RPGM1 (b) Random waypoint

(c) Manhattan

Figure 2.23: Throughput versus number of sources, 10 m/s



(a) RPGM1



(b) Random waypoint

Figure 2.24: Response time versus number of sources, 1 m/s

The third algorithm, which we label as *Distribution* in the plots, is similar to DSR-SHARC except how residual lifetime values are obtained. Instead of estimating the residual lifetime values from past history, the lifetime values are obtained from the previous simulations with the same parameters (including the random seeds used). The lifetime distributions are embedded in all the nodes at the start of the simulation so that when a node receives a request packet, it can base its decision on the exact neighbor lifetime distributions.

Figures 2.22 and 2.23 shows the throughput of long-lived TCP with varying number of sources and corresponds to maximum speeds of 1 m/s and 10 m/s, respectively . In most of the cases, DSR-SHARC always performs better than the baseline cases of *stability* and *DSR*. At low speed (1 m/s), we can see that the improvement varies from 10% to 30%. At higher speeds (10 m/s), however, we can see that the improvement varies from 10% to 45%. In many cases, an algorithm that takes into account stability only performs worse than DSR. This is consistent with previous works like SSA, and work by Sridhar et al., [75]. *When stability is added as an enhancement to the hop-count metric, however, performance can be improved.*

The plots also show that the throughput values of DSR-SHARC are closer to the scheme that operates with complete knowledge of residual lifetimes. Although the estimation error is large in the beginning of the experiment, as the simulation time progresses and the node collects more information of neighbor lifetimes, the estimates can be substantially better. Nevertheless, even with the rough approximations obtained using historical data, it is possible to perform close to the ideal algorithm in some cases. The performance gap is between 1% to 10%.

Next, in order to evaluate the performance with respect to delay values, we consider the response time of web-traffic. We randomly select a node to be a HTTP server, and associate a cache node with it. We use the web-traffic model of NS-2 [11], in which pages are maintained as page pool, with configurable parameters such as page objects, expiry time, request interarrival times. In our simulation we use page with just a single object, with average expiry time of 5 secs. The interarrival time of request from the client has an average value of 10 secs.

Figure 2.24 shows the response time with varying number of web-clients. Figure 2.24(a) corresponds to single group (RPGM1) mobility model, whereas Figure 2.24(b) corresponds to random waypoint. Both the models operate in low-speed (1 m/s). The results are similar to the throughput plots. DSR-SHARC has lower delay values compared to both the baseline algorithms. The improvements are 10% to 40% over DSR and 5% to 50% over a stability-only algorithm.

Energy and Processing Overheads

Energy overheads refers to additional energy consumptions due to mechanics involved in the protocol. Whereas, processing overheads refers to additional processing a node carries out when it uses the protocol. In SHARC, these overheads are introduced in following cases: due to neighbor management mechanism, due to maintenance of the history of neighbor lifetimes and finally due to link residual lifetime estimation.

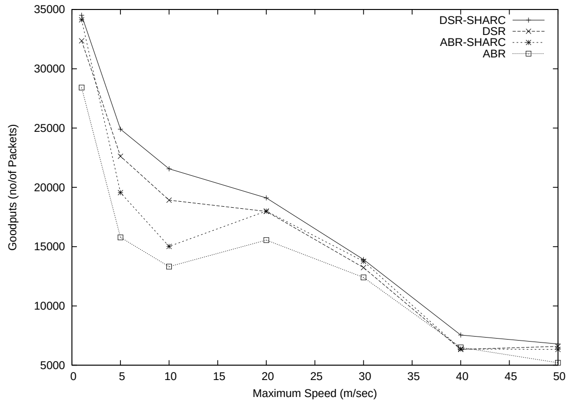
Many routing protocols like ABR and AODV include neighbor management mechanisms using "Hello" messages. Therefore, when these protocols are enhanced with SHARC, overhead of neighbor management mechanism is not involved. When other protocols like DSR is enhanced with SHARC, however, the overhead of neighbor management exists. This overhead with respect to processing is negligible, and with respect to energy consumption, is less.

Earlier simulations have shown that at any given time, the average neighbor density (number of neighbors) of any node varies from 5 - 30 (for high to low mobility scenarios). Further the average number of lifetime values will vary from approximately 20 - 500 (for low to high mobility scenarios). In our implementation, we consider bins of 10 secs, and the lifetime values are added into corresponding bins. Hence, the maximum processing overhead would be to consider all the bins, which is negligible.

Effect of Routing Protocol Mechanics

In the preceding section we mentioned that SHARC is a route computation mechanism and can be included in majority of the routing protocols. We believe that the routing mechanics such as source/hop-by-hop routing, local/global error recovery, caching schemes, however, play a major role. To study this property, we implement SHARC in two different routing protocol framework: ABR and DSR. In case of ABR we replaced ABR's associativity technique with SHARC, and in case of DSR we enhanced it with SHARC route computation. We study these two versions of SHARC with original ABR and DSR versions. Figure 2.25(a) and 2.25(b) shows the goodput values for these 4 mechanisms with varying maximum speed for both random waypoint and RPGM mobility models. RPGM is the RPGM1 mobility model as explained in the previous sections. The traffic used is 10 best-effort traffic flows. Other simulation parameters remain the same as used in previous sections.

Considering the Figure 2.25(a), DSR versions perform better than ABR versions for low speed traffics in random waypoint model, whereas for RPGM1 mobility model, ABR versions perform better than DSR counterparts. This may be due to caching schemes which helps in random scenarios, and their effect is not substantial with group mobility models. In addition, ABR's local recovery helps in group mobility models. SHARC-versions of ABR and DSR, however, always performs better than ABR and DSR without SHARC. Further, for lower mobility scenarios of random waypoint, ABR's associativity idea (longer they remain as longer, possibility of them remaining as neighbors for longer time is high) may be wrong, resulting in poor performance. These results substantiate our claim in preceding section.



(a) Random waypoint (b) RPGM1

Figure 2.25: Response time versus number of sources, 1 m/s

2.6 Summary

In this chapter, we studied various temporal properties of link lifetimes. From the simulations, we found that a single distribution model cannot be applied for all the mobility patterns. In particular, we found that for random scenarios lognormal distribution happens to be the best fit and for group mobilities Weibull distributions seems to be the most appropriate fit. Further, we found that residual link lifetime is a function of current link age, mobility speed and mobility pattern, and does not vary monotonically with age. Nevertheless, the residual lifetime can still be useful if it can be estimated approximately. We proposed a stability and hop-count based routing algorithm, called *SHARC*, which finds the most stable route among the set of shortest hop routes. Performance evaluation of SHARC shows that it performs better than purely stability-based and purely hop-count based algorithms in terms of throughput of long-lived flows and response time of short data transfers.

2.A Appendix: Study of Link Stability based Routing

We use the term *stability* based routing to refer to the class of routing protocols that use some representation of link-quality as the routing metric. The terms associativity, longevity and availability are used to refer link stability. For example, higher the associativity or availability more stable the link/path is. Further, there can be various parameters (practically measurable) that are used to measure the associativity, longevity or availability. Typical parameters include - number of beacon exchanges, link/path lifetime, signal strength.

In this section, we will carry out the performance study of a stability-based routing mechanism. We consider ABR for the following reasons: It is the first stability-based mechanism proposed for MANETs, and the idea was also patented and used in commercial products. In addition, SSA [20] was found to have disadvantages with its location-stability approach. Other mechanisms like RABR and link availability based routing that were existent during the time of study were similar to that of ABR.

2.A.1 Comparative Study with AODV

Ad hoc On-demand Distance Vector (AODV) is one such routing protocol that has been studied and optimized over years. AODV incorporates the on-demand mechanism of route-discovery similar to Dynamic Source Routing (DSR), and uses hop-by-hop routing. Sequence numbers are used to prevent route loops. In AODV, routing tables are used to maintain the route information, one entry per destination. A routing table entry expires if not used recently, which is achieved by maintaining timer-based state information. There are other optimizations proposed for AODV, which can be found in the AODV draft.

The outline of this section is: we first compare the performance of ABR and AODV for real-time traffic, with varying pause-time, and varying the number of real-time traffic sources. Similar simulations are carried out for best-effort traffic. Next we carry out simulations to compare the energy consumption behavior of these two routing protocols. We provide a discussion reasoning out the performance differences among the protocols.

Simulation Environment

Each mobile host has a transmission range of 250 meters and shares an 11 Mbps radio channel with its neighbors. The simulation includes a two-ray ground reflection model and IEEE 802.11 MAC protocol. All the simulations are run for 900 seconds. A multihop network with 50 mobile nodes is simulated. The network area has a rectangular shape of 1500 m x 300 m. The mobility pattern used is the random waypoint model. Each node selects a random destination and moves with a random speed up to a maximum speed of 20 m/s (72 km/hr), pausing for a given "pause time" when the destination is reached. Average end-to-end packet delay, average goodput, packet delivery ratio and percentage energy consumption are the performance metrics that are considered. All experiments are carried out with 12 replications.

Implementation of ABR on NS-2

We have implemented and tested ABR on NS-2 [11, 76]. The implementation of ABR is in line with AODV implementation. ABR beacons are implemented as control message, with just ABR base header. This is to support the associativity of ABR protocol. An important decision criterion is the duration for which destination waits before sending the reply. This duration is made dynamic, i.e., the duration is made proportional to the number of hops traversed by the first request message that is received. If the number of hops are less, the waiting time will also be less.

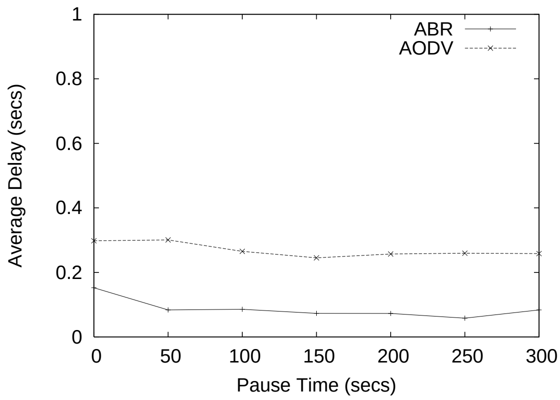
All nodes maintain three tables: routing, neighbor and seen tables for forwarding, neighbor information and to avoid duplicate processing, respectively. The current implementation uses the associativity ticks and relaying load at each node as routing metrics.

Real-Time Traffic

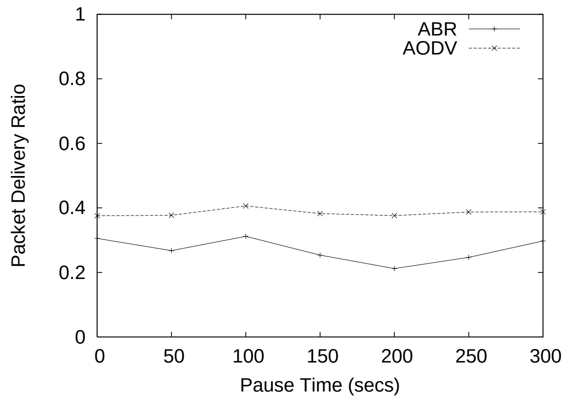
The real-time traffic is modelled as video traffic, which is 200 kbps Constant Bit Rate (CBR) traffic with a packet size of 512 bytes. We first consider average delay and packet delivery ratio, for 10, 20 and 40 sources, with varying pause-time as shown in Figure 2.26. From Figure 2.26(a), for l0 CBR sources, we can see that ABR has better delay performance than AODV, whereas Figure 2.26(b) shows that AODV has better packet delivery ratio than ABR. Figures 2.26(c) and 2.26(d) show the results for 20 CBR sources. The performance difference is more for 10 CBR sources when compared to 20 CBR sources. When the number of sources is increased to 40, the pattern reverses with AODV having better delay values and ABR having better packet delivery ratios. This is shown in Figures 2.26(e) and 2.26(f). This behavior is pronounced in Figure 2.27. Figures 2.27(a) and 2.27(b) show average delay and packet delivery ratios, respectively, with varying number of CBR sources, for pause time of 0 secs. Figure 2.27(c) shows the average delay of real-time traffic for varying CBR sources with 300 secs pause time. From Figure 2.27(a), we can notice that ABR performs better with respect to average delay for lesser CBR sources ($<$ 40 sources). In addition, there are larger variations in ABR delay values. From Figure 2.27(b), ABR has better packet delivery ratio than AODV, though the advantage is not significant. Packet delivery ratio performance was similar for 300 secs pause time, which is not shown here.

AODV uses hop-wise path length as the metric to choose among alternate routes. In AODV, destination replies only to the first arriving request packet, by which, it will be favoring the route with less congestion. Whereas, in ABR the metrics used are longevity of the route and relaying load. ABR protocol's advantage in the delay for real-time traffic, as in Figure 2.26(a), is mainly attributed to these metrics. Route discovery latency is high in case of ABR because the destination node will wait for some duration, which is proportional to the number of hops the first request has traversed, before it sends a reply. This waiting time affects the average delay when the number of sources increases, as seen in Figures 2.26(c) and 2.26(a). This is also a reason for ABR's high variation in average delay as shown in Figure 2.26(a). This latency also contributes for decrease in packet delivery ratios of CBR traffic, as in Figure 2.26(b).

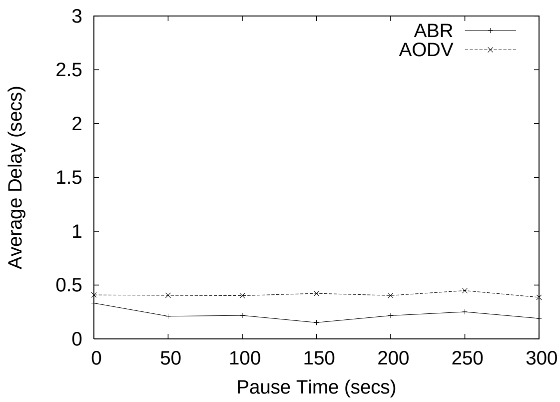
AODV protocol has lesser packet delivery ratio, as the number of CBR sources increase (Figures 2.26(f) and 2.26(b)), which can be attributed to stale information due to intermediate node replying to the route requests. This stale information results in packet losses which in turn affects the total packets received. The slightly better performance of the ABR protocol with respect to packet delivery ratio for the best-effort traffic (Figure 2.26(b)) is again due to its route selection method. Lesser the number of times a given route breaks, higher will be the packet delivery ratio. The route selection technique in ABR helps in finding a route which has longer life-time (longevity of the route). The stale information in case of AODV is the reason for its lesser packet delivery ratio.



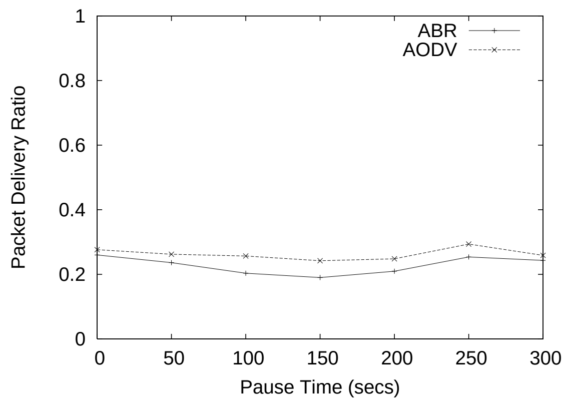
(a) Average delay of real-time traffic vs mobility (10 CBR soures).



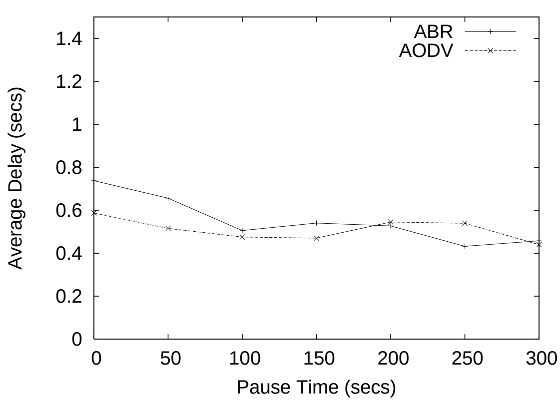
(b) Packet delivery ratio of real-time traffic vs mobility (10 CBR soures).



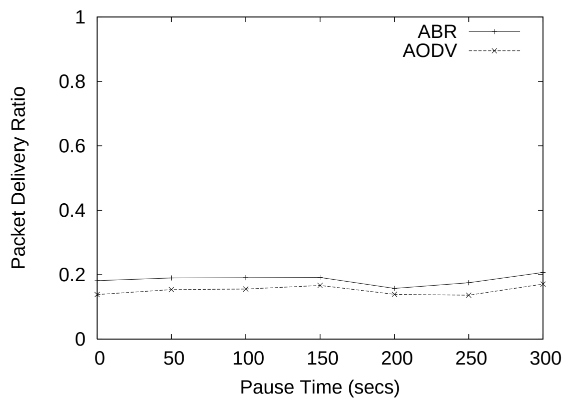
(c) Average delay of real-time traffic vs mobility (20 CBR soures).



(d) Packet delivery ratio of real-time traffic vs mobility (20 CBR soures).

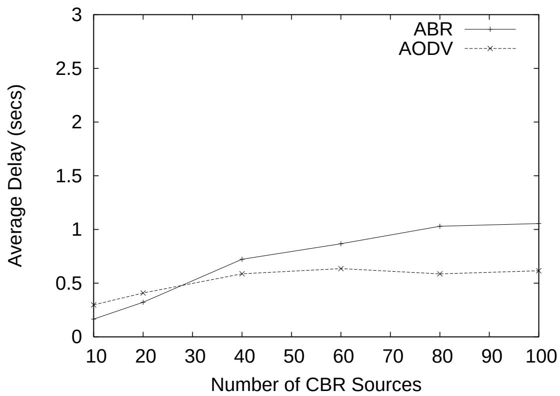


(e) Average delay of real-time traffic vs mobility (40 CBR soures).

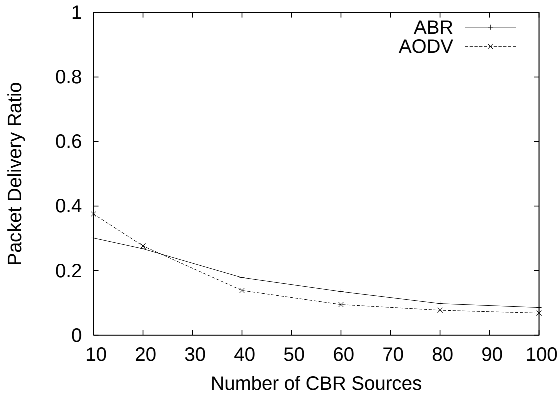


(f) Packet delivery ratio of real-time traffic vs mobility (40 CBR soures).

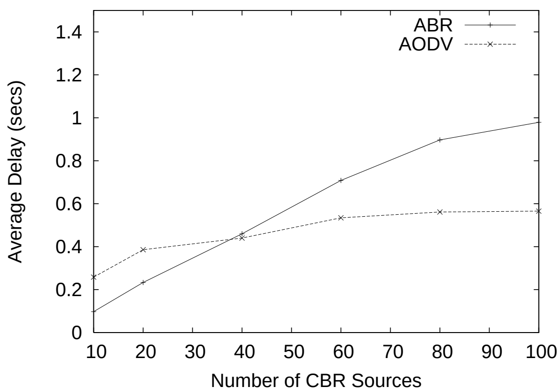
Figure 2.26: Effect of varying mobility with fixed number of CBR sources



(a) Average delay of real-time traffic vs number of CBR sources (0 secs pause time).

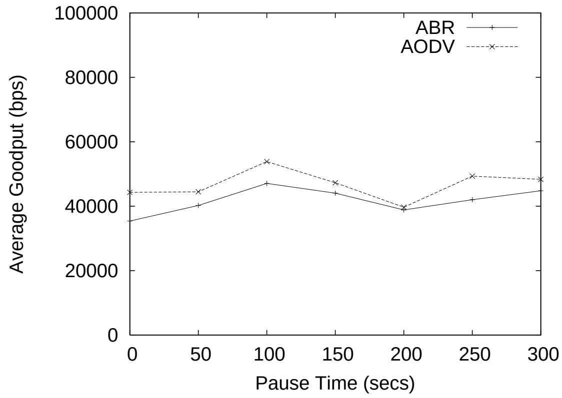


(b) Packet delivery ratio of real-time traffic vs number of CBR sources (0 secs pause time).

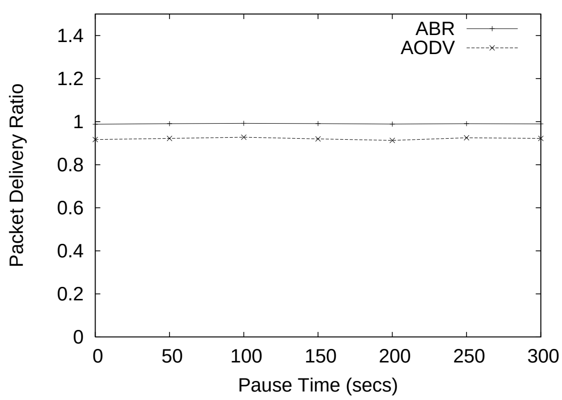


(c) Average delay of real-time traffic vs number of CBR sources (300 secs pause time).

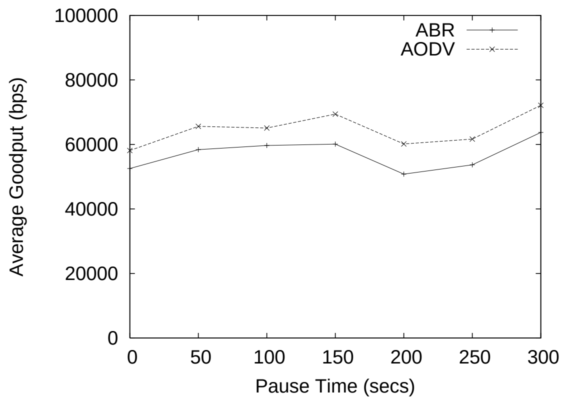
Figure 2.27: Effect of varying number of CBR sources with varying mobility



(a) Average goodput of best-effort TCP traffic vs pause time (10 tcp sources)



(b) Packet delivery ratio of best-effort TCP traffic vs pause time (10 tcp sources



(c) Average goodput of best-effort TCP traffic vs pause time (40 tcp sources).

Figure 2.28: Effect of varying pause time with fixed number of TCP flows.

Best-Effort Traffic

The best-effort traffic is a set of greedy FTP sessions. Figures 2.28(a) and 2.28(b) show average goodput and packet delivery ratio, respectively, for 10 TCP sources, with varying pause-time. Whereas, Figure 2.28(c) shows average goodput with varying pause-time for 40 TCP sources. As expected, average goodput increases with increase in number of TCP sources. From Figure 2.28(b), we can notice that with varying pause time, there is no effect on the packet delivery ratio, it remains almost constant. Experiments were carried out for 20 TCP sources. The results obtained were similar to the results with 10 and 40 TCP sources. AODV performed constantly better than ABR with respect to average goodput, though the difference is not significant. ABR performed better than AODV with respect to packet delivery ratio.

The impact of mobility (varying pause-time) on delay and goodput is due to the route discovery latency and congestion on new route. Route discovery latency impacts the end-to-end packet delay. Though ABR has higher route discovery latency compared to AODV, the congestion and number of route discoveries are less. The ability of ABR to consider the load at each node and associativity between nodes facilitates reduction in congestion and number of route discoveries.

Energy Consumption

The design of energy efficient protocols is an important requirement for MANETs, as the nodes are constrained with battery power. This constraint makes designers of protocols for MANETs to concentrate on improving the energy efficiency along with other performance metrics. There have been proposals of routing protocols for MANETs that consider energy or battery power consumption [77, 78]. In this section, we concentrate on the energy consumption aspects of ABR and AODV routing protocols.

Energy consumption model considered and the calculations of transmitted and received energies are similar to [79]. Considering voltage and current values which correspond to 2,400 MHz WaveLan implementation of IEEE 802.11, the following equations represent the energy used (in Joules) [79]:

Transmitted Energy

$$Energy_{(tx)} = (330 * 5 * PacketSize)/2 * 1000000)$$

Received Energy

$$Energy_{(rx)} = (230 * 5 * PacketSize)/2 * 1000000)$$

Transmitted and received energies refer to energy consumed during transmission and reception, respectively. Packet size is in bits. Similar to [79], we assume that energy consumed is null during listening mode. In addition, we maintain the Radio Frequency (RF) value at 281.8 mW, which corresponds to the RF energy required for transmission range of 250 m. This energy value determines successful or failed packet reception.

The simulation environment is similar to the previous experiments. We used 40 sources, which generated CBR data traffic, each with a sending rate of 4pkts/sec, using a packet size of 512 bytes. Each simulation lasted for 1000 secs, and each value in graph is obtained after 12 replications. We estimated the following quantities: total energy consumed, energy consumed during packet transmissions and receptions, and energy consumed (in both transmission and reception) for MAC and routing control packets. Request To Send (RTS), Clear To Send (CTS) and acknowledgement (ACK) are the MAC control packets that are considered. Routing control packets include route request, route reply, Route Delete, Route Notification, and Route Error. Energy consumed neither for transmission nor for reception of data packets were considered.

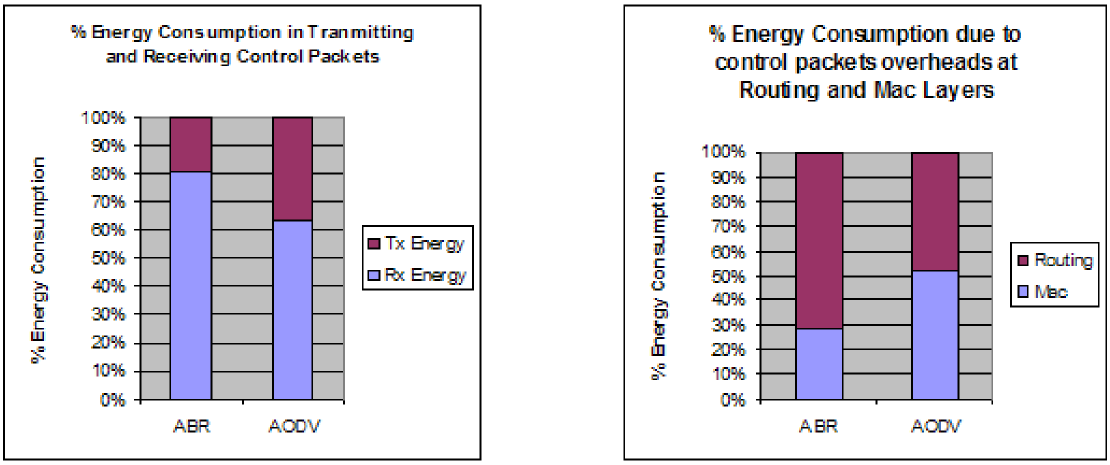


Figure 2.29: Percentage of total energy consumption

Figure 2.29 shows the percentage energy consumption in transmitting and receiving the packets (both routing and MAC). Reception process consumes more energy than transmission process, with respect to the total energy consumption, for both AODV and ABR. Figure 2.29 also shows the percentage energy consumed as a function of packet type. For both AODV and ABR, the energy consumed due to MAC protocol control packets affects significantly the total energy consumed. These results are similar to those presented in [79]. In terms of energy solely consumed by routing control packets, AODV performs better than ABR. This is due to size of routing control packets size, which is large in ABR. When we consider total amount of energy consumed by all nodes involved in either transmission or reception of control packets, the results are different, however.

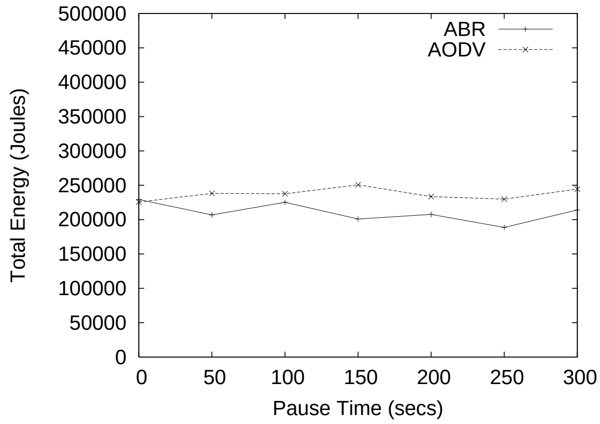
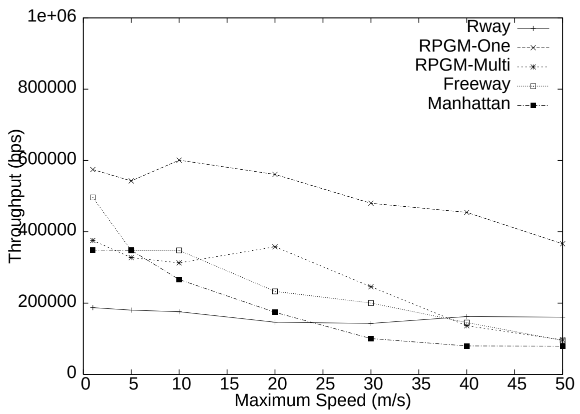
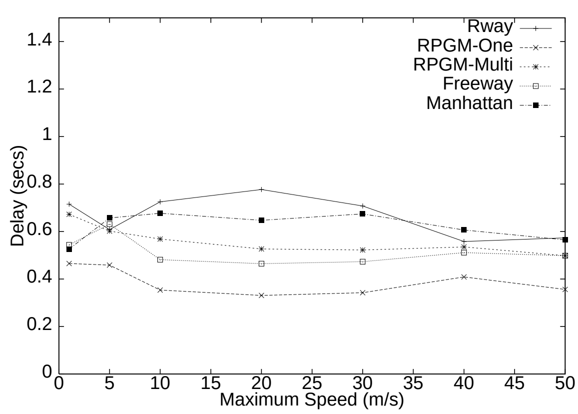


Figure 2.30: Total energy consumed versus mobility with 40 CBR sources.

Figure 2.30 shows the effect of varying node mobility on energy consumption. We vary the pause time of nodes, and with 40 CBR sources, we measure the total energy consumed by all the nodes involved in transmission or reception. ABR has marginal advantage over AODV. For both the protocols, the variation in the energy consumption is less as pause-time is increased.



(a) Average throughput of best-effort traffic vs maximum speed



(b) Average delay of real-time traffic vs maximum speed.

Figure 2.31: ABR across various mobility models

Because of the routing packets' size, one would expect ABR to perform badly with respect to energy consumption by control packets. Though ABR has larger control packets, the number of times the route re-discovery or maintenance occurs is far less than that of

AODV. This compensates for its jumbo control packets. Figure 2.30 shows that ABR has slightly better values of energy consumption, with varying pause time. Therefore, the larger size of ABR control packets does not have considerable impact on the energy consumption of ABR.

2.A.2 Scenario Based Evaluation of ABR

In this section, we study the performance of ABR under various mobility scenarios. For our experiment we consider four models: random waypoint, reference point group mobility, freeway and Manhattan.

Random waypoint is the most used mobility model in research community. At every instant, a node randomly chooses a destination and moves towards it with a chosen speed uniformly distributed in [0, Vmax], where Vmax is the maximum velocity allowed for every mobile node. After reaching the destination, the node stops for a duration defined by the "pause time" parameter.

Reference Point Group Mobility (RPGM) mobility model can be used to model most of the group movements [64]. Here, each group has a logical center (group leader) that determines the group's motion behavior. Initially each member of the group is uniformly distributed in the neighborhood of the group leader. Subsequently, at each instant, every node has a speed and direction that is derived by randomly deviating from that of a group leader [64].

Freeway and Manhattan models are similar to the ones used in the literature by Fan Bai et al. [80]. Freeway model emulates the motion of nodes on a freeway. Applications of this mobility model include exchanging of traffic status information or tracking of a vehicle. Manhattan model, however, is used to emulate the movement pattern of mobile nodes on streets defined by maps. This model is used to model movement of nodes in an urban area. The details of these two models can be found in [46].

For RPGM we used the mobility generator from [65]. Mobility generator for Freeway and Manhattan models was borrowed from [67].

Environment

For all mobility patterns, 50 mobile nodes moved in an area of 1000 m x 1000 m for a period of 900 secs. We use two versions of Reference Group Mobility Model (RPGM) in our experiments: RPGM-1 and RPGM-Multi. RPGM-1 refers to RPGM with single group of 50 nodes. RPGM-Multi refers to RPGM with 10 groups, with 5 nodes in each group. For RPGM maximum distance between the logical center and reference point is 200 m, whereas maximum distance between the reference point and mobile node is 100 m. In the freeway and Manhattan models, the node movement is controlled as per the specification of the models. Freeway model includes 2 freeways, with 4 and 2 lanes in first and second freeway, respectively. Manhattan model includes a grid pattern with 4 x 4 streets, with each street having 2 lanes. The maximum speed was set to 1, 5, 10, 20, 30, 40, 50 m/s.

Evaluation

Figures 2.31(a) and 2.31(b) show, respectively, the throughput of best-effort traffic and delay of real-time traffic, versus maximum speed. With respect to both delay and throughput, ABR performs best in the case of RPGM-1 and badly in the case of random mobility. ABR performance with Freeway and Manhattan mobility patterns are similar. There are minimum variations of delay and throughput with varying speed, for ABR with random waypoint model. The variations of average delay with varying maximum speed are relatively less for all mobility models.

ABR shows a throughput improvement up to 40% with RPGM-1, as seen in Figure 2.31(a). In addition, there is a similar difference in the average delay across various mobility models. In [46] it was shown that DSR performs worse in case of Manhattan and best in case of RPGM-1. Like DSR, ABR performs best in case of RPGM-1 but the worst-case performance is not evident. ABR performs relatively bad in case of random waypoint.

2.A.3 Study of ABR with Service Differentiation Mechanism

In this section, we present the study of ABR with the service differentiation mechanism SWAN [59].

Stateless Wireless Ad hoc Network (SWAN) model [59] uses distributed algorithms to deliver service differentiation in MANETs. In particular, it uses local rate control for TCP best-effort traffic, and source-based admission control for UDP real-time traffic. It also uses Explicit Congestion Notification (ECN) to dynamically regulate admitted real-time traffic. As nodes do not maintain any state information, there is no need for signaling or state control mechanisms like update, refresh, release. This is the reason why it is termed as a stateless approach as compared to other stateful approaches.

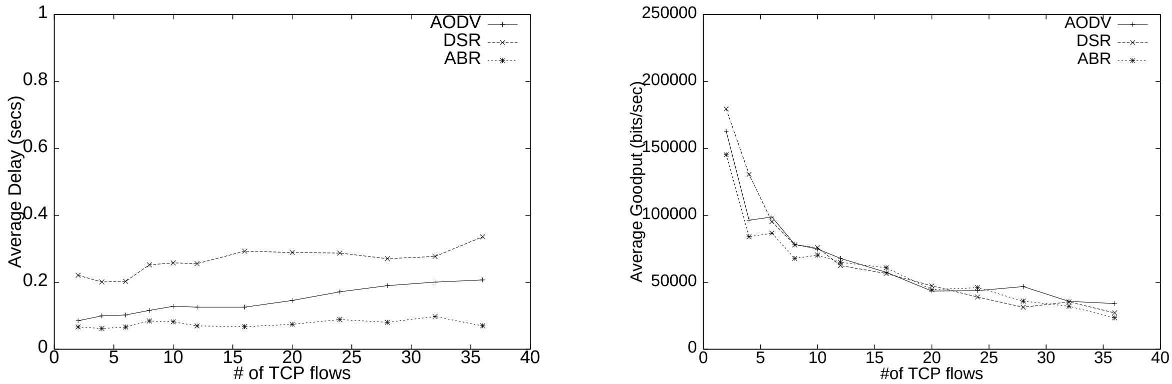
In this section, we compare the effect of AODV, ABR and DSR on SWAN. All the routing protocols share similar on-demand behavior, i.e., initiating routing activities only on the arrival of data packets in need of a route. These routing protocols, however, differ in many of their routing mechanics (route recovery, route maintenance, route reply technique).

2.A.4 Effect of Varying Best Effort and Real Time Traffic

Figures 2.32(a) and 2.32(b) show the average delay for real-time traffic and the average goodput of TCP best-effort traffic for increasing amount of TCP traffic, with a fixed amount of real-time traffic. The pause time for the nodes is 100 secs. With varying number of TCP flows, AODV provides better delays than DSR, and ABR provides better delays than both DSR and AODV. The delay variation with increasing TCP traffic in the case of ABR is also minimum. The variation of average goodput of best-effort traffic with varying number of TCP flows for DSR, ABR and AODV are almost the same.

Figures 2.33(a) and 2.33(b) show the average delay of real-time packets and the average goodput of TCP best-effort flows, respectively, for increasing amount of real-time traffic, with a fixed amount of best-effort traffic (10 TCP flows). Noteworthy points from these graphs are: greater delay variation with AODV and better delays for DSR compared to AODV when the amount of real-time traffic is high. ABR, as in the case of increasing TCP traffic, provides lesser delays than DSR and AODV. In almost all the cases, ABR has the best delay values. AODV has lesser delay compared to DSR. This advantage in delay for AODV and ABR can be attributed to their local route recovery mechanism, when compared to global route recovery mechanism of DSR. The advantage can also be traced to the features of adapting to the mobility via beacon/hello messages from neighbors.

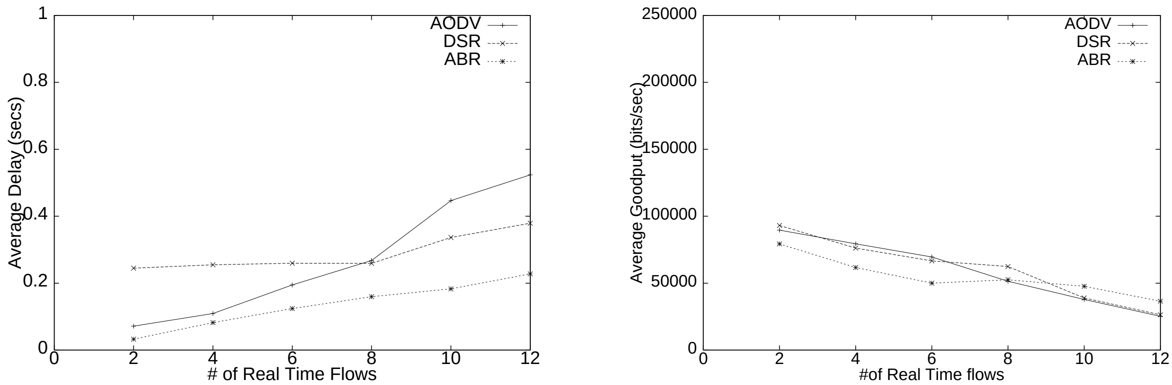
With lesser amount of real-time traffic, ABR provides relatively lower goodput values than AODV and DSR. With increased real-time traffic, however, ABR provides higher goodput values.



(a) Average delay of real-time traffic vs number of TCP flows.

(b) Average goodput of best-effort TCP traffic vs number of TCP flows.

Figure 2.32: Effect of varying best-effort traffic



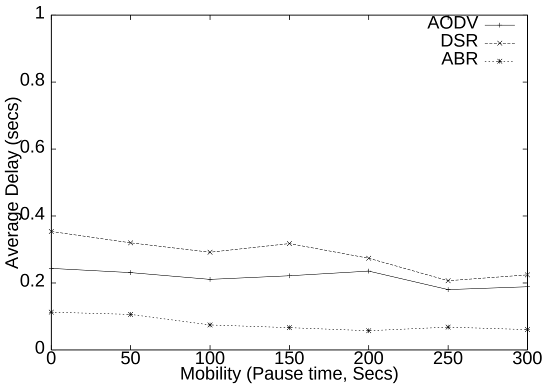
(a) Average delay of real-time traffic vs number of real-time flows.

(b) Average goodput of best-effort TCP traffic vs number of real-time flows.

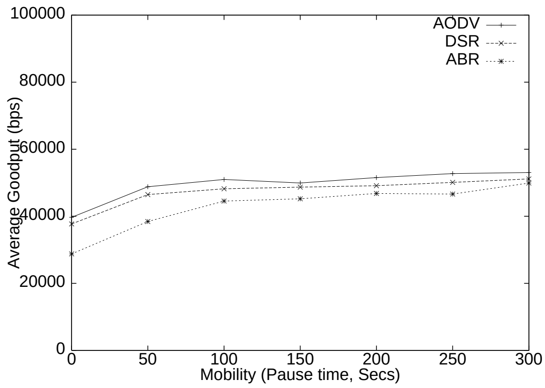
Figure 2.33: Effect of varying real-time traffic

2.A.5 Effect of Varying Mobility

The impact of node mobility is illustrated in Figures 2.34(a) and 2.34(b). The background traffic is a mixture of TCP best-effort and real-time traffic. The numbers of real-time and best-effort flows remain as 4 and 20, respectively. The pause time of each node is varied from 0 (continuously moving) to 300 secs. Default values of the SWAN AIMD parameters are used. All the protocols follow the same pattern - with increasing average delay as the mobility increases. At low mobility scenarios the average delay remains less. Similarly, the goodput is low for high mobility scenarios. As the mobility decreases, the goodput increases with all the three protocols. With varying mobility, ABR has better delay values and AODV has better goodput values.



(a) Average delay of real-time traffic vs mobility.



(b) Average goodput of best-effort TCP traffic vs mobility.

Figure 2.34: Effect of varying mobility

The impact of mobility on delay and goodput is due to the route discovery latency and congestion on new route. Route discovery latency impacts the end-to-end packet delay. Though ABR has more route discovery latency compared to AODV and DSR, the congestion and number of route discoveries are less. The ability of ABR to consider the load at each node and associativity between nodes facilitates reduction in congestion and number of route discoveries.

In the case of ABR, destination will wait for some duration, which is proportional to the number of hops the first request has traversed, before it sends a reply. This waiting time affects the throughput of the network. This is the reason for its poor performance with respect to goodput of best-effort traffic.

Experiments were conducted to see the effect of varying SWAN parameters. Varying the increment rate does not have any impact on the real-time traffic delay, whereas varying the decrement rate does have. ABR with better delay values performs best with lesser decrement rates.

In summary, we have seen that ABR can perform better than AODV and DSR with respect to many parameters. This is mainly because, shorter route may not always prove to be the best route. In addition, a stability based routing as part of a resource management solution can also prove to be useful. Therefore, considering a "correct" representation of link quality as a routing metric can be advantageous.

2.B Appendix: Lifetime Distribution Models

The theoretical model used to describe the unit link lifetimes are link lifetime distribution model [81]. This distribution model can be any probability density function (PDF) $f(t)$ defined over the range of time $t = 0$ to $t = t_1$. Ideally $t_1 = \infty$, as we collect data from simulation we consider t_1 as duration of simulation time. The corresponding CDF ($F(t)$) is useful as it gives the probability that a randomly selected link will fail by time t.

In the field of reliability engineering, the lifetime of entities are typically described using a "graphical representation" called *Bathtub Curve* [82, 83], as shown in Figure 2.35. By definition, bathtub curve is a plot of instantaneous failure rates versus time, which is used to classify the failure rates. Bathtub curve describes the relative failure rate of an entire collection of entities over a period, and not just a single entity. A bathtub curve can be described as consisting of three sections: infant mortality failures, random failures, and wear-out failures.

Typical distribution models used for lifetimes are exponential, Weibull, lognormal and gamma [62]. The exponential distribution has a single parameter whereas other distributions have 2 parameters. We mainly focus on Weibull and lognormal distributions.

The exponential distribution was used for link lifetime in majority of the previous works [84]. The probability density function of X having exponential distribution is given as:

$$f(x) = \lambda e^{-\lambda x}$$

The failure rate (also termed as Mean Time To Failure (MTTF)) reduces to the constant λ for any time. The exponential distribution is the only distribution to have a constant failure rate.

The exponential distribution is an excellent model for the second section (random failures, with constant failure rate) of the bathtub curve. These type of failures are sometimes also referred to as “Intrinsic” failures [62]. Exponential model is perfectly suited when early failure and wear-out failures are not of a concern. In our work, we show that the link failures in ad hoc networks matches well with wear-out failure model than with random failure model.

The Weibull is a flexible lifetime distribution model with **two** parameters [85]. The probability density function of X having Weibull distribution is given as:

$$f(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta}\ exp^{-\left(\frac{t}{\alpha}\right)^{\beta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Where β is called the shape parameter and is typically between 0.5 and 8, which determines the shape of the Weibull probability density function as shown in Figure 2.36. From the Figure 2.36 we can notice that several of the probability density functions displayed looks familiar. This is indeed true, as shown in Table 2.2, which shows β values and corresponding distribution which Weibull is identical to. α is the scale parameter, and is also known as characteristic life (63.2 percent of the link population fails by the characteristic life point, regardless of the value of β) [85].

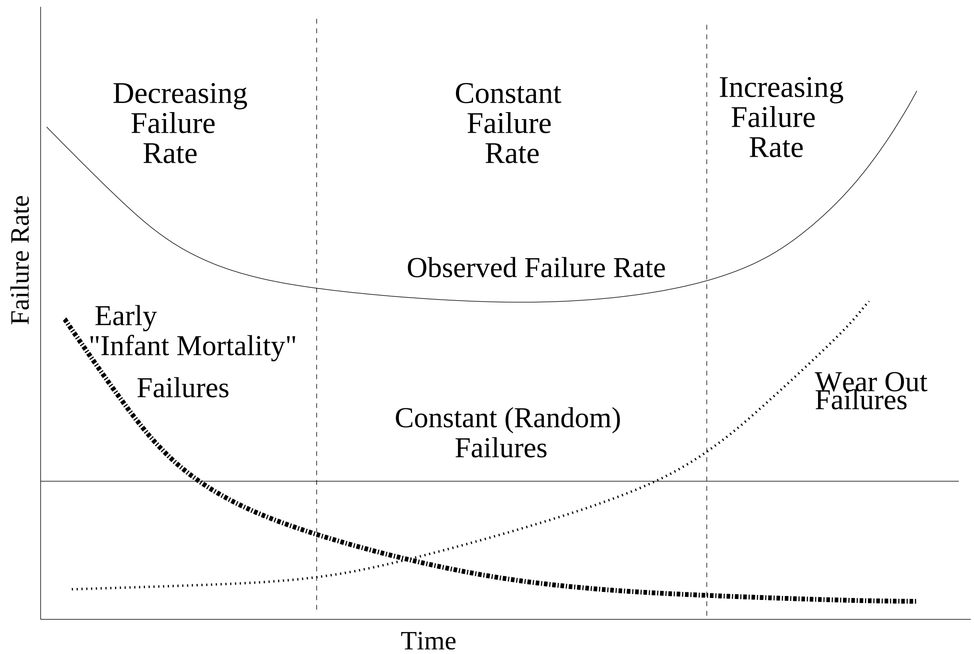


Figure 2.35: Bathtub curve

According to the central limit theorem (CLT), the probability distribution of a variable, that is the product of many independent random variables (none of which dominates the result) is lognormal [74]. That is, typically lognormal distributions are generated by processes that follow the law of proportionate effect (multiplicative process) [86]. The probability

β	Weibull Identical to
1	exponential
2	Rayleigh
3.5	lognormal
3.6	Normal
5	Peaked Normal

Table 2.2: Distributions that Weibull is identical to

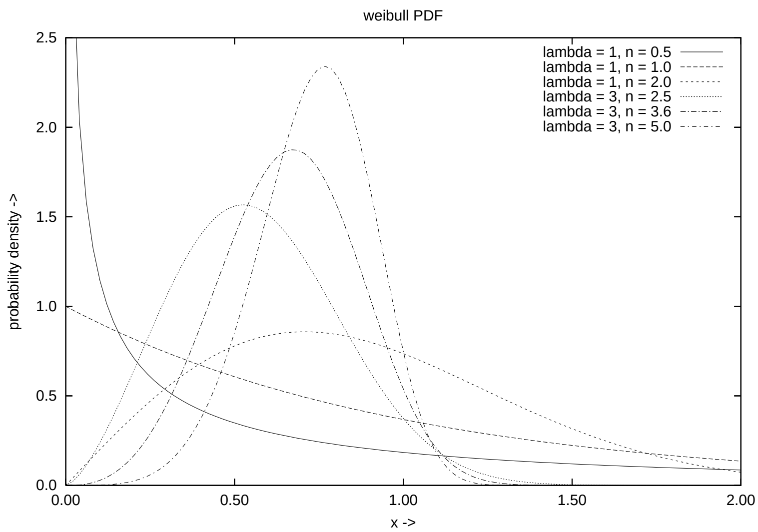


Figure 2.36: Weibull probability function

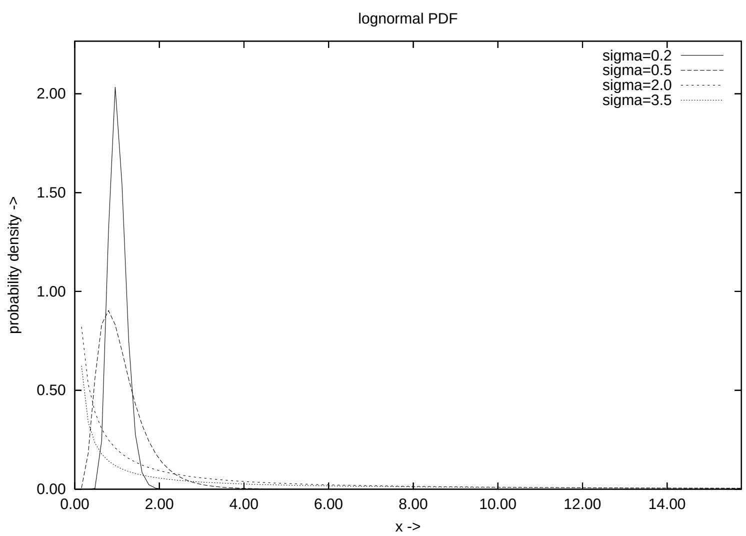


Figure 2.37: Lognormal probability function

density function of X having lognormal distribution is given as:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \, exp \, \left\{ -\frac{(ln \; x-\mu)^2}{2\sigma^2} \right\} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Here x is the variate, μ is the mean value of the log of the variate, and σ^2 is the variance of the log of the variate. To understand the flexibility of the distribution, we should consider the PDF plot, as shown in Figure 2.37, and vary the parameters. We vary the parameter σ^2, which can also be termed as shape parameters (similar to β in Weibull distribution), keeping the scale parameter (μ) fixed. Considering the conclusions that we make in succeeding sections, for smaller sigmas, lognormal shapes are similar to Weibull shapes with larger βs, and for larger sigmas give plots that are similar to Weibull shapes with smaller βs. Therefore, both the distributions are flexible and a careful decision should be made when choosing which one to use (especially for smaller samples). For larger samples, however, this difficulty on choice is reduced drastically [74].

Chapter 3

Call Admission Control

In the case of mobile ad hoc networks, some form of radio resource management is typically required to allocate the available resources to the contending flows/stations in accordance with the needs, priorities and some notion of fairness. A critical property of any admission control scheme proposed for ad hoc networks is the ability to monitor the resource usage and to determine the proper share among the contending stations. In this chapter, we will focus on the admission control mechanism, whose goal is resource management. We propose novel techniques for bandwidth estimations, and flexible fairness criteria based on equal share.

3.1 Introduction

Call admission control (CAC) problem is both unique and important in wireless networks in general, and ad hoc networks in particular. CAC schemes are critical to the growth of mobile ad hoc networks. An important problem in designing CAC scheme for ad hoc networks is that the bandwidth available to a node is not only constrained by the amount of bandwidth consumed locally, but also by the nodes which lie within the interference-range of a node. That is, in an ad hoc network, when a node is transmitting to one of its neighbors, other neighbors (within the transmission range) need to be silent. Further, a transmission by one node will affect other transmissions happening within its interference range, mainly due to the interference in the medium [87]. This feature of ad hoc network should be given importance in designing CAC scheme.

Admission control schemes typically include two processes - measurement and decision. Measurement process is the actual measuring and monitoring of the resources (bandwidth). Whereas, a decision process is a process where a set of criteria decides whether to admit a new flow. This decision is based on an appropriate model of flows, fairness criteria and measured quantities. In our work, the novelty lies more in measurement process and less in decision process.

The decision process, which is dependent on the CAC policy, is one of the important design considerations in ad hoc networks. Our decision process is based on following principle: “in mobile ad hoc networks, it is desirable that any overloaded (using more than one’s fair-share of bandwidth or hogging) or under-served (starving) user change its throughput”. That is, we believe that call admission control can be beneficial to prevent some sources from injecting “excess” traffic into the network, and to ensure that some sources do not starve in the network. Whereas, our measurement process relies on not treating the interference uniformly but classifying the interference based on estimates of the position of the interfering nodes, and noise levels at the sender and receiver.

With the rapid growth of ad hoc networks, it is expected that multimedia applications will be popular in personal networks or other collaborative scenarios [88]. There have been a class of applications developed, which are flexible, tolerant to the delay variations (email, paging [89]), and adaptive to dynamics of underlying network [90,91]. In any case, in order to obtain good network performance, some form of admission control coupled with rate control can be desirable since traffic imbalance in any part of the network can lead to localized congestion, resulting in excessive packet loss and high packet delay. Many existing unfairness and performance problem in 802.11 network can also be attributed to many interfering nodes transmitting at a high rate [92]. Further, when a flow requests some form of service, it must characterize its traffic so that the network can make a decision whether to admit the flow or not. Typically, such flows are characterized as peak rate, average rate or minimum rate. These requirements usually provide upper bounds on the traffic generated by the source, which many researchers have used to provide some sort of guarantees. Considering the MANET environment, it would be feasible if we focus on applications which are flexible, and consider only the characterization of "minimum rate". That is, when admitting a new flow, not only must the admission control mechanism decide whether the flow can get the service requested (the minimum rate), but it must also decide if admitting the flow will prevent flows which are previously admitted from getting unfair share.

Generally, call admission control should ensure that in accepting a new flow, performance of on-going flows will not be affected. The performance measure can be of different forms, including total network throughput, total network utilization, end-to-end delay, fairness among calls/users. In this work, our goal is to enhance the network throughput while maintaining low end-to-end delay and packet loss. A good admission control scheme is one which admits as many requests as possible without compromising the performance of existing requests. An overly conservative scheme maintains good performance by admitting far too few flows, while an overly optimistic scheme allows all requests to be admitted without any regard to the performance of existing flows. This is better understood by considering the classic fairness versus utilization tradeoff in ad hoc networks.

3.1.1 Fairness and Utilization Conflict

Spatial reuse of available bandwidth is useful for increasing the utilization. In mobile ad hoc network environments, however, multiple flows may transmit simultaneously, and the transmission of a flow in a " contention-region" might impact other flows existing in the same "contention-region". This feature introduces a conflict between achieving fairness and increasing the channel utilization. For example, consider a topology with three contending flows (f_1, f_2, f_3) within the same contention region, which has a maximum channel capacity of C, as shown in Figure 3.1. In order to achieve maximum utilization, we have to starve any two of the three flows and allow only one (one end of the solution space). By doing this, we can achieve maximum utilization of 'C'. Whereas, to achieve fairness, if we allow all the three flows (other end of the solution space), then it is not difficult to notice that the maximum utilization will be definitely less than 'C'. This is because of the inherent shared nature of wireless medium, which will result in collisions and losses, when multiple flows are transmitted simultaneously within a contention region. Therefore, it is important to have some sort of trade-off between these two conflicting criteria.

In this chapter, we address this trade-off between achieving fairness and increasing channel utilization. We try to increase both the channel utilization and the fairness achieved by compromising on the amount of channel allocated to each of the flows (i.e., by limiting the transmission rates). We enforce a notion of fairness that ensures that each flow (among the set of contending flows) receives proportionately fair and a minimum channel allocation. Having this constraint, we try to enhance the aggregate channel utilization. Our fairness notion is similar to the local fairness model (topology dependent) of the literature [93]. We use both position estimation and rate control to achieve the trade-off between the fairness and channel utilization.

We adopt a fluid flow model, which mandates that when a set of flows F share a channel, a flow i with rate r_i receive a channel allocation of $(C\frac{r_i}{\sum_{j \epsilon B(t)} r_j}\delta t)$ over any small time window δt, where C is the channel capacity and $B(t)$ is the set of backlogged (contending) flows at time t. We approximate this model by each node having the information of the number of contending flows. This information is used to share the bandwidth among the contending flows proportionately. We admit as many flows as possible, as long as the allocation $(C\frac{r_i}{\sum_{j \epsilon B(t)} r_j}\delta t)$ is greater than the minimum required bandwidth.

This fluid model fair share is advantageous because they will provide fairness among backlogged flows, and provide a minimum throughput for backlogged flows. By enforcing the constraint, the losses and delays will also be bounded.

3.1.2 Rate and Power Control

Rate adaptation is a process where sources increase or decrease rate depending on some feedback. This feature is often used in the majority of congestion control techniques. The origin of feedback can be either within the node or from the network. The feedback can also be either positive or negative. Therefore, depending on the type of the feedback, sources either decrease or increase the sending rate. The rate control process can be used in conjunction

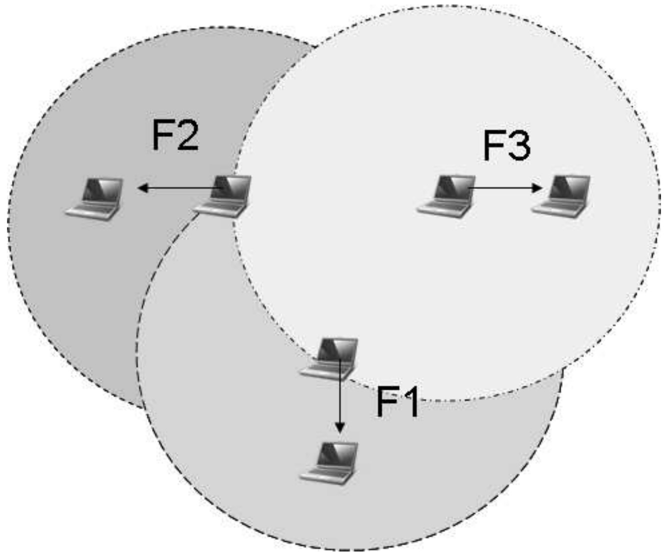


Figure 3.1: Fairness versus utilization

with the admission control mechanism. In such a case, rate control can be used to reduce or increase the traffic rate according to the available bandwidth. Hence, measurements of available bandwidth made by admission control scheme can act as a feedback to rate control scheme.

Power control scheme provides an intelligent way of determining transmitting power to achieve different goals in wireless networks. Transmitter power control in wireless communication networks makes it possible for information carrying signals to reach their intended receivers. This allows for more mobiles to coexist in the same channel, increasing the capacity of the network. Intelligent allocation of power is crucial in wireless networks for both longer battery life of the mobile devices and for increased utilization of the limited resource. In this work, we use power control to enhance the coverage area, and only for one particular message. The details of when the rate adaption and transmission power increase is carried out is provided in the succeeding sections.

3.1.3 Goals and Design Choices

There are various reasons or goals for which the CAC schemes are used for. This goal typically defines the decision policy. For wired networks, CAC schemes are used as a tool for either congestion control or QoS provisioning. Whereas, for wireless networks the CAC schemes are used for QoS provisioning in terms of signal quality, call blocking, delays, loss rate. In our work, important goals or reasons for using admission control are: control of transmission rate (attempt to provide flows with minimum transmission rate), and fair resource sharing.

CAC schemes can be classified based on various design choices. Some of the important design choices are: centralization, information scale, decision time [94]. Centralization refers to where the actual decision is taken, and it can be either centralized or distributed. Information scale refers to the span of the network from which information about the resource is collected. This span can be local, semi-local or global. Semi-local may refer to one or two hops, or a single cluster if a network is organized as clusters. Finally, CAC schemes can either be proactive (based on some set of predefined parameters) or reactive (based on active/passive measurements). The CAC scheme proposed in our work is distributed, semi-local and measurement-based (reactive) technique.

3.1.4 Multihop Considerations

In mobile ad hoc networks, due to the existence of multihop paths, it is not just enough to consider the available bandwidth at one particular node, but availability at all nodes in the path should be considered. Various ways have been proposed for this end-to-end consideration. Most popular technique is by considering available bandwidth as a concave metric. That is, end-to-end available bandwidth is computed as the minimum value over a path P and is formally defined as $B_E = min(B_i), i \epsilon P$, where B_E is end-to-end available bandwidth, and B_i is the bandwidth available at node i along the path P. In our work, for multihop scenarios we compute the end-to-end available bandwidth by using the same approach.

Call admission control (CAC) has been extensively studied for wired networks. Although admission control solutions have been proposed for wireless ad hoc network [59], most of them are solutions aiming for QoS provisions, and are solutions for wired network ported for ad hoc networks. Inherent features of wireless ad hoc networks such as multiple access interference from nodes in the transmission and sensing range, change in topology and existence of multiple hops, make CAC in wireless ad hoc networks a difficult task. We believe that this difference between wired and wireless networks should be given importance in designing the solutions. In this chapter, we introduce a call admission control mechanism for wireless ad hoc networks called interference-based call admission control (iCAC). *Shared wireless medium* feature of MANETs is considered in iCAC.

iCAC is unique because it does not treat interference uniformly instead classifies interference based on estimates of the position of the interfering nodes, and noise level at the sender and the receiver. In addition, a simple fair allocation scheme that takes into account of these estimates is also used to provide a fair bandwidth allocation for each admitted flow. These features help iCAC to have a better estimate of available bandwidth, and also prevent overloading of network.

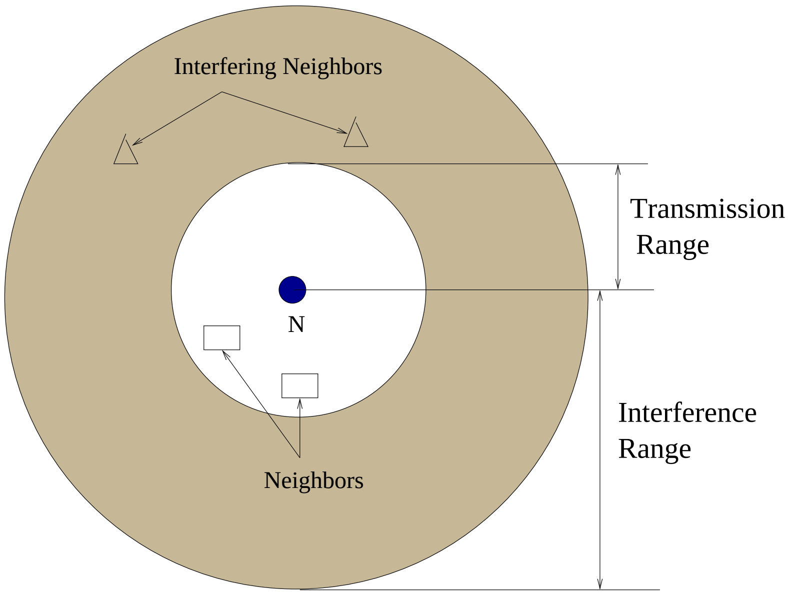


Figure 3.2: Approximate ranges for a wireless node N

The rest of the chapter is organized as follows. Section 3.2 presents some background information and all the important related works. Section 3.3 describe the model used for bandwidth measurement. In Section 3.4, we describe the model for bandwidth sharing. Section 3.5 explains in detail the available bandwidth estimation process. Section 3.6 presents the iCAC algorithm, describing the overall admission control process. Section 3.7 provides the evaluation of iCAC in comparison with the existing call admission control scheme. Finally, in Section 3.8, we conclude by summarizing the chapter.

3.2 Background and Related Works

Before describing the related works, we would first provide some background information about the various "ranges" (distances), for packet transmissions and receptions, and also the effectiveness of IEEE 802.11 RTS/CTS mechanism. We believe that this helps in understanding both the proposed and the related work.

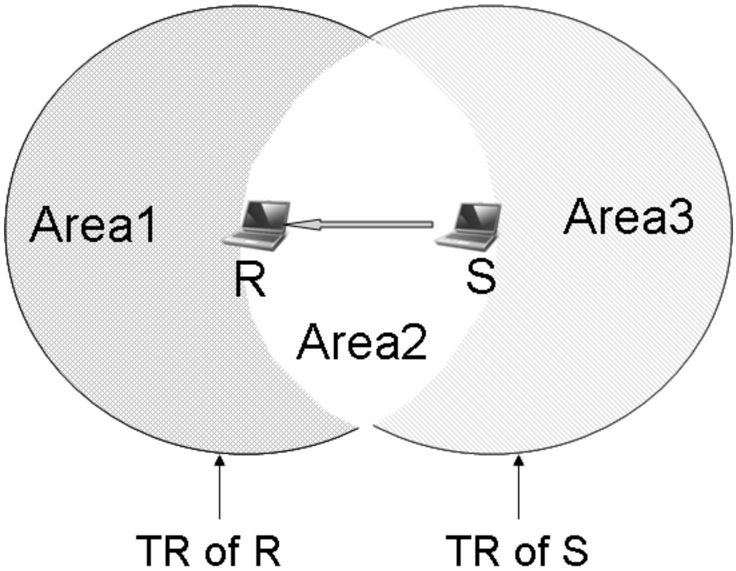


Figure 3.3: Effectiveness of IEEE 802.11

From Figure 3.2, we can see that for any node N, there are 2 notable ranges for wireless communication. Each range is important for measuring the channel utilization and predicting the available bandwidth. The smaller of the ranges is termed as transmission range (TR), sometimes also referred to as reception range. Nodes present within this range are termed as neighbors, and node N can communicate with these neighboring nodes directly. The bigger range in Figure 3.2 is referred to as interference range (IR) (also termed as carrier sense range). Nodes that are within the carrier sense range (termed a interfering neighbors or far neighbors) of node N can "sense" the packet transmission of N. By sense we mean that these nodes can detect the packet but may not be able to decode the packet. We know that larger IR prevent multiple transmissions to occur simultaneously, whereas, smaller IR allows more spatial reuse. For correct reception, the area surrounding the receiver must be free of interfering transmissions.

To understand the effectiveness of RTS/CTS mechanism of IEEE 802.11, let us consider three areas for existence of interfering flow I_f within the transmission ranges of sender S

and receiver R, as shown in Figure 3.3. In Area-1, receiver R can receive packets from the I_f, which are above receive threshold (minimum power for the packet to be received, also termed as Rx Threshold) value. Therefore, RTS/CTS packets sent by I_f can be received by R, which will prevent R from sending CTS to sender S. Similarly I_f can receive the CTS sent by R, and avoid initiating transmission while S is transmitting the data packets. Therefore, RTS/CTS mechanism will provide some sort of co-operation in this case. In Area-2, both S and R can receive the packets from I_f, which are above the receive threshold, and I_f can receive packets above receive threshold from both S and R. This is the case where “maximum” co-operation can be achieved between nodes, through RTS/CTS mechanism. Area-3 is similar to Area-1, but the co-operation level is much higher compared to Area-1 and lower compared to Area-2. Here, sender S can receive packets from I_f above receive threshold, and I_f can receive packets that are above receive threshold from sender S. The RTS packet sent by sender S is received by I_f in Area-3, and will defer from initiating transmission. Similarly, sender can receive RTS/CTS from I_f in Area-3, and defer from initiating transmission.

The whole idea is, if either the sender or the receiver can hear and understand the interfering transmission, then they can have some control over the interfering flow, and achieve some sort of co-operation and resource sharing. With this basic understanding, we proceed to explain all the important related works.

We first present the related works, which consider resource management through call admission control for wireless ad hoc networks. Further, we also present works which study different aspects of ad hoc network capacity. We include these studies (network capacity) because, it gives insights to important aspects such as achievable throughput, role played by

MAC protocols, and the effects of interference and mobility on the capacity. Further, many simulation studies in our work matches with some of the important conclusions given by the earlier works on network capacity.

The literature [94] provides a detailed survey on CAC schemes for wireless networks. Typically, call admission is carried out by having a measure of how much the resource has been used by the existing flows and how much resource is available for a new flow. The key concepts in these admission control algorithms are how resource availability or network utilization is measured. The various proposals for the measurement usually involves one or more of the following parameters:

- Bandwidth estimation.
- Throughput
- Transmission delay.
- Queue length or load at the node.
- Collisions.
- Power control.
- Signal to interference ratio.

A common approach to estimate the available bandwidth measurement is to measure the channel busy time [95–97]. Let T be the sampling time-window, T_{idle} be the duration for which radio was in idle state in the last time-window T, T_{busy} be the duration for which radio was in busy state in the last time-window T, and W be the maximum bandwidth. ABW, the available bandwidth can be computed as follows.

$$ABW = \frac{(T - T_{busy})}{T} * W \tag{3.1}$$

The authors in [98] measure the throughput (Th) of a transmitting packet as:

$$Th = \frac{S}{(t_r - t_s)} \tag{3.2}$$

where S is the size of the packet, t_s is the time-stamp at which the packet is ready at the MAC layer, and t_r is the time-stamp at which an ACK is received. They claim that the time duration $t_r - t_s$ includes the channel busy and contention time. They maintain separate throughput estimates for each and every neighbor. This throughput measurement is assumed to reflect the available bandwidth for the new flow. For this technique, it is important to make the throughput measurement independent of the size of the packet by normalizing the packet sizes.

Proposal by Sun et al. [88] considered both the load at each node and predicted delay values to measure the network utilization, and used these information to carry out admission control mechanism. Each node maintains a neighbor set, and also the load information of each neighbor. Load information is in terms of number of service flows, and is also associated with *confirmed, pending, and unknown* states. When a request for new flow comes, based on this flow information of all neighbors, a node will predict the delay value. There have been proposals of constraint based routing which consider the load at each node [99]. Further SWAN [59] also considers load at each node for admission control decision.

Measure of average collision ratio is another technique used to estimate the network utilization. Similar to bandwidth measurement, a sampling period T is maintained. The average collision ratio (R_c) is defined as number of collision occurred over the total number of transmissions (including retransmissions). Therefore,

$$R_c = N_c/N_t \tag{3.3}$$

where N_c is the number of collisions, N_t is total number of transmissions. The collision ratio indicates how much the network is loaded. Typically thresholds are set for the collision ratios, and the admission decision is taken depending on these threshold values [100]. There are also works (Dent et al. [101] and Xu et al. [95]), which study the network utilizations focusing purely on IEEE 802.11 MAC protocol.

Power control techniques for call admission control are proposed for CDMA networks. Power control techniques typically refer to controlling the transmission power. It can be carried out in two ways: when a new call arrives, it is allowed to transmit with the limited power, followed by gradual increases. If the target SIR is achieved, the call is admitted, otherwise it is blocked. The second approach is to have a measure of received power to indicate the interference level. Based on this measure, call admission decision is taken [102, 103].

In our work, we consider network utilization by estimating the bandwidth available by using the measure of busy times. We, however, differ in the busy time definition from other approaches, and consider two versions of busy times, which will be explained in the succeeding section. *Along with busy time measurements, we also measure the noise values at both sender and receiver to estimate the position of interfering nodes, which also helps in* *available bandwidth estimation.* This estimation of position of interfering nodes for available bandwidth measurement has not been used by any of the previous approaches. We also consider the notion of fairness in admission control scheme, which is rarely considered by previous approaches.

In the remaining part of this section, we discuss some of the works on ad hoc network capacity. Gupta and Kumar [104] first derived an upper bound on the maximum possible transmission capability achievable by any static ad hoc wireless network. The results, though ignores lot of practical difficulties, offers important insights to the problem. They consider two types of networks, arbitrary (node positions, source and destinations, and traffic requirements are all arbitrary), and random (sources and their destinations are randomly chosen). The authors consider a protocol model (involving only distances) and a physical model (involving transmission power, SINR, and distances), and give an upper bound on the transmission capacity. They show that average available throughput per node decreases as the square root of the number of nodes n, equivalently, the network capacity increases as at most $\sqrt{n}$. The result holds true irrespective of the topology, power control policy or any transmission scheduling strategy. Further, authors also showed that adding relay-nodes in the network increases the total network capacity, thus increasing the share of available bandwidth to each sender.

Grossglauser and Tse [105] showed that the mobility helps in increasing the capacity of the ad hoc network. They also showed that it is possible for each sender-receiver pair to obtain a constant fraction of the total available bandwidth, independent of the number of pairs (note that delay can be arbitrarily large).

Li et al., [106] provide good results on the capacity of the ad hoc networks. The authors consider different network topologies - 2 node, chain, and lattice. They study the throughput obtained and maximum channel utilization varying different parameters such as interference range and number of nodes.

Hekmat et al., [107] compute SIR by considering the number of nodes, node density, multi-hop characteristics, and the amount of relay traffic. This is computed considering a regular lattice, and authors conclude that the interference is upper-bounded in ad hoc networks that use carrier sensing for medium access control.

3.3 Model for Bandwidth Measurement

In this section, we will first present the radio state model, followed by the key concepts used in bandwidth measurements.

3.3.1 Radio State Transition

We will introduce the radio state transition diagram which is derived from the radio layer implementation of the GloMoSim [12] simulator. We will use this state diagram to explain our busy period consideration for bandwidth measurements.

Let us consider the state diagram as shown in Figure 3.4. The radio initially starts with the *Idle* state, from which it can either go to the *Receive*, *Transmit* or *Sensing* state. When a signal sent by any of the neighboring nodes arrives, it compares the signal power with the receive threshold. The thresholds can be better understood by referring to Figure 3.5, which

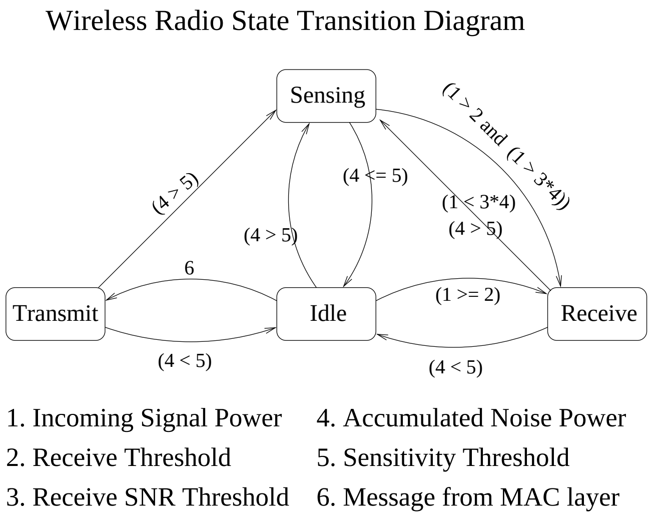


Figure 3.4: Radio state transition

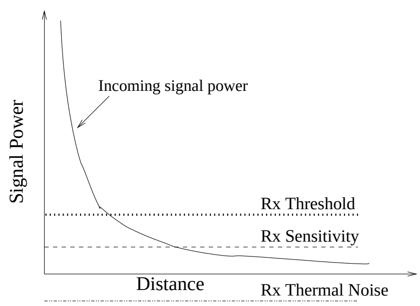


Figure 3.5: Physical parameters to determine communication range

indicates the usage of physical layer parameters to estimate the communication range. If the incoming signal power is greater than the receive threshold, it moves to the *Receive* state, or else it accumulates this power value as the noise value. If this accumulated noise value is greater than the radio sensitivity threshold (minimum power for a packet to be sensed - Rx Sensitivity), then the radio moves from the *Idle* state to the *Sensing* state. From the *Idle* state, if the node receives a message from upper layers to transmit, it moves to the *Transmit* state. Radio can change its state from *Sensing* to *Receive* if the incoming signal power is greater than both the receive threshold, and SNR threshold times the accumulated noise. The state, however, changes back from *Sensing* to *Idle*, if the accumulated noise is less than the sensitivity threshold.

After the transmission, radio changes its state from *Transmit* to either *Idle* or *Sensing*, depending on whether the accumulated noise is lesser or greater than the sensitivity threshold, respectively. Similarly, radio state can change from *Receive* to either *Idle* or *Sensing*, under the same conditions.

A key observation in this radio transition model is that the power of interference and noise is calculated as the sum of all the signals arriving at the radio, along with the one being received, and adding it with thermal noise. The resulting power is used as the base of SNR, which determines the probability of successful reception of the signal. The noise and interference model used can be mathematically written as [12]:

$$SINR = \frac{P_{incoming}}{(\sum P_{othersignals}) + (F * K * T * B)} \tag{3.4}$$

Where K is the Boltzman constant, T is the temperature, F is the noise factor of the radio and B is the effective noise bandwidth.

As a result, it is possible for two flows within interference range to transmit at an aggregate throughput much higher than the case where flows are within transmission range, with low probability of packet corruption (due to noise).

3.3.2 Use of Bandwidth with Sensing as Idle (BSI)

As explained earlier, measuring the radio busy duration is the popular approach for measuring the available bandwidth. In our work, we also take the busy duration measurement approach, as given by equation 3.1. The way we measure the busy duration is different and is explained in the succeeding paragraphs.

All the earlier proposed schemes on bandwidth measurement using the measure of busy times consider *the sensing state same as the receive state*. Therefore, *the sensing period is considered as part of the busy period*. Such assumption, however, is highly conservative as sensing state is different from receive state, and if detailed classification is performed, the available bandwidth can be increased substantially.

In our work, we call the available bandwidth measurements using equation 3.1 as *BSB* (*Bandwidth considering Sensing as Busy*) and BSI (*Bandwidth considering Sensing as Idle*). BSB provides the lower bound on available bandwidth, whereas BSI provides the upper bound. Depending on the other measurements to be presented, actual available bandwidth is somewhere in between these two values.

Besides providing an upper bound on the available bandwidth, BSI plays an important role in providing some hints on the position of the interfering nodes, which can be used to improve the available bandwidth estimate. For example, let MAX be the maximum bandwidth available. If $BSI = MAX$, and $BSB < MAX$ all interfering nodes are outside the transmission range. On the other hand, if $BSB = BSI$ and the measuring node senses little noise, then all interfering nodes are within the transmission range. In addition, if BSI is y and BSB is x, then we know that the bandwidth $y - x$ is consumed by the interfering nodes outside the transmission range. The relationships between BSB and BSI are used in first level classification in the algorithm presented in Section 3.6.

3.4 Model For Bandwidth Sharing

In a wireless environment, proper co-operation is possible only when two nodes are within each other's transmission range, and when MAC protocols like 802.11 where CSMA/CA coupled with RTS/CTS are used for coordination.

The implementation of bandwidth sharing depends on the notion of fairness that is adopted. There have been various fairness notions in wired network, such as max-min fairness and proportional fairness [108]. The fairness problem is exacerbated in MANET scenarios because of the dynamic changes in topology. Even if the topology remains static, however, interference within transmission and *sensing range* make the problem of fairness more difficult to solve in the wireless network domain. In mobile ad hoc network, the contention nature, the MAC protocol used and the positions of interfering nodes decide the possible extent of fairness that can be achieved. It would be difficult, if not impossible, to implement fairness directly at the network layer, without control over the MAC layer or link scheduling algorithm. Therefore, we do not aim to provide any bandwidth guarantees such as max-min fair allocation of bandwidth, but we aim to provide some notion of fairness so that no node will be starved unnecessarily.

In this section, we will present our model for estimating bandwidth consumption. Our model is unique in the following ways. We classify interference of neighboring nodes into different categories and each of these categories are treated differently. In addition, we attempt to share the bandwidth "fairly" by estimating the fair share bandwidth available.

The overall available bandwidth computation is based on the concept of fair (or equal) share. Fair share also ensures that no admitted flow will be starved. In addition, a fair allocation has the advantage of encouraging better spatial reuse. Consider, for example, the simple case where 6 nodes (A-B-C-D-E-F) are arranged in a straight line. The distance between the nodes is such that the neighboring nodes are within transmission range (e.g. A and B) and nodes one hop away are beyond transmission range but within interference range (e.g. A and C). Nodes separated by two or more hops (e.g. B and E) are beyond each other's interference range. Consider three active flows, flow 1 goes from A to B, flow 2 goes from C to D and flow 3 goes from E to F. Maximum bandwidth available is 2 Mbps. If flow 2 is allocated 2 Mbps by the admission control algorithm, then flows 1 and 3 will be starved. If a fair share of 1 Mbps is given to all three flows, then the aggregated throughput is 3 Mbps (50% increase). It is true that if flows 1 and 3 are allocated 2 Mbps, the aggregated throughput will be 4 Mbps but flow 2 will be starved.

In iCAC, each node of the route computes the fair share of the bandwidth available by estimating the number of active flows (*senders*) within its transmission or interference range, depending on the estimation of position of interfering nodes. If N flows are estimated to be within range, a flow i with rate r_i will receive a channel allocation of $C\frac{r_i}{\sum_{j \epsilon N} r_j}\delta t$ over some time window δt, where C is the channel capacity and N is the set of backlogged flows. The available bandwidth for a particular (multi-hop) flow is the minimum available bandwidth over all hops.

iCAC admits as many flows as possible, as long as the allocation is greater than the minimum required bandwidth. That is, we define the following utility function:

$$Maximize\ F,\ such\ that,\ r_i >= MIN_i$$

$$and\ SUM(r_i) <= C, where\ i\ \in\ F.$$

where, F is the total number of flows admitted, $F \in N$, r_i is the rate allocated to flow i and MIN_i is the minimum rate required for flow i.

The basic idea of fair sharing is simple, but the implementation is not straight forward because in order to compute the correct fair share, the number of interfering sender and receivers, and their relative positions needs to be known. In our approach, using BSB, BSI and noise measurements, a node obtains the information of all the other contending nodes and depending upon the location of the contending nodes, the node decides how sharing should be done.

In our scheme, when all the interfering flows (senders) are within the transmission range of either the sender or the receiver, we take advantage of the effectiveness of the RTS/CTS handshake mechanism. For all the other cases, however, we need to have an estimate of number of existing flows, and based on this information we carry out the fair bandwidth sharing. The detailed description of the various cases considered are presented in the succeeding section.

3.5 Estimating Available Bandwidth

In the previous sections, we described the model for bandwidth sharing, and bandwidth measurement. Apart from BSI and BSB values, in this work we also use the noise values at sender to estimate the positions of the interfering flows, and decide on the available bandwidth. In this section, we will first describe the importance of the noise level values at sender and receiver. Further, we will explain our bandwidth estimation technique based on these measurements, considering various cases.

3.5.1 Measurement Setup

In order to illustrate the effect of noise level and position of interfering nodes, the model shown in Figure 3.6 will be used. We will use this topology also to highlight various cases of available bandwidth measurements. Let nodes S and R be the sender and receiver, respectively. All the measurements are carried out on S and R. Remaining nodes are termed as interfering nodes. In the model there are 3 circles around S - the innermost circle with nodes represent the set of nodes within the transmission range, numbered as $I_i \dots I_n$.

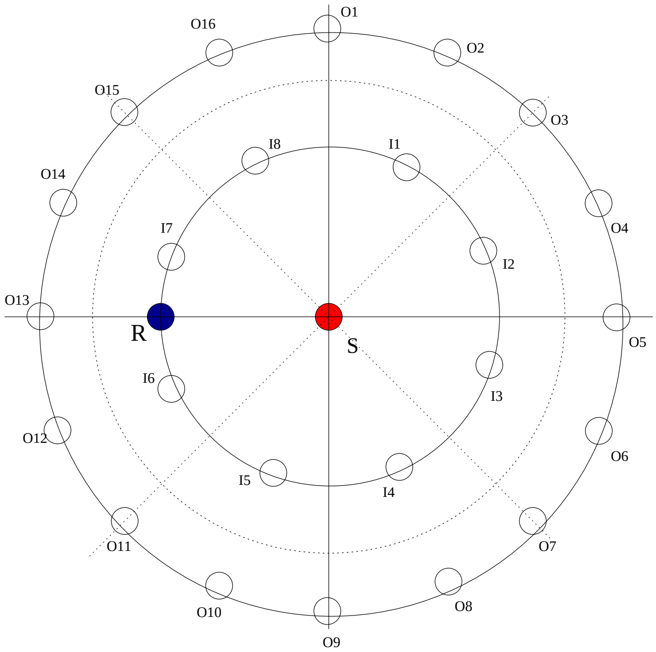
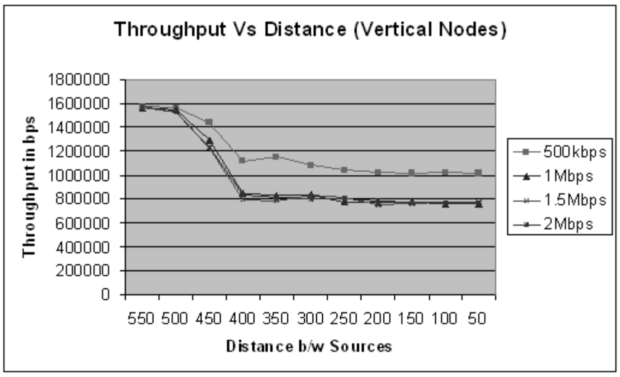


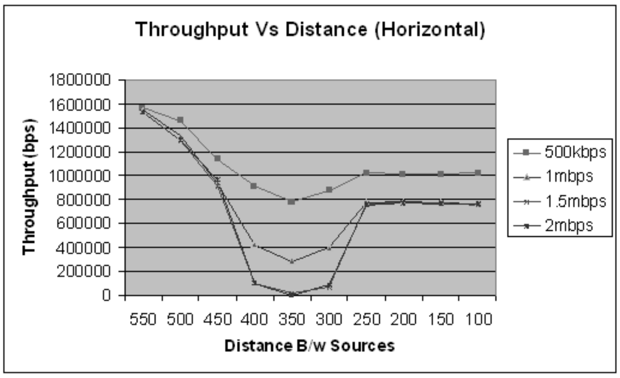
Figure 3.6: Topology used for illustrations

The next dotted circle represents the transmission range of sender S. The outermost circle represents the interfering nodes that are outside the transmission range and within the sensing range, and are numbered as $O_i \dots O_m$. In our model we consider $n = 8$ and $m = 16$, with the diameter of outermost circle as 300 m, and inner circle is of diameter 200 m. The transmission range of all the nodes are 250 m. The transmission between interfering nodes always occur from $I_i \rightarrow I_{i+1}$ or $O_i \rightarrow O_{i+1}$. Unless mentioned otherwise, in all cases we will consider only one interfering pair interfering with the transmission between S and R at any given time.

The distance between the sender node S and the interfering flows (200 m and 300 m in the above setup) play an important role. The Figure 3.7 shows the plots for throughput obtained by sender S (transmitting at 2 Mbps) and varying the distance between the sender S and the interfering flow sender, and considering for different transmission rates of interfering flows. Figures 3.7(a) and 3.7(b) show the plots varying vertical and horizontal distances,



(a) Vertical distance



(b) Horizontal distance

Figure 3.7: Effect of distance between S and interfering flows

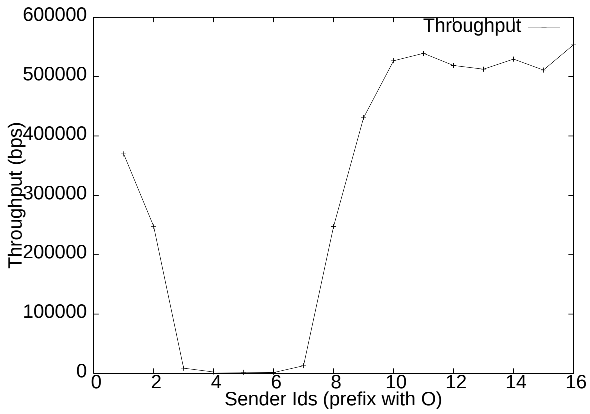


Figure 3.8: Throughput for different interfering pairs

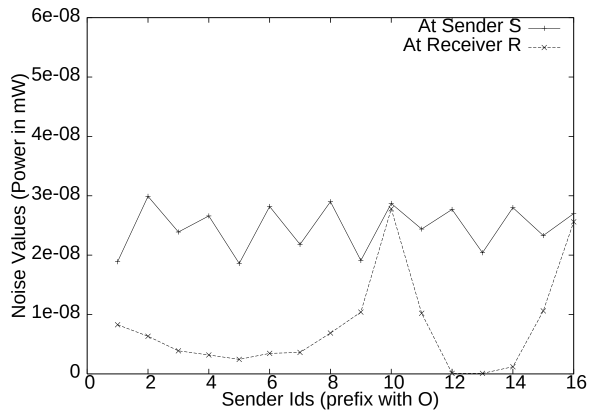


Figure 3.9: Noise values for different interfering pairs

respectively. Therefore, the distances of 200 m and 300 m were chosen considering these two plots.

3.5.2 Noise Levels at Sender and Receiver

Let us consider interfering nodes out of transmission range of the sender S, that is the nodes on the outermost circle (O_1 to O_{16}). Let the noise level at the sender be denoted by N_S and the noise level measured at the receiver be NR. As mentioned earlier, at any given time we consider only one interfering pair, and also any node with id O_i transmits to its neighbor O_{i+1}. Figure 3.8 shows the throughput obtained for S. The x-axis shows the interfering sender id. The farther the interfering nodes are from the receiver, the lower the throughput is. In the cases where the interfering nodes are out of the carrier sense range of the receiver R (O_3 to O_7), the throughput degrades significantly to almost zero. The reason for such low throughput is that since R does not sense the signal from the interference pair at all, the packet R sent has a high probability of collision during the reception by S. For the other positions (O_1, O_2, O_8 and O_9), R can sense the interference and hence can achieve higher throughput. For positions O_{10} to O_{16}, R is within transmission range of the interfering pair and the RTS/CTS protocol works correctly, thus achieving the highest throughput. The result is the same when the positions of S and R are swapped.

There are three notable cases where the noise level at the receiver R is zero or low. The first case is the trivial case where there is no interfering pair within the sender and the receiver. In this case, $NS = NR = 0$. The second case is where the receiver is out of the sensing range of the interfering pair, as described above. In this case, $NS >> NR > 0$. Finally, when the interfering pair is within the transmission range of the receiver as the RTS/CTS protocol will coordinate accesses ($NS >> NR = 0$).

Figure 3.9 shows the noise level for the sender and receiver and provides another illustration of these observations. When the noise level is strong enough or almost zero, the throughput is high, as accesses are coordinated. When the noise level is below the detection threshold, as in the case of O_3 to O_7, interference is not taken into account and accesses are completely uncoordinated, resulting in low throughput.

From these observations, we see that when the sender S has a measure of available bandwidth and finds that the interfering nodes are out of transmission range, it has to also check the noise level experienced by itself and its receiver. This noise level helps in finding whether the interfering nodes are close to receiver. In that case, sender can have better throughput compared to interfering nodes being far away from the receiver. In summary, the noise levels or the difference in noise levels between sender and receiver will help in predicting the position of the interfering nodes.

Further, the reason why noise at R drops when O_{11} is transmitting as compared to O_{10} is due to boundary effect. Position of O_{10} and O_{11} are boundary cases (similarly O_{15} and O_{16}), where some packets received at the receiver R are higher than the *receive_threshold*, and some are lesser than the threshold. This causes the difference in noise values.

Finally, we would like to mention that the measurements presented here are obtained from the GloMoSim simulator [12]. We have run the same experiments on NS-2 [11] but have obtained different results. As we believe that GloMoSim provides a accurate radio propagation simulation model, we will only use these results.

In the remaining part of this section, we will describe how iCAC estimates bandwidth based on the number of active flows, values of BSB and BSI, and noise measurements at sender and receiver. To simplify notations, we will refer to the set of all interfering flows as I_f, and the sender and receiver node as S and R, respectively. In general, estimation and coordination is possible only when a interfering sender is present within the transmission range or either S or R. We classify positions of senders in I_f with respect to S and R, whereby estimation and coordination are effective, into the following categories:

- **Case 1:** All senders of I_f are within the transmission range of S.
- **Case 2:** All senders of I_f are beyond the transmission range and within the interference

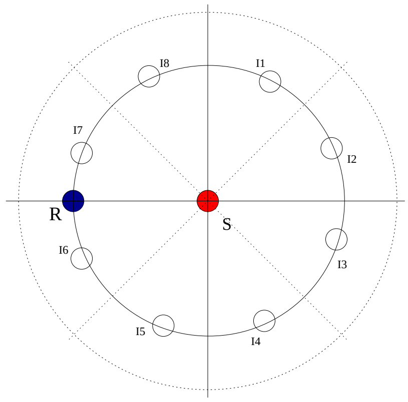


Figure 3.10: Interfering pairs inside the TR of S

range of S.

- **Case 2A:**All senders of I_f are within the transmission range of R
- **Case 2B:**Some sender(s) beyond the transmission range of R.

- **Case 3:** Senders are both inside and outside the transmission range of S (but within the interference range).

The above listed cases are complete (it covers all possible cases of existence of interfering flows) and also that the effectiveness of RTS/CTS is limited to cases 1 and 2A, whereas for cases 2B and 3 it is not effective.

3.5.3 Case 1: All Senders of I_f are Within the Transmission Range of S

In this case, the senders of the interfering flows are within the transmission range of the sender S, as shown in Figure 3.10. The positions of the interfering receivers do not matter. It should be obvious that this case can be identified by two conditions. First, the noise values at sender S is zero (because the interfering senders are within the transmission range). Second, both the bandwidth measurements should be equal and greater than 0, $BSI = BSB > 0$. Further,

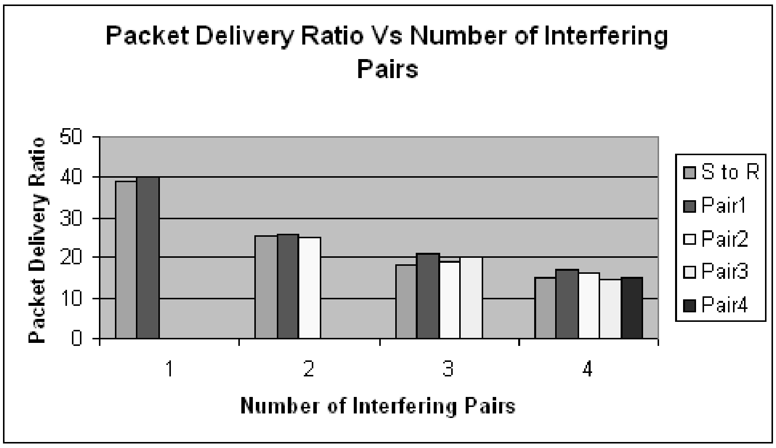


Figure 3.11: Case 1:All nodes within the transmission range

this is the case where the RTS/CTS scheme is most effective. For example, according to the Figure 3.11, if there is an increase in the number of interfering pairs, within the transmission range of sender S, the individual share that each flow gets is reduced proportionately.

There are still two potentially notable cases, however, where the I_f may be outside the transmission range of R (I_2, I_3), or inside the transmission range of R $(I_1, I_4, I_5, I_6 \text{ and } I_8)$. In these two cases, BSB and BSI at sender will almost be equal. Whereas, in the former case NR will be greater than NS. We carried out a detailed study of both cases and found that irrespective of I_f being inside or outside the transmission range of receiver R, as long as it is within the transmission range of sender S, S and I_f can share the bandwidth equally. We allow new flow, which will share the bandwidth with other flows, through IEEE 802.11 RTS/CTS mechanism.

Let us see the former case ($I_f$outside the transmission range of R) in detail. Consider the plot as shown in Figure 3.12, which shows the throughput achieved by S with varying transmission rates, with I_2's transmission rate fixed to 2 Mbps. We can see that as we increase the transmission rate of S, the throughput achieved also increases, affecting the

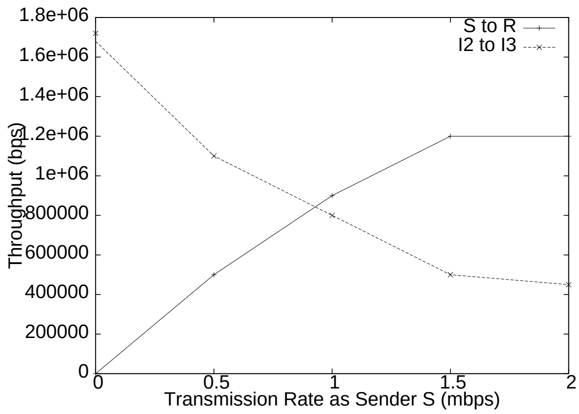


Figure 3.12: Case 1: I_f outside TR of R

transmission of I_2. The reason why S gets a better share is that, when S sends the RTS packet all the nodes (R, I_2 and I_3) can receive the RTS packet, and when I_2 sends the RTS packet, only I_3 and S can receive it. The chances are high that I_3 after receiving 2 RTS packets, may restrain/delay from sending the CTS packet to I_2, whereas R (which will receive only one RTS) will send the CTS packet to S. This will make S gaining the better share of bandwidth. Therefore, it is important in this case that S limits its rate so that both I_2 and S have fair share.

For all the scenarios in Case 1, the available bandwidth estimation is achieved by continuously monitoring the number of neighbors who are senders, and setting the available bandwidth to $\frac{maximum-bandwidth}{number-of-senders+1}$.

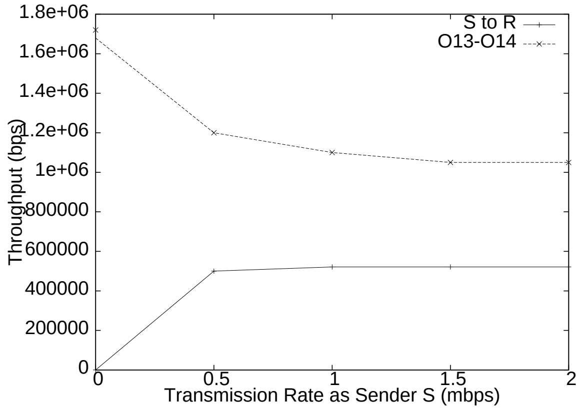


Figure 3.13: Case 2A: I_f inside the transmission range of R

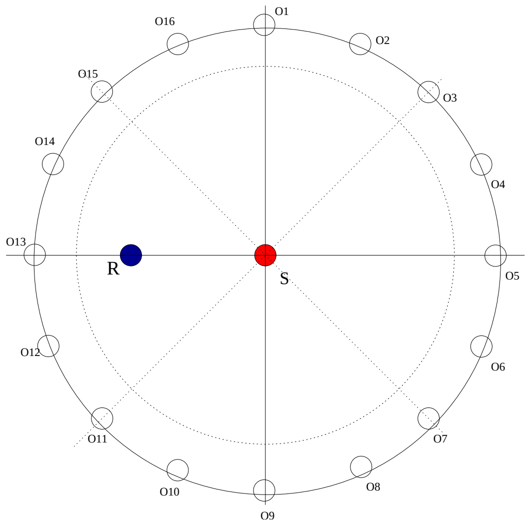


Figure 3.14: Interfering pairs outside the transmission range of S

3.5.4 Case 2: All Senders of I_f are Beyond the Transmission Range and Within the Interference Range of S

Case 2A: All Senders of I_f are Within the Transmission Range of R and Outside the Transmission range of S.

When the senders of I_f are beyond the transmission range of S but within the transmission range of R, the flow from S to R is at the mercy of the senders in I_f.

According to our BSI definition, we can see that this case can be identified by BSI = MAX (no interfering nodes within transmission range of sender) and the noise level at the receiver R is zero (or low). The case where only interfering receivers are present is not considered because coordination will not be effective.

From Figure 3.9, we can see that there are two interesting cases where the noise level at the receiver R is zero or low. The first case is where the receiver is out of the sensing range of the interfering pair, O_4, O_5 and O_6 in the Figure 3.14. In this case, $NS >> NR > 0$. In the second case, the interfering pair is within the transmission range of the receiver and the

RTS/CTS protocol will coordinate accesses $NS >> NR = 0$ (O_{10} and O_{11} in the Figure 3.14). This is the case where coordination is possible.

Let us describe this case (interfering nodes are within the transmission range of the receiver R) with an example. According to the Figure 3.14, when O_{12} or O_{13} transmits, the noise at R will be null and S will be high. In this case, the achievable bandwidth is ideally greater than BSB. The throughput achieved by O_{13} and S with increasing rate at S is shown in Figure 3.13. In this plot, O_{13} sends at maximum rate and transmission rate at S is varied. If both flows are allowed to send at the maximum rate, O_{13} will get a higher share of the bandwidth because R will receive RTS from both S and O_{12} and is therefore more restrained from replying (to send CTS) to S.

The fair share for this case is computed as setting the achievable bandwidth to (MAX/(number-of-senders in transmission range of R + 1)).

Case 2B: Some Interference Senders are Beyond the Transmission Range of R

This case considers the scenario where the interfering pairs are beyond the transmission range of sender S and there are some interfering sender(s) beyond the transmission range of R and within the interference range of S. The case is identified by $BSI = MAX$ and $NR > 0, NS > 0$. In the earlier two cases (Case 1 and Case 2A) we took advantage of effectiveness of RTS/CTS scheme. Whereas, in this and the following cases (Case 3), we consider scenarios where RTS/CTS scheme is not effective.

We propose to handle this case by two methods. In the first method we follow the same model of bandwidth sharing (fair share) as we did for the previous cases. Whereas, in the second method we will take a different approach: request-reply technique. We will first describe the first approach, and term it as *method-1*, and the second approach is termed as *method-2*.

Method-1

In this method, the sender sends out a probe packet (a small packet with a single field) with increased transmission power, such that the packet reaches the nodes which are beyond the transmission range and within the interference range. We follow the technique proposed in [95] to determine the interference range to be used and the corresponding transmission power.

Interfering nodes only beyond the transmission range of S node will respond to the probe packet, *if and only if it is an active sender*. The node, which had sent the probe packet, when it receives the response, will store the node as far-neighbor. Therefore, a node also has to maintain a table of far-neighbors. This table, however, will lose its contents whenever a new probe packet is sent. The "available bandwidth estimation algorithm" will decide when to initiate a probe packet. This probing mechanism will result in sender S obtaining the information of the number of senders beyond transmission range and within the interference range (OSC). Therefore, the "fair" allocation for S is computed as $\frac{MAX}{OSC+1}$.

Method-2

This method is based on request-response technique. In this technique, we also rely on the power-control scheme. We describe this technique by considering the sender (request) side, and the receiver (request) side (any interfering sender outside the transmission range). We also consider few examples to describe the operation of this method.

At the Sender (request) Side

First the node should determine if it can send a request message. This is decided based on two conditions: (1) the available bandwidth is less than ‘m’ times the minimum required bandwidth (2) majority of the bandwidth is consumed by nodes outside the transmission range. These two conditions can be checked (especially the second one) by using the BSI and BSB values. Once the node identifies the case, it will create a request message (short message) and broadcast it. This message is sent only once and with increased power, as it has to reach the nodes outside the transmission range. After sending, it has to wait for the reply message. If the reply arrives, during the time of the decision the call will be admitted even if the bandwidth is less than the required bandwidth, because at least one of the interfering nodes has promised to reduce the rate. The call will be blocked until the reply is received.

At the Receiver (request) Side

Whenever a node receives a request message, first it has to decide if it can send the reply message. This decision is again based on the following two conditions: (1) the sender of the request message should not be the node’s neighbor (2) the node itself is consuming ‘n’ times more than the minimum required bandwidth. If these conditions are satisfied, the node sends the reply message, however, with certain amount of delay. The amount of delay is inversely proportional to the amount of bandwidth it is consuming. This approach will ensure that a node consuming highest amount of bandwidth among the interfering nodes

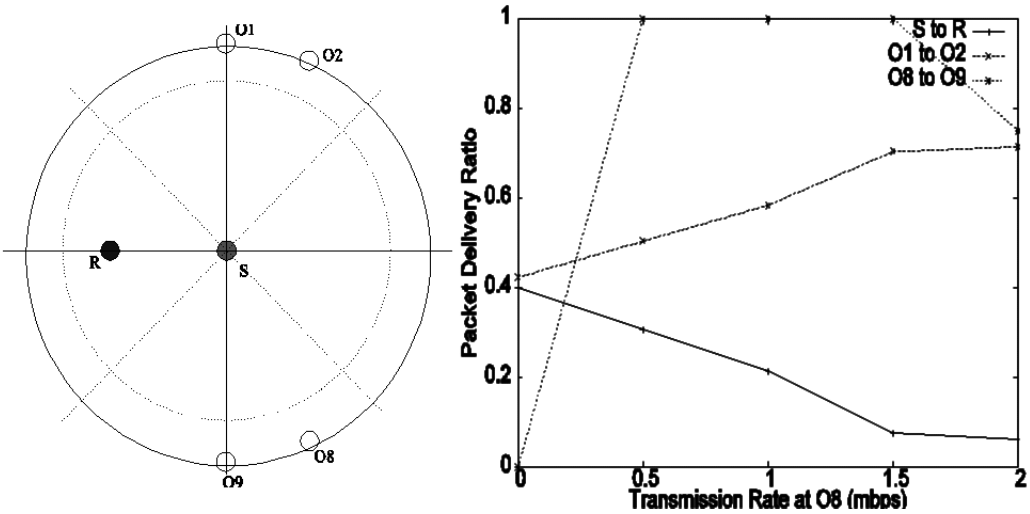


Figure 3.15: Topology and packet delivery fraction with varying rate (two interfering flows)

will send the reply first. This reply will also be received by other interfering nodes, who will abort the process of sending the reply. Therefore, only one interfering node within a “region” will send the reply message. It is also possible that two interfering nodes (lying in two different regions) can send reply to the sender, and reduce the bandwidth. This, however, is a desirable property than a disadvantage, which is shown in the succeeding paragraphs.

We will provide few example scenarios to highlight some advantages and disadvantages of this scheme (method-2). Referring to the topology as shown in Figure 3.14, let us consider an interfering flow from node O_1 and O_2, transmitting at a rate of 1.5 Mbps. Now, if sender S wants to initiate, according to method-1, it will assign the available bandwidth as 1 Mbps ($frac2Mbps2$). Whereas, in method-2, S initiates a request to node O_1. Assuming that minimum rate as 200 kbps, O_2 replies and reduces bandwidth by 200 kbps. We found that with method-1, receiver R receives 5162 packets, whereas with method-2 it receives 5361. That is, in method-2 the initial losses due to convergence-delay does not exist, because of the request-reply exchange.

Another interesting scenario would be to have two interfering flows outside each other's interference range, and interfering a single flow. For example, considering the topology shown in Figure 3.15, if O_1 (flow 1) is transmitting and S is also transmitting (flow 2), where S is getting less share as it cannot control the O_1's flow. Now when O_8 (flow 3), which is also outside the transmission range of flow S and outside the sensing range of flow O_1, is added, flow S will have to compete with both flow O_1 and flow O_8, which are outside its transmission range. In this case, when flow O_8 wants to join, it should make sure that flow S's share is not reduced drastically.

Let us assume that first O_1 starts transmission to O_2 at a rate of 1.5 Mbps. Later S starts transmission to R. From simulation, depending on the actual distance between the interfering pair and S, the range of bandwidth achievable is between 100 - 400 kbps. Now if O_8 wants to transmit to O_9, what rate should it use? For O_8, the BSB is around 1.5 - 1.8 Mbps, as it will not be affected by transmission of O_1 to O_2. Figure 3.15 shows the packet delivery fraction achieved by S, by varying the transmission rate of O_8 sending to O_9. From the Figure 3.15, transmission rate of 0.5 Mbps or 1.5 Mbps is good for O_8, though a higher rate causes the flow from S to R to degrade.

Using method-1, we found that the receiver R did not receive any packet, whereas both interfering flows did not experience any loss. Using method-2, flow from S to R, will receive some throughput (500 kbps, as both O_1 and O_8 reduces their rate), but this will result in losses for the two interfering flows. The combined losses in method-2 is higher than method-1, whereas, in method-2 the flow from S to R was not starved. If we can tune the parameters dynamically, then method-2 can achieve better performance. This parameter set tuning, however, is a difficult task.

Advantages of this scheme (method-2) are as follows:

- Minimize the initial losses which occurred in the previous technique (method-1).
- Overhead of maintaining the number of interfering nodes outside the transmission range are reduced.
- Suitable for low mobility scenarios.
- Provides better fair share than the method-1 for some cases.

Some of the disadvantages of this scheme (method-2) are:

- For high to average mobility scenarios the settling period is high, and can result in under utilization.
- Parameters ('m' and 'n') which decide when to reduce and how much to reduce are not easy to decide.
- For some cases, the number of errors may be higher compared to method-1.
- Fairness problem can still exist (when there are many interfering pairs)

In summary, we can see that both the methods have both advantages and disadvantages. *Method-2, however, is more applicable to static scenarios, whereas method-1 is more applicable to dynamic scenarios.*

In the remaining part of this section on Case 2B, we highlight two notable cases. Unfortunately, we cannot uniquely identify these two cases, and default sharing will have to be used.

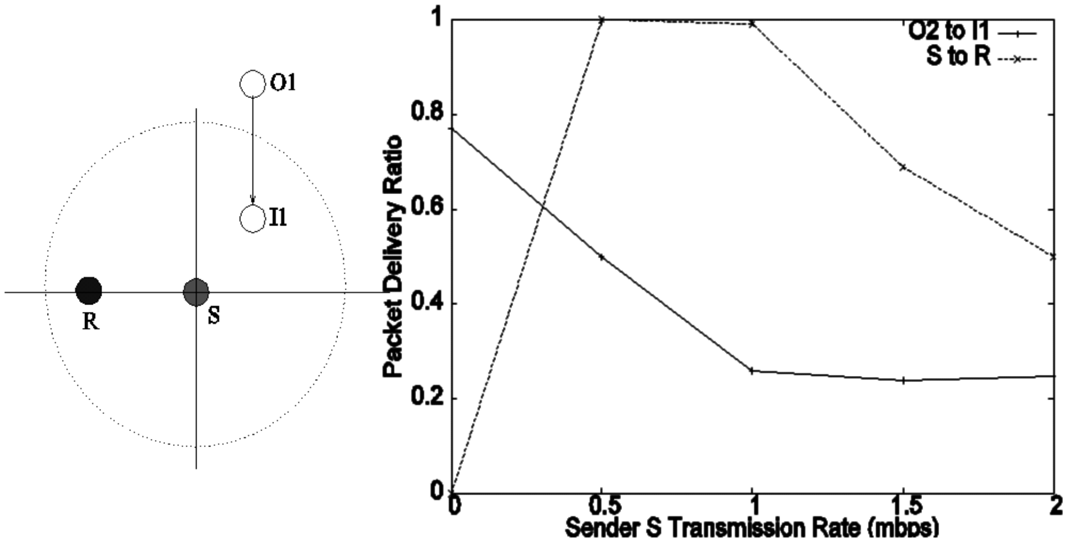


Figure 3.16: Topology and packet delivery fraction with varying rate (Receiver of I_f within the transmission range)

Case 2B-1

In this section, we will describe a interesting scenario where all the interfering senders are beyond the transmission range of sender S and their corresponding receivers are within the transmission range of sender S. This is also a case where RTS/CTS scheme is effective, and it is not necessary to carry out any probing or request-response procedures. We found that this scenario, however, is impossible to identify as unique case, as this scenario will not result in any unique parameter values (BSI, BSB or noise values). Therefore, we consider this scenario as part of Case 2B.

Let us illustrate this scenario (where receivers of interfering nodes being within the transmission range of sender S) in detail. Here, S has control over the interfering pairs through the receivers (control refers to obtain a share of bandwidth). For example, consider the topology and the packet delivery fraction plot shown in Figure 3.16. Let O_1, which is outside the transmission range of sender S transmit to I_1, which is inside the transmission range. Now when S (flow 2) starts, "ideally" it should incrementally increase and obtain its share, and also at the same time it should see that O_1 also gets an equal share of the bandwidth. That is, O_1's share is not reduced below its own share of bandwidth. As mentioned before, it is not easy to uniquely identify this case.

3.5.5 Case 3: Nodes of I_f Beyond and Within the Transmission Range of S

In this case, the interfering flows are both beyond and within the transmission range of S, making the coordination, if not impossible, difficult. This case is identified when $BSI, BSB < MAX$ and $NS, NR > 0$. Even in this case, sender S carries out the probing technique, similar to Case 2B (method-1), to obtain the number of interfering senders beyond the transmission range of S or OSC. Further, we also have number of senders within the transmission range (SC). We achieve better sharing by setting the available bandwidth as $\frac{maximum-bandwidth}{SC+OSC+1}$. This setting ensures that no one flow takes the complete bandwidth share making other flows to suffer.

In summary, in cases where the number of interfering senders can be detected directly, for example, in Cases 1 and 2A, a simpler and efficient method is used. Whereas, for cases where the RTS/CTS mechanisms are not effective we proposed two techniques (method-1 and method-2 of Case 2B). We would also like to highlight that while the estimations are rather coarse, we believe that we have included sufficient important cases such that the improvement will be substantial compared to a scheme that uses only BSB.

3.6 Available Bandwidth Measurement Algorithm

In this section, we combine all the cases described before and present iCAC - an interference-based call admission control scheme.

The admission control mechanism has four components: local bandwidth measurement, end-to-end bandwidth measurement, admission and rate-control, and bandwidth re-computation. Local bandwidth estimation is carried out by all nodes along the route by carrying out the algorithm explained in the preceding section. For end-to-end bandwidth estimation, the routing mechanism performs the task.

Local available bandwidth measurement include measuring the busy periods (considering both sensing as busy and idle), and noise values. Further, both the probing (method-1 of Case 2B) and request-reply message (method-2 of Case 2B) are part of this measurement.

We modify just the Route-Reply (RREP) packet of DSR, with following fields- bottleneck bandwidth (BB) value, and noise-value (NV). The destination node, when it initiates the RREP message, adds its local bandwidth measurement into BB field and its noise-value into NV field, and sends to the next hop. The next hop when it receives this reply message, uses the NV value of the reply-packet to compute its local available bandwidth. It next checks if its local bandwidth value is lesser than the BB field of the packet, and if it is, it replaces the BB field value with its bandwidth value. In addition, it replaces the NV value of the packet with its noise value. This process continues till the packet reaches the source node.

We also enhance the routing table information with the bottleneck bandwidth value, which is associated with each and every route it stores. Further, we include a neighbor management mechanism in DSR, which includes exchange of hello messages between nodes for every fixed duration (5 secs). This message includes the status of the neighbor (whether the node is a sender or not), along with its ID.

The admission control mechanism decides if the required bandwidth is less than or equal to the bottleneck (minimum available) bandwidth of the route. If it is, then the call will be admitted, and if not, the call will be either blocked or the rate of the transmission is reduced. Rate control reduces or increases the traffic rate according to the feedback provided by admission control scheme. If the bottleneck bandwidth falls below some minimum value the call will be completely blocked.

Bandwidth re-computation is performed after a call is admitted, and is triggered on two conditions. First, when the number of senders among the neighboring nodes change (increases or decreases), second, when the noise values change by certain fixed amount (increases or decreases). Note that in the current framework, a flow will not be given more than the minimum of its end-to-end fair share, which can under-utilize the network. We believe that fully utilizing all the bandwidth is too aggressive and ensuring that all bandwidths are assigned similar to the max-min assignment described in [109] requires multiple iteration and is too expensive in terms of messages required and admission control duration.

In the flowchart of our available bandwidth measurement algorithm as shown in Figure 3.17, a sender node S on which this algorithm is run and a receiver node R is considered. In the flowchart, AB represents the available bandwidth, NS represents the noise value at the node Sender, whereas NR represents the noise value at the receiver. BSB and BSI are the same terms as explained in the previous sections. SC and OSC represents the number of senders within the transmission range and beyond the transmission range (within the interference range) of node S, respectively. m represents the number of senders within the transmission range of receiver R. Finally, MAX is the maximum available bandwidth. The detailed description of flow chart is excluded as majority of description is covered in preceding sections.

iCAC starts by checking if $BSI = BSB$ and $NS = 0$, and this is the case (Case 1) when all the interfering pairs are within the transmission range of the sender. The AB is set to $\frac{MAX}{SC+1}$. If this case is not satisfied, it proceeds to check if $BSI = MAX$ and $BSB < MAX$. This is the case (Case 2) where all the interfering nodes are outside the transmission range of the sender S. Further, it checks if $NR = 0$ (Case 2A), and if true the AB is set to $\frac{MAX}{m+1}$. If false (Case 2B), we can take any one of the two approaches. In method-1 AB is set to $\frac{MAX}{OSC+1}$, whereas, in method-2 AB is set to BSB. If both Case 1 and Case 2 were not satisfied, then algorithm sets the AB value as $\frac{MAX}{SC+OSC+1}$ (Case 3).

3.7 Evaluation of Admission Control Mechanism

For simulations, we modified dynamic source routing (DSR) [34] protocol to carry out end-to-end bandwidth estimation. Evaluations are performed for both single-hop and multi-hop scenarios. We have considered method-2 of Case 2B for only the fairness evaluation.

We compare iCAC with Perceptive Admission Control (PAC) [97] and the legacy IEEE 802.11 [6] mechanism without any admission control process. PAC scheme depends on self estimation of the available bandwidth, which is performed by changing the range of the

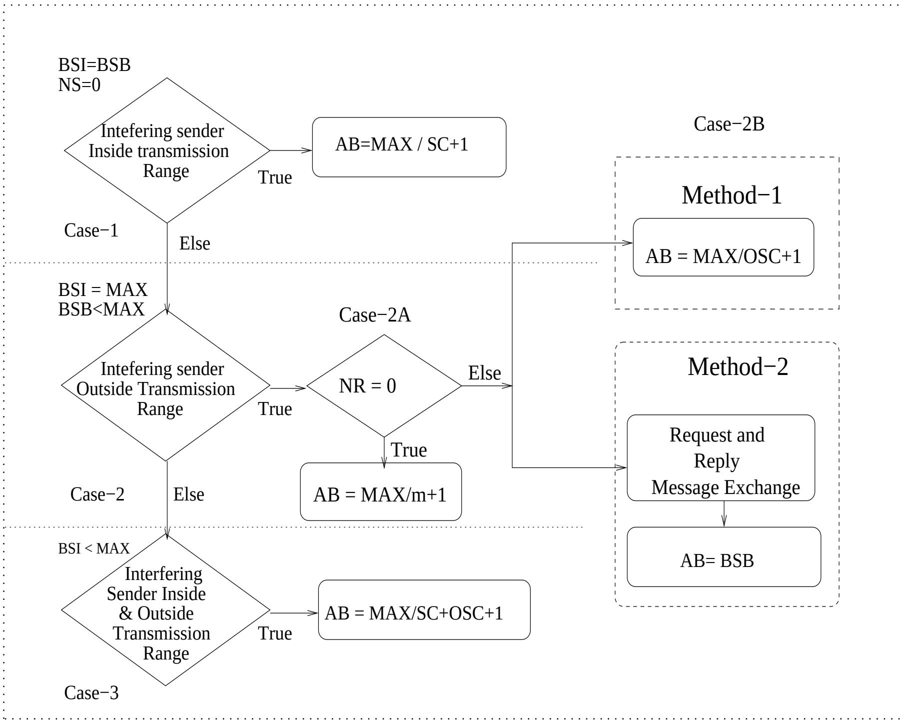


Figure 3.17: Flowchart of available bandwidth measurement algorithm

bandwidth measurement. This technique relies on enhancing the range and letting each node measure the bandwidth without the need for communicating with the other nodes. In PAC, the *sensing range* of the node is enhanced to the distance of $2*RxR+RID$, where RxR is the transmission/reception range and RID is the receiver interference distance. RID is the distance between a receiver and any other sender, such that the corresponding receiver's ability to decode a packet from its sender is not affected. The authors of PAC believe that at any distance greater than $2*RxR+RID$, two ongoing transmissions will not interfere with the packet receptions, and therefore a node can make decisions (on admitting new flows) based on its available bandwidth (by considering this large range).

PAC is designed for single-hop networks, therefore in this section we consider topologies with single-hop ad hoc network. The evaluation of our scheme for multihops is provided in Section 3.7.

All our evaluations are carried out on GloMoSim [12] simulator. All the parameters for PAC (RxR = 250 m, RID = 440 m, Sensing range = 940 m) are the same as used in [97]. Each mobile host has a transmission range of 250 m and shares a 2 Mbps radio channel with its neighbors. The simulation includes a two-ray ground reflection model and IEEE 802.11 MAC protocol. All the simulations are run for 200 seconds. A single hop network with 50 mobile nodes (25 pairs) is simulated. The network area is 2000 m x 2000 m. For all the pairs, the nodes are within the transmission range of each other, so that we can focus only on the effects of admission control scheme. Nodes move together, and this movement happens only 2 to 3 times. The link between the two nodes of a pair is always intact.

3.7.1 Single Hop Evaluation

We consider a traffic load with 25 flows. The source and destination nodes of each flow are within transmission range. We consider real-time UDP traffic with a packet size of 1460 bytes, with varying transmission rate. We consider three transmission rates 100 kbps, 200 kbps, 500 kbps. The flow arrivals are 5 seconds apart. Therefore, after 125 seconds of simulation time all the sender-receivers pairs are active. This evaluation part is similar to the one used in [97].

The results are summarized in the Tables 3.1, 3.2, 3.3 and 3.4. The tables show the average end-to-end delay, number of calls admitted, number of packets delivered and packet losses respectively for the transmission rates of 100, 200 and 500 kbps, for the three schemes- PAC, 802.11 and iCAC.

A good admission control scheme is one which admits as many requests as possible without compromising on the performance of existing requests. A conservative scheme maintains good performance by admitting far too few requests, and an optimistic scheme allows all requests to be admitted without any regard to the performance. We admit as many requests as possible according to our bandwidth sharing model described in Section 3.4.

For low loads, all three schemes have similar performance, though iCAC has the lowest delay and loss compared to both PAC and IEEE 802.11. At a medium load, the performance gaps start to appear. iCAC admits slightly more requests than PAC, has low average delay and attains higher throughput. iCAC, however, does have a small amount of losses. Compared to iCAC, the IEEE 802.11 scheme admits the same number of requests but has a much higher delay and packet loss.

For high loads, iCAC admits almost twice as many requests as PAC (23 versus 11) and slightly less than IEEE 802.11 (23 versus 25). In addition, in spite of the fact that the overall traffic load is much higher, by using a better local bandwidth estimation and rate control, iCAC can provide fairly low end-to-end delay, high throughput and low packet losses.

We carried out detailed simulations on the percentage ratio of number of times the probing technique is involved over total number of time the available bandwidth is computed. We found that as the flow-density increases, this percentage ratio increases. With 40 to 50 flows, the ratio is about 70%.

Table 3.1: Average end-to-end delay

	100 kbps	200 kbps	500 kbps
iCAC	7.8 ms	8.5 ms	0.206 secs
PAC	8.0 ms	11.5 ms	0.105 secs
IEEE 802.11	9 ms	15 ms	2.98 secs

Table 3.2: Average number of calls admitted

	100 kbps	200 kbps	500 kbps
iCAC	25	25	23
PAC	25	23	12
IEEE 802.11	25	25	25

Table 3.3: Average number of packets delivered

	100 kbps	200 kbps	500 kbps
iCAC	28905	56799	92476
PAC	28905	45633	45862
IEEE 802.11	28901	57696	108335

3.7.2 Fairness Evaluation (Single-Hop)

In this section, we evaluate how fair our call admission control algorithm is. For our evaluation, we also consider the fair allocation algorithm presented in [109]. This allocation algorithm computes the fair share allocation for every flow by considering the flow contention graph and building cliques out of the flow contention graph. The major drawback of this scheme is that, it considers only the transmission range of nodes in developing the flow contention graph and cliques, and does not consider the interferences and noises due to flows outside the transmission range, which would affect the transmission. Therefore, its estimation is highly optimistic and sometimes far from reality. We, however, included it to

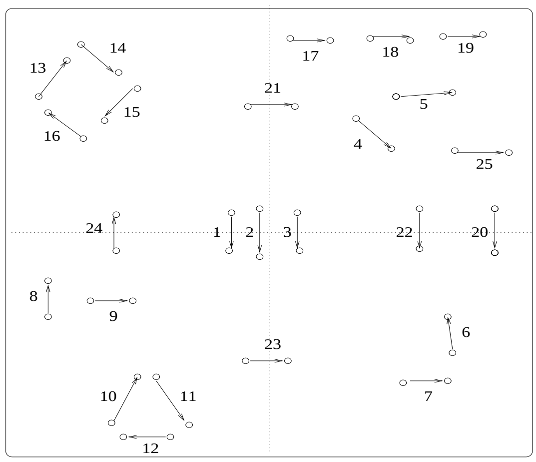
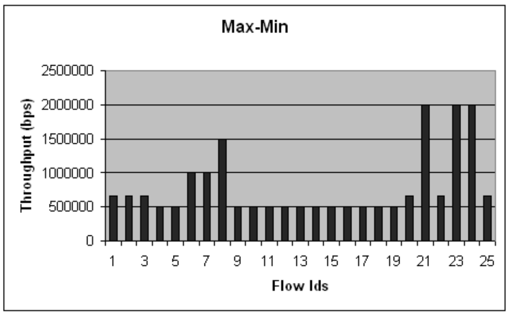
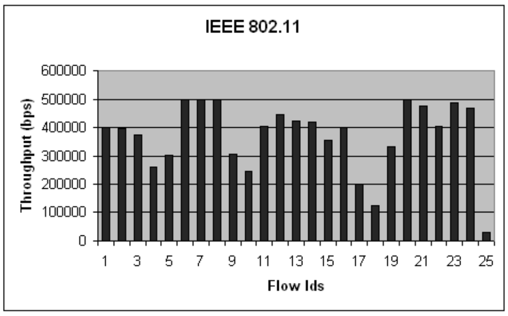


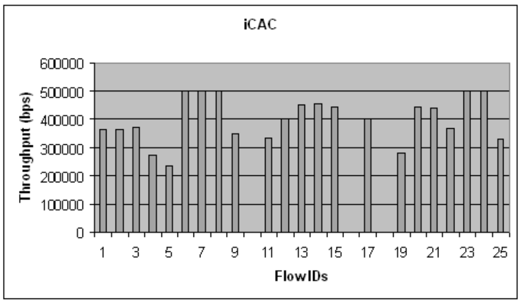
Figure 3.18: Simulated topology for fairness



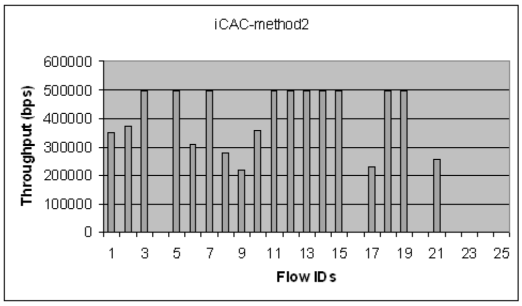
(a) Max-Min fairness



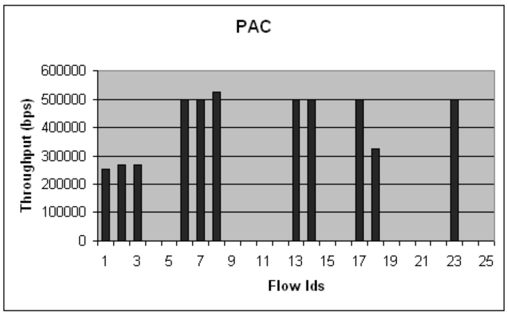
(b) IEEE 802.11



(c) iCAC



(d) iCAC-Method2



(e) PAC

Figure 3.19: Comparison of flow shares by various approaches

Table 3.4: Average number of packet losses

	100 kbps	200 kbps	500 kbps
iCAC	0	12	1021
PAC	0	0	346
IEEE 802.11	7	240	43290

see how our algorithm performs relative to this fair allocation.

We consider the similar simulation settings as previous sections. The simulation area is 2000 m x 2000 m with 50 nodes (25 pairs) randomly placed. Nodes are static and flow is single hop. Flow contention graph is developed for this topology and using algorithm 1 of [109], we compute the fair share for each flow. In the topology, 14 cliques with different degrees were formed. The allocation using this algorithm is shown in Figure 3.19(a). For the other three algorithms, the traffic load per flow is 500 kbps and the allocations for IEEE 802.11 without admission control, iCAC, PAC are shown in Figure 3.19(b), 3.19(c) and 3.19(e), respectively. Allocation of iCAC with method-2 for Case 2B is shown in Figure 3.19(d). Table 3.5 summarizes the performance results (delay, packet loss, packets received, calls admitted) for IEEE 802.11, PAC, iCAC and iCAC-method2, for the scenario considered.

From Figure 3.19(a), we see that the ideal max-min allocation, which does not take into account of interference would accept all calls and at the same time provide at least 500 kbps to all flows. Once interference is taken into account, however, the bandwidth is much lower as indicated by Figure 3.19(b), which shows the performance of IEEE 802.11 without admission control. Using IEEE 802.11, flow 25 gets little bandwidth, average delay is more than 2 seconds and packet loss rate is more than 30%. With PAC, only 11 out of 25 requests are accepted. Out of these 11 requests, 3 requests have rates below 300 kbps. Thus, the control is both too conservative and unfair. Finally, iCAC admits 22 out of 25 calls, and all requests admitted have more than 200 kbps. The total throughput achieved is almost double that of PAC and within 83% of IEEE 802.11. Loss rate and delay are slightly higher than PAC but this is unavoidable since the throughput is much higher. Considering the topology in Figure 3.18, and Figure 3.19(c) we can see that whenever the flows are within the transmission range {for example, flows 1,2, 3 and flows 13,14, 15, 16 }, nodes tend to share the available bandwidth fairly. From Figure 3.19(d) we can see one of the disadvantages mentioned earlier with method-2 for Case 2B. Flows 4, and 22-25 are blocked, because none of the interfering flows agree to reduce the rate because of the parameter set (m=1, n=3).

Table 3.5: Fairness evaluation of iCAC

	Avg Delay (sec)	Total Pkt loss	Total Pkts Rcvd	No/of calls Admitted
iCAC	0.231	176	97833	22
iCAC-method2	0.1821	106	71439	18
PAC	0.136	20	47371	11
802.11	2.72	36115	118215	25

3.7.3 Multi-Hop Evaluation with Random Mobility

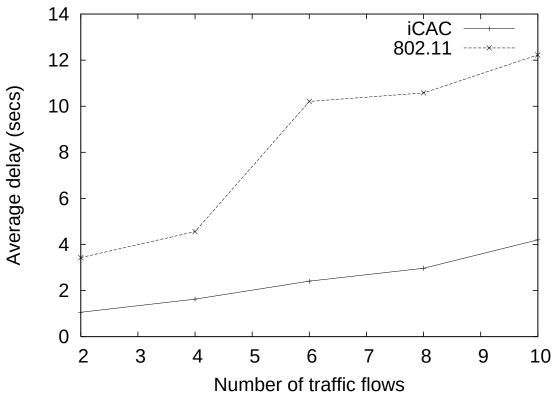
In this section, we present the evaluation of our admission control scheme in multihop scenarios with random mobility. The node capabilities are similar to the previous simulations. The simulation area is 1000 m x 1000 m, with 25 nodes. Random waypoint mobility model is used with a maximum speed of 5 m/s and with pause time of 50 secs, which is a relatively slow moving scenario. We compare our admission control scheme with IEEE 802.11 without admission control scheme. We vary the number of traffic flows in the network from 2 to 10 flows. The source and destination are chosen randomly. The transmission rate is 500 kbps for all the sources.

Figures 3.20(a), 3.20(b), 3.20(c) show the average delay, number of packets delivered and the number of packet loss with varying number of traffic flows. From the figures we can see that iCAC performs much better than the mechanism without admission control with respect to average delay and packet losses. The number of packets delivered, however, are slightly lesser compared to IEEE 802.11 without CAC. The decrease in packets delivered is mainly due to flows that are blocked. This helps in reduction of delay and packet losses, which are crucial for real-time applications. Sources pumping the traffic at a very high rate and link failures are the reasons behind the larger delay values.

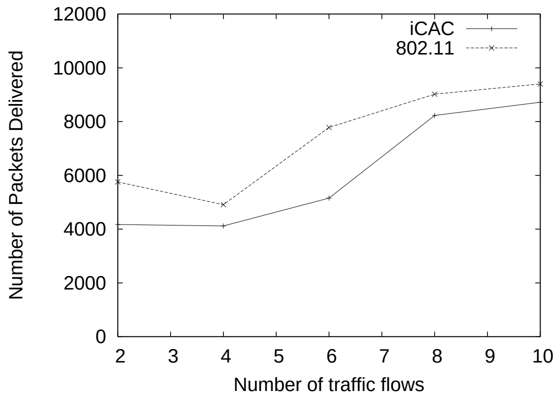
Similar to the single-hop evaluation, we also carried out a detailed simulation study on the percentage ratio of number of times the probing technique is involved with respect to number of times the available bandwidth is computed. Approximately, 60% to 70% of the time, the algorithm computed a different fair share.

3.8 Summary

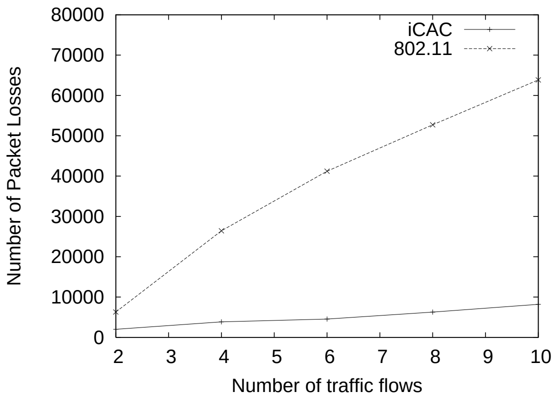
In this chapter, we described a call admission control mechanism for wireless ad hoc networks called interference-based call admission control (iCAC). iCAC is unique because it does not treat interference uniformly instead classifies interference based on estimates of the position of the interfering nodes. iCAC relies on two novel techniques: (1) estimation of position of the interfering nodes (2) fair allocation using bandwidth acquisition and rate control. By incorporating these techniques, iCAC can increase the estimated available bandwidth substantially without overloading the network. We compared iCAC with Perceptive Admission Control (PAC) [97] and IEEE 802.11 without admission control. Simulation results show



(a) Average delay



(b) Number of packets delivered



(c) Number of packet losses

Figure 3.20: Performance of iCAC and IEEE 802.11 in multihop scenarios

that iCAC admits substantially more requests than PAC, achieves more than 80% of the throughput of IEEE 802.11, and maintains low packet loss rate and average delay comparable to PAC. Considering the practical implications, existing wireless LAN Driver implementation [110] maintains the states of wireless radio, which can be probed by upper layers. The frequent probing of states and noise values introduces a negligible overhead of computation.

Chapter 4

Scheduling

In the previous chapters we discussed routing and admission control mechanisms. In this chapter, we concentrate on the packet scheduling mechanism, which determines which queued packet is to be processed next. Packet scheduling has a major impact on the performance of mobile ad hoc networks, and is also an important service differentiation component. In this chapter, we describe a novel scheduling mechanism, which is designed specifically for MANETs. We incorporate both end-to-end channel condition and congestion information into the packet scheduling decision.

4.1 Introduction

Wireless medium is a shared and scarce resource, which is used by all nodes in the network. Efficiently controlling the access to this scarce resource is a complicated task. Resource management schemes play a major role in achieving this task. Packet scheduling is one such resource management scheme, which controls the allocation of bandwidth among multiple flows.

In wireless ad hoc networks, the significance of packet scheduling rises from following challenges:

- Need to support a variety of applications, with wide range of requirements.
- Apart from congestion, dynamic changes in topologies due to node mobility result in sudden changes in network connectivity.
- Decentralized access to a shared and scarce wireless medium.
- Existence of multiple hops.

For Internet, scheduling as a resource management scheme mainly focuses on supporting a wide variety of applications. Whereas, for MANETs, it is important to consider all the challenges mentioned above. These challenges will decide the approach one should take in designing the packet scheduling mechanism for MANETs.

An important functionality in scheduling mechanism is the *scheduling decision*, which decides which packet to send from a set of queues. This decision is driven by the objective of the scheduler. Typical scheduling objectives include reducing the packet delay, increasing the throughput, enhancing spatial reuse and achieving fairness. In our work, the main objective is to enhance the performance (reduced delay and increased throughput) by overcoming the challenges involved in MANETs.

As mentioned in earlier chapters, our main focus is to consider the features unique to MANETs in designing the mechanisms. Adhering to this line of study, in our scheduling mechanism we consider the *changing topology*, *multihops* and *shared wireless medium* features of MANETs. Our scheduling scheme considers these features by using the "channel-aware" approach. The term "channel-aware" in our work refers to having the knowledge of both end-to-end and local channel *conditions*. The term *condition* refers to the quality of the channel which can be measured in terms of suitable metrics. Terms "channel state" and "channel condition" are used interchangeably.

The concept of channel-aware scheduling has been considered by many researchers in the context of WLANS [111–114]. Base stations maintain a set of queues, where each queue corresponds to one destination. The basic idea is to choose packets from queue corresponding to a destination such that wireless channel associated to that destination is "good". There are different proposals, which have different definitions for "good". For example, "good" can be defined as signal to noise ratio measurements on the channel being above a threshold value.

By considering channel-awareness, the focus is to solve problems associated with multiple sessions, within a single node, sharing the wireless link. One such problem is the repeated retransmission attempts of the head of line (HOL) packet due to channel failures (caused by mobility, fading, interference from other users and shadowing from objects), blocking the transmission of packets to other receivers. Since the wireless links to various destinations are statistically independent, packets for other destinations could be successfully transmitted during this interval. This is in fact the problem, which all the proposed channel-aware scheduling mechanisms for WLANs address. For MANETs, however, the problem due to the presence of multihops makes the study challenging and interesting.

Existing work on packet scheduling over MANETs focuses on providing performance guarantees (e.g. throughput and delay) and fairness. There has not been any work on packet

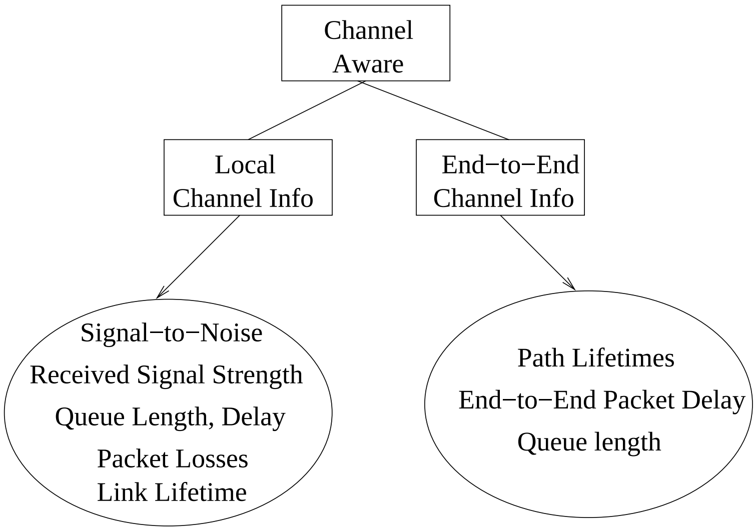


Figure 4.1: Channel-state awareness

scheduling that considers the impact of mobilities and path breakages. In the MANET environments, it is important to consider the inherent characteristic of MANETs such as path breakages, multihops, shared medium, in the scheduling mechanism. In this chapter, we present, *CaSMA*, a scheduling mechanism for mobile ad hoc networks (MANETs) that takes into account both the congestion state and end-to-end *path duration*. Our scheduling mechanism is termed Channel aware Scheduling for Mobile Ad hoc networks (CaSMA), where the term channel-aware is used to indicate both the *congestion state* and the *end-to-end path duration*. CaSMA is complimentary to packet scheduling scheme that utilizes only local channel information and can be added to these schemes.

During the path setup, the estimates of the path lifetimes are collected and stored. This path lifetime value is used as a parameter to represent the end-to-end channel condition. During packet scheduling, scheduler selects packets, which has high probability of reaching the destination, and takes into account the cost of a link break by giving priority to flows that have a longer normalized (with path residual lifetime) backlog queue. We show that CaSMA approximates an ideal scheduling mechanism in terms of maximizing the goodput and sharing the throughput(losses) fairly among the contending flows. Further, the simulation results show that both average delay for CBR flows and throughput for TCP can be improved substantially compared to FIFO.

CaSMA can be deployed as a link layer solution on all mobile nodes in an ad hoc network. As CaSMA relies heavily on other protocols like routing and neighbor management to be "channel-aware", it is important to have these protocols in place before deploying CaSMA. CaSMA scheduling method can also be used in conjunction with the transport layer techniques for improving the performance over the wireless channels.

Before describing CaSMA in detail, we first explain the concept of local and end-to-end channel awareness, so that the understanding of CaSMA becomes easier.

4.1.1 Local Versus End-to-End Channel Conditions

Channel conditions in wireless networks can be broadly classified as local and end-to-end channel conditions, as shown in Figure 4.1. For mobile ad-hoc networks, unlike wireless LANs, local and end-to-end channels are different. The difference between the local and end-to-end channel information can be better understood by considering their typical characteristics. We can consider 4 key categories as shown in Table 4.1: frequency, granularity, accuracy and measured-time with respect to packet delivery.

Frequency category represents the monitoring frequency of the channel state. Monitoring at a higher frequency induces larger overhead or at times more consumption of bandwidth. As a result, there should be a trade-off in the frequency of monitoring. Local channel state is monitored with higher frequency, whereas end-to-end channel state enforces low-frequency monitoring due to the above mentioned tradeoff. *Granularity* refers to the representation of channel state. One method is to use 2-values (good/bad or 1/0), whereas other method is to express as a direct value of signal strength or SNR. Local channel state is represented as 2-state in majority of earlier works on WLANs. There are numerous works, however, representing local channel state with multiple values. Whereas, for end-to-end channel state it may not be efficient to consider 2-state representation.

Measured time category represents the duration in time at which the monitoring of channel state is carried out. Typically local channel state is monitored just before or at the time of the packet delivery. Whereas, end-to-end channel state is measured much before the packet delivery time. *Accuracy* defines correctness of the measurements that represent the channel state. For local channel state, the accuracy is high compared to end-to-end channel state. As local channel state is monitored just when it is used, the information will not become stale. Further lower-layers (MAC/Physical) and upper layers (Network) can give exact representation (measurements) of the local channel state.

As shown in Figure 4.1, typical parameters that are used to represent the local channel information are received signal strength, signal-to-noise values, queue-length, burst-error mode, packet losses, single hop delay and link lifetime. Whereas, parameters that could possibly represent the end-to-end channel conditions are: path lifetime, end-to-end packet delay and queue-length at every node.

In our work, we focus on the end-to-end channel awareness and represent the end-to-end channel quality in terms of path lifetimes.

Category	Local	End-to-End
Accuracy	High	Low
Granularity	2-Values	Multiple Values
Time with respect to delivery	Closer	Farther
Frequency	High	Low

Table 4.1: Local versus end-to-end channel awareness

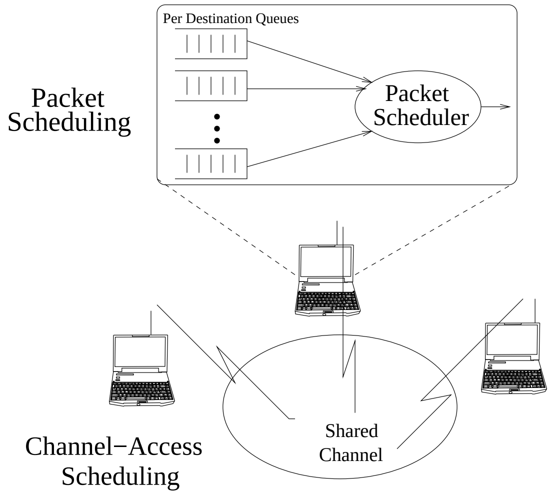


Figure 4.2: Packet and channel access scheduling

The remaining part of the chapter is organized as follows. In Section 4.2, we discuss the related works, covering scheduling mechanisms in MANETs, and describing the contributions of our work. CaSMA is described in Section 4.3. We begin with describing the motivation for using channel awareness in general, and path lifetimes in particular. Further, we describe the approach taken in CaSMA, the framework, algorithm used for packet selection and the limitations of CaSMA. We conclude the Section 4.3 with the experimental evaluation of CaSMA. The chapter ends with summary and concluding remarks in Section 4.4.

4.2 Related Works

The term "scheduling" in multihop wireless networks usually refers to two problems, which is depicted in Figure 4.2:

1. Packet scheduling: which flow should be served among the set of backlogged flows within a node? Typical goals of packet scheduling include to minimize the packet drops in queue, and provide a fair share of bandwidth among the backlogged flows. To achieve this, it is necessary to have a proper understanding of the queueing dynamics under different degrees of mobility, traffic loads. Packet scheduling also involves prioritizing different kinds of packets.

2. Channel access scheduling: which node should get access among the set of competing nodes in a "contention region"? In channel access scheduling, fair distribution of bandwidth among the contending nodes and maximization of resource utilization are identified as two important goals [115]. Achieving both fairness and maximization of channel utilization, however, is particularly challenging in wireless ad-hoc networks. It is also desirable for a channel access scheduling to be developed as a fully distributed and scalable mechanism.

4.2.1 Packet Scheduling

A detailed study on packet scheduling algorithms for MANETs was carried out by Baker et al. [116]. The algorithms they study are:

- No priority scheduling- no differentiation is made between data and control packets, and First In First Out (FIFO) is used.
- Priority scheduling- control packets are given higher priority than data packets. Within this priority scheduling scheme various schemes are studied.
  - Shortest-Path-Length-first scheduling (SPL)

- Fewest-Remaining-Fops-first scheduling (FRH)
- Round Robin scheduling (RR)
- Greedy Scheduling (GS)

In their work, they first show the importance of providing priority to control packets over data packets. They found that there is little advantage in using priority scheduling over non-priority scheduling, with respect to average goodput of best-effort traffic. Scheduling mechanisms SPL and FRH perform better than any of the mechanisms with respect to average end-to-end delay. It was also found that the average goodput for best-effort traffic was similar for all the scheduling mechanisms.

Majority of the previous works consider packet scheduling along with channel-access scheduling. Therefore, in the remaining part of this section we will describe various channel-access scheduling mechanisms.

4.2.2 Channel Access Scheduling

Luo et al., [93, 117, 118] have proposed fair-scheduling mechanisms based on timestamp. All the mechanisms aim to support QoS guarantees and requires a flow graph (contending flow graph) to be generated first. In a flow graph, a vertex indicates a flow, and an edge must be added when two flows are contending for resource. For each newly arrived packet n of flow i at time $A(t_{i,n})$, two timestamps are assigned, namely, the start tag ($S_{i,n}$) and finish tag ($F_{i,n}$) as follows: $S_{i,n} = max\{V(A(t_{i,n})), F_{i,n-1}\}$; $F_{i,n} = S_{i,n} + L_p/r_i$, where $V(x)$ is the virtual arrival time, L_p is the length of the packet, and r_i is the weight of flow i. One of the timestamps can serve as the service tag. The packet with the least service tag will be transmitted first.

In fact [93] is an extended version of [118]. In [93] 2-tier (global-fairness and local-fairness) service model is proposed. The global fairness model assumes having a complete knowledge of the flows in the entire network, whereas local fairness assumes the knowledge of local contending flows. Both [93] and [118] achieve spatial reuse by assigning backoff values. These backoff values are proportional to the number of contending flows (in [93] also depends whether global fairness or local fairness model is used). In [117] the spatial reuse is achieved by using graph coloring theory. Flows marked with same colors are transmitted simultaneously. These timestamp based mechanisms suffer from disadvantages like complexity in building flow graphs, complete knowledge of topology, sorting of packets and assigning timestamps.

Chao et al., [119] propose a credit-based fair scheduling mechanism in which they assign credits to the flows. High priority is assigned to flows which use less bandwidth. They assume a TDMA based multiple-access system. They have cluster architecture, where a 2-tier hierarchy is used for assigning timeslots, which is termed as credit-based slot allocation protocol (CSAP). Each node maintains a Flow Allocation Table (FAT) for flow scheduling (a *scheduler* assigns next time slot to the node), and scheduler at nodes maintain an extra table called Node Allocation Table (NAT) for node scheduling (the node assigns the time slot to the flow).

Chao et al., [120] also compared their work with timestamp based mechanisms. They also propose *flow weight assignment* technique to timestamp based mechanism. They found that CSAP performs better than any other mechanism. The major drawback of credit-based mechanism is the architecture itself, where a *scheduler* is assumed for each cluster, which makes it hard to implement.

Wu et al., [121] have proposed a centralized fair scheduling scheme based on considering bottlenecks in ad hoc networks. Authors first predict the achievable throughput under the strict notion of fairness. Based on this, they identify bottleneck links, and give higher priority to flows belonging to a bottleneck "locality". In order to differentiate between the severities of the bottlenecks they propose a parameter termed as *contending power* of a flow based on local flow weights and topology information.

Kanodia et al., [115] have proposed two mechanisms-"distributed priority scheduling" and "multihop coordination" for scheduling in MANETs. In distributed priority scheduling, every node constructs a scheduling table based on the overheard information and incorporates the estimate of its relative priority into the IEEE 802.11 MAC. Initially, a priority index to the packet is assigned based on a locally computed parameter "deadline" (considering a delay bound). The node piggybacks the priority index of the head-of-the-line (HOL) packet into the RTS/CTS handshake. Further, the node assesses the priority of its own HOL packet in reation to its neighbors' HOL packets, and assigns the relative priority. They exploit this relative priority information to modify the backoff scheme of IEEE 802.11 to approximate a "global" dynamic priority schedule. The multihop coordination scheme relies on downstream nodes compensating for upstream nodes. The priority index of each packet at downstream depends on the index at the upstream nodes and the delay experienced. Nodes in the network co-operate to provide end-to-end service.

Bao et al., [122] proposed three channel-access scheduling (problem-1) for ad hoc networks. The idea behind these three proposals is to resolve contention by choosing (deterministically) one or multiple senders ("winners") for transmission, within a contention region. All three proposals depend on common scheme called "neighbor-aware contention resolution" (NCR). The three channel access protocols are: (1) NAMA (node-activation multiple access) - based on NCR, and distributed TDM scheme (2) LAMA (link activation multiple access) - based on NCR, DSSS, and time-slotted code division scheme (3) PAMA (pairwise-link activation protocol) - similar to LAMA, but code is assigned for a sender-receiver pair, and computed at every time slot.

Majority of the previous proposals try to solve both the problems at the same time. In this process, however, the focus is more on the second problem (channel-access scheduling) rather than the first (packet scheduling). The standard approach used in all the earlier proposals for packet scheduling can be summarized as follows.

1. Choose a parameter that is locally computed and reflects only local conditions. Call this as *LC* (locally computed), which is a parameter, used to reflect local channel condition. For example, *LC* used by Kanodia et al. [115] in their scheme is termed as "priority index" or the "deadline" for each packet (considering a delay bound).
2. Choose a flow with a set of backlogged flows within a node using LC (minimum or maximum) values.
3. Modify the MAC protocol to approximate the global ideal scheduler. That is, priority to global minimum/maximum of LC is approximated by giving priority to the flow with local minimum/maximum of LC within a contention region.

There are various disadvantages with these approaches. To begin with, the flow model considered by previous works does not consider the validity-period of the flow, which depends on the quality of the path it is taking. The *LC* chosen by previous works are typically used for Internet (tagging), with slight modifications, which may not be suited for ad hoc networks. For example, majority of the works do not show how the chosen *LC* helps in achieving the objective (either fairness or throughput/delay bounds). Further, the approximation of the "local minimum/maximum of LC" by modifying the MAC protocol to achieve "global (within a contention region) maximum/minimum of *LC*" is less accurate as *LC* does not reflect end to end behavior. Earlier works on channel quality aware scheduling [111–113, 123, 124] have considered only the local channel states. This is mainly because the channel state dependent research has focused more on wireless LANs (single hop networks), and less on multihop wireless networks. Finally, the impact of mobility is not investigated in earlier packet scheduling schemes. Mobility directly affects residual path lifetime, which is an end-to-end parameter.

Our work is different in the following ways. We focus only on the problem of packet scheduling. We do not propose any channel-access mechanisms. Any of the existing channel-access mechanisms can be used with our scheme. We propose a novel *LC* that reflects end-to-end conditions. According to our knowledge there is *no work considering end-to-end channel information in packet scheduling.*

4.3 Congestion and Path Lifetime Aware Packet Scheduling for Mobile Ad-hoc Network

In this section, we describe *CaSMA*, the scheduling mechanism for mobile ad hoc networks (MANETs) that takes into account local congestion information and end-to-end *path duration* information. We begin with describing the importance of considering channel awareness in general and path durations in particular. Further, we formally define the problem and describe the approach to solve the problem in detail.

4.3.1 Motivation for Using Channel Aware Scheduling

We begin our study of channel-aware scheduling scheme for MANETs by considering the existing stability-based routing protocol. The channel condition parameters we used were also the parameters which the routing protocol uses. We will first describe this simple approach, and then proceed to describing CaSMA.

In this study, we use ABR as the routing protocol. We consider the local channel information that is obtained by the neighbor state information maintained in the routing protocol. The end-to-end information is the stability value as defined by the ABR routing protocol. We call the scheduling mechanism as Stability and Neighbor State Dependent Scheduling (SNSDS). We define stability of a route 'r' as: $Stability(r) = Associativity(r)$

The term associativity is same as it is defined in ABR. Associativity of a route is the minimum of associativity of each pair of nodes along that route. Let Associativity(i,j) be the associativity of link (i,j). For a route 'r', with source 'i' and destination 'm' and with links (i,j), (j,k)... (l,m), where 'j', 'k' and 'l' are intermediate nodes, associativity of route r is defined as,

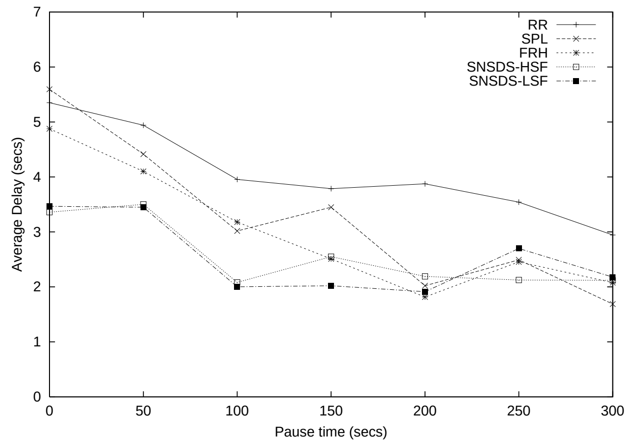
$$Associativity(r) = min[Associativity(i, j), Associativity(j, k).....Associativity(l, m)].$$

A set of data queues are maintained at each node, with a single data queue for every source-destination pair, and a separate queue for routing packets.

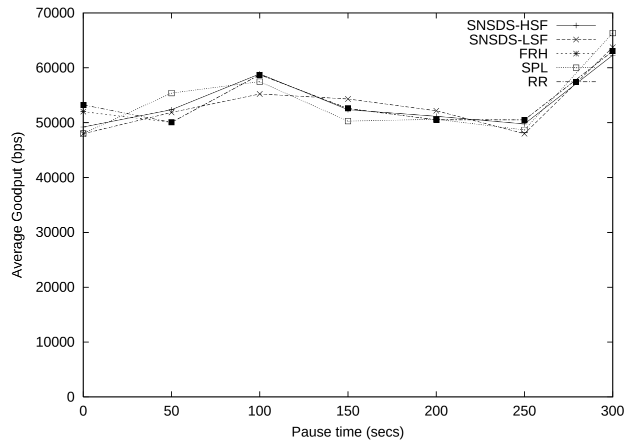
ABR provides both the stability factor and the neighbor status to the scheduling mechanism. Stability factor for a route is calculated at destination node, at the time of route discovery mechanism. This stability information is passed along with other information in the reply packet, so that each and every intermediate node stores the stability of that route. Two possible variants of this scheduling mechanism can be either least-stability-first (LSF) or highest-stability-first (HSF).

The evaluation of the scheduling mechanism is similar to evaluations carried out in [116]. We use the terms SNSDS-HSF and SNSDS-LSF for HSF and LSF variants, respectively. We compare these mechanisms with Round Robin (RR), Shortest-path-length-first (SPL), and Fewest-remaining-hops (FRH) first scheduling mechanisms, that are studied in [116]. The simulation environment is similar to that of the earlier experiments. The results presented here are for the case of 40 CBR sources and 40 TCP sources. We evaluate average delay and average goodput, respectively, for CBR and TCP traffics.

Figure 4.3(a) shows the average delay with varying pause time for CBR traffic. Both SNSDS-HSF and SNSDS-LSF have an advantage over other mechanisms for higher mobility scenarios(0 - 100 secs pause time). After pause time of 150 secs the performance of all mechanisms are similar except RR. There is not much difference between the two variations of SNSDS. SNSDS performs better than FRH and SPL, which performed best in [116]. Figure 4.3(b) shows the average goodput with varying pause time for TCP traffic. The results are similar to the one obtained in [116]. There is not much difference among different mechanisms with respect to the goodput values. All scheduling mechanisms perform similarly. For high mobility scenarios, use of end-to-end stability and neighbor state in scheduling does not have advantage even with respect to average end-to-end delay.



(a) Average delay of real-time traffic vs pause time (40 cbr sources)



(b) Average goodput of best effort TCP traffic vs pause time (40 cbr sources).

Figure 4.3: Scheduling mechanisms with real-time and best effort traffic

In summary, this initial work motivated us to explore the area of channel aware scheduling in MANETs. In the remaining part of this chapter, we describe the enhanced version: congestion and path lifetime aware scheduling mechanism for MANETs.

4.3.2 Motivation for Considering Path Residual Lifetime

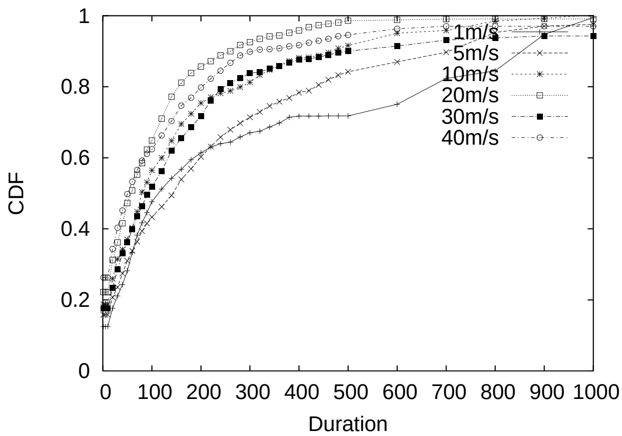
Before presenting the CaSMA algorithm, we would like to motivate the importance of considering *path duration* (residual lifetime). We argue that the problem of varying topology due to mobility has a significant impact on throughput, delay and fairness. Impact of mobility is usually studied in the context of routing protocols. In previous works on scheduling mechanism, the mobility and hence the path lifetime attribute is largely overlooked.

There are three problems that we would like to investigate. First, how mobility affects the duration of the period between link breakage? We call this duration, as the *flow on-times*, which is also termed as continuous period in our flow model. Second, how often does link breakage result in end-to-end route repair instead of local route repair? Finally, how expensive is end-to-end recovery as compared to local recovery?

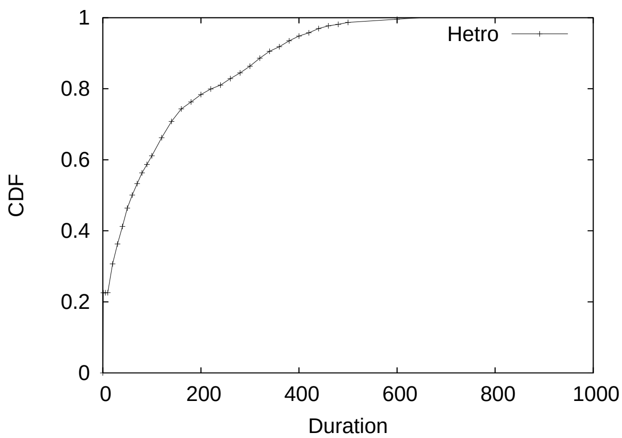
Impact of Mobility on Flow On-Times

Figure 4.4 shows the CDF of the duration of link lifetimes. Figure 4.4(a) shows CDF for random waypoint with maximum speed varying from 1 to 40 m/s. Figure 4.4(b) shows the CDF where the maximum node speed is heterogeneous (different nodes have different maximum speeds, again chosen from 1 to 40 m/s). In both Figures 4.4(a) and 4.4(b), we can see that the lifetime of the links can vary widely. This means that, at any node, if there are n flows, the lifetimes of those flows are unlikely to be similar and can vary over a large range.

Therefore, considering these lifetime values of the routes as a parameter for scheduling can be useful, and can play a significant role in improving the performance.



(a) Random waypoint



(b) Heterogenous speed

Figure 4.4: CDF of flow on-times for different speeds

Ratio of End-to-end and Local Route Repair

Due to dynamic nature, many routing protocols like AODV have in-built mechanisms for local route-recovery (route repair). Typically local recoveries are triggered when routing protocol at any intermediate node gets packet transmission failure message from its MAC layer. Whereas, end-to-end recovery is triggered when routing protocol at source node receives error-message from any intermediate node. Intermediate nodes send these error messages to

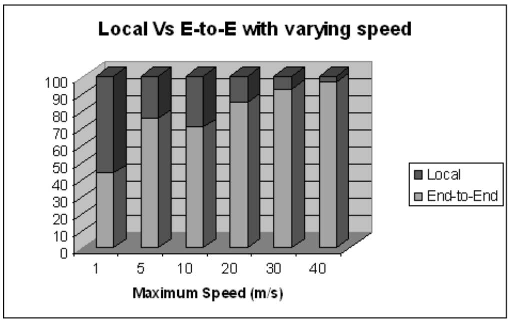


Figure 4.5: Local versus end-to-end route repairs with varying speed

source node, when it fails (fails to obtain a route) in local recovery process. The advantages of local recovery are, shorter route recovery time and can vary between 5 ms to 100 ms with potentially smaller number of packet drops. We carried out a simulation to find out the number of local and end-to-end route recovery for different mobilities.

We considered the random waypoint model, with maximum node speeds varying from 1 to 40 m/s. Among the common mobility models used, the random waypoint model results in the largest number of link breaks for a given maximum speed. We varied number of nodes from 50 - 800, and proportionately varied network area from 1 km x 1 km to 3 km x 3 km, but maintained same neighborhood density. As shown in Figure 4.5, we found that for lower mobility (1 m/s) there were more local route-recoveries compared to end-to-end recoveries. As the node mobility increases, however, end-to-end recovery dominates, and the number of local recoveries becomes small for high speeds (40 m/s).

Impact of End-to-end Recovery

The impact of end-to-end recovery comes in two forms. First, recovery time is longer, in the order of milliseconds to tens of seconds. Second, the packets buffered by the nodes on the path before any link breakage will be lost. Therefore, the cost of a link breakage is also determined by the amount of data buffered by the nodes as these packets are lost, and depending on the application may have to be retransmitted. Hence, it is important to include queue size in the scheduling decision so that the backlog for each flow is reduced.

The conclusion we can draw from the study in this section is that, packet scheduling must take into account the end-to-end channel conditions. In addition, in cases where end-to-end recovery is common, it is also important to minimize the amount of backlog data in the flow.

4.3.3 End-to-End Channel State Representation in CaSMA

In this section, we will describe how we represent or incorporate the end-to-end channel information in CaSMA. We focus on end-to-end channel awareness, and do not consider local channel information explicitly in this work. Local channel information as 2-state values (GOOD/BAD) is typically considered in many of the channel-aware schemes proposed for WLANs. This can be added to our mechanism with minor modifications. Therefore, we make this important assumption that any packet will experience loss if it is transmitted over a link whose *SNR* values (or any other multi-valued metrics) are below some threshold values (for cases where local channel quality is represented in terms of multiple values) or whose state is *Bad* (for cases where local channel quality is represented in terms of 2-values), and focus more on the end-to-end channel awareness.

End-to-End Channel State

One of the key ideas in CaSMA is to represent end-to-end channel quality in terms of path lifetimes. In this section, we describe how considering end-to-end channel state can be viewed as a variation of earliest lifetime first (ELF) approach. The residual lifetime of a path reflects the current end-to-end channel state. Since the channel state continually keeps changing, the end-to-end path has temporal interval for which they are valid. We use the term *path validity* to define the time interval for which the path associated for a flow is valid.

For a path $P = n_1, n_2, \ldots, n_k$ consisting of k nodes, at time t_0, path duration is the time interval $[t_0, t_1]$ during which each of the $k-1$ links between the nodes exist. Path duration is the minimum of the duration of the $k-1$ links at time t_0. That is, if the lifetime of each and every link of path $\mathbf{P}$ from node i to node j is estimated as $l_1, l_2, \ldots, l_n$, then the path validity $P_{ij} = min(l_1, l_2, \ldots, l_n)$. This path validity value is referred as *path lifetime* D_{ij}.

We consider path lifetime value in our scheduling process by using the shortest path lifetime first approach. This lifetime value is typically obtained by estimation techniques. The residual lifetime estimation techniques can be broadly classified as: measurement-based [55–57,125] and probabilistic-based [84,126]. There have been various proposals for both the techniques. In our work, we incorporate a measurement-based lifetime estimation technique. The details of our technique can be found in Chapter 2.

Proposition 1: In a non-preemptive queue with equal service times, the CaSMA discipline which chooses packets based on minimum-residual-lifetime first, minimizes the maximum of the lateness experienced [127].

Proof: Let us consider two queues for two flows A and B. For simplicity, let us consider the complete route as a single link. This will result in two links a and b with deadline D_a and D_b (assume $D_a < D_b$). CaSMA would schedule from queue A first. If we consider a different scheduling scheme $\overline{CaSMA}$ which chooses packets from B instead of A. If the service times are T, the packets from B will experience the lateness of $l_b = t_0 + T - D_b$. Packets from A will get opportunity some time later than $t_0 + nT$, n is some number of packets. Packets of queue A will have lateness $l_a = t_0 + (n+1)T - D_a$. For CaSMA, however, the lateness values will be $l_a^{'} = t_0 + T - D_a$ and $l_b^{'} = t_0 + (n+1)T - D_b$. We can see that $l_a^{'} < l_a$ and $l_b^{'} > l_a$. Therefore maximum of $(l_a^{'}, l_b^{'})$ is smaller than maximum of (l_a, l_b).

4.3.4 Problem Formulation

We formalize our problem as follows. Each request or flow i running through a path is described by a 6-tuple $(T_i, C_i, s_i, e_i, \{o_i\}, \{b_i\})$, where T_i is the minimum packet inter-arrival time, C_i is the maximum packet transmission time over a link, s_i and e_i are the start and termination period of a flow, finally $\{o_i\}$ and $\{b_i\}$ are the sets of continuous and breakpoint periods, respectively. We use o_i to represent single continuous period of flow i. The relation between s, e and o_i, b_i is as shown in Figure 4.6. Let us denote o_i^s and o_i^t as starting and termination of a continuous period of flow i.

Let us define the *span* of a flow f as the interval $[s, e]$. The flow f can only be served within this span. Let us also define a schedule instance I, as a sequence $(f_1, f_2, \ldots, f_n)$. How the flows are served is described by the schedule.

Formally a *schedule* for I can be seen as a function H, which can be defined as

$$H : R \rightarrow \{f_1, f_2, \ldots, f_n\} \bigcup \{\emptyset\}$$

where $H(t \subseteq span(f_k)) = f_k$. That is, k'th flow is served at time t. Further $H(t) = \emptyset$ means no flow is being served.

At any moment t_i, if a packet belonging to flow f receives a service ($H(t_i) = f$) at any of the nodes (except the penultimate node) in the path p, is said to be *partially served.* If this service is at the penultimate node of the path p, then the packet is said to be *completely served.* We also define $cs(f)$, which is the finite union of service (completely served) received by all the packets of flow f. $cs(f)$ is directly related to the goodput of a flow.

Further, to denote the *pending* state of any flow f, indicating the amount of workload remaining to be served for the queue at any time moment t, at any node, we define residue of flow as $\gamma(f, t)$.

Lastly, we define the important optimizing factors called *merit* and *backlog* of a schedule. The *backlog* is defined as the amount of packets that remain in the network at the end of their respective continuous period of all the flows.

$$\sum_{j=1}^{n} \gamma(j, o_j^t)$$

The merit M(H) of a schedule is defined as:

$$\sum_{j=1}^{n} cs(j)$$

For each flow, the scheduler gains the merit based on the number of completely served

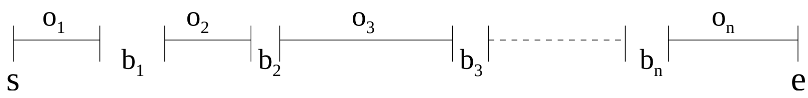


Figure 4.6: Flow model

packets. Packets which do get transmitted for a few hops and get dropped at any of the intermediate nodes will not contribute for the merit of the schedule.

The problem is to design a schedule, which over a period attains maximum *merit* and minimum *backlog*, and also fairly distributes the achieved merit among all flows. Minimizing backlog can serve two purposes. First, it reduces the delay, second, it reduces the loss due to link breakages.

In the remaining part of this chapter, we will focus only on the important flow parameters, and in this regard we will reduce the flow representation from 6-tuple to 3-tuple: (T_i, C_i, o_i). This is mainly because at any given time the scheduler is aware of a single continuous period value, and the other three parameters s, e, b_i are not accessible to the scheduler.

4.3.5 Ideal Global Scheduler and Approximation

A flow along with its span (start and end times), is also defined by its *breakpoints* and *continuous period*, as indicated in the flow model (Figure 4.6). A breakpoint is the duration of time during which an attempt to transmit a packet of that flow will result in failure. In this work, neglecting packet loss due to congestion, we will consider mobility as purely the reason behind the loss of link between the transmitter and the receiver. This duration can be in the order of seconds to minutes. The occurrence of breakpoints is more frequent in ad hoc networks compared to wireless LANs. Channel aware scheduling schemes like CaSMA are designed exactly to handle these breakpoints. Therefore, the channel aware scheduling in ad hoc networks plays a more significant role compared to wireless LANs.

For any given moment of time, we can only deal with single value of breakpoint and continuous period for any given flow, as we do not have any information about the future values. We consider the cost of not completing the service before the breakpoint as one or more of the following:

- Packet queued at the intermediate node may not reach the destination after the continuous period. They may be dropped or may have to be retransmitted.
- Any attempt to transmit these queued packets at the intermediate nodes may result in wastage of resources.
- Packets might reach the destination unordered.

This effect can be mathematically described as follows. Let us define α_i as the amount of service that is required to make the flow schedulable in any i^{th} continuous period. Let us also define β_i as the actual amount of service that is received during the i^{th} continuous period. Then, at any continuous period j of a flow f, we define γ difference between the expected and the actual service as:

$$\alpha(f, j) + \gamma(f, j-1) - \beta(f, j) = \gamma(f, j)$$

where $\gamma(f, -1) = 0$. γ models the impact of not completing the service within the continuous period.

Global Ideal Scheduler

Let us consider a simple model with multiple flows over a single bottleneck link where we have a single scheduler. After the single shared link (with infinite lifetime), these flows use different (non-shared) links with different lifetime.

Let us assume a global scheduler - S_i, which schedules these flows ("m" flows). Let us consider a single continuous period c of "m" flows, with arrivals within this continuous period, and no further arrivals. That is, let us take a single snapshot in time of m flows with each flow having single continuous period of varying durations. For simplicity, let all flows have same $T = 1$, and $C = 1$. Therefore, S_i can schedule at most r_{max} packets in o_{max} period, where o_{max} is the maximum continuous period of any flow (or total interval of the snapshot). r_i represents the number of packets existing for flow i.

We, however, know that maximum number of packets existing in all the queues is $\sum_{i=1}^{n} r_i$. Let us call this value as r_{sum}. Therefore, percentage ratio of throughput would be $\frac{r_{max}}{r_{sum}}$. Now, we adopt a fairness criterion, where this ratio is maintained across all the flows. In other words, the losses/backlog is proportionately distributed across all the flows. The idea here is that all sharers constrained by the same problem are treated fairly by assigning the proportionally equal throughputs. That is, for each flow i, the throughput it would receive is

$$r_i * \frac{r_{max}}{r_{sum}}$$

Further, losses at each queue would be

$$r_i(1 - \frac{r_{max}}{r_{sum}})$$

The rationale behind having this formulation for an ideal scheduler is based on the argument that shorter continuous periods of flows are purely due to the inherent property of ad hoc networks. Therefore, we believe in not penalizing flows which suffer due to the inherent property of the network. Further, the scheduler will not be aware of the amount of service a flow has received in the previous continuous period (if existed) or the amount of service a flow will receive in the next continuous period. Therefore, we go by the assumption that providing equal proportion of service in the current set of continuous periods would probably prove to be advantageous.

Approximation

Our approximation to the global scheduler has two steps. First, we show that use of the parameter $\frac{QS}{RLT}$, where QS is queue size and RLT is the residual life time, approximate the ideal-scheduler described above. Second, we describe the schedulability list technique, which shows how the decision made at first node would be sufficient enough, and encompasses the decision for the whole path. Schedulability list technique also results in maximizing the merit of the scheduler.

Use of $\frac{QS}{RLT}$ to Approximate Ideal Scheduler

We have to show that when we schedule using $\frac{QS}{RLT}$, and when scheduling mechanism can schedule at most r_{max} packets, the number of packets served for each flow i is approximately $r_i * (\frac{r_{max}}{r_{sum}})$. The important idea here is to solve two problems: (1) provide higher priority to flows which take short-lived paths, and (2) proportion of service received for each flow will remain similar.

Let us consider the scheduling approach where queues with maximum value of $\frac{QS}{RLT}$ is always chosen, where QS is queue size and RLT is residual lifetime (also called continuous period o). Serving every queue considering $\frac{QS}{RLT}$ is similar to rate monotonic scheduling (RMS). RMS is an optimal, static-priority scheduling used in hard real-time systems [128]. Higher priority is given to a flow which has higher request rate. RMS aims at maximizing the number of tasks meeting its deadlines. In CaSMA, $\frac{QS}{RLT}$ acts as a request rate. Therefore, serving queues which has higher $\frac{QS}{RLT}$ values first will result ing providing higher priority to flows which take short-lived paths.

Now we have to solve the second problem of providing equal proportion of service. This is because, considering only $\frac{QS}{RLT}$ may not provide equal proportion of service. Let us consider a simple model, which is a single snap-shot in time, where we have n flows with each flow i having workload (number of packets) as r_i. Let the service time for all packets be same (1 time unit). Further, as all flows have T and C set to 1, then $o_i = r_i$, i.e., RLT for each flow will be the same as r_i (to begin with all flows have same request rate). Let the maximum number of packets the scheduler serves in the given time duration(or the maximum duration of time snap-shot) be maximum of r_i values, termed as r_{max}. Let us term the number of packets served for flow i be X_i, and use r_{sum} to represent sum of all r_is.

We know that QS either decreases or remain the same (as we consider single snapshot in time and no further arrivals), and RLT is strictly decreases. Therefore request rate ($\frac{QS}{RLT}$) can either remain the same or increase. Using the above model and notations we can rewrite request rate as

$$\frac{r_i - X_i}{r_i - \sum_n X_j}$$

Let us define another parameter α ($0 < \alpha < 1$), which is the proportion(percentage) of r_i of service that any flow i receives at any given time within the considered time duration. The important point to note here is that, there is no one-to-one mapping between the request rate considered and proportion of service received (α). That is, if a flow i has greater request rate than the other flow j, then it may not mean that amount of the service (proportionately, α) received by the flow i is lesser than j. In fact, when the o_is varies to a larger extent, the proportionate amount of services received by flows can also vary to a larger extent (flows with shorter continuous periods (o_i) will receive proportionately greater service).

For a special case where o_is are same, if a flow has received lesser proportion of service than the other flow, then its request rate will always be higher than the other flow. Under these conditions (similar o_is), it can be shown that serving by $\frac{QS}{RLT}$, results in fair distribution of service.

We have seen in the preceding section (Section 4.3.2) that o_i values can vary to a great extent. Therefore, we need to avoid the condition where short-lived flows can receive proportionately greater service. We achieve this by having an additional parameter termed as $eligible - service$, for each flow. This $eligible - sevice$ for any flow i is equivalent to $\frac{r_{max}}{r_{sum}}$, and is computed by considering the r_is, which is given as follows:

$$\frac{(r_i * \frac{C_i}{T_i})}{\sum_{j=1}^{n} r_j * \frac{C_j}{T_j}} * (r_{max} * \frac{C_{max}}{T_{max}})$$

C_{max} and T_{max} indicates maximum possible C and T, respectively. The first term indicates the ratio of the work to be performed for a flow i and the total amount of work considering all flows. Whereas, the second term indicates the maximum work that can be done, and this term, in practice, is related to the maximum wireless link rate.

We update this parameter ($eligible - service$) only when new flows arrive or existing flows leave. The priority is given to flows considering both the request rate and eligible-service. Higher priority is given to flows whose request rate is high, and which has not yet received its eligible-service. This parameter will ensure that flows do not receive greater service (in proportion) at the cost of other flows.

In the remaining part of this section, we will describe how we enhance the approximation of ideal scheduler by considering end-to-end packet scheduling.

Schedulability

A set of flows Γ is said to be "schedulable" (S) if none of the flows has packets queued in the intermediate nodes at the end of their respective continuous periods. Any set of flows at a node that are schedulable over a link is termed as "schedulable set".

We consider the following two problems related to flow schedulability. First, we have to consider that given a set of n flows $\Gamma = (T_i, C_i, o_i)$, $i = 1, 2, \ldots n$, how many of them (m, $m \leq n$) are schedulable over a link? (schedulable set). Second, suppose there are n flows $\Gamma = (T_i, C_i, o_i)$, $i = 1, 2, \ldots n$, of which m flows form a schedulable set ζ. Now, given a new flow j, what is the maximum value of its continuous period (o_j), such that the new flow will be subset of the schedulable set (may result in preemption of a flow existing in the current schedulable set).

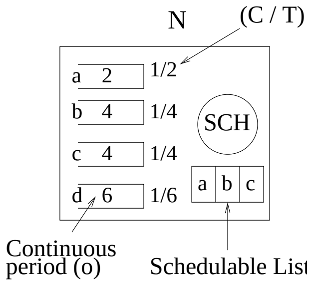


Figure 4.7: Schedulable set example

We will provide the solution for the above two problems, which will be used in our scheduling algorithm. First, let us begin with the schedulable set (ζ). A schedulable set (set of flows that are schedulable) is derived as follows. Let us assume that a node has n flows, of which it has to choose m flows to form a schedulable set. We use the classic result of real-time scheduling [129], and define the necessary condition for a set of flows to be schedulable over a link is given as

$$\sum_{i=1}^{m}\left(\frac{C_i}{T_i}\right) \leq 1 \tag{4.1}$$

In addition, we know that there are different combinations that are possible in choosing m flows out of n flows (C_n^m). We know that the value of m is dependent on the C_i and T_i values. For example, value of m becomes smaller for smaller values of T_i. Hence, we have to decide on a specific way to choose m flows out of n flows.

In our work, we choose the m flows considering the residual lifetime values of the flows. Scheduling based on residual lifetime is similar to earliest deadline scheduling (EDF). Therefore, based on the results from EDF scheduling [128] and adhering to the approach of choosing smallest residual lifetime first, we sort all the n flows in terms of the increasing residual lifetime, and from this sorted set we choose the first m flows. These m flows form our schedulable set ζ.

To simplify the understanding, let us take an example, as shown in Figure 4.7. Let $\{a, b, c, d\}$ be the flows at any node 'N'. Let $\{2, 4, 4, 6\}$ and $\{1/2, 1/4, 1/4/, 1/6\}$ be their continuous periods and rates $(\frac{C}{T})$, respectively. SCH indicates scheduler at node 'N'. Node 'N' chooses flows $\{a, b, c\}$ as schedulable following the condition given by equation 4.1. Flows $\{a, b, c\}$ are chosen considering their continuous periods and the rates. We can see that an addition of flow d will violate the condition, that is, summation of the rate values $(\frac{C_i}{T_i})$ will be greater than 1.

We can also rewrite the above necessary condition in terms of the packets scheduled. Since the minimum packet inter-arrival time of a flow i is T_i, there are at most $\frac{(o_i)}{T_i}$ packets arrived over channel i during the interval, and which need at most $\frac{(o_i)}{T_i}C_i$ units of time to transmit. Now the summation of this time for all the m flows should be less than the r_{max} (maximum number of packets the scheduler serves), which is written as

$$\sum_{i=1}^{m}(\frac{o_i}{T_i})C_i \leq r_{max}$$

The solution to the second problem follows the solution of the first problem. If a node has a set of flows Γ passing through it, we define a schedulable set ζ $(\zeta \subseteq \Gamma)$ where ζ is the set of flows which are schedulable at that particular node. Let the maximum continuous period in the set ζ be o_j of some flow j. The schedulable set ζ also satisfies the necessary condition provided above. Now the maximum value of continuous period for a new flow, say k to be schedulable is to be lesser than o_j, and the arrival rate is lesser than or equal to $j\acute{s}$ arrival rate. That is, a new flow k with continuous period o_k will be schedulable, iff $o_k < o_j$ and $\frac{C_k}{T_k} \leq \frac{C_j}{T_j}$. This is because, the schedulable set is built considering two conditions - residual lifetime and the necessary condition as given above.

If the continuous period of the new flow (k) is lesser than the continuous period of a flow (j), where flow j is both a member of the existing schedulable set and has a maximum continuous period in the set, then the new flow (k) will be added into the schedulable set at the expense of this existing flow (j, which had maximum continuous period will be preempted). In addition, the second condition ($\frac{C_k}{T_k} \leq \frac{C_j}{T_j}$) is important to make sure that the new schedulable set does not violate the condition given by the equation 4.1. Therefore, for a flow to become eligible as a member of the existing schedulable set is that its continuous period be lesser than the maximum continuous period in the existing schedulable set.

Considering the example in Figure 4.7, if a new flow has to become schedulable then its continuous period has to be < 4, and request rate has to be $\leq \frac{1}{4}$.

The solution to the second problem leads to the notion of a flow i being "schedulable" (S) at node l. This notion provides an important parameter in our analysis, as it is used in two ways: (1) An end-to-end measure of this value during the path set-up helps the source to decide on initiating the traffic (2) Intermediate nodes make their scheduling decision based on these values, which can be updated by the downlink neighbors whenever value changes.

We know that if a flow is schedulable at all the intermediate nodes, then it is schedulable over the path. The idea is analogous to the *series of traffic lights*. It is useful to turn the first light green when all the remaining lights will turn green within some acceptable duration. This technique helps in increasing the *merit* (as described in the problem formulation section) of a scheduler, as priorities are given to packets which will be "completely served".

The notion of schedulability takes on only binary values. When we use this parame-

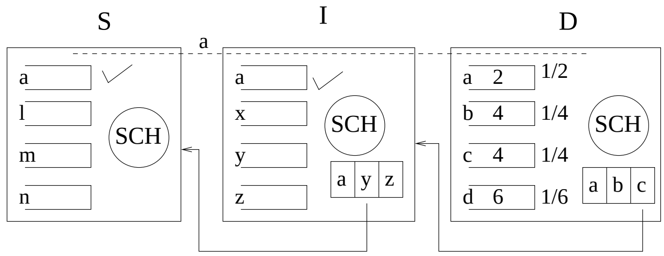


Figure 4.8: Schedulability example

ter in the algorithm, the mechanism just makes the decision for given values and existing conditions. This decision process is used to build the schedulable-list message, as described below, in continuation with the example in Figure 4.7.

Consider three nodes S, I, and D as shown in Figure 4.8. We will focus on a single flow 'a' starting at node 'S', with intermediate node 'I' and terminating at node 'D'. Let {a, b, c, d} be the flows at 'D'. Let {2, 4, 4, 6} and {1/2, 1/4, 1/4/, 1/6} be their continuous periods and rates ($\frac{C}{T}$), respectively. Node 'D' chooses flows {a,b,c} as schedulable following the condition given by equation 4.1, and creates a schedulability-list message (list of flows schedulable), which is transmitted to the upstream neighboring nodes. When 'I' receives this message, marks flow 'a' as schedulable (at downstream) and builds its own schedulable-list (let it be {a, y, z}) and transmits it to its upstream neighbors. In this manner, the schedulable-list message flows upstream until it reaches source node 'S', which upon receiving will mark flow 'a' as schedulable (at downstream). If either the destination node or any of the intermediate nodes does not include flow 'a' in their schedulable list message, then the source node will not set flow 'a' as schedulable (at downstream).

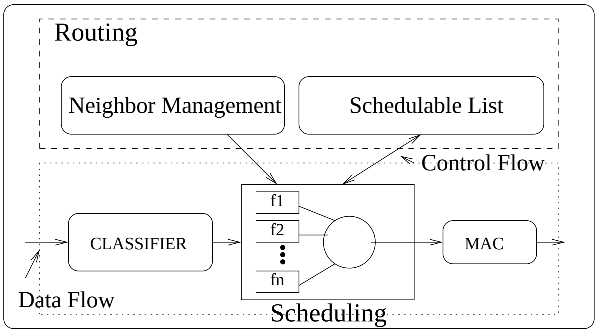


Figure 4.9: Framework of CaSMA

4.3.6 Approach, Framework, Algorithm and Limitation

In this section, we present the framework and implementation of CaSMA. CaSMA is designed to give preference to those packets that are determined to be urgent, where urgency depends on factors such as residual lifetime, queue size of a flow, and throughput received. Therefore, it necessarily involves time-varying properties. Our scheme implicitly assumes that a packet urgency increases with the imminence of its lifetime.

It could be argued that the parameter set used in CaSMA, can be used as either "weighted round-robin" or use them as priorities in dynamic priority scheduling. We believe that the parameter-set is such that, it is more applicable to dynamic priority scheduling, than the weighted round-robin techniques, as round-robin techniques try to achieve fairness, and the problem of Head-of-Line (HoL) blocking will persist.

Residual life-time is measured end-to-end whereas queue size is measured locally. Assigning priorities based on queue-size also has a significant effect on end-to-end performance. This combination also mitigates the inaccuracies associated with the end-to-end measurements to some extent. Further, we have proved in the earlier section that this approach also approximates the ideal case. We compute the eligible-service for each flow using the equation given in Section 4.3.5. We also associate schedulable-list with each flow, and at every node. This helps in approximating the ideal global scheduler. Apart from these measurements, we also include “fixed priorities” of the flow, based on the flow-types, and throughput measurement for every flow, which are used to break the tie. Exact arbitration criterion used in CaSMA is better explained in the algorithm described below.

The framework, which is used for the realization of CaSMA is as shown in Figure 4.9. This framework includes three mechanisms: routing, classifier, and scheduling mechanism. The routing protocol used in this work is Dynamic Source Routing (DSR) [34]. DSR is used purely for the ease of implementation, and any other routing mechanism can be modified to include CaSMA. DSR is enhanced with two schemes: *neighbor management scheme*, which is used for lifetime estimations, as described in [125] and also in Chapter 2, and *schedulable-list scheme* is used to implement the schedulable-list technique as described in Section 4.3.5. The classifier classifies the arriving packet to one of the different per-destination queues maintained in the scheduling module. A single queue is maintained for every destination of the flows that a node carries, i.e., different flows to the same destination are enqueued in the same queue.

Typically aggregation of traffic flows follows either class-level (per-class) or path-level (per-flow). Per-class approach is appropriate when the traffic density is high, whereas, for low and moderate traffic density per-flow approach would suffice. Per-flow approach does scale well in comparison with per-class approach. Earlier research has shown that per-flow approach has better bandwidth management (in context of heterogeneous wireless networks) than per-class approach. Per-class approach is easier to develop in comparison to per-flow approach, as per-flow approach requires complex algorithms. In our work, path-level aggregation is used. An important practical concern is that when the number of flows increases (in terms of hundreds and thousands) in the network, the computation complexity also increase proportionately. Therefore, the solution is feasible for low to medium size networks. Majority of the existing MANET systems are either low or medium sized networks. In addition, unlike wired network, we do not include costly per-flow state maintenance process.

If we have per-neighbor queues or per-flow queues, the processing becomes complex. This is mainly because a single neighbor can be associated with many destinations, whereas a destination is typically associated with a single neighbor. Further, in these two cases a node has to differentiate packets twice to consider different parameters (local, end-to-end channel information, and fixed-priorities). Whereas, for per-destination queues, packet differentiation will happen only once. The scheduler chooses appropriate packets, following the algorithm as described below, and passes the packet to MAC to continue the transmission process.

Algorithm

The algorithm is as shown in **Algorithm** 2. The algorithm used to choose the packet from queue is as shown below. Among the set of flows (with or without priority), queues which are a subset of the schedulable-list are chosen. If none of the queues satisfies the condition of being the subset of its schedulable-list, then all the queues which were chosen first are considered. From this chosen set, queues with highest $\frac{QS}{RLT}$ value and those who have not received their eligible-service are chosen. If there is a tie, queue that has received least throughput is chosen.

Algorithm 2 Packet selection in CaSMA

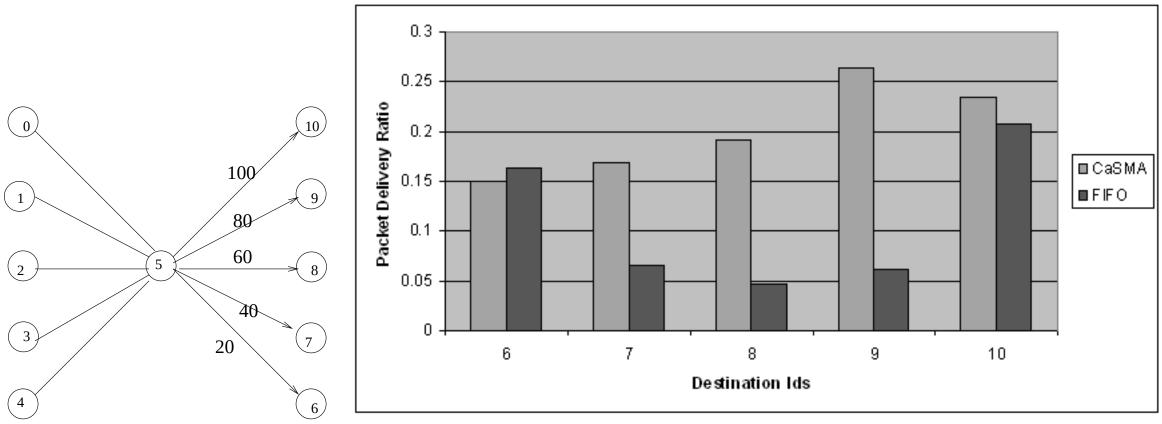
Require: Initialize

- Per-destination queues are maintained,
- Each queue has [queue size, residual life-time, eligible-service, neighbor's schedulable-list (SL) (flows schedulable at downstream), priority, and throughput received]

repeat
  Consider a set of high-priority (real-time) queues
  From this set of high priority flows {HP}.
  Select the queue q such that $q \subset SL$
  if No queue satisfies the condition ($q \subset SL$) **then**
    Select all the queues {HP}.
  end if
  From these selected queues:
  Select queue q such that the value $\frac{QS(q)}{RLT(q)}$ is the maximum, and who have not yet received *eligible-service*
  In case of tie select flow that has received least *throughput*
until all queues are empty

Limitations of CaSMA

- CaSMA assumes a path/link lifetime estimation technique. As no standard technique exists till date, CaSMA's performance varies as the accuracy of link estimation varies.
- Neighbor management and schedulable-list scheme can add overhead with respect to bandwidth consumption, especially for high-mobility scenarios.
- Some instances of flow-level unfairness still exists, such as, long lived (longer residual lifetime) flows can suffer, purely because of the existence of a lot of short lived (short residual lifetime) flows.
- If the topology changes too rapidly, estimations can be inaccurate, resulting in lower performance.



(a) Topology (b) TCP throughput

Figure 4.10: Packet delivery ratio for different flows

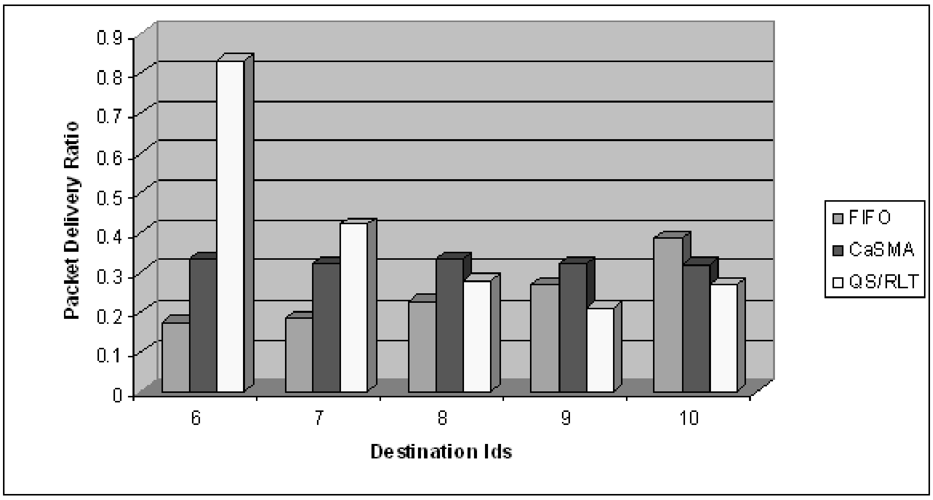


Figure 4.11: Packet delivery ratio for different flows

4.3.7 Experimental Evaluation

In this section, we describe the experimental evaluation of CaSMA. In the first part of the simulation we consider a scenario where the scheduler has perfect knowledge of the link lifetimes. The goal of this part is to provide the reader a better understanding of the advantages of CaSMA, when there are no lifetime estimation errors. In the second part of the simulation, we consider scenarios where link lifetime is estimated, and we compare the performance of various scheduling mechanisms. All our evaluations are carried out on NS-2 [11] simulator. Each mobile host has a transmission range of 250 m and shares a 2 Mbps radio channel with its neighbors. The simulation includes a two-ray ground reflection model and IEEE 802.11 MAC protocol.

Performance Comparison with Known Path Lifetime

In this section, we focus on understanding the significance of the parameters considered (QS, RLT and eligible-service). We considered a simple topology of 11 nodes, and simulation duration of 100 seconds. The topology is as shown in Figure 4.10(a). The source-destination pairs are [(0,6),(1,7),(2,8),(3,9),(4,10)], with single intermediate node 5. In Figure 4.10(a), the numbers shown on links between node 5 and {6, 7, 8, 9 and 10} indicate the respective link lifetimes.

We consider CBR flows transmitting at 400 kbps. The packet delivery ratios are shown in Figure 4.10(b). The delivery ratios for CaSMA are both even and higher compared to FIFO. For flow [0-6], FIFO has slightly better delivery ratio than CaSMA, but it performs badly for other flows. The delivery ratios are higher for CaSMA because CaSMA does not make an attempt to transfer those flows, whose link lifetime has expired. This shows that CaSMA is designed to provide service to the flows within their "lifetime" and not beyond that.

To focus on the importance of *eligible-service*, we slightly modified the source-destination pairs. Now, all the 5 flows initiate from node 5, flowing towards same destination, with same RLTs. The transmitting rate, however, is increased from 400 kbps to 600 kbps. Figure 4.11, shows the packet delivery ratios for FIFO, CaSMA and $\frac{QS}{RLT}$ (without eligible-service). We can see that CaSMA, achieves both better packet delivery ratio and proportionate share. Though $\frac{QS}{RLT}$ (without eligible-service) performs better than FIFO, the division of share is not fair (flow [5-6] gets proportionately greater share). This is precisely the case for which eligible-service is included to handle, which results in providing fair share. It can also be seen that performance trend of FIFO and $\frac{QS}{RLT}$ tend to be opposite. That is, FIFO's performance increases with link lifetimes, whereas $\frac{QS}{RLT}$'s performance decreases with link lifetime.

In summary, CaSMA is designed to perform such that flows with lesser residual lifetime get higher preference, and the losses (throughput) will remain proportionately same for all contending flows.

Performance Comparison with Estimated Path Lifetime

In our second part of the simulation, we consider scenarios where lifetimes are estimated. This mobile environment is considered to emphasize on the advantage of using schedulable-list scheme along with the other parameters ($\frac{QS}{RLT}$ and eligible-service). We consider a network with 50 mobile nodes, with area 1000 m x 1000 m. All the simulations are run for 1000

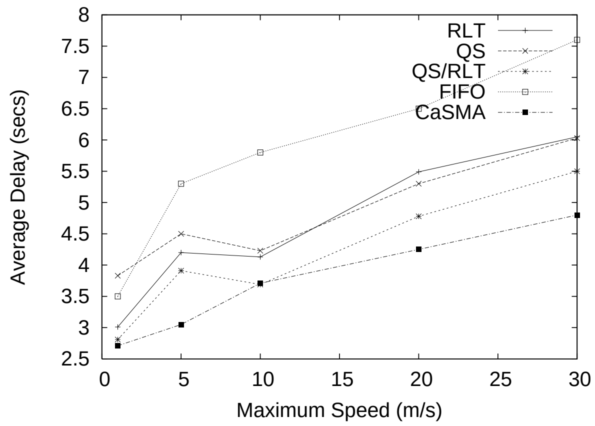


Figure 4.12: Average delay versus maximum speed

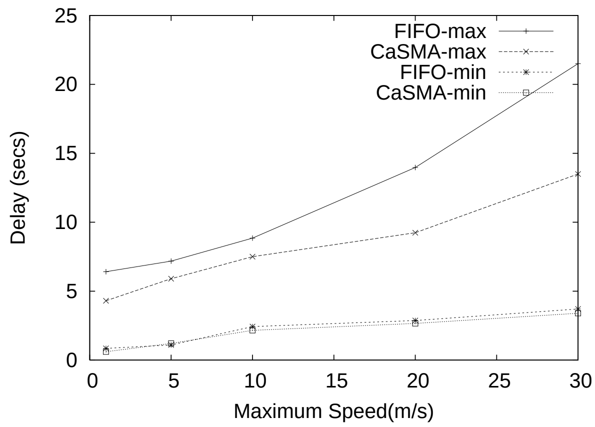


Figure 4.13: Max and min delay versus maximum speed

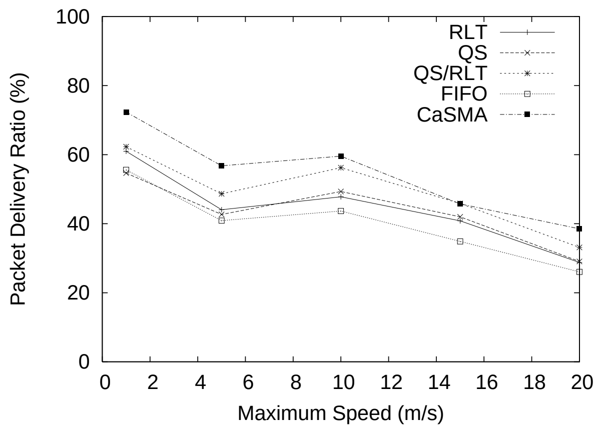


Figure 4.14: Packet delivery ratio versus maximum speed

seconds, with 8 replications. The actual traffic flow, however, is started only after the 2nd half of the simulation duration. This is to allow the routing protocol to collect enough history of link lifetimes, so that the estimation of residual lifetime is accurate. We had seen in our earlier work that greater the amount of history available, better the estimation would be. In this part of evaluation, maximum speed of the node is varied from 1 m/s to 20 m/s.

The five mechanisms chosen are: FIFO (First In First Out), RLT (considering only residual life time), QS (considering only queue size), QS/RLT (considering both queue size and RLT), CaSMA (considering, queue size, RLT and schedulable-list).

For the first set of plots, we use 10 CBR flows with the transmission rate of 500 kbps. Figure 4.12 shows the plot of average delay values for all the five mechanisms. Considering the Figure 4.12, CasMA performs best among all the schemes. This can be attributed to both the parameter chosen and the schedulable-list technique. Considering only the performance between QS/RLT and CaSMA, we can see the advantage of using schedulable-list technique.

To have a better understanding of the advantage we note the maximum and minimum of delay values, considering only FIFO and CaSMA. Figure 4.13 shows that the maximum values of CaSMA are also lesser compared to FIFO, whereas minimum values are almost the same. The main reason behind the reduction in delay values (average and maximum) is due to a reduction in the backlogs (or γ values, as described in preceding sections). The increase in backlogs can result in transmissions after a route-recovery delay. The backlog increase also has effect on the losses.

Figure 4.14 shows the average packet delivery ratios with varying maximum node speed.

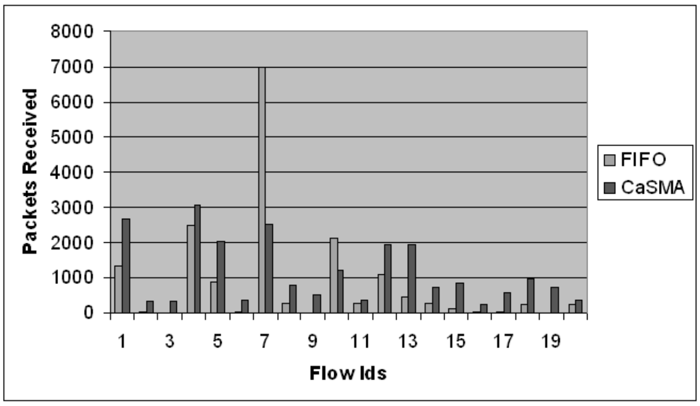


Figure 4.15: Throughput versus maximum speed

We can see from the plot that CaSMA outperforms other schemes. The performance difference is not considerably high, however. To understand better, we considered a similar scenario with 20 flows (start at random times), and studied the bandwidth share allocated for each flow. Figure 4.15 shows the plot for number of packets received for FIFO and CaSMA. We can see that, using FIFO flows {2, 3, 7, 9, 16 and 19} were almost starved. Whereas, with CaSMA the sharing of bandwidth is more fairer and better compared to FIFO. Further, CaSMA has 25% less packet loss compared to FIFO.

Figure 4.16 shows the number packets that are dropped at the queue due to link breakages. This parameter is directly related to the amount of backlog. From the figure, we can see that the backlogs using CaSMA is reduced by more than 30% - 40%. Further, we can see that increasing the frequency of topology changes, the amount of backlog also increases.

We further consider 10 TCP flows, and study the TCP performance in such scenarios. TCP flows are considered because, if the scheduler attempts to schedule a packet whose path residual lifetime has expired, with high probability, it will result in dropping. This dropping will force TCP to reduce the congestion window, and in turn reduce the throughput. Figure 4.17 shows the throughput performance of various schemes. We found that the TCP

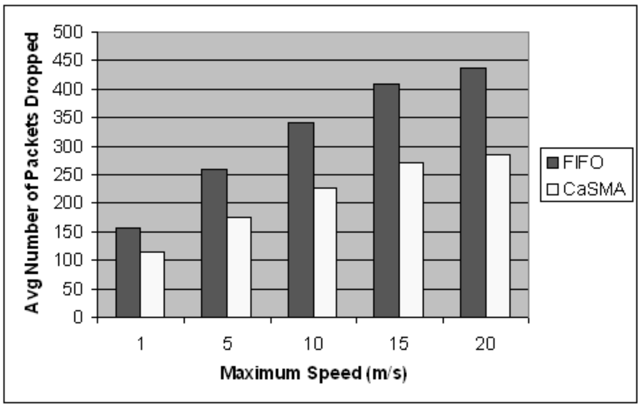


Figure 4.16: Number of packets dropped at queue due to link breakage versus maximum node speed

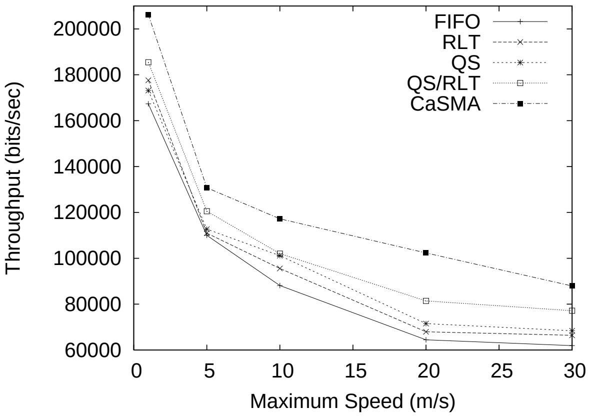


Figure 4.17: Throughput versus maximum speed

throughput for CaSMA increased in some cases up to 50% over FIFO. The reasons behind better TCP performance are the same as provided in the first part of this section.

4.4 Summary

In this chapter, we proposed a novel scheduling mechanism considering the inherent feature (existence of multihops) of MANETs. We consider end-to-end channel condition represented as residual lifetime for channel-awareness, and also included a queue size parameter to make the scheduling scheme congestion-aware. This combination of parameters avoids the congestion and reduces the accumulation of packets (backlogs) at the end of flow on-times. Further, we included a schedulable-list technique, which apart from providing better end-to-end co-ordination and approximation to an ideal scheduler, also increases the merit (number of completely served packets) of the scheduler.

Chapter 5

UNIFIED Service Differentiation Solution

In the previous three chapters we described three components of our service differentiation solution - route computation, admission control and path scheduling. In this chapter, we will describe the service differentiation solution architecture combining the three components. Our motivation behind developing this solution is to study the combined performance of the three components and also to demonstrate the flexibility of the schemes developed. Service differentiation solution is used as a case study, and we perform a comparative study with the existing service differentiation solution proposal. We will also provide the detailed evaluation of the combined solution.

5.1 Introduction to Protocol Architecture

Protocol architecture specifies the decomposition of a system into functional modules, and also the semantics of the individual modules [130]. Decomposing into modules provides necessary abstractions for a designer to understand the overall system. It further provides flexibility in designing and developing individual components in parallel. Therefore, a key architectural requirement is flexible decomposition. Further, it is important to consider proper architecture design as a performance optimization, on a long run. von Neumann architecture for computer systems in a classic example of good architecture design, and shows the advantages of providing importance to architecture [89].

In the context of computer networks, however, ISO/OSI layered architecture on which the Internet architecture is based is another example of excellent architecture design. Some researchers do conclude that success of Internet is mainly due to its layered architecture [89]. Detailed descriptions of this layered architecture can be found in [130]. The success of layered architecture for wireline networks has also influenced wireless networks design, and it has become the default architecture for designing wireless networks. There have been various works which show that layered approach is indeed a good option for base architecture design for even wireless networks. There are also proposals, however, focussing on the optimization of design architecture. One such popular optimization approach is cross layer design [89].

The idea behind cross layer design is to explore a variety of ways of interaction across layers that are possible. The rationale is to address the trade-off between the performance (short term advantages) and the architecture (long term advantages).

Our solution architecture also incorporates few cross layer interaction features. From the description of the architecture provided in the Section 5.4, it can be noticed that we take into consideration of the cautions pointed out by Kawadia et al. on cross-layer designs. We term our solution as UNIFIED (Unique Features Influenced) solution for service differentiation for MANETs.

5.2 Introduction to Service Differentiation

The future Internet, or any network, will be dominated by the mobile, hand held devices. Nodes will be required to support application of different types (best-effort and real-time). MANETs would be no exception in carrying data from different applications [131]. Different applications have different requirement. For example, real-time applications such as Voice on Demand (VoD), Virtual Classrooms (VC), and Telephony are sensitive to packet loss and delay, and may have minimum bandwidth requirements. For any network to support such applications, there should be mechanisms, which provide better service to these real-time traffics. Therefore, there should be a differentiation in service provided when multiple traffics are supported. The idea of service differentiation is depicted in Figure 5.1. A node or router can receive different traffics, which can have different set of requirements. Based on the type of traffic, the node has to provide different service to different traffics. For example, a real-time traffic may have requirement of minimum amount of bandwidth. The mechanisms within a node will try to satisfy this requirement by providing higher priority and greater share of the bandwidth. In other words, when a node/router includes mechanisms to handle different traffics with different requirements in a different way, then it is said to be providing service differentiation. In the context of Internet, the difference in service is achieved by Quality of Service (QoS) models, where approaches like guaranteed services, differentiated services and flow protections are included [132]. When the future MANETs needs to support real-time services, it should also include mechanisms which can provide service differentiation. Whether any of the solutions proposed for Internet QoS would be applicable to MANETs, however, is still an open research problem.

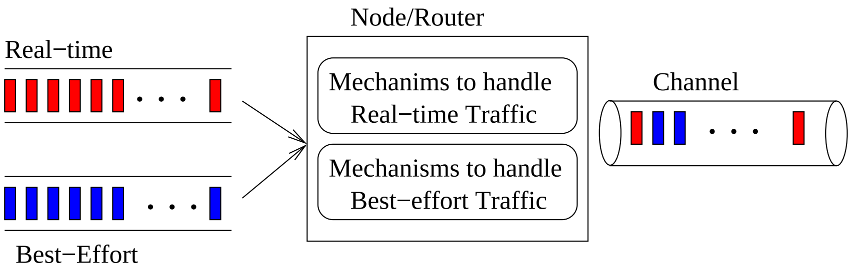


Figure 5.1: Service differentiation

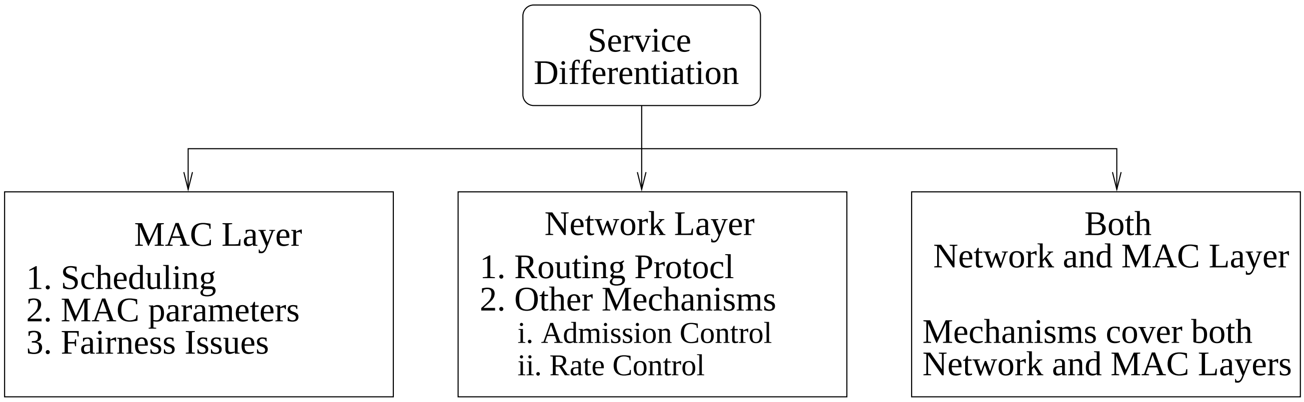


Figure 5.2: Related works classification

5.3 Related Works

5.3.1 Resource Management in MANETs

In this section, we describe the related works on resource management schemes for MANETs.

For easy understanding of related works we classify them based on the scope of the work. Figure 5.2 shows our classification criteria. We classify it considering the layered architecture and the focus of the work. We consider three cases: MAC Layer - aim to provide service differentiation by including/modifying the mechanisms of MAC layer. Network Layer - routing and mechanisms like signalling, admission control, which function at network layer and focus on achieving service differentiation in MANETs. In the last case we consider those works whose focus is on both MAC and Network layers.

MAC Layer

In this section, we discuss works in which the main focus is to achieve service differentiation through MAC layer mechanisms. In this category, we consider works, which modify IEEE 802.11 to provide differentiation, which propose MAC protocols that provide service differentiation, and which modify scheduling mechanism to achieve differentiation.

Unlike other wireless networks such as cellular networks, there is a lack of centralized control and global synchronization in ad hoc wireless networks. This attribute makes Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) schemes unsuitable. There have been various proposals for MAC protocols for wireless environment [6,133–136]. Most of the popular MAC protocols are based on Carrier Sense Multiple Access (CSMA) paradigm. All aim to solve popular problems that exist in wireless scenarios- the hidden and exposed node problems. Two notable examples are MACA [133,134] and IEEE 802.11 [6]. MACA introduces reservations using RTS/CTS exchange. A variation of MACA, namely MACAW, also recognizes the importance of congestion, and exchange of knowledge about congestion level among entities.

The most popular among MAC protocols used for simulation and experiments currently is IEEE 802.11. This has the RTS/CTS dialogue similar to MACA or MACAW. A study [92] has revealed the failure of 802.11 in ad hoc multihop wireless environment. It showed that the effect of hidden/exposed node problem in multihop environment is definitely noticeable and sometimes high, and also 802.11 has different kinds of unfairness associated with it. There are proposals, however, to improve IEEE 802.11 to provide service differentiation. Work by Aad et al. [137] propose two differentiation mechanisms for IEEE 802.11. The first proposal is based on tuning the contention window according to traffic time, and second proposal is based on tuning interframing spacing values for different users.

A QoS MAC protocol besides dealing with hidden and exposed node problems must also provide resource reservation and QoS guarantees to real-time traffic. There are few proposals for QoS MAC [138, 139]. In [139], authors differentiate stations into high and low priority stations. For low priority stations, the access method is the normal CSMA/CA mechanism used in the legacy 802.11. For high priority stations, when station tries to send packets, it first senses the medium to verify whether it has been idle for an interval of time PIFS. Once the medium is sensed idle, it can send packets. If the medium is busy, the station will wait for the medium to be idle for PIFS and then enter a Black Burst (BB) contention period. In BB contention period, the station sends a black burst signal to jam the channel. After transmitting the black burst signal, the station listens to the medium for a short period to see if some other station is sending a longer BB which would imply that the other station has waited longer and thus should access the medium first. If the medium is idle, the station will send its packet; otherwise it will wait until the medium becomes idle again and enter another BB contention period. After each successful transmission, the station will schedule the next transmission for a fixed time interval.

In [138] similar to BB contention scheme, a different MAC protocol is used for different types of traffic. For non-real-time traffics, MAC protocol is the same as the legacy 802.11. There is a major difference between BB contention and MACA/PR for real-time traffic. MAC protocol used in MACA/PR is based on a specific reservation scheme. To record the reservations successfully made, each station maintains a table called Reservation Table (RT). The whole protocol is composed of two parts, namely Reservation Setup (RS) and

Reservation Maintenance (RM).

Other MAC protocol proposals, which provides service differentiation are [140, 141]. To support QoS in WLAN, IEEE 802.11 Task Group E recently defined enhancements to 802.11-WLAN, called IEEE 802.11e [140], which introduces a channel access mechanism termed as Hybrid Coordination Function (HCF). HCF includes access mechanism termed as contention-based channel access, also referred to as Enhanced Distributed Coordination Function (EDCF). EDCF enhances the original DCF to provide prioritized QoS support. It differentiates the level of QoS through the introduction of four Traffic Classes (TCs). Each TC has its own transmit queue and its own set of TC parameters. The prioritized QoS is realized by setting different values for the TC parameters. SEEDEX [141] is based on an idea to employ random schedule scheme, which is controlled by a pseudo-random-number generator. Exchanging the seeds of random-generator, nodes publish their schedules to all and potential hidden and exposed nodes. This proposal provide service differentiation by choosing different value of ‘p’, where ‘p’ is probability of node being in “possibly transmit” mode.

Scheduling of frames for timely transmission to support service differentiation in MANETs is difficult [1]. There have been various proposals for scheduling with QoS Support proposed for mobile ad hoc networks. Majority of them are timestamp based. The idea behind timestamp based mechanisms is to assign timestamp to the incoming packets. Based on this timestamp value, backoff value is calculated, which determines when the packet will be sent.

Luo et al., [93, 117, 118] have proposed mechanisms based on timestamp. All the mechanisms aim to support QoS guarantees and requires a flow graph (contending flow graph) to be generated first. In fact [93] is an extended version of [118]. In [93] 2-tier service model is proposed. The global fairness model assumes having a complete knowledge of the flows of the entire network, whereas local fairness assumes the knowledge of local contending flows. Both [93, 118] achieve spatial reuse by assigning backoff values. This backoff value is proportional to the number of contending flows. In [93] the backoff values also depend on whether global fairness or local fairness model is used. In [117] the spatial reuse it achieved by using graph coloring theory. Flows marked with same colors are transmitted simultaneously.

These timestamp based mechanisms suffer from disadvantages like complexity in building flow graphs, complete knowledge of topology, sorting of packets, and assigning timestamps.

Chao et al., [119] propose a credit-based scheduling mechanism in which they assign credits to the flows. High priority is assigned to flows which uses less bandwidth. The authors assume cluster architecture, with each cluster having scheduler assigning time slots to nodes, which in turn assign timeslots to flows. Chao et al. [120] enhance this work by comparing timestamp based and credit based mechanisms. They propose flow weight assignment to timestamp based mechanism and credit-based slot allocation protocol (CSAP) to credit-based mechanism. They found that CSAP performs better than any of the other mechanisms. The major drawback of credit-based mechanism is the architecture itself, where a "scheduler" is assumed for each cluster, which makes the feasibility difficult.

Network Layer

In this section, we see works which mainly focus on mechanisms at Network layer. We cover QoS routing (constraint based routing) mechanisms and signalling mechanism.

Basic idea behind QoS routing is that given a QoS request (may be in terms of bandwidth, reliability, delay and jitter) of a flow, routing mechanism should return a route that is most likely to be able to meet the requirements. This is one of the approaches of service differentiation followed in Internet. There have been many proposed solutions for QoS routing in MANETs [38–41]. The authors of [38] propose a self organizing routing structure called "core". The core node establishes a route that satisfies QoS requirements on behalf of other nodes. CEDAR is hierarchial and similar to OSPF, which is a routing protocol proposed for Internet. In [41], AODV routing protocol is modified to consider the bandwidth as each node while finding the route. Works by [39, 40] are similar where a source node has the information of available bandwidth to all destination or finds the available bandwidth to a particular destination, respectively. In summary, none of the constraint based routing approaches significantly differ from those proposed for Internet. Important point to consider is that, earlier works do not consider the inherent features of MANET and design the routing mechanism accordingly.

Further, we consider two important works by Columbia University's COMET group [142]. First work is state-based service differentiation using signalling (INSIGNIA), whereas second work is the stateless service differentiation framework (SWAN) [59].

The first and foremost signaling protocol proposed for MANETs is INSIGNIA [143,144], which is similar to Resource reSerVation Protocol (RSVP) [145], a signaling protocol for wired networks. INSIGNIA is an IP based QoS signaling protocol, designed specifically for MANETs. Goal of INSIGNIA is to support adaptive services, which aim to provide minimum bandwidth assurances to real-time applications. The important idea is a strict separation of routing, signaling and forwarding. INSIGNIA supports fast reservation, restoration, end-to-end adaptation, service differentiation and distributed resource control. All the techniques are designed to determine an adaptive real-time service in MANETs environment. INSIGNIA is categorized into a class of in-band signaling protocols. It maintains flow state information for the realtime flows on an end-to-end basis. INSIGNIA is just a signaling protocol and there is a necessity to involve a routing protocol such as ABR and DSR, which track changes in ad hoc topology and make updates to routing tables.

With its unique features of in-band signaling and soft-state approach, INSIGNIA is the widely accepted signaling protocol. There has been a study [144], which evaluates the performance of INSIGNIA with routing protocols such as Ad hoc On-demand Distance Vector (AODV), DSR and Temporally Ordered Routing Algorithm (TORA). This study also indicates an improvement in TCP/UDP performance with INSIGNIA.

Stateless Wireless Ad hoc Networks (SWAN) is a stateless network model, which uses distributed control algorithms to achieve service differentiation in MANETs [59]. SWAN model includes a number of components like classifier, shaper, and admission controller to support flow admission of real-time traffic rate regulation of best-effort traffic. Detailed discussion about all the components is available in the literature [59].

SWAN supports per-hop and end-to-end control algorithms that primarily rely on the efficient operation of TCP/IP protocols. SWAN uses rate control for UDP and TCP best-effort traffic, and source-based admission control for UDP real-time traffic. It uses explicit congestion control to regulate admitted real-time traffic. The interesting part is that SWAN is designed to support real-time services over best-effort MACs, without the need to install and maintain costly QoS states at MANET nodes. The authors have presented a detailed simulation and performance analysis work on SWAN in [59].

Both MAC and Network Layers

In this section, we consider works that propose a framework or models covering both network and MAC layers.

The first proposal on QoS model for MANETs is Flexible QoS Model for MANETs (FQMM) [146]. FQMM defines three types of nodes, namely the ingress, core and egress, with single host playing multiple roles, similar to DiffServ, and the difference being, the type has nothing to do with physical location in the network. Node is ingress, if it is transmitting data; core, if it is forwarding; and egress, if it is receiving. The basic idea behind the model is that it uses both the per flow state property of IntServ [147] and class differentiation of DiffServ [148]. This idea can be viewed as hybrid provisioning: per-flow and per-class scheme, in which higher priority flows take per-flow and lower priority ones are handled per-class.

Another important feature of FQMM is adaptive traffic conditioning, which polices the traffic according to the traffic profile. Traffic profiling in FQMM is defined as a relative percentage of the effective link capacity. Bandwidth allocation is used as the relative service differentiation parameter. This model is based on the assumption that not all packets in network are actually seeking for highest priority, if not, this model would result in a model similar to IntServ. FQMM has its own limitations and problems, among them the important ones are the implementation and scalability problems.

Sun et al., [149] propose to achieve service differentiation by using a system approach that involves coordinated changes at the MAC and IP layers. At the MAC layer they propose priority based scheduling mechanism to provide service differentiation. Their queue structure is similar to IEEE 802.11e EDCF traffic classes. In IEEE 802.11e traffic classes are differentiated by assigning different congestion window values. In their work, they propose to combine the collision rate (number of retransmissions) with the backoff scheme (termed as adaptive backoff scheme). By doing this, different traffic with different priority levels will have different backoff behaviors when collisions occur. Low priority traffic will backoff for longer after collision occurs compared to high priority traffic.

The authors further propose a delay model based on the adaptive backoff scheme. This delay model is used to estimate the available bandwidth, which is used by the admission control protocol at the Network layer. Therefore combining the mechanisms at IP and MAC layer they achieve service differentiation, which was found to provide bounded latency and low jitter for real-time traffic.

5.3.2 Cross-layer Design Architectures

There have been various works which uses cross layer design for wireless networks [123, 150–152]. Madueno et al., [153] propose a broadcast protocol based on physical-MAC cross layer design. They exploit the signal-separation principles of physical layer and aim to provide higher capacity to exchange information among neighbors. The medium access scheme considered is time slotted. Pham et al, [154] propose a joint physical-MAC layer cross design approach to improve performance at MAC layer for ad hoc networks. In their work, they rely on a method of predicting the "future state" of channel under Rayleigh fading, and also based on the history of signal strength measurements. Based on this prediction (Good/Bad), MAC layers decides to carry out the transmission.

Mung Chian, [155] study the joint power and congestion control cross layer design. The work proposes a distributive power control algorithm that combines with TCP congestion control to improve on end-to-end throughput and energy efficiency. Author shows that a simple utilization in the physical layer of the buffer occupancy should be enough to achieve the joint optimum of the design. Author also shows that the coupled system converges to the global optimum of joint power and congestion control.

In [156], design aspects associated with providing multimedia (video/voice) service over wireless networks is studied. They focus on enhancing the transmission of video over wireless through an adaptive scheme, which involves interaction across multiple protocol layers. There are also proposals [157] which focus on providing QoS over wireless networks, and making use of cross-layer design.

5.4 Unified Service Differentiation Solution Architecture

In this section, we will describe the architecture of the UNIFIED service differentiation solution. Our architecture also follows the layered approach along with some cross-layer designs. We will describe UNIFIED solution architecture considering both the control flow and the data flow.

5.4.1 Control Flow

The major parts of the UNIFIED solution architecture are shown in Figure 5.3. Figure 5.4 shows the data flow in our architecture. In the control flow (Figure 5.3), four layers - adaptive application, Network, MAC, and Physical are shown along with component entities.

An important reason that makes the architecture design challenging for wireless networks in general and ad hoc network in particular is that the information which actually belong to one particular layer needs to be exchanged between different functionalities at different layers. In this regard, in our work we broadly classify cross-layer interactions into two categories. In the first category, we have a separate cross-layer component which consists of services that are used by entities at multiple layers. In the second category, the services provided by an entity at one layer is not just accessed by neighboring layer entities but also by entities which are across multiple layers. In our architecture the first category cross-layer component is termed as Neighbor Management Mechanism, which is shown vertically across Network and MAC layers. Whereas, the second category cross-layer interaction, which is between admission control and physical layer, is shown as a directed arrow interaction.

The *neighbor table* is the main repository of the cross-layer component (neighbor management mechanism). This enables the cooperation between Network and MAC layers. An entry in the neighbor table usually consists of the address of the neighbor, link quality(lifetime), and also its status (active sender). Apart from neighbor table, neighbor management mechanism also maintains a history of link lifetimes which are used to estimate the residual link lifetime. The scheduling mechanism uses the information of link quality and lifetimes to make the scheduling decision. It can be seen that both the routing and scheduling are dependent on the parameters at neighbor management mehanism, and are accessed at a different time scales. When a neighbor expires, the neighbor management scheme informs about the same to both routing at Network layer and scheduling at MAC layer.

The interaction between the admission control and the Physical layer is to probe the wireless radio states to measure BSB and BSI (as described in the admission control chapter) and also to obtain the noise value measurements. Therefore, both the Physical layer and the Network layer uses these common set of parameters, in a periodical fashion, and independent of each other.

When multiple components are combined, it is important to study for any unintended consequences. In this regard, we consider two important aspects: stability and fairness, and try to understand the interactions in detail. These two aspects are considered because they are addressed in more than one components. Link stability is addressed in both route-computation mechanism (SHARC) and scheduling mechanism (CaSMA). In SHARC, routes with highest stability value (longest residual lifetime) are chosen for every flow. When all flows at a node contend for the resources, from this set of most-stable routes, routes whose lifetime is shorter is chosen at CaSMA. Both SHARC and CaSMA solve different problems at different scales, and do not contradict each-other. As a route-computation mechanism SHARC's goal is to choose most stable routes, whereas, as scheduling mechanism, and to be fair with all the flows within a node, CaSMA chooses least stable from the set of most-stable flows. By combining these two components the consequences are not unwanted, on the contrary, the two mechanisms are complementary to each-other.

In UNIFIED, two fairness notions exists each at admission control scheme (iCAC) and scheduling scheme (CaSMA). Though both address similar problem, the important difference between the notions is that they address it at two different planes. In iCAC, fairness is achieved among the contending flows, which are at different nodes within a "contention region". Whereas, in CaSMA the fairness is achieved among the contending flows, which are within are node. Therefore, when they combine, UNIFIED solution aims to achieve fairness at both the levels - contending region and contending node.

From the service differentiation functionality point of view, routing mechanism chooses the most stable route for both best-effort and real-time flows. Whereas, admission control scheme is applicable only to real-time flows. In the scheduling mechanism, higher priority is given for real-time flows compared to best-effort traffic. The three mechanisms of the UNIFIED solution provide better performance for real-time traffic, and there does not exist any specific mechanism to handle best-effort traffic. To provide a better comparative study of the mechanism, with the existing service differentiation proposals, we add a traffic shaper, which is similar to the SWAN [59] proposal, and which regulates the transmission rate of best-effort traffic. The shaper, however, is comparatively simple and uses the packet-loss as the feedback. Whereas, in SWAN the shaper is proactive and robust, and uses MAC delays as feedback for regulation. The shaper functions in additive increase and multiplicative decrease fashion, and [59] explains this operation in detail.

5.4.2 Data Flow

Considering the data flow diagram as shown in Figure 5.4, packets first arrive at the routing protocol from the upper layers. If the routing protocol does not have route to the destination

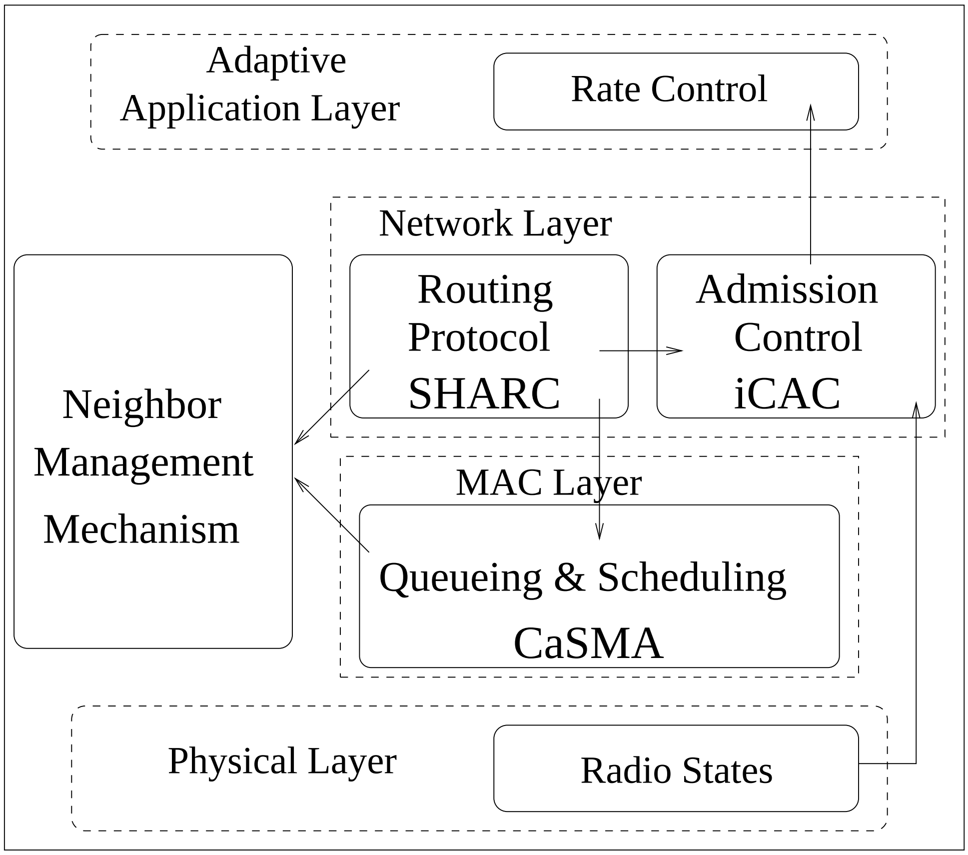


Figure 5.3: UNIFIED solution architecture control flow

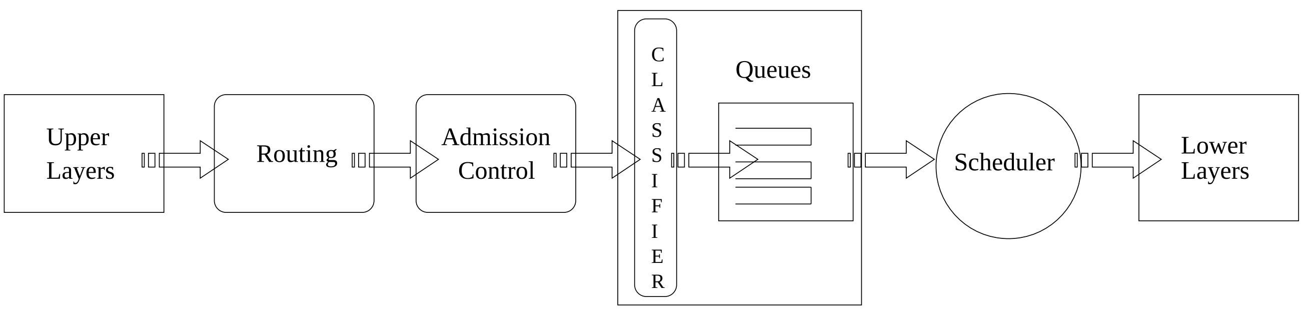


Figure 5.4: UNIFIED architecture data flow

to which the packet is destined, it initiates the route request message to find the route. In our architecture, routing protocol uses the SHARC route-computation approach to determine the best route. Once the route is known, and if the packet belongs to real-time flow, it has to pass through the admission control scheme. Best-effort traffic packets do not pass through admission control scheme, and they are directly sent to classifier. Classifier classifies the packets based on the destination ids. Admission control scheme uses the iCAC algorithm to determine if it can admit the real-time packet. Once, the admission control scheme admits, it is put into a proper per-destination queue by the classifier. Best-effort packets arrive at their respective queues after been regulated by the shaper. Further, packets from these set of per-destination queues are chosen according the CaSMA algorithm. All end-to-end measurements(lifetime, bandwidth and noise) are carried out using the routing protocol.

In the next section, we proceed with describing few implementation details. Further, we will provide the evaluation results of UNIFIED solution.

5.4.3 Implementation

Recall from the previous chapters that route computation mechanism (SHARC) and scheduling mechanism (CaSMA) were both implemented on NS-2, whereas admission control scheme iCAC was implemented on Glomosim. The main reason is that the wireless channel model implementation in NS-2 is too simplistic and not easily modifiable. Further, in our admission control scheme we depend heavily on the wireless radio states (in NS-2, only one state "Idle" is properly maintained), and this disadvantage with NS-2 left us to use Glomosim. Therefore, we implemented all the mechanisms of the UNIFIED solution on Glomosim. As simulators like Glomosim are less frequently used, comparison of the results with work of other researchers becomes difficult.

We consider DSR routing protocol and IEEE 802.11 MAC protocol in our work. The neighbor management mechanism and admission control mechanism are developed as separate modules and added as an extension to the DSR protocol. Neighbor management also gets information about the link status from the MAC protocol. We maintain per-destination multiple interface queues. Queues are added/removed as and when new flows are added/removed. For support of multihop scenarios, we have three different end-to-end measurement process, path lifetimes (SHARC), end-to-end available bandwidth (iCAC), schedulable list (CaSMA), which are added as separate modules of DSR.

5.4.4 Configurable Parameters

In our design, there are 5 parameters which are configurable. The neighbor management mechanism, which is the cross layer component, has 3 parameters: beacon interval, maximum interval without beacons, and neighbor flushing period. In neighbor management mechanism, beacons (small HELLO messages) are sent periodically. This interval duration (*beacon interval*) is made configurable, and in our implementation we make it a random value varying between 2 - 4 secs. Whenever a new neighbor is added or a beacon is received from existing neighbor, the expiry period in updated. This period of expiry is set as current time plus the *maximum interval without beacons*. Further, the neighbor information maintained in the neighbor table needs to be periodically cleared of stale information (remove expired neighbors). We carry out this flushing operation every *neighbor flushing period*.

The admission control scheme iCAC has a configurable parameter: *the minimum required rate for a flow*. This parameter decides when to completely block a flow. If the available bandwidth is less than this minimum required rate, then the flow will be blocked. In our implementation we set this value as 100 kbps.

Our final configurable parameter is the *maximum number of flows at a node*. This parameter is used by the scheduling mechanism CaSMA. In CaSMA, per-destination queues are maintained. In our implementation, we do not change the number of queues dynamically, but the number of queues is pre-defined depending on this configurable parameter. Other parameters used in available bandwidth measurements such as noise threshold are considered based on separate experimental studies.

5.4.5 Evaluation

In this section, we will provide the evaluation results of UNIFIED solution. We present the study of UNIFIED with the existing service differentiation mechanism Stateless Wireless Adhoc Network (SWAN) [59]. As described in Chapter 2 SWAN model [59] uses distributed algorithms to deliver service differentiation in MANETs. We compare and summarize the features of SWAN and UNIFIED in the Table 5.1.

Simulation Environment

Each mobile host has a transmission range of 250 meters and shares an 11 Mbps radio channel with its neighbors. The simulation includes a two-ray ground reflection model and IEEE 802.11 MAC protocol. All the simulations are run for 1000 seconds. A multihop network with 50 mobile nodes is simulated. The network area has a square shape of 1000 m x 1000 m. The mobility pattern used is the random waypoint model. Each node selects a random destination and moves with a random speed up to a maximum speed of 1 m/s to 20 m/s (72 km/hr), pausing for a random period up to a maximum duration of 50 secs, when the destination is reached. Average end-to-end packet delay, average throughput, percentage of packet losses, routing overhead and number of routes chosen are the performance metrics that are considered. The number of routes chosen parameter indicates count of routes chosen by a node, which is dependent on the number of flows that particular node carries and also any optimizations the routing mechanism incorporates. All experiments are carried out with 9 replications.

Effect of Varying Real Time Traffic

In this part of the evaluation, we consider 5 FTP background flows, and vary the number of CBR sources.

Figures 5.5(a) and 5.5(b) show the average delay and aggregate percentage of losses for real-time traffic, respectively. Whereas, Figure 5.5(c) shows the average throughput of TCP best-effort traffic for increasing amount of CBR traffic, with a fixed amount of best-effort traffic (5 TCP flows). The maximum node speed is set to 5 m/s. and pause time varies from 0 to 50 secs. With varying number of CBR flows, UNIFIED provides better delays than SWAN, and UNIFIED provides lesser percentage of losses than SWAN. The delay variation with increasing CBR traffic in case of UNIFIED is also less compared to SWAN.

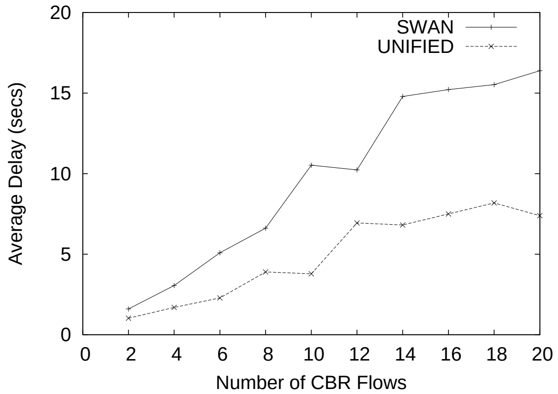
The variation of average throughput of best-effort traffic with varying number of CBR flows for both UNIFIED and SWAN are almost the same. SWAN has better throughput for lesser CBR flows, which is attributed to the better rate-control scheme used for best-effort traffic. As the number of CBR flows increases, however, UNIFIED has better throughputs compared to SWAN.

Figures 5.5(d) and 5.5(e) show the routing overhead and number of routes chosen for both real-time and best-effort traffic, respectively, for increasing amount of real-time traffic, with a fixed amount of best-effort traffic (5 TCP flows). Interesting points to be noted from these graphs are that UNIFIED, as expected, has higher number of control (routing) packets, and greater number of routes chosen compared to SWAN. This is mainly because UNIFIED includes hello-packets and probe-packets(in iCAC). The difference is still lesser because, there is an increase in the number of control packets for SWAN, which is due to many route setups and route error exchanges. There is an increase in the number of routes chosen because, in UNIFIED, DSR tries to find routes, which are both shortest and stable. In this process, it tends to pick different routes when available.

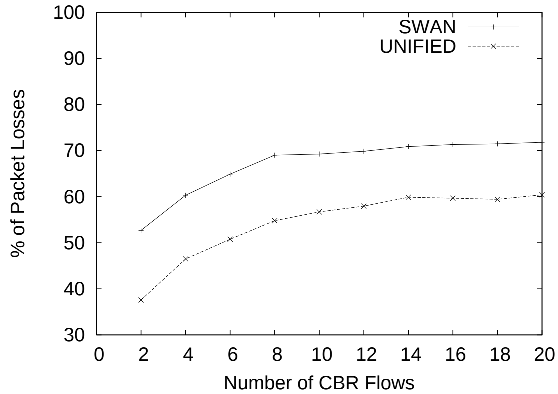
Figure 5.6 shows the packet delivery ratio for every flows, considering 20 flows case. This plot shows that UNIFIED tends to provide both higher packet delivery ratio, and better minimum packet delivery for each flow. This can be attributed to the fairness policy incorporated at both admission control and scheduling for UNIFIED.

Effect of Varying Speed

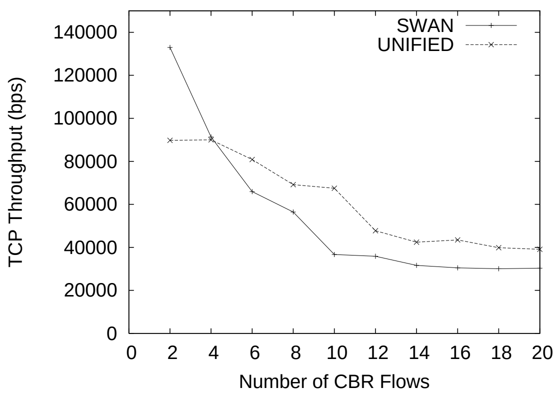
In this part of evaluation, we consider 5 FTP background flows and 8 CBR flows, with varying maximum node mobility speed.



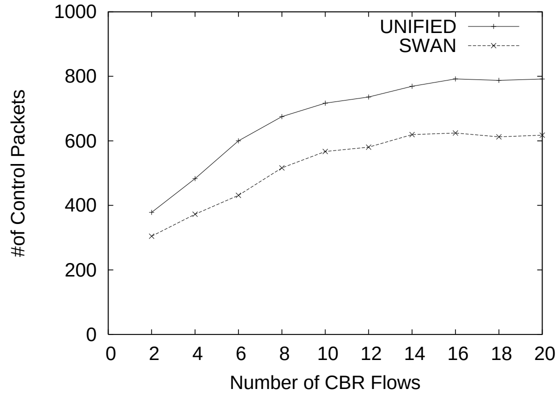
(a) Average delay of real-time traffic



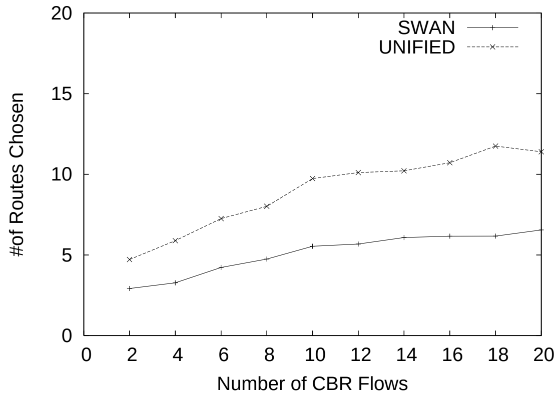
(b) Percentage of packet losses of real-time traffic



(c) Average throughput of best-effort TCP traffic



(d) Routing Overhead



(e) Number of routes chosen

Figure 5.5: Effect of varying real-time traffic

Figures 5.7(a) and 5.7(b) show the average delay and aggregate percentage of losses for real-time traffic. Whereas, Figure 5.7(c) shows the average throughput of TCP best-effort traffic for increasing , with a fixed amount of best-effort and real-time traffic (8 CBR and 5 TCP flows). The maximum node speed is varied from 1 m/s to 20 m/s. and pause time

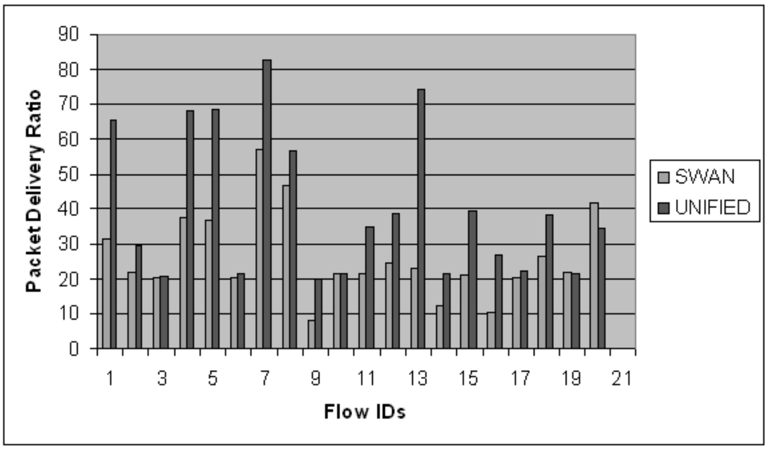
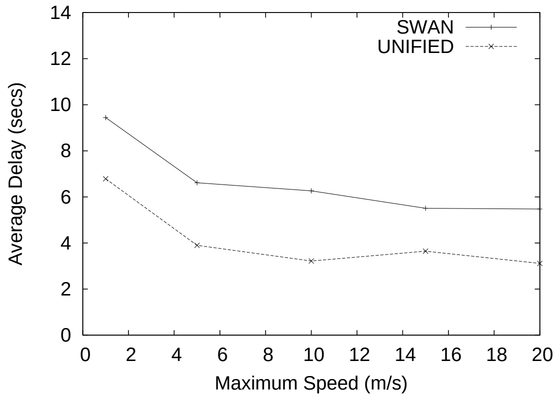


Figure 5.6: Percentage of share each flow gets

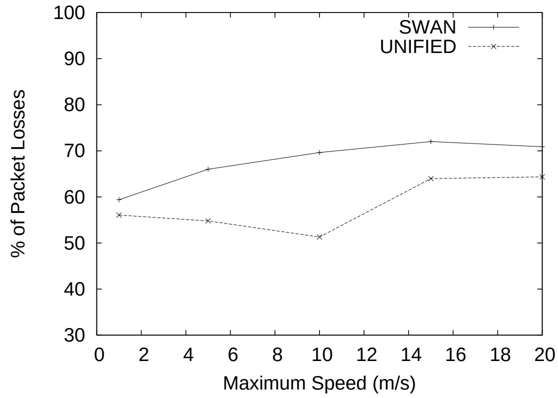
varies from 0 to 50 secs. With varying node speed, UNIFIED provides better delays than SWAN, and UNIFIED provides lesser percentage of losses than SWAN. The delay variation with increasing speed in the case of both UNIFIED and SWAN is similar, it decreases with increasing speed. This decrease in average delay is because the goodput of CBR traffic decreases with increasing speed.

The variation of average throughput of best-effort traffic with increasing speed for both UNIFIED and SWAN are almost the same. As the speed increases, UNIFIED has better throughputs compared to SWAN.

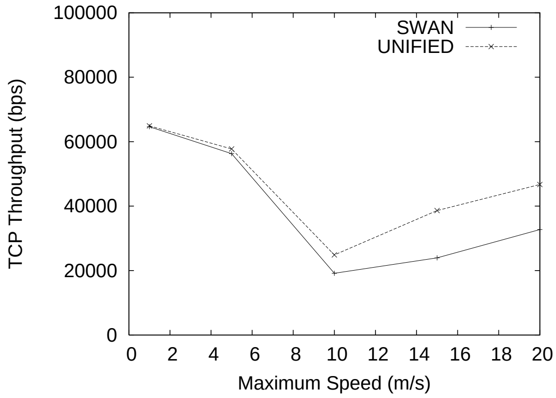
Figures 5.7(d) and 5.7(e) show the routing overhead and number of routes chosen for both real-time and best-effort traffic, respectively, for increasing maximum node speed, with a fixed amount of best-effort and real-time traffic (5 TCP and 8 CBR flows). Similar to the CBR traffic variations plot, UNIFIED has slightly higher overhead, and greater number of routes chosen compared to SWAN.



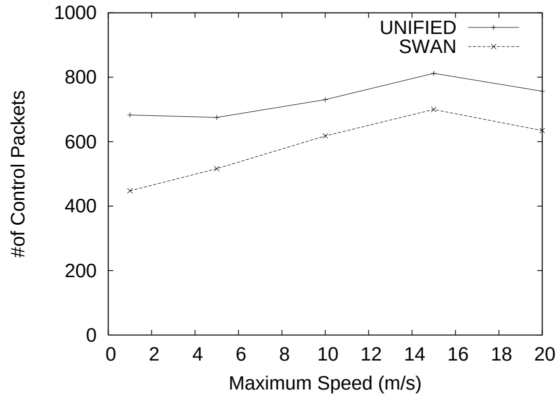
(a) Average delay of real-time traffic



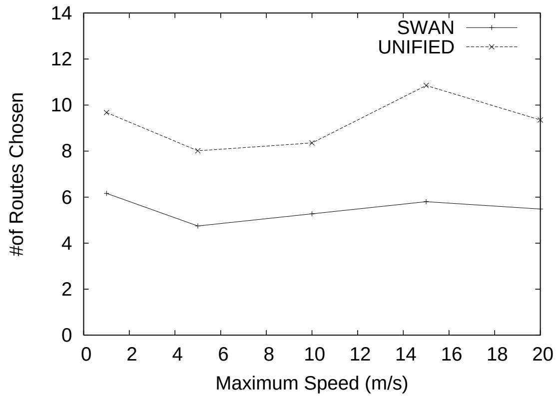
(b) Percentage of packet losses of real-time traffic



(c) Average throughput of best-effort TCP traffic



(d) Routing overhead



(e) Number of routes chosen

Figure 5.7: Effect of varying node maximum speed

Discussion

The reasons behind various advantages/disadvantages of SWAN and UNIFIED can be better understood considering their differences in approach as summarized in table 5.1. In this section, we briefly list the important differences.

In UNIFIED, routes chosen are based on lifetimes, and priority is given for most stable routes. Among these set of stable routes, at scheduling, priorities are given to those flows whose lifetime is lesser. This combined approach helps in increasing the packet deliveries (in turn decreases the packet losses) and improves the fairness. Whereas for SWAN, the routing is only based on hop-count, and scheduling is based on FIFO. Further, a novel and adaptive admission control mechanism of UNIFIED have better estimation of available bandwidth compared to the available bandwidth mechanism of SWAN. This difference in admission control scheme, also attributes to various performance advantages of UNIFIED solution as described above.

We would like to highlight that an adaptive rate-control scheme of SWAN independent of routing and scheduling, may not prove advantageous. In addition, a weak bandwidth measure of SWAN also proves to be insufficient. Further, better rate-control for best-effort traffic scheme of SWAN helps when the network conditions are not overloaded and the topology changes are also lesser. As the number of flows increases or topology changes become frequent, however, only rate-control scheme may not suffice.

Table 5.1: SWAN versus UNIFIED

*	SWAN versus UNIFIED	*
Category	SWAN	UNIFIED
Mechanisms	Rate Control, Admission Control, Shaper (focus on Rate Control)	Routing, Admission control and Scheduling
Routing	Hop-count	stability and hop-count
Admission Control		
Probing	End-to-End	End-to-End
Bandwidth Measurement	Instantaneous rate and Conservative rate	iCAC Algorithm
Neighbor Transmission	NO	YES
Decision Location	Source	Source
MAC		
Packet Scheduling	FIFO	Channel and Congestion aware
MAC Protocol	IEEE 802.11	IEEE 802.11
Feedbacks		
Source	Network, MAC	Network, MAC, Neighboring Nodes
Packet Marking	Yes (Avoid shaping for RT, Connection Re-establishment)	No
Rate Control		
Real Time	NO (connection is re-established)	Explicitly Specified
Best Effort	AIMD (Feedback: MAC delays)	AIMD (Feedback: Packet Losses)
Mobility and Multihop Support		
Transmission of packets of same flow at neighboring nodes	Not Considered	Considered
Priority based on local and End-to-End	Not Considered	Considered
Bandwidth Re-computation due to change in topology	Weak	Strong

Table 5.1: (continued)

*	SWAN versus UNIFIED	*
Packet Differentiation	Best Effort: Shaper Real Time: No Shaping, Packet Marking	Best Effort: Shaper Real Time: Higher Priority in Scheduling
Overhead	Lesser	higher
Congestion Control	Estimation and Notification	Estimation, Prevention and Notify
Complexity	Simple	Average
Approach	Stateless	Stateless

5.5 Summary

In this chapter, we presented UNIFIED service differentiation solution combining the mechanisms described in the previous chapters. We presented a cross-layer design architecture for UNIFIED solution. We carried out a comparative performance evaluation of UNIFIED with SWAN. We showed the importance of considering link-stability based route-computation scheme and channel-quality based scheduling scheme for service differentiation in MANETs. Further, we also showed that it is important to have an admission control scheme whose bandwidth measurement scheme is robust and adaptive to the MANET environments.

Chapter 6

Conclusions and Future Directions

In this work, we develop individual resource management schemes and a service differentiation architecture combining the schemes for mobile ad hoc networks to achieve efficient utilization of scarce available channel bandwidth.

We have developed the mechanisms focusing on the challenges and inherent aspects of mobile ad hoc networks. In particular, we focus on the features of ad hoc networks such as shared wireless medium, multihop, node mobility, and time varying channel quality in developing routing (SHARC), admission control (iCAC) and packet scheduling schemes (CaSMA). We carried out detailed study on important inherent features such as node mobility and its effects on wireless link characteristics, interference and its effects channel bandwidth measurements. Link lifetime, one of the characteristics of wireless link was analyzed following the approach used in reliability engineering studies. Link lifetime studies revealed that earlier assumptions such as, longer the two nodes have remained as neighbors, the probability that the two nodes continue to remain as neighbors for longer time is high, may not apply for all the mobility patterns. In some cases it may be the opposite. Besides, link lifetime distribution models are different for different mobility patterns, and exponential model (as considered by majority of previous works) is not a suitable fit for all the mobility patterns. Further, link failures are never random, and for majority of the mobility patterns link failures are similar to “wear-out” failures. In addition, it is difficult to have an accurate measure of the residual link lifetime, and heuristics based estimation of link lifetimes perform considerably better across various mobility patterns.

Based on the link lifetime studies, we develop a route computation mechanism termed as Stability and Hop-count based Approach for Route Computation (SHARC), and a packet scheduling scheme termed as Channel-aware Scheduling for MANETs (CaSMA). SHARC can be built into existing routing protocols, and which considers the link quality (represented as residual lifetime) as a metric, designed for ad hoc network environments. CaSMA considers end-to-end channel conditions in making the scheduling decisions. For efficient resource allocation, we found that it is advantageous to consider the end-to-end channel quality along-with local channel quality while making the scheduling decisions. Combining both link lifetime and congestion level helps in modeling the end-to-end channel conditions effectively.

We develop a novel call admission control scheme termed as interference based Call Admission Control (iCAC), which relies on estimating the positions of interfering nodes, and adheres to a fairness notion of equal-and-fair share. For position estimation, we exploit the wireless radio antenna states and noise measurements.

Finally, we combine above three schemes into single service differentiation solution, termed as UNIFIED, to evaluate the combined performance, to demonstrate the flexibility of the schemes and to have comparative study with the existing proposals.

In summary, our findings show that it is important to develop mechanisms specifically for MANETs focusing mainly on the challenges and inherent features of MANETs. Such mechanisms either used individually or combined into a resource management solution perform better across various scenarios.

6.1 Future Directions

In this section, we describe the future research directions considering the three individual components (routing, admission control and scheduling) described in the previous chapters.

Routing

We propose to further the study of routing mechanics, which impacts the performance of stability-based routing. This could be a challenging work, as till date, there is no work which describes what an ideal stability-based routing protocol should include. It is not difficult to see that estimation process can be improved in different ways, one example could be dual estimation- both the nodes of the link (two nodes which form the link) to estimate the residual lifetime, and choose the appropriate. What is "appropriate" is also not clear, however. We believe that this approach would be accurate and might eliminate some wrong estimations.

Another important question, answer for which might take careful and detailed study, is "min Hop or min Hop+1?". Let us consider the scenario as shown in Figure 6.1. It is a thought-provoking question whether a stability based routing should choose $\{a, b, c\}$ or $\{a, d, e, c\}$. Further, what if "b" is the weakest link (shortest residual lifetime)? Therefore, it would be a challenging future work to make SHARC find such situations and take decisions.

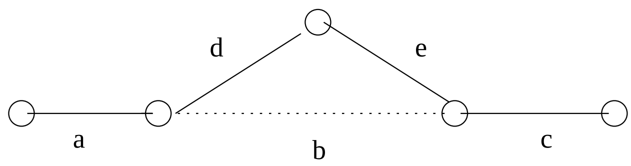


Figure 6.1: Minhop or minhop+1?

Finally, if we have many accurate estimations of the link, then we can provide some sort of guarantees, such as "route that can last for "t" seconds". There are applications which expects routing protocol to provide a route with bounds on delay and jitter. There might also be applications which require bounds on lifetime of the route. These are applications, if interrupted would have to start from first. This interruption would result in wastage of already spent resources like node energy and bandwidth. Therefore, for such applications, it would be a significant work to study if SHARC can provide a solution.

Admission Control

Enhancing the fairness aspect of iCAC could be an interesting future work. Fairness model proposed in [109] could be enhanced to consider interference due to flows outside the transmission range. In addition, study of fairness achieved by iCAC in multihop scenarios would be a challenging future work. Apart from the study of fairness, another important future work would be to study the performance iCAC across different mobility scenarios, to understand how mobility affects the measurements involved in iCAC.

Packet Scheduling

Apart from reducing the limitations of CaSMA, we identify two areas, which require a detailed study: study of impact of link lifetime estimation error. Our initial studies have shown that estimation error affects the "schedulable-region". Detailed study is required, which will be part of our future work. There are various flow-level fairness goals (both per-hop and end-to-end) in CaSMA, which are yet to be achieved. The other area, which is yet to be explored as part of our future work is the latest starting time (waiting time) at any node for any packet belonging to any flow f_i. We propose to follow the work of Martin et al., [158] in this regard.

Variety of mobility models have been proposed for ad hoc networks [44]. Pei et.al., [159] mentions that realistic models are a necessity to model the mobility patterns, while carrying out simulations, in order to effectively capture the protocol performance. As a supplement to our work, we have developed a realistic mobility pattern generator (*EGRESS- Environment of Generation of REalistic Scenarios for Simulations*) [160]. Therefore, a considerable future work would be to evaluate our protocols in realistic scenarios generated by EGRESS.

Related Papers

- K. N. Sridhar and L. Jacob, "Interplay of service differentiation and routing protocols: A performance study," in *Proceedings of IEEE International Phoenix Computing, and Communications Conference (IPCCC) 2004*, Phoenix, Arizona, October 2004, pp. 40–46.

- K. N. Sridhar, L. Jacob, and R. Shorey, "Performance evaluation and enhancement of link stability based routing for MANETs," in *Proceedings of IEEE International Conference on Parallel Processing (ICPP) Workshops 2004*, Montreal, Quebec, Canada, August 2004, pp. 133–140.

- K. N. Sridhar and L. Jacob, "Performance evaluation and enhancement of link stability based routing protocol for MANETs," *International Journal of High Performance Computing and Networking*, vol. 4, no. 1-2, pp. 66–77, 2005.

- K. N. Sridhar and M. C. Chan, "Stability and hop-count based approach for route computation in MANET," in *Proceedings of IEEE International Conference on Computer Communications and Networks (ICCCN) 2005*, San Diego, California, October 2005, pp. 25–31.

- K. N. Sridhar and M. C. Chan, "Interference based call admission control for wireless ad hoc networks," in *Proceedings of IEEE/ACM MOBIQUITOUS*, San Jose, California, July 2006, pp. 424 – 432.

- K. N. Sridhar and S. Hao and M. C. Chan, A. L. Ananda "EGRESS: Environment for Generating REalistic Scenarios for Simulations," in *Proceedings of Tenth (IEEE) International Symposium on Distributed Simulation and Real-Time Applications, (DS-RT06)*, Malaga, Spain, October 2006, pp. 15–24.

- K. N. Sridhar and M. C. Chan, "Link lifetimes: distribution, estimation and applications," *In preparation.*

Bibliography

[1] C.-K. Toh, *Ad Hoc Wireless Networks: Protocols and Systems.* Upper Saddle River, NJ: Prentice Hall PTR, 2001.

[2] P. Mohapatra and S. Krishnamurthy, *Ad hoc Networks: Technologies and Protocols.* Boston, MA: Springer, 2004.

[3] C. S. R. Murthy and B. S. Manoj, *Ad Hoc Wireless Networks.* Upper Saddle River, NJ: Prentice Hall International, May 2004.

[4] J. Jubin and J. Tornow, “The DARPA Packet Radio Network Protocols,” *Proceedings of the IEEE*, vol. 75, no. 1, pp. 21–32, 1987.

[5] “IETF MANET working group,” http://www.ietf.org/html.charters/manet-charter.html.

[6] IEEE, “802.11 working group, wireless LAN medium access control (MAC) and physical layer (PHY) specifications,” 1999.

[7] A. Ahmad, *Wireless and Mobile Data Networks.* NY: Wiley-Interscience, John Wiley & Sons, 2005.

[8] F. Xue and P. R. Kumar, “Scaling laws for ad hoc wireless networks: an information theoretic approach,” *Foundations and Trends in Networking*, vol. 1, no. 2, pp. 145–270, 2006.

[9] A. Ephremides, “Ad hoc networks: not an ad hoc field anymore.” *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 441–448, 2002.

[10] M. Gerla, M. Kazantzidis, G. Pei, F. Talucci, and K. Tang, “Ad hoc, wireless, mobile networks: The role of performance modeling and evaluation,” in *Performance Evaluation*, 2000, pp. 51–65.

[11] F. K. and V. K., “NS notes and documentation,” The VINT Project, UC Berkely, LBL, USC/ISI, and Xerox PARC, Available from http://http://www.isi.edu/nsnam/vint/, 1997.

[12] X. Zeng, R. Bagrodia, and M. Gerla, “GloMoSim: A library for parallel simulation of large-scale wireless networks.” in *Workshop on Parallel and Distributed Simulation*, Banff, Alberta, Canada, May 1998, pp. 154–161.

[13] R. Hekmat, *Ad-hoc Networks: Fundamental Properties and Network Topologies.* Boston, MA: Springer, November.

[14] C. E. Perkins, S. R. Alpert, and B. Woolf, *Mobile IP; Design Principles and Practices.* Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1997.

[15] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 132–144, 2006.

[16] C. E. Perkins and E. M. Belding-Royer, "Ad-hoc on-demand distance vector (AODV) routing." in *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications (WMCSA) 1999*, 1999, pp. 90–100.

[17] T. Clausen and P. Jacquet, "Optimized link state routing protocol (OLSR)," Published Online, Internet Engineering Task Force, RFC 3626, October 2003.

[18] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing DSDV for mobile computers." in *Proceedings of ACM Special Interest Group on Data Communications (SIGCOMM) 1994*, London, UK, 1994, pp. 234–244.

[19] C.-K. Toh, "A novel distributed routing protocol to support ad-hoc mobile computing," in *Proceedings of IEEE International Phoenix Computing, and Communications Conference (IPCCC) 1996*, Phoenix, Arizona, October 1996, pp. 480–486.

[20] R. Dube, C. D. Rais, K.-Y. Wang, and S. K. Tripathi, "Signal stability based adaptive routing (SSA) for ad-hoc mobile networks," University of Maryland at College Park, Institute for Advanced Computer Studies, Tech. Rep. UMIACS-TR-96-34, 1996.

[21] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks." in *Proceedings of The Conference on Computer Communications, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 1997*, Kobe, Japan, April 1997, pp. 1405–1413.

[22] Z. J. Haas, "A new routing protocol for the reconfigurable wireless networks," in *IEEE International Conference on Universal Personal Communications - ICUPC*, vol. 2. San Diego, CA: IEEE, October 1997, pp. 562–566.

[23] S. Murthy and J. J. Garcia-Luna-Aceves, "An efficient routing protocol for wireless networks." *ACM Baltzer Mobile Networks and Applications (MONET)*, vol. 1, no. 2, pp. 183–197, 1996.

[24] C. E. Perkins, *Ad hoc Networking.* Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 2001.

[25] D. A. Maltz, J. Broch, J. Jetcheva, and D. B. Johnson, "The effects of on-demand behavior in routing protocols for multi-hop wireless ad hoc networks," *IEEE Journal on Selected Areas in Communications, special issue on mobile and wireless networks*, vol. 17, no. 4, pp. 1439–1453, August 1999.

[26] V. Bharghavan, "Performance evaluation of algorithms for wireless medium access," in *IEEE International Computer Performance and Dependability Symposium*, vol. 00, Durham, North Carolina, September 1998, pp. 86–96.

[27] S. R. Das, C. E. Perkins, and E. M. Belding-Royer, "Performance comparison of two on-demand routing protocols for ad hoc networks," in *Proceedings of The Conference on Computer Communications, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) 2000*, Tel-Aviv, Israel, March 2000, pp. 3–12.

[28] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. G. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 1998*, Dallas, Texas, October 1998, pp. 85–97.

[29] S.-J. Lee, M. Gerla, and C.-K. Toh, "A simulation study of table-driven and on-demand routing protocols for mobile ad hoc networks," *IEEE Networks*, vol. 13, no. 4, pp. 48–54, July-August 1999.

[30] A. Nasipuri, R. Castañeda, and S. R. Das, "Performance of multipath routing for on-demand protocols in mobile ad hoc networks," *ACM Baltzer Mobile Networks and Applications (MONET)*, vol. 6, no. 4, pp. 339–349, 2001.

[31] E. M. Royer and C.-K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46–55, April 1999.

[32] C.-K. Toh, R. Chen, M. Delwar, and D. Allen, "Experimenting with an ad hoc wireless network on campus: insights and experiences," *SIGMETRICS Performance Evaluation Review*, vol. 28, no. 3, pp. 21–29, 2000.

[33] S. Giordano, "Mobile ad hoc networks," *Handbook of wireless networks and mobile computing New York, NY: John Wiley & Sons, Inc.*, pp. 325–346, 2002.

[34] D. B. Johnson, D. A. Maltz, and J. Broch, "DSR: the dynamic source routing protocol for multihop wireless ad hoc networks," *Ad hoc networking Boston, MA: Addison-Wesley Longman Publishing Co., Inc.*, pp. 139–172, 2001.

[35] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *mobicom03*. New York, NY, USA: ACM Press, 2003, pp. 134–146.

[36] C.-K. Toh, "Associativity-based routing for ad hoc mobile networks," *Wireless Personal Communications*, vol. 4, no. 2, pp. 103–139, 1997.

[37] S. Agarwal, A. Ahuja, J. P. Singh, and R. Shorey, "Route-lifetime assessment based routing RABR protocol for mobile ad-hoc networks," in *Proceedings of IEEE International Conference on Communications (ICC) 2000*, New Orleans, LA, 2000, pp. 1697–1701.

[38] P. Sinha, R. Sivakumar, and V. Bharghavan, "CEDAR: a core-extraction distributed ad hoc routing algorithm." in *Proceedings of The Conference on Computer Communications, Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 1999*, New York, NY, March 1999, pp. 202–209.

[39] S. Chen and K. Nahrstedt, "Distributed QoS routing with imprecise state information." in *Proceedings of IEEE International Conference on Computer Communications and Networks (ICCCN) 1998*, Los Angeles, California, October 1998, pp. 614–623.

[40] C. R. Lin, "An on-demand QoS routing protocol for mobile ad hoc networks." in *Proceedings of The Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2001*, Anchorage, Alaska, April 2001, pp. 1735–1744.

[41] S. R. D. Charles E. Perkins, Elizabeth M. Royer, "Quality of service for ad hoc on-demand distance vector routing," Published Online, Internet Engineering Task Force, Internet Draft draft-perkinsmanet-aodvqos-00.txt, November 2001.

[42] M. Mirhakkak, N. Schult, and D. Thomson, "Dynamic quality-of-service for mobile ad hoc networks," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC) 2000*. Piscataway, NJ: IEEE Press, 2000, pp. 137–148.

[43] S. Corson and J. Macker, "RFC 2501: Mobile ad hoc networking (manet): Routing protocol performance issues and evaluation considerations," January 1999.

[44] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 2003*, San Diego, California, September 2003, pp. 217–229.

[45] X. Perez-Costa, C. Bettstetter, and H. Hartenstein, "Toward a mobility metric for comparable & reproducible results in ad hoc networks research," *SIGMOBILE Mobile Computing and Communication Review*, vol. 7, no. 4, pp. 58–60, 2003.

[46] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proceedings of The Conference on Computer Communications, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2003*, San Francisco, California, April 2003, pp. 825–835.

[47] S. Cho and J. P. Hayes, "Impact of mobility on connection stability in ad hoc networks." in *Proceedings of IEEE Wireless Communications & Networking Conference (WCNC) 2005*, vol. 3, New Orleans, Louisiana, March 2005, pp. 1650–1656.

[48] V. Lenders, J. Wagner, and M. May, "Measurements from an 802.11b mobile ad hoc network," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) 2006*, Buffalo, New York, June 2006, pp. 519–524.

[49] J. Tsumochi, K. Masayama, H. Uehara, and M. Yokoyama, "Impact of mobility metric on routing protocols for mobile ad hoc networks," in *Proceedings of Pacific Rim Conference on Communications.* Taipei, Taiwan: IEEE, 2003, pp. 1000 – 1003.

[50] Y. Han, R. J. La, A. M. Makowski, and S. Lee, "Distribution of path durations in mobile ad-hoc networks: Palm's theorem to the rescue," *Computer Networks*, vol. 50, no. 12, pp. 1887–1900, 2006.

[51] S. Arbindi, K. Namuduri, and R. Pendse, "Statistical estimation of route expiry times in on-demand ad hoc routing protocols," in *Proceedings of Second IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) 2005*, Washington, DC, November 2005, pp. 467– 471.

[52] D. Turgut, S. K. Das, and M. Chatterjee, "Longevity of routes in mobile ad hoc networks." in *Proceedings of IEEE Vehicular Technology Conference (VTC) Spring 2001*, Rhodes, Greece, 2001, pp. 2833–2837.

[53] Y.-C. Tseng, Y.-F. Li, and Y.-C. Chang, "On route lifetime in multihop mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 366–376, October 2003.

[54] I. Gruber and H. Li, "Link expiration times in mobile ad hoc networks," in *Proceedings of IEEE Conference on Local Computer Networks (LCN) 2002*, Tampa, Florida, November 2002, pp. 743–750.

[55] M. Gerharz, C. de Waal, M. Frank, and P. Martini, "Link stability in mobile wireless ad hoc networks," in *Proceedings of IEEE Conference on Local Computer Networks (LCN) 2002*, Tampa, Florida, November 2002, pp. 30–42.

[56] Z. Cheng and W. B. Heinzelman, "Exploring long lifetime routing LLR in ad hoc networks," in *Proceedings of ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM) 2004*, Venice, Italy, October 2004, pp. 203–210.

[57] N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy, "PATHS: analysis of path duration statistics and their impact on reactive MANET routing protocols," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC) 2003*, Annapolis, Maryland, June 2003, pp. 245–256.

[58] Y.-C. Hu and D. B. Johnson, "Caching strategies in on-demand routing protocols for wireless ad hoc networks," in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 2000*, Boston, Massachusetts, August 2000, pp. 231–242.

[59] G.-S. Ahn, L.-H. Sun, A. Veres, and A. T. Campbell, "SWAN: Service differentiation in stateless wireless ad hoc networks," in *Proceedings of The Conference on Computer Communications, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2002*, pp. 127–141.

[60] J. L. Stanford and S. B. Vardeman, *Statistical methods for physical science.* Methods of experimental physics, San Diego CA: Academic Press, 1994.

[61] "e-Handbook of statistical methods," http://www.itl.nist.gov/div898/handbook/, Retrieved August 2006.

[62] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data.* NY, New York: John Wiley & Sons, Inc, 1998.

[63] L. L. Cam, "Maximum likelihood: An introduction," *International Statistical Review*, vol. 58, no. 2, pp. 153–171, August 1990.

[64] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002.

[65] "Toilers," Available from http://toilers.mines.edu.

[66] "Bonnmotion," Available from http://web.informatik.uni-bonn.de/IV/mitarbeiter/dewaal/bonnmotion/.

[67] "Important," Available from http://nile.usc.edu/important/.

[68] C. Quesenberry and J. Kent, "Selecting among probability distributions used in reliability," *American Statistical Association, Technometrics*, vol. 24, no. 1, pp. 59–65, Februaray 1982.

[69] A. W. Marshall, J. C. Meza, and I. Olkin, "Can data recognize its parent distribution?" *Journal of Computational and Graphical Statistics*, vol. 10, no. 3, pp. 555–582, September 2001.

[70] C. A. Drossos and A. N. Philippou, "A note on minimum distance estimates," *Annals of the Institute of Statistical Mathematics*, vol. 32, no. 1, pp. 121–123, December 1980.

[71] R. Mullen, "The lognormal distribution of software failure rates: Application to software reliability growth modeling," in *ISSRE '98: Proceedings of the The Ninth International Symposium on Software Reliability Engineering.* Washington, DC: IEEE Computer Society, 1998, pp. 134–143.

[72] "Easyfit 3.0," MathWave Technologies, Available from http://mathwave.com/downloads.html, 2005.

[73] W. Q. Meeker, L. A. Escobar, and C. J. Lu, "Accelerated degradation tests: Modeling and analysis," *American Statistical Association, Technometrics*, vol. 40, no. 2, pp. 89–99, Februaray 1998.

[74] E. L. Crow and K. Shimizu, *Lognormal Distributions Theory and Applications.* NY, New York: Drekker, 1988.

[75] K. N. Sridhar and L. Jacob, "Performance evaluation and enhancement of link stability based routing protocol for MANETs," *International Journal of High Performance Computing and Networking*, vol. 4, no. 1-2, pp. 66–77, 2005.

[76] "The CMU monarch project," The CMU Monarch Project's Wireless and Mobility Extensions to NS, August 1998. Available from http://www.monarch.cs.cmu.edu/.

[77] J.-H. Chang and L. Tassiulas, “Energy conserving routing in wireless ad-hoc networks,” in *Proceedings of The Conference on Computer Communications, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) 2000*, Tel-Aviv, Israel, March 2000, pp. 22–31.

[78] S. Singh, M. Woo, and C. S. Raghavendra, “Power-aware routing in mobile ad hoc networks,” in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 1998*, Dallas, Texas, October 1998, pp. 181–190.

[79] J.-C. Cano and P. Manzoni, “A performance comparison of energy consumption for mobile ad hoc network routing protocols,” in *Proceedings of IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS) 2000*, San Francisco, California, September 2000, pp. 57–64.

[80] F. Bai, N. Sadagopan, B. Krishnamachari, and A. Helmy, “Modeling path duration distributions in manets and their impact on reactive routing protocols.” *ACM Wireless Networks*, vol. 22, no. 7, pp. 1357–1373, 2004.

[81] J. F. Lawless, *Statistical models and methods for Lifetime Data.* NY, New York: John Wiley & Sons, Inc, 1982.

[82] G.-A. Klutke, P. C.Kiessler, and M.A.Wortman, “A critical look at the bathtub curve,” *IEEE Transactions on Reliability*, vol. 52, no. 3, pp. 125–129, 2003.

[83] “Bathtub curve image,” Available from http://commons.wikimedia.org/wiki/Image Bathtub_curve.jpg.

[84] J. Shengming, H. Dajiang, and R. Jianqiang, “A prediction-based link availability estimation for mobile ad hoc networks,” in *Proceedings of The Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2001*, Ancorage, Alaska, April 2001, pp. 1745–1752.

[85] W. Waloddi, “A statistical distribution function of wide applicability,” *Journal of Applied Mechanics*, vol. 18, pp. 293–297, 1951.

[86] J. Aitchison and J. Brown, *The Lognormal Distribution.* Cambridge, UK: Cambridge University Press, 1957.

[87] P. Stavroulakis, *Interference Analysis and Reduction for Wireless Systems.* Norwood, MA: Artech House, June 2003.

[88] Y. Sun, E. M. Belding-Royer, X. Gao, and J. Kempf, “A priority-based distributed call admission protocol for multi-hop wireless ad hoc networks,” Department of Computer Science, University of California, Santa Barbara, Technical Research Report 2004-20, 2004.

[89] V. Kawadia and P. R. Kumar, “A cautionary perspective on cross-layer design,” *IEEE Wireless Communications*, vol. 12, no. 1, pp. 3–11, February 2005.

[90] P. M. Ruiz and E. Garcia, "Improving user-perceived QoS in mobile and wireless IP networks using real-time adaptive multimedia applications," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2002*, vol. 3, September 2002.

[91] E. Huang, W. Hu, J. Crowcroft, and I. Wassell, "Towards commercial mobile ad hoc network applications: a radio dispatch system," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC) 2005*. Urbana-Champaign, IL: ACM Press, 2005, pp. 355–365.

[92] S. Xu and T. N. Saadawi, "Revealing the problems with 802.11 medium access control protocol in multi-hop wireless ad hoc networks," *Computer Networks*, vol. 38, no. 4, pp. 531–548, 2002.

[93] H. Luo, S. Lu, and V. Bharghavan, "A new model for packet scheduling in multihop wireless networks," in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 2000*, Boston, Massachusetts, August 2000, pp. 76–86.

[94] M. H. Ahmed, "Call admission control in. wireless networks: A comprehensive survey," *IEEE Communications Surveys*, vol. 7, no. 1, pp. 2–21, October 2005.

[95] K. Xu, K. Tang, R. Bagrodia, M. Gerla, and M. Bereschinsky, "Adaptive bandwidth management and QoS provisioning in large scale ad hoc networks," in *Proceedings of Military Communications Conference (MILCOM) 2003*, Boston, Massachusetts, October 2003, pp. 41–46.

[96] Y. Yang and R. Kravets, "Contention-aware admission control for ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 4, pp. 363–377, April 2005.

[97] I. D. Chakeres and E. M. Belding-Royer, "PAC: Perceptive admission control for mobile wireless networks." in *Proceedings of Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE) 2004*, Dallas, Texas, October 2004, pp. 18–26.

[98] S. H. Shah, K. Chen, and K. Nahrstedt, "Dynamic bandwidth management for single-hop ad hoc wireless networks." in *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom)2003*, Dallas, Texas, March 2003, pp. 195–204.

[99] I. Jawhar and J. Wu, "A race-free bandwidth reservation protocol for QoS routing in mobile ad hoc networks." in *Proceedings of Hawaii International Conference on System Sciences (HICSS) 2004*, vol. 3, September 2004, pp. 294–302.

[100] D. Gu and J. Zhang, "A new measurement-based admission control method for IEEE802.11 wireless local area networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2003*, vol. 3, Beijing, China, September 2003, pp. 2009–2013.

[101] J. Deng, B. Liang, and P. K. Varshney, "Tuning the carrier sensing range of IEEE 802.11 MAC," in *Global Telecommunications Conference*, vol. 5, Dallas, Texas, December 2004, pp. 2987–2991.

[102] M. Andersin, J. Zander, and Z. Rosberg, "Soft and safe admission control in cellular networks." *IEEE/ACM Transactions on Networks*, vol. 5, no. 2, pp. 255–265, April 1997.

[103] J. Kuri and P. Mermelstein, "Call admission on the uplink of a CDMA system based on total received power," in *Proceedings of IEEE International Conference on Communications (ICC) 1999*, vol. 3, Vancouver, BC, Canada, June 1999, pp. 1431–1436.

[104] P. Gupta and P. R. Kumar, "The capacity of wireless networks." *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[105] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks." in *Proceedings of The Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2001*, Anachorage, Alaska, April 2001, pp. 1360–1369.

[106] J. Li, C. Blake, D. S. J. D. Couto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks." in *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MOBICOM) 2001*, Rome, Italy, July 2001, pp. 61–69.

[107] R. Hekmat and P. V. Mieghem, "Interference in wireless multi-hop ad-hoc networks and its effect on network capacity." *ACM Wireless Networks*, vol. 10, no. 4, pp. 389–399, September 2004.

[108] D. Bertsekas and R. Gallager, *Data networks.* Upper Saddle River, NJ: Prentice-Hall, Inc., 1987.

[109] X. L. Huang and B. Bensaou, "On max-min fairness and scheduling in wireless adhoc networks: analytical framework and implementation." in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC) 2001*, Long Beach, California, October 2001, pp. 221–231.

[110] "The linux orinoco driver," Available From http://www.nongnu.org/orinoco/.

[111] P. Bhagwat, P. P. Bhattacharya, A. Krishna, and S. K. Tripathi, "Using channel state dependent packet scheduling to improve TCP throughput over wireless LANs," *ACM Wireless Networks*, vol. 3, no. 1, pp. 91–102, 1997.

[112] J. Gomez and A. T. Campbell, "Havana: Supporting application and channel dependent QoS in wireless packet networks," *ACM Wireless Networks*, vol. 9, no. 1, pp. 21–35, 2003.

[113] H. Aida, Y. Tamura, Y. Tobe, and H. Tokuda, "Wireless packet scheduling with signal-to-noise ratio monitoring," in *Proceedings of IEEE Conference on Local Computer Networks (LCN) 2000.* Washington, DC: IEEE Computer Society, November 2000, pp. 32–44.

[114] S. C. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Transactions on Networks*, vol. 13, no. 3, pp. 636–647, 2005.

[115] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. W. Knightly, "Distributed priority scheduling and medium access in ad hoc networks," *ACM Wireless Networks*, vol. 8, no. 5, pp. 455–466, 2002.

[116] B.-G. Chun and M. Baker, "Evaluation of packet scheduling algorithms in mobile ad hoc networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 6, no. 3, pp. 36–49, July 2002.

[117] H. Luo and S. Lu, "A topology-independent fair queueing model in ad hoc wireless networks," in *Proceedings of IEEE International Conference on Network Protocols (ICNP) 2000*, Osaka, Japan, November 2000, pp. 325–335.

[118] H. Luo, P. Medvedev, J. Cheng, and S. Lu, "A self-coordinating approach to distributed fair queueing in ad hoc wireless networks," in *Proceedings of The Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2001*, Anchorage, Alaska, April 2001, pp. 1370–1379.

[119] H.-L. Chao and W. Liao, "Credit-based fair scheduling in wireless ad hoc networks." in *Proceedings of IEEE Vehicular Technology Conference (VTC) Fall 2002*, Vancover, Canada, September 2002, pp. 1442–1446.

[120] H.-L. Chao, J.-C. Kuo, and W. Liao, "Fair scheduling with QoS support in ad hoc networks," in *Proceedings of IEEE Conference on Local Computer Networks (LCN) 2002*, Tampa, Florida, November 2002, pp. 502–507.

[121] X. Wu, C. Yuen, Y. Gao, H. Wu, and B. Li, "Fair scheduling with bottleneck consideration in wireless ad-hoc networks," in *Proceedings of IEEE International Conference on Computer Communications and Networks (ICCCN) 2001*, Phoenix, Arizona, October 2001, pp. 714–723.

[122] L. Bao and J. J. Garcia-Luna-Aceves, "Distributed dynamic channel access scheduling for ad hoc networks," *Journal of Parallel and Distributed Computing*, vol. 63, no. 1, pp. 3–14, 2003.

[123] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) 1999*, August 1999, pp. 35–42.

[124] M. Andrews, "Probabilistic end-to-end delay bounds for earliest deadline first scheduling," in *Proceedings of The Conference on Computer Communications, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) 2000*, Tel-Aviv, Israel, March 2000, pp. 603–612.

[125] K. N. Sridhar and C. M. Choon, "Stability and hop-count based approach for route computation in MANET," in *Proceedings of IEEE International Conference on Computer Communications and Networks (ICCCN) 2005*, San Diego, California, October 2005, pp. 25–31.

[126] A. Mcdonald and T. Znati, "A path availability model for wireless ad-hoc networks," in *Proceedings of IEEE Wireless Communications & Networking Conference (WCNC) 1999*, pp. 235–244.

[127] T. M. CHEN, J. WALRAND, and D. G. MESSERSCHMITT, "Dynamic priority protocols for packet voice," *IEEE Journal on Selected Areas in Communications, special issue on mobile and wireless networks*, vol. 7, no. 5, 1989.

[128] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," *Journal of ACM (JACM)*, vol. 20, no. 1, pp. 46–61, 1973.

[129] J. P. Lehoczky, L. Sha, and Y. Ding, "The rate monotonic scheduling algorithm: Exact characterization and average case behavior," in *Proceedings of the Tenth IEEE Real-Time Systems Symposium*, IEEE. Santa Monica, California: Computer Society Press, Dec 1989, pp. 166–171.

[130] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet Package.* Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 2000.

[131] D. D. Perkins and H. D. Hughes, "A survey on quality-of-service support for mobile ad hoc networks." *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 503–513, 2002.

[132] X. Xiao and L. M. Ni, "Internet QoS: a big picture," *Network, IEEE*, vol. 13, no. 5, pp. 8–18, 1999.

[133] V. Bharghavan, A. J. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in *Proceedings of ACM Special Interest Group on Data Communications (SIGCOMM) 1994*, London, UK, August 1994, pp. 212–225.

[134] P. Karn, "MACA - a new channel access method for packet radio," in *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, London, Ontario, Canada, September 1990, pp. 134–140.

[135] F. Talucci, M. Gerla, and L. Fratta, "MACA-BI (MACA by invitation): A receiver oriented access protocol for wireless multihop networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 1997*, Helsinki, Finland, September 1997, pp. 425–429.

[136] C.-K. Toh, V. Vassiliou, G. Guichal, and C.-H. Shih, "March: A medium access control protocol for multihop wireless ad hoc networks," in *Proceedings of Military Communications Conference (MILCOM) 2000*, Los Angeles, CA, October 2000, pp. 512–516.

[137] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11." in *Proceedings of The Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2001*, Anchorage, Alaska, April 2001, pp. 209–218.

[138] C. R. Lin and M. Gerla, "Asynchronous multimedia multihop wireless networks." in *Proceedings of The Conference on Computer Communications, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 1997*, Kobe, Japan, March 1997, pp. 118–125.

[139] J. Sobrinho and A. S. Krishnakumar, "Quality-of-service in ad hoc carrier sense multiple access wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1353–1368, 1999.

[140] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e wireless LAN for quality of service." in *European Wireless*, Florence, Italy, February 2002, pp. 32–39.

[141] R. Rozovsky and P. R. Kumar, "SEEDEX: a mac protocol for ad hoc networks." in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC) 2001*, Long Beach, California, October 2001, pp. 67–75.

[142] "Comet," Available from http://comet.columbia.edu.

[143] S.-B. Lee, G.-S. Ahn, X. Zhang, and A. T. Campbell, "INSIGNIA: An IP-based quality of service framework for mobile ad hoc networks," *Journal of Parallel and Distributed Computing*, vol. 60, no. 4, pp. 374–406, 2000.

[144] S.-B. Lee, G.-S. Ahn, and A. T. Campbell, "Improving UDP and TCP performance in mobile ad hoc networks with INSIGNIA," *IEEE Communications Magazine*, pp. 374–406, June 2001.

[145] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, and S. Jamin, "RFC 2205: Resource ReSerVation Protocol (RSVP) — version 1 functional specification," September 1997, status: PROPOSED STANDARD.

[146] X. Hannan, C. K. Chaing, and S. K. G. Winston, "Quality of service models for ad hoc wireless networks," *The handbook of ad hoc wireless networks*, pp. 467–482, 2003.

[147] R. Braden, D. Clark, and S. Shenker, "RFC 1633: Integrated services in the Internet architecture: an overview," June 1994, status: INFORMATIONAL.

[148] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "RFC 2475: An architecture for differentiated services," December 1998, status: PROPOSED STANDARD.

[149] Y. Sun, E. M. Belding-Royer, X. Gao, and J. Kempf, "Real-time traffic support in large-scale mobile ad hoc networks," in *BROADNETS workshop BROADWIM*, San Jose, CA, October 2004, pp. 243–253.

[150] G. Pau, D. Maniezzo, S. Das, Y. Lim, J. Pyon, H. Yu, and M. Gerla, "A cross-layer framework for wireless LAN QoS support," in *Proceedings of IEEE International Conference on Information Technology: Research and Education (ITRE)2003*, Newark, New Jersey, August 2003, pp. 331–334.

[151] Q. Wang and A.-R. M. Ali, "Cross-layer signalling for next-generation wireless systems," in *Proceedings of IEEE Wireless Communications & Networking Conference (WCNC) 2003*, New Orleans, Louisiana, March 2003, pp. 1084–1089.

[152] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, 2005.

[153] M. Madueno and J. Vidal, "Joint physical-MAC layer design of the broadcast protocol in ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 65–75, 2005.

[154] P. P. Pham, S. Perreau, and A. Jayasuriya, "New cross layer design approach to ad hoc networks under rayleigh fading," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 28–39, 2005.

[155] M. Chiang, "To layer or not to layer: Balancing transport and physical layers in wireless multihop networks." in *Proceedings of The Conference on Computer Communications, Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM) 2004*, Hong Kong, China, March 2004, pp. 2525 – 2534.

[156] H. Zheng, "Optimizing wireless multimedia via cross layer design," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME) 2003*, Baltimore, Maryland, July 2003, pp. 185–189.

[157] J. Chen, T. Lv, and H. Zheng, "Joint cross-layer design for wireless QoS content delivery," in *Proceedings of IEEE International Conference on Communications (ICC) 2004*, Paris, France, June 2004, pp. 4243–4247.

[158] S. Martin, P. Minet, and L. George, "The trajectory approach for the end-to-end response times with non-preemptive FP/EDF." in *Software Engineering Research, Management and Applications*, Los Angeles, CA, 2004, pp. 229–247.

[159] G. Pei, M. Gerla, X. Hong, and C.-C. Chiang, "A wireless hierarchical routing protocol with group mobility," in *Proceedings of IEEE Wireless Communications & Networking Conference (WCNC) 1999*, no. 1, IEEE. New Orleans, LA: IEEE, September 1999, pp. 1538–1542.

[160] "EGRESS: Environment for generating realistic scenarios for simulations," Available from http://www.comp.nus.edu.sg/s̃ridhark/mantra/.