ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

CAPACITY ESTIMATION AND OPTIMAL RELAY SELECTION IN IEEE 802.16 MULTI-HOP WIRELESS NETWORKS

GE YU

School of Electrical & Electronic Engineering

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

2010

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Acknowledgment

I would like to take this opportunity to express my sincere gratitude to those who supported me during the pursuit of my PhD degree. Without their continuous guidance, encouragement and support, I would not have gone this far.

Foremost, I would like to thank my supervisor, Dr. Ang Yew Hock, for his constant dedication, guidance, advice and inspiration. I greatly appreciate all the insights, time and tremendous efforts he has given me to improve the quality of this research and thesis.

I would like to extend my appreciation to my friends and colleagues in Institute for Infocomm Research (I^2R), especially, Gu Zhenghui, Wang Haiguang, Su Wen, Sun Peng and Cheng Wang Cho, for their support in these years. Special thanks go to Dr. Chen Zhining in I^2R, for his encouragement and sharing of research methodologies in the research path.

Last but not least I thank my family, including my husband, Zhu Shihao, my daughter, Geri, and my parents, for unconditionally supporting my decision to pursue PhD, for selflessly loving me, and for tirelessly encouraging me whenever I am feeling the sky is grey. I dedicate this thesis to my family.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Contents

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Figures

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Tables

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Abbreviations

AMC	Adaptive Modulation and Coding
AODV	Ad Hoc On-Demand Distance Vector Routing
AP	Access Point
BR	Bottleneck Region
BS	Base Station
BZ	Bottleneck Zone
CBR	Constant Bit Rate
CSI	Channel State Information
DSRC	Dedicated Short-Range Communication
IE	Information Element
LP	Linear Programming
LPG	Local Peer Group
MAC	Medium Access Control
MCMP	Multi-Channel, Multi-Path
MCSP	Multi-Channel, Single-Path

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

MR	Multihop Relay
MSH-DSCH	Mesh Distributed Schedule Control Message
PHY	Physical
PMP	Point-to-MultiPoint
RREP	Route Reply
RREQ	Route Request
RSs	Relay Stations
RSS	Received Signal Strength
RSU	Road Side Unit
SCMP	Single-Channel, Multi-Path
SCSP	Single-Channel, Single-Path
SNR	Signal-to-Noise-Ratio
SS	Subscriber Station
VBR	Variable Bit Rate
VRC	Vehicle-Roadside Communication
WL	Weakest Link
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network
WSN	Wireless Sensor Network
WPAN	Wireless Personal Area Network

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Summary

In the design of IEEE 802.16 multi-hop wireless networks, accurate estimation of end-to-end capacities plays a critical role in Quality of Service (QoS) provisioning. Capacity information has been applied to various network control functions such as admission control, routing, and packet scheduling for assuring service guarantees. In IEEE 802.16 multi-hop relay networks, optimal relay selection, in relation to end-to-end capacity, forms an essential function in achieving good network performance in capacity-constrained networks. In this thesis, research issues relating to end-to-end capacity estimation and optimal relay selection are studied in details for improved network performance in IEEE 802.16 multi-hop wireless networks.

In the first research issue, we have studied the problems associated with capacity estimation in IEEE 802.16 mesh networks. Unfortunately, in such a communication mode, accurate determination of dynamic capacity in wireless mesh networks (WMNs) becomes a challenging issue due to the distributed environment of the mesh network and the interference between communicating nodes.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In contrast, existing studies on capacity estimation are mostly either theoretical studies of capacity without factoring in the effects of transmission overheads in wireless protocols, or empirical studies based on capacity measurements made on specific networks but not those in IEEE 802.16 mesh networks. In this thesis, a solution for computing the dynamic link capacity between two IEEE 802.16 mesh nodes is proposed. The solution is further extended to determine the dynamic end-to-end capacity bounds of a multi-hop route in an IEEE 802.16 mesh network. Our solution introduces a new concept in capacity estimation based on *Bottleneck Zones (BZs)*. A BZ is a region of mesh nodes that has the least available capacity and forms the most resource-constrained network region through which capacities of end-to-end flows are computed or measured. Using our proposed capacity estimation model, we have verified through extensive simulations that the capacity estimation model is accurate, responsive and robust in defining the lower bounds of the capacities, and it can adapt to various network deployment patterns.

In our second research issue studied, we have considered the optimal relay selection problem in IEEE 802.16j multi-hop relay networks. Unlike existing solutions, which are designed for generic networks with no practical consideration on usage scenarios, we have proposed a relay selection solution based on specific mobility patterns in IEEE 802.16j vehicular networks, targeting at maximizing end-to-end capacity. By incorporating highway mobility patterns and the Erceg terrain model, the relay selection problem can be formulated into a non-linear optimization problem, and solved for the optimal location of a relay node that

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

guarantees maximum end-to-end capacity for a vehicular node. Our numerical results show that by selecting the corresponding optimal relay node, the expected end-to-end capacity for subscriber nodes can be increased by nearly 50% as compared to that without the use of relay nodes. Combining both solutions developed in thesis, we are able to obtain reliable capacity estimation and optimal relay selection for QoS provisioning in IEEE 802.16 wireless multi-hop networks.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 1

Introduction

1.1 Background and Motivations

In this era of rapid growth of the Internet community and services, multi-hop wireless networks enable ubiquitous connectivity and a flexible means of communication for both data and multimedia applications. The IEEE 802.11 and IEEE 802.16 are two standards that support multi-hop wireless technologies and multi-service networks to be implemented over community and wide metropolitan areas. Although both standards support multi-hop wireless technologies, they are distinctly different in many perspectives including radio frequencies, physical layer operations and medium access control methods. The IEEE 802.11 uses a contention-based access method and hence is bandwidth inefficient as contention back-off time increases exponentially with increase in network traffic load. The IEEE 802.16, on the other hand, uses contention-free coordinated access mechanism for transmission scheduling, and hence is more efficient in the usage of radio

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

resources. Besides, the IEEE 802.16 PHY specification supports longer-distance links compared to the IEEE 802.11 PHY specification, and thus IEEE 802.16 networks potentially have wider coverage than IEEE 802.11 networks. Furthermore, IEEE 802.16 supports both centralized and distributed scheduling algorithms and therefore allows more flexible bandwidth allocation in multi-hop networks, and also multiple service classes in the Medium Access Control (MAC) layer for various applications. For these reasons, the IEEE 802.16 has emerged as a more promising standard as compared to the IEEE 802.11 standard for supporting ubiquitous mobile broadband wireless services of metropolitan scales.

The IEEE 802.16 standard specifies the data/control and management functions of the MAC and Physical (PHY) layers, as shown in Figure 1.1. The MAC layer comprises of three sublayers: the service-specific convergence sublayer that interfaces higher layers and maps application data to the proper MAC service flow, the MAC common part sublayer that provides the core MAC functionalities such as network access, bandwidth allocation and connection maintenance, and the security sublayer that offers authentication, secure key exchange and encryption. The PHY layer interfaces the MAC layer and handles transmission and reception over a radio interface.

The early versions of the IEEE 802.16 standard support only single-hop point-to-multipoint (PMP) operation mode, where transmissions are assumed to occur only between a base station (BS) and subscriber stations (SSs). In such a scenario, SSs must be within the transmission range of a BS, which inadvertently results in poor receptions at the BS cell boundary due to weak received

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

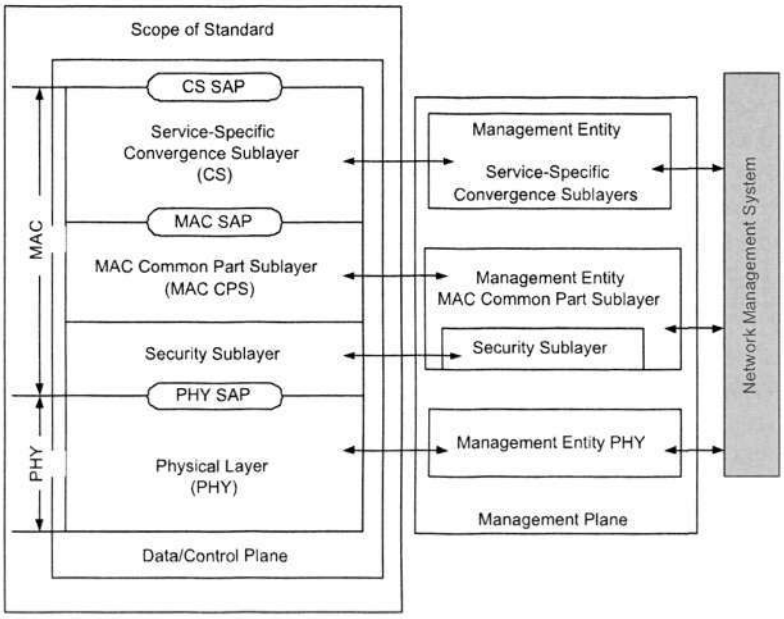


Figure 1.1: IEEE 802.16 protocol layering.

signal power and inter-cell interference. To extend the coverage and enhance the capacity of the network, IEEE 802.16 defined a family of standards [1] for multi-hop extensions in two network configurations, namely, the mesh network and tree network.

The IEEE 802.16-2004 standard [2] defines a mesh operational mode, where peer-to-peer communications between SSs are allowed, and transmissions can be relayed by other SSs across multiple hops. In general, a wireless mesh network (WMN) consists of mesh routers, which cooperatively form a multi-hop backbone wireless network as gateways to provide Internet access to end users [3, 4]. This thesis focuses on IEEE 802.16-2004 WMNs where mesh routers communicate with each other over IEEE 802.16 radio interfaces, and end users connect to mesh routers via other interfaces such as Ethernet and IEEE 802.11 wireless LAN, as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

shown in Figure 1.2. IEEE 802.16-2004 mesh networks are more advantageous

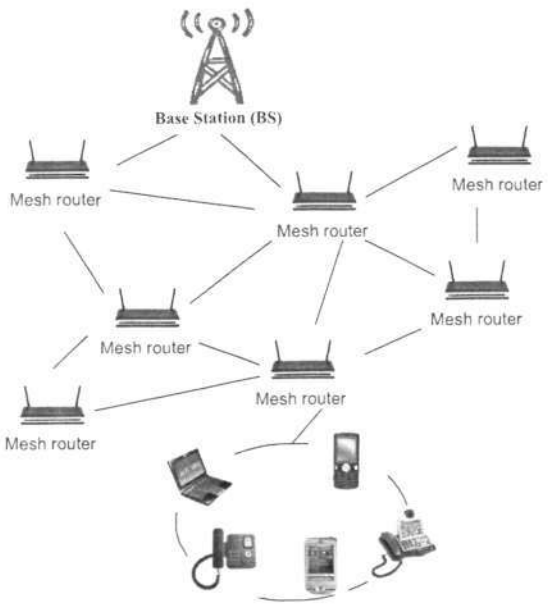


Figure 1.2: The architecture of IEEE 802.16 wireless mesh networks.

and scalable than traditional single-hop PMP networks for the following reasons: (1) they can be deployed more cost effectively than traditional PMP networks, (2) they are inherently more reliable, as each node is able to reach the Internet via multiple routes, (3) they are more bandwidth-abundant, as capacity-optimal paths can be selected for users to reach the Internet, and (4) they are more extendable, as additional mesh nodes can be deployed for wider area coverage.

Separately, the IEEE 802.16j [5] standard defines a tree-based Multi-hop Relay (MR) operation for improving the network performance of IEEE 802.16e [6] in PMP wireless communications. Existing IEEE 802.16e networks are inherently

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

problematic due to signal attenuation causing low signal-to-noise-ratio (SNR) at cell boundaries and signal blockage causing coverage holes within the transmission regions. The IEEE 802.16j MR standard [5] is being proposed to overcome the inherent problems in IEEE 802.16e networks by deploying relay stations (RSs) for relaying control messages and data packets between BSs and SSs. IEEE 802.16j networks are compatible and interoperable with IEEE 802.16e networks in terms of the operations of BSs and SSs. An existing IEEE 802.16e BS can be conveniently upgraded to an IEEE 802.16j MR-BS (multi-hop relay BS), and an IEEE 802.16e SS can still work gracefully in an IEEE 802.16j network without noticing the existence of the RSs. In this thesis, the MR-BS and BS are semantically the same. Unlike BSs, RSs are not directly connected to a wire-line backhaul, thus they can be easily and rapidly deployed without the complications and restrictions of wired lines. The RSs are functionally less complex than BSs and therefore are less costly to deploy and also reduce the overall network cost as fewer BSs are needed to be deployed. This makes IEEE 802.16j an attractive standard for wide area coverage and economical deployment. In an IEEE 802.16j network, as shown in Figure 1.3, an SS either connects directly to a BS or to an RS, depending on the location of the SS and other routing criteria.

IEEE 802.16j multi-hop relay networks present several merits: (1) the number of BSs in the system can be significantly reduced by introducing RSs. As RSs do not need to directly connect to a wired backbone and are less complex compared to BSs, the deployment costs are significantly reduced, (2) the transmission ranges of both SSs and BSs can be reduced due to multi-hop relaying,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

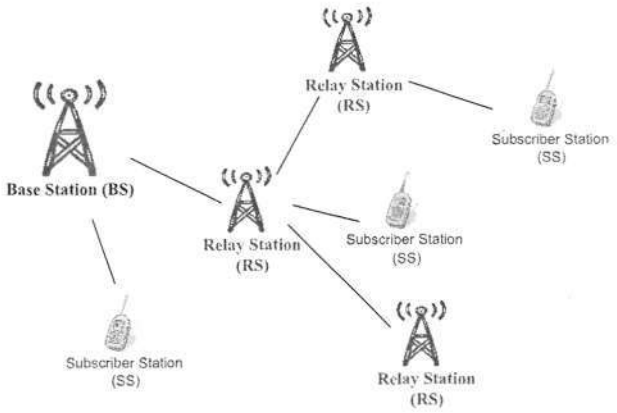


Figure 1.3: The architecture of IEEE 802.16j system.

(3) the transmission powers can be reduced due to the shorter transmission range requirement and hence the network is more energy efficient and the lifetime of mobile users can be prolonged, and (4) multi-hop paths via RSs are less vulnerable than single-hop paths as high-quality links can be chosen among many relay paths for improved data transmission.

Multi-hop wireless networks are expected to support a wide variety of applications including data and multimedia services. These applications demand multi-services with different network Quality of Service (QoS) requirements, such as end-to-end packet delay and loss ratio, and pose even a greater challenge in capacity scarce wireless networks. In this thesis, we aim to find solutions for optimizing the use of limited network resources to best meet the conflicting demands of multi-services. We focus our research on two key issues of capacity related problems, namely, real-time end-to-end capacity estimation and capacity

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

optimized relay selection for radio resource management in capacity-constrained multi-hop wireless networks.

1.2 Thesis Objectives

To achieve the aim of our research, we set out to achieve two objectives in this thesis: (1) to develop a practical capacity estimation solution that takes into consideration the interference and deployment patterns in the computation of end-to-end capacity bounds in IEEE 802.16 wireless mesh networks, and (2) to develop an optimal relay selection solution based on capacity analysis to achieve maximized end-to-end capacity for users in IEEE 802.16 multi-hop relay networks.

1.2.1 Capacity Estimation in IEEE 802.16 Wireless Mesh Networks

Network capacity is a crucial parameter for QoS support in resource scarce multi-hop wireless networks. The ability to accurately determine the end-to-end capacity forms an important function in radio resource management and provides effective network control for QoS guarantees in wireless mesh networks. Capacity information has been widely used in many network control mechanisms such as capacity-aware routing [7, 8], admission control [9–11], packet scheduling [12, 13], server selection [14–16], and so on.

For example, to support QoS in resource scarce wireless networks, effective admission control must be put in place to prevent overwhelming of limited

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

radio resource by accepting too many flows. Admission control can be easily implemented in wired networks where available link capacity can be determined without the problem of the interference from neighboring nodes. However, admission control schemes used in wired networks cannot be directly applied to wireless networks as co-channel interference in wireless networks makes the determination of link and end-to-end capacity difficult, especially in wireless mesh networks.

Existing capacity studies in wireless networks, in general, are limited to the following three aspects: (1) providing theoretical scaling laws in capacity estimation, without considering the effects of wireless protocol overhead, (2) predicting or measuring capacity through simulations or experiments in other wireless networks but not in IEEE 802.16 mesh networks, and (3) theoretical studies of MAC layer capacity, without considering the practical requirements of real-time computation and distributed environment.

In this research, we aim at developing a practical solution for capacity estimation based on our new concept of Bottleneck Zones. Our objectives are to provide accurate and reliable real-time end-to-end capacity estimation for radio resource management, and optimal relay selection for QoS provisioning in IEEE 802.16 multi-hop wireless networks. To support these two objectives, the solutions for capacity estimation must satisfy the following requirements. Firstly, the process of capacity estimation must be executed in a distributed manner. This will avoid the need for global information knowledge which are hard to obtain in wireless mesh networks. Secondly, the capacity estimation process should incur as little protocol overhead as possible in view of the scarce resources in wireless net-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

works. Thirdly, it must take into account of different interference and deployment patterns peculiar to the type of wireless networks. Fourthly, the capacity must be computed in real-time so that network control can be administered promptly to avoid network congestions from building up.

In this thesis, the term *capacity* refers to the available bandwidth in a link or route that can be used for future data transmissions. This definition is consistent with that used by Chobanyan et al [17] and by Jain and Dovrolis [18] in their studies on network capacity in the Internet. We also consider a wireless mesh network as one described by Robinson and Knightly [3]. They described a wireless mesh network as one usually deployed in two tiers, where an access tier allows a customer's device to access the network through a mesh node, and a backhaul tier allows mesh nodes to connect the access tier to a wired gateway. Based on this network model, our study focuses on the backhaul tier of the IEEE 802.16 wireless mesh network, and where the coordinated distributed scheduling mechanism is present.

1.2.2 Capacity-Aware Optimal Relay Selection in IEEE 802.16 Multi-hop Relay Networks

The IEEE 802.16j standard defines various usage models [19] for different deployment scenarios. The usage models include: (1) fixed infrastructure usage model, whereby fixed RSs are deployed at the cell boundary to provide coverage extension, (2) in-building coverage usage model, whereby RSs are deployed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to provide better coverage and higher throughput for users in a building, tunnel or underground, (3) temporary coverage usage model, whereby nomadic RSs are deployed temporarily to provide additional coverage or capacity in cases of emergency, disaster recovery and events, and (4) coverage on mobile vehicle usage model, whereby SSs are traveling on mobile vehicles. The usage models dictate where the coverage is being provided and do not specify the manner in which RSs are to be deployed. As the network deployment patterns and network characteristics may be significantly different in various usage models, an optimal relay selection solution for generic usage models may not be practical and efficient for a specific one.

In recent years, the demand for wireless vehicular communications is increasing. Many vehicle manufacturers have shown interest in integrating communication devices with their vehicles for the purpose of safety, entertainment, and mobile commerce. It is foreseeable that one of the major goals of network operators in future vehicular networks is to provide QoS for both safety and non-safety related applications [20]. In safety applications, dedicated short-range communication (DSRC) has been proposed as a key enabling technology [21] based on the IEEE 802.11 wireless protocol [21]. Unfortunately, the performance of this protocol is not satisfactory in vehicular environments due to the contention and non-deterministic nature of IEEE 802.11 transmissions [22, 23]. As such, IEEE 802.16 wireless vehicular networks are becoming more popular and replacing IEEE 802.11 networks.

In this thesis, we focus on IEEE 802.16j coverage based on mobile vehicle

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

usage model that is applicable to vehicle-infrastructure communications for in-vehicle Internet access of various transport systems, such as highways, railways, near-land rivers, and seas. Our objective is to develop an optimal relay selection solution for optimizing the performance of end-to-end flows in IEEE 802.16 MR networks. We aim to provide a theoretical derivation of the optimal solution based on actual mobile vehicle usage model. As network topology and node mobility pattern are captured in the actual usage model and we also consider the specific wireless channel model pertaining to the actual usage scenario, our developed solution is expected to be realistic and practical as well.

Based on the IEEE 802.16j network of Figure 1.3, the proper selection of an RS to serve a particular vehicular SS is crucial in achieving good end-to-end performance. In developing our solution for finding the optimal RS to achieve maximized end-to-end data rate for a particular vehicular SS, we require the solution to be computationally simple and can be conducted in a proactive way. In our proposed solution, the RSs may be stationary units mounted on towers and buildings, or mobile units carried by vehicles. In both cases, we can assume that the battery life of RSs to be long or unlimited as they can readily tap on energy sources at their fixed locations or on-board the vehicles. In such cases, and similarly in those studied by So and Liang [24], there is no need to consider any energy saving strategy in the solution.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.3 Thesis Contributions

This thesis addresses two critical capacity-related issues in IEEE 802.16 multi-hop wireless networks. The first issue relates to obtaining an accurate end-to-end capacity estimation [25, 26] and the second issue selects a relay node to optimize the QoS guarantees of end-to-end flows [27, 28]. The research extends existing theoretical studies on capacity-related issues in generic multi-hop wireless networks to the development of practical solutions for IEEE 802.16 networks. These major contributions are summarized as follows:

1. End-to-End Capacity Estimation. The contribution in capacity estimation is described as follows. Firstly, the bandwidth availability status in a specific IEEE 802.16 mesh node is analyzed along with various configurations of interference range in multi-hop wireless networks. Based on this, the real-time link capacity between two neighboring IEEE 802.16 mesh nodes is computed in accordance to the manner in which the coordinated distributed scheduling process operates in the IEEE 802.16 standard. Using existing control messages in IEEE 802.16 for disseminating capacity information among mesh nodes, the solution for computing link capacity does not incur extra transmission overhead in the network.

   Secondly, a new concept on *Bottleneck Zone* is introduced to take into account the different ranges of transmission interference. Based on the *Bottleneck Zone* concept, an interference-aware scheme is proposed to obtain real-time end-to-end capacity bounds for a given route. We adopt the K-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

hop interference model as proposed by Sharma et al [29], which itself is an extension of the Protocol Model proposed in [30]. In the K-hop interference model, all nodes within K-hop distance from a receiver are considered as interfering nodes to this receiver. The value of K is configurable based on specific network deployments. Hence, the K-hop interference model can be adapted and applied to networks with different configurations. The information on the capacity bounds provides the means for QoS provisioning such as admission control and capacity-aware routing.

Thirdly, real-time end-to-end capacity bounds are computed based on actual network deployment configurations including optimal and random deployment patterns with different node densities. Hence, the real-time capacity bounds obtained are practical.

2. Optimal Relay Selection. The contribution in optimal relay selection is described as follows. Firstly, the highway mobility model is studied in terms of the generic characteristics of transport systems and the probabilistic distribution of data transmission rates is obtained based on highway mobility pattern.

   Secondly, link adaptation technology has been used to form and solve the relay selection problem. The objective of the relay selection problem is to maximize the end-to-end capacity for an SS by selecting the optimal RS. We have formulated the selection problem as a non-linear optimization problem based on a vehicular SS moving along a highway and solving it by

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

optimizing the end-to-end capacity for users.

Thirdly, the actual Erceg terrain model is used to generate the link adaptation data which are applied to compute the optimal relay selection. As actual terrain models are used in the optimization problem, the solution obtained is practical.

1.4 Thesis Organization

The structure of this thesis is as follows. Chapter 2 presents the literature survey of capacity studies and optimal relay selection solutions in wireless multi-hop networks. Following a brief analysis of different interference models in wireless networks, a detailed survey on different capacity study solutions in wireless networks is presented. Most of the existing capacity study solutions will be briefly discussed and analyzed in order to justify the need of a new real-time capacity estimation solution. Subsequently, various optimal relay selection schemes are analyzed based on their network configurations. Chapter 3 presents the proposed scheme to measure link capacity between two neighboring nodes in an IEEE 802.16 mesh network, and describes the proposed *Bottleneck Zone* based solution for computing the end-to-end capacity bounds for a given route, considering different network deployment patterns. Chapter 4 presents extensive simulation results to demonstrate the effectiveness and accuracy of the proposed capacity estimation solutions. Chapter 5 explores the problem of selecting the optimal RS for a vehicular SS in order to maximize its end-to-end capacity, assuming that

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the location of the vehicular SS is known. Chapter 6 shows numerical results of the optimal relay selection scheme, including the connectivity and achieved end-to-end capacity in the network. Finally, Chapter 7 concludes this thesis and discusses future work.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2

Literature Review

2.1 Introduction

Capacity information plays an important role in radio resource management, particularly in wireless networks where capacity resources are scare. Information on the availability of network capacity provides valuable insight into the state of network resource utilization and allows proper traffic control into networks for optimal network performance. In this chapter, we present a critical review on capacity estimation as well as optimal relay selection for achieving good end-to-end performance in multi-hop wireless networks.

2.2 Capacity Estimation in Wireless Networks

Capacity information is required for optimal radio resource management in wireless networks for proper network control and traffic regulations; such as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

in QoS routing, admission control and packet scheduling. Existing studies on capacity estimation can be broadly classified into model-based approaches and measurement-based approaches. In the model-based approaches, capacity scaling laws governing node density are derived using specific assumptions about the network. In the measurement-based approaches, capacities of given networks are obtained from actual measurements through field tests and simulations.

2.2.1 Preliminaries

In capacity study, *transport capacity* is defined as the total bit-distance product per second that can be transported by the network [30]. *Throughput capacity* is defined as the maximum common throughput that can be provided to each node with a randomly chosen destination [30].

Research on capacity in wireless networks can be categorized into two groups: (1) model-based approach, where capacity is studied based on assumptions of specific models like interference models, mobility models, and traffic models, etc., and (2) measurement-based approach, where capacity is measured through experiments with certain network configurations.

2.2.1.1 Interference Models in Wireless Networks

Interference models capture effects of interference on the success probability of packet transmission in a network. An accurate modeling of the interference pattern is fundamental in capacity estimation in wireless networks. Two interference models are popularly applied in capacity estimation for wireless networks:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Protocol Model, which is a graph-based model, and Physical Model, which is a SINR-based model. Both models have been formally described by Gupta and Kumar [30].

a) Protocol Model

Gupta and Kumar [30] described the Protocol interference model as follows: suppose node X_i transmits over the m-th subchannel to node X_j, then this transmission is successfully received by node X_j if $|X_k - X_j| \geq (1+\Delta)|X_i - X_j|$ for every other node X_k simultaneously transmitting over the same subchannel, where $|X_k - X_j|$ denotes the distance between node X_k and node X_j. The quantity $\Delta > 0$ depicts the guard zone.

The Protocol Model is widely used by researchers in their capacity studies [30–32]. For example, Keshavarz-Haddad and Riedi [31] studied broadcast capacity considering the Protocol Model for multi-hop wireless networks.

The simulation results of Gronkvist and Hansson [33] as well as the analytical results of Behzad and Rubin [34] indicated that the Protocol Model does not provide a comprehensive view of the effects of actual interference, but it gives a tractable and useful approximation to the interference in the wireless networks. Therefore, the Protocol Model, though may be simplistic in scope, does in fact simplify the theoretical performance analysis of network protocols.

The k-hop interference model, proposed by Sharma et al [29], is an extension of the Protocol Model. The k-hop interference model considers all transmitters within a hop distance of k from the receiver as interferering nodes.

Another variant of the Protocol Model is the Fixed Power Protocol Model

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

suggested by Wang et al [35]. In contrast to the Protocol Model which assumes the transmission range and interference range are homogenous in the network, the Fixed Power Protocol Model allows each node to have different transmission range and interference range from each other. However, the two ranges for a specific node must be fixed for a certain period of time.

b) Physical Model

Gupta and Kumar [30] described the Physical interference model as: for a successful transmission from node X_i to node X_j, the SINR received by X_j must be above a threshold, that is

$$SINR_{X_j} = \frac{\frac{P_i}{|X_i - X_j|^\alpha}}{N + \sum_{\forall k \neq i} \frac{P_k}{|X_k - X_j|^\alpha}} \geq \beta \tag{2.1}$$

where P_i denotes the transmission power of node X_i, N denotes the noise power, α denotes the attenuation factor over the wireless channel, and β denotes the SINR threshold.

The Physical Model has also been used by many researchers in their studies on capacity estimation [36,37]. A variant of the physical model, based on a *virtual coordinate* system, was proposed by Lim et al [38]. It uses a *virtual distance* measure based on the received signal strength (RSS) between mesh nodes to indicate the level of interference between them. The virtual coordinates form a square matrix and represent the interference level of each pair of nodes.

Agarwal and Kumar [39] proved that the Protocol Model has stricter constraints in comparison to the Physical Model. When a configuration satisfies the constraints of the Protocol Model by choosing an appropriate guard zone, then

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

such a configuration also satisfies the constraints of the Physical Model with appropriate settings. Compared to the Protocol Model, the Physical Model is a more realistic model although it is also more complex and makes the capacity analysis more challenging.

Besides the two basic interference models, specific wireless protocols define their own interference models based on their context. For example, RTS/CTS interference model was specifically used in IEEE 802.11 [35,40], whereby all nodes that are within the interference range of either the transmitter or the receiver are not allowed to transmit.

c) Graph Models in Capacity Estimation

Jain et al [41] modeled interference in wireless networks using a *conflict graph*, indicating the mutually interfered links that cannot be active simultaneously. Using the concept of *conflict graph*, the upper and lower bounds on the optimal throughput were computed for a given network and traffic load. The concept of *conflict graph* was also applied by Yang and Vaidya [42,43] for priority scheduling in wireless networks. In their context, the conflict graph is defined in terms of flows rather than links.

Stuedi and Alonso [44] proposed a *schedule graph* in capacity study, where a probabilistic approach was proposed to compute the transport capacity of a network under the Physical Model and the achievable throughput per communication pair is modeled as a random variable. The achievable link throughput is linked to the connection probability of the two nodes, which depends on a given node distribution and transmission range.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Similarly, Wu et al [45] introduced a *capacity graph* concept, which contains the information of the concurrency of the transmissions among the links in the network using a random packing strategy. The *capacity graph* can be used in scheduling and capacity estimation.

2.2.1.2 Topologies of Wireless Mesh Networks

In general there are two broad categories of mesh network topologies: regular and random network topologies.

a) Regular topologies

As defined by Robinson and Knightly [3], in a regular mesh topology, neighborhood nodes form vertexes of specific shapes with interconnecting links. They considered three regular tessellations of triangular, square, and hexagonal as the baseline mesh topologies.

b) Random topologies

Robinson and Knightly [3] also considered random mesh topologies that were modeled in two ways, i.e. random process and perturbation. In the first way, random mesh topologies were modeled as homogeneous Poisson processes. The second way of modeling the randomness was to introduce perturbation to an ideal regular grid topology through displacing each mesh node by a random distance and angle. The random mesh network based on node perturbations approximates the practical network as perfect grid placement cannot usually be achieved due to the constraints of physical environment.

Robinson and Knightly [3] studied the performance of wireless mesh networks

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(WMNs) and concluded that network topology can strongly influence the mesh network performance. They have drawn their conclusions based on extensive Monte Carlo simulations. The conclusions drawn from Robinson and Knightly's study [3] can be summarized as follows:

- A regular topology provides up to 20% higher coverage on average as compared to a random topology.
- A random topology is unsuitable for a large-scale mesh deployment, since such a deployment requires doubled node density.
- A moderate level of perturbations from ideal grid placement has a minor impact on network performance.
- The routes in a random network are on average one hop longer than the routes in a square grid network.

Camp et al [46] explored the performance of various WMNs using the empirical data gained from the single-hop and multi-hop measurements in a two-tier urban scenario. The study showed that a regular grid topology achieves up to 50% higher throughput than random node placement. They further measured the achievable rate for a flow of length i hops and found that spatial reuse can be achieved four hops away from the transmitter with two-hop neighbors interfering.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.2 Model-Based Capacity Estimation

This subsection presents numerous studies on capacity scaling laws, and factors affecting capacity estimation in wireless networks such as transmission interference, transmission power, scheduling schemes, and mobility issues.

Two types of networks have been considered in model-based capacity estimation:

1. Arbitrary networks, where nodes are arbitrarily located and traffic patterns are arbitrary. In an arbitrary network, nodes are arbitrarily located in a disk of unit area in the plane. Each node sends traffic at an arbitrary rate to an arbitrarily chosen destination.

2. Random networks, where nodes are randomly located and traffic patterns are random. In a random network, nodes are randomly located, i.e. independently and uniformly distributed in a disk of unit area in the plane. Each node sends traffic at a random rate to a randomly chosen destination.

For practical deployment, an arbitrary network allows the network designer to choose the locations of nodes and traffic patterns without restrictions. Hence, the arbitrary network can be viewed as the best case scenario and a reference network for optimal use of network capacity as compared to other topological mesh networks.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.2.1 Capacity Scaling Laws

Capacity scaling laws are important guidelines in network planning, strategic design and optimization to achieve certain predefined performance benchmark. The contributions of the studies on capacity scaling laws are to provide the generic capacity bounds in multi-hop wireless networks.

Gupta and Kumar made a milestone contribution on theoretical capacity study in wireless networks [30]. Assuming that n nodes are located in a region with a unit area, their main results are:

1. The transport capacity of an arbitrary network under both of the Protocol Model and Physical Model is in the order of $\sqrt{n}$ bit-meters per second in the optimal scenario, where the nodes are optimally placed, the traffic pattern and range of each transmission are optimally chosen.

2. In random networks, the order of the network transport capacity is $\sqrt{n/\log n}$ under both of the Protocol Model and Physical Model.

In wireless networks, node density is a critical factor that affects network connectivity and transport capacity. Generally, connectivity improves when the node density increases, but the effect of the node density on the transport capacity differs in various interference models and deployment patterns. For example, the transport capacity in the Protocol Model has been proven to scale at the rate of $\sqrt{n}$ in an arbitrary network, where n is node density, but at the lower rate of $\sqrt{n/\log n}$ in a random network [30].

Dousse et al [47] studied the effect of the radio attenuation function on the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

achievable capacity in multi-hop wireless networks under the Physical Model. They showed that when signal attenuates faster over distance, the connectivity decreases but the capacity could increase due to less interference incurred by a transmission. Therefore, there is a fundamental trade-off between connectivity and capacity.

The transport capacity of wireless networks was further studied by Xue et al [48], taking multipath fading into considerations. They worked out an upper bound on the transport capacity considering channel models with flat and slow fading, and a lower bound on the transport capacity, covering the case of rapidly varying channels. They concluded that the transport capacity of a network can grow no faster than linearly in the size of this network, which conforms to Gupta and Kumar's results.

Franceschetti et al [49] closed a previous gap between upper and lower bounds on the optimal per-node throughput capacity presented by Gupta and Kumar under the Physical Model. Their interesting finding showed that nodes in a random network can achieve the asymptotic scaling as nodes in an arbitrary network in the high signal attenuation regime.

Assuming that the network topology and traffic patterns are known a priori, Kodialam and Nandagopal [50] derived the capacity upper bounds under the Protocol Model by using constrained optimization method in networks with multiple channels and multiple interfaces. A similar problem was explored by Alicherry et al [40] except that they focused on constrained mesh networks where traffic is only forwarded between gateway node and mesh node.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Agarwal and Kumar [39] generalized the analysis of the lower bound on the transport capacity for the Protocol Model and Physical Model. By showing that the Protocol Model is more restrictive than the Physical Model, they proved that lower bound capacities under the Protocol Model continue to hold under the Physical Model.

Chafekar et al [51] considered the problem of characterizing the achievable capacity for arbitrary networks under the SINR interference model. The throughput capacity was approximated by means of a linear programming (LP) formulation, where similar approaches using LP were proposed by Kodialam and Nandagopal [52], and by Kumar et al [53] for capacity estimation under the graph-based interference model.

Sun et al [54] computed the capacity in IEEE 802.11 networks, where the packet transmission attempts were assumed to be Poisson distributed. Bianchi [55] computed the capacity on the assumption of a probability distribution of successful transmissions over a link.

Initially, research on capacity study assumed that each node transmits at a certain uniform rate over a common wireless channel, thus adaptive transmission was not considered. Later, some researchers considered multiple transmission rates in their capacity study. For example, Toumpis and Goldsmith [56] captured the rate over multi-hop route considering adaptive transmission using SINR interference models.

Zhou et al [57] studied the asymptotic theoretical throughput capacity of infrastructure wireless mesh networks (WMNs) under the Protocol Model. They

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

explored the appropriate number of mesh routers and gateways that need to be deployed in a WMN in order to achieve high transport capacity. Their results can be used as guidelines for protocol design and deployment of WMNs.

Besides the capacity scaling laws, researchers have explored the factors that could affect capacity estimation in wireless networks. In the following subsections, factors discussed include transmission power, scheduling schemes, multi-channel and mobility issues.

2.2.2.2 Transmission Power Issues

Some researchers considered transmission power in capacity study. Gomez and Campbell [58] extended Gupta and Kumar's results [30] by adding variable transmission power in the analysis. They showed that in the optimum case when each node chooses its own optimum transmission power, the average throughput capacity per node keeps constant even if the node density of the network increases. Their results further showed the average capacity per node increases as the common transmission range increases up to a certain point. After that point the average available capacity per node decreases sharply due to the domination of signaling overhead from routing protocols. Jovicic et al [59], as well as Xie and Kumar [60], studied upper bound capacity using information-theoretic methods, which take power constraints into considerations.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.2.3 Scheduling Scheme Issues

In wireless networks, transmission scheduling deals with resource allocation over wireless links, and it therefore significantly affects network capacity. Lin [10] pointed out that in TDMA-based multi-hop wireless networks, the end-to-end capacity is decided by not only the available slots on each link, but also the way the slots are scheduled.

Different scheduling algorithms lead to different fairness levels, which in turn result in different system capacity. The relationship between fairness level and system capacity was studied by Kumar et al [53]. They found that there is a trade-off between fairness level and system capacity. When the fairness level approaches one, implying that all flows have almost identical individual capacity, the aggregate system capacity can only be half of that in the system with very low fairness level in which some flows are starved.

2.2.2.4 Multi-channel Issues

Gupta and Kumar's results [30] are applicable to single channel wireless networks or multi-channel wireless networks where there is a dedicated interface for each channel and hence no switching between channels. Kyasanur and Vaidya [61] extended Gupta and Kumar's results to a more generic and practical case where a few of channels share an interface and channel switching occurs frequently. They concluded that the transport capacity scaled down to $1/\sqrt{k}$ compared to Gupta and Kumar's result in arbitrary networks under the Protocol Model, where k is the ratio of the number of channels to interfaces. While in random networks,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the transport capacity does not scale down to $1/\sqrt{k}$ when k is less than $\log n$, where n is the number of nodes in a unit area region.

Tam and Tseng [62] studied the capacity in multi-channel wireless networks, and showed that an ideal multi-channel MAC protocol could achieve a maximum end-to-end throughput of $\frac{1}{2}$ of the effective MAC transmission rate in a chain topology. They then analyzed the upper bounds of the end-to-end throughput, which were summarized as follows, assuming the effective MAC layer data rate being r:

- In the single-channel, single-path (SCSP) scenario, an ideal MAC protocol could only achieve an end-to-end throughput up to $r/3$, assuming a receiver is not interfered by a transmitter which is not the receiver's neighbor.
- In the multi-channel, single-path (MCSP) scenario, an ideal multi-channel MAC protocol could achieve end-to-end throughput up to $r/2$, due to the half-duplex property of radio.
- In the single-channel, multi-path (SCMP) scenario, the achievable end-to-end throughput is slightly higher than the SCSP scenario.
- In the multi-channel, multi-path (MCMP) scenario, the ideal end-to-end throughput can be up to r.

2.2.2.5 Mobility Issues

Grossglauser and Tse [63] introduced mobility into the capacity study in wireless networks. They found that data transmissions can be made with fewer

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

hops through node mobility. As a result, the average long-term throughput can be maintained constant at the cost of increased end-to-end delay.

Lea et al [64] quantified the mobility level in cellular mobile networks using the average number of handoffs during the lifetime of a call. They then studied the capacity of cellular mobile networks using analytical model and concluded that channel utilizations drops when the mobility level increases.

Sharma et al [65] analyzed the trade-off between delay and capacity in mobile networks. They studied the effect of various mobility models on the capacity of wireless networks [66, 67] under the assumption of the graph-based interference model. They concluded that a network showed a smoother delay-capacity trade-off curve in the pattern that the nodes move over long distances without changing directions, as compared to the pattern that the nodes change directions frequently.

By exploring a cluster-based routing scheme where only the cluster head is responsible for communicating with mobile nodes, Bansal and Liu [68] observed that the throughput obtained is independent of the velocity of the mobile nodes.

The discussion regarding model-based capacity estimation research is summarized in Table 2.1, which considers the following factors: network type, interference model, transmission power, scheduling schemes, and mobility.

2.2.3 Measurement-Based Capacity Estimation

Capacity estimation in wireless networks can be analyzed based on actual measurements of throughput in the prototype networks. Such prototype networks are implementations of IEEE 802.11 or IEEE 802.16 protocol with specific node

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

deployment and traffic patterns.

2.2.3.1 Capacity Scaling Laws through Measurements

Li et al [69] studied end-to-end capacities in multi-hop wireless networks through simulation and found that the end-to-end capacity of a network with a chain topology is $\frac{1}{7}$ of the single-hop throughput, which is much less than the expected $\frac{1}{4}$ assuming two-hop interference; the end-to-end capacity of a regular lattice network is $\frac{1}{17}$ of the single-hop throughput, which is less than the expected $\frac{1}{12}$. The authors further found that if the local communications (communication with small hops) dominates, the network can scale in achievable capacity.

Dong et al [70] found that the maximal channel utilization is related to node density in IEEE 802.11 networks. From the simulation results, they derived an approximate formula, indicating that the reciprocal of maximum channel utilization linearly decreases with an increase in network node density.

2.2.3.2 Capacity Measurements using Contention Discovery Approach

Xiao et al [71] introduced a concept of *transmission budget*, which indicates the residual transmission time for a user. They then proposed a solution to measure *transmission budget* during beacon intervals of IEEE 802.11 frames for each node, and used that for a measurement-based admission control scheme.

Yang and Kravets [11] proposed capacity-based routing and admission control schemes for IEEE 802.11 ad hoc networks. Their approaches were based on knowledge of both local capacity at a node and the contention of the neighboring

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

nodes, which they referred to as *c-neighbors*.

2.2.3.3 Capacity in Terms of the Amount of Application Flows

Some researchers studied capacity based on the amount of application flows that are supported in a network such as Shin and Schulzrinne's study on VoIP flows [72] and Cho's study on video flows [73]. Shin and Schulzrinne [72] studied the capacity in terms of the number of VoIP calls supported in a network through experimental measurement using delay and loss ratio of the VoIP calls as the capacity benchmark. Hillestad et al [74] studied the number of simultaneous video-on-demand users that an IEEE 802.16 PMP rural network can support with 3.5 GHz frequency band and 20 MHz channel bandwidth. In Harsha's study [75], the effect of TCP flows on the VoIP capacity was explored.

2.2.3.4 Metrics and Probing-Based Capacity Measurements

Network metrics such as delay have been used to estimate the achievable capacity. Kazantzidis et al [76] proposed a measurement method to compute achievable bandwidth by considering the aggregate delay, which consists of link transmission delay, MAC queueing time, collision avoidance phase, etc.

Ahn et al proposed an end-to-end capacity probing method [77], in which the source node sends a probing request message that contains a *bottleneck bandwidth* field towards the destination node. An intermediate node updates the *bottleneck bandwidth* field in the message if the available bandwidth at the node is less than the current value of the field. When the destination node receives the probing

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

request message, it then sends a probing response message back to the source node with the *bottleneck bandwidth* field copied from the probing request message.

Measurement-based capacity estimation approaches are summarized in Table 2.2, where the following key issues are considered: network type, deployment, transmission power, traffic pattern, and mobility.

Chua et al [78] compared two capacity estimation approaches: parameter-based and measurement-based approaches, and concluded that the results obtained from a measurement-based approach (using Time Sliding Window [79]) are more accurate than those obtained from a parameter-based approach. In their context, the parameter-based approach is similar to the model-based approach described in this chapter.

2.2.4 Capacity Estimation in IEEE 802.16 Wireless Networks

With the popularity of IEEE 802.16 technology, researchers have explored capacity in such a wireless network with various configurations.

2.2.4.1 Theoretical Analysis on Capacity Estimation

Redana and Lott [80] performed theoretical analysis on the throughput in IEEE 802.16a networks by means of an analytical approach. They found that, for two-hop transmissions, centralized scheduling outperformed distributed scheduling in terms of throughput due to the lower overhead of the centralized scheme. Hoymann [81] showed that a cell throughput of 10 Mbps is achievable in a multi-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

user and multi-mode scenario with 15 users in the system.

Hincapie et al [82] studied raw link capacity based on received signal power. They proposed to use high rate link in path selection because transmissions through a higher rate link require fewer minislots than through a lower rate link, and hence the chosen route is the one that minimizes the total amount of minislots required for end-to-end transmissions.

Niyato and Hossain [83] proposed a bandwidth allocation and admission control solution using game-theoretic formulation, queueing model and Poisson arrival traffic pattern. Elayoubi et al [84] built up an analytical model to calculate the Erlang capacity where the threshold of the blocking probability is based on the SNR for a specific modulation and coding scheme.

2.2.4.2 Effects of Scheduling Mechanisms on Capacity

The scheduling mechanism has significant effects on network performance and capacity in IEEE 802.16 wireless networks. Cicconetti et al [85] studied the performance of IEEE 802.16 networks in PMP mode under the loading of several traffic types. They found that the end-to-end delay is highly dependent on the delay introduced by the bandwidth request mechanism. Similar results have been shown by Cao et al [86] for the coordinated distributed scheduling mechanism of the IEEE 802.16 mesh mode in that the three-way-handshake mechanism for bandwidth allocation significantly affects network performance in terms of throughput and delay. Cao et al [87] then studied the distribution and mean intervals between successive transmission opportunities of scheduling messages

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

assuming two-hop neighboring interference.

2.2.4.3 Factors Affecting Capacity Estimation

Nuaymi and Noun [88] explored the affecting factors on capacity of an IEEE 802.16 PMP network with a dense urban network deployment through simulations. By calculation using the received signal strength, they showed that the capacity is closely related to the number of active users, the application traffic profile and the spatial reuse pattern. Subsequently, Belghith and Nuaymi [89] found that IEEE 802.16 capacity is also highly dependent on the algorithm for scheduling transmissions of different traffic classes. Pries et al [90] found that the downlink/uplink subframe ratio significantly affects network capacity because network traffic is asymmetric by nature. They then proposed a capacity enhancement solution in PMP mode in that an adaptive framing algorithm was introduced which adjusted the downlink/uplink subframe ratio according to the measured traffic load. Tarhini et al [91] computed system capacity in IEEE 802.16 networks in PMP mode based on the blocking probabilities of streaming and elastic flows, considering adaptive transmission and inter-cell collisions.

2.2.4.4 Fairness in Capacity Allocations

Cicconetti et al [92] proposed a fair end-to-end bandwidth allocation algorithm (FEBA) for IEEE 802.16 nodes to negotiate bandwidth in a multi-channel wireless mesh network. Bandwidth requests and grants at each neighbor are proportionally assigned in a round-robin fashion so that fairness can be achieved in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

a multi-user environment.

2.2.4.5 Capacity in Indoor Coverage Case

Recently, IEEE 802.16 femtocell has been introduced [93–96] to provide indoor coverage. Such a deployment was designed to support low mobility users requiring high data rate communications in indoor environments. Femtocell BSs are simplified IEEE 802.16 BSs connected to residential broadband access port. Yeh et al [93] studied femtocell capacity for long-term downlink transmissions, considering path loss, slow fading, and interference. Their study showed that the areal capacity, defined as the total bits per second that can be transmitted over a specific area, can be achieved up to 300 times in dense deployments. Such capacity gain comes from the the reuse of the frequency of macro-BSs at femtocell APs and high SINR at users due to little signal attenuation.

The areal capacity gain comes from two aspects. The first one is that Femto-APs reuse the same bandwidth as the , so the available bandwidth per unit area increases. The other reason for higher spectrum efficiency is that most users associated with Femto-APs experience little signal attenuation, which results in high SINR and correspondingly high SE for these users.

2.2.5 Survey Summary and Open Issues

As pointed out by Ahmed [97], knowledge of capacity is critical in wireless networks in order to control network resources effectively. Existing capacity studies can be generally categorized into two groups: model-based and measurement-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

based approaches. Model-based approaches analyze capacity according to specific assumptions on interference models, network and traffic patterns. They provide capacity scaling laws for networks which could be guidelines for network design and deployment. However, model-based approaches may not be able to provide real-time information in a dynamic environment for network control. Measurement-based approaches, on the other hand, can provide real-time capacity information based on actual measurements. The problem of measurement-based approaches is the accuracy of the measurement results. Active measurement schemes, through probing messages, inject additional packets and introduce measurement inaccuracy. Passive measurement methods, through monitoring the performance of existing traffic, may also incur measurement inaccuracy because the measurement module may not be able to obtain enough samples. Therefore, there is a need to investigate a capacity estimation solution that can provide accurate capacity information for real-time network control and resource management.

2.3 Optimal Relay Selection in Multi-hop Wireless Networks

The relay selection scheme has a critical impact on the performance of application flows for individual users and the system. According to different design purposes for different networks, the objectives of optimal relay selection strategies fall into several categories: (1) maximizing the lifetime of wireless networks, (2)

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

maximizing the data rate, and (3) minimizing the number of relay nodes. The literature survey of existing relay selection schemes is presented in the following sections.

2.3.1 Relay Selection Schemes in Wireless Networks

Earlier, much research effort has been made to develop relay node selection strategies in various wireless networks with different application scenarios.

2.3.1.1 Wireless Sensor Networks

Researchers have explored the relay node placement problem for wireless sensor networks. Some focused on maximizing the lifetime of wireless networks, such as the study from Hou et al [98], Xin et al [99], Himsoon et al [100], and Wang et al [101]. Other researchers targeted at maximizing the data rate received at the BS, as in Guo and Huang's study [102], and minimizing the number of relay nodes, as in Wang et al's study [103]. Xin et al proposed the use of simple relay nodes to alleviate the energy bottleneck issue [99], where computational complexity of relay nodes are considered to be even simpler than sensor nodes. Himsoon et al investigated the cooperative diversity issues in wireless sensor networks by considering both location and energy criteria [100]. Guo et al [102] investigated the problem of optimal relay node placement for two-tiered wireless sensor networks to maximize the data received at the BS and to prolong the lifetime of the network subject to a given bit error rate constraint on each node.

Unlike the schemes applied in sensor networks, energy saving is not the focus

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of this thesis which targets to find the optimal relay node for improved network performance.

2.3.1.2 Cellular Networks and WLANs

The relay node placement problem has also been studied in cellular networks and WLAN. Lin and Hsu [104] introduced the concept of multi-hop cellular networks based on IEEE 802.11 technology. They analyzed the system throughput by modeling the packet departure process in the system as a renewal process. They concluded that the system throughput can be improved by using peer mobile nodes to relay data between BS and mobile nodes, when compared to the single hop cellular network. The results are consistent with Hossain et al's findings [105]. Hossain et al [105] compared the delay and throughput performance of mobile cellular networks with and without relaying capabilities and concluded that using relay stations can indeed improve the system throughput and delay performance.

Long and Hossain [106] presented an overview of different relay architectures and research challenges in multi-hop cellular networks. Wu et al proposed the iCar system [107] for balancing of network load and improving system performance by using relay nodes to forward traffic from a congested cell to less congested neighboring cells. So and Liang [24,108] explored different relay strategies and the effects of relay node placement on the expected throughput capacity of WLAN. The benefits of multi-hop cellular networks for capacity enhancement have been evaluated by Mukherjee and Viswanathan [109].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.3.1.3 IEEE 802.16 Mesh Networks

Related work on route selection in IEEE 802.16 mesh networks has been reported in the literature, such as the study of Wei et al [110], Tao et al [111], Fu et al [112], and Nahle et al [113]. They studied generic algorithms to construct routing trees based on IEEE 802.16 centralized scheduling, including the constructing and adjustment of routing tree, to achieve higher end-to-end throughput. Wei et al [110] defined a blocking metric for route selection, which is the total number of one-hop neighbors of the nodes (source, destination and relay nodes) along a path, thus reflecting the level of transmission interference along the path. The algorithm to compute the least interference route is similar to the shortest path routing algorithm.

Taking full advantage of spatial reuse, the approaches of Tao et al [111] and Fu et al [112] to construct routing trees are similar to Wei et al's approach [110], except that the first two groups of researchers used the number of interfering links as the route metrics, whereas Wei et al [110] used the number of interfering nodes as the routing metrics. Besides, Jiao et al [114] emphasized on routing tree construction in IEEE 802.16 mesh networks using various metrics such as hop count, modulation order, energy per bit and interference level.

Shetiya and Sharma [115] also studied the problem of tree-based routing with centralized scheduling for IEEE 802.16 mesh networks, targeting to achieve minimum end-to-end delay for flows. In their study, per hop delay was viewed as the link cost to form routing trees using shortest path algorithms.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.3.1.4 IEEE 802.16 Multi-hop Relay Networks

Recently, research on IEEE 802.16 multi-hop relay networks, which is this thesis's focus, has emerged. Soldani and Dixit [116] summarized the use cases of IEEE 802.16 multi-hop relay networks. Pabst et al [117] presented an overview of multi-hop relay networks and their analytical results showed that range extension and throughput enhancement resulting from using RSs are substantial. As pointed out by Pabst et al [117], the number of tiers in a forwarding path is bounded by the capacity available from the BS, the capacity provision for the users in a specific service area, and the end-to-end delay requirements. In general, it is advisable not to exceed two tiers of relays in networks, as smaller topologies are preferred in most deployments, for lower complexity and lower cost in the overall system [118–120].

Theodoros et al [121] measured the achievable throughput between an SS and a BS at different distances. Lin et al [122] studied the single RS placement problem for static SSs in IEEE 802.16j networks. They assumed the Two-Ray Ground path loss model and incorporated cooperative relaying technology such as Decode-forward (D-F) or Compress-Forward (C-F) to formulate the RS placement problem into an optimization problem. Lo et al [123] proposed an optimal relay selection algorithm, assuming partial decode-and-forward transmission in the system. In their proposed framework, the source node applies a layered coding strategy, which allows each receiving node to partially decode the message. Another study attempted to find the set of BSs and RSs that accommodates user

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

demands at the lowest cost [124]. Fu et al [125] studied the deployment of RSs in the Manhattan-like environment and compared the cell capacity using different radio resource reuse schemes.

Li and Sampalli [126] studied intra-cell mobility and its impact on resource consumption, taking link adaptation into consideration. They identified that the time-varying distance between a user and the BS plays a major role in the resource consumed by the user. Ann et al [127] proposed a path selection scheme, which considers link available bandwidth, signal-to-noise ratio and path hop count.

Adaptive transmission has been adopted by radio resource management schemes for effective and efficient network control. In adaptive transmission, the transmission mode used between two stations is determined by estimating the SNR at the receiver and mapping this to one of the thresholds specified. As the path loss model that this thesis assumes is distance dependent, the maximum distance between stations in different transmission modes must be known in adaptive transmissions. The following section provides the existing empirical studies on adaptive transmission in wireless networks.

2.3.2 Empirical Studies on Adaptive Transmission in Wireless Networks

Researchers have studied the relationship between link capacity and node distance when networks operate in different frequency bands, for example, Hoymann [81] focused on 5 GHz frequency band, Hillestad et al [74], Bian et al [128],

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

and Joseph and Martens [129] focused on 3.5 GHz frequency band. The maximum distance between two stations depends on the transmission mode adopted by the stations. The IEEE 802.16 standard [2] proposes switching points between transmission modes according to the SNR at the receiver. Based on this, Hoymann [81] calculated the distance between stations as follows:

$$d = \frac{\lambda * 10^{\frac{P_t[dBm]-SNR[dB]-N[dBm]}{20}}}{4\pi} \tag{2.2}$$

where λ is wavelength, P_t is transmission power, N is noise level.

According to equation (2.2), Hoymann [81] calculated the switching points, in terms of node distance, between transmission modes in 5 GHz operation frequency with 20 MHz channel bandwidth. Using the similar approach, Hillestad et al [74] suggested the maximum transmission range between two stations in 3.5 GHz operation frequency with 20 MHz channel bandwidth. In comparison with Hoymann's results [81] in 5 GHz frequency band, Hillestad's results [74] in 3.5 GHz frequency band indicated that transmission over the lower frequency band achieves longer range when the same transmission mode is adopted by two communicating nodes.

Joseph and Martens [129] have studied the coverage and throughput in IEEE 802.16 fix networks. They focused on the path loss models for different terrains in 3.5 GHz band. Assuming the transmission coverage ratio being at least 90%, they studied the PHY throughput and the coverage range of various transmission modes considering different path loss models.

As this thesis focuses on the optimal relay selection problem in vehicular

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

networks for transport systems, such as highways, railways, near-land rivers, and seas, the related work on vehicular networks are presented in the subsequent section.

2.3.3 Related Work on Vehicular Networks

Vehicular communications have attracted much research interest and effort in recent years. Biswas et al [130] provided an overview of inter-vehicle communications for highway cooperative collision avoidance. They discussed the current state and challenges of the research in protocol design in vehicular communications. Wu et al [131] studied various configurations of vehicular networks and discussed their performance using the metrics like coverage, hop limit and traffic load. Besides the research characterizing the challenges of vehicular networks, much effort has been put on connectivity and routing performance of such networks.

Chen et al [132] have conducted a thorough survey on various routing schemes in vehicular networks, including geocast, broadcast, multicast and unicast routing protocols. Artimy et al [133] conducted a study on the minimum transmission range that is required to maintain the connectivity in vehicular ad hoc networks, considering different road patterns. Santi and Blough [134], Bettstetter [135], and Wan and Yi [136] studied the minimum transmission range that ensures formation of a connected network using probability analysis and graph theory. They assumed that nodes are independently and uniformly distributed in a specific region.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Namboodiri et al [118] studied the connectivity characteristics in vehicular networks and evaluated the effectiveness of the prediction-based routing scheme using simulations. In their study, link lifetime is used as a prediction metric for preemptive actions to prevent path breakage. Fiore et al [137] studied vehicular mobility patterns and their impact on the network topology and connectivity using metrics such as link duration, nodal degree and the number of clusters. Yuen [138] has shown that both the fraction of connection time and data rate increase with node mobility in a mobile infostation network in highway environments, where renewal reward process was applied to get the long-run average data rate. Based on a generic model for random graphs, Hekmat and Mieghem [139] analyzed node connectivity assuming log-normal shadowing radio propagation model, which takes into account the statistical dynamics of radio signals caused by obstructions and irregularities in the surroundings of the transmission path.

Some researchers proposed group-based routing approaches to improve network performance in vehicular networks [22, 140–142]. Chen et al [141, 142] introduced the local peer group (LPG) architecture. Yang et al [22] designed a multi-hop cluster-based protocol to increase the end-to-end throughput while ensuring fairness guarantee among segments in IEEE 802.16 vehicular networks. They calculated the optimal transmission time allocated for each segment with mesh configurations.

Heterogeneous wireless networks for vehicular communications have also been explored. Hung et al [143] proposed a routing protocol for heterogeneous vehicular wireless networks. They calculated link lifetime based on mobility pat-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

terns and a vehicle-RSU (road side unit) route can then be switched accordingly between IEEE 802.11 and IEEE 802.16e networks for reliability. Preemptive handoff schemes for heterogeneous wireless networks have been proposed by Ling et al [144] for selecting the best available network interface to achieve optimal network capacity in vehicular networks.

The discussion of various relay and path selection schemes is summarized in Table 2.3, which considers network type, deployment, design objective, mobility, and traffic pattern.

2.3.4 Survey Summary and Open Issues

Relay selection solutions have been reviewed in this chapter, which are based on different design objectives including network lifetime maximization, end-to-end throughput maximization and the number of relay node minimization. Empirical studies on adaptive transmission in wireless networks, which have been used for relay selection, are also reviewed. Since this thesis focuses on the optimal relay selection problem in vehicular networks, the state-of-art on vehicular network research is presented. As generic relay selection solutions may not work effectively and efficiently in a specific application scenario, this thesis focuses on the relay selection solution in vehicular environments with highway mobility pattern.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.4 Summary

Real-time capacity information is important for supporting radio resource management in resource-limited and distributed multi-hop wireless networks. Since the direct adaptation or extension of the capacity measurement schemes in wired networks to multi-hop wireless networks appears unfeasible due to co-channel interference, much work has been done to study capacity for wireless networks. In this chapter, a comprehensive survey has been conducted on the existing capacity study solutions for multi-hop wireless networks. Despite the numerous studies available, a satisfactory solution for capacity study is still not evident for effective radio resource management and there remain a number of issues and open problems that require further investigation and research, such as accuracy, responsiveness, scalability, bandwidth consumption, and so on. In this research, the responsiveness and accuracy issues will be highlighted in the capacity study.

Optimal relay selection is critical for achieving good end-to-end throughput performance in relay-based multi-hop wireless networks. Although much effort has been put on the relay selection schemes, the solution specifically for vehicular networks with high-way mobility pattern is still an open issue. This thesis focuses on such an application scenario and explores the optimal relay selection problem for IEEE 802.16 multi-hop relay vehicular networks. Based on the literature survey, the research question is raised and solutions will be proposed in the following chapters.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 2.1: Comparison of model-based approaches in capacity study.

	Gupta and Kumar [30]	Dousse et al [47]	Franceschetti et al [49]	Kodialam and Nandagopal [50]	Agarwal and Kumar [39]	Gomez and Campbell [58]	Kumar et al [53]	Kyasanur and Vaidya [61]
Networks type	Arbitrary and random	Random	Random	Arbitrary	Random	Random	Arbitrary	Arbitrary and random
Interference model	Protocol and Physical	Physical	Physical	Protocol	Protocol and Physical	Protocol	Protocol	Protocol
Transmission power	Fixed	Fixed	Fixed	Fixed	Fixed	Optimally selected	Variable	Fixed
Scheduling schemes	Optimal	Optimal	Optimal	Static and dynamic	Optimal	Optimal	Link-flow scheduling	Optimal
Mobility	No	No	No	No	No	Yes	No	No
Other issues	N.A.	N.A.	N.A.	Multiple channels	N.A.	N.A.	N.A.	Multiple channels
Remarks	Scaling laws in terms of node density	Trade-off between connectivity and capacity	Capacity scaling in random networks converges to arbitrary networks in the high attenuation regime.	Jointly optimize routing, link channel assignments and scheduling to derive capacity bounds.	Protocol Model is more restrictive than Physical Model in capacity bounds.	Average capacity per node keeps constant regardless of node density.	Maximum e2e capacity is studied considering fairness together with routing and scheduling.	Transport capacity scaled down to $1/\sqrt{k}$ compared to Gupta and Kumar's result (k is ratio of number of channels to interfaces).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 2.2: Comparison of measurement-based approaches in capacity study.

	Li et al [69]	Yang and Kravets [11]	Shin and Schulzrinne [72]	Hillestad et al [74]	Ahn et al [77]	Yeh et al [93]	Cicconetti et al [92]
Networks type	IEEE 802.11	IEEE 802.11	IEEE 802.11	IEEE 802.16	IEEE 802.11	IEEE 802.16	IEEE 802.16
Deployment	Chain and regular mesh	Random	Heterogeneous star	Heterogeneous star	Random	Indoor star	Chain and star
Transmission power	Fixed	Variable	Fixed	Fixed	Fixed	Variable	Fixed
Mobility	No	Yes	No	No	Yes	No	No
Traffic pattern	Locality of communication preferred in achievable capacity	Random CBR traffic	CBR and VBR VoIP traffic	VBR video streaming	Mixed TCP and UDP flows	Mixed TCP and UDP flows	TCP flows
Remarks	Compared to one-hop capacity, $\frac{1}{7}$ in chain topology and $\frac{1}{17}$ in regular mesh	Interference and contention inside carrier-sensing range is considered.	VoIP capacity is measured through testbed, simulation and theoretical studies.	Video on Demand capacity is measured with typical cinematic configuration	End-to-end capacity is detected using probing messages.	Areal capacity can be increased significantly in dense indoor deployments.	Capacity is studied and allocated based on fairness considerations.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 2.3: Comparison of relay selection schemes.

	Himsoon et al [100]	Guo and Huang [102]	Wang et al [103]	Lin and Hsu [104]	Lin et al [122]	Wei et al [110]	Yang et al [22]
Networks type	WSNs/WPANs	WSNs	WSNs	Cellular and IEEE 802.11	IEEE 802.16j	IEEE 802.16 mesh	IEEE 802.16 tree
Deployment	Tree	Two-tier, two-hop	Random	Tree	Single RS placement	Tree	Tree
Objective	Maximizing lifetime	Maximizing the data received at BS	Minimizing number of relay nodes	Maximizing system throughput	Maximizing achievable rate	Minimizing interference in the path	Increase end-to-end throughput while maximizing fairness level
Mobility	No	No	No	No	No	No	Vehicle speed follows a Gaussian distribution.
Traffic pattern	Multiple source one destination	Multiple source one destination	Multiple source one destination	Multiple source multiple destination	Multiple nodes to/from one node	Multiple nodes to/from one node	Multiple nodes to/from one node
Remarks	Considering cooperative transmission and energy constraint	Placing the relay nodes to the optimal locations to best serve edge nodes	Considering cost efficiency in the optimal problem	Considering adding relay nodes in cellular networks to improve throughput	Considering cooperative relay strategy	Considering the number of interfering nodes in the path	Ensuring fairness guarantee in bandwidth usage among road sections

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 3

Capacity Estimation in IEEE 802.16 Wireless Mesh Networks

This chapter discusses the issues in capacity estimation and presents solutions for link capacity estimation. The solutions are applied to real-time estimation of capacity for end-to-end flows in IEEE 802.16 wireless mesh networks.

3.1 Introduction

We first describe a distributed solution for determining the link capacity between two neighboring IEEE 802.16 mesh nodes. Subsequently, an interference-aware scheme based on *bottleneck zone* concept is proposed to obtain real-time end-to-end capacity for a given route, considering different network deployment patterns. We define the term *capacity* as the available bandwidth that can be used for future data transmission, unless specified otherwise. Our research work is distinguished from earlier studies of Gupta and Kumar [30], Stuedi and Alonso [44], and Jain et al [41] in that we focus on the real-time capacities at a given link and path in a dynamic system where traffic loads may vary with time. The following sections present the assumptions and background in capacity estimation, link ca-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

pacity estimation, and real-time end-to-end capacity estimation in IEEE 802.16 mesh networks.

3.1.1 IEEE 802.16 Mesh Network Implementation

Figure 1.2 shows a typical IEEE 802.16 wireless mesh network, which is usually deployed in two tiers, i.e. an access tier, where a customer's device accesses the network through a mesh node, and a backhaul tier, where mesh nodes connect the access tier to the wired gateway. Our research focuses on the backhaul tier in a wireless mesh network. For traceability of our research results, we developed our solutions in capacity estimation based on the following implementations of IEEE 802.16 standard:

1. Mesh nodes are connected by bi-directional links (symmetric transmission) such that IEEE 802.16 MAC control messages can be heard by both of two neighboring mesh nodes over the link. Hence, these two neighboring nodes are able to schedule data transmissions between them.

2. The coordinated distributed scheduling method in IEEE 802.16 standard is used to schedule data transmissions over mesh links.

3. The k-hop interference model as proposed by Sharma et al [29] is used in the analysis of interference among neighboring mesh nodes for capacity estimation. The k-hop interference model is an extension of the Protocol Model used by Gupta and Kumar for capacity analysis [30], whereby all nodes within a hop distance of k from the receiver are considered as interfering nodes.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4. Mesh nodes are static or normadic, which change their locations only occasionally, if at all. This is a necessary assumption for real-world implementation of mesh networks, since IEEE 802.16 mesh operation does not provide mobility support.

We adopt the same definitions of neighbor and extended neighbor in IEEE 802.16. Hence, any station with a direct link to a node is a neighbor of that node. A node's neighbor is considered to be "one hop" away from the node, and neighbors of a node form a neighborhood. A neighbor's neighbor is an extended neighbor and is a two-hop neighbor.

3.1.2 IEEE 802.16 Mesh Mode Operation

The IEEE 802.16 MAC is TDMA-based, and transmissions take place in allocated minislots in frames that are synchronized through a ranging process during the network entry procedure for a node. A frame is divided into control and data subframes. Control messages and data packets are transmitted in control subframe and data subframe, respectively. There are two types of control subframes in mesh mode: network control and schedule control subframes. The network control subframe facilitates frame synchronization and the initial network entry for a new node. The schedule control subframe coordinates the scheduling of data transmissions between mesh nodes. A simplified mesh frame structure showing the schedule and network control subframes is illustrated in Figure 3.1. Detailed descriptions of the frame structure are presented in [2] and [80].

In IEEE 802.16 mesh mode, each SS can transmit packets to another SS,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

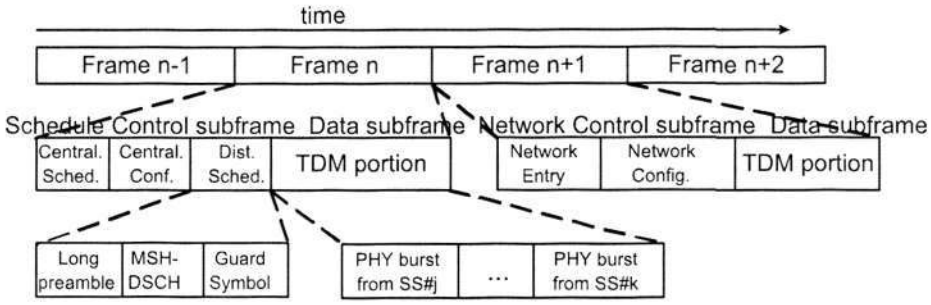


Figure 3.1: IEEE 802.16 MAC mesh frame structure.

possibly relayed by one or more intermediate SSs. The standard defines two scheduling mechanisms for data transmissions in mesh mode: centralized and distributed scheduling. In centralized scheduling, the BS determines the sharing of the channel among multiple SSs to avoid collisions over the links in the routing tree between the BS and SSs. The scheduling procedure is simple, but the time required to set up connections is long. In distributed scheduling, on the other hand, each SS schedules transmissions in a distributed manner without the control of the BS. Data minislots are allocated based on three-way handshaking between the sender and receiver. Distributed scheduling is more complex than centralized scheduling, however, it is also more flexible and efficient in mesh mode operation. In coordinated distributed scheduling, all nodes (BS and SSs) coordinate their transmissions with their neighbors. In such scheduling, data transmissions are scheduled by the sender SS and the receiver SS. Each SS decides the transmission of control messages from the information received in its schedule control subframe. To avoid collision of control messages, a pseudo-random mesh election algorithm is used [2,87,145]. The "Request-Grant-Grant confirmation" three-way

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

handshake is used in the scheduling of data transmissions between the sender and receiver through the exchange of MSH-DSCH control messages. As shown in Figure 3.2, when a sender requires minislots to transmit a data burst, it computes the required number of minislots and sends a *Request* Information Element (IE) in an MSH-DSCH message to the receiver. The *Request* IE contains information such as the number of minislots required and the number of frames that the minislots are required for (also known as *persistence* [2]). The receiver then checks the availability information (see next subsection) of both the sender and the receiver itself. Based on the availability information, the receiver allocates minislots where both the sender can transmit and the receiver can receive. This allocation is carried in a *Grant* IE contained in an MSH-DSCH message. The *Grant* IE conveys information on the actual granted minislot range and the number of frames required for data transmission. Once a *Grant* is received by the sender, it replies with a *Grant confirmation* in an MSH-DSCH message to acknowledge the receipt of the *Grant*. Subsequently, data packets can be transmitted by the sender in the allocated minislots.

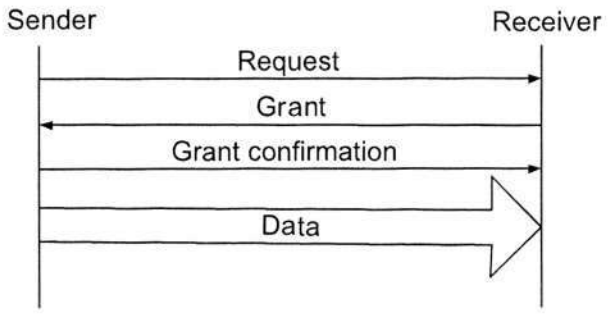


Figure 3.2: Three-way handshake of the coordinated distributed scheduling in IEEE 802.16.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.2 Link Capacity Estimation

In wired networks, the capacity of a link is not affected by data transmission in the adjacent links. Hence, the capacity of a link is simply the difference between the total and occupied bandwidth of the link. However, in wireless networks, the broadcast nature of radio propagation gives rise to interference between nodes transmitting on common channels. Such interference is referred to as co-channel interference and has significant impact on transmission scheduling. A transmission can be successful only if there is no co-channel interference or the co-channel interference is below a threshold level. Co-channel interference affects bandwidth utilization efficiency in mesh networks and poses challenges to capacity estimation. The following subsections first analyze in detail the effects of co-channel interference between neighboring nodes for a set of different transmission scenarios, and then propose a reliable solution for computing the real-time link capacity in IEEE 802.16 mesh networks.

3.2.1 Transmission Scheduling in Wireless Mesh Networks

Figure 3.3 shows a snapshot of network transmissions where the link from s to d is considered as the link of interest. As defined in IEEE 802.16 [2], the neighbors of a node have direct links with this node, and these neighbors are considered to be "one hop" away from the node. In this figure, each of the smaller circles represents the transmission boundary of a sender and all nodes within the smaller circle are neighbors of that sender. The larger circle represents the interference boundary of sender s. Consider the scenario illustrated in the figure, where there

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

are three on-going transmissions in a given minislot of an IEEE 802.16 data subframe, i.e. $A \rightarrow B$, $C \rightarrow G$ and $s \rightarrow d$. To determine the real-time capacity of the link of interest s-d, the conditions are studied under which a transmission in the link of interest can be successful. Suppose sender s is requesting to transmit data to receiver d, then receiver d determines a minislot t_i, which can be granted for the transmission only if all of the following conditions are satisfied:

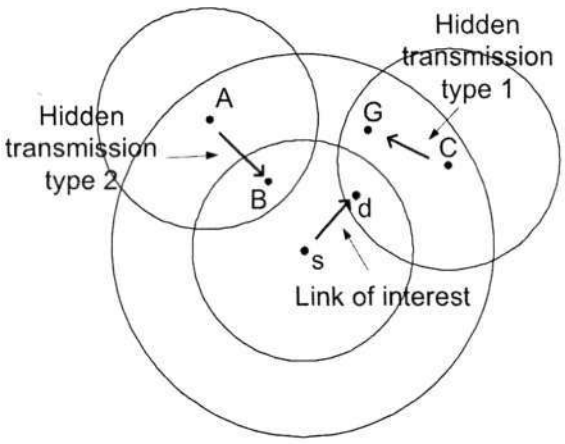


Figure 3.3: The collision/interference scenario in wireless networks.

(a) Node d is not within the interference boundary of any sending node in minislot t_i. A transmission initiated by any neighbor of d (except s) is named as *Hidden Transmission Type 1 (HTT1)*. It is a *hidden transmission* because sender s may not detect this transmission as s is out of the transmission range of d's neighbor (for example, s is out of the transmission range of C in Figure 3.3). If s and C concurrently transmit data, collision will happen at d.

(b) All nodes within the interference range of the sender s (for example, B) are

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

not scheduled to receive in minislot t_i. Any transmission targeted at the neighbors of s is named as as *Hidden Transmission Type 2 (HTT2)*. It is a *hidden transmission* because the sender s may not detect this transmission as s is out of the transmission range of B's neighbor (for example, s is out of the transmission range of A in Figure 3.3). If s and A concurrently transmit data, collision will happen at B.

(c) Both s and d have not scheduled to transmit or receive data in minislot t_i. This condition implies that the minislot t_i is not used by s or d themselves.

Conditions (a) and (b) address the hidden terminal problem in wireless networks. The above conditions conform to the k-hop interference model proposed by Sharma et al [29]. Based on these conditions, a solution is proposed for computing the real-time link capacity in IEEE 802.16 networks.

3.2.2 State Transition of Minislots

IEEE 802.16 [2] defines four states to identify the availability of a minislot. The states are: (1) available for both transmission and reception (state *11*), (2) available for transmission only (state *01*), (3) available for reception only (state *10*), and (4) not available (state *00*). The information on the states of minislots is encapsulated in *Availability* IEs and broadcasted as part of an MSH-DSCH message. When an MSH-DSCH message is received from a neighbor, the local node (or "node of interest") computes the state of availability for each minislot and composes the corresponding *Availability* IEs. However, the mechanism to determine the state of availability of a minislot at a given node is left undefined

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

in the standard. We have developed a state transition model for deriving the states of minislots and used this information to compute the link capacity.

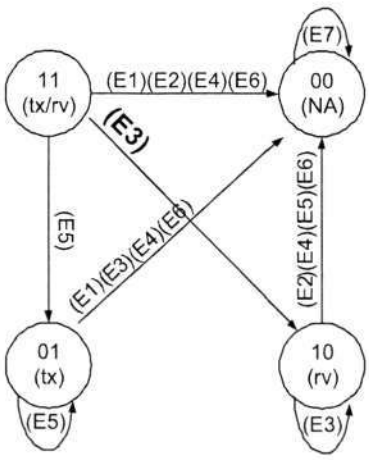
It is expected that the state transition of a minislot is greatly affected by the ratio of interference range to transmission range. Here, we focus on the analysis on the effects of different ratios of interference to transmission range on the state transition of a minislot. The state transition processes can be categorized into two cases: **capacity-aggressive** case (see Figure 3.4) and **capacity-conservative** case (see Figure 3.6). In the former case, the interference range is assumed to be close to the transmission range, and hence the co-channel interference is considered not severe, so that high link utilization can be achieved in the system through greater spatial reuse. In the latter case, the interference range is assumed to be no shorter than twice that of the transmission range, and hence the effect of co-channel interference will result in a lower link utilization in the system.

Figure 3.4 illustrates the process of state transition among the four possible minislot states in the **capacity-aggressive** case, where an incoming event (such as $(E1), (E2), ...$) is the reception of an MSH-DSCH message containing various IEs (such as *Grant* and *Grant confirmation*; see Section 3.1.2 Figure 3.2). Figure 3.5 presents various scenarios causing state transitions in a minislot at node Q, arising from different positions of neighboring nodes A and B relative to Q. Take the example of an incoming event $(E3)$ arriving at a local node Q, which has an initial state *11* (available for both reception and transmission) at minislot t_i. The local node Q then receives a *Grant* from node B[1]. This *Grant* implies

[1]In this case, node B is one of the neighbors of node Q.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

that the minislot t_i has been assigned to a sender node A, which is neither the local node Q nor the neighbor of node Q[2]. Hence, following the state transition process, the state of the minislot t_i will transit to *10*, which indicates that node Q is now available for reception only.



(E1) Receive a Grant for itself.
(E2) Transmit a Grant/receive a Grant confirmation for itself.
(E3) Receive a Grant for others and the granted transmitter is NOT its neighbor (HTT2) (example) .
(E4) Receive a Grant for others and the granted transmitter is its neighbor (HTT1).
(E5) Receive a Grant confirmation for others and the granted receiver is NOT its neighbor.
(E6) Receive a Grant confirmation for others and the granted receiver is its neighbor.
(E7) Receive anything.

Figure 3.4: State transitions of a minislot in an IEEE 802.16 frame at a particular node for the case where the interference range is close to the transmission range (capacity-aggressive case).

Figure 3.6 illustrates the state transition of a minislot at node Q in the **capacity-conservative** case. In this case, as long as a node receives a *Grant* or *Grant confirmation*, the granted transmitter/receiver must be one of the three types of nodes: the local node, the neighbor of the local node, or the two-hop neighbor of the local node[3]. Thus, the local node must be within the interference range of the granted transmitter and receiver and therefore this local node must not transmit or receive so as to avoid interference and collision. This is illustrated

[2]In this case, node A is not the neighbor of node Q.
[3]A neighbor's neighbor is denoted as a two-hop neighbor.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

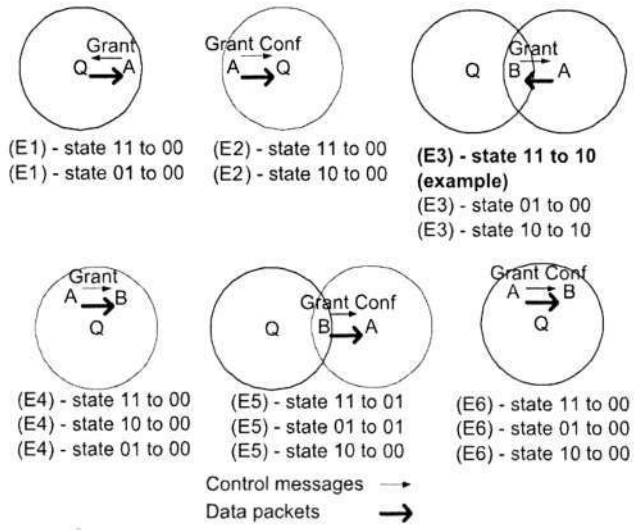


Figure 3.5: Examples of scenarios causing state transitions of a minislot at node Q, with respect to different positions of neighboring nodes A and B relative to Q (capacity-aggressive case).

in Figure 3.6, where node Q receives a *Grant* which implies that minislot t_i has been assigned to sender A. In the capacity-conservative case, both node A and node B are within the interference range of node Q. Therefore, node Q should not transmit or receive and the state of the minislot t_i at node Q will transit to *00*.

In other words, if the interference range is close to the transmission range in

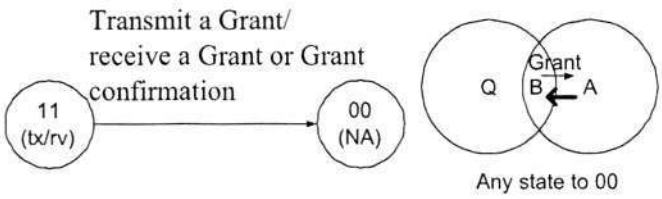


Figure 3.6: State transitions in a minislot of an IEEE 802.16 frame at a particular node Q assuming that interference range is not shorter than twice that of the transmission range (capacity-conservative case).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the neighborhood of the sender, the state transition model of Figure 3.4 is applied to achieve higher capacity in the **capacity-aggressive** case. On the other hand, if the interference range is not shorter than twice that of the transmission range in the neighborhood of the sender, the state transition model of Figure 3.6 is applied to avoid interference in the **capacity-conservative** case.

3.2.3 Link Capacity Computation

The link capacity can be computed in real-time once the state of each minislot in a data subframe is known. Consider the scenario illustrated in Figure 3.3, where the link of interest is the link s-d joining the source node s and the destination node d. To find the capacity on the link s-d, let w be the data transmission rate of link s-d, and N_{data} be the number of data minislots in one MAC frame. Then, $\frac{w}{N_{data}}$ is the data rate per minislot in a MAC frame. Let C_T^s be the capacity available for node s to transmit data. C_T^s is the summation of data rates of all minislots that satisfy conditions (b) and (c) of Section 3.2.1, and it can be determined as follows

$$C_T^s = \frac{w}{N_{data}} \sum_{i=1}^{N_{data}} \prod_{k=1}^{D_s} [1 - \Phi_{ki}^s] \tag{3.1}$$

where

$$\Phi_{ki}^s \equiv \begin{cases} 0, & \text{if } t_i^s \vee r_i^s \vee z_{ki}^s = 0 \\ 1, & \text{otherwise} \end{cases} \tag{3.2}$$

and $k \in [1, D_s]$, $i \in [1, N_{data}]$. We denote D_Q as the number of neighbors of node Q, where $Q \in \{s, d\}$; t_i^Q as the status of the transmission mode of node Q in minislot i ($Q \in \{s, d\}$); r_i^Q as the status of the receiving mode of node Q in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

minislot i ($Q \in \{s, d\}$); z_{ki}^{s} as the status of the receiving mode of neighbor k of node s in minislot i, and y_{ji}^{d} as the status of the transmission mode of neighbor j of node d in minislot i. The meanings of these notations are summarized in Table 3.1.

Table 3.1: Notations of link capacity computation

Notation	Description
D_Q	The number of neighbors of node Q ($Q \in \{s, d\}$)
N_{data}	The number of data minislots in one frame
w	The data subframe transmission rate in link s-d
y_{ji}^{d}	The status of the transmission mode of neighbor j of node d in minislot i
z_{ki}^{s}	The status of the receiving mode of neighbor k of node s in minislot i
t_i^Q	The status of the transmission mode of node Q in minislot i ($Q \in \{s, d\}$)
r_i^Q	The status of the receiving mode of node Q in minislot i ($Q \in \{s, d\}$)
C_T^s	The capacity available for node s to transmit data
C_R^d	The capacity available for node d to receive data

Note that t_i^Q, r_i^Q, z_{ki}^{s} and y_{ji}^{d} are binary indicator variables with 1 and 0 representing presence and absence of the condition described in their definitions, and are derived from state transition of minislots at the local and neighboring nodes. In expressions (3.1) and (3.2), the value of Φ_{ki}^{s} is zero when the conditions (b) and (c) of Section 3.2.1 are satisfied at node s in minislot i.

Similarly, let C_R^d be the capacity available for node d to receive data. C_R^d is the summation of data rates of all minislots that satisfy conditions (a) and (c) of Section 3.2.1, and it can be determined as follows

$$C_R^d = \frac{w}{N_{data}} \sum_{i=1}^{N_{data}} \prod_{j=1}^{D_d} [1 - \Lambda_{ji}^{d}] \tag{3.3}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

where

$$\Lambda_{ji}^{d} \equiv \begin{cases} 0, & \text{if } t_i^d \vee r_i^d \vee y_{ji}^d = 0 \\ 1, & \text{otherwise} \end{cases} \tag{3.4}$$

and $j \in [1, D_d]$, $i \in [1, N_{data}]$. In expressions (3.3) and (3.4), the value of Λ_{ji}^{d} is zero when the conditions (a) and (c) of Section 3.2.1 are satisfied at node d in minislot i.

Note that node s can transmit data to node d in a specific minislot only if this minislot is available for both transmissions at node s and receptions at node d. We then compute the available link capacity $C^{s,d}$ on link $s \rightarrow d$ as follows

$$C^{s,d} = \frac{w}{N_{data}} \sum_{i=1}^{N_{data}} (Z_i^s \times Y_i^d) \tag{3.5}$$

where $Z_i^s \equiv \prod_{k=1}^{D_s}[1 - \Phi_{ki}^s]$ and $Y_i^d \equiv \prod_{j=1}^{D_d}[1 - \Lambda_{ji}^d]$.

In link capacity computation, all the available data minislots are assumed to be usable for future data transmissions. In practice, data minislots in a subframe may become "fragmented" due to inefficient selection or suboptimal assignment of minislots[4], which will result in isolated minislots which are unusable. In such case, the actual available link capacity will be reduced.

3.3 End-to-end Capacity Estimation

In a wireless mesh network, packets in a flow are forwarded from the source to the destination through a dedicated route comprising of a series of nodes. A node on the route may be interfered, either by its upstream and downstream

[4]Design of optimal data scheduling algorithm to achieve maximum channel utilization in TDMA MAC is our on-going work and it is outside the scope of the thesis.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

nodes along the route[5] or other nodes which are not on the route but in the vicinity of this node[6]. To analyze and develop solutions for end-to-end route capacity estimation, we adopt the k-hop interference model proposed by Sharma et al [29], which is an extension of the Protocol Model [30] to cover all possible transmission interference from nodes within the k-hop distance from the receiver node. The following subsections first analyze the issue of concurrent transmissions in wireless networks and then develop a solution for estimating the end-to-end capacity bounds of a route.

3.3.1 Issue of Concurrent Transmissions in Wireless Networks

Let us consider a case that the interference range is twice that of the transmission range in a system (similar case considered by Robinson and Knightly [3]) and a multi-hop route as shown in Figure 3.7, where each node is a transceiver. In the figure, node 1 has a route to node 6, through intermediate nodes 2, 3, 4, and 5, and each circle represents the transmission boundary of a node along the route.

Now, suppose node 1 is transmitting to node 2 in a given set of minislots, then other transmissions in the same minislots must be scheduled in the following manners: (i) Node 2 cannot transmit because it cannot transmit and receive at the same time. (ii) Node 3 cannot transmit because any transmission from node 3 will collide with the transmission from node 1 to node 2. (iii) Node 4 cannot

[5]This type of interference is denoted as intra-flow interference [38].
[6]This type of interference is denoted as inter-flow interference [38].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

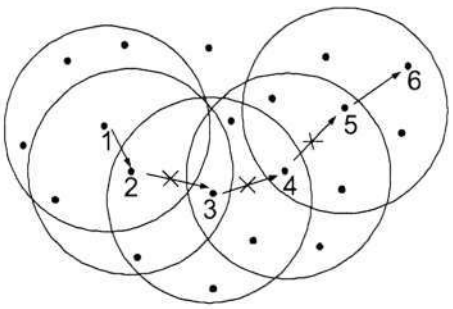


Figure 3.7: Flow scheduling in a multi-hop route (assuming that the interference range is twice that of the transmission range).

transmit because its transmission will interfere with the transmission from node 1 to node 2, since node 2 is within the interference range of node 4. (iv) Node 5 can transmit because its transmission will neither collide nor interfere with the transmission from node 1 to node 2, since node 2 is out of the interference range of node 5.

From the above analysis, it can be seen that, based on the assumption that interference range is twice that of transmission range, no more than one link can be active for any four successive links on a route in a given data minislot. Camp et al [46] made the same assumption and measured the achievable rate for a flow transmitted over multiple hops and found that the multi-hop throughput decreases more gradually when a node is four hops away from a gateway and flows can have spatial reuse at this point.

We can now generalize the above case by defining the interference range as a ratio of its transmission range; i.e. the ratio of interference range to transmission range to be N in a system, where $N \geq 1$. This concept is similar to the K-hop interference model in [29]. In this model, every node has the same transmission

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

range a; the interference range of a node is then $N \times a$. As shown in Figure 3.8, a route from node 1 to node $H+1$ comprises of H hops, with relay node $2, 3, \ldots,$ and H in sequence. In the figure, node 1 is transmitting packets to node 2 in minislot t. In order to ensure that such a transmission is successful, node 2 must be outside the interference boundary of any node in the path which is supposed to transmit packets in minislot t.

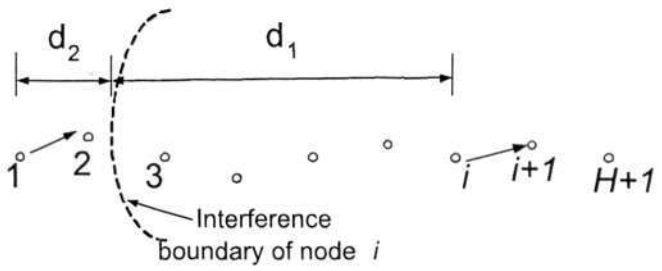


Figure 3.8: Valid concurrent transmissions in a multi-hop path.

In Figure 3.8, let node i be the next transmitting node along the path that will not interfere with the reception at node 2, and permit spatial reuse of the wireless channel. Without loss of generality, we can conclude that along the path from node 1 to node $H+1$, only one link can be active within the distance of $d_1 + d_2$, where d_1 is the interference range of node i ($d_1 = N \times a$), and d_2 is the distance of node pair 1-2 plus the distance between node 2 and the interference boundary of node i. Now, let $h(d_1)$ and $h(d_2)$ be the number of hops along the path within distances d_1 and d_2, respectively. Then, following our earlier descriptions of Figure 3.8, $h(d_2)$ can be expressed as

$$1 < h(d_2) \leq 2 \tag{3.6}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

From Vural and Ekici's analysis [146], as the expected one-hop distance changes with node density in the network, $h(d_1)$ is a random variable. They have computed the upper and lower bounds on the number of hops between two nodes for a given node distance x assuming that nodes are uniformly distributed in a specified area. From their results, the number of hops between two nodes with distance x, $h(x)$, is bounded by

$$\lfloor \frac{x}{a} \rfloor < h(x) \leq 2\lfloor \frac{x}{a} \rfloor + 1 \tag{3.7}$$

Since $d_1 = N \times a$, from expression (3.7), the following expression can be obtained

$$\lfloor N \rfloor < h(d_1) \leq 2\lfloor N \rfloor + 1 \tag{3.8}$$

From expressions (3.6) and (3.8),

$$\lfloor N \rfloor + 1 < h(d_1) + h(d_2) \leq 2\lfloor N \rfloor + 3 \tag{3.9}$$

When the network is dense or the nodes are optimally deployed, the average one-hop distance approximates the node transmission range[7]. In this case, $\lfloor N \rfloor + 2$ is the smallest integer that meets the requirement of $h(d_1)+h(d_2) > \lfloor N \rfloor + 1$. Thus, when the network is dense or the nodes are optimally deployed, $h(d_1)+h(d_2)$ can achieve a lower bound of $\lfloor N \rfloor + 2$.

From the above analysis, we can conclude the following: (1) if nodes are densely or optimally deployed, two concurrent transmitters are $\lfloor N \rfloor + 2$ hops apart; (2) if nodes are randomly deployed, the hop distance between two concurrent transmitters has an upper bound of $2\lfloor N \rfloor + 3$.

[7]When the network is dense, shortest path routes with minimal hops are more achievable between two nodes; when nodes are optimally deployed, the number of nodes deployed can be minimized in order to achieve a predetermined network coverage. In both cases, the average one-hop distance approximates the node transmission range.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.3.2 Transmission Bottlenecks in Wireless Networks

We introduce in this subsection a few new concepts and definitions used in the derivation of a solution for computing end-to-end capacity bounds in wireless mesh networks.

Definition 1: A ***Weakest Link (WL)*** is defined as the link which has the lowest (available) capacity on a route, where capacity is given by equation (3.5) in Section 3.2. There may be more than one WLs on a route.

Definition 2: A ***Bottleneck Zone (BZ)*** on a route is defined as a group of successive links, including at least one WL, among which only one link can be active for transmissions. In a BZ, concurrent transmissions are not allowed.

Definition 3: A ***Bottleneck Region (BR)*** on a route is defined as an area that contains all Bottleneck Zones (BZs) sharing the same WL.

These definitions are further described and illustrated in Figure 3.9, based on the assumption of $\lfloor N \rfloor = 2$. As illustrated, a route from node 1 to node 9 contains seven intermediate nodes $2, 3, \ldots$, and 8. Let $l_{i \to i+1}$ denote the link from node i to node $i+1$ and hence link $l_{5 \to 6}$ is identified as the WL on the route. The Bottleneck Zones (BZs) denoted by BZ_1, BZ_2, BZ_3 and BZ_4 are those that cover the WL $l_{5 \to 6}$. The Bottleneck Region (BR) is one that covers all the BZs that share the same WL and is composed of seven links: $l_{2 \to 3}$, $l_{3 \to 4}$, $l_{4 \to 5}$, $l_{5 \to 6}$, $l_{6 \to 7}$, $l_{7 \to 8}$ and $l_{8 \to 9}$. It is illustrated in Figure 3.9 that the BR covers links along the transmission path and some of those links may interfere or collide with its WL $l_{5 \to 6}$.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

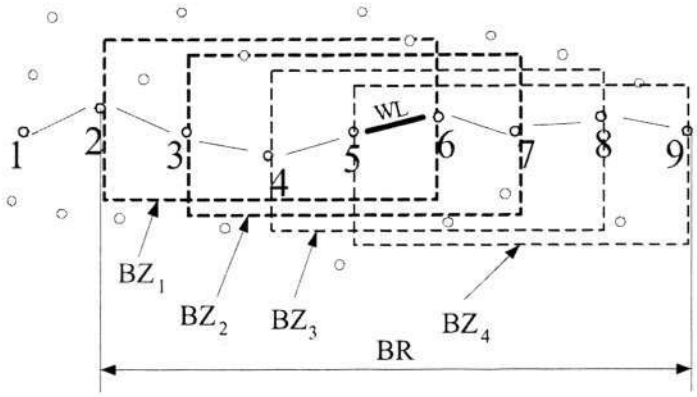


Figure 3.9: An example of *Bottleneck Zones* and *Bottleneck Region* (assuming $\lfloor N \rfloor = 2$).

Vural and Ekici's study [146] has shown that the expected one-hop distance varies with the node density of the network and tends towards the transmission range when nodes are densely or optimally deployed. Following their study, we consider two deployment scenarios in our capacity analysis:

1. ***Scenario I***, where nodes are densely or optimally deployed in the network.

2. ***Scenario II***, where nodes are randomly deployed in the network with various node densities[8].

The end-to-end capacity bounds for a given route are computed based on the Bottleneck Zone (BZ) concept. Let the number of links (hops) on a route be H; the link with index i be $l_i, i \in [1, H]$; the data transmission rate of link l_i be r_i; the available capacity of link l_i be C_i; the number of the Weakest Links

[8] Actually, Scenario I is a special case of Scenario II.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(WLs) on the route be M; the index of the WL be w_m, where $m = 1, 2, \ldots$, and M; the duration of a frame be T_f; the duration of a data minislot be T_μ; the number of data minislots in a frame be N_{data}; the number of data minislots that have been utilized per frame in the WL l_{w_m} be $N^u_{w_m}$; the number of data minislots required to transmit P bits over link l_i be N_i; the set of data minislots assigned to transmit P bits over link l_i be $S(N_i)$; the number of links in one BZ be L_{BZ}; the number of links in one BR be L_{BR}; the number of BZs in one BR be L^Z_{BR}. The notations are summarized in Table 3.2.

Table 3.2: Parameters for end-to-end capacity computation

Notation	Description
H	The number of link (hops) in a route
N	The ratio of interference range to transmission range.
C_i	The available capacity of link i, where $i \in [1, H]$
M	The number the Weakest Links (WL) in the route
w_m	The index of the Weakest Link (WL), where $m = 1, 2, \ldots$, and M
r_i	The data transmission rate of link i
l_i	The link with the index i
T_f	The duration of a frame
T_μ	The duration of a data minislot
N_{data}	The number of data minislots in a frame
$N^u_{w_m}$	The number of data minislots that have been utilized per frame in the WL w_m
$S(N^u_{w_m})$	The set of data minislots that have been utilized per frame in the WL w_m
N_i	The number of data minislots required to transmit P bits over link i
$S(N_i)$	The set of data minislots assigned to transmit P bits over link i
L_{BZ}	The number of links in one BZ
L_{BR}	The number of links in one BR
L^Z_{BR}	The number of BZs in one BR

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.3.3 End-to-End Capacity Bounds in Dense Networks or Optimally Deployed Networks

In ***Scenario I***, which considers that nodes are densely or optimally deployed in a network, the expected one-hop distance is known to be close to the node transmission range in the network.

Lemma 1: *In Scenario I, the number of links in one BZ is expressed as* $L_{BZ} = \min\{H, \lfloor N \rfloor + 2\}$.

Proof: From the analysis in Section 3.3.1, in Scenario I at most one transmission can happen in successive $\lfloor N \rfloor + 2$ links on a route. If the number of hops on the route H is fewer than $\lfloor N \rfloor + 2$, at most one transmission can happen on this route. Therefore, the number of links in one BZ is expressed as $L_{BZ} = \min\{H, \lfloor N \rfloor + 2\}$.

Lemma 2: *In Scenario I, the number of links in one BR is bounded by* $\min\{H, \lfloor N \rfloor + 2\} \leq L_{BR} \leq 2\lfloor N \rfloor + 3$.

Proof: The proof is as follows: (1) Lower bound: If $H \leq \lfloor N \rfloor + 2$, there is only one BZ in the BR, $L_{BR} = H$; if $H > \lfloor N \rfloor + 2$ and the WL index $w_m = 1$ or $w_m = H$ (the WL w_m is the first or last link on the route), $L_{BR} = \lfloor N \rfloor + 2$. Thus, $L_{BR} \geq \min\{H, \lfloor N \rfloor + 2\}$, and (2) Upper bound: Let BZ_1 be the first BZ in the BR, then BZ_1 contains the first link of the BR; let BZ_k be the last BZ in the same BR, then BZ_k contains the last link of the BR. According to the definition of BR, any two BZs in one BR must share links, at least the WL. When the two BZs share the WL only, the number of links in the BR is maximum. In this case, the combined sets of two BZs ($S(BZ_1)$ and $S(BZ_k)$) satisfies $S(BZ_1) \cap S(BZ_k) = l_{w_m}$. As in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Lemma 1, the upper bound of the number of links in one BZ is $\lfloor N \rfloor + 2$. Therefore, the upper bound of the number of links in one BR is $2(\lfloor N \rfloor + 2) - 1 = 2\lfloor N \rfloor + 3$.

Lemma 3: *In Scenario I, the number of BZs in one BR is bounded by* $1 \leq L_{BR}^{Z} \leq \lfloor N \rfloor + 2$.

Proof: In the case of the lower bound in Lemma 2, there is only one BZ in the BR, i.e. $L_{BR}^{Z} = 1$. In the case of the upper bound in Lemma 2, BZ_1 contains the first link of the BR and BZ_k contains the last link of the BR. The WL is the last link of BZ_1 and the first link in BZ_k. Let the index of the first link of the BR be e. The first link l_e and the WL l_{w_m} are the first and the last link in BZ_1 so that $e = w_m - (\lfloor N \rfloor + 1)$. The first link of BZ_k is the WL l_{w_m}. Thus, the maximal number of BZs in one BR can be represented by the index difference of the first links of the two BZs, i.e. BZ_1 and BZ_k. Such an index difference is $w_m - e + 1 = \lfloor N \rfloor + 2$. Therefore, $1 \leq L_{BR}^{Z} \leq \lfloor N \rfloor + 2$.

3.3.3.1 End-to-End Capacity Upper Bound

The capacity bounds are computed based on the Lemmas, starting with the analysis on the upper bound of the end-to-end capacity. Lin [10] has pointed out that in TDMA-based multi-hop wireless networks, the end-to-end capacity is decided by not only the available slots on each link, but also the way the slots are scheduled. Following that, we consider the data minislots allocation for a bottleneck region (BR) in an IEEE 802.16 network where inevitably there will be some minislots that are unavailable in the weakest link (WL), due to the allocations for transmissions in the WL or in neighboring links. In general, some

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of these minislots are reusable for transmissions by other links in the same BR. For example, in the scenario illustrated in Figure 3.9, link $l_{5\rightarrow 6}$ is assumed to be the WL. Hence, transmissions from node 4 to node 5 block some minislots in link $l_{5\rightarrow 6}$, even though those minislots are reusable for transmissions in link $l_{8\rightarrow 9}$, since it is valid for concurrent transmissions over links $l_{4\rightarrow 5}$ and $l_{8\rightarrow 9}$[9]. In the best case, all of these minislots are reusable in the links other than the WL in the BR. Those non-WLs can share the unavailable minislots of the WL for future allocations and the WL can allocate all of its available minislots for future data transmissions. Such case leads to the upper bound of the route capacity.

Theorem 1: *The upper bound of the end-to-end capacity of a route in dense networks or optimally deployed networks is*

$$\frac{r_{w_m} T_\mu (N_{data} - N^u_{w_m})}{T_f} \tag{3.10}$$

where r_{w_m} is the transmission rate of the WL with index w_m.

Proof: The capacity for a Bottleneck Zone (BZ) that covers the WL, l_{w_m}, is first considered. In the best case, all the available minislots of the WL can be allocated to the WL for its own transmissions, instead of being blocked by the transmissions of neighboring links. In such case, the links in the BZ, other than the WL, can re-use the minislots that are unavailable for the WL, i.e. $S(N^u_{w_m})$. Let P be the number of bits that can be transmitted over the BZ. The condition for the best case is that the set of data minislots assigned to transmit P bits over

[9]In this case, the unavailable minislots in link $l_{5\rightarrow 6}$ can be reusable for transmissions by link $l_{8\rightarrow 9}$, and the two links are in the same BR.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

link $l_i(i \neq w_m)$ in the BZ must belong to $S(N^u_{w_m})$, that is

$$S(N_b) \cup S(N_{b+1}) \ldots \cup S(N_{b+L_{BZ}-1}) \in S(N^u_{w_m}) \tag{3.11}$$

where b is the index of the first link in a BZ. The link indexes in expression (3.11) are $b, b+1, \ldots, b+L_{BZ}-1$, excluding the WL index w_m. As in Lemma 1, the number of links in one BZ, L_{BZ}, is $\min\{H, \lfloor N \rfloor + 2\}$ (including the WL). To avoid concurrent transmissions in a BZ, the following condition must be satisfied

$$S(N_b) \cap S(N_{b+1}) \ldots \cap S(N_{b+L_{BZ}-1}) = \emptyset \tag{3.12}$$

where the link indexes are $b, b+1, \ldots, b+L_{BZ}-1$, including the WL index w_m. From expression (3.11), the total number of minislots that can be allocated in the BZ (except the WL) for future transmissions must be less than $N^u_{w_m}$, as

$$\sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} N_i \leq N^u_{w_m} \tag{3.13}$$

As $N_i = \lceil \frac{P}{T_\mu r_i} \rceil$, expression (3.13) can be written as

$$\sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} \lceil \frac{P}{T_\mu r_i} \rceil \leq N^u_{w_m} \tag{3.14}$$

The following expression always holds

$$\sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} \lceil \frac{P}{T_\mu r_i} \rceil \geq \sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} \frac{P}{T_\mu r_i} \tag{3.15}$$

Combining expressions (3.14) and (3.15), the following expression is obtained

$$\sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} \frac{P}{T_\mu r_i} \leq N^u_{w_m} \tag{3.16}$$

Manipulating expression (3.16) leads to the following expression

$$P \leq T_\mu N^u_{w_m} \lfloor \sum_{i=b, i\neq w_m}^{b+L_{BZ}-1} \frac{1}{r_i} \rfloor^{-1} \tag{3.17}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The number of bits that can be transmitted in a frame over the WL can be expressed as

$$P = (N_{data} - N^u_{w_m}) r_{w_m} T_\mu \tag{3.18}$$

If the conditions expressed in (3.11) and (3.12) are satisfied, the maximal capacity for the BZ is

$$C^{max}_{BZ} = \frac{P}{T_f} = \frac{r_{w_m} T_\mu (N_{data} - N^u_{w_m})}{T_f} \tag{3.19}$$

To achieve the maximal capacity for the BZ (the best case), we derive, using expressions (3.17) and (3.18), the following form:

$$r_{w_m} \leq \frac{N^u_{w_m}}{N_{data} - N^u_{w_m}} [\sum_{i=b, i \neq w_m}^{b+L_{BZ}-1} \frac{1}{r_i}]^{-1} \tag{3.20}$$

Expression (3.20) is the condition to achieve the upper bound capacity for the BZ.

Next, we formulate the capacity for a Bottleneck Region (BR) that covers the WL with index w_m, subject to the best case condition that the set of data minislots assigned to transmit P bits over link $l_i (i \neq w_m)$ in the BR must belong to $S(N^u_{w_m})$. That is,

$$S(N_e) \cup S(N_{e+1}) \ldots \cup S(N_{e+L_{BR}-1}) \in S(N^u_{w_m}) \tag{3.21}$$

where e is the index of the first link in a BR and the link indexes in expression (3.21) are $e, e+1, \ldots, e+L_{BR}-1$ excluding w_m. As in Lemma 2, the number of links in one BR, L_{BR}, is bounded by $\min\{H, \lfloor N \rfloor + 2\} \leq L_{BR} \leq 2\lfloor N \rfloor + 3$.

Expression (3.21) implies that the links in the BR, other than the WL, can re-use the minislots that are unavailable for the WL, i.e. $S(N^u_{w_m})$. To avoid

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

concurrent transmissions in any BZ of a BR, similar condition expressed in (3.12) must hold. In the best case, the condition expressed in (3.13) must be satisfied for each BZ in a BR. Given such condition, it is unnecessary to sum up all N_i in expression (3.13) in a BR to ensure the sum is no more than $N^u_{w_m}$. This is because links which are not in the same BZ can allocate the same minislots used in other BZs, due to spatial reuse. If the index of the BZ in a BR is included in expression (3.17), it can be rewritten as

$$P_j \leq T_\mu N^u_{w_m} [\sum_{i=b_j, i\neq w_m}^{b_j+L_{BZ}-1} \frac{1}{r_i}]^{-1} \tag{3.22}$$

where j is the index of the BZ in a BR and b_j is the index of the first link of the BZ j. To achieve the maximal capacity for the BR (the best case), we derive, using expressions (3.22) and (3.18), the following form:

$$r_{w_m} \leq \frac{N^u_{w_m}}{N_{data} - N^u_{w_m}} min_j [\sum_{i=b_j, i\neq w_m}^{b_j+L_{BZ}-1} \frac{1}{r_i}]^{-1} \tag{3.23}$$

Expression (3.23) is the condition to achieve the maximal capacity for a BR, C^{max}_{BR}, which is expressed as

$$C^{max}_{BR} = \frac{P_{max}}{T_f} = \frac{r_{w_m} T_\mu (N_{data} - N^u_{w_m})}{T_f} \tag{3.24}$$

The above derivations can be generalized to a generic case where there are $M (M \geq 1)$ WLs along a route and M BRs which may overlap one another. If in this case, the condition (3.23) is satisfied in every BR along the route, then

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the end-to-end capacity is a maximum, and can be expressed as

$$\begin{aligned} C_{e2e}^{max} &= \frac{r_{w_1} T_\mu (N_{data} - N_{w_1}^u)}{T_f} = \frac{r_{w_2} T_\mu (N_{data} - N_{w_2}^u)}{T_f} \\ &= \ldots = \frac{r_{w_M} T_\mu (N_{data} - N_{w_m}^u)}{T_f} \end{aligned} \tag{3.25}$$

From expression (3.25), it can be seen that the maximum end-to-end capacity is the capacity of the WLs[10]. In order to compute the real-time capacity bound expressed in (3.25), r_{w_m} and $N_{w_m}^u$ must be computed iteratively during run-time, whereas N_{data} and T_f are fixed system parameters.

■

3.3.3.2 End-to-End Capacity Lower Bound

In this subsection, the lower bound of end-to-end capacity is derived following similar analysis made on the upper bound capacity in the previous subsection. In ensuring maximum link utilization in IEEE 802.16 networks, unavailable minislots in the WL may be reusable for transmissions by non-WLs within the same BR. In the worst case, none of those minislots are reusable by the non-WLs such as in the case of $H \leq \lfloor N \rfloor + 2$. In such case, those non-WLs can only share the remaining available minislots of the WL for future allocations, and hence set the lower bound on the route capacity.

***Theorem 2**: The lower bound of the end-to-end capacity of a route in dense networks or optimally deployed networks is*

$$\frac{T_\mu (N_{data} - max_m\{N_{w_m}^u\} - \min\{H, \lfloor N \rfloor + 2\})}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{\min\{H, \lfloor N \rfloor + 2\}} \tag{3.26}$$

[10]According to Definition 1, all WLs have the same capacity.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

where $m = 1, 2, \ldots,$ and M.

Proof: The capacity for a Bottleneck Zone (BZ) covering the WL with index w_m is first considered. In the worst case, none of the minislots in $S(N^u_{w_m})$ are reusable in the non-WLs in the BZ. Those non-WLs must share the available minislots of the WL for future allocations. Such a case leads to the lowest capacity for the BZ. The condition of the worst case is then

$$\{S(N_b) \cup S(N_{b+1}) \ldots \cup S(N_{b+L_{BZ}-1})\} \cap S(N^u_{w_m}) = \emptyset \tag{3.27}$$

where b is the index of the first link in a BZ and the link indexes in expression (3.27) are $b, b+1, \ldots, b+L_{BZ}-1$, including the WL index w_m. To avoid concurrent transmissions in a BZ, the following condition must be satisfied

$$S(N_b) \cap S(N_{b+1}) \ldots \cap S(N_{b+L_{BZ}-1}) = \emptyset \tag{3.28}$$

where the link indexes are $b, b+1, \ldots, b+L_{BZ}-1$, including the WL index w_m.

We can generalize the above derivations to a generic form where the minislots in $S(N^u_{w_m})$ are reusable, at least partially, by the non-WLs. In its generic form, the total number of minislots that can be allocated in the BZ (including the WL), in order to transmit P bits, must be at least $N_{data} - N^u_{w_m}$, i.e.

$$\sum_{i=b}^{b+L_{BZ}-1} N_i \geq N_{data} - N^u_{w_m} \tag{3.29}$$

As $N_i = \lceil \frac{P}{T_\mu r_i} \rceil$, expression (3.29) can be written as

$$\sum_{i=b}^{b+L_{BZ}-1} \lceil \frac{P}{T_\mu r_i} \rceil \geq N_{data} - N^u_{w_m} \tag{3.30}$$

The following expression always holds

$$\sum_{i=b}^{b+L_{BZ}-1} \lceil \frac{P}{T_\mu r_i} \rceil \leq \sum_{i=b}^{b+L_{BZ}-1} (\frac{P}{T_\mu r_i} + 1) \tag{3.31}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Combining expressions (3.30) and (3.31), the following expression is obtained

$$\sum_{i=b}^{b+L_{BZ}-1}(\frac{P}{T_\mu r_i}+1) \geq N_{data}-N_{w_m}^u \tag{3.32}$$

Manipulating expression (3.32),

$$P \geq T_\mu(N_{data}-N_{w_m}^u-L_{BZ})[\sum_{i=b}^{b+L_{BZ}-1}\frac{1}{r_i}]^{-1} \tag{3.33}$$

Therefore, the number of bits that can be transported in the BZ in a frame is expressed as

$$\frac{P}{T_f} \geq \frac{T_\mu(N_{data}-N_{w_m}^u-L_{BZ})}{T_f}[\sum_{i=b}^{b+L_{BZ}-1}\frac{1}{r_i}]^{-1} \tag{3.34}$$

For any link i in the BZ, $\frac{1}{r_i} \leq \frac{1}{\min\{r_i, i\in[b,b+L_{BZ}-1]\}}$, thus, the following expression holds

$$[\sum_{i=b}^{b+L_{BZ}-1}\frac{1}{r_i}]^{-1} \geq \frac{\min\{r_i, i\in[b,b+L_{BZ}-1]\}}{L_{BZ}} \tag{3.35}$$

Combining expressions (3.34) and (3.35), the following expression is obtained

$$\frac{P}{T_f} \geq \frac{T_\mu(N_{data}-N_{w_m}^u-L_{BZ})}{T_f} \times \frac{\min\{r_i, i\in[b,b+L_{BZ}-1]\}}{L_{BZ}} \tag{3.36}$$

Logically, $\frac{P}{T_f} \geq 0$. The lower bound capacity for a BZ, C_{BZ}^{min}, is therefore

$$C_{BZ}^{min} = \frac{T_\mu \cdot \max\{0,(N_{data}-N_{w_m}^u-L_{BZ})\}}{T_f} \times \frac{\min\{r_i, i\in[b,b+L_{BZ}-1]\}}{L_{BZ}} \tag{3.37}$$

Next, we formulate the capacity for a Bottleneck Region (BR) that covers the WL with index w_m, subject to the condition of the worst case that the set of data minislots assigned to transmit P bits over link $l_i(i \neq w_m)$ in the BR cannot belong to $S(N_{w_m}^u)$. That is,

$$\{S(N_e) \cup S(N_{e+1}) \ldots \cup S(N_{e+L_{BR}-1})\} \cap S(N_{w_m}^u) = \emptyset \tag{3.38}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

where e is the index of the first link in a BR and the link indexes in expression (3.38) are $e, e+1, \ldots, e+L_{BR}-1$, excluding w_m.

Expression (3.38) implies that the non-WLs in the BR cannot re-use the minislots that are unavailable for the WL, i.e. $S(N^u_{w_m})$. To avoid concurrent transmissions in any BZ of a BR the condition expressed in (3.28) must hold. In the worst case, the condition expressed in (3.29) must be satisfied for each BZ in a BR. Given such condition, it is unnecessary to sum up all N_i in expression (3.29) in a BR. This is because links which are not in the same BZ can allocate the same minislots used in other BZs due to spatial reuse. If the index of the BZ, j , is included in expression (3.29), it can be rewritten as

$$\sum_{i=b_j}^{b_j+L_{BZ}-1} N_i \geq N_{data} - N^u_{w_m} \tag{3.39}$$

Following similar derivation procedure (from expression (3.29) to expression (3.36)), the number of bits that can be transported in a BR during a frame is expressed as

$$\frac{P}{T_f} \geq \frac{T_\mu(N_{data} - N^u_{w_m} - L_{BZ})}{T_f} \times \frac{min_j\{r_i, i \in [b_j, b_j + L_{BZ} - 1]\}}{L_{BZ}} \tag{3.40}$$

According to expression (3.40), the minimum capacity of the BR which covers the WL with index w_m, C^{min}_{BRm}, can be expressed as

$$C^{min}_{BRm} = \frac{T_\mu(N_{data} - N^u_{w_m} - L_{BZ})}{T_f} \times \frac{\min\{r_i, i \in [e, e + L_{BR} - 1]\}}{L_{BZ}} \tag{3.41}$$

where e is the index of the first link of the BR.

We can generalize the above derivations to a generic case where there are $M(M \geq 1)$ WLs along a route and M BRs which may overlap one another along a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

route. If in this case, the condition in expression (3.39) holds, then the end-to-end capacity is a minimum, and can be expressed as

$$C_{e2e}^{min} = min_m\{C_{BRm}^{min}\} \geq \frac{T_\mu(N_{data} - max_m\{N_{w_m}^u\} - L_{BZ})}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{L_{BZ}} \tag{3.42}$$

where $m = 1, 2, \ldots,$ and M. According to Lemma 1, $L_{BZ} = \min\{H, \lfloor N \rfloor + 2\}$. Expression (3.42) is then depicted as

$$C_{e2e}^{min} \geq \frac{T_\mu(N_{data} - max_m\{N_{w_m}^u\} - \min\{H, \lfloor N \rfloor + 2\})}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{\min\{H, \lfloor N \rfloor + 2\}}$$

where $m = 1, 2, \ldots,$ and M.

■

3.3.4 End-to-End Capacity Bounds in Random Networks

In ***Scenario II*** where nodes are randomly deployed in a network with a specific node density, the one-hop distance is a random variable depending on the transmission range and node density in the network.

Lemma 4: *In random networks (Scenario II), the number of links in one BZ is bounded by*

$$\min\{H, \lfloor N \rfloor + 2\} \leq L_{BZ} \leq 2\lfloor N \rfloor + 3$$

Proof: The proof is as follows: (1) Lower bound: Scenario I is a special case of Scenario II. In Scenario I, the one-hop distance is maximum and the number of links in one BZ is minimum, i.e. $L_{BZ} \geq \min\{H, \lfloor N \rfloor + 2\}$ (Lemma 1), and (2) Upper bound: from the analysis in Section 3.3.1, in a random network, two

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

concurrent transmitters are at most $2\lfloor N \rfloor + 3$ hops apart on a route. Therefore, the upper bound of the number of links in one BZ is $2\lfloor N \rfloor + 3$.

Lemma 5: *In random networks, the number of links in one BR is bounded by*

$$\min\{H, \lfloor N \rfloor + 2\} \leq L_{BR} \leq 4\lfloor N \rfloor + 5$$

Proof: The proof is as follows: (1) Lower bound: If there is only one BZ in the BR, the number of links in one BR is minimum, i.e. $L_{BR} = L_{BZ} \geq \min\{H, \lfloor N \rfloor + 2\}$ (Lemma 4), and (2) Upper bound: Let BZ_1 be the first BZ in the BR and BZ_k be the last BZ in the same BR. BZ_1 contains the first link of the BR and BZ_k contains the last link of the BR. According to the definition of BR, any two BZs in one BR must share at least the WL. When the two BZs share only the WL, the number of links in the BR is maximum. In this case, the combined sets of two BZs ($S(BZ_1)$ and $S(BZ_k)$) satisfies $S(BZ_1) \cap S(BZ_k) = l_{w_m}$. As the upper bound of the number of links in one BZ is $2\lfloor N \rfloor + 3$ (Lemma 4), the upper bound of the number of links in one BR is thus $2(2\lfloor N \rfloor + 3) - 1 = 4\lfloor N \rfloor + 5$.

Lemma 6: *In random networks, the number of BZs in one BR is bounded by* $1 \leq L_{BR}^{Z} \leq 2\lfloor N \rfloor + 3$.

Proof: The proof is as follows: (1) Lower bound: In the case of the lower bound in Lemma 5, there is only one BZ in the BR, i.e. $L_{BR}^{Z} = 1$, and (2) Upper bound: In the case of the upper bound in Lemma 5, BZ_1 contains the first link of the BR and BZ_k contains the last link of the BR. The WL l_{w_m} is the last link of BZ_1 and the first link of BZ_k. Let the first link index of BZ_1 be e. The upper bound of the number of BZs in one BR is then $w_m - e + 1$, which is the upper

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

bound of the number of links in one BZ. According to Lemma 4, $L_{BR}^{Z} \leq 2\lfloor N \rfloor + 3$.

Based on the above proofs, we can now derive the capacity bounds for a given route in random networks. We note that the condition to achieve the upper bound capacity is similar to that in Scenario I, and that is when all of the unavailable minislots of a WL are reusable in the non-WLs of a BR. In such case, all unavailable minislots of the WL are available for reuse in future data transmissions, and hence maximum spatial reuse is achieved in the random network.

***Theorem 3**: The upper bound of the end-to-end capacity of a route in random networks is*

$$\frac{r_{w_m} T_{\mu} (N_{data} - N_{w_m}^{u})}{T_f} \tag{3.43}$$

where r_{w_m} is the transmission rate of the WL with index w_m.

The proof of the upper bound capacity in random networks is similar to that of Theorem 1, except that the values of L_{BZ} and L_{BR} are different from those in the optimal networks and random networks. Since Scenario I is a special (optimal) case of Scenario II, the upper bound capacity in Scenario I also provides the upper bound capacity in Scenario II.

The lower bound of the end-to-end capacity of a route in random networks is computed based on the worst case where non-WLs cannot reuse any of the unavailable minislots of a WL in a BR. This is similar to the worst case in Scenario I where those non-WLs must share the available minislots of the WL for future allocations, and therefore spatial reuse is limited to its minimum.

***Theorem 4**: The lower bound of the end-to-end capacity of a route in ran-*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

dom networks is

$$\frac{T_\mu(N_{data} - max_m\{N^u_{w_m}\} - 2\lfloor N \rfloor - 3)}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{2\lfloor N \rfloor + 3} \tag{3.44}$$

where $m = 1, 2, \ldots,$ and M.

Proof: The conditions to achieve the lower bound capacity are the same as those in Scenario I, and are expressed in (3.27) and (3.28). The procedures to derive the lower bound capacity follow the same (from expressions (3.29) to (3.42)) in Scenario I. Hence, the minimal end-to-end capacity is expressed as

$$C^{min}_{e2e} = min_m\{C^{min}_{BRm}\} \geq \frac{T_\mu(N_{data} - max_m\{N^u_{w_m}\} - L_{BZ})}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{L_{BZ}} \tag{3.45}$$

where $m = 1, 2, \ldots,$ and M. As described in the proof of Theorem 2, the lowering of capacity bound tends towards the generic case where there are $M (M \geq 1)$ WLs on a route. In such a case, there are M BRs on the route, which may overlap one another and according to Lemma 4, $L_{BZ} \leq 2\lfloor N \rfloor + 3$. Hence, the lower bound capacity in random networks is expressed as

$$\begin{aligned} C^{min}_{e2e} &= min_m\{C^{min}_{BRm}\} \geq \frac{T_\mu(N_{data} - max_m\{N^u_{w_m}\} - L_{BZ})}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{L_{BZ}} \\ &\geq \frac{T_\mu(N_{data} - max_m\{N^u_{w_m}\} - 2\lfloor N \rfloor - 3)}{T_f} \times \frac{\min\{r_i, i \in [1, H]\}}{2\lfloor N \rfloor + 3} \end{aligned}$$

where $m = 1, 2, \ldots,$ and M.

■

3.3.5 Capacity Discovery Protocol

A capacity discovery procedure is proposed for soliciting and collating local capacity information for end-to-end capacity based on solutions derived in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

earlier subsections. The procedure for discovering end-to-end capacity bounds can be one that uses a proprietary protocol or piggybacks onto a route discovery protocol. This research proposes a capacity discovery procedure using the later approach, where capacity information is encapsulated in the messages of routing protocols. In proactive routing, a route is set up on an off-line basis before actual flows arrive. In reactive routing, a route is set up from a local node to the destination on demand. Proactive routing is more suitable for a network with relatively stable topology, whereas reactive routing can adapt to changes in network topology and is more suitable for networks with unstable topology. Our proposed capacity discovery procedure can be adapted for both proactive and reactive routing protocols. In the following subsections, we present a procedure to determine the end-to-end capacity in the reactive routing case.

Reactive routing is also called source-initiated on-demand routing. A well-known reactive routing protocol is the Ad Hoc On-Demand Distance Vector Routing (AODV) [147]. When a source node has data to send to a destination node and does not have an existing route to that destination, the source node initiates a path discovery process. The source node broadcasts a route request (RREQ) message to its neighbors, which then forward the RREQ to their neighbors, and so on, until either the destination or an intermediate node with a "sufficiently fresh" route to the destination receives the RREQ message.

Upon receiving an RREQ, intermediate nodes record in their route tables the neighbor from which the first copy of the broadcast packet is received, thereby establishing a reverse path. Once the RREQ reaches the destination or an inter-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

mediate node with a valid route, the destination or intermediate node responds by unicasting a route reply (RREP) message back to the node from which it first received the RREQ. As the RREP is routed back along the reverse path of the related RREQ, nodes along this path set up forward route entries in their routing tables which point to the node from which the RREP was received.

We extend the AODV routing protocol to simultaneously discover a route and also the end-to-end capacity associated with the route. We refer to the extended protocol as **AODV-C** (AODV with capacity extension), in which information elements carrying capacity information are piggybacked onto routing control messages (RREQ and RREP) for efficient information exchange and minimum message overheads.

To request for a new route with end-to-end capacity estimation, a *C flag* in the *Reserved* field of the RREQ message is set at the source. We define the 14th bit in the reserved field of the RREQ message to be the *C flag* bit. We also define four message extensions in both RREQ and RREP messages for carrying capacity information, as shown in Figure 3.10. The message extensions are designed to carry the information required to compute end-to-end capacity bounds according to Theorems 1-4. In Figure 3.10, Type 3 extension carries the value of $N^u_{w_m}$, as defined in Table 3.2.

When the flow originator initiates an RREQ message, with the *C flag* set and the message extensions, the RREQ message is processed by intermediate nodes until it reaches the destination. The destination will reply with a corresponding RREP message, carrying the capacity information back to the source. With this

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

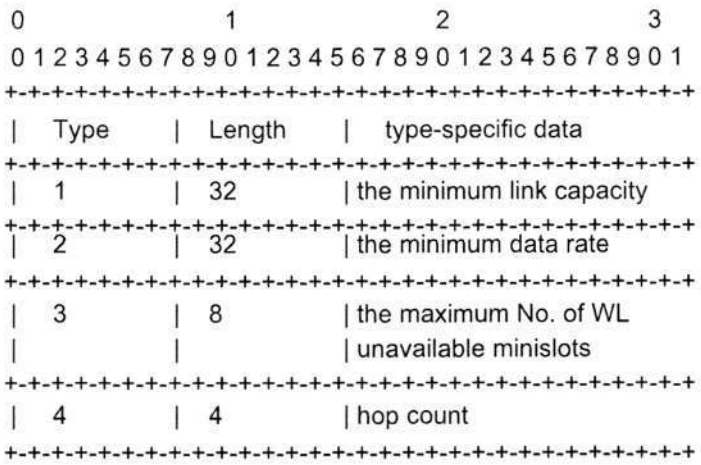


Figure 3.10: AODV-C capacity retrieval extensions.

information, the source will know the end-to-end capacity for a specific route. The process flow for the operation of the AODV-C protocol is described as follows:

A) Route Request with Capacity Information

A flow originator initiates an RREQ message, setting the *C flag* and resetting the *hop count* field to be zero.

B) Capacity Information Initialization

A node i, on receiving an RREQ message with its *C flag* set, and supporting the capacity extension, will compute and pass on its link capacity to its neighboring nodes. If node i does not support the capacity extension, it resets the *C flag* and treats the RREQ message as a normal AODV message. The originator then sets up the route without end-to-end capacity information.

If the *hop count* field is *zero*, indicating that the previous hop is the node that initiates the RREQ, node i will need to perform the following actions:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1. insert the capacity of the link from node $i-1$ to node i into *the minimum link capacity* field;

2. insert the data transmission rate of the link from node $i-1$ to node i into *the minimum data rate* field;

3. insert the number of unavailable minislots in the link from node $i-1$ to node i into *the maximum WL unavailable minislots number* field;

4. increase the value of the *hop count* field by one.

C) Capacity Information Updates

If the *hop count* field is not *zero*, indicating the previous hop is not the node that initiates the RREQ, node i will need to:

1. compare the value from *the minimum link capacity* field with the capacity of the link from node $i-1$ to node i, and put the smaller value into *the minimum link capacity* field;

2. compare the value from *the minimum data rate* field with the data transmission rate of the link from node $i-1$ to node i, and insert the smaller value into *the minimum data rate* field;

3. if there is no change to *the minimum link capacity* field, *the maximum WL unavailable minislots number* field will also remain unchanged; otherwise, insert the number of unavailable minislots in the link from node $i-1$ to node i into *the maximum WL unavailable minislots number* field;

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4. increase the value of the *hop count* field by one.

D) End-to-End Capacity Estimation

In order to obtain end-to-end capacity information, AODV-C RREQ messages must be able to reach the destination node, and only the destination node can reply to the RREQ message. As such, an intermediate node cannot reply to the RREQ message even if it has a valid route to the destination node. Hence, the route setup delay will be increased accordingly[11].

When generating RREP messages, the destination node copies the four extension fields from the RREQ message onto its RREP message. An intermediate node receiving a RREP message will forward it in the reverse path, keeping the four capacity extension fields unchanged. When the source node receives the RREP message, the node can then complete the computation of the lower bound and upper bound of the route capacity based on Theorems 1-4 and the values of $N_{w_j}^u$ and $\min_i\{r_i\}$.

3.4 Summary

In this chapter, a distributed solution is proposed for estimating real-time link and end-to-end capacities in IEEE 802.16 wireless mesh networks. The pro-

[11]From our simulation, in a random topology network with 60 nodes, by disabling the intermediate node replying, the average route setup delays are increased by 51%, 57%, and 64% for two-hop, three-hop, and four-hop flows.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

posed solution considers both intra-flow and inter-flow interference[12] and hence is robust for real-world deployment of mesh networks for supporting wide range of traffic patterns. The solution is efficient as it does not introduce extra overhead over and above the basic scheme defined in the IEEE 802.16 specifications. The real-time end-to-end capacity bounds in IEEE 802.16 wireless mesh networks are derived based on the concept of *Bottleneck Zone* introduced in the chapter. The end-to-end capacity are affected by the size of the *Bottleneck Zone*, which in turn depends on network deployment patterns. As such, two scenarios which illustrate the specific and generalized deployment patterns are presented. The procedure for capacity discovery in wireless mesh networks is also presented using piggy-backed message exchange to convey capacity information over the well-known AODV routing protocol.

[12]The end-to-end capacity computation takes intra-flow interference into consideration, while the link capacity computation considers both intra-flow and inter-flow interference, because the computing of minislot states, which forms the foundation of link capacity computation, does not differentiate flows that the minislots are allocated for (as in Figures 3.4, 3.5 and 3.6).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 4

Performance Evaluation of Capacity Estimation

This chapter presents and discusses the performance of our proposed solutions in capacity estimation for IEEE 802.16 mesh networks described in Chapter 3. Simulation results are obtained on capacity bounds using the QualNet 3.9.5 simulator [148] and used to evaluate the accuracy and robustness of our proposed solutions for capacity estimation.

4.1 Simulation Setup

The simulations are carried out using the QualNet 3.9.5 simulator [148]. In order to take into consideration the radio interference from neighboring nodes, we develop and build a MAC protocol for mesh operation in the simulation based on the IEEE 802.16 specification [2], which defines the use of a coordinated dis-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

tributed method for scheduling data transmissions.

In the simulation, the channel bandwidth is set at 16 MHz and the raw physical layer data rate is 11.423 Mbps. The OFDMA FFT size is 512 in the physical layer. The path loss is based on the two-ray ground reflection model and nodes communicate with a transmission power of 15 dBm using omni-directional antennas. Received signals with powers below -83 dBm (prior to signal amplification by the antenna at the receiver) are not delivered. The MAC frame length is 10 ms with 16 minislots in the control subframe and 256 minislots in the data subframe. In the simulation, the ratio of interference range to transmission range is close to two. Therefore, the state transition of a minislot follows the conservative case (illustrated in Figure 3.6), to avoid interference among neighboring links. The buffer size in the transmission node is configured to be 200 packets at the routing layer. For ease of computation and analysis of simulation results, we consider only constant bit rate (CBR) input flows in our simulations[1].

In order to evaluate the performance of our proposed capacity estimation solution for practical implementation in IEEE 802.16 mesh networks, we incorporate our enhanced Ad hoc On-Demand Distance Vector (AODV) protocol with capacity extension, AODV-C (see Chapter 3.3.5), in the simulations. In our simulations, route capacities are discovered by iteratively computing and updating link capacity information, i.e. parameter values of Theorems 1-4, such as $N_{w_j}^u$ and $\min_i\{r_i\}$ at each mesh node along the route. The sending node, on receiv-

[1]For the variable bit rate (VBR) case, computing capacities is much more complicated as compared to the CBR case even in wired networks.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ing the capacity information from all mesh nodes along the path, calculates the end-to-end capacity bounds for the route according to Theorems 1-4.

4.2 Simulation Results

We evaluate our proposed solutions in capacity estimation by comparing the capacity bounds with actual throughput in our simulations. Two deployment scenarios, as described in Chapter 3, for implementing our solutions are simulated for evaluation. In Scenario I, where nodes are densely or optimally deployed in a mesh network, we evaluate our solutions based on the string and square mesh network topologies. In Scenario II, where nodes are randomly deployed in a mesh network, we evaluate the solutions for different node densities.

4.2.1 Optimal Mesh Nodes Deployment

In Scenario I, where the nodes are optimally deployed so that the expected one-hop distance is close to node transmission range, we evaluate the capacity estimates for different sets of string and mesh topologies.

4.2.1.1 Capacity Estimation in String Topology Networks

As shown in Figure 4.1(a), $n(n = 2, 3, \ldots, 8)$ nodes form a string topology network with $n - 1$ hops and inter-node distance of 1000 meters. Each node can only receive packets from its immediate neighbor(s). For each string topology of different number of hops, node 1 transmits CBR flows to node n with various set

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

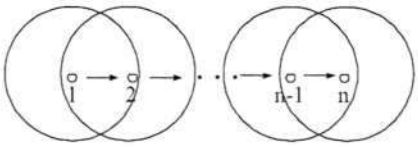
of input loads. For each input load, three metrics are measured: (1) the upper bound of the end-to-end capacity from node 1 to node n, (2) the lower bound of the end-to-end capacity from node 1 to node n, and (3) the throughput of the route from node 1 to node n^2. The simulation results obtained are presented in Figure 4.1.

a) Capacity Bounds and Throughput Comparison

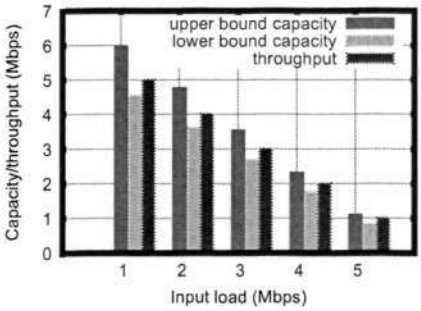
From Figure 4.1, it can be seen that the measured throughput is within the upper and lower capacity bounds in most cases of input loads. It can also be observed that the measured throughput tends closer to the lower bound than the upper bound. This is because the upper bound is an idealized capacity estimate which is attainable only under the conditions of (1) with centralized scheduling decisions, (2) when nodes are optimally deployed, and (3) when traffic pattern is optimized. In a centralized scheduling, global knowledge on traffic patterns at all nodes are known to the scheduling node and hence the capacity estimate computed at a node represents an upper bound that is not practically attainable in IEEE 802.16 networks where distributed scheduling is employed. Similarly, the lower bound represents the upper limit of the worst case attainable capacity estimate based on distributed scheduling decisions, because the under-utilization due to minislot fragmentation is not taken into considerations. Hence, in our

[2]As capacity bounds are computed in terms of available capacities in respond to specific input loads, to compare the available throughput and capacity bound, the throughput shown in each figure is the available throughput which is the total throughput minus the throughput of the pre-configured input load flows.

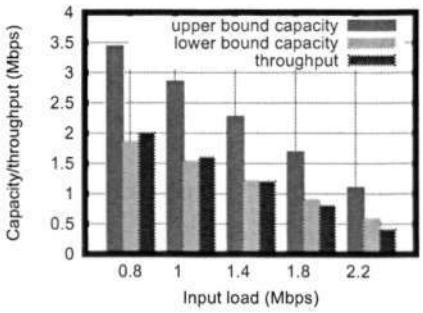
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



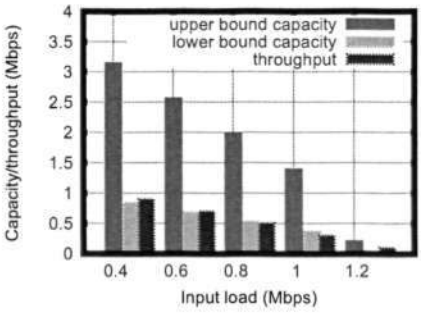
(a) String topologies.



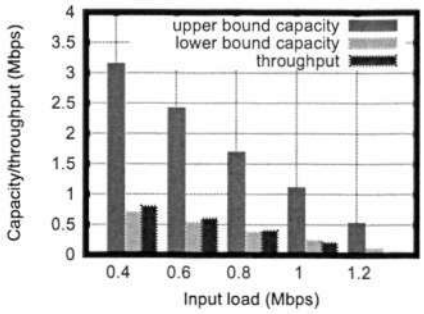
(b) One-hop string topology.



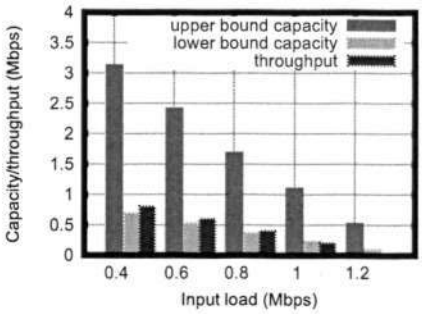
(c) Two-hop string topology.



(d) Four-hop string topology.



(e) Five-hop string topology.



(f) Seven-hop string topology.

Figure 4.1: Capacity and throughput comparisons in string topologies.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

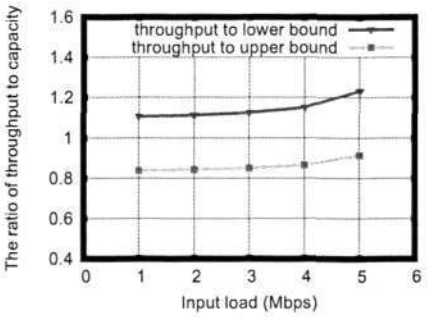
simulations of IEEE 802.16 networks which employ distributed scheduling, the measured throughput obtained is typically lower than the theoretical upper limit and closer to the lower bound. When the input loads are high, the throughput may be lower than the lower bound because the available minislots may be fragmented and unusable in data transmissions. Such data minislot fragmentation may be even more severe when the input loads approach saturation, hence the deviation of throughput below the lower bound is greater.

b) Capacity Estimation Accuracy

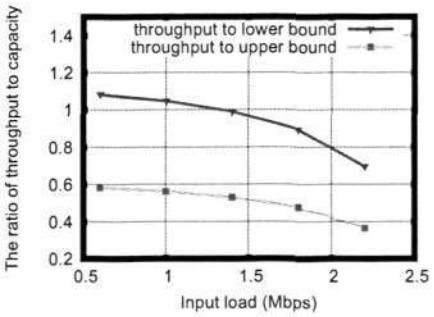
To determine the accuracy in capacity estimation, we compare the measured throughput with the lower capacity bound in the string topology networks. The average ratios of throughput to lower capacity bound are computed for each of the string topologies. As shown in Figure 4.2(f), the throughput-lower bound capacity ratio varies between 1.028 and 1.184 for hop count less than or equal to five, and increases significantly when the hop count is more than five. This indicates that the measured throughput closely tracks the lower bound capacity when the nodes are optimally deployed in a mesh network with a string topology and short hop-distance. Hence, the results show that our proposed solution is sufficiently accurate in estimating the lower bound of the available capacity in a mesh network with string topology.

As shown in Figure 4.2(f), the differences between the upper bound and the lower bound capacities are greater in networks with more hops. This is because the upper bound capacity represents an idealized capacity estimate and is not practically attainable in multi-hop networks.

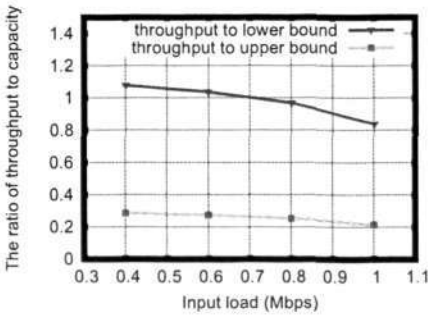
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



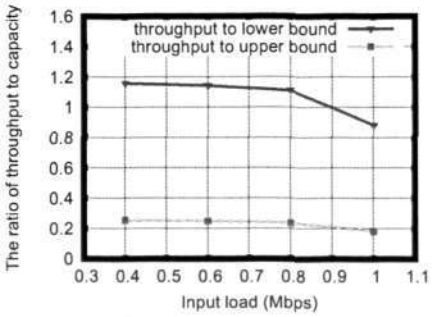
(a) One-hop string topology.



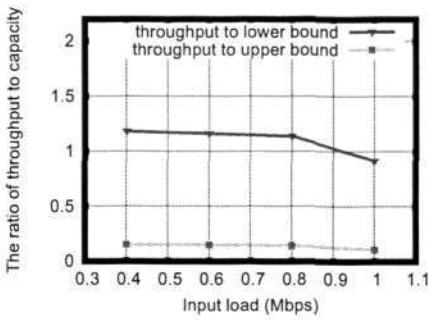
(b) Two-hop string topology.



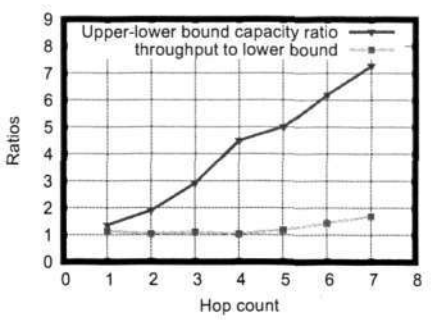
(c) Four-hop string topology.



(d) Five-hop string topology.



(e) Seven-hop string topology.



(f) Ratios of throughput-to-capacity bound in string topologies.

Figure 4.2: Ratio of throughput to capacity in string topologies.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

c) Effects of Transmission Interference on Capacity Estimation

To observe the effects of transmission interference on capacity estimation, the ratios of throughput-to-lower and throughput-to-upper bound capacities are calculated for each case of the string topologies. From Figs. 4.2(a) - 4.2(e), it can be observed that throughput-capacity ratios tend to decrease with increase in input loads for string topologies with more than one hop. In capacity estimation, it was assumed that all available data minislots could be allocated to subsequent data flows and thus these data minislots constitute the available capacity. In practice, minislot allocation usually results in "fragmented" data subframes due to interference from multiple transmitters and non-optimum selection of minislots. This results in isolated data minislots which are difficult to use. In the one-hop string topology, only node 2 (the receiver) schedules the transmission, so that fragmentation of minislots is minimized, hence the throughput-capacity ratio is less sensitive to the effect of minislot fragmentation. However, in the a string topology network with more than two nodes, data minislots are allocated across multiple links, which inevitably increases the fragmentation of minislots. Hence, in practical implementations of IEEE 802.16 networks, where the effect of minislot fragmentation is expected to be greater at higher input loads, the attainable throughput will be correspondingly lower than the capacity bounds.

d) Effects of Spatial Reuse on Capacity Estimation

Figure 4.3 compares the ratios of end-to-end throughput to one-hop throughput for string topologies with different number of hops. The end-to-end throughput is normalized by its one-hop throughput in order to determine the effects of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

spatial reuse on capacity estimation. It can be seen that the end-to-end throughput drops sharply when the number of hops in the string topology network increases from one to two and gradually less so as the number of hops in the network increases to four hops. This is due to link capacity sharing among the increasing number of links in the network and hence less throughput in data transmission. When the number of nodes in the network is five or more, the end-to-end throughput is less affected and remains almost constant. This is because nodes at distance greater than two-hop away from the sending node are out of the two-hop transmission interference range of the sender and therefore they do not affect the overall end-to-end throughput.

e) Effect of Hop Count on Capacity Estimation

To understand the effect of hop count (number of hops) on capacity estimation, the average throughput-lower bound capacity ratio is computed for each of the string topologies. As shown in Figure 4.2(f), the throughput-lower capacity ratio increases with increasing number of hops in the string topology. This implies that the throughput deviates more from the lower bound capacity. This is due to more spatial reuse in the network with more hops which results in higher throughput, while the lower bound capacity is computed when the minimal level of spatial reuse is considered. Hence, the accuracy of capacity estimation decreases in a string topology with more hops.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

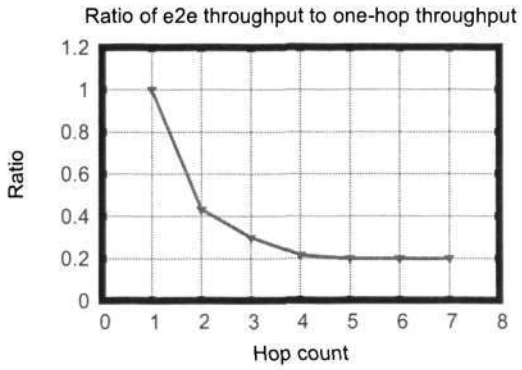


Figure 4.3: The ratio of e2e throughput to one-hop throughput.

4.2.1.2 Capacity Estimation in Mesh Topology Networks

We extend our evaluations in capacity estimation to mesh topology networks using similar metrics as those in the string topology networks. The evaluation is based on the same assumptions as in Scenario I, which is the nodes are optimally deployed in a mesh topology network so that the expected one-hop distance is close to the node transmission range. We simulate the mesh network with a regular square mesh topology defined by Robinson and Knightly [3]. In such a mesh topology, neighborhood nodes form vertexes of squares with interconnecting links, as shown in Figure 4.4(a).

In our simulations, the inter-node distance is configured to be 1000 meters, and is close to the node transmission range. The physical channel characteristics are the same as those in string topologies. The input traffic is a CBR flow of 800 Kbps and is transmitted from node 13 to node 16 between $t = 40$ s and $t = 80$ s. The real-time end-to-end capacity bounds are measured for each row of nodes in the network, i.e. from node 9 to node 12, node 5 to node 8, and node 1 to node 4.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

As described earlier in Chapter 4.1, route capacities are discovered by iteratively computing $N_{w_j}^u$ and $\min_i\{r_i\}$ at each mesh node along the route. The sending node then calculates the end-to-end capacity bounds for the route according to Theorems 1-4. The measurement results for lower and upper bound capacities are presented in Figs. 4.4(b) and 4.4(c), respectively.

a) Capacity Bounds and Throughput Comparison

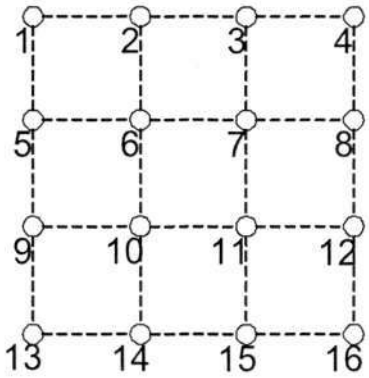
It can be observed in Figs. 4.4(b) and 4.4(c) that both lower and upper capacity bounds drop drastically during the initial period just after $t = 40$ s. The reason is that AODV-C is used as the routing protocol, which is a reactive routing scheme where a path is set up only on demand. During the route discovery period, because there is no existing route between the source and destination nodes, packets are enqueued in buffers at the network layer. After the path has been set up, the enqueued packets, together with the new arrived packets from the application layer, are released into the network. Hence, in the initial period just after $t = 40$ s, more bandwidth is consumed to transmit data packets compared to the application layer sending rate.

The comparisons between the throughput and the capacity bounds are shown in Figure 4.5(a). It can be seen that the measured throughput is within the upper and lower capacity bounds. As for the case of string topologies, the measured throughput in mesh topology is similarly closer to the lower bound capacity than to the upper bound capacity.

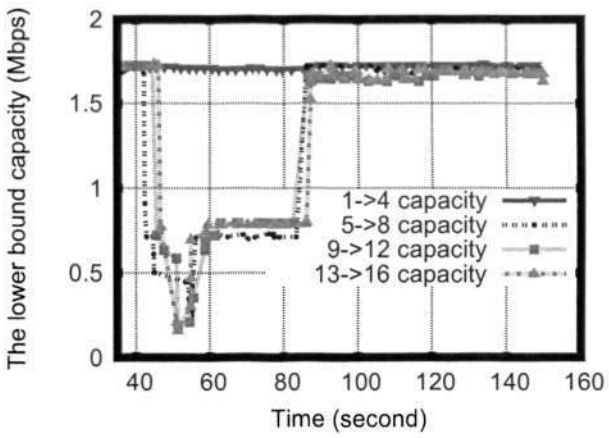
b) Capacity Estimation Accuracy

The accuracy of the capacity estimation solutions for mesh networks is eval-

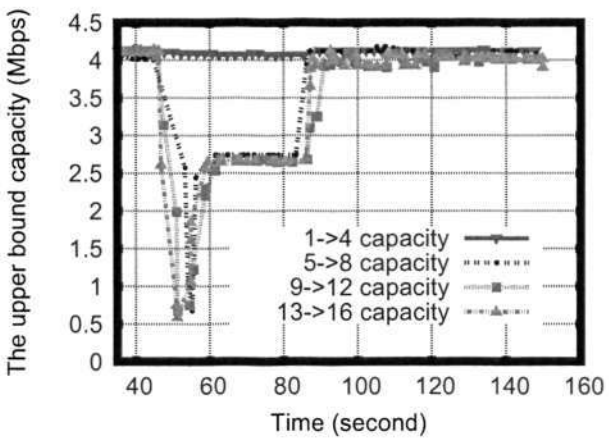
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) The mesh topology in simulation.



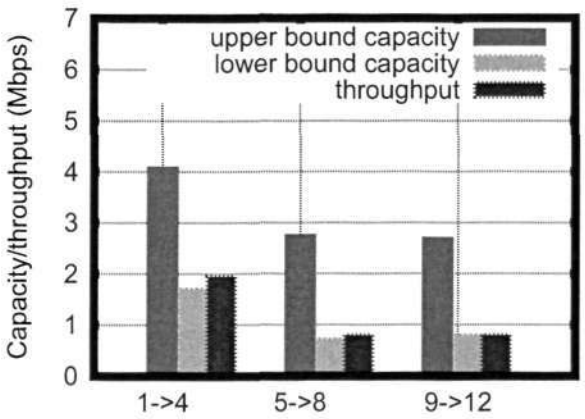
(b) Lower capacity bounds of different paths.



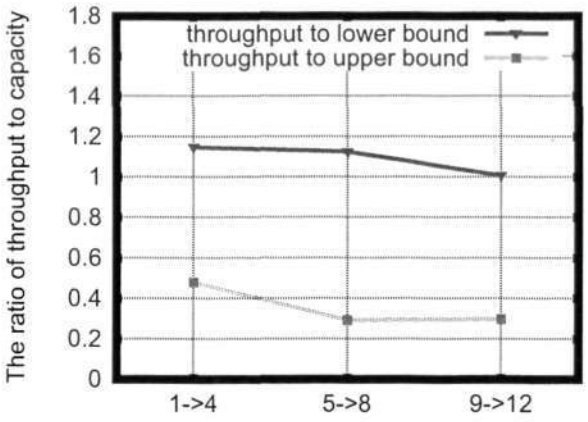
(c) Upper capacity bounds of different paths.

Figure 4.4: The capacity bounds in a mesh topology network.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) The throughput and capacity bounds comparison.



(b) The ratios of throughput to capacity bounds.

Figure 4.5: Comparison of capacity bounds and throughput in a mesh topology network.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

uated by measuring the throughput of node pairs $1 \rightarrow 4$, $5 \rightarrow 8$ and $9 \rightarrow 12$ between $t = 60$ s and $t = 80$ s, and comparing the throughput with the capacity bounds based on the same background traffic flowing from node 13 to node 16. The ratios of the throughput to capacity bounds are computed and shown in Figure 4.5(b). It can be seen that the ratios of the throughput to lower bound capacity are between 1.01 to 1.14, indicating that the capacity estimation method is sufficiently accurate in defining the lower bound capacities in the regular mesh topology.

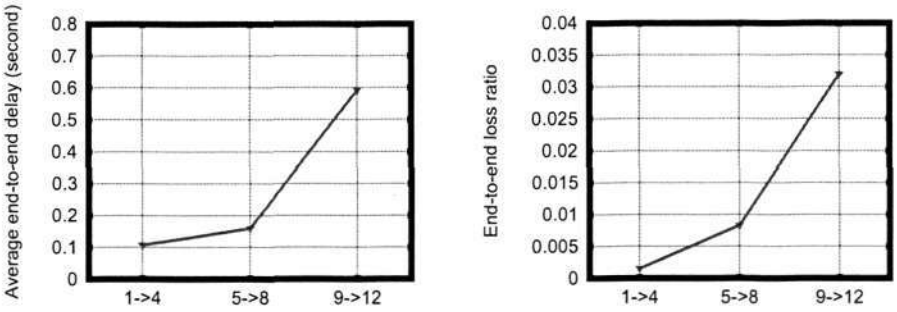
c) Effects of Transmission Interference on Capacity Estimation

It can be seen in Figs. 4.4(b) and 4.4(c) that during the period of the active CBR flow, the real-time capacity bounds drop sharply for all node pairs except node pair $1 \rightarrow 4$. The drop in capacity bounds for node pair $13 \rightarrow 16$ is attributed to the capacity being used for the transmission of the CBR flow from node 13 to node 16. The drop in capacity bounds for node pairs $9 \rightarrow 12$ and $5 \rightarrow 8$ are the results of transmission interference of neighboring nodes in conveying the CBR flow. Such interference leads to a reduction in the number of usable minislots, and hence a corresponding drop in capacity bounds. The interference occurs because nodes 5 to 12 are within the two-hop interference range of nodes 13 to 16.

d) Effects of Spatial Reuse on Capacity Estimation

It can be seen from Figures 4.4(b) and 4.4(c) that the capacity bounds for node pair $1 \rightarrow 4$ are not affected by the CBR flow, because nodes 1 to 4 are not within the interference range of nodes 13 to 16, and hence the capacity of node pair $1 \rightarrow 4$ is not affected by the CBR transmission in nodes 13 to 16. In such a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) The average end-to-end delays in mesh topology.

(b) The end-to-end loss ratios in mesh topology.

Figure 4.6: Flow performance in a square mesh network.

case, spatial reuse can be achieved in node pair $1 \rightarrow 4$.

e) Effects of Interference on Flow Performance

The performance of the mesh network is studied by measuring the average packet delay and loss ratio with the same background traffic flowing from node 13 to node 16. From the results presented in Figure 4.6, it can be seen that the flows at node pair $1 \rightarrow 4$ achieve the minimal packet delay and loss ratio as compared to the flows at node pairs $5 \rightarrow 8$ and $9 \rightarrow 12$. This is due to the fact that flows at node pair $1 \rightarrow 4$ are farther away from the background traffic as compared to the flows at other two node pairs, and thus are least affected by the transmission interference from the background traffic, resulting in lower packet delay and loss ratio. It can also be observed that the flow performance at node pair $9 \rightarrow 12$ is worse than that at node pair $5 \rightarrow 8$. This is due to the same reason that the former is closer to the background traffic than the latter, and hence is most affected by transmission interference from the background traffic.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.2 Random Mesh Nodes Deployment

In Scenario II, where the nodes are randomly deployed, we evaluate our capacity estimation solutions for different node densities in the network. In our simulations, $n(n = 20, 40, 60, 80, 100)$ nodes are randomly deployed with a uniform distribution in a 5000 m x 5000 m square area. The performance of the random mesh network is evaluated at a randomly selected communicating node pair, namely node 7 and node 13. In this simulation, the locations of node 7 and node 13 are taken from the network with 20 nodes. For the ease of evaluations, the locations of the two nodes are fixed in other random networks. The nodes other than node 7 and node 13 are randomly deployed with a uniform distribution in networks with $n(n = 40, 60, 80, 100)$ nodes. Figure 4.7 shows mesh networks with various number of nodes each. The nodes are represented as square dots and their locations are defined by their (x, y) coordinates.

a) Effects of Node Density on Path Hop Count in Random Mesh Networks

In the simulation, CBR flows are sent from node 7 to node 13, and the routed path between them is shown as a connected line linking the intermediate nodes in Figure 4.7. The hop counts of the path from node 7 to node 13 are compared with those in other networks of different node densities.

It is observed that for a given pair of communicating nodes in a dense network, there are fewer hops along the path than in a sparse network. This is because in a dense network, the shortest path can be found more optimally with richer connectivity than in a network with sparse neighboring nodes. In a densely

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

connected network, there is little benefit in increasing the node density, because any further increase in the number of nodes does not correspondingly reduce the number of hops of the path. This is due to the fact that the hop count for a given communicating pair node i and node j is lower bounded by $\lceil \frac{D_{i,j}}{a} \rceil$ [146], where $D_{i,j}$ is the distance between node i and node j, and a is the transmission range of the nodes.

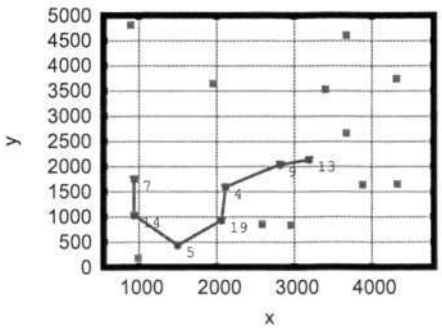
b) Effects of Path Hop Count on Capacity Estimation in Random Mesh Networks

The throughput is measured and compared with the capacity bounds computed using our capacity estimation solution for the randomly deployed networks. As shown in Figure 4.8(a), when the node density increases, both the throughput and lower bound capacity increase. This implies that rich connectivity in dense network can actually improve the end-to-end capacity because flows will incur less intra-flow and inter-flow interference over fewer path hop count. However, further increase in node density in dense networks may not necessarily reduce the path hop count, instead, the competition of the control channel among neighboring nodes will be more severe, and hence the actual route capacity may be reduced.

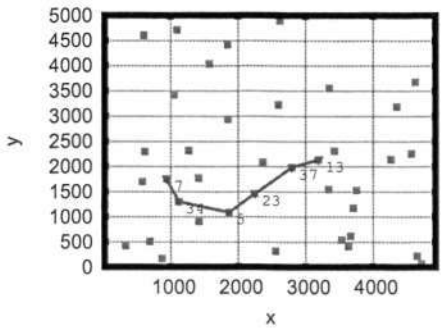
c) Accuracy of Capacity Estimation in Random Mesh Networks

It is observed in Figure 4.8(a) that the throughput tends closer towards the lower bound capacity than towards the upper bound capacity, similar to the string topology networks and the square mesh network. As presented in Figure 4.8(b), the ratio of throughput to lower bound capacity decreases with increase in node

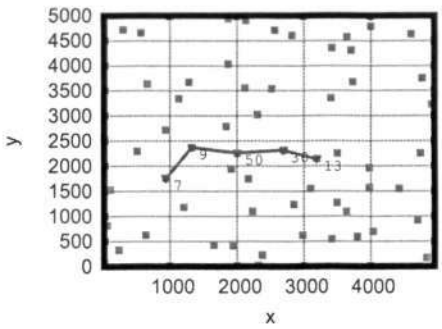
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



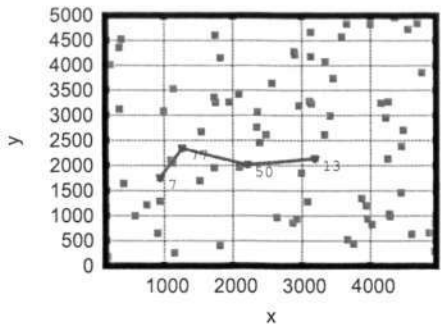
(a) 20 nodes.



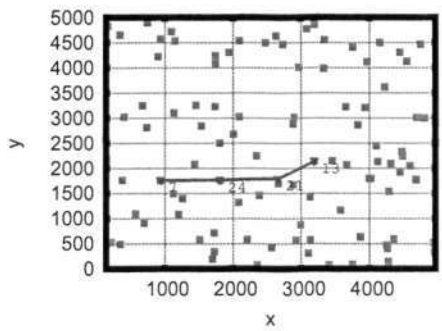
(b) 40 nodes.



(c) 60 nodes.



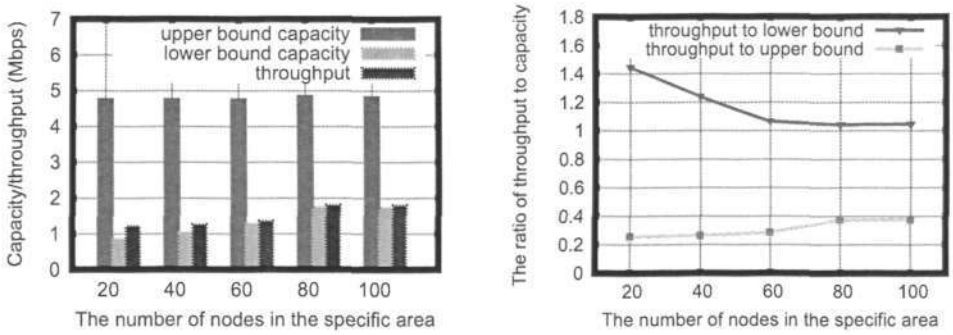
(d) 80 nodes.



(e) 100 nodes.

Figure 4.7: Shortest paths in mesh networks with random topologies.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) The throughput/capacities in random topologies.

(b) The ratios of the throughput to the capacity bound in random topologies.

Figure 4.8: Throughput/capacities comparisons in random topologies.

density, from 1.44 to 1.04. That is, the gap between throughput and lower bound capacity narrows as the number of nodes increases. This implies that lower bound capacities computed on sparse networks are more conservative than that on dense networks. This result is consistent with that in the string topologies in that the lower bound capacity estimates are more conservative in networks with more hops (see Figure 4.2(f)).

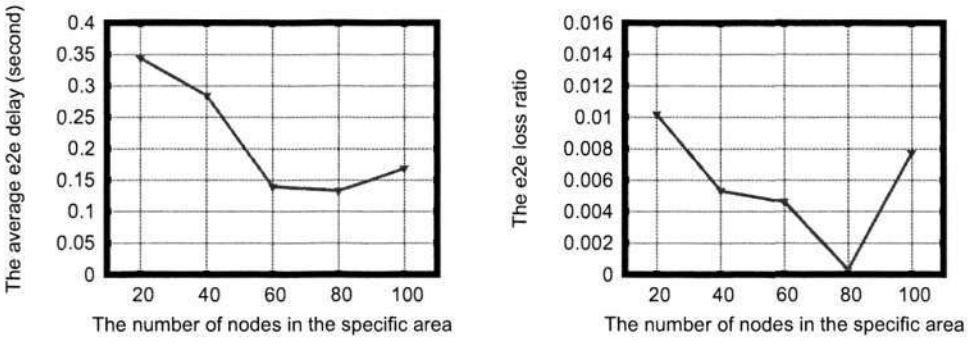
The throughput-lower bound ratio is between 1.04 and 1.237 when the number of nodes is between 40 and 100, which implies that our capacity estimation solution is accurate when a network has relatively rich connectivity.

d) Performance of Random Mesh Networks

The performance of random mesh networks is studied in terms of average packet delay and loss ratio for a CBR flow emitting at node 7 and receiving at node 13. As shown in Figure 4.9, both average delay and loss ratio[3] decrease

[3]We only consider intra-flow interference in random networks, this is because the inter-flow interference is not comparable in random network cases when they differ in node deployment.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) The average end-to-end delays.

(b) The end-to-end packet loss ratios.

Figure 4.9: The performance of data flows in random topology networks.

with increase in the node density up to 80 nodes, whereas the flow performance becomes worse in a network with 100 nodes as compared to that in a network with 80 nodes. This is due to the fact that flows are transmitted over fewer hops in denser networks as shown earlier, which leads to better flow performance. However, the link connection setup time, caused by three-way handshaking for data scheduling, increases significantly with the total node number [87], especially in high node density situation. Hence, when the number of nodes increases from 80 to 100, the higher link connection setup time results in increased packet delay and packet loss ratio due to buffer overflow.

4.3 Summary

This chapter presented the simulations conducted for evaluating performance of our proposed capacity estimation solutions for mesh networks with different deployment patterns. The simulations were conducted using the QualNet simulator. Simulation results show that the proposed capacity estimation solutions

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

are accurate in defining lower bound capacities, responsive to the dynamics of network traffic load, and robust in networks with different topologies. The results show that the upper bound capacity is optimistic and hard to achieve in distributed multi-hop networks because it requires global information knowledge to make scheduling decisions and optimal traffic pattern. It is also observed that the addition of new nodes in randomly deployed networks can actually improve the end-to-end capacity when the node density is not high. This is because the path from a source node to a sink node is shorter in denser networks contributed by increased level of node connectivity. However, increasing the node density in a dense network will increase the link connection setup time for data scheduling, which may in turn significantly increase packet delay and loss ratio.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 5

Capacity-Aware Optimal Relay Selection in IEEE 802.16 Multi-hop Relay Networks

This chapter presents the solution for optimal relay station (RS) selection in an IEEE 802.16j Multi-hop Relay (MR) vehicular network. The object of the optimal RS solution is to maximize the end-to-end throughput in vehicle-to-infrastructure communications between a mobile vehicular subscriber station (SS) and a fixed base station (BS) in the vehicular network.

5.1 Introduction

Vehicular networks provide wireless communications for road safety, Internet access and mobile commerce. In vehicular networks, one of the challenges to network operators is to provide reliable network services to support critical real-time applications such as in road safety emergencies and non-safety applications such as mobile commerce [20]. Although there has been existing communication technology proposed for safety applications [21] using the IEEE 802.11-based Dedicated Short-Range Communication (DSRC), the performance of the protocol

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

has so far not been satisfactory due to the contention and nondeterministic nature of IEEE 802.11 transmissions [22, 23] in a vehicular environment. As such, IEEE 802.16-based vehicular networks are emerging as more promising alternatives.

IEEE 802.16 [1] networks are promising technology for ubiquitous mobile broadband wireless services of a metropolitan scale due to its long range and high data rate characteristics. One of the important usage models defined in the latest version of the standard, IEEE 802.16j [5], is for transport systems [19], such as highways, railways, near-land rivers and seas. In these applications, IEEE 802.16j has been proposed as the standard for supporting mobile and multi-hop relay in vehicular communications.

5.2 System Model of Vehicular Networks

IEEE 802.16j Multi-hop Relay (MR) technology has been proposed for Vehicle-Roadside Communications (VRCs) to provide in-vehicle Internet access. In a vehicular MR network, the strategic deployment of relay stations (RSs) is crucial in ensuring good network performance. In this chapter, the problem of selecting the optimal RS for maximizing end-to-end capacity to a vehicular subscriber station (SS) is presented. In this work, it is assumed that the locations of vehicular SSs are known at a specific time for making decisions to select a relay station. The subsequent sections in this chapter present the system model in vehicular communications, communication protocol overview, link adaptation mechanisms, and optimal relay selection solution.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

5.2.1 Communication Protocol

IEEE 802.16j MR [5] networks exhibit many attractive features for supporting vehicular communications such as their multi-hop and mobility support, wide coverage and richness in QoS support. Figure 5.1 shows an example of the IEEE 802.16j vehicular network, with its three system components: BS[1], RS and SS. The BS is directly connected to the wired backhaul and provides connectivity, management, and control of RSs and SSs. The RSs are responsible for relaying data between the SSs and the BS. When the link condition between a BS and SS is poor, the data rate of one-hop transmissions between them will be compromised. In such a situation, IEEE 802.16j [5] MR allows one or more RSs to be deployed between BSs and SSs; thereby extending the coverage and enhancing the performance of the network.

Unlike BSs, RSs are not directly connected to a wire-line backhaul, thus rapid deployment can be achieved. The functionalities of RSs are considerably less complex than those of BSs and therefore deployment cost is also expected to be much less than that required to deploy more BSs in the system. This makes IEEE 802.16j system an attractive solution for wide area coverage and economical deployment.

Figure 5.1 depicts an example of an IEEE 802.16j vehicular network for highway deployment, where each vehicle is equipped with an IEEE 802.16 SS (and RS if possible). An SS either connects directly to a BS or to an RS, depending

[1]IEEE 802.16j standard defines base station as MR-BS (multi-hop relay BS), which is simply called BS in this chapter.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

on the location of the SS and other routing criteria. In this thesis, the RS can either be a stationary road side unit (RSU) [141], or a device carried on board a mobile vehicle, similar to the routing protocol proposed by Taleb et al [149]. As in So and Liang's study [24], the battery life of an RS is considered to be long or unlimited and hence the issue of energy saving strategy is not considered here.

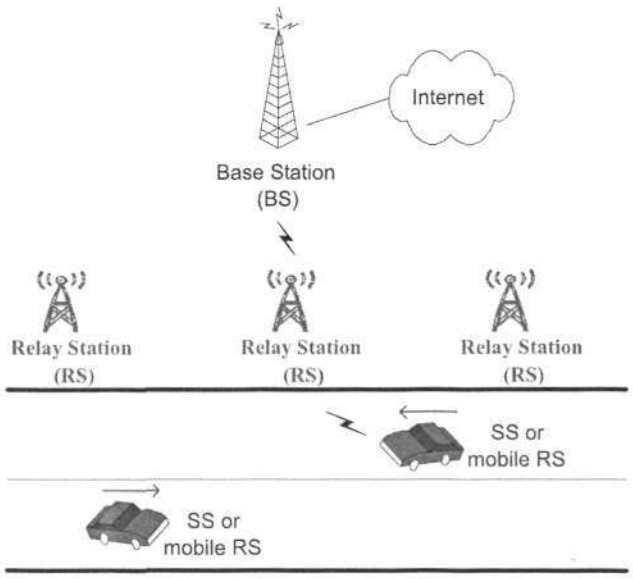


Figure 5.1: An IEEE 802.16j vehicular network.

We focus on the transport system usage model [19] defined in IEEE 802.16j. Most of these transport systems have a common highway mobility pattern. In this thesis, we present solutions in optimal relay selection based on our analytical model of highway mobility patterns.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

5.2.2 Link Adaptation in Wireless Networks

In wireless networks, link transmission rate is one of the important factors in the deployment of RSs. The IEEE 802.16j adopts a link adaptation technique for adaptation of link transmission rates to different channel conditions using an adaptive modulation and coding (AMC) scheme [150]. The objective of AMC is to maximize the data rate by adjusting transmission parameters to match varying channel conditions, while maintaining a prescribed packet error rate. In link adaptation, channel state information (CSI) carrying the signal-noise-ratio (SNR) at the receiver is fed back to the sender. On receiving the CSI, the sender adjusts its transmission mode to an appropriate modulation and coding type.

In AMC, the received SNR is divided into N non-overlapping regions assuming that the system supports N transmission modes. In the remainder of this chapter, "a node transmits with rate r_i" is used interchangeably with "a node is in transmission mode M_i". The system is assumed to support N data rates, denoted by r_i, where $i = 1, 2, \ldots, N$, and $r_1 < r_2 < \ldots < r_N$. When the received SNR at a specific time t, $SNR(t)$, is estimated to be in the i-th region ($i = 1, 2, \ldots, N$), i.e. $SNR(t) \in [SNR_i, SNR_{i+1})$, transmission mode M_i is assigned to the link and the data transmission rate over this link is r_i. When $SNR(t) < SNR_1$, there is no direct connection between the two nodes. When $SNR(t) \geq SNR_N$, the highest order of transmission mode M_N is applied to the link. The AMC scheme is depicted in Figure 5.2.

When the SNR increases at the receiver, the sender will adopt a higher order

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

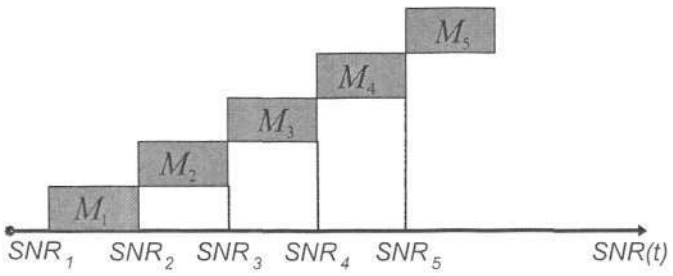


Figure 5.2: The Adaptive modulation and coding scheme.

transmission mode to transmit at a higher achievable rate of the link. Similarly, as the SNR gets worse, the sender switches its transmission mode to a lower order to adapt to the degraded channel condition.

In this thesis, the capacity-awareness of the optimal relay selection scheme means that data transmission rate is considered a key factor in the relay selection decision. We also consider the data transmission rate of a link to be dependent on its transmission mode. Further, we refer to the term "end-to-end transmission rate" of a path as the maximum number of bits that can be delivered over a path in a unit time and ignore the effect of any protocol overheads.

5.3 Optimal Relay Selection

5.3.1 Study of Highway Mobility Pattern

In AMC, the transmission mode, which affects the link transmission rate, is derived through SNR measurements of the received messages. When nodes are mobile, the distance between two nodes changes with time. The changes in the inter-node distances affect the received SNR according to the specific path loss model of the wireless channel. That is, the transmission mode of a link is a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

function of the link length. The effects of link length on transmission rates and modes have been studied by Joseph and Marten for different terrains [129].

In our analysis of the link transmission adaptation model, the corresponding coverage distances of link rate $r_1, r_2, \ldots$, and r_N are denoted by $l_1, l_2, \ldots$, and l_N, and $l_1 > l_2 > \ldots > l_N$, as shown in Figure 5.3. Any two nodes communicate using rate r_i when their distance L meets $l_{i+1} < L \leq l_i, i = 1, 2, \ldots, N-1$, and use rate r_N when $L \leq l_N$.

In Figure 5.3, it is assumed that there is a point H on the highway with the shortest distance, h, to the BS. For every distance l_i, there is a mapping point on the highway with distance a_i to the reference point O on the highway.

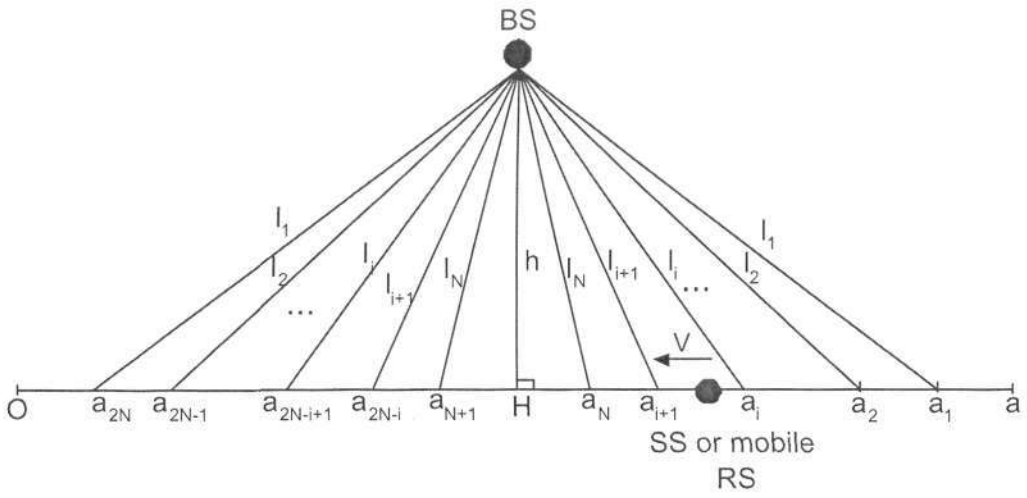


Figure 5.3: The generic model of adaptive transmission in highway mobility.

When an SS moves from one end of the highway (point a in Figure 5.3) to H, the distance between the BS and SS decreases, following the sequence of $l_1, l_2, \ldots$, and l_N. The transmission mode is in turn progressively switched to a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

higher order when the SS crosses switching locations $a_1, a_2, \ldots$, and a_N towards H. When the SS subsequently moves pass location H and proceeds towards the opposite end of the highway (point O in Figure 5.3), the distance between the BS and SS increases, following the sequence of $l_N, \ldots, l_2$, and l_1. In this case, the transmission mode will be progressively switched to a lower order when the SS crosses switching locations $a_{N+1}, \ldots, a_{2N-1}$, and a_{2N}.

Let L be the distance between the BS and SS, and R be the data rate between the BS and SS. The probability of the two nodes communicating with rate r_i, $P_i = P(R = r_i)$, follows

$$P_i = P(l_{i+1} < L \leq l_i) \tag{5.1}$$

where l_i is determined based on empirical measurement or a specific path loss model. An SS switches to rate r_i at two locations, a_i and a_{2N-i+1}, which are symmetric with reference to the location H, with the exception that the node switches to rate r_N only at a_N, as shown in Figure 5.3. Let X be the distance between the SS and the reference point O, and the distance between switching points and O be $a_1, a_2, \ldots$, and a_{2N}. The probability of the BS and the SS communicating with rate r_i, P_i, is

$$P_i = \begin{cases} P(a_{i+1} < X \leq a_i) + P(a_{2N-i} < X \leq a_{2N-i+1}), & \text{for } i = 1, 2, \ldots, N-1; \\ P(a_{N+1} \leq X \leq a_N), & \text{for } i = N \end{cases} \tag{5.2}$$

In the following part, the probability $P(R = r_i)$ is calculated by deriving the probabilistic distribution of node locations in the highway mobility pattern.

In our analysis, it is expected that vehicles traveling on a highway are con-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

strained to move along a fix path and in a certain direction, although the vehicular speed may vary from time to time. A mobile node carried by a vehicle moves at a certain speed on a highway and may change its speed along the way; however, it may only be allowed to change its direction at a designated location such as at the end of the highway segment or at a "U-turn" point of the highway. Figure 5.4 depicts the generic highway mobility model. It is assumed that a node initially chooses a random destination E_1 on a highway and moves at a constant speed v_1 towards E_1. Once the node reaches E_1, it randomly chooses another destination E_2 ahead and moves at another constant speed v_2 towards E_2. This process is repeated until the node reaches the end of the highway. Highway mobility model is different from random waypoint mobility model. In the latter model, a node may choose a random destination within a two-dimensional area, whereas the former allows a node to choose a random destination which is ahead of the current location along the highway. For the purpose of the analysis, if the highway is not a straight line, it will be modeled as a few straight-line highway segments.

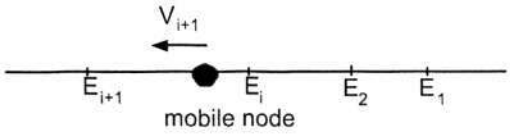


Figure 5.4: The highway mobility model.

It is now assumed that a node moves at a speed $\mathbf{V}$ on a straight-line highway segment with distance a. A random variable X denotes the location of the node, where $X \in [0, a]$. The speed $\mathbf{V}$ is a random variable and its magnitude may

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

change at any location along the highway segment. A *movement period* is defined as the trajectory of a node when the node travels at a constant speed. In other words, a movement period is the trajectory of a node between any two neighboring shifts of speed. Let random variables B and E denote the beginning and ending locations of a movement period, with distances b and e to the reference point at the left end of the highway, as shown in Figure 5.5. Since the node moves in a single direction, $b > e$ holds in any movement period. B and E are randomly chosen on the highway segment, i.e. their probability density functions are

$$f_B(b) = \begin{cases} \frac{1}{a}, & \text{for } 0 \leq b \leq a; \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

$$f_E(e) = \begin{cases} \frac{1}{b}, & \text{for } 0 \leq e \leq b; \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

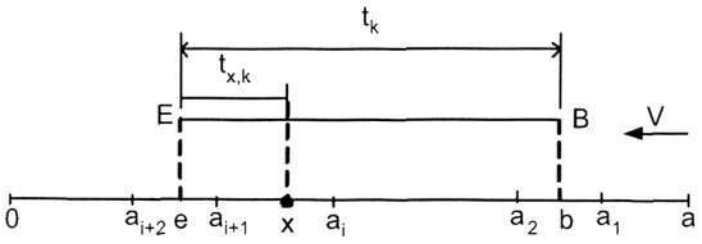


Figure 5.5: A movement period of a mobile node.

In order to derive $f_X(x)$, let us first calculate the cumulative distribution function (CDF) $F_X(x) = P(X \leq x)$, which denotes the probability that a node is located in $[0, x]$ at an arbitrary instant of time.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For each movement period k, let t_k denote the duration of this period and $t_{x,k}$ denote the duration that the node spends within $[0, x]$ during period k. If the k-th movement period does not intersect $[0, x]$, $t_{x,k} = 0$. Figure 5.5 shows the concept of t_k and $t_{x,k}$.

By observing the process for K movement periods, as K goes to infinity, the time that a node spends within $[0, x]$ during K movement periods, divided by the total duration of the K movement periods, converges towards $P(X \leq x)$. That is

$$P(X \leq x) = \lim_{K \to \infty} \frac{\sum_{k=1}^{K} t_{x,k}}{\sum_{k=1}^{K} t_k} \tag{5.5}$$

Let D denote the distance traveled by the mobile node over an entire movement period, V denote the speed of the node during the movement period, and T denote the duration that the node stays in the movement period. Similarly, let D_x denote the distance of a movement period that intersects with $[0, x]$, and T_x denote the duration that the node spends within $[0, x]$ during the movement period. D, V, T, D_x, and T_x are random variables. As V and D are independent, V and D_x are also independent, and hence from equation (5.5), the following expression is obtained

$$P(X \leq x) = \lim_{K \to \infty} \frac{\sum_{k=1}^{K} t_{x,k}}{\sum_{k=1}^{K} t_k} = \frac{E[T_x]}{E[T]} = \frac{E[D_x]}{E[D]} \tag{5.6}$$

From equation (5.6), $E[D]$ and $E[D_x]$ are computed in order to calculate $P(X \leq x)$.

***Theorem 1**: The expected distance traveled by a mobile node over an entire movement period, in a line segment $[0, a]$, is $\frac{a}{4}$.*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Proof: As shown in Figure 5.5, a movement period has a starting point B and an ending point E in the line segment $[0, a]$. Let $D(b, e)$ be the distance of the movement period, the expected distance of movement period is expressed as

$$E[D] = \int_{e=0}^{a} \int_{b=e}^{a} D(b, e) f_B(b) f_E(e) dbde \tag{5.7}$$

Combining equations (5.3), (5.4) and (5.7) yields

$$E[D] = \int_{e=0}^{a} \int_{b=e}^{a} (b - e) \frac{1}{a} \frac{1}{b} dbde = \frac{a}{4} \tag{5.8}$$

■

Theorem 2: *The expected distance traveled by a mobile node over a movement period in a line segment* $[0, a]$ *that intersects with* $[0, x]$ *is*

$$E[D_x] = \frac{x^2}{2a}(\ln a - \ln x) + \frac{x^2}{4a} \tag{5.9}$$

Proof: For a movement period k, $e < b$ always holds. The location of a node, x, may have various location relative to b and e, as shown in Figure 5.6.

From Figure 5.6, the time that the node spends within $[0, x]$ during period k, $t_{x,k}$, and the distance of movement k that intersects with $[0, x]$, $D_{x,k}$, can be expressed as follows

$$\begin{cases} t_{x,k} = 0, \ D_{x,k} = 0, & \text{for } x \leq e < b; \\ t_{x,k} > 0, \ D_{x,k} = x - e, & \text{for } e < x < b; \\ t_{x,k} = t_k, \ D_{x,k} = b - e, & \text{for } e < b \leq x; \end{cases} \tag{5.10}$$

The expected D_x can be expressed as

$$E[D_x] = \int_{e=0}^{a} \int_{b=e}^{a} D_{x,k} f_B(b) f_E(e) dbde \tag{5.11}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

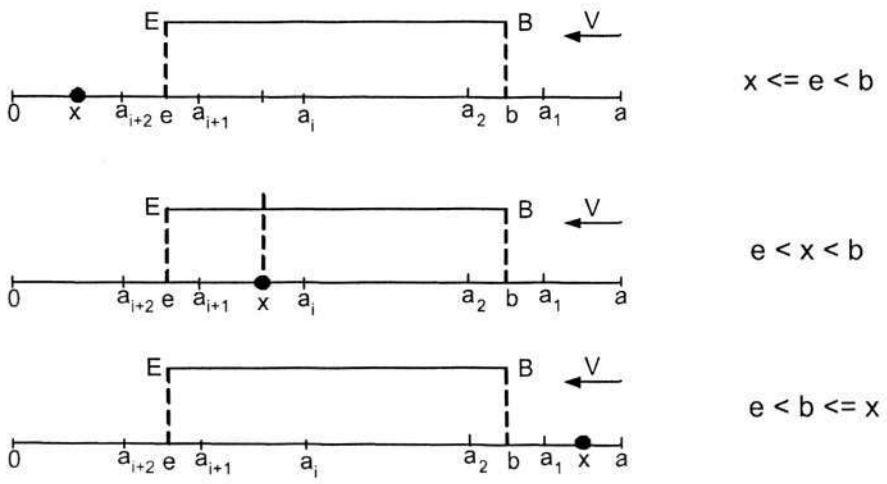


Figure 5.6: The various locations of a node.

Combining equations (5.11) and (5.10) yields

$$\begin{aligned} E[D_x] &= \int_{e=0}^{a}\int_{b=e}^{a} D_{x,k} f_B(b) f_E(e) dbde \\ &= \int_{e=0}^{x}\int_{b=x}^{a}(x-e)\tfrac{1}{a}\tfrac{1}{b}dbde + \int_{e=0}^{x}\int_{b=e}^{x}(b-e)\tfrac{1}{a}\tfrac{1}{b}dbde \\ &= \tfrac{x^2}{2a}(\ln a - \ln x) + \tfrac{x^2}{4a} \end{aligned} \tag{5.12}$$

■

From equations (5.5), (5.8) and (5.9), the following expression is obtained

$$P(X \le x) = \frac{E[D_x]}{E[D]} = \frac{\frac{x^2}{2a}(\ln a - \ln x) + \frac{x^2}{4a}}{\frac{a}{4}} = \frac{2x^2}{a^2}(\ln a - \ln x) + \frac{x^2}{a^2} \tag{5.13}$$

Therefore, the following theorem can be concluded:

***Theorem 3**: The cumulative distribution function of the location of a mobile*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

node in a line segment $[0, a]$ *is*

$$F_X(x) = P(X \leq x) = \begin{cases} \frac{2x^2}{a^2}(\ln a - \ln x) + \frac{x^2}{a^2}, & \text{for } 0 < x \leq a; \\ 0, & \text{for } x \leq 0; \\ 1, & \text{for } x > a \end{cases} \tag{5.14}$$

According to equation (5.2), the probability of a node being assigned a transmission mode i, $P_i = P(R = r_i)$, can be expressed as

$$P_i = \begin{cases} F_X(a_i) - F_X(a_{i+1}) + F_X(a_{2N-i}) - F_X(a_{2N-i+1}), & \text{for } i = [1, N-1]; \\ F_X(a_N) - F_X(a_{N+1}), & \text{for } i = N \end{cases} \tag{5.15}$$

From equations (5.14) and (5.15), the following theorem is obtained:

Theorem 4*: The probability of a mobile node being in transmission mode* i*, in a line segment* $[0, a]$*, is*

$$\begin{aligned} P_i(i \neq N) &= \frac{(2\ln a+1)(a_i^2 - a_{i+1}^2)}{a^2} - \frac{2a_i^2 \ln a_i - 2a_{i+1}^2 \ln a_{i+1}}{a^2} \\ &+ \frac{(2\ln a+1)(a_{2N-i}^2 - a_{2N-i+1}^2)}{a^2} - \frac{2a_{2N-i}^2 \ln a_{2N-i} - 2a_{2N-i+1}^2 \ln a_{2N-i+1}}{a^2} \end{aligned} \tag{5.16}$$

where a_i, a_{i+1}, a_{2N-i} and a_{2N-i+1} are the switching points of transmission mode i, and

$$P_i(i = N) = \frac{(2\ln a + 1)(a_N^2 - a_{N+1}^2)}{a^2} - \frac{2a_N^2 \ln a_N - 2a_{N+1}^2 \ln a_{N+1}}{a^2} \tag{5.17}$$

Using the definition of probability density function (PDF): $f_X(x) = \frac{\partial F_X(x)}{\partial x}$, the PDF of the location X of a mobile node moving in a line segment $[0, a]$ can be computed as

$$f_X(x) = \frac{4}{a^2}(x \ln a - x \ln x) \tag{5.18}$$

for $0 < x < a$ and 0 otherwise.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

5.3.2 Location of Optimal Relay Station

If each SS communicates with a BS directly, the expected transmission rate in the system is

$$E[R] = \sum_{i=1}^{N} P_i r_i \tag{5.19}$$

where N is the number of transmission rates that the system supports and P_i is the probability of a node transmitting with rate r_i, i.e. $P_i = P(R = r_i)$. According to existing path loss models for different terrains [129], the transmission rate drops faster than linear when the transmitter-receiver distance increases.

In this research, only one-hop or two-hop transmissions are considered between BSs and SSs as smaller topologies are preferred in most deployments, for lower complexity and lower cost in the overall system [118, 119]. In such configuration, the orthogonal resource allocation is usually applied, in which only one link is active in the network. As such, the problem of maximizing the end-to-end throughput is simplified to the problem of maximizing the end-to-end transmission rate.

Due to the fast attenuation of transmission rate over distance, transmitting through relay nodes is more efficient than transmitting directly over a long distance link. Through relaying, a single low SNR link is replaced with multiple higher SNR links over which higher transmission rates can be supported. Hence, higher end-to-end transmission rates can be achieved for SSs. In multi-hop relaying, selecting different RSs can result in different end-to-end transmission rates achieved for a particular SS. This research targets to solve the problem of lo-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

cating the optimal RS such that the end-to-end transmission rate of an SS is maximized. Relay nodes can be those carried on moving vehicles on the highway, or fixed/nomadic nodes which are mounted on light poles or towers.

Figure 5.7 depicts the generic optimal relay locating problem for the highway mobility pattern. As shown in Figure 5.7, a BS is installed at a location with a vertical distance h from the highway. H is the point on the highway with the shortest distance h to the BS. At a specific time, an SS is located x_0 away from H on the highway, and the distance between the SS and BS is d_2. The candidate RS is located between H and the SS.

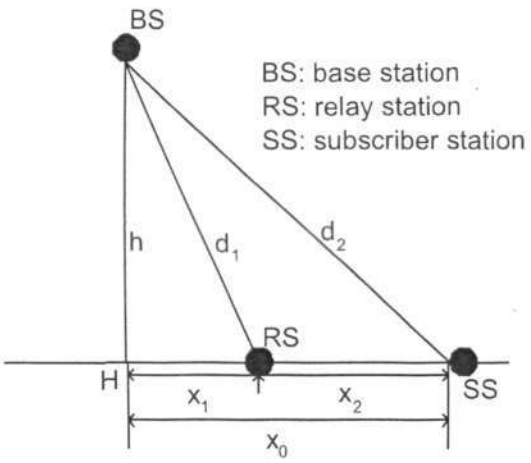


Figure 5.7: Locating the optimal relay station.

Let us denote the distance between an RS and H as x_1, the distance between the RS and a specific SS as x_2, and $x_1 + x_2 = x_0$. The distance between the BS and the RS is denoted as d_1. The problem in locating an optimal RS can then be defined as: given a path loss pattern and an SS with distance x_0 from H, find the optimal RS with a distance x_1^* from H that maximizes the end-to-end

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

transmission rate between the BS and SS.

Let the two-hop transmission rate between a BS and an SS be C_{e2e}, the one-hop transmission rate between the BS and the RS be C_1, and the one-hop transmission rate between the RS and the SS be C_2. It is assumed that orthogonal bandwidth allocation is applied such that only one link can transmit at a time within a cell controlled by a BS and there is no intra-cell interference. If the BS transmits S bits to the SS over two-hop transmissions, the BS first sends the bits to the RS and the RS in turn relays the S bits to the SS. Since only one node can transmit at a time in the cell, the transmissions over the relay link and the access link are non-concurrent. According to Kim and Liu's analysis [151], the end-to-end transmission rate is

$$C_{e2e} = \frac{C_1 C_2}{C_1 + C_2} \tag{5.20}$$

The problem in locating the optimal RS is expressed as

$$\text{Maximize: } C_{e2e} = [\frac{1}{C_1} + \frac{1}{C_2}]^{-1} = \frac{C_1 C_2}{C_1 + C_2} \tag{5.21}$$

$$\text{subject to: } x_1 + x_2 = x_0 \tag{5.22}$$

As IEEE 802.16 MAC is TDMA-based, transmissions are conducted in the allocated minislots. The number of minislots used in a given transmission is therefore inversely proportional to the transmission rate of the link. Hence, from equation (5.21), the optimal RS is the one that minimizes the total number of data minislots required by the entire route to deliver a certain amount of bits. Here, the link cost metric is defined as the inverse of the link transmission rate. Hincapie et al [82] and Shetiya et al [115] have used a similar metric in their studies.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The time-varying distance between two communicating nodes plays a major role in the link transmission rate. The rate-distance relationship can be modeled as a step function. Let x be the distance between a communicating pair, then the probability that data rate C is equal to r_i is $P(C = r_i) = P(l_{i+1} < x \leq l_i)$. The switching distances l_i and l_{i+1} are either computed according to specific path loss model, or measured empirically. $C(x)$ is a step function and can be written as

$$C(x) = \sum_{i=1}^{N} r_i \cdot 1_{A_i}(x), \ x \in (0, +\infty) \tag{5.23}$$

where N is the number of transmission rates that the system supports and $A_1, A_2, \ldots$, and A_N is a sequence of non-intersecting intervals, defined as

$$\begin{cases} A_i := (l_{i+1}, l_i], & \text{for } i = 1, 2, \ldots, N-1 \\ A_N := (0, l_N] \end{cases} \tag{5.24}$$

$1_{A_i}(x)$ is the indicator function of A_i, defined as

$$1_{A_i}(x) = \begin{cases} 1, & \text{if } x \in A_i \\ 0, & \text{otherwise.} \end{cases} \tag{5.25}$$

If transmission rates are expressed using the step function in equation (5.23), it is difficult to obtain a solution which has a generic form for the optimization problem expressed in equation (5.21). Nevertheless, using a fitted power function $C(x) = Ax^B$ [152] to express the rate-distance relationship, it becomes more tractable to find a generic form of the solution. Let the one-hop transmission rate between the BS and the RS (C_1) be modeled as $C_1(x_1) = A_1 x_1^{B_1}$. As shown in Figure 5.7, x_1 is the image of the distance d_1 on the highway segment and $x_1 = \sqrt{d_1^2 - h^2}$. Let the one-hop transmission rate between the RS and the SS

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(C_2) be modeled as $C_2(x_2) = A_2 x_2^{B_2}$, where x_2 is the distance between the RS and SS. Using the two functions ($A_1 \neq A_2$ and $B_1 \neq B_2$), the optimization problem in equation (5.21) is rephrased as

$$\text{Maximize: } C_{e2e} = \frac{A_1 x_1^{B_1} \cdot A_2 x_2^{B_2}}{A_1 x_1^{B_1} + A_2 x_2^{B_2}} \tag{5.26}$$

$$\text{subject to: } x_1 + x_2 = x_0 \tag{5.27}$$

Applying the method of Lagrange multipliers to the original problem, the optimization problem can be expressed as in a standard form as

$$\text{Minimize: } Z = -\frac{A_1 x_1^{B_1} \cdot A_2 x_2^{B_2}}{A_1 x_1^{B_1} + A_2 x_2^{B_2}} \tag{5.28}$$

$$\text{subject to: } x_1 + x_2 - x_0 = 0 \tag{5.29}$$

where $Z = -C_{e2e}$. To solve the optimization problem, the Lagrange function is

$$L = f(X) + \lambda h(X) = -\frac{A_1 x_1^{B_1} \cdot A_2 x_2^{B_2}}{A_1 x_1^{B_1} + A_2 x_2^{B_2}} + \lambda(x_1 + x_2 - x_0) \tag{5.30}$$

Then,

$$\frac{\partial L}{\partial x_1} = \frac{-A_1 A_2^2 B_1 x_1^{B_1 - 1} x_2^{2B_2}}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^2} + \lambda = 0 \tag{5.31}$$

$$\frac{\partial L}{\partial x_2} = \frac{-A_1^2 A_2 B_2 x_1^{2B_1} x_2^{B_2 - 1}}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^2} + \lambda = 0 \tag{5.32}$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 - x_0 = 0 \tag{5.33}$$

From equations (5.31) and (5.32), the following expression can be obtained

$$A_2 B_1 x_1^{-1} x_2^{B_2} = A_1 B_2 x_1^{B_1} x_2^{-1} \tag{5.34}$$

From equation (5.33),

$$x_2 = x_0 - x_1 \tag{5.35}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Substituting equation (5.35) into equation (5.34) leads to the following expression

$$x_1 = (\frac{A_2 B_1}{A_1 B_2})^{\frac{1}{B_1+1}} \cdot (x_0 - x_1)^{\frac{B_2+1}{B_1+1}} \tag{5.36}$$

Equation (5.36) is an implicit expression and it is prohibitively difficult to obtain a generic form of the solution. Fortunately, in a real system, A_1, A_2, B_1 and B_2 are known parameters. An approximate solution can be obtained using an iterative method subject to an error bound.

5.3.3 Control Messages for Location Updates in IEEE 802.16j

Two types of control messages are proposed in IEEE 802.16j standard for location updates: *MR_LOC-REQ* and *MR_LOC-RSP*. *MR_LOC-REQ* messages are transmitted by a BS to an RS to request the location information of the RS. This type of message can also be transmitted by an RS to the BS to request the location of a neighboring station by providing the MAC address of that station. The update can be carried out once or periodically. A node responds to a location request through *MR_LOC-RSP* messages, which contain the location information encapsulated in the *LLA_IE* information elements. The *LLA_IE* contains three items: Latitude, Longitude and Altitude. Besides the location information updates on RSs, an SS can also specify its capabilities for the provision of location information, through a *REG-REQ* message during the SS initialization.

In actual deployments of vehicular networks, link transmission rates at different geographical locations can be obtained based on specific terrain models or measurement results. These data can be fitted into the power functions

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$C_1(x) = A_1 x^{B_1}$ and $C_2(x) = A_2 x^{B_2}$ [152] so that the parameters A_1, B_1, A_2 and B_2 in equation (5.26) are known. When these parameters are substituted into equation (5.36), the location of the optimal RS x_1 can be solved with a given SS location x_0.

Therefore, using the proposed solution, the BS can compute the location of the optimal RS for a specific transport segment on an off-line basis. When a vehicular SS enters the segment, the BS can then decide immediately on the optimal RS selection for the particular SS based on the pre-computed data.

5.4 Summary

In this chapter, a solution for capacity-aware optimal relay selection has been proposed. The solution can be implemented to locate and select an optimal RS for a vehicular SS given the locations of the SS. We apply link adaptation technology in our analysis of link transmission rates based on channel conditions in a vehicular network. The probabilistic distribution of transmission rates are calculated based on highway mobility pattern. To achieve the maximum end-to-end transmission rate for a vehicular SS, a non-linear optimization problem is formulated and solved using the Lagrange method.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 6

Performance Analysis of Optimal Relay Selection in Multi-hop Relay Networks

6.1 Introduction

We evaluate the performance of our solution for optimal relay selection by computing the numerical results on the optimal relay selection in MR networks based on the Erceg Terrain model [153].

As the solution for optimal relay selection in vehicular networks can be applied to many transport systems, such as highways, railways, near-land rivers and seas, we focus our evaluations on two applications scenarios. In the highway vehicular network scenario, we analyze the optimal relay selection in an environment where the transmission modes are based on the Erceg terrain model. We compare the end-to-end transmission rates with those obtained for one-hop transmissions and using other relay algorithms.

In the maritime vehicular network scenario, we analyze the connectivity and routing using actual ship mobility traces in the sea lanes off the East Coast of Singapore. In our analysis, there are possibly more than one routes available

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to reach a base station for each connected node. The probabilistic distributions of the end-to-end routes are analyzed in detail, including the distribution of the number of routes for a specific node and their path lengths. The results of the connectivity study [154] have been applied to the routing protocol design [155,156] in maritime wireless mesh networks [157].

In the next subsection, the performance of the optimal relay selection solution, described in Chapter 5, will be analyzed using a case study based on the Erceg model.

6.2 Performance Analysis on Optimal Relay Selection

6.2.1 A Case Study Using Erceg Terrain Model

The performance of our solution for optimal relay selection is analyzed using a case study. The case study allows us to evaluate our solution not just for its performance but also the feasibility of our solution for practical deployment. Indeed, for this reason, we have performed our case study, based on an actual terrain model, known as the Erceg terrain model [153].

6.2.1.1 The Erceg Terrain Model

The path loss model proposed by Erceg et al [153] (named as Erceg model) is recommended by the IEEE 802.16 standard. The Erceg model covers three most common suburban terrain categories found across the United States. The path loss model for each terrain is named Erceg A, Erceg B and Erceg C. Terrain category A is hilly terrain with moderate-to-heavy tree densities. Terrain category

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

C is mostly flat terrain with light tree densities. Terrain category B is between A and C. Among the three models, the transmission rate in the Erceg A model attenuates most rapidly over distance while the transmission rate in the Erceg C model attenuates least rapidly. The Erceg B model captures the intermediate path loss condition, which is considered in the case study.

6.2.1.2 The Optimal Solution

The case study illustrated in Figure 6.1 is based on the generic optimal relay selection problem for the highway mobility pattern, as shown in Figure 5.7. In Figure 6.1, $g_i, i = 1, 2, ...$ are distances to H; they are the boundaries where transmission modes are switched, and are calculated using $g_i = \sqrt{l_i^2 - h^2}$.

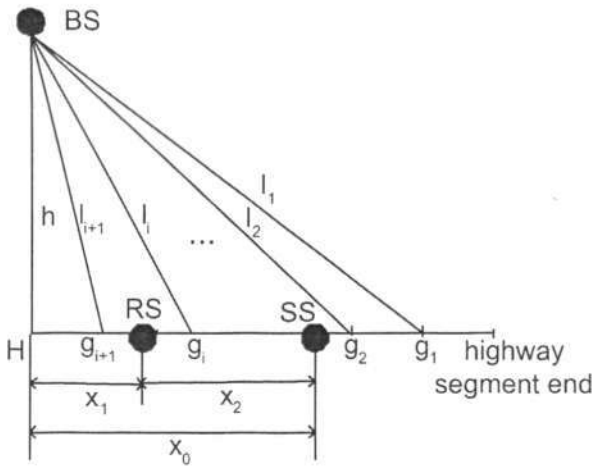


Figure 6.1: The case study of the optimal relay selection.

The Erceg B data computed by Joseph and Martens [129] has been re-tabulated in Table 6.1, assuming $h = 1$ km and 10% of MAC overhead [81].

Assuming that the length of the highway segment is 7 km and $h = 1$ km, the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 6.1: The relationship between transmission mode and distances in Erceg B model

modulation and coding scheme	transmission rate (Mbps)	distance (l_i) (km)	distance (g_i) (km)
BPSK 1/2	1.269	3.2	3.0397
QPSK 1/2	2.538	2.7	2.5080
QPSK 3/4	3.816	2.5	2.2913
16-QAM 1/2	5.085	1.9	1.6155
16-QAM 3/4	7.623	1.7	1.3748
64-QAM 2/3	10.161	1.3	0.8307
64-QAM 3/4	11.439	1.2	0.6633

probabilities of one-hop transmission rate in Erceg B model are calculated using Theorem 4 as in Chapter 5. The probabilities are shown in Table 6.2. In such a configuration, 3.64% of nodes do not have direct connection to the BS.

Table 6.2: The probability distribution of one-hop transmission rates

modulation and coding scheme	Data rate (Mbps)	Probability
Null (unconnected)	0	0.0364
BPSK 1/2	1.269	0.1004
QPSK 1/2	2.538	0.0537
QPSK 3/4	3.816	0.2032
16-QAM 1/2	5.085	0.0823
16-QAM 3/4	7.623	0.1995
64-QAM 2/3	10.161	0.0641
64-QAM 3/4	11.439	0.2604

Using a curve fitting software, the data in Table 6.1 is fitted into a power function $C(x) = Ax^B$, in order to get the optimal solution for equation (5.36). In the power functions used in equation (5.26), the fitted parameters are: $A_1 = 7.638$, $B_1 = -1.109$, $A_2 = 16.01$, $B_2 = -1.706$.

Substituting A_1, B_1, A_2 and B_2 into equation (5.36), the following expression is obtained

$$x_1 = 0.05852 \times (x_0 - x_1)^{6.4771} \tag{6.1}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For a given x_0, an iterative method is used to solve x_1 in equation (6.1). For example, given $x_0 = 3.0$ and an error not more than 1%, the solution of equation (6.1) is $x_1 = 1.372$. This implies that the solution for the optimization problem in equation (5.26) is: $x_1^* = 1.372$, $x_2^* = 1.628$. The interpretation of this solution is: when an SS is 3.0 km away from location H (possibly on both sides), and an RS with a distance of 1.372 km from H is selected, the end-to-end transmission rate between the BS and SS is maximum.

We have proved the sufficient condition exists for the solutions to be optimized. The proof can be found in Appendix A.

The solution $x_1^* = 1.372$ and $x_2^* = 1.628$ minimizes Z in the optimization problem expressed in equation (5.28) and maximizes C_{e2e} in the original optimization problem expressed in equation (5.26).

6.2.1.3 Selection Margin of the Optimal RS Location

In a practical system, the optimal solution x_1^* generally would fall between two switching points, i.e. $g_{i+1} < x_1^* \leq g_i$ for some i, where a specific capacity r_i of BS-RS link is sustained. Therefore, the optimal RS's distance to H can be anywhere between x_1^* and g_i without compromising the capacities of both BS-RS link and RS-SS link[1]. We call the region of $[x_1^*, g_i]$ in Figure 6.1 *selection margin*. As long as the optimal RS is in the *selection margin*, the target end-to-end capacity can be achieved. The concept of the selection margin is depicted in Figure 6.2.

[1]This is because if a point x_1 between x_1^* and g_i^* is chosen, the rate of BS-RS link remains unchanged in the region of $[x_1^*, g_i]$ and the rate of SS-RS link may increase due to the decrease of x_2 ($x_1 + x_2$ is a constant).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

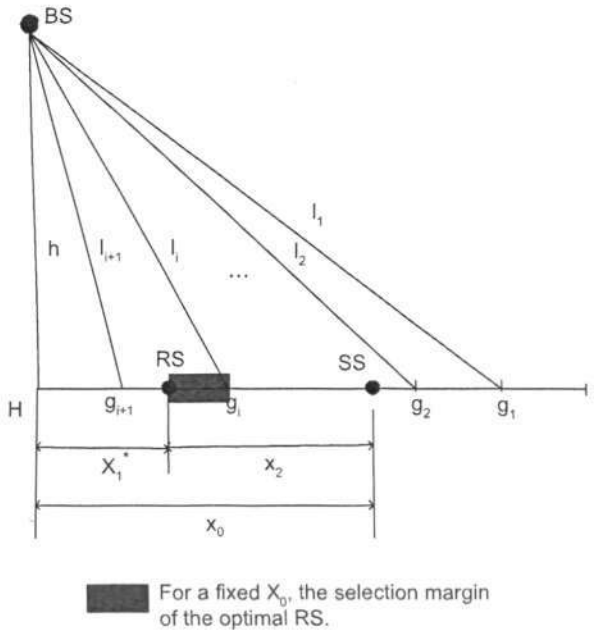


Figure 6.2: The concept of selection margin in optimal relay selection.

6.2.2 Performance Evaluation on Optimal Relay Selection Using a Case Study

We evaluate the performance of our solution for optimal relay selection in a case study and present numerical results to validate the correctness of our analysis. Based on Table 6.1 and equation (5.21), a two-hop transmission can achieve a maximal end-to-end rate of 5.719 Mbps, given that both of the two-hop links have a maximal transmission rate 11.439 Mbps. Thus, if the one-hop transmission rate between the BS and SS is larger than 5.719 Mbps, there is no gain in transmission rate when applying two-hop relaying. According to the data in Table 6.1, if the one-hop transmission rate between the BS and SS is equal to or larger than 7.623 Mbps, one-hop transmission is better than two-hop transmission.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The achievable end-to-end transmission rate is computed between the SS and BS when the SS is located somewhere on the highway. The location of SS is represented by the distance between the SS and the reference point O on the highway, as shown in Figure 5.3.

Figure 6.3 shows the end-to-end transmission rate comparison when using one-hop transmission or optimal two-hop relaying method. From Figure 6.3, if the SS's location falls into the range of $[2.1252, 4.8748]$, the one-hop transmission rate between the BS and SS is equal to or larger than 7.623 Mbps, thus one-hop transmission is applied. Otherwise, if the SS's location falls outside the range of $[2.1252, 4.8748]$, two-hop transmission achieves higher end-to-end transmission rate. From Figure 6.3, it can also be seen that the SS which is farther away from the BS can gain more from the optimal two-hop relaying, in general, compared to the SS which is closer to the BS.

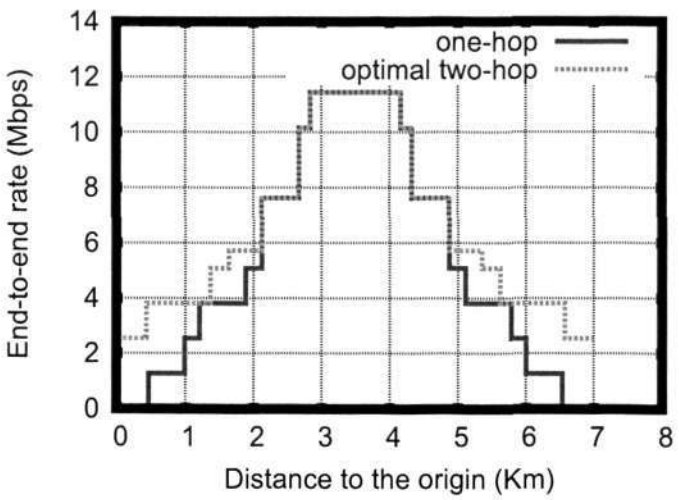
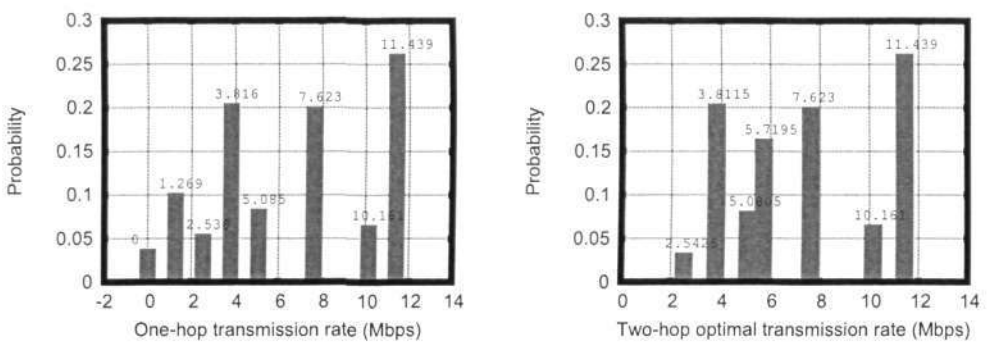


Figure 6.3: The end-to-end transmission rate of an SS at different locations.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The probabilities of the one-hop transmission rates and end-to-end transmission rates with optimal two-hop relaying are calculated and shown in Figure 6.4, where the numbers on top of each column are the transmission rates in Mbps. From Figure 6.4, as compared to the one-hop transmission rates, the end-to-end transmission rates with optimal two-hop relaying are improved significantly. This is even more obvious for the SSs with relatively lower one-hop transmission rates, where the expected end-to-end transmission rate can be increased by 49.86%. Furthermore, optimal two-hop relaying extends the network coverage from 96.36% to 100%[2].



(a) One-hop transmission. (b) Optimal two-hop relaying.

Figure 6.4: The probability distributions of the SS-BS transmission rates.

As Figure 6.4 shows the individual probabilistic distribution of SS-BS rates with one-hop and optimal two-hop transmissions, it may not be able to clearly illustrate the performance difference. The cumulative distributions of the SS-BS transmission rates are calculated using one-hop transmissions and two-hop

[2]With one-hop transmission only, 3.64% of the highway area are uncovered (with rate 0), as shown in Table 6.2.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

optimal relaying, as shown in Figure 6.5. It can be seen that with one-hop transmission, 90% of the nodes achieve 0.6 Mbps or larger transmission rate, while with optimal two-hop transmission, 90% of the nodes can achieve 3 Mbps or larger transmission rate. Comparably, with one-hop transmission, 70% of the nodes achieve 3.2 Mbps or larger transmission rate, while with optimal two-hop transmission, 70% of the nodes can achieve 5.1 Mbps or larger transmission rate. Since relaying is not applied when SS-BS one-hop transmission rate is equal to or larger than 7.623 Mbps, the probability of the SS-BS transmission rate being at least 7.623 Mbps does not change. It can be concluded from Figure 6.5 that applying optimal two-hop relaying can increase SS-BS transmission rate significantly.

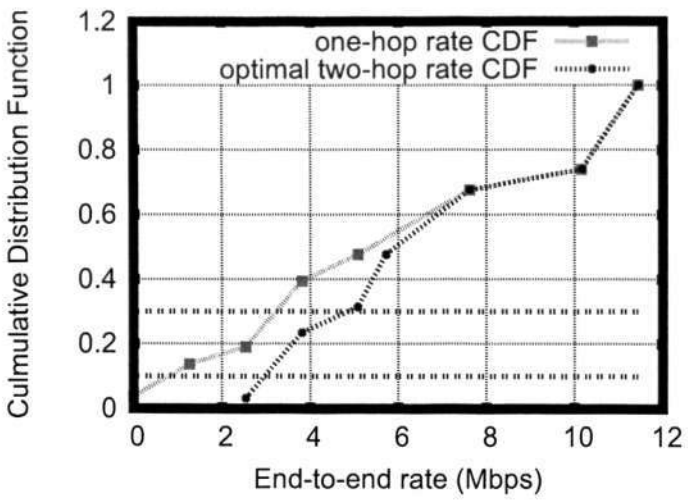


Figure 6.5: Comparison of the CDF of SS-BS transmission rates using one-hop and two-hop optimal relaying.

There are different methods for selecting an RS for an SS. The RS may be randomly located between H and itself, or be located at the middle-point

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between H and itself ($x_1 = 0.5x_0$)[3]. The SS-BS transmission rates achieved using different methods are computed: (1) using one-hop transmission only, (2) randomly selecting an RS, (3) selecting the RS at the middle point, and (4) selecting the optimal RS. Table 6.3 shows the SS-BS transmission rates achieved using different methods given an SS with distance X_0 from H. From Table 6.3, it can be seen that optimal two-hop relaying method always achieves the highest transmission rates as compared to other methods.

Table 6.3: The SS-BS transmission rate comparison using different relaying methods.

SS-BS rate (Mbps)	$x_0 = 1.8$ (km)	$x_0 = 3.0$ (km)	$x_0 = 3.5$ (km)
one-hop transmission	3.816	1.269	0
middle-point two-hop relaying	4.574	3.05	2.18
random two-hop relaying	5.381	2.7742	2.077
optimal two-hop relaying	5.7195	3.8115	2.543

6.3 Connectivity and Routing Analysis in Maritime Vehicular Networks

The feasibility study of multi-hop relay networks in terms of the connectivity and routing analysis can be seen in the Appendix.

6.4 Summary

This chapter presents the evaluation results on our solution for optimal relay selection in multi-hop relay wireless networks in terms of feasibility and network performance.

[3]The notations are shown in Figure 5.7.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In the performance analysis of our solution for optimal relay selection using a case study based on the Erceg terrain model, our numerical results obtained shows that the end-to-end transmission rate can be significantly increased by optimally selecting the appropriate relay. Our solution for optimal relay selection is applicable to a generic class of vehicular communication systems for deployment on highways, trains, rivers and near land seas. Although our study was based on IEEE 802.16j networks, in principle, the method developed here can also be applied to other vehicular networks such as IEEE 802.11 and multi-hop cellular networks, provided that the link transmission rate and the terrain models are known in the system. Our approach in the feasibility study and performance analysis are generic, and can be applied to any other transport systems, so long as the mobility traces and terrain model are available for that transport system.

In the feasibility study as depicted in the Appendix, a maritime vehicular network usage scenario was considered. Through connectivity and routing analysis, this study confirms the feasibility of a maritime vehicular network in the shipping lanes at the Singapore East Coast. Such an analysis is unique as it uses the actual ship mobility traces and not simply based on a statistical mobility model. The results verify the possible use of multi-path routing protocols, and the short path length implies a potentially acceptable end-to-end delay for real-time traffic.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 7

Conclusions and Future Work

We conclude this thesis by summarizing our findings and highlighting our contributions in the areas of end-to-end capacity estimation and optimal relay selection in IEEE 802.16 wireless multi-hop networks. We also suggest some further work.

7.1 Conclusions and Contributions

The research work in this thesis focuses on capacity estimation for QoS provisioning and optimal relay selection for efficient QoS routing in IEEE 802.16 multi-hop wireless networks. We have completed critical reviews of existing related work and proposed new solutions for accurate capacity estimation and optimal relay selection. We have carried out extensive simulations to establish the correctness and accuracy of our proposed solutions. Our solutions when implemented will contribute to improved network performance in IEEE 802.16 wireless multi-hop networks.

We highlight our research findings with our contributions in the following areas:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1. End-to-End Capacity Estimation. In order to assure end-to-end service guarantees, accurate and reliable determination of capacities is crucial for robust network control and management. We proposed in Chapter 3 new solutions for real-time link and end-to-end capacity estimations in IEEE 802.16 wireless mesh networks. We introduced the concept of *Bottleneck Zone* to estimate the end-to-end capacity by taking multi-hop interference into consideration. Hence, our solutions for capacity estimation is both robust and accurate in capacity estimation. Furthermore, the proposed solutions can be fully incorporated in existing routing protocols for capacity discovery and efficient computation of real-time end-to-end capacity in a distributed manner. We have also developed our solutions based on mesh networks with vastly different optimal and random node placement patterns, and hence the solutions developed here remain applicable to wide range of wireless configurations.

   Extensive simulations were conducted to study and evaluate the performance of the proposed capacity estimation scheme. The simulation results, presented in Chapter 4, show that the proposed capacity estimation solution is accurate, responsive and robust in defining lower bound capacities, and it can adapt to various network topologies and deployment patterns. The results also show that the upper bound capacity is optimistic and hard to achieve in distributed networks because it requires global information knowledge and optimal traffic pattern to make scheduling decisions. Furthermore, we observed that the addition of new nodes in randomly deployed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

networks can actually improve the end-to-end capacity due to the availability of shorter paths from a source node to a sink node in denser networks caused by better connectivity.

2. Optimal Relay Selection. We explored the optimal relay selection problem in Chapter 5 to maximize the end-to-end capacity for vehicle-to-infrastructure communications in IEEE 802.16j multi-hop wireless networks. We developed an analytical model for locating and selecting the optimal RS based on geographical locations of vehicular SSs. Our relay selection model optimizes the route capacity between the BS and a particular SS by selecting the higher capacity links and thus minimizes the number of minislots required for transmissions between the BS and SS.

   We verified our optimal relay selection model with a case study based on link capacity data extracted from the Erceg terrain model in Chapter 6. Our analysis on connectivity and route statistics in multi-hop vehicular networks are based on actual mobility traces and our results confirmed the feasibility of deploying and operating multi-hop vehicular networks. We established that the optimal selection of RS can lead to an increase of end-to-end data rate for individual SSs by nearly 50% as compared to direct transmissions without using relays.

   In actual implementations, the optimal RS for a specific transport segment can be pre-computed off-line by the BS using our proposed solution. When a vehicular SS enters the segment, the BS can then decide immediately on

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the optimal RS selection for the particular SS based on the pre-computed data. Our optimal relay selection solution is applicable to vehicular communication systems deployed on highways, trains, rivers and seas. Although this study was based on IEEE 802.16j networks, in principle, the solution developed here can also be applied to other vehicular networks such as IEEE 802.11 and multi-hop cellular networks.

7.2 Future Work

We present discussions on some areas in this research that need further study and suggestions for future work. Some of the areas which should be explored in the framework of this thesis are suggested below.

1. Impact of scheduling algorithms on capacity estimation. It was observed that scheduling algorithms affect network capacity in multi-hop wireless networks. Scheduling algorithms are open issues in wireless network standards such as IEEE 802.16 and IEEE 802.11, and thus network developers possibly adopt scheduling algorithms that significantly differ from one another. The algorithms for scheduling data transmissions for multi-user and multi-class of traffic over multi-hop links could have a variety of design objectives, such as maximizing throughput, maximizing fairness among users, maximizing revenue, etc. The impact of scheduling algorithms on capacity estimation in multi-hop wireless networks has seldom been explored and needs further investigation.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2. Impact of interference models on capacity estimation. As discussed in Chapter 2, various interference models have been assumed in capacity studies in the literature. In the capacity estimation research presented in this thesis, we assumed the K-hop interference model which is an extension of the Protocol Model. As described in Chapter 2, the Physical Model is more practical as compared to the Protocol Model; but the former is more complicated than the latter in capacity analysis. As a result, the Protocol Model and its extensions may be more suitable for networks deployed in homogenous terrains. A possible extension to capacity estimation in this thesis is to explore capacity in WMNs considering other interference models.

3. Impact of mobility on channel capacity in rate-based optimal relay selection. In mobile environments, a vehicular SS achieves different data rates even when the SS is at the same distance from the BS but moving towards or away from the BS due to Doppler effect. In such a case, the rate switching points in the highway model as shown in Figure 5.3 become asymmetric. Consider this, our current model is a special case when vehicles move in relatively slow speeds where Doppler effect on link capacity is weak and can be ignored in the analysis. The asymmetric properties of link capacities due to Doppler effect in the highway model should be studied in future work.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

APPENDIX

A. Proving That the Sufficient Condition Holds in the Multi-variable Optimization Problem

This section proves that the sufficient condition [158] holds in the following multi-variable optimization problem

$$\text{Minimize: } Z = -\frac{A_1 x_1^{B_1} \cdot A_2 x_2^{B_2}}{A_1 x_1^{B_1} + A_2 x_2^{B_2}}$$

$$\text{subject to: } x_1 + x_2 - x_0 = 0$$

Sufficient condition: *Let X^* be a solution satisfying $h(X^*) = 0$, and a vector $\lambda^* = [\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*]$ such that $\nabla_X L(X^*, \lambda^*) = 0$. Suppose the matrix $\nabla^2_{x_1 x_2} L(X^*, \lambda^*)$ is positive definite on $M = \{Y : \nabla h(X^*)^T Y = 0\}$* , that is, for $Y \in M$, $Y \neq 0$ there holds $Y^T \nabla^2_{x_1 x_2} L(X^*, \lambda^*) Y > 0$. Then $Z = f(X^*)$ is a strict local minimum subject to $h(X) = 0$[1].

Proof: To find $M = \{Y : \nabla h(X^*)^T Y = 0\}$.

$$\nabla h(X^*) = \begin{pmatrix} 1 & 1 \end{pmatrix}^T \tag{7.1}$$

[1]The detail description of the sufficient condition can be found in [158]

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

and

$$\nabla h(X^*)^T Y = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ \\ Y_2 \end{pmatrix} = Y_1 + Y_2 = 0 \tag{7.2}$$

That is: $M = \{Y : Y_1 + Y_2 = 0\}$. We note that

$$\nabla^2_{x_1 x_2} L = \begin{pmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \\ \frac{\partial^2 L}{\partial x_2^2} & \frac{\partial^2 L}{\partial x_2 \partial x_1} \end{pmatrix} \tag{7.3}$$

From equation (5.30), we obtain

$$\frac{\partial^2 L}{\partial x_1^2} = \frac{A_1 A_2^2 B_1 x_2^{2B_2} x_1^{B_1-2}(A_1 B_1 x_1^{B_1} - A_2 B_1 x_2^{B_2} + A_2 x_2^{B_2} + A_1 x_1^{B_1})}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^3} \tag{7.4}$$

$$\frac{\partial^2 L}{\partial x_1 \partial x_2} = \frac{-2A_1^2 A_2^2 B_1 B_2 x_1^{2B_1-1} x_2^{2B_2-1}}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^3} \tag{7.5}$$

$$\frac{\partial^2 L}{\partial x_2 \partial x_1} = \frac{-2A_1^2 A_2^2 B_1 B_2 x_1^{2B_1-1} x_2^{2B_2-1}}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^3} \tag{7.6}$$

$$\frac{\partial^2 L}{\partial x_2^2} = \frac{A_1^2 A_2 B_2 x_1^{2B_1} x_2^{B_2-2}(A_2 B_2 x_2^{B_2} - A_1 B_2 x_1^{B_1} + A_1 x_1^{B_1} + A_2 x_2^{B_2})}{(A_1 x_1^{B_1} + A_2 x_2^{B_2})^3} \tag{7.7}$$

From equation (7.2),

$$Y_1 + Y_2 = 0 \Rightarrow Y_2 = -Y_1 \tag{7.8}$$

Thus,

$$\begin{aligned} Y^T \cdot \nabla^2_{x_1 x_2} L \cdot Y &= \begin{pmatrix} Y_1 & -Y_1 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \\ \frac{\partial^2 L}{\partial x_2^2} & \frac{\partial^2 L}{\partial x_2 \partial x_1} \end{pmatrix} \begin{pmatrix} Y_1 \\ \\ -Y_1 \end{pmatrix} \\ &= Y_1^2 (\tfrac{\partial^2 L}{\partial x_1^2} + \tfrac{\partial^2 L}{\partial x_2^2} - 2 \times \tfrac{\partial^2 L}{\partial x_1 \partial x_2}) \end{aligned} \tag{7.9}$$

Since $Y_1 \neq 0$, $Y_1^2 > 0$, therefore in order to prove $Y^T \cdot \nabla^2_{x_1 x_2} L \cdot Y > 0$, we need to prove

$$\frac{\partial^2 L}{\partial x_1^2} + \frac{\partial^2 L}{\partial x_2^2} - 2 \times \frac{\partial^2 L}{\partial x_1 \partial x_2} > 0 \tag{7.10}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

From equations (7.4-7.7), we obtain

$$\begin{aligned}\frac{\partial^2 L}{\partial x_1^2}+\frac{\partial^2 L}{\partial x_2^2}-2\times\frac{\partial^2 L}{\partial x_1\partial x_2} &= \frac{A_1A_2^2B_1x_2^{2B_2}x_1^{B_1-2}(A_1B_1x_1^{B_1}-A_2B_1x_2^{B_2}+A_2x_2^{B_2}+A_1x_1^{B_1})}{(A_1x_1^{B_1}+A_2x_2^{B_2})^3}\\ &+\frac{A_1^2A_2B_2x_1^{2B_1}x_2^{B_2-2}(A_2B_2x_2^{B_2}-A_1B_2x_1^{B_1}+A_1x_1^{B_1}+A_2x_2^{B_2})}{(A_1x_1^{B_1}+A_2x_2^{B_2})^3}\\ &+\frac{4A_1^2A_2^2B_1B_2x_1^{2B_1-1}x_2^{2B_2-1}}{(A_1x_1^{B_1}+A_2x_2^{B_2})^3}\end{aligned} \tag{7.11}$$

In the generic form of $C_1(x_1) = A_1x_1^{B_1}$ and $C_2(x_2) = A_2x_2^{B_2}$, $C_1(x_1)$ and $C_2(x_2)$ are non-negative and monotonically decreasing with respect to increases in distance x_1 and x_2. Thus, $A_1, A_2 > 0$ and $B_1, B_2 < 0$ for $x_1, x_2 > 0$, and hence

$$(A_1x_1^{B_1}+A_2x_2^{B_2})^3 > 0 \tag{7.12}$$

In order to prove that expression (7.11) is larger than 0, and since the denominator is already larger than 0 (expression (7.12)), we need to prove that the numerator Δ is also larger than 0, as

$$\begin{aligned}\Delta &= A_1A_2^2B_1x_2^{2B_2}x_1^{B_1-2}(A_1B_1x_1^{B_1}-A_2B_1x_2^{B_2}+A_2x_2^{B_2}+A_1x_1^{B_1})\\ &+ A_1^2A_2B_2x_1^{2B_1}x_2^{B_2-2}(A_2B_2x_2^{B_2}-A_1B_2x_1^{B_1}+A_1x_1^{B_1}+A_2x_2^{B_2})\\ &+ 4A_1^2A_2^2B_1B_2x_1^{2B_1-1}x_2^{2B_2-1} > 0\end{aligned} \tag{7.13}$$

With the parameters $A_1 = 7.638$, $B_1 = -1.109$, $A_2 = 16.01$, $B_2 = -1.706$, and the optimal solution obtained previously $x_1^* = 1.372$, $x_2^* = 1.628$ for a given $x_0 = 3.0$, we now obtain the value of Δ as

$$\Delta = 1341.0523 > 0 \tag{7.14}$$

Therefore, the sufficient condition holds and x_1^*, x_2^* are the optimal solutions.

■

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

B. Connectivity and Routing Analysis in Maritime Networks

B.1 Background

In maritime vehicular networks, the coverage extension can be achieved by using ships as relay nodes for other ships that are further away from the land. The maritime MR network connects to the terrestrial networks via land stations at shore as illustrated in Figure 7.1.

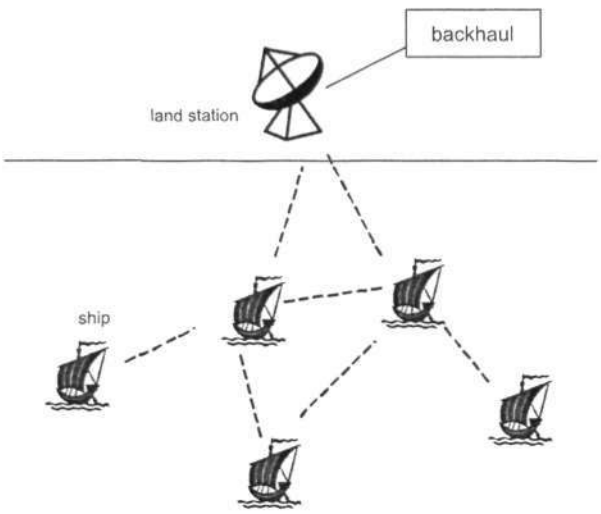


Figure 7.1: A maritime vehicular network.

In this study, the term *node connectivity* is defined as the percentage of time that a node is connected to the land station. Then, (system level) *connectivity* is defined as the average value of *node connectivities* for all nodes in the network. In a maritime vehicular network, the nodes which are ships may move in an autonomous way according to navigation schedules, rules and regulations. With this dynamic topology, the connectivity may be affected and varies from time

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to time. An acceptable level of connectivity is necessary to ensure feasibility of the network. While it is subjective to judge which level of connectivity is needed to make a network feasible, Santi and Blough [134] have suggested that the connectivity level of 1.0, 0.9 and 0.1 are indicative of three different application scenarios that a multi-hop ad hoc network must satisfy. In the first level, the network is used for life-critical applications such as the one for public safety. In the second level, occasional network disconnections are tolerable for non-critical applications. In the last level, the network is disconnected most of the time, but temporary connection periods can be used to exchange data among nodes for applications like environmental monitoring.

In this study, a maritime communication network is considered feasible when 90% of nodes have node connectivity of 0.9. In order to capture the conditions of the feasibility to have such a network at the Singapore East Coast, this study analyzes the connectivity based on actual mobility traces of ships for the area. Literature survey shows that none of the existing connectivity studies has been done for maritime communication networks. More importantly, none of them is based on actual mobility traces of ships.

B.2 System Model

The system is modeled for maritime vehicular communications in the shipping lanes of the Singapore East Coast located in the rectangular region as shown in Figure 7.2. The north east corner of the region is located at latitude 1.3167 and longitude 104.15. The south west corner of the region is located at latitude

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.20 and longitude 103.867. Figure 7.3 shows an enlargement of the region of interest, and the number inside each square bracket is the identifier of a ship in this region. The land station is located at the northernmost site (at latitude 1.3117 and longitude 104.0), labeled as "[1]" in Figure 7.3. The dotted lines define the shipping lanes where ship movements are supposedly confined to. In this figure, some ships are outside the boundary of the map, which implies that those ships are not present in the region of interest at that moment.

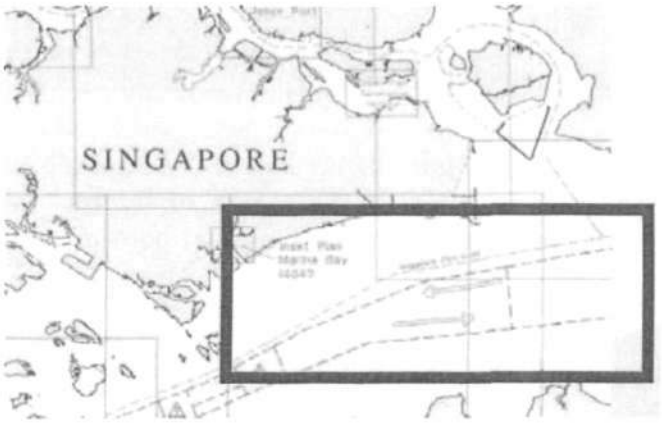


Figure 7.2: The region of the maritime communication network at the Singapore East Coast.

Instead of using any statistical mobility model to describe nodes in the shipping lanes, actual mobility traces of ships observed in the region are used. The actual ship mobility traces are derived from the Automatic Identification System (AIS) [159] which contains instantaneous information of the ships. Since January 2005, all ships over 300 gross space units are required by the International Maritime Organization (IMO) to install AIS equipment. Through the AIS, each ship broadcasts its information which includes position, velocity, length of the ship,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

destination port and expected time of arrival. The AIS contributes primarily to traffic safety on shipping lanes, and such data can also be used in communication services because ship data is always up-to-date and can be retrieved at any time.

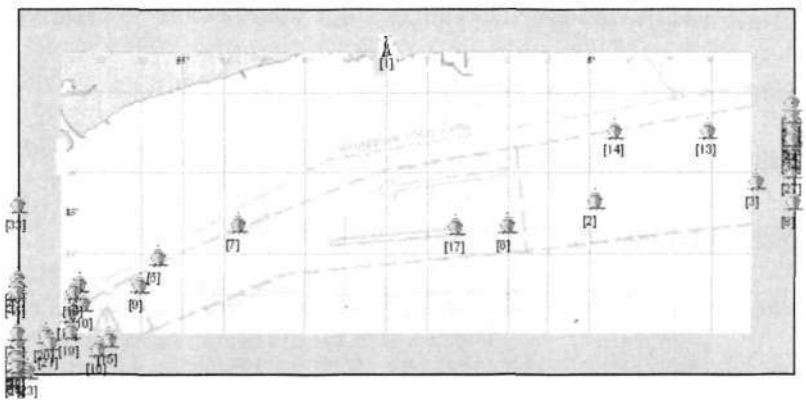


Figure 7.3: The enlargement of the region of interest for the maritime communication network.

The AIS data that is used in this study are recorded in two days, each for a duration of ten hours from 0:00 UTC to 10:00 UTC. From all the collected AIS data, only the trace which is within the region of interest is kept for processing. Ships which travel at speeds slower than 1 knot (1.852 km/hour) are not considered. After the filtering process, 132 and 104 ships appear in the region in day 1 and day 2, respectively.

B.3 Connectivity Analysis

Connectivity is considered to be independent of the traffic load in a network. In an actual system, two connected nodes may not communicate with each other at any given moment in time due to transmission interference caused by simultaneous communications between other nodes in the network. In this connectivity

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

study, we consider a topology-based scheme which does not consider interference in order to provide a guideline for network deployment and configurations. Therefore, two nodes are considered connected if they are within transmission range of each other. The transmission range is assumed to be a perfect circle around the transmitting node.

The study focuses on the connectivity between the nodes and the land station. A node is considered connected to the land station if it has at least one route to the land station, and the path length must not exceed three hops to minimize any excessive end-to-end delay as suggested by Pabst et al [117].

Intuitively, connectivity depends on the transmission range of a node and increasing the transmission range will improve the connectivity. The impact of transmission range on connectivity is quantified in this study. We refer to a node on board a ship as a subscriber station (SS) and a land station as a base station (BS). BS-SS transmission range is assumed to be symmetric between the BS and an SS. Similarly, SS-SS transmission range is assumed to be symmetric between any two SSs.

In our analysis, BS-SS transmission range is fixed at 10 km and the connectivity is computed for different SS-SS transmission ranges. This process is repeated for different BS-SS transmission ranges at 15 km and 20 km. Results in Figure 7.4 confirm that connectivity does increase with larger values of BS-SS and SS-SS transmission ranges. Notably, the connectivity is almost 1.0 for all values of SS-SS transmission ranges larger than 0.5 km when BS-SS transmission range is 20 km. This is due to the fact that almost all SSs can be connected to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the BS directly at such a large BS-SS transmission range. In order to achieve average node connectivity of 0.9, the SS-SS transmission range must be not shorter than 7 km for the BS-SS transmission range of 10 km. While, to achieve the same average connectivity for a BS-SS transmission range of 15 km, the SS-SS transmission range must be not shorter than 2.4 km.

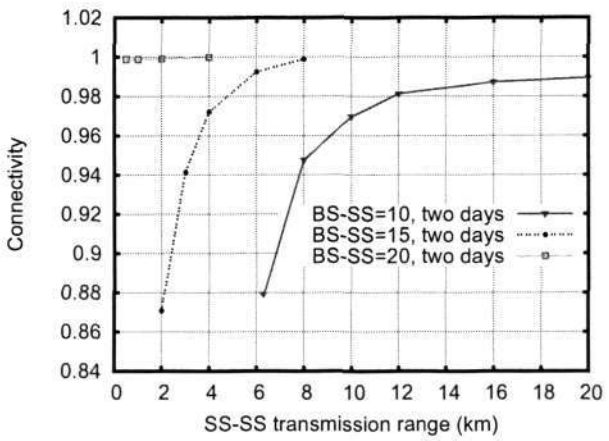


Figure 7.4: Connectivity at different transmission ranges.

Figure 7.4 presents the mean node connectivity, whereas Figure 7.5 presents the cumulative distribution functions (CDFs) of the node connectivity. When BS-SS and SS-SS transmission ranges are 10 km and 6.3 km, respectively, from Fig 7.5(a), it can be seen that more than 90% of the nodes has more than 0.6 node connectivity. This means 90% of the ships can connect to the land station at least 60% of time. The 90 percentile node connectivity increases to 0.77, when SS-SS transmission range increases to 8 km as illustrated in Figure 7.5(a). The 90 percentile node connectivity increases further to 0.87, when SS-SS transmission range increases to 12 km, as depicted in Figure 7.5(a). When BS-SS transmission

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

range is increased to 15 km, the CDF of node connectivity reveals the same trend. For BS-SS transmission range being 20 km, the connectivity approaches to 1.0 when the SS-SS transmission ranges are larger than 0.5 km, which conforms to the results as shown in Figure 7.4.

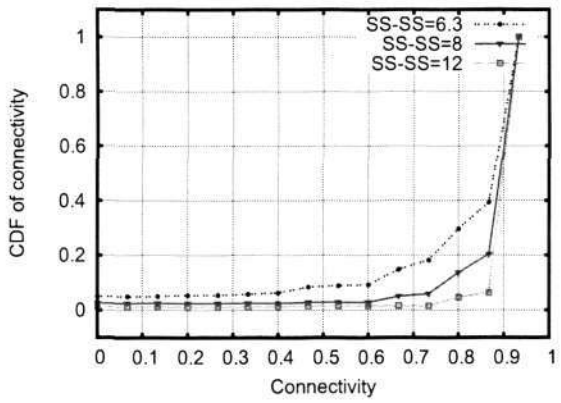
From the observations in this study, if the node density is low in the vicinity of an SS and the SS-SS transmission range is short, the SS is easily disconnected from the BS for a period of time. However, if BS-SS transmission range is large enough, the SS can connect to BS directly with one hop. As a result, with a large BS-SS transmission range, connectivity can be high even though SS-SS transmission range is short. This can be seen when BS-SS transmission range is 20 km, almost all SS have 1.0 node connectivity, even though the SS-SS transmission range is only 0.5 km (c.f Figure 7.5(c)).

From this study and the connectivity results, it can be concluded that it is feasible to have a maritime vehicular network at the Singapore East Coast for specific BS-SS and SS-SS transmission range configurations.

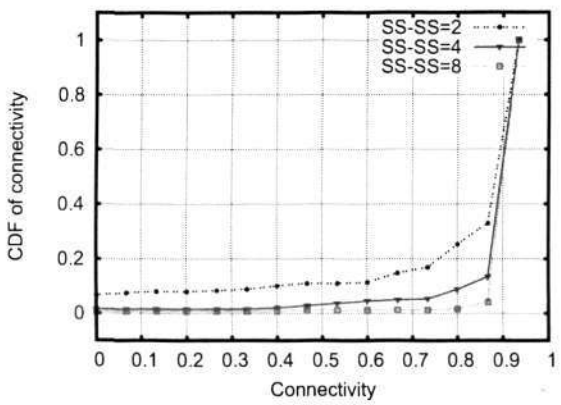
B.4 Routing Analysis

Wireless links are unstable to some degree, resulting in end-to-end route instability. If there exist multiple routes between a ship and the land station, multi-path routing can help in achieving high transport reliability because of route redundancy. If one route fails temporarily, another route can be used to forward packets. As such, route redundancy can help in improving connectivity. The number of node-disjoint paths of an SS is denoted as the sum of all paths

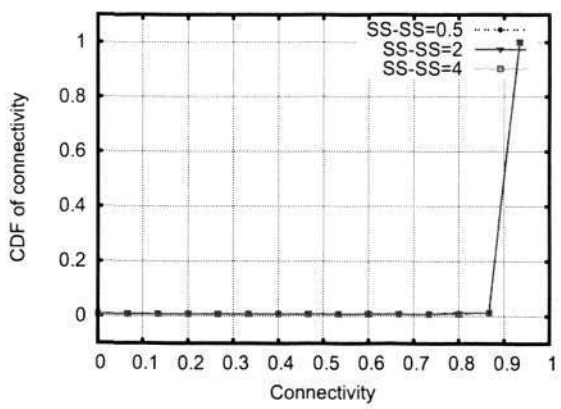
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) L_a=10km



(b) L_a=15km



(c) L_a=20km

Figure 7.5: CDFs of node connectivity (L_a is BS-SS transmission range).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

without any common relay node between this SS and the BS. Therefore, the larger the number of node-disjoint paths, the higher the route redundancy in the system.

B.4.1 Node-disjoint Path Analysis

In this study, the path length is considered to be not more than three hops. The average number of node-disjoint paths of all SSs in the system are illustrated in Figure 7.6, for different BS-SS and SS-SS transmission ranges. Generally, the larger the transmission ranges, the more node-disjoint paths are available. It can be seen that if the BS-SS transmission range is relatively large, for example, 15 km and 20 km, the number of node-disjoint paths has a similar pattern. This is because when the BS-SS transmission range is relatively large, such a range is a dominant factor on the number of node-disjoint paths, because the BS has more one-hop neighbors to act as relay nodes for other nodes. When the BS-SS transmission range is short, for example, 10 km, increasing SS-SS transmission range only marginally increases the number of node-disjoint paths.

Figure 7.6 shows the average number of node-disjoint paths per SS in the system, whereas Figure 7.7 shows the probability density function (PDF) of the number of node-disjoint paths per SS. It can be seen that for the BS-SS and SS-SS transmission ranges being 15 km and 4 km, respectively, the majority of SSs have less than 20 node-disjoint paths. If the SS-SS transmission range increases to 8 km, the majority of SSs have less than 30 node-disjoint paths. This result shows that increasing the SS-SS range helps to improve route redundancy. In the case of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

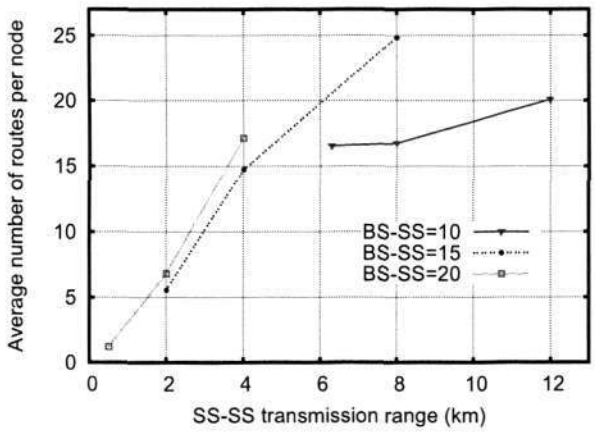
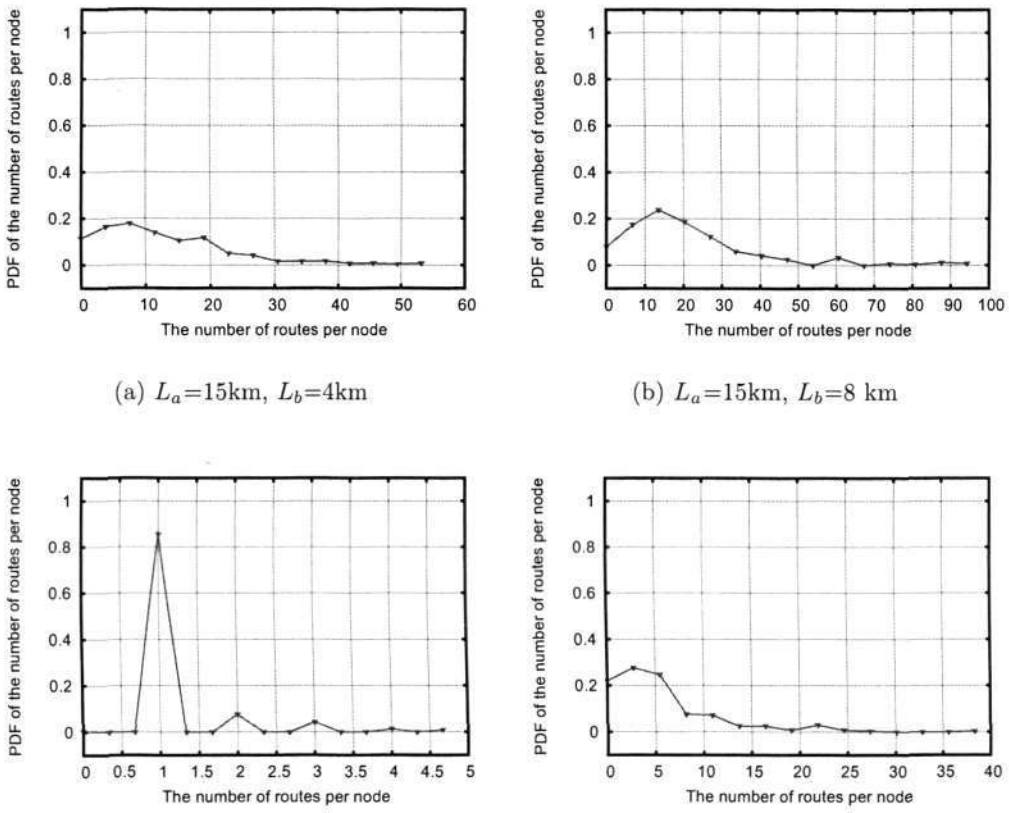


Figure 7.6: Average number of node-disjoint routes between an SS and the BS.

BS-SS transmission range being large, i.e. 20 km, even if the SS-SS transmission range is short, redundant paths might exist for an SS. In an extreme case when BS-SS and SS-SS transmission ranges are 20 km and 0.5 km, respectively, 80% of the SSs have one path to the BS, as shown in Figure 7.7(c). This is because when the SS-SS transmission range is as short as 0.5 km, there are rare links between two SSs, and hence an SS is hard to relay packets for other SSs. As a result, the path between the SS and the BS is the direct link between them. This is further illustrated in Figure 7.9(c) that 80% of the SSs have one-hop path to the BS. In another case that the SS-SS transmission range is slightly increased as shown in Figure 7.7(d), when BS-SS and SS-SS transmission ranges are 20 km and 2 km, respectively, the majority of the SSs have three to eight node-disjoint paths to the BS, which could lead to good end-to-end reliability for an SS. This result further proofs that increasing the SS-SS range helps to improve route redundancy.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) L_a=15km, L_b=4km

(b) L_a=15km, L_b=8 km

(c) L_a=20 km, L_b=0.5 km

(d) L_a=20 km, L_b=2 km

Figure 7.7: Probability density function of the number of node-disjoint paths (L_a and L_b are BS-SS and SS-SS transmission ranges, respectively).

B.4.2 Average Path Length Analysis

We analyze the average path length, i.e. the average number of hops per route, which can be a guideline for network deployment. The results are shown in Figure 7.8. It can be seen that with a shorter BS-SS transmission range configuration, the routes have longer average length, indicating that an SS requires more nodes to relay the traffic to BS. With a relatively shorter BS-SS transmission range, i.e. 10 km, increasing SS-SS transmission range does not result in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

much change to the average path length. That implies the BS-SS transmission range is an important factor in the route statistics. Figure 7.8 further shows that if the BS-SS transmission range is relatively large, for example, 15 and 20 km, increasing the SS-SS transmission range will increase the average path length because more new routes with more hops are included in this case.

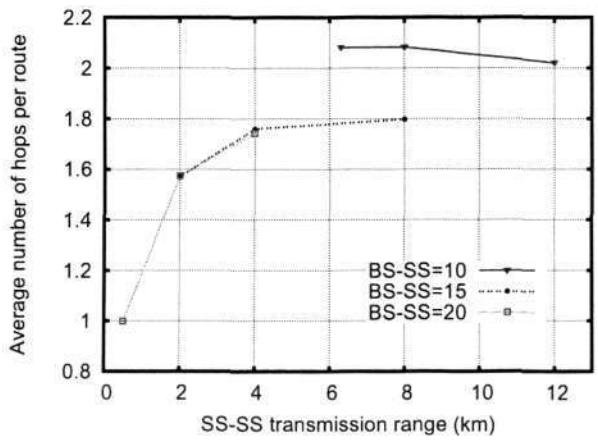
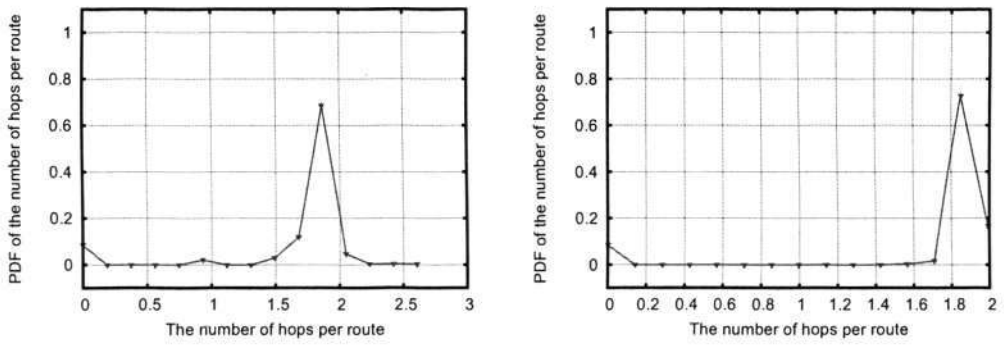


Figure 7.8: Average number of hops per route.

Figure 7.8 shows the average path length in the network, whereas Figure 7.9 shows the probabilistic distribution of the average path length per SS. From the figure, when BS-SS transmission range is 15 km, the majority of the average path length per SS is around 1.8, and this pattern becomes sharper around 1.8 with the increase in SS-SS transmission range. When BS-SS and SS-SS transmission ranges are 20 km and 0.5 km, respectively, approximate 80% of the average path length per SS is one. This result is intuitive because most of the SSs are able to connect to the BS directly and there are only a few links between SSs due to short SS-SS transmission range. In this case, increasing SS-SS transmission

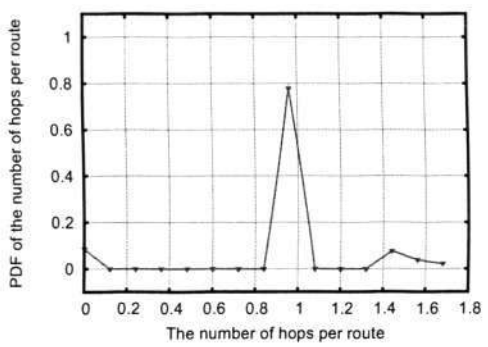
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

range results in more two-hop routes between an SS and the BS, as shown in Figure 7.9(d).

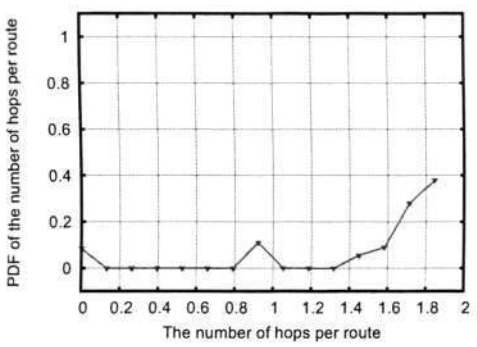


(a) L_a=15 km, L_b=4 km

(b) L_a=15 km, L_b=8 km



(c) L_a=20 km, L_b=0.5 km



(d) L_a=20 km, L_b=2 km

Figure 7.9: Probability density function of the average number of hops per route (L_a and L_b are BS-SS and SS-SS transmission ranges, respectively).

From the connectivity and routing analysis, it can be concluded that higher transmission range results in better connectivity and higher route redundancy. With some configurations, there are always multiple candidate routes between an SS and the BS, and the average path length of the candidate routes is 1.8. This calls for possible use of multi-path routing protocols, and the short path length implies a potentially acceptable end-to-end delay for real-time traffic. The results

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

further show that BS-SS transmission range has a more important effect on the connectivity and route redundancy, compared to SS-SS transmission range in the maritime communication environment. When BS-SS transmission range is short, increasing SS-SS transmission range only marginally improves route redundancy.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Author's Publications

- Yu Ge, Su Wen, Yew-Hock Ang, and Ying-Chang Liang, "Optimal Relay Selection in IEEE 802.16j Multihop Relay Vehicular Networks", accepted by IEEE Transactions on Vehicular Technology.

- Yu Ge, Chen-Khong Tham, Peng-Yong Kong, and Yew-Hock Ang, "Dynamic End-to-end Capacity in IEEE 802.16 Wireless Mesh Networks", accepted by Computer Networks Journal (Elsevier).

- Yu Ge, Su Wen, and Yew-Hock Ang, "Analysis of Optimal Relay Selection in IEEE 802.16 Multihop Relay Networks", in Proceedings of IEEE WCNC 2009, April 2009.

- Peng-Yong Kong, Jaya Shankar Pathmasuntharam, Haiguang Wang, Yu Ge, Chee-Wei Ang, Wen Su, Ming-Tuo Zhou, and Hiroshi Harada, "A Routing Protocol for WiMAX Based Maritime Wireless Mesh Networks", in Proceedings of IEEE VTC2009-Spring, April, 2009.

- Jaya Shankar Pathmasuntharam, Peng-Yong Kong, Yu Ge, Haiguang Wang, Chee-Wei Ang, Wen Su, and Hiroshi Harada, "TRITON: High Speed Mar-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

itime Mesh Networks", in Proceedings of SocialNets'2008 (PIMRC workshop), September 2008.

- Yu Ge, Chen-Khong Tham, Peng-Yong Kong, and Yew-Hock Ang, "Capacity Estimation for IEEE 802.16 Wireless Multi-Hop Mesh Networks", in Proceedings of IEEE WCNC 2008, March/April 2008.

- Peng-Yong Kong, Haiguang Wang, Yu Ge, Chee-Wei Ang, Wen Su, Jaya Shankar Pathmasuntharam, Mingtuo Zhou, and Hoang Vinh Dien, "A Performance Comparison of Routing Protocols for Maritime Wireless Mesh Networks", in Proceedings of IEEE WCNC 2008, March/April 2008.

- Yu Ge, Peng-Yong Kong, Chen-Khong Tham, and Jaya Shankar, "Connectivity and Route Analysis for a Maritime Communication Network", in Proceedings of IEEE ICICS 2007, December 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Bibliography

[1] *The IEEE 802.16 Working Group on Broadband Wireless Access Standards webpage, http://www.ieee802.org/16/.*

[2] *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE 802.16-2004. October 2004.*

[3] J. Robinson and E. Knightly, "A Performance Study of Deployment Factors in Wireless Mesh Networks," in *Proceedings of IEEE INFOCOM 2007*, May 2007.

[4] I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks Journal (Elsevier)*, vol. 47, pp. 445–487, March 2005.

[5] *Draft Amendment to IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Multihop Relay Specification. IEEE P802.16j/D5. May 2008.*

[6] *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, IEEE 802.16e-2005. February 2006.*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[7] C. Lin and J. Liu, "QoS routing in ad hoc wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1426 – 1438, August 1999.

[8] L. Chen and W. Heinzelman, "QoS-aware routing based on bandwidth estimation for mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 561 – 572, March 2005.

[9] S. Jamin, P. Danzig, S. Shenker, and L. Zhang, "A measurement-based admission control algorithm for integrated service packet networks," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 56 – 70 , February 1997.

[10] C. Lin, "Admission control in time-slotted multihop mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1974 – 1983, October 2001.

[11] Y. Yang and R. Kravets, "Contention-aware admission control for ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 4, pp. 363 – 377, July-August 2005.

[12] K. Choi, D. Jeong, and W. Jeon, "Packet Scheduler for Mobile Communications Systems with Time-Varying Capacity Region," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1034 – 1045, March 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[13] Q. Li and R. Negi, "Prioritized Maximal Scheduling in Wireless Networks," in *Proceedings of IEEE GLOBECOM 2008*, pp. 1 – 5, November/December 2008.

[14] T. Dreibholz, X. Zhou, and E. Rathgeb, "A Performance Evaluation of RSerPool Server Selection Policies in Varying Heterogeneous Capacity Scenarios," in *Proceedings of IEEE 33rd EUROMICRO Conference on Software Engineering and Advanced Applications, 2007. SEAA 2007*, pp. 157 – 166, August 2007.

[15] R. Carter and M. Crovella, "Server selection using dynamic path characterization in wide-area networks," in *Proceedings of IEEE INFOCOM 1997*, vol. 3, pp. 1014 – 1021, April 1997.

[16] S. Penz, "SLP-based service management with QoS server selection in mobile ad-hoc networks," in *Proceedings of IEEE WoWMoM 2008*, pp. 1 – 3, June 2008.

[17] A. Chobanyan, M. Mutka, V. Mandrekar, and N. Xi, "End-to-end available bandwidth as a random autocorrelated QoS-relevant time-series," *Computer Networks Journal (Elsevier)*, vol. 52, pp. 1220 – 1237, April 2008.

[18] M. Jain and C. Dovrolis, "Path selection using available bandwidth estimation in overlay-based video streaming," *Computer Networks Journal (Elsevier)*, vol. 52, pp. 2411 – 2418, August 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[19] *Harmonized Contribution on 802.16j (Mobile Multihop Relay) Usage Models, http://wirelessman.org/relay/docs/80216j-06_015.pdf.*

[20] T. Mak, K. Laberteaux, and R. Sengupta, "A multi-channel VANET providing concurrent safety and commercial services," in *Proceedings of ACM VANET 2005*, pp. 1 – 9, September 2005.

[21] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, and R. Herrtwich, "Design of 5.9 ghz dsrc-based vehicular safety communication," *IEEE Wireless Communications*, vol. 13, pp. 36 – 43, October 2006.

[22] K. Yang, S. Ou, H. Chen, and J. He, "A Multihop Peer-Communication Protocol With Fairness Guarantee for IEEE 802.16-Based Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 3358 – 3370, November 2007.

[23] J. Zhu and S. Roy, "MAC for dedicated short range communications in intelligent transport system," *IEEE Communications Magazine*, vol. 41, pp. 60 – 67, December 2003.

[24] A. So and B. Liang, "Enhancing WLAN Capacity by Strategic Placement of Tetherless Relay Points," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 474 – 487, May 2007.

[25] Y. Ge, C. Tham, P. Kong, and Y. Ang, "Capacity Estimation for IEEE 802.16 Wireless Multi-Hop Mesh Networks," in *Proceedings of IEEE WCNC 2008*, pp. 2651 – 2656, March/April 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[26] Y. Ge, C. Tham, P. Kong, and Y. Ang, "Dynamic End-to-end Capacity in IEEE 802.16 Wireless Mesh Networks," *Accepted by Computer Networks Journal (Elsevier).*

[27] Y. Ge, W. Su, and Y. Ang, "Analysis of Optimal Relay Selection in IEEE 802.16 Multihop Relay Networks," in *Proceedings of IEEE WCNC 2009*, April 2009.

[28] Y. Ge, W. Su, Y. Ang, and Y. Liang, "Optimal Relay Selection in IEEE 802.16j Multihop Relay Vehicular Networks," *Accepted by IEEE Transactions on Vehicular Technology.*

[29] G. Sharma, R. Mazumdar, and N. Shroff, "On the complexity of scheduling in wireless networks," in *Proceedings of ACM MobiCom 2006*, pp. 227 – 238, September 2006.

[30] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388 – 404, March 2000.

[31] A. KeshavarzHaddad and R. Riedi, "On the Broadcast Capacity of Multihop Wireless Networks: Interplay of Power, Density and Interference," in *Proceedings of IEEE SECON 2007*, pp. 314 – 323, June 2007.

[32] S. Ramanathan, "A unified framework and algorithm for (T/F/C)DMA channel assignment in wireless networks," in *Proceedings of IEEE INFOCOM 1997*, vol. 2, pp. 900 – 907, April 1997.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[33] J. Gronkvist and A. Hansson, "Comparison Between Graph-Based and Interference-Based STDMA Scheduling," in *Proceedings of ACM MobiHoc 2001*, pp. 255 – 258, October 2001.

[34] A. Behzad and I. Rubin, "On the Performance of Graph-Based Scheduling Algorithms for Packet Radio Networks," in *Proceedings of IEEE GLOBECOM 2003*, pp. 3432 – 3436, December 2003.

[35] Y. Wang, W. Wang, X. Li, and W. Song, "Interference-Aware Joint Routing and TDMA Link Scheduling for Static Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, pp. 1709 – 1726, December 2008.

[36] P. Rickenbach, R. Wattenhofer, and A. Zollinger, "Algorithmic models of interference in wireless ad hoc and sensor networks," *IEEE/ACM Transactions on Networking*, vol. 17, pp. 172 – 185, February 2009.

[37] T. Moscibroda, R. Wattenhofer, and A. Zollinger, "Topology control meets SINR: the scheduling complexity of arbitrary topologies," in *Proceedings of ACM MobiHoc 2006*, pp. 310 – 321, May 2006.

[38] H. Lim, C. Lim, and J. Hou, "A coordinate-based approach for exploiting temporal-spatial diversity in wireless mesh networks," in *Proceedings of ACM MobiCom 2006*, pp. 14 – 25, September 2006.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[39] A. Agarwal and P. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 71 – 81, July 2004.

[40] M. Alicherry, R. Bhatia, and L. Li, "Joint Channel Assignment and Routing for Throughput Optimization in Multi-Radio Wireless Mesh Networks," in *Proceedings of ACM MobiCom 2005*, pp. 58 – 72, August 2005.

[41] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proceedings of ACM MobiCom 2003*, pp. 66 – 80, September 2003.

[42] X. Yang and N. Vaidya, "Priority scheduling in wireless ad hoc networks," in *Proceedings of ACM MobiHoc 2002*, pp. 71 – 79, June 2002.

[43] X. Yang and N. Vaidya, "Priority scheduling in wireless ad hoc networks," *Kluwer Wireless Networks*, vol. 12, pp. 273 – 286, May 2006.

[44] P. Stuedi and G. Alonso, "Modeling and computing throughput capacity of wireless multihop networks," *Computer Networks Journal (Elsevier)*, vol. 52, pp. 116 – 129, January 2008.

[45] Y. Wu, P. Chou, Q. Zhang, K. Jain, W. Zhu, and S. Kung, "Network planning in wireless ad hoc networks: a cross-Layer approach," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 136 – 150, January 2005.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[46] J. Camp, J. Robinson, C. Steger, and E. Knightly, "Measurement driven deployment of a two-tier urban mesh access network," in *Proceedings of ACM MobiSys'06*, pp. 96 – 109, June 2006.

[47] O. Dousse and P. Thiran, "Connectivity vs capacity in dense ad hoc networks," in *Proceedings of IEEE INFOCOM 2004*, vol. 1, pp. 476–486, March 2004.

[48] F. Xue, L. Xie, and P. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 834 – 847, March 2005.

[49] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the Gap in the Capacity of Wireless Networks Via Percolation Theory," *IEEE Transactions on Information Theory*, vol. 53, pp. 1009 – 1018, March 2007.

[50] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proceedings of ACM MobiCom 2005*, pp. 73 – 87, August 2005.

[51] D. Chafekar, V. Kumar, M. Marathe, S. Parthasarathy, and A. Srinivasan, "Approximation Algorithms for Computing Capacity of Wireless Networks with SINR Constraints," in *Proceedings of IEEE INFOCOM 2008*, pp. 1166 – 1174, April 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[52] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem," in *Proceedings of ACM MobiCom 2003*, pp. 42 – 54, September 2003.

[53] V. A. Kumar, M. Marathe, S. Parthasarathy, and A. Srinivasan, "Algorithmic aspects of capacity in wireless networks," in *Proceedings of ACM SIGMETRICS 2005*, pp. 133 – 144, June 2005.

[54] Y. Sun, X. Gao, E. Belding-Royer, and J. Kempf, "Model-based resource prediction for multi-hop wireless networks," in *Proceedings of IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, pp. 114 – 123, October 2004.

[55] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535 – 547, March 2000.

[56] S. Toumpis and A. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 736 – 748, July 2003.

[57] P. Zhou, X. Wang, and R. Rao, "Asymptotic Capacity of Infrastructure Wireless Mesh Networks," *IEEE Transactions on Mobile Computing*, vol. 7, pp. 1011 – 1024, August 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[58] J. Gomez and A. Campbell, "A case for variable-range transmission power control in wireless multihop networks," in *Proceedings of IEEE INFOCOM 2004*, vol. 2, pp. 1425 – 1436, March 2004.

[59] A. Jovicic, P. Viswanath, and S. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 50, pp. 2555 – 2565, November 2004.

[60] L. Xie and P. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, pp. 2313 – 2328, June 2006.

[61] P. Kyasanur and N. Vaidya, "Capacity of multichannel wireless networks under the protocol model," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, pp. 515 – 527, April 2009.

[62] W. Tam and Y. Tseng, "Joint Multi-Channel Link Layer and Multi-Path Routing Design for Wireless Mesh Networks," in *Proceedings of IEEE INFOCOM 2007*, pp. 2081 – 2089, May 2007.

[63] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proceedings of IEEE INFOCOM 2001*, vol. 3, April 2001.

[64] C. Lea and M. Zhang, "On the mobility/capacity conversion in wireless networks," in *Proceedings of IEEE Wireless Communications and Networking, 2003. WCNC 2003.*, vol. 3, pp. 1433 – 1438, March 2003.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[65] X. Lin, G. Sharma, R. Mazumdar, and N. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with Brownian mobility," *IEEE Transactions on Information Theory*, vol. 52, pp. 2777 – 2784, June 2006.

[66] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and Capacity Trade-Offs in Mobile Ad Hoc Networks: A Global Perspective," in *Proceedings of IEEE INFOCOM 2006*, pp. 1 – 12, April 2006.

[67] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and Capacity Trade-Offs in Mobile Ad Hoc Networks: A Global Perspective," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 981 – 992, October 2007.

[68] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *Proceedings of IEEE INFOCOM 2003*, vol. 2, pp. 1553 – 1563, March/April 2003.

[69] J. Li, C. Blake, D. D. Couto, H. Lee, and R. Morris, "Capacity of Ad Hoc Wireless Networks," in *Proceedings of ACM Mobicom 2001*, July 2001.

[70] Y. Dong, D. Makrakis, and T. Sullivan, "Effective admission control in multihop mobile ad hoc networks," in *Proceedings of ICCT 2003*, vol. 2, pp. 1291 – 1294, April 2003.

[71] Y. Xiao and H. Li, "Local data control and admission control for QoS support in wireless ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 1558 – 1572, September 2004.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[72] S. Shin and H. Schulzrinne, "Experimental Measurement of the Capacity for VoIP Traffic in IEEE 802.11 WLANs," in *Proceedings of IEEE INFOCOM 2007*, pp. 2018 – 2026, May 2007.

[73] Y. Cho, S. Karande, K. Misra, H. Radha, J. Yoo, and J. Hong, "On Channel Capacity Estimation and Prediction for Rate-Adaptive Wireless Video," *IEEE Transactions on Multimedia*, vol. 10, pp. 1419 – 1426, November 2008.

[74] O. Hillestad, A. Perkis, V. Genc, S. Murphy, and J. Murphy, "Delivery of on-demand video services in rural areas via IEEE 802.16 broadband wireless access networks," in *Proceedings of ACM WMuNeP'06*, pp. 43 – 52, October 2006.

[75] S. Harsha, A. Kumar, and V. Sharma, "An Analytical Model for the Capacity Estimation of Combined VoIP and TCP File Transfers over EDCA in an IEEE 802.11e WLAN," in *Proceedings of IEEE IWQoS 2006*, pp. 178 – 187, June 2006.

[76] M. Kazantzidis, M. Gerla, and S. Lee, "Permissible throughput network feedback for adaptive multimedia in AODV MANETs," in *Proceedings of IEEE ICC 2001*, vol. 5, pp. 1352 – 1356, June 2001.

[77] G. Ahn, A. Campbell, A. Veres, and L. Sun, "Supporting service differentiation for real-time and best-effort traffic in stateless wireless ad hoc networks

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(SWAN)," *IEEE Transactions on Mobile Computing*, vol. 1, pp. 192 – 207, July/September 2002.

[78] K. Chua, H. Xiao, and K. Seah, "Relative service differentiation for mobile ad hoc networks," in *Proceedings of IEEE WCNC 2003*, vol. 2, pp. 1379 – 1384, March 2003.

[79] D. Clark and W. Fang, "Explicit allocation of best-effort packet delivery service," *IEEE/ACM Transactions on Networking*, vol. 6, pp. 362 – 373, August 1998.

[80] S. Redana and M. Lott, "Performance Analysis of IEEE 802.16a in Mesh Operation Mode," in *Proceedings of the 13th IST Mobile & Wireless Communications Summit*, June 2004.

[81] C. Hoymann, "Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16," *Computer Networks Journal (Elsevier)*, vol. 49, pp. 341 – 363, October 2005.

[82] R. Hincapie, J. Sierra, and R. Bustamante, "Remote locations coverage analysis with wireless mesh networks based on IEEE 802.16 Standard," *IEEE Communications Magazine*, vol. 45, pp. 120 – 127, January 2007.

[83] D. Niyato and E. Hossain, "A game-theoretic approach to bandwidth allocation and admission control for polling services in IEEE 802.16 broadband wireless networks," in *Proceedings of ACM QShine 2006*, August 2006.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[84] S.-E. Elayoubi, B. Fourestie, and X. Auffret, "On the capacity of OFDMA 802.16 systems," in *Proceedings of the IEEE ICC 2006*, vol. 4, pp. 1760 – 1765, June 2006.

[85] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 26–38, January 2007.

[86] M. Cao, W. Ma, Q. Zhang, and X. Wang, "Analysis of IEEE 802.16 Mesh Mode Scheduler Performance," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1455 – 1464, April 2007.

[87] M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Proceedings of ACM MobiHoc'05*, pp. 78 – 89, May 2005.

[88] L. Nuaymi and Z. Noun, "Simple Capacity Estimations in WIMAX/802.16 System," in *Proceedings of the IEEE PIMRC 2006*, pp. 1 – 5, September 2006.

[89] A. Belghith and L. Nuaymi, "WiMAX Capacity Estimations and Simulation Results," in *Proceedings of the IEEE VTC Spring 2008*, pp. 1741 – 1745, May 2008.

[90] R. Pries, D. Staehle, and D. Marsico, "IEEE 802.16 Capacity Enhancement Using an Adaptive TDD Split," in *Proceedings of the IEEE VTC Spring 2008*, pp. 1539 – 1543, May 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[91] C. Tarhini and T. Chahed, "On capacity of OFDMA-based IEEE802.16 WiMAX including Adaptive Modulation and Coding (AMC) and inter-cell interference," in *Proceedings of IEEE LANMAN 2007*, pp. 139 – 144, June 2007.

[92] C. Cicconetti, I. F. Akyildiz, and L. Lenzini, "Bandwidth Balancing in Multi-Channel IEEE 802.16 Wireless Mesh Networks," in *Proceedings of IEEE INFOCOM 2007*, pp. 2108 – 2116 , May 2007.

[93] S. Yeh, S. Talwar, S. Lee, and H. Kim, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," *IEEE Communications Magazine*, vol. 46, pp. 58 – 65, October 2008.

[94] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, pp. 59 – 67, September 2008.

[95] S. Ortiz, "The Wireless Industry Begins to Embrace Femtocells," *IEEE Computer*, vol. 41, pp. 14 – 17, July 2008.

[96] R. Baines, "The Need for WiMAX picocell & Femtocells," *WiMax London 2007*, pp. 1 – 36, April 2007.

[97] M. Ahmed, "Call admission control in wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 7, pp. 49 – 68 , First Quater 2005.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[98] Y. Hou, Y. Shi, H. Sherali, and S. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2579 – 2590, September 2005.

[99] Y. Xin, T. Guven, and M. Shayman, "Relay Deployment and Power Control for Lifetime Elongation in Sensor Networks," in *Proceedings of IEEE ICC'06*, pp. 3461 – 3466, June 2006.

[100] T. Himsoon, W. Siriwongpairat, Z. Han, and K. Liu, "Lifetime maximization via cooperative nodes and relay deployment in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 306 – 317, February 2007.

[101] Q. Wang, G. Takahara, H. Hassanein, and K. Xu, "On relay node placement and locally optimal traffic allocation in heterogeneous wireless sensor networks," in *Proceedings of IEEE LCN'05*, p. 8 pp., November 2005.

[102] W. Guo and X. Huang, "Optimal Relay Node Association for Two-Tiered Wireless Networks," in *Proceedings of IEEE GLOBECOM'07*, pp. 5185 – 5189, November 2007.

[103] Q. Wang, K. Xu, G. Takahara, and H. Hassanein, "Device Placement for Heterogeneous Wireless Sensor Networks: Minimum Cost with Lifetime Constraints," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 2444 – 2453, July 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[104] Y. Lin and Y. Hsu, "Multihop cellular: a new architecture for wireless communications," in *Proceedings of IEEE INFOCOM 2000*, vol. 3, pp. 1273 – 1282, March 2000.

[105] M. Hossain, A. Mammela, and H. Chowdhury, "Impact of Mobile Relays on Throughput and Delays in Multihop Cellular Network," in *Proceedings of IEEE the Fourth International Conference on Wireless and Mobile Communications, 2008. ICWMC'08*, pp. 304 – 308, July/August 2008.

[106] L. Long and E. Hossain, "Multihop Cellular Networks: Potential Gains, Research Challenges, and a Resource Allocation Framework," *IEEE Communications Magazine*, vol. 45, pp. 66 – 73, September 2007.

[107] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 2105 – 2115, October 2001.

[108] A. So and B. Liang, "Effect of relaying on capacity improvement in wireless local area networks," in *Proceedings of IEEE WCNC 2005*, vol. 3, pp. 1539 – 1544, March 2005.

[109] S. Mukherjee and H. Viswanathan, "Analysis of throughput gains from relays in cellular networks," in *Proceedings of IEEE GLOBECOM'05*, vol. 6, pp. 3471 – 3476, November/December 2005.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[110] H. Wei, S. Ganguly, R. Izmailov, and Z. Haas, "Interference-aware IEEE 802.16 WiMAX mesh networks," in *Proceedings of IEEE VTC 2005-Spring*, vol. 5, pp. 3102 – 3106, May 2005.

[111] J. Tao, F. Liu, Z. Zeng, and Z. Lin, "Throughput enhancement in WiMAX mesh networks using concurrent transmission," in *Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing, 2005*, vol. 2, pp. 871 – 874, September 2005.

[112] L. Fu, Z. Cao, and P. Fan, "Spatial reuse in IEEE 802.16 based wireless mesh networks," in *Proceedings of IEEE ISCIT 2005*, vol. 2, pp. 1358 – 1361, October 2005.

[113] S. Nahle, L. Iannone, B. Donnet, and N. Malouch, "On the construction of wimax mesh tree," *IEEE Communications Letters*, vol. 11, pp. 967 – 969, December 2007.

[114] W. Jiao, P. Jiang, R. Liu, and M. Li, "Centralized Scheduling Tree Construction Under Multi-Channel IEEE 802.16 Mesh Networks," in *Proceedings of IEEE GLOBECOM'07*, pp. 4764 – 4768, November 2007.

[115] H. Shetiya and V. Sharma, "Algorithms for routing and centralized scheduling to provide QoS in IEEE 802.16 mesh networks," in *Proceedings of ACM WMuNeP'05*, pp. 140 – 149, October 2005.

[116] D. Soldani and S. Dixit, "Wireless relays for broadband access," *IEEE Communications Magazine*, vol. 46, pp. 58 – 66, March 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[117] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, pp. 80 – 89, September 2004.

[118] V. Namboodiri, M. Agarwal, and L. Gao, "A study on the feasibility of mobile gateways for vehicular ad-hoc networks," in *Proceedings of ACM VANET 2004*, pp. 66 – 75, October 2004.

[119] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16J relay-based wireless access networks: an overview," *IEEE Wireless Communications*, vol. 15, pp. 56 – 63, October 2008.

[120] C. Lo, R. Heath, and S. Vishwanath, "The Impact of Channel Feedback on Opportunistic Relay Selection for Hybrid-ARQ in Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 1255 – 1268, March 2009.

[121] T. Theodoros and V. Kostantinos, "WiMax Network Planning and Systemys Performance Evaluation," in *Proceedings of IEEE WCNC 2007*, pp. 1948 – 1953, March 2007.

[122] B. Lin, P. Ho, L. Xie, and X. Shen, "Optimal relay station placement in IEEE 802.16j networks," in *Proceedings of ACM IWCMC 2007*, pp. 25–30, August 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[123] C. Lo, S. Vishwanath, and R. Heath, "Relay Subset Selection in Wireless Networks Using Partial Decode-and-Forward Transmission," in *Proceedings of IEEE VTC Spring 2008*, pp. 2395 – 2399, May 2008.

[124] Y. Yang, S. Murphy, and L. Murphy, "Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks," in *Proceedings of IEEE CCNC 2008*, pp. 922 – 926, January 2008.

[125] I. Fu, W. Sheen, and F. Ren, "Deployment and radio resource reuse in IEEE 802.16j multi-hop relay network in Manhattan-like environment," in *Proceedings of IEEE 6th International Conference on Information, Communications & Signal Processing, 2007. ICICS 2007*, pp. 1 – 5, December 2007.

[126] J. Li and S. Sampalli, "Cell Mobility Based Admission Control for Wireless Networks with Link Adaptation," in *Proceedings of IEEE ICC'07*, pp. 5862 – 5867, June 2007.

[127] S. Ann, K. Lee, and H. Kim, "A Path Selection Method in IEEE 802.16j Mobile Multi-hop Relay Networks," in *Proceedings of IEEE the Second International Conference on Sensor Technologies and Applications, 2008. SENSORCOMM'08*, pp. 808 – 812, August 2008.

[128] Y. Bian and A. Nix, "Mobile WiMAX: Multi-Cell Network Evaluation and Capacity Optimization," in *Proceedings of the IEEE VTC Spring 2008*, pp. 1276 – 1280, May 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[129] W. Joseph and L. Martens, "Performance evaluation of broadband fixed wireless system based on IEEE 802.16," in *Proceedings of IEEE WCNC 2006*, vol. 2, pp. 978 – 983, April 2006.

[130] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *IEEE Communications Magazine*, vol. 44, pp. 74 – 82, January 2006.

[131] H. Wu, R. Fujimoto, M. Hunter, and R. Guensler, "An architecture study of infrastructure-based vehicular networks," in *Proceedings of ACM MSWiM 2005*, pp. 36 – 39, October 2005.

[132] W. Chen, R. Guha, T. Kwon, J. Lee, and I. Hsu, "A survey and challenges in routing and data dissemination in vehicular ad-hoc networks," in *Proceedings of IEEE International Conference on Vehicular Electronics and Safety, 2008. ICVES 2008*, pp. 328 – 333, September 2008.

[133] M. Artimy, W. Phillips, and W. Robertson, "Connectivity with static transmission range in vehicular ad hoc networks," in *Proceedings of IEEE Communication Networks and Services Research Conference 2005*, pp. 237 – 242, May 2005.

[134] P. Santi and D. Blough, "The critical transmitting range for connectivity in sparse wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 2, pp. 25 – 39, January 2003.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[135] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in *Proceedings of ACM MobiHoc 2002*, pp. 80 – 91, June 2002.

[136] P. Wan and C. Yi, "Asymptotic critical transmission ranges for connectivity in wireless ad hoc networks with Bernoulli nodes," in *Proceedings of IEEE WCNC 2005*, vol. 4, pp. 2219 – 2224, March 2005.

[137] M. Fiore and J. Harri, "The networking shape of vehicular mobility," in *Proceedings of ACM MobiHoc 2008*, pp. 261 – 271, May 2008.

[138] W. Yuen, R. Yates, and C. Sung, "Effect of node mobility on highway mobile infostation networks," in *Proceedings of ACM MSWiM 2003*, September 2003.

[139] R. Hekmat and P. Mieghem, "Connectivity in wireless ad-hoc networks with a log-normal radio model," *Kluwer Mobile Networks and Applications 2006*, vol. 11, pp. 351 – 360, June 2006.

[140] T. Taleb, M. Ochi, A. Jamalipour, N. Kato, and Y. Nemoto, "An efficient vehicle-heading based routing protocol for VANET networks," in *Proceedings of IEEE WCNC 2006*, vol. 4, pp. 2199 – 2204, April 2006.

[141] J. Chennikara-Varghese, W. Chen, T. Hikita, and R. Onishi, "Local Peer Groups and Vehicle-to-Infrastructure Communications," in *Proceedings of IEEE Globecom Workshop: 2nd IEEE Automotive Networking and Applications (AutoNet), 2007*, pp. 1 – 6, November 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[142] W. Chen and S. Cai, "Ad hoc peer-to-peer network architecture for vehicle safety communications," *IEEE Communications Magazine*, vol. 43, pp. 100 – 107, April 2005.

[143] C. Hung, H. Chan, and E. Wu, "Mobility Pattern Aware Routing for Heterogeneous Vehicular Networks," in *Proceedings of IEEE WCNC 2008*, pp. 2200 – 2205, March/April 2008.

[144] Y. Ling, W. Chen, T. Hsing, and O. Altintas, "Preserving HTTP Sessions in Vehicular Environments," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 2697 – 2712, September 2007.

[145] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance evaluation of the mesh election procedure of IEEE 802.16/WiMAX," in *Proceedings of ACM MSWiM'07*, pp. 323 – 327, October 2007.

[146] S. Vural and E. Ekici, "Analysis of hop-distance relationship in spatially random sensor networks," in *Proceedings of ACM MobiHoc'05*, pp. 320 – 331, May 2005.

[147] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing, IETF RFC 3561," July 2003.

[148] *The Qualnet simulator webpage, http://www.scalable-networks.com/.*

[149] T. Taleb, E. Sakhaee, A. Jamalipour, K. Hashimoto, N. Kato, and Y. Nemoto, "A Stable Routing Protocol to Support ITS Services in VANET

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Networks," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 3337 – 3347, November 2007.

[150] M. Alouini and A. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels," *Kluwer Wireless Personal Communications: An International Journal*, vol. 13, pp. 119–143, May 2000.

[151] Y. Kim and H. Liu, "Infrastructure Relay Transmission With Cooperative MIMO," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 2180 – 2188, July 2008.

[152] D. Cox and H. Lee, "Physical relationships," *IEEE Microwave Magazine*, vol. 9, pp. 89 – 94, August 2008.

[153] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1205 – 1211, July 1999.

[154] Y. Ge, P. Kong, C. Tham, and J. Pathmasuntharam, "Connectivity and Route Analysis for a Maritime Communication Network," in *Proceedings of IEEE ICICS 2007*, pp. 1 – 5, December 2007.

[155] P. Kong, H. Wang, Y. Ge, C. Ang, W. Su, J. Pathmasuntharam, M. Zhou, and H. Dien, "A Performance Comparison of Routing Protocols for Maritime Wireless Mesh Networks," in *Proceedings of IEEE WCNC2008*, pp. 1 – 5, April 2008.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[156] P. Kong, J. Pathmasuntharam, H. Wang, Y. Ge, C. Ang, W. Su, M. Zhou, and H. Harada, "A Routing Protocol for WiMAX Based Maritime Wireless Mesh Networks," in *Proceedings of IEEE VTC2009-Spring,* , , pp. 1 – 5, April 2009.

[157] J. Pathmasuntharam, P. Kong, Y. Ge, H. Wang, C. Ang, W. Su, and H. Harada, "TRITON: High Speed Maritime Mesh Networks," in *Proceedings of IEEE PIMRC 2008*, pp. 1 – 5, September 2008.

[158] M. Bazaraa, H. Sherali, and C. Shetty, "Non-linear Programming: Theory and Algorithms," *WileyBlackwell; 3rd Edition edition*, 2006.

[159] *Automatic Identification System webpage, http://www.navcen.uscg.gov/enav/AIS/default.htm.*