A SOLUTION TO OPTIMAL AND FAIR RATE ADAPTATION IN WIRELESS MESH NETWORKS

by

Pieter Albertus Jansen van Vuuren

Submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Electronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology
UNIVERSITY OF PRETORIA

August 2013

"Let your light so shine before men, that they may see your works, and give glory to your Father who is in heaven." - Matthew 5:16.

ACKNOWLEDGEMENTS

I would like to extend my sincerest gratitude to the following people and organisation for their assistance:

- My supervisor and mentor Prof B.T. Maharaj, for his invaluable guidance and advice.
- My co-supervisor and mentor Prof A.S. Alfa from University of Manitoba, Canada, for his valuable advice and assistance.
- The Broadband Wireless Multimedia Communications (BWMC) group at the University of Pretoria for their financial support during my studies.
- Marthinus Jansen van Vuuren, my father, Petronella Jansen van Vuuren, my mother and Tinus Jansen van Vuuren, my brother, for their selfless love, prayers, encouragement and support.
- Dr. Juanita Nel, for all her advice and support during the course of my Masters research.
- My colleagues and friends, Thinus Prinsloo, Simon Barnes, Philip Botha, Nikolai de Figueiredo and Anika Muller with regards to various aspects of my Masters research.
- Mrs Iauma Cooper, for critically and professionally editing the manuscript.

SUMMARY

A SOLUTION TO OPTIMAL AND FAIR RATE ADAPTATION IN WIRELESS MESH NETWORKS

by

Pieter Albertus Jansen van Vuuren

Supervisor(s):	Prof. B.T. Maharaj
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Master of Engineering (Electronic Engineering)
Keywords:	Fairness, game theory, mathematical programming, rate adaptation, wireless mesh networks

Current wireless networks still employ techniques originally designed for their fixed wired counterparts. These techniques make assumptions (such as a fixed topology, a static enviroment and non-mobile nodes) that are no longer valid in the wireless communication environment. Furthermore, the techniques and protocols used in wireless networks should take the number of users of a network into consideration, since the channel is a shared and limited resource. This study deals with finding an optimal solution to resource allocation in wireless mesh networks. These networks require a solution to fair and optimal resource allocation that is decentralised and self-configuring, as users in such networks do not submit to a central authority.

The solution presented is comprised of two sections. The first section finds the optimal rate allocation, by making use of a heuristic. The heuristic was developed by means of a non-linear mixed integer mathematical formulation. This heuristic finds a feasible rate region that conforms to the set of constraints set forth by the wireless communication channel. The second section finds a fair allocation of rates among all

the users in the network. This section is based on a game theory framework, used for modelling the interaction observed between the users. The fairness model is defined in strategic form as a repeated game with an infinite horizon.

The rate adaptation heuristic and fairness model employs a novel and effective information distribution technique. The technique makes use of the optimized link state routing protocol for information distribution, which reduces the overhead induced by utilising multi-point relays. In addition, a novel technique for enforcing cooperation between users in a network is presented. This technique is based on the Folk theorem and ensures cooperation by threat of punishment. The punishment, in turn, is executed in the form of banishment from the network.

The study describes the performance of the rate adaptation heuristic and fairness model when subject to fixed and randomised topologies. The fixed topologies were designed to control the amount of interference that a user would experience. Although these fixed topologies might not seem to reflect a real-world scenario, they provide a reasonable framework for comparison. The randomised network topology is introduced to more accurately represent a real-world scenario. Furthermore, the randomised network topologies consist of a significant number of users, illustrating the scalability of the solution. Both data and voice traffic have been applied to the rate adaptation heuristic and fairness model.

It is shown that the heuristic effectively reduces the packet loss ratio which drops below 5% after about 15 seconds for all fixed topologies. Furthermore, it is shown that the solution is near-optimal in terms of data rate and that a fair allocation of data rates among all nodes is achieved. When considering voice traffic, an increase of 10% in terms of data rate is observed compared to data traffic. The heuristic is successfully applied to large networks, demonstrating the scalability of the implementation.

OPSOMMING

'N OPLOSSING VIR OPTIMALE EN BILLIKE TEMPO-AANPASSING IN DRAADLOSE MAASNETWERKE

deur

Pieter Albertus Jansen van Vuuren

Studieleier(s):	Prof. B.T. Maharaj
Departement:	Elektriese, Elektroniese en Rekenaar-Ingenieurswese
Universiteit:	Universiteit van Pretoria
Graad:	Magister in Ingenieurswese (Elektroniese Ingenieurswese)
Sleutelwoorde:	Draadlose maasnetwerke, regverdigheid, spelteorie, tempo-aanpassing, wiskundige programmering

Draadlose netwerke benut tans steeds tegnieke wat oorspronklik ontwerp is vir hul vaste eweknieë. Hierdie tegnieke maak aannames wat nie meer geldig is in die draadlose kommunikasie-omgewing nie. Verder moet die tegnieke en protokolle wat in draadlose netwerke gebruik word die aantal gebruikers van 'n netwerk in ag neem, aangesien die kanaal 'n verdeelde en beperkte hulpbron is. Hierdie studie ondersoek 'n optimale oplossing vir die toekenning van hulpbronne in draadlose maasnetwerke. Hierdie netwerke vereis 'n oplossing vir billike en optimale toekenning van hulpbronne wat gedesentraliseerd en self-ingestel is, weens die feit dat die gebruikers in sulke netwerke nie 'n sentrale gesag gehoorsaam nie.

Die oplossing bestaan uit twee afdelings. Die eerste afdeling vind die optimale tempo-toekenning deur gebruik te maak van 'n heuristiek. Die heuristiek is ontwikkel deur gebruik te maak van 'n nie-linieêre gemengde heelgetal- wiskundige program. Die heuristiek vind 'n haalbare tempogebied wat voldoen aan die stel van beperkings, soos dit uiteengesit word deur die draadlose kommunikasiekanaal. Die tweede afdeling

vind die regverdige toekenning van datatempo vir al die gebruikers in die netwerk. Hierdie gedeelte is gebaseer op 'n spel-teoretiese raamwerk, wat gebruik word vir die modellering van die interaksie wat waargeneem word tussen die gebruikers. Die regverdigheidsmodel is gedefinieer in strategiese vorm as 'n herhaalde spel met 'n oneindige horison.

Die tempo-aanpassingheuristiek en regverdigheidsmodel maak gebruik van 'n nuwe en doeltreffende tegniek vir die verspreiding van inligting. Die tegniek maak gebruik van die optimale skakel-toestand-roeteprotokol vir die verspreiding van inligting, wat die oorhoofse inligting verminder deur gebruik te maak van verskeie puntaflosse. Verder is 'n nuwe tegniek vir die handhawing van samewerking tussen die gebruikers in 'n netwerk aangebied. Hierdie tegniek is gebaseer op die Folk-stelling en verseker samewerking deur die dreigement van straf. Die straf behels ballingskap uit die netwerk.

Die studie beskryf die werksverrigting van die tempo-aanpassingheuristiek en regverdigheidsmodel wanneer onderhewig aan vaste en ewekansige topologieë. Die vaste topologieë is ontwerp om die hoeveelheid inmenging wat 'n gebruiker kan ervaar, te beheer. Alhoewel hierdie vaste topologieë nie 'n werklike omgewing kan beskryf nie, verskaf hulle 'n raamwerk vir vergelykingstudies. Die ewekansige netwerktopologie is ingestel om 'n meer akkurate verteenwoordiging van 'n werklike omgewing te beskryf. Verder bestaan die ewekansige netwerktopologieë uit 'n groot aantal gebruikers, sodat die skaalbaarheid van die oplossing geïllustreer kan word. Beide data- en stemverkeer is toegepas op die tempo-aanpassingheuristiek en regverdigheidsmodel.

Daar is aangetoon dat die heuristiek die pakkie-verliesverhouding binne 15 sekondes verminder na 5% vir alle vaste topologieë. Verder is bewys dat die oplossing byna optimaal is wat datatempo betref. Billike toekenning van datatempo vir alle gebruikers word behaal. 'n Toename van 10% in datatempo kan waargeneem word vir stemverkeer. Daar is ook bewys dat die heuristiek suksesvol toegepas kan word op groot netwerke.

LIST OF ABBREVIATIONS

AMR	Adaptive Multi Rate
AODV	Ad Hoc On Demand Distance Vector
ASIC	Application-Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
BCJR	Bahl Cocke Jelinek Raviv
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BW	Bandwidth
CDMA	Code Division Multiple Access
CRC	Cyclic Redundancy Check
CSMA-CA	Carrier Sense Multiple Access with Collision Avoidance
DCF	Distributed Coordination Function
DoS	Denial of Service
DSSS	Direct Sequence Spread Spectrum
EVRC	Enhanced Variable Rate Codec
FEC	Forward Error Correction
FPGA	Field Programmable Gate Array
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
ICC	International Conference on Communications
IGW	Internet Gateway
IMAP	Internet Message Access Protocol
ISI	Institute for Scientific Information
ISM	Industrial, Scientific and Medical
ITU	International Telecommunications Union
LDPC	Low Density Parity Check

LP	Linear Programming
MAC	Medium Access Control
MAN	Metropolitan Area Network
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MPDU	MAC Protocol Data Unit
MPR	Multi-Point Relay
NFJ	Null Frequency Jamming
NIC	Network Interface Cards
NLP	Non-Linear Programming
NP	Non Polynomial
NTP	Network Time Protocol
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	Orthogonal Frequency Division Multiple Access
OLSR	Optimised Link State Routing
OMNET	Open Modular Network Simulator
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
PAN	Personal Area Network
PLR	Packet Loss Ratio
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
REFOT	Relative Fairness and Optimized Throughput
RHF	Rate Heuristic and Fairness Model
RIP	Routing Information Protocol
RTP	Real-time Transport Protocol
SINR	Signal to Interference Noise Ratio

SNR	Signal to Noise Ratio
SPF	Shortest Path First
TCP	Transport Control Protocol
TDMA	Time Division Multiple Access
TFTP	Trivial File Transfer Protocol
UDP	User Datagram Protocol
VoIP	Voice-over Internet Protocol
WCP	Wireless Control Protocol
WCPcap	Wireless Control Protocol with Capacity Estimation
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

As the demand for wireless communication increases, so does the need for effective techniques and protocols used in these environments. Current wireless networks still employ techniques originally designed for their fixed wired counterparts [1]. These techniques make assumptions that are no longer valid in the wireless communication environment. Furthermore, the techniques and protocols used in wireless networks should take the amount of users of a network into account, seeing as the channel is a shared and limited resource [2]. Wireless mesh networks (WMNs) are a popular solution to the high demand for flexible wireless networks [3].

Although WMNs share similar characteristics with their wired counterparts, WMNs differ in their ad hoc nature. WMNs do not follow a fixed topology and are often dynamic, in the sense that nodes could be mobile [4]. These differences provide unique challenges, such as the requirement of self-configuring and self-organising nodes. Nodes in WMNs are often required to perform functions traditionally executed by routers, which leads to the notion of a multi-hop relay [4]. A multi-hop relay in a WMN can be defined as the forwarding of information from one node to another by making use of other nodes in the network. Nodes could have multiple wireless interfaces, all connecting to multiple devices [5].

Several methods have been developed to address these challenges. Of these, game theory is one of the most active methods [6]. The issue of developing self-configuring and self-organising nodes in an optimal manner can be achieved by developing models based on game theory. Game theory deals with the interaction of nodes in such a network, based on the assumption that they will act rationally [7]. This assumption is justified by considering that nodes act as they are programmed to act, which in turn, is based on the standards governing the development of these devices.

Resource allocation refers to the distribution of resources that are often limited among entities [8]. Distributing these resources optimally is an on going problem being researched in the field of operation research. Traditionally, such problems are dealt with by using mathematical programming techniques. Although these techniques lead to exact and optimal solutions, several real-world problems cannot be solved with these techniques. Such problems are referred to as non-polynomial time (NP)-hard problems, and do not have exact and optimal solutions. NP-hard problems are often addressed using heuristics. Heuristics are algorithms, which often rely on past information or events to effectively alter future events, and frequently result in sub-optimal solutions [9].

Chapter 2 discusses WMNs, game theory, resource allocation and the standards used in the wireless networking environment, in more detail.

1.2 MOTIVATION

Several factors affect the performance of wireless communications in a network environment, including the rate at which topology changes occur; number of nodes in the network; current channel conditions, and physical attributes of the communication device. Some of these factors are overlooked when designing algorithms for increasing performance [10]. The transmission control protocol (TCP) employs techniques that assume the capacity of the channel does not vary. However, the capacity depends on the signal to interference noise ratio (SINR) on that link. The SINR, in turn, depends on the transmission power of the communication device as well as the transmission power of all the devices surrounding that device. As the devices in the network alter their transmission power, the channel capacity varies, ultimately leading to a reduction in performance when employing traditional TCP techniques [11].

One prominent attribute of a WMN is that no central authority exists. This, in turn, leads to a challenging problem, namely the lack of coordination between the nodes [5]. This together with the fact that these nodes share a limited resource (the channel) results in nodes being non-cooperative. The performance of a WMN depends on how effective protocols can operate, and these protocols often depend on the cooperation of nodes in such a network. One such protocol is the routing protocol, which largely relies on the cooperation of nodes to forward messages in a multi-hop environment.

An effect regularly found in WMNs is the unfair allocation of resources across a network. When considering transmission power, nodes often attempt to increase their own performance by increasing their transmission power. This increase of transmission power leads to an increase in the amount of interference experienced by surrounding nodes [9]. This frequently leads to nodes being starved of communication in dense areas, whilst others experience high throughput. Thus the non-cooperative nature of nodes in a WMN leads to sub-optimal performance, as well as an unfair allocation of resources across the network.

As WMNs find more areas of application, the number of users in such a network is bound to increase. Metropolitan areas normally consist of a dense population and users in these areas typically demand reliable communication with a high throughput [3]. Although traditional protocols offer reliable communication between nodes in such a network, many times they are not scalable to larger and dense networks. The convergence time and amount of overhead required in such a network increases significantly when the number of nodes in the network increases. The increased convergence time leads to the degradation of reliability, especially when topology changes occur frequently [12]. The increased overhead would lead to a decrease in the total throughput in the network.

This situation motivated the researcher to investigate and find a solution to resource allocation, in terms of data rate that is both optimal and fair. An additional requirement would be to find a solution that is reliable, by reducing the probability of packet loss, and could be extended to large networks. These challenges require a solution that enforces cooperation, whilst maintaining an optimal operating point. Finding such a solution would enable the demand for wireless communication in larger networks to be met.

1.3 RESEARCH OBJECTIVE

The purpose of this study was to develop a method that would enable an optimal and fair allocation of resources across a WMN in a decentralised manner. In this study, the allocation of resources refers to data rate which is seen as a measure of throughput. Rate adaptation refers to the alteration of data rate, whilst a device is actively communicating with another device. Thus a mathematical model for rate adaptation is required, that takes the channel capacity, current transmission power, probability of loss and physical layer design parameters into account. This would ensure that the allocation of data rate is achieved by taking the factors previously omitted by traditional protocols into account.

To achieve its purpose, the study had several objectives. Allocating resources across a network in a fair manner would require knowledge of how nodes interact with each other. Decision theory traditionally deals with a single node that attempts to maximise some utility. Game theory differs from decision theory based on the number of nodes participating in the decision process, and can thus be used to model the interactions between nodes. However, game theory is merely a set of mathematical tools, and a framework tailored to the rate adaptation problem is required. Accordingly, finding and tailoring a form of game theory that applies to the rate adaptation problem, forms part of the objective of this study.

Since there is no central authority ensuring fair play among nodes, enforcing fairness in WMNs is required. This, in turn, would require a method of punishment, which should be enforced by the nodes themselves. This additional enforcement should be included in the game theory model, as it makes use of the interaction between the nodes in the network. The Folk theorem would provide a method for ensuring cooperation between nodes in a network, but, would need to be tailored and applied to the rate adaptation problem. Furthermore, an appropriate definition of fairness would also be required. There are several definitions of fairness and some are tailored to specific applications and depend on the preference of the users in their specific area. Consequently, the definition of fairness should be both feasible and accurate for the rate adaptation problem.

Most algorithms for resource allocation require information about the current state of the resources. In a network that makes use of a decentralised approach, this information should be distributed among the nodes. In this regard, the challenge is distributing this information with the lowest possible amount of overhead. The problem becomes more challenging when considering that the resource allocation method requires information about the current state of resource allocation, as well as the current state of fairness. Several routing protocols facilitate the ability to distribute information among nodes in a network. Consequently, another objective was finding a method that effectively makes use of existing protocols for the distribution of resource allocation information.

A reliable method for resource allocation in large and dense networks is required. As WMNs gains more interest in metropolitan areas, the number of users would increase. This would require a method that was both scalable and operated effectively. The reliability of a network is often measured in terms of the effectiveness to reduce the probability of loss. Due to the demand for high throughput in dense networks, the data rate of users in such a network should still be allocated optimally. The study thus aimed to overcome this challenge.

1.4 CONTRIBUTION

The study contributes to a novel and effective heuristic that finds the near-optimal allocation of data rates in a WMN. The heuristic finds a near-optimal operation point whilst remaining within the feasible region, resulting in high throughput whilst reducing the probability of packet loss. The study provides a mathematical framework based on mathematical programming that was used for developing the rate adaptation heuristic. This model follows a non-linear mixed integer programme. The mathematical programming model makes use of the objective function, decision variables and constraints as identified in the communication system. The framework captures the limitations and important properties of a communication system with regard to rate adaptation. Furthermore, the framework could be used to develop further mathematical solutions to the rate adaptation problem in WMNs.

The study presents a novel method for achieving fair data rate allocation. This method is referred to as the fairness model and makes use of game theory. A game theory model, identified as an infinite horizon repeated game, was developed. The infinite horizon repeated game was described in mathematical form by making use of the utility, actions and stage gain found in the rate adaptation problem when subject to a wireless network. This model captures the interaction between nodes in a network, and provides an effective method of decision making that each node could use to improve performance. Moreover, the game theory model provided a method for enforcing cooperation shown to lead to improved fairness. In this model, the punishment for non-cooperation, is being banished from the network for a specific period. The period is selected in such a way that the gain achieved by not cooperating does not outweigh the loss experienced in the banishment period.

The study provides a method for achieving effective information distribution. The optimised link state routing (OLSR) protocol enables information passing between the nodes in a wireless network. A novel method making use of the reserved fields in an OLSR frame was used to distribute power and rate information between the nodes in a WMN. By making use of the OLSR protocol, overhead reduction is achieved. This is due to the fact that the OLSR protocol makes use of multi-point relays (MPRs), which are seen as the centroids of the network. MPR nodes are the only nodes allowed to rebroadcast topology control information, which now also includes power and rate information. This method effectively reduces the required overhead whilst still achieving a fast convergence time.

The study introduces a platform for analysing the performance of rate adaptation algorithms. The platform was developed by making use of open modular network simulator (OMNET++) and contains three fixed topologies, as well as a randomised topology consisting of 50, 75 and 100 nodes. This platform can easily be extended to include any fixed topology consisting of numerous of nodes. The platform provides a method for increasing the amount of interfering links and performing comparative performance evaluations. The comparative performance evaluation itself can also be seen as a novel contribution. In addition, the scalability of almost any protocol can be tested by applying it to a large randomised network, provided that the simulation is repeated a significant number of times.

1.5 DISSERTATION OUTLINE

The study employed a top-down approach, where the top most level consists of the network topology and lowest level consists of physical layer techniques for rate adaptation. The study included all the relevant layers as defined by the open systems interconnection (OSI) layering model. Two different types of traffic and their related application layer protocols were illustrated. Details of the network layer, specifically the routing protocol has been discussed, as well as how these protocols could be used to aid rate adaptation. The physical layer has been presented with specific reference to rate adaptation and how game theory plays a role in this technique. Finally results have been presented, illustrating the effectiveness of the rate adaptation heuristic and game theory model.

Chapter 2 discusses the background information required for WMNs; game theory; resource allocation, and The IEEE communication standards. The advantages and challenges of WMNs are stated here, along with the intended applications for these types of networks. Game theory has been discussed in terms of the feasibility to wireless communication networks and the challenges encountered in such networks. Different types of techniques and the feasibility and advantages of these techniques are presented in the section entitled resource allocation. It is also shown that heuristics provide feasible solutions to NP-hard problems. Finally, the physical layer standards for communication devices are presented with specific reference to the standard most commonly used in WMNs; the IEEE 802.11 standard.

Chapter 3 presents the topologies used for evaluating networks and forms the top level in the top-down approach. Three fixed topologies and a randomised topology are presented. Fixed topologies follow a particular structure and provide significant advantages in terms of evaluating the performance of rate adaptation in an interfering environment. These fixed topologies do not reflect real world situations; the random topology on the other hand, better represents a real world environment. The random topology was presented in three forms, namely, a network consisting of 50 nodes, 75 nodes and 100 nodes all of which was randomly generated in a fixed area. The randomised topology provides advantages in terms of the performance of the rate adaptation technique when subject to large and dense networks.

Chapter 4 discusses and illustrates the layers used to model the communication device. The model consists of an application, presentation, session, transport, network, data link and physical layer. The application layer consists of two protocols, namely, file transfer protocol (FTP) and voice over Internet protocol (VoIP). These protocols handle different types of traffic; data traffic is normally used in conjunction with the FTP protocol, where VoIP is typically used for voice traffic. The next significant layer discussed is the network layer, responsible for the effective routing of information from source to destination. This layer makes use of the OLSR protocol which is explained in greater detail in Chapter 5. The physical layer in its most simplistic form consists of an encoder, modulator, channel, demodulator and decoder. The physical layer is illustrated in terms of the IEEE 802.11 standard.

Chapter 5 describes the routing protocol used in the network model, namely the OLSR protocol. The primary and auxiliary functions of the OLSR protocol are discussed with specific reference to the proactive nature of the protocol, as well as how the protocol creates routing tables and how link state information is distributed. The auxiliary functions, which are used to improve the performance of this protocol are described. Auxiliary functions of the OLSR protocol includes MPRs which are used to reduce the amount of overhead required for routing. The discussion covers how the OLSR protocol achieves shortest path routing, as well as how headers are structured. Finally, it is shown that the frame structure for this protocol contains fields that could be used by other algorithms (such as a rate adaptation heuristic) for passing messages between nodes in a WMN.

Chapter 6 defines the rate adaptation heuristic that takes place at the physical layer. The rate adaptation problem is defined in terms of the objective function, the decision variables and the constraints found in a wireless communication system. A formal mathematical problem formulation is derived and the problem is identified as a non-linear mixed integer programming problem. It is shown that an exact solution to the problem is not feasible and a heuristic is required. This heuristic requires messages passing between nodes, which is achieved by making use of the OLSR protocol. Lastly, it is noted that the heuristic does not attempt to find the fair allocation of resources among nodes, and game theory is necessary to achieve this.

Chapter 7 presents the game theory model that achieves a fair allocation of data rates among all nodes in the network. The different measures of fairness most commonly found in resource allocation problems; the advantages and disadvantages of each metric, and why Jain's fairness metric is the most feasible metric for the rate adaptation problem are discussed. This chapter then continues by defining the game theory model by illustrating how it follows a static non-cooperative model. The chapter indicates that the problem could be modelled as a repeated game with an infinite horizon. Finally, a mathematical analysis of the model is described and it is shown how cooperation could be achieved by making use of the Folk theorem.

Chapter 8 illustrates the performance of the rate adaptation heuristic and rate fairness model. In this chapter it is shown how the heuristic and fairness model performed when subject to different amounts of interference. This was achieved by means of the different fixed topologies as was described in Chapter 3. The rate adaptation heuristic and fairness model are evaluated by making use of both data and voice traffic (discussed in Chapter 4). The chapter indicates that a fair and near-optimal rate allocation can be achieved. Furthermore, it is shown that by taking advantage of lower source coding, an increase of about 10% in terms of data rate can be achieved (compared to data traffic). This chapter also illustrates that the rate adaptation heuristic and fairness model can be applied to large networks.

1.6 PUBLICATIONS

This section describes a list of peer reviewed conference and institute for scientific information (ISI) accredited web of science journal papers. The publications deal with the advances made in rate adaptation for wireless networks and the current necessity thereof.

1.6.1 Conference on Rate Adaptation

The following conference paper was submitted to IEEE International Conference on Communications (ICC) for publication. This paper deals with rate adaptation when subject to uniform traffic.

P.A. Jansen van Vuuren, A.S. Alfa, and B. T. Maharaj, “A fair solution to optimal rate adaptation in wireless mesh networks," *in IEEE ICC 2014 – Mobile and Wireless Networking Symposium (ICC 14 MWN)*, Sydney, Australia, Jun. 2014, in review.

1.6.2 Journal on Rate Adaptation

The following journal article was submitted to Elsevier Ad Hoc Networks journal for publication. This paper deals with the effective rate adaptation in a wireless network when subject to data and voice traffic. The work presented in this paper is based on the work presented in this research project.

P.A. Jansen van Vuuren, B.T. Maharaj and A.S. Alfa, "Fair Rate Adaptation for Wireless Mesh Networks using Data and VoIP Traffic," *Ad Hoc Networks*, Aug. 2013, in review.

1.6.3 Other ISI Accredited Journal Publication

The following journal article was co-authored and accepted for publication in Elsevier Measurements journal. This work deals with the measurement of spectrum occupancy in South Africa. The motivation for an improved rate adaptation technique is partially based on this article.

S.D. Barnes, P.A. Jansen van Vuuren, and B.T. Maharaj, "Spectrum occupancy investigation: Measurements in south africa," *Measurement*, vol. 46, no. 9, pp. 3098 - 3112, 2013.

CHAPTER 2

WIRELESS MESH NETWORKS

2.1 INTRODUCTION

This chapter discusses WMNs, including their advantages and challenges, and telecommunication standards. WMNs have several advantages and significant challenges. Similar to the nodes in a traditional fixed wired network, nodes in a WMN are developed by making use of a layered model. The most commonly used layering model is the 7 layer OSI model. Although WMNs and their wired counterparts share similar characteristics, WMNs differ in their ad hoc nature. WMNs follow no fixed topology and are often dynamic, in the sense that nodes could be mobile [4]. These differences provide unique challenges, such as the requirement of self-configuring and self-organising nodes. Several layers in the traditional fixed network are no longer applicable to WMNs. One of the most significant challenges, lies in the network layer, where routing is required to deal with these dynamic changes [5].

Several methods have been developed to address these challenges. One of the most active methods is game theory. The issue of developing self-configuring and self-organising nodes in an optimal manner can be addressed by models based on game theory. Game theory deals with the interaction of nodes in a network, based on the assumption that they would act rationally [7]. This assumption is justified by considering that nodes act as they are programmed, which in turn, is based on the standards governing the development of these devices.

Resource allocation deals with the optimal distribution of limited resources among competing nodes. Traditionally, these problems are dealt with by using mathematical programming techniques. Linear programming (LP), non-linear programming (NLP), quadratic, convex and integer programming are examples of mathematical programming used from resource allocation [13]. Although these techniques lead to exact and optimal solutions, several real-world problems cannot be solved using them. These problems are referred to as NP-hard problems, and do not have exact and optimal solutions. NP-hard problems are often addressed using heuristics [14]. Heuristics are algorithms that rely largely on past information or events to effectively alter future events and often result in sub-optimal solutions.

Several standards for telecommunication exist, and all have major effects on the devices making use of them. The IEEE has developed some of the most significant telecommunication standards. When it comes to networking, four standards prevail; IEEE 802.3 for fixed wired networks, such as Ethernet, IEEE 802.11 for wireless local area networks (WLAN), IEEE 802.15 for personal area networks (PAN), typically used in Bluetooth applications, and IEEE 802.16 for metropolitan area networks (MAN).

2.2 OVERVIEW OF WIRELESS MESH NETWORKS

A communication network refers to the connection of multiple communication devices, such as; computers, telephones or radios, among other things. These devices usually connect to one central relay in a star topology or directly interface each other using a token ring topology [15]. Wireless networks are communication networks where the channel is free space (or air) and typically connects to a base station or mobile station. WMNs are regarded as the interconnection of peer-to-peer communication devices or nodes. Figure 2.1 (a) depicts the structure of a WMN in terms of communication devices (nodes) and links. These nodes only form connections with other nodes if they are within the communication range and thus form an arbitrary topology. Figure 2.1 (b) illustrates the communication range of each device in the same WMN. These nodes are required to be self-organising and self-configuring, allowing communication in an environment with no infrastructure [16]. Nodes in WMNs are often required to perform functions traditionally executed by routers, and this leads to the notion of a multi-hop relay [4]. A multi-hop relay in a WMN may be defined as the forwarding of information from one node to another by making use of other nodes in the network. Nodes could have multiple wireless interfaces, all connecting to multiple devices using different wireless access technologies [5].

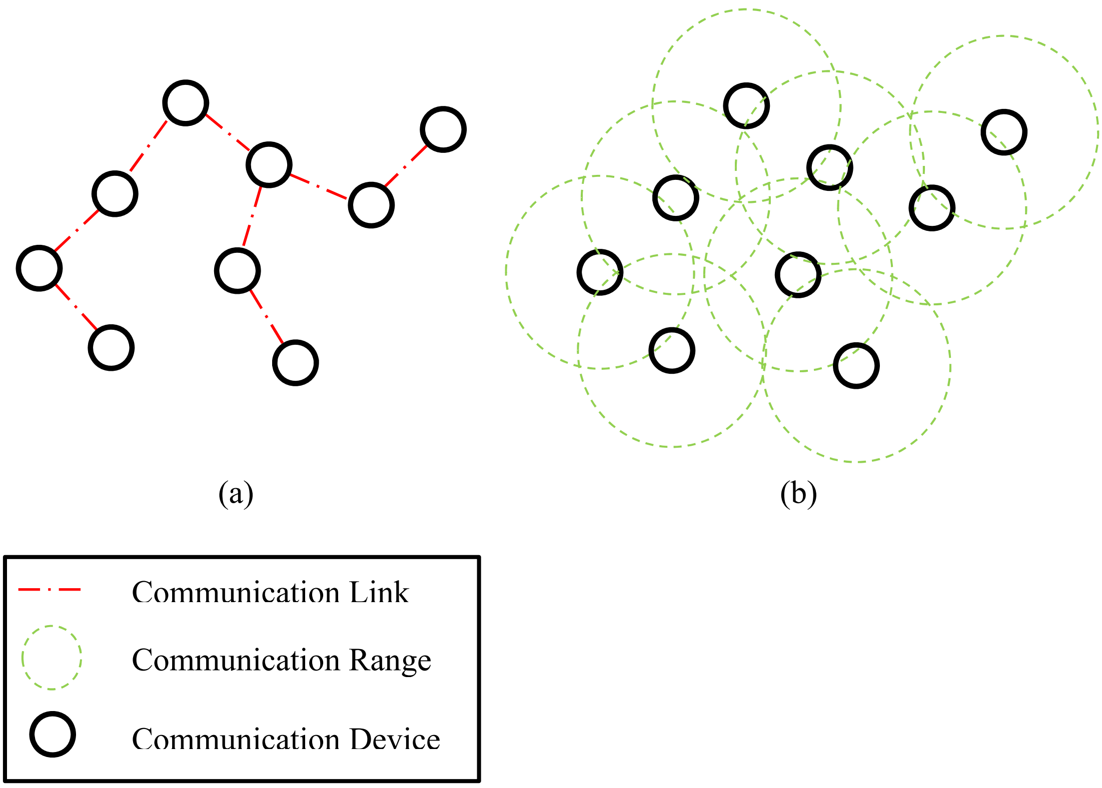


Figure 2.1: (a) An illustration of the formation of communication links between nodes in a peer-to-peer connected WMN. (b) The communication range of nodes in the same peer-to-peer connected WMN.

WMNs have several advantages compared to their fixed wired counterparts. The fact that no infrastructure is required leads to easy deployment and low setup cost of WMNs [16]. The maintenance is low due to the self-configuring nature of WMNs;

adding or replacing a node requires no additional wires or hardware (other than the node itself) [5]. Large area coverage can be achieved when several nodes are spread out in a WMN. In addition, this coverage will be achieved at a lower transmission power compared to a situation making use of a base station [17]. As users have become more accustomed to communication at any destination and at any time, mobility has become an expected feature of the communication device. WMNs provide mobility in the sense that the node can move and maintain communication capability as long as the node is within communication range of another node, that is connected with the remainder of the network. Lastly, when a link between nodes or the nodes themselves become unreliable, the surrounding nodes can detect these changes and adapt accordingly, thus nodes are considered self-configuring [5].

WMNs find ideal applications in rural areas where no infrastructure for communication is available [18, 19, 20]. Metropolitan areas normally consist of a dense population, and fixed infrastructures in these areas often fail to meet the needs of the clients. WMNs provide an alternative solution to supply the demand for communication of clients in these areas, as they provide a large coverage area [21, 22, 23, 24, 25]. Furthermore, users in a metropolitan area expect mobility and as demand increases, normal wireless communication infrastructure may become insufficient. Natural or man-made disasters lead to the destruction of communication infrastructure. In many cases, these disasters lead to situations where communication becomes imperative. Due to the ease and low cost of installation, WMNs become the natural choice for communication in these situations [26, 27]. In a military environment, soldiers are required to communicate while moving into enemy territory. In such cases, a WMN can be deployed and soldiers can move freely whilst communicating, provided that proper encryption and concealment has been implemented [4, 28, 29].

Although, WMNs provide several significant advantages, it also provides several challenges. Scalability has always been a challenging problem faced when considering routing protocols [30]. Routing protocols are responsible for creating efficient routing tables that can be used by nodes for establishing paths between the source and destination of a message. These tables grow larger as more nodes join the network. Routing protocols achieve network convergence when all nodes have obtained a complete routing table. However, as these tables become larger, the overhead of the network increases and the effective performance of the network decrease. Another problem encountered when dealing with WMNs is that no central governing entity exists [31]. This leads nodes to become non-cooperative and there is no central authority forcing these nodes into cooperation. The non-cooperative nature of nodes in a WMN can be observed when considering that different nodes might be part of different service providers and constraints such as memory, capability and battery power create incentives to not cooperate. WMNs are of a dynamic nature; links form and break as time goes on; nodes join and leave the network, and channel conditions change [5]. Lastly, nodes in a WMN share the communication channel (a limited resource), resulting in a competitive environment where selfish behaviour has been shown to lead to sub-optimal performance [6].

The performance in terms of end-to-end delay of a WMN is defined by the effectiveness of its routing protocol [30]. However, routing protocols applied to large networks often take long to converge and result in high overhead [32]. Shortest path and minimum cost trees are two routing approaches used when developing routing protocols. The minimum cost tree approach leads to an NP-complete problem that takes exponentially longer to solve as the network becomes larger [33]. As the quality of service (QoS) becomes more stringent, the minimum cost tree approach becomes more infeasible, as it results in longer delays. Although the shortest path tree approach can be solved in a shorter time compared to the minimum cost tree, overhead becomes a major challenge when developing routing protocols using this approach. OLSR offers a solution that minimises the overhead which leads to fast topology adaptation and network convergence [34].

OLSR makes use of the shortest path tree approach which is computed using the Dijkstra algorithm and could be applied to large networks [34]. OLSR takes care of the overhead instigated by routing protocol by making use of MPRs [35]. MPRs are nodes selected specifically for forwarding routing table information; all other nodes are not allowed to forward this information. Convergence is achieved in the same time (compared to situations not making use of MPRs) by selecting the MPRs as the centroids of the networks (nodes select their own MPRs). Furthermore, OLSR is a proactive routing protocol and thus forces the network into performing routing functions periodically [34]. This ensures that dynamic changes like link forming and node movement in the network are detected and properly accounted for.

It is clear from the description of OLSR that there is no central coordination or authority regulating the nodes. Every node relies on other nodes to forward packets of information, although constraints such as memory, capability and battery power still play an important role [5]. Although OLSR might work well when all the nodes forming the network belong to a single entity (like a service provider), having multiple entities participate in the same network would lead to a non-cooperative environment. Game theory has been applied to WMNs as a mathematical tool for analysing non-cooperative situations or developing methods that ensure cooperation [36].

2.3 GAME THEORY IN WIRELESS NETWORKS

Game theory can be considered as a set of mathematical tools used for modelling the interactions between entities, from here on referred to as users [7]. These users are considered competing users, as they often share some limited resource. The action of one user typically affects the action of its competitors. The accurate representation of the action a user would take, relies on the most fundamental assumption of game theory: all users are assumed to be rational [6]. This assumption allows for the prediction of the action users would take. Rationality is easily challenged, however, as it is difficult to prove that all users will act as rational users [37]. Traditionally, game theory has been applied in economics, psychology, sociology and politics and users are often individuals. In these fields the rational assumption is difficult to defend, as human beings do not always act rationally. Preference is an important factor when considering rationality. In a traditional sense, it can be said that a businessman is acting as a rational individual when he prefers higher profit to lower profit. More recently, game theory received pronounced application in the field of telecommunications. Rationality in this area is more easily justified, as communication systems are designed according to standards [38]. Although communication devices could be designed and manufactured by several different companies, standards define the end goal and constraints placed on these devices. These goals are fixed for the communication device, and in a way define the rationality of the device. In this case, the communication device becomes the user and the user is designed to operate rationally by attempting to maximize its utility or minimizing its cost. Thus when considering rationality in terms of preference, it can be said that a rational user would prefer higher throughput to lower throughput.

A game typically consists of three parts: users, actions and utilities, as illustrated in Equation 2.1 [6].

$$G = (N, \mathbf{A}_{i \in N}, \mathbf{U}_{i \in N}), \tag{2.1}$$

where N represents the number of users, $\mathbf{A}$ represents the action set and $\mathbf{U}$ represents the utility set. A single game could be made up of several users. Users could be represented as messages being passed between nodes (communication devices), flows of information from source to destination, paths between source and destination or the nodes themselves, among other things. The act of selecting a specific path or setting a specific rate can be seen as an action taken by a user, which in turn is contained within an action set. An action set lists all the possible actions a user can take, whilst considering the actions taken by other users [7]. The utility functions of a game are those functions that users would attempt to maximise [39]. Each user is allowed to have a utility function completely unique from other users, although this is not a requirement and all the users may also have the same utility function [40]. The utility function is considered a crucial aspect of the game as it describes the preference of the users.

The Nash equilibrium is defined as the state at which no user could improve their utility by changing their action as illustrated in Equation 2.2.

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \quad \forall\, a_i \in \mathbf{A}_i, \tag{2.2}$$

where a_i^* is the best possible action for user i and a_{-i}^* is the best possible action for all other users except i. A rational user would not want to change their current action at this state (a_i^*), as it would lead to a utility lower than their current utility [41]. In some cases it can be shown that this equilibrium is not optimal. The term “Pareto optimal” refers to the state where no user can change their action in an attempt to improve their own utility, without lowering the utility of other users. This term is used to define a situation which is optimal in a system-wide sense [42].

Non-cooperative game theory is the most popular type of game theory applied in telecommunications [43, 44, 45]. Non-cooperate game theory deals with interaction of users in a competitive environment. Examples of such situations include selecting paths with high throughput but lowering their remaining capacity; choosing to forward messages at the cost of energy in terms of transmission power, or selecting higher power level effectively inducing more interference. Cooperative game theory, on the other hand, mainly deals with those situations where users have agreed to a set of rules [46, 47, 48]. Cooperation in this case is achieved when all users adhere to these rules and this leads to two branches of game theory, namely bargaining theory and coalition games [6]. It should be noted, however, that users in a non-cooperative game can cooperate, although this cooperation should be self-sustaining [7].

In a static non-cooperative game, users take actions only once as the user has no incentive to change its action. Static games normally imply that users have no information about the actions of other users and thus no incentive to change their original action [6]. In a dynamic non-cooperative game, users reselect their actions at every stage of the game [49]. For a dynamic game Equation 2.1 would have to include a stage variable, as illustrated in Equation 2.3:

$$\Gamma^r = (N, \mathbf{A}_{i \in N}, \mathbf{U}_{i \in N}), \tag{2.3}$$

where r represents the current stage of the game. In dynamic games, time plays an important role as it defines the stages of the game. At each stage, each user would receive information about the actions of the other users in the game. The notion of a strategy arises when a user takes an action based on the current information received [7]. If all the users know all the possible actions and utility of each other, the game is considered a dynamic game with complete information. It is not always possible for all users to have complete information, however, and techniques for achieving complete information often lead to added complexity (if estimation is required) or overhead (if message passing is required) [50]. Games with incomplete information are referred to as situations where users only have partial information regarding actions and utility of other users [6].

A repeated game is a static non-cooperative game that is repeated over time [45]. In a repeated game, users gain information on past actions of the other users. This past information often leads to a game with incomplete information, thus a repeated game can be considered a type of dynamic game. The past information is captured in a history and each user builds a history set [38]. The history set allows users to observe the actions of other users and plan strategies accordingly. A repeated game with infinite horizon is a game that would seem to continue forever. The general form of the utility function for a repeated game is given in Equation 2.4:

$$u_i = (1-\delta)\sum_{k=0}^{\infty}(\delta)^k g_i(a^k), \tag{2.4}$$

where k is the number of the current stage game and $g_i(a^k)$ is the kth stage utility or payoff for user i as a function of the action profile a^k. Real-world applications do not continue forever, however, in an infinite horizon repeated game, the users only need to believe the next stage of the game is not the final stage. Cooperation in an infinite horizon repeated game can be achieved by making use of the folk theorem [6]. This is normally achieved by punishment; users do not deviate from cooperation as the utility gained by doing so does not outweigh the utility lost through punishment [51]. Cooperation in these situations is shown to be self-sustaining and leads to the Pareto optimal solution [38]. Figure 2.2 summarises the different types of game theory models discussed in this section.

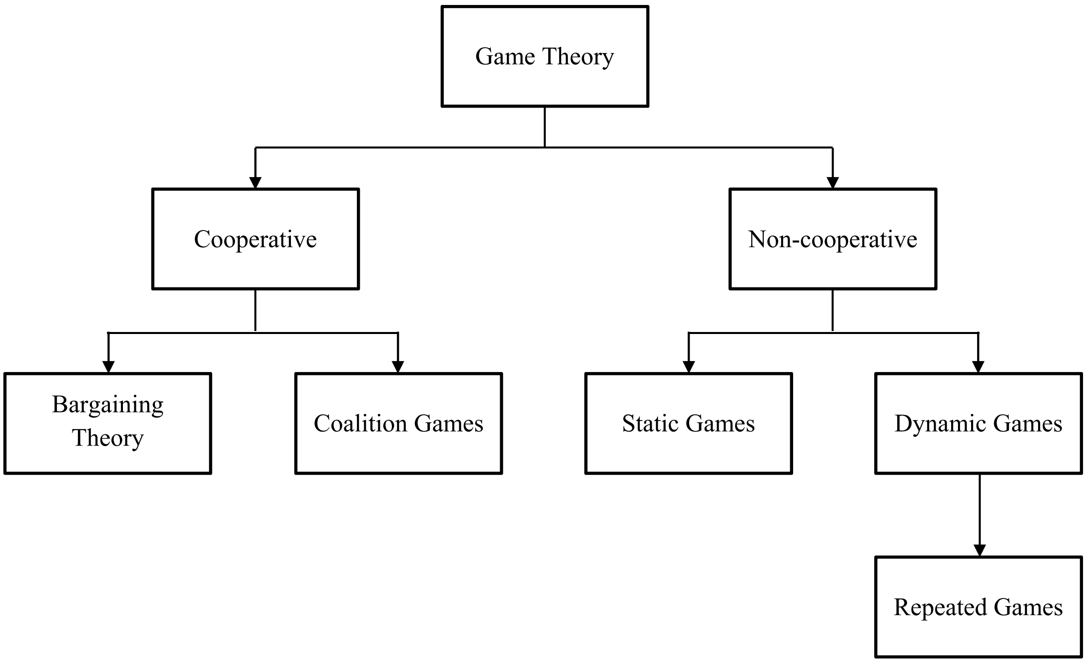


Figure 2.2: A summary of the different game theory models.

Game theory has been applied in a variety of applications in telecommunications. Game theory has been used to model call admission problems at the transport layer [52, 53, 54, 55]. Here a fair and efficient policy for controlling admission is modelled as non-cooperative game. These policies determine when certain traffic is allowed access to the network and typically attempt to maximise the QoS and balance the load. The network layer deals with routing traffic efficiently, and it is thus easy to see how game theory could be applied here [45, 56, 57]. In WMN, users depend on each other to forward their information from source to destination. Forwarding a message does not necessarily produce any benefit to the user doing the forwarding (and could even be considered a loss). Nevertheless, if no user is willing to forward a message, the network would collapse and no communication would be possible. The selection of the shortest and least congested path is another example where game theory could be applied. In this instance, the more users use a specific path (normally the perceived shortest path), the more congested that path becomes and less attractive it becomes. Traditionally, game theory is applied in an environment where users act selfishly. When considering network security, game theory could be applied to model malicious users [58].

When considering the physical layer, game theory has been applied to the power control problem [43, 47, 59]. The SINR is the main constraint in these problems and users are presented with a conflict of interest. When one user increases its transmit power levels, it will enjoy a higher SINR and ultimately better performance in terms of throughput. Higher transmit power levels for one user leads to more interference for other users and ultimately lowers their SINR. Game theory is used in these problems as a tool for modelling the non-cooperative nature of the users in such environments.

2.4 RESOURCE ALLOCATION IN WIRELESS NETWORKS

Resource allocation refers to the distribution of resources that are often limited among entities (that often compete with each other) [8]. Distributing these resources in an optimal manner is an active field of research in the telecommunications area. LP is a mathematical tool traditionally used in operation research for solving optimisation problems [60]. These problems are modelled by describing three properties, namely the objective function, decision variables and constraints. The objective function defines the utility that a user would like to optimise, for example, the objective function of a communication device might be to minimise packet loss. This example indicates that an optimisation model could consist of several objective functions, as a node might also want to maximise its throughput. The decision variables are the variables the users could alter in pursuit of optimising their objective function. Again, more than one decision variable could exist, and thus the objective function could be a function of more than one variable. The constraints determine the extent of the alteration the user could perform on the decision variables. Typically, several constrains exist as they relate the physical limitations of the real-world application to the mathematical model.

LP problems are traditionally solved using the simplex algorithm, although in some simple cases, LP problems can be solved graphically [13]. Figure 2.3 illustrates how a LP problem can be solved graphically: corner points (A and B) indicate optimal points, the hatched area indicates the feasible region. Objective functions consisting of two decision variables can be solved graphically. Although most problems consist of more than two decision variables, these solutions provide insightful results. The simplex algorithm does not require the number of decision variables to be limited. The first step of the simplex algorithm is defining the problem in standard form, followed by generating more canonical forms of the problem in an attempt to find the feasible solution [60]. Several other methods for solving LP problems include the interior-point method (or more specifically the path following method). This method is an iterative method that moves closer to the optimal solution at each step.

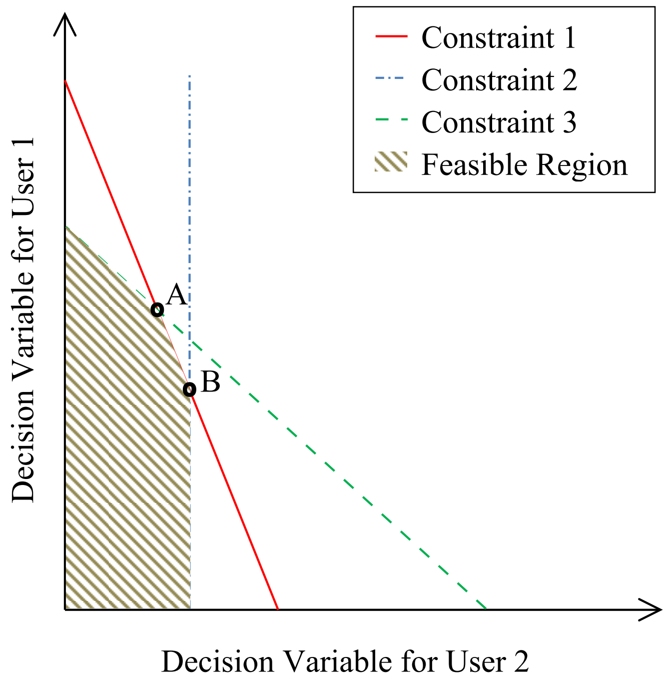


Figure 2.3: An illustration of how LP problems can be solved graphically.

Integer programming is a special case of LP, where each decision variable can only take on integer values [13]. Many real world problems can only be solved using integer programming. The optimization of travellers on a bus route at a specific time is an example of integer programming. Using LP to solve such a problem might lead to situations requiring fractions of travellers to take certain bus routes. This is not possible because people cannot be divided into fractional portions. Information packets in a WMN often take on specific sizes, as node buffers have a limitation on the number of packets they can queue. Thus the assignment of a specific information packet to a specific route forms another integer problem. Integer programmes are traditionally solved by making use of column generation [61].

In several real-world problems, the objective function or some of the constraints are not linear [62, 63]. In these situations, solutions such as the simplex algorithm cannot be applied, thus NLP arises. In some special cases of NLP, the objective function is allowed to contain products of variables [13]. In these cases quadratic programming may be used to model the problem. A broader case states that as long as the objective function is convex, a solution can be found [14]. This is referred to as convex programming and employs the prime-dual theorem and interior-point method [13]. Although LP and NLP provide solutions to several optimisation problems, many real-world problems are known to be NP-hard or NP-complete. These problems do not have an exact solution and the use of heuristics to find near-optimal solutions becomes necessary.

Resource allocation has been used in WMNs for load balancing. Generally, balancing loads in a network refers to the fair allocation of bandwidth to users making use of the network. This problem is challenging as WMNs are expected to serve many users and thus high volumes of traffic are generated. A solution to load balancing at an Internet gateway (IGW) level has been proposed in [64]. This solution involves a heuristic that attempts to mitigate congestion. The allocation of a set of subcarriers and total transmit power in a multi user orthogonal frequency division multiple access (OFDMA) system to users is another example of resource allocation [65]. Here the amount of sub carriers assigned to each user or node is determined by the required data rate of the traffic traversing that node. The transmit power is then allocated across the subcarriers based on the channel. The allocation of traffic to certain nodes in an attempt to balance the traffic load of buffer queues is an active field of research when considering WMNs. The assignment of packets to specific routes based on the buffer queue of the node has been investigated by making use of queuing theory [66]. The design follows a cross-layer approach and the problem is formulated as a NLP. Time division multiple access (TDMA) is a method used for dividing users into separate time slots, in an attempt to reduce the chances of users interfering with each other. A technique for the optimal allocation of time slots to users, along with spatial reuse in a WMN has been developed in [67]. The technique is based on integer programming and is extended to approximate dynamic programming.

2.5 IEEE 802.11 STANDARD

Standards play a major role in the development of communication systems. Traditionally, IEEE 802.3 is the most common standard when developing a wired network [68]. IEEE 802.3 defines the physical and data link layers for Ethernet networks. IEEE 802.11 can be seen as the wireless counterpart of IEEE 802.3, as both specify requirements on the physical and data link layers in a similar fashion (both standards make use of code division multiple access [CDMA]) [69]. Other communication standards include IEEE 802.16; which is used in wireless MAN and is often referred to as the worldwide interoperability for microwave access (WiMAX) standard [70]. IEEE 802.16 utilises orthogonal frequency division multiplex (OFDM) at the physical layer and includes multiple input multiple output (MIMO) features. IEEE 802.15 is used in PAN and is typically employed in systems making use of Bluetooth [71]. Although there are several communication standards, IEEE 802.11 is the most applicable standard for WMNs as it is traditionally applied to WLAN, due to the distances that are defined in the standard as well as some physical layer specifications. However, the data link layer often requires alterations [72].

IEEE 802.11 features, such as distributed coordination function (DCF) and carrier sense multiple access with collision avoidance (CSMA-CA), offer solutions that are well suited for WMNs [73]. The CSMA-CA technique offers a method for sensing the occupancy of a channel. IEEE 802.11 utilises direct sequence spread spectrum (DSSS) techniques which allow simultaneous channel access for multiple users in the network. Traditionally IEEE 802.11 operates at 2.4 GHz, however the standard also allows operation in the 5 GHz band [69]. These techniques allow devices to communicate with each other without a centralised authority. One disadvantage is that the standard medium access control (MAC) layer together with TCP can lead to severe unfairness across the network [72].

The IEEE 802.11 standard has been used as a backbone network for many WMN applications. A novel link adaptation technique called MAC Protocol Data Unit (MPDU) link adaptation has been developed and is based on IEEE 802.11 [74]. This technique is table driven and relies on the DCF functionality provided by this standard. Both the 2.4 GHz and 5 GHz bands can be utilised in a WMN making use of IEEE 802.11 standard [75]. Here the channel assignment is achieved by making use of multiple network interface cards (NICs) and combined with routing to achieve improved load balancing across a WMN. An analytical model for queue delays on a node in a WMN has been developed [76]. The nodes were modelled with G/G/1 queues and the channel access delay, packet collision and packet size were taken into consideration. The MIMO features have also found application in WMNs, where optimal antenna selection algorithms are applied in an attempt to increase the channel capacity [77]. These techniques suffer from a trade-off between complexity added to the device and the performance gain.

As indicated earlier, IEEE 802.11 utilises collision avoidance techniques. This is done to ensure that the wireless devices in the network do not interfere with each other. The hidden node problem provides a challenging problem when using this technique. When two or more nodes in the network are actively sensing the channel, all these nodes might sense no occupancy thereby leading to simultaneous transmission [72, 78]. DSSS techniques such as CDMA provide multiple users access to the channel simultaneously by assigning each user an orthogonal code. The typical implementation

Table 2.1: Summary of the MCS for IEEE 802.11.

MCS	**Modulation**	**FEC**	**Receiver Sensitivity**
0	BPSK	1/2	-82 dBm
1	QPSK	1/2	-79 dBm
2	QPSK	3/4	-74 dBm
3	16-QAM	1/2	-71 dBm
4	16-QAM	3/4	-70 dBm
5	64-QAM	2/3	-66 dBm
6	64-QAM	3/4	-65 dBm
7	64-QAM	5/6	-64 dBm

of CDMA in a WMN involves a 11 chip Barker sequence and results in a maximum capacity of 1 Mbps [79]. Forward error correction (FEC) coding is also implemented for the further enhancement of performance in IEEE 802.11. These techniques lead to a degradation in the throughput due to the addition of parity bits. Additionally, a cyclic redundancy check (CRC) technique is applied which detects errors made on a frame level.

Standards define the structure and features of communication devices (at least at the physical and data link layers for IEEE 802 standards), but also define the constraints placed on these devices. By allocating the 2.4 GHz band (the industrial, scientific and medical [ISM] band) as the transmission frequency of the device, the device becomes constricted. According to the international telecommunications union (ITU), all devices operating in this band are allowed a maximum of 100 mW of transmission power. This, in turn, restricts the range at which the device can communicate [80]. Other constraints include factors such as the bandwidth, which is fixed to 20 or 40 MHz. The bandwidth effectively determines the available channel capacity. The modulation type is either OFDM or CDMA. In either case the complexity of modulation is defined and places a constraint on the processing time [81]. Table 2.1 summarises the modulation and coding schemes (MCSs) for IEEE 802.11.

2.6 CONCLUDING REMARKS

This chapter discussed the development, advantages and challenges of WMNs, game theory, resource allocation and telecommunication standards in detail. Chapter 3 covers network topologies.

CHAPTER 3

NETWORK TOPOLOGIES

3.1 INTRODUCTION

This chapter describes network topologies and how they are used to evaluate the performance of the rate adaptation heuristic and fairness model. The network topology plays an important role in the performance of heuristics and protocols that are applied to the network. In the rate adaptation problem, the topology defines the location of neighbouring users and the effect they have on each other, therefore the amount of interference experienced by a user depends on the network topology [82]. Interference occurs when two or more users are transmitting simultaneously, given that somewhere along the path, they are connected via a link.

There is an associated node for each user. This node represents the communication device used for the exchange of information. All the nodes in the network are usually assumed to be heterogeneous, which requires them to be a specific device. The model in this study does not require the devices to be the same, although the devices have to follow the network model as discussed in Chapter 4. Nodes are connected to each other via links and are typically defined by the channel used, which in this case is an additive white Gaussian noise (AWGN) channel. A path is a set of links used to connect a source node to a destination node. There is an associated flow with each user, which defines the motion of information from a source to a destination node.

Furthermore, each flow is not limited to a single path, and multiple paths (with the same minimal distance) can be used to convey information. Table 3.1 summarises the parameters in the network topology. In this study three fixed topologies are presented.

Table 3.1: Entities found in all topologies.

Entity	**Symbol**	**Description**
Node	N	A device capable of wireless communication, following the model described in Chapter 4. N_1 refers to node 1.
Link	L	A wireless link connecting two nodes that is defined by the channel, which in this case is an AWGN channel. L_{12} refers to the link between N_1 and N_2.
Path	P	A set of links connecting a source and destination node. P_{123} refers to a path connecting N_1 (the source node) and N_3 (the destination node).
Flow	F	The flow of information from a source to a destination node, but is not limited to a specific path. F_{13} refers to the flow of information between N_1 and N_3.

Although these topologies do not reflect real world situations, they do provide significant advantages in terms of evaluating the performance of rate adaptation in an interfering environment. Furthermore, a randomised topology is presented, which provides more realistic representation of a WMN.

3.2 STACK TOPOLOGY

The stack topology is the simplest topology. This topology consists of nine nodes, of which three nodes (N_1, N_2 and N_3) connected with two links (L_{12} and L_{23}) are stacked on three other nodes (N_4, N_5 and N_6) also connected with L_{45} and L_{56}. The set of six nodes are stacked on another three nodes (N_7, N_8 and N_9) in the same way. The first three nodes are connected to the next set of nodes by a common link (L_{25}) and the second set is connected in the same way.

Figure 3.1 depicts the stack topology and three flows, namely an outer flow F_{13}, an inner flow F_{46}, and another outer flow F_{79}. In this topology, each flow can only move along a single path. F_{13} and F_{46} interfere with each other along L_{25}, while F_{79} and F_{46} likewise interfere with each other along L_{58}. It should be noted is the fact that the outer flows place a constraint on the inner flow due to interference, while the inner flow likewise places a constraint on each of the outer flows. The outer flows, however, do not place any constraints on each other. Significantly, in this topology

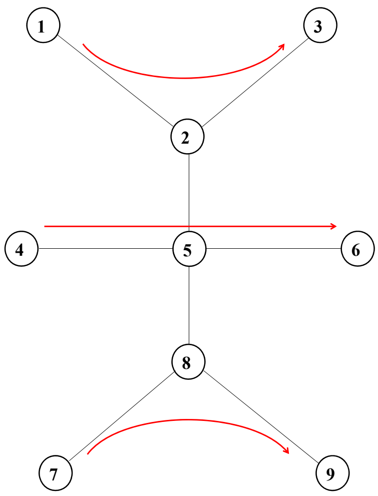


Figure 3.1: Stack topology illustrating three flows with nine nodes, adapted from [10].

each flow only has a single shortest path. Thus F_{13} can only convey information along path P_{123}; F_{46} can only make use of P_{456}, and F_{79} can only use P_{79}. If these flows select any other path that is not the shortest path, the end-to-end delay will increase and performance will decrease. This discussion might seem trivial, however it becomes more significant when comparing this topology to other topologies.

3.3 HALF DIAMOND TOPOLOGY

The half diamond topology is similar to but slightly more complex than the stack topology. The half diamond topology also consists of nine nodes. Again three nodes are stacked on another three nodes and those six nodes are stacked onto another three nodes. The half diamond topology differs from the stack topology by adding two additional links, first between N_2 and N_4 (namely L_{24}) and then between N_2 and N_6 (namely L_{26}).

Figure 3.2 illustrates the half diamond topology and three flows can be identified; F_{13}, F_{46} and F_{79}. This topology has an additional path available for F_{46}. Information on this flow can select both P_{456} and P_{426} with the same delay in terms of number of hops. However, there is a distinct difference between selecting one path compared to another. When considering the congestion on a specific path, F_{46} can alternate between using P_{456} and P_{426}, which in turn could lead to higher throughput. This performance gain, however, relies on the assumption that the buffer of N_2 can handle the traffic generated from both F_{13} and F_{46}. Although the extra

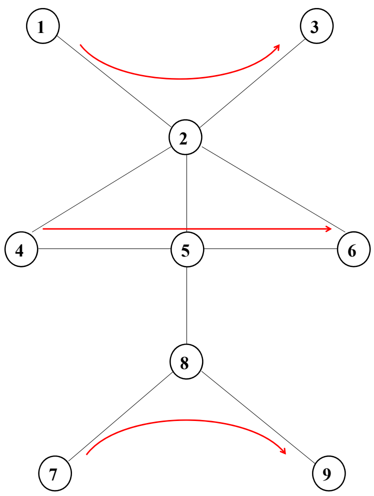


Figure 3.2: Half diamond topology illustrating three flows with nine nodes, adapted from [10].

links appear to offer an advantage in terms of additional paths, the links also induce more interference. For example, when the worst case path is selected (which in this instance is P_{456} along L_{45} and L_{56} when assuming an infinite buffer size) F_{46} induces more interference compared to the equivalent flow in the stack topology. Thus F_{46} will interfere with F_{13} along L_{24}, L_{25} and L_{26}. This increased interference, in turn, leads to a reduction in performance when considering throughput. The remaining flow, F_{79} acts similarly to its equivalent flow in the stack topology; F_{79} and F_{46} only interfere with each other along L_{58}.

3.4 FULL DIAMOND TOPOLOGY

The full diamond topology is the most complex fixed topology, although it simply extends the stack and half diamond topologies. Again, nine nodes are stacked on each other, but with an additional four links compared to the stack topology. These links, namely L_{24}, L_{26}, L_{48} and L_{68}, are used to connect nodes N_2, N_4, N_6 and N_8.

Figure 3.3 illustrates the full diamond topology with the same three flows. The additional links in this case lead to two additional shortest paths that F_{46} could use. Taking congestion into consideration and by assuming an infinite buffer at N_2 and N_8, the performance can be improved by alternating between paths. However, the two additional paths lead to more interference. In this topology, when the worst case path is selected (which is P_{456} along L_{45} and L_{56}), F_{46} induces more interference compared to the equivalent flow in both the stack and half diamond topologies. By selecting this path, F_{46} would interfere with F_{13} along L_{24}, L_{25} and L_{26}. Furthermore, F_{46} would also interfere with F_{79} along L_{48}, L_{58} and L_{68}. Although these fixed topologies do not reflect a real-world scenario, they provide a framework for comparison with each other. These topologies may also be used to compare different algorithms and heuristics with each other. Finally, these topologies have the advantage of being fairly simple and thus easy to implement. The next section describes of how a more realistic topology can be found.

3.5 RANDOMISED TOPOLOGY

A randomised network topology can be defined as a network with no fixed topology. The location of each node, in relation to other nodes is chosen at random. This leads to a random set of links forming between nodes. A unit disk approach is used to

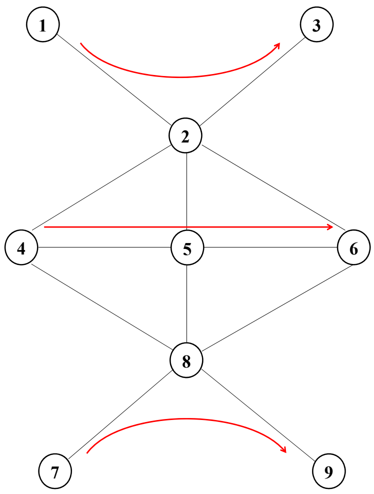


Figure 3.3: Full diamond topology illustrating three flows with nine nodes, adapted from [10].

determine where links should form [80]. In conventional graph theory, a network refers to a graph, and nodes and links to vertices and edges, respectively. When modelling a network with a unit disk approach, the communication range determines the radius of the disk, which is centred on a node. A link between two nodes can only be formed if one node falls within the disk of another node. A unidirectional link is formed when one node is within the disk of another node but the second node does not fall within the first one's disk. A bidirectional link is formed when both nodes fall within each of the other's disks. The actual communication range depends on the device and the propagation of electromagnetic waves. Assuming free space propagation loss, Friis law describes the restriction on the communication range of a node [83]. The equation below illustrates Friis law.

$$P_{RX}(d) = P_{TX}\frac{1}{4\pi d^2}A_{RX}, \tag{3.1}$$

where $P_{RX}(d)$ is the power of the receiving node as a function of the distance, P_{TX} is the power of the transmitting node, A_{RX} is the area of the antenna on the receiving node and d is the distance between the nodes. The distance can be calculated as 140 m, when it is assumed that the minimum receiver sensitivity is -64 dBm. In this study, a network containing 50 nodes was randomly distributed in a 500 m by 500 m area.

The simulation of the random network was then repeated with 75 and 100 nodes in the same area. Each simulated network was repeated 200 times with different seeds (used for random number generators). Figure 3.4 illustrates the randomised network topology of a twelve-node network, repeated three times. The length of the simulation continued for 20 minutes. Table 3.2 summarises the parameters used for the simulation.

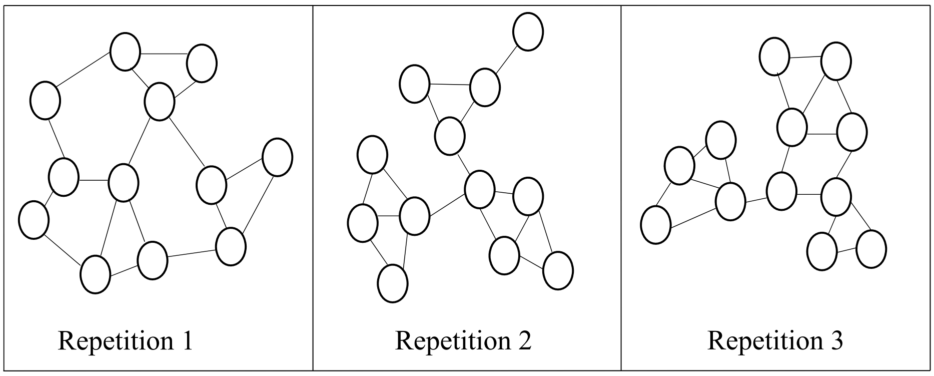


Figure 3.4: A twelve-node randomised network repeated three times.

Table 3.2: Parameters used for the random topology simulation.

Nodes	50, 70 and 100 Nodes
Area	250000 m^2
Number of repetitions	200 times
Simulation time	20 minutes or 1200 seconds

3.6 CONCLUDING REMARKS

This chapter describes the topologies used for simulating the performance of the rate adaptation heuristic and rate fairness model. Three fixed topologies were described, namely, the stack, half diamond and full diamond topology. The differences between these topologies were highlighted in terms of congestion and interference. Finally, the simulation of a random topology was described, and the parameters of the randomised network explained. Chapter 4 covers the network model.

CHAPTER 4

NETWORK MODEL

4.1 INTRODUCTION

Communication devices in a WMN typically follow the seven-layer OSI model. The layers are (from top to bottom); the application, presentation, session, transport, network, data link, and physical layer, as illustrated in Figure 4.1 (a). The application layer provides an interface between the user and the remaining layers. Protocols developed in this layer are specifically designed according to the type of functions used in the network by the users. The two protocols used in this layer are VoIP and FTP. The presentation layer is required for compression/decompression, encryption/decryption, and source encoding/decoding. Several of these features are included in the application layer in this network model. The session layer is responsible for maintaining sessions and reporting errors. The transport layer manages connections and facilitates reliable data transfer. In this layer, TCP and user datagram protocol (UDP) have been selected as transport protocols. Routing and addressing packets are done at the network layer, where OLSR is used. OLSR provides an open shortest path method for finding optimal routing tables in terms of minimum hops [34]. The data link layer enables features such as frame allocation and data flow control. The physical layer forms the connection between the physical channel and the other layers, and is responsible for the physical transmission of information. The IEEE 802.11 defines the structure implemented at this layer. Figure 4.1 (b) illustrates the interconnection of the implemented protocols relative to their layers used in this network model.

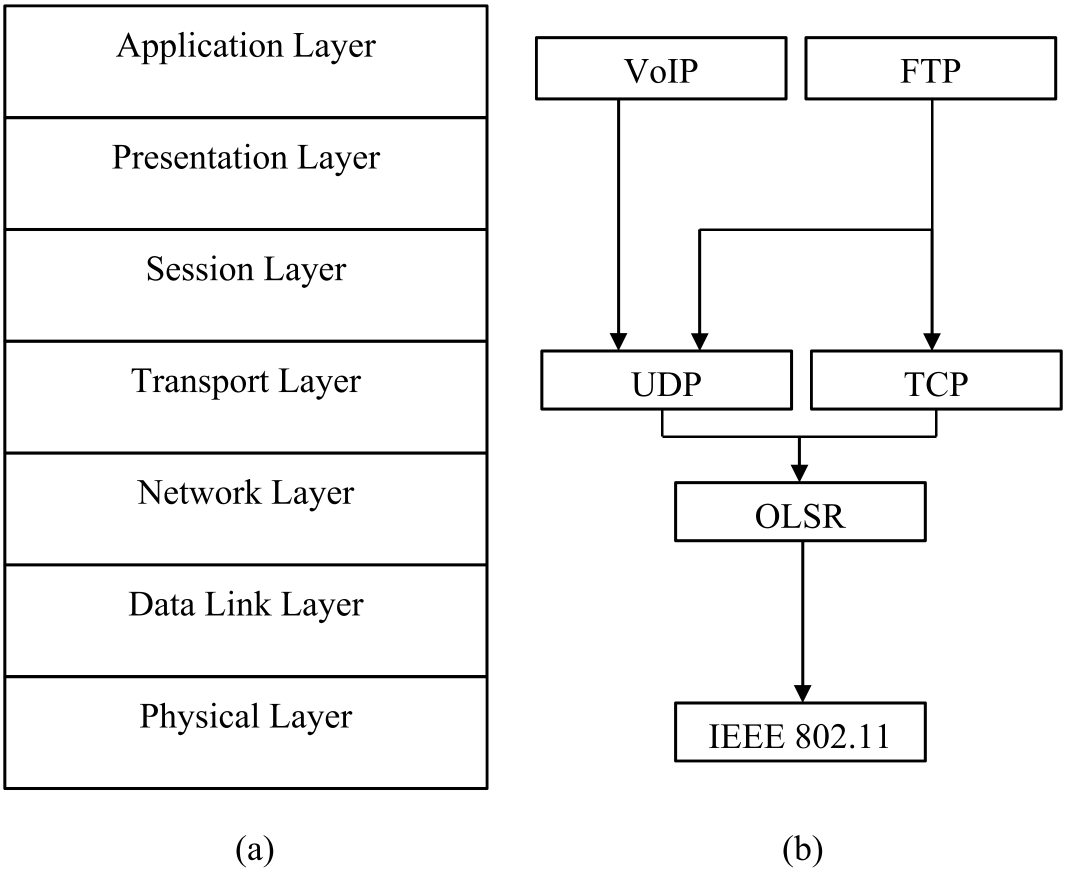


Figure 4.1: (a) The traditional seven-layer OSI model. (b) The network model described in this section.

4.2 APPLICATION LAYER

The application layer utilises two protocols: FTP and VoIP. Originally, FTP was created as a method for isolating various file storage systems while maintaining reliable and effective data transfer [84]. FTP was selected as the application layer protocol for representing data traffic. Traditionally, VoIP was used in networks where voice calls are made and requires unique features compared to FTP. Packet loss results in serious performance degradation in terms of QoS, as high packet loss would lead to voice clipping and skips [85]. Another challenge is that unlike FTP, VoIP is not delay tolerant and large delays would lead to large periods of waiting time between users listening and talking. These delays add to the performance degradation of the QoS. Factors such as the packet structure and voice digitisation (including encoding and compression) also add to the performance degradation when implemented incorrectly. VoIP was selected as application layer protocol for representing voice traffic.

4.2.1 Data Traffic

Although other protocols such as hypertext transfer protocol (HTTP) or Internet message access protocol (IMAP) could be used to represent data traffic at the application layer, FTP was selected as it is most commonly found in peer-to-peer WMN [84]. In this network model, every node will transmit eight packets of data, each consisting of 64 bytes, every second. This assigns each node a data rate of 4 kbps. Each node in the network will select a random destination, defining the end receiver of these information packets.

In this case, that all packets have the same priority, therefore none of the packets can be considered more important than another. FTP allows three main transmission modes: stream, block and compressed mode. In stream mode, files are sent using packets that are constantly sent to the destination. In block mode, files are divided into packets with sequence numbers and asynchronous reception of these packets at the destination is allowed. Compressed mode is used in dense networks where bandwidth is limited [84]. This network model makes use of FTP in block mode and the header structure is indicated in Figure 4.2. The header consists of 24 bits, where 8 bits are used for the descriptor and 16 bits are used to represent the byte count.

Generally, FTP is used in environments where large amounts of data are transferred. In this model, the network actively transfers data for 1200 seconds. If the payload is set to 64 bytes per packet, each node would have transferred 4.6 Mb of data. The reliability of this transfer depends on the transmission mode as well as protocols used in other layers. The reliability of block mode transfer is higher

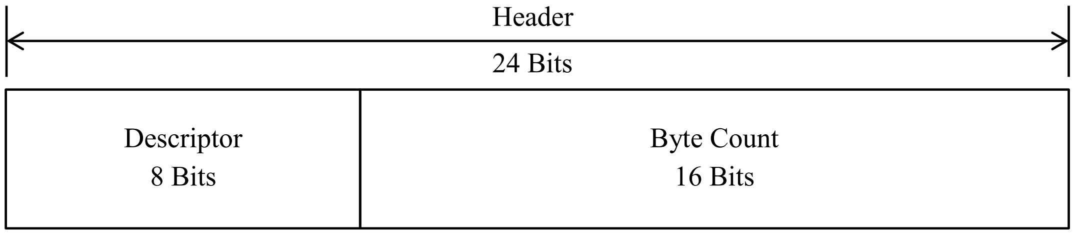


Figure 4.2: The FTP header structure.

compared to stream mode, as stream mode offers no time to receive acknowledgements of reception, unlike block mode. Clearly, then, there is a trade-off between reception time and reliability, as block mode provides reliability at the cost of waiting for acknowledgments. The reliability of transmission also depends on the protocol used at the transport layer. TCP offers acknowledgement of reception features, where UDP offers no such features. Traditionally, TCP is used in combination with FTP, but there is no restriction on FTP that does not allow its use in combination with UDP.

4.2.2 VoIP Traffic

The first step required when utilising VoIP is the digitisation of voice (an analogue signal) using a vocoder. Traditionally, an adaptive multi-rate (AMR) codec or enhanced variable rate codec (EVRC) is implemented to achieve this digitisation. These codices provide compression and encoding. After digitisation, packet construction is required, which is usually achieved by using real-time transport protocol (RTP) in conjunction with UDP [86]. UDP in voice communication is preferred to TCP, as the requirement on the maximum end-to-end delay is stringent. For acceptable QoS, a maximum end-to-end delay of 250 ms is required [85].

In addition to smaller packet size, VoIP typically generates traffic at different priorities. Two different types of packets can be identified for VoIP traffic, this depending on how individuals communicate verbally. Voice packets contain actual voice messages, whilst silent packets contain only silence (generated at times when individuals wait for responses from others). The AMR codec, adapts its source rate depending on the channel conditions [87]. The source coding rate could decrease (thereby leading to lower quality of voice), allowing the channel coding rate to increase. This mechanism allows for effective communication in severe channel conditions at the same data rate, thus meeting the strict delay requirements.

In a similar fashion to the AMR codec, packets with lower priority could be assigned lower source coding rates; silent packets being at the lowest priority are assigned the lowest source coding rate. When considering the AMR codec, the lowest source coding rate would equal 12.5% of the rate of the highest source coding. Thus a simplified VoIP model based on the AMR codec can be developed. In contrast to the AMR codec, this VoIP model always assigns a lower source rate to silent packets, irrelevant of the channel conditions. This leads to unused channel resources if the channel coding rate is not locally adapted according to the source rate [88]. However, instead of wasting this resource by increasing the channel coding rate locally, the rate adaptation heuristic, located at the physical layer (explained in Section 4.4), could take advantage of this unallocated resource, by assigning it to another node currently generating voice or data packets.

Figure 4.3 illustrates a simplified VoIP model using a two-state Markov model. The first state is the voice state, where packets generated are assigned the maximum source coding rate, as they contain the most information. The second state is the silence state; where packets are assigned a source rate of 12.5% compared to that of the highest source rate. The following equation indicates the state transition probability matrix:

$$P = \begin{bmatrix} P_S & P_{VS} \\ P_{VS} & P_V \end{bmatrix} \tag{4.1}$$

Four probabilities can be defined. P_{VS} is the probability that the traffic type moves from the voice state to the silence state, whilst the traffic type is currently in the voice state. P_V is the probability that the traffic type stays in the voice state. P_{SV} is the

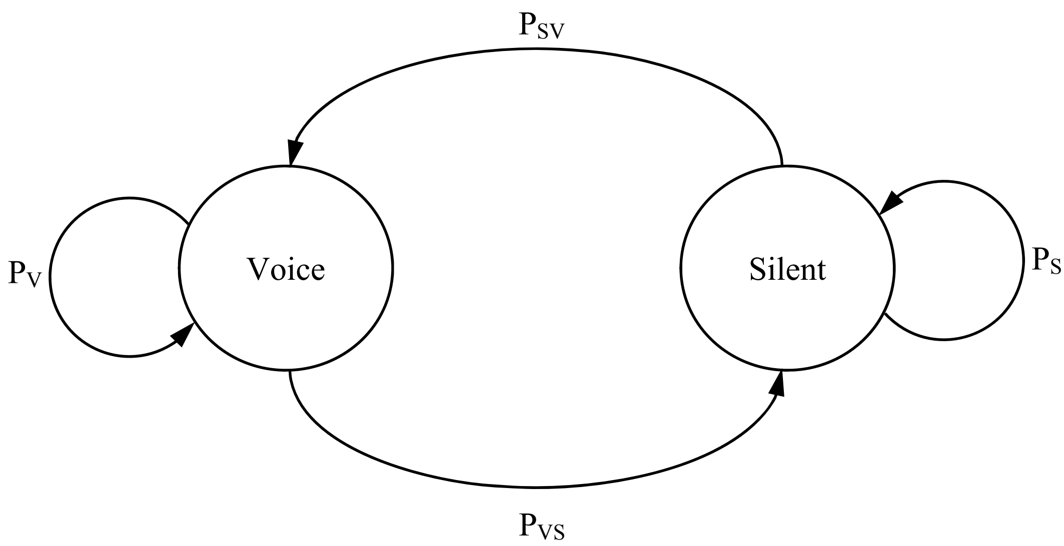


Figure 4.3: A simplified VoIP model making use of a Markov model [7].

probability that the traffic type moves from the silence state to the voice state, while currently in the silence state. P_S is the probability that the traffic type stays in the silence state.

4.3 NETWORK LAYER

This section describes the protocols of the transport and network layer. These protocols provide vital functionality as the network would collapse if they were omitted. TCP and UDP, which are transport layer protocols, are responsible for reliable data transfer. OLSR is responsible for routing information between source and destination in the network.

4.3.1 Transport Control Protocol

TCP is typically used in situations where the reliability of data reception outweighs the requirement for minimum end-to-end delay. Thus, TCP is widely applied in combination with FTP, HTTP and IMAP. Reliability is achieved by making use of positive acknowledgements. Each packet is thus assigned a sequence number and the receiving node sends an acknowledgement of reception back to the transmitting node. Acknowledgement messages are based on the sequence number, notifying the transmitting node which packets have been received and which packets are lost [11]. Error recovery is another feature provided by TCP and is achieved through the retransmission of lost packets. Retransmission occurs when no acknowledgement message is received within the given timeout period.

Flow control features are also included in TCP. When receiving nodes send acknowledgement packets back to the transmitting node, the available buffer size in terms of bytes could be added to the acknowledgement packet. This effectively indicates the number of packets the transmitting node can still send through these receiving nodes. There are two disadvantages in utilising TCP. First, large delays in data transfer are encountered, as the transmitter is forced to wait for acknowledgments from receivers [88]. Secondly, the transmission of acknowledgments along with the required retransmissions induces severe overhead.

The features provided by TCP lead to two requirements. First, transmitting nodes are required to keep track of the time it takes from transmitting a message and receiving the corresponding acknowledgement. Secondly, each transmitting node is required to store all messages until their acknowledgement messages are received, in order to ensure retransmission is possible [11]. Figure 4.4 depicts the header structure for TCP used in this network model. The source field is used by the receiving node as a method for replying (with acknowledgements) to the transmitting node. The destination field is used (by receiving nodes) as a method for forwarding messages to the correct nodes. Both source and destination fields consist of 16 bits.

4.3.2 User Datagram Protocol

In contrast to TCP, UDP provides a method for data transfer with low end-to-end delay. This is achieved by omitting reliability features and thus negating the need to

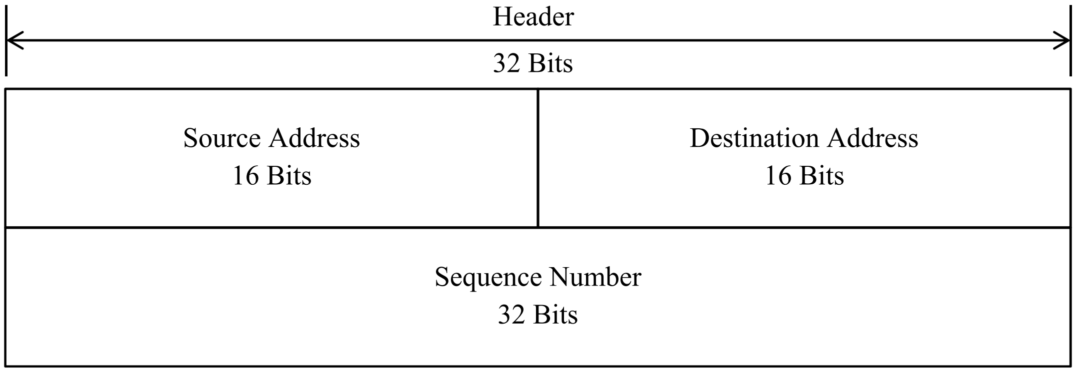


Figure 4.4: The TCP header structure.

wait for acknowledgements. However, this leads to packet loss which cannot be corrected. Furthermore, no flow control methods can be implemented at this layer (using UDP); no information is sent back from receiver nodes. This lightweight protocol offers other advantages (along with low end-to-end delay), due to the lack of reliability, error correction and flow control features, UDP results in low overhead. These advantages make UDP an ideal transport layer protocol for trivial file transport protocol (TFTP), network time protocol (NTP) and VoIP. Figure 4.5 depicts the header structure for UDP in this network model. Significantly, this header only contains the source and destination addresses and these addresses are only 8 bits wide.

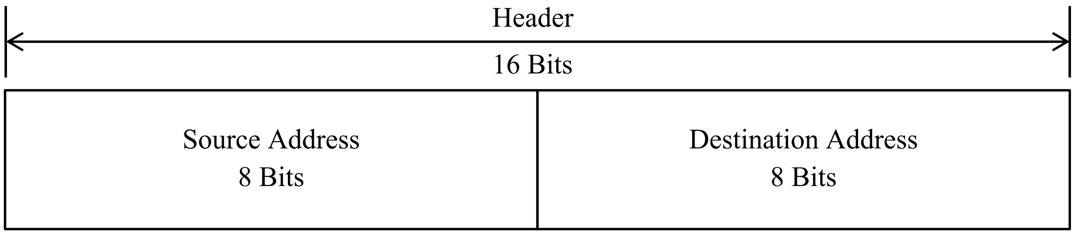


Figure 4.5: The UDP header structure.

4.3.3 Optimised Link State Routing

The network layer deals with the effective addressing and routing of messages. The performance of WMNs is often defined by the effectiveness of their routing protocols [89]. Routing is the selection of the optimal path a message should take. The optimality of the path is measured according to the delay between transmitting and receiving the message. A path consists of a set of links formed between nodes; each link could possibly constitute a different delay compared to the other links. Excepted transmission time, round trip time and hop count are examples of the metrics used for delays. Other measures of effectiveness include the time it takes for the protocol to adapt to changes and how effective resources can be assigned. Hop count is the most commonly used metric as it is the simplest metric. A hop count is the measure of links between nodes. Hop count does not take into account the time it takes for a message to travel across a link. This simplification leads to a metric that is less accurate compared to others, but ensures lower overhead due to smaller routing tables.

Routing functionality is generally executed by routers only. However, WMNs require nodes to be self-organising, therefore nodes are required to perform these routing functions themselves. Routing protocols can be classified into three types: reactive, proactive and hybrid protocols, depending on how the protocol adapts to changes. Protocols that adapt to topology changes when a change is detected, are known as reactive protocols. This type of protocol requires the nodes to broadcast messages when existing links have been broken or when new links are formed. These protocols lead to lower overhead compared to the other types, but also react more slowly. Proactive protocols require nodes to broadcast messages periodically, leading to higher overhead compared to the reactive protocols, but can respond to changes quickly [34]. The hybrid types use features from both proactive and reactive protocols, leading to the highest overhead of all types and quickest response time.

Routing protocols can be further classified according to how the protocol achieves network discovery and how the construction of routing tables is implemented. In this classification, there are two protocols, namely distance vector routing and link state routing. Distance vector routing requires each node to construct and maintain a distance vector table. This table describes the path from the node to all other destination nodes and is broadcast to all neighbouring nodes. This type of protocol leads to easy implementations but requires large overhead (distance vector tables become large), especially as the nodes in the network increase. Routing information protocol (RIP) is the most commonly used distance vector routing protocol. Link state routing requires nodes to construct topology tables, normally containing only the next node, destination node and distance of the path. Shortest path first (SPF) algorithms reduce topology tables to routing tables, which are broadcast to neighbouring nodes. These protocols are slightly more complex and lead to less overhead. Ad hoc on demand distance vector (AODV) and OLSR are two of the most commonly used link state routing protocols [34, 90].

OLSR is a proactive routing protocol that makes use of MPRs for overhead reduction. This reduction is achieved by only allowing MPRs to broadcast routing table information. OLSR is specifically designed for networks where topology changes occur regularly. OLSR was chosen as routing protocol in this network model and is described in more detail in the following section.

4.4 PHYSICAL LAYER

4.4.1 Physical Structure and IEEE 802.11

Typical physical layer wireless communication systems consist of channel encoders (FEC), constellation modulators, and high frequency mixing at the transmitter design. The receiver design typically consists of down converting, synchronisation, channel estimation, constellation demodulation, and channel decoding. More advanced features such as CDMA and OFDM are also included in standards such as IEEE 802.11. Standards define the frequency band at which these communication devices have to operate. IEEE 802.11 provides two possible frequencies for operation, namely 2.4 and 5 GHz.

IEEE 802.11 devices traditionally operate at 2.4 GHz, corresponding to the ISM band. Regulator bodies such as ITU define the maximum allowed transmission power that a communication device can utilise at a specific band. For the ISM band this maximum is set to 100 mW. Assuming free space propagation loss, the actual communication range can be defined in terms of Friis law, as follows:

$$P_{RX}(d) = P_{TX}\frac{1}{4\pi d^2}A_{RX}, \tag{4.2}$$

where $P_{RX}(d)$ is the power of the receiving node as a function of the distance; P_{TX} is the power of the transmit node; A_{RX} is the area of the antenna on the receiving node, and d is the distance between the nodes. The distance can be calculated as 140 m, when it is assumed that the minimum receiver sensitivity is -64 dBm. Figure 4.6 depicts a simplified transmitter and receiver structure. These physical layer structures are normally implemented on field programmable gate arrays (FPGAs) or other application specific integrated circuit (ASIC) platforms. The IEEE 802.11 standard requires FEC

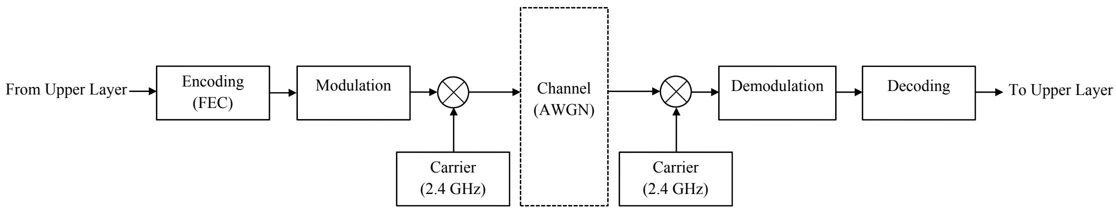


Figure 4.6: A simplified physical layer structure, filters, synchronisation and channel estimation omitted.

using convolutional encoding (as opposed to block coding such as Reed-Solomon), which is typically decoded using Viterbi decoders. Furthermore, IEEE 802.11 allows channel coding rates of $\frac{1}{3}$, $\frac{1}{2}$, $\frac{3}{4}$ and $\frac{5}{6}$. All quadrature amplitude modulation (QAM) schemes as well as binary phase shift keying (BPSK) are incorporated in the standard structure. The channel in this model is assumed to be AWGN. Table 4.1 summarises other parameters defined by IEEE 802.11.

Table 4.1: Parameters used in the physical layer.

Frequency	2.4 GHz
Bandwidth	20 MHz
Data Rate	4 kbps
Modulation Type	CDMA
Range	140 m
Transmission Power	100 mW

4.4.2 Physical Layer Limitations

The purpose of the physical layer is to provide an interface between the upper layers of a node and the upper layers of other nodes. In a fixed wired network, this interface connects to a channel consisting of copper wires or an optical fibre line. Electrical signals and light waves form the medium in which the information is conveyed in these cases. In a wireless environment, the channel is assumed to be free space and the information is conveyed using electromagnetic waves [83]. The wireless nature of WMNs provides advantages such as mobility and low infrastructure, but also leads to a serious disadvantage, namely the channel in a WMN is a shared resource and nodes often compete against each other.

Shannon's channel capacity theorem defines the constraints found due to the limited nature of the channel resource. The first constraint is the maximum transmission rate that a user could achieve. This constraint can be represented mathematically as follows:

$$Ri_{MAX} = log(1 + \frac{Pi}{\Sigma Pj + N_O}), \tag{4.3}$$

where Ri_{MAX} is the maximum achievable transmission rate for user i; Pi is the transmission power of user i; $\sum Pj$ is the sum of the transmission powers of all the other users (thus j is the subset of users excluding user i), and N_O is Gaussian noise. The second constraint is the sum capacity of the channel, and illustrates that the total throughput cannot exceed the capacity of a point-to-point AWGN channel. Equation

4.4 illustrates this constraint.

$$\Sigma Ri < log(1 + \frac{\Sigma Pi}{N_O}). \tag{4.4}$$

Figure 4.7 illustrates an example of these constraints when considering two users ($i = [1, 2]$). Constraints 1 and 2 can be defined by Equations 4.5 and 4.6, respectively.

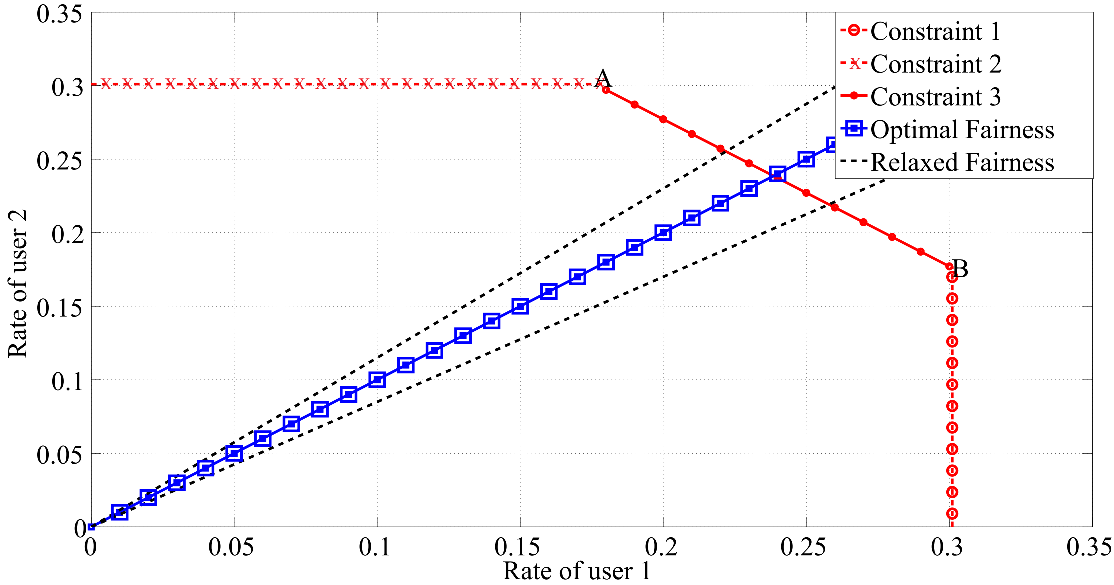


Figure 4.7: An example of the constraints found in a wireless environment when considering two users.

$$R_1 < log(1 + \frac{P_1}{P_2 + N_O}), \tag{4.5}$$

$$R_2 < log(1 + \frac{P_2}{P_1 + N_O}). \tag{4.6}$$

These equations relate to the maximum transmission rate, as defined in Equation 4.3. The line connecting points A and B (constraint 3) in Figure 4.7 can be represented mathematically by the following equation:

$$R_1 + R_2 < log(1 + \frac{P_1 + P_2}{N_O}). \tag{4.7}$$

This equation relates to the sum capacity equation illustrated in Equation 4.4. The symmetrical capacity can be defined as the case where the rate of the first user is equal to the rate of the second user. This situation results in optimal fairness as illustrated by the 45° line in Figure 4.7. The relaxed fairness bounds illustrate the deviation from exact fairness that would be tolerated by all users in the system. Equation 4.8 illustrates this tolerance:

$$|R_1 - R_2| < \Delta f, \tag{4.8}$$

Here Δf represents the arbitrary amount of unfairness all users would allow.

4.5 CONCLUDING REMARKS

This chapter described the model of the communication device. The model closely follows the seven-layer OSI model consisting of an application, presentation, session, transport, network, data link and physical layer. The application layer consists of two protocols, namely, FTP and VoIP. The protocols handle different types of traffic; data traffic is normally used in conjunction with the FTP protocol, where VoIP is typically used for voice traffic. The network layer is responsible for the effective routing of information from destination to source. The physical layer in its simplest form consists of an encoder, modulator, channel, demodulator and decoder. The physical layer was illustrated in terms of the IEEE 802.11 standard. Chapter 5 discusses the routing protocol.

CHAPTER 5

ROUTING PROTOCOL

5.1 INTRODUCTION

The OLSR protocol was selected as routing protocol because of its decentralised and fast adaptive nature (See Table 5.1). This is a proactive link state routing protocol and makes use of the Dijkstra algorithm for finding the shortest path [34]. The algorithm requires each node to broadcast a discovery packet. These packets contain little information (as only the source field is used). The discovery of neighbouring nodes is the primary goal of these packets; each node that receives a discovery message will know that it has a neighbour on at least a unidirectional link. A unidirectional link is a link that can only be used in one direction. A node connected to a unidirectional link can only receive or transmit on that link, but not both. A bidirectional link is a link used for both receiving and transmitting messages.

The protocol continues with all nodes building a table containing all their neighbouring nodes (referred to as a one-hop table). This table is then sent to all neighbouring nodes by broadcasting the message. Calculating if a bidirectional link exists can be achieved by requiring each node to compare the newly received tables with its own one-hop table. If the nodes find, themselves within their neighbour's table and they find their neighbour within their own table, a bidirectional link must exist. Each node can also build a two-hop table using the one-hop table received from its neighbours. This table can be used when a node would like to locate the neighbour of a neighbouring node.

Table 5.1: Features of the OLSR protocol and their respective locations in this chapter.

Step	Function	Section location and short description
1	One-hop Discovery	Section 5.2.2 - Broadcast one-hop information.
2	One-hop Table	Section 5.2.2 - Create one-hop table.
3	Two-hop Discovery	Section 5.2.2 - Broadcast two-hop information.
4	Two-hop Table	Section 5.2.2 - Create two-hop table.
5	MPR Calculation	Section 5.3.1 - Select MPR set.
6	MPR Notification	Section 5.3.1 - Notify the MPRs of their task.
7	Topology Discovery	Section 5.2.3 - Broadcast link state information.
8	Topology Table	Section 5.2.3 - Create topology table.
9	OSPF Calculation	Section 5.3.2 - Compute shortest paths to destinations.

The OLSR protocol provides important overhead reduction features. This is achieved by making use of MPRs, (see Section 5.3.1 for discussion of the functionality of MPRs). The two-hop table is used for calculating which nodes should be selected as MPRs. Once this calculation has been completed, each node nominates a set of nodes as possible MPRs and a notification is sent to each node. Nodes with the highest amount of nominations or nodes nominated as a single nominee become MPRs. Single nominee MPRs occur when nodes only have a single neighbour to select as MPR. The overhead reduction is achieved by only allowing MPRs to broadcast topology control messages [35].

Topology control messages are used for propagating the link state information (also known as the topology information) throughout the network. Once all the nodes in the network receive the complete topology table (which will occur when all MPRs stop broadcasting topology control messages), an open shortest path first (OSPF) algorithm can be applied to the table. These algorithms reduce the topology table by eliminating sub-optimal paths [33]. This leads to the routing table which contain only the shortest path to a specific destination. The OLSR protocol makes use of the Dijkstra algorithm for finding the shortest paths to all destinations.

5.2 OLSR PRIMARY FUNCTIONS

This section describes the most essential functions required for successfully implementing the OLSR protocol. Without these functions, the nodes in the network would not be able to communicate with each other. Moreover, auxiliary functions depend on the successful implementation of primary functions, such as the creation of one- and two-hop tables and the distribution of link state information.

5.2.1 Proactive Nature of OLSR

As mentioned previously, OLSR is a proactive routing protocol. Proactive refers to the periodic flooding of the network. One-hop, two-hop and topology control messages are all included in the periodic flooding feature, and are often referred to as control messages. Proactive protocols are specifically applicable to situations where topology changes occur frequently. These situations include mobile networks as well as dense networks. A network that makes use of proactive protocols provides nodes with immediate access to routing information, in contrast to reactive protocols, which attempt to discover routes only when requested to do so [91]. This feature provides the unique advantage of low end-to-end delay (when actual communication takes place). The disadvantage is that the routing information might be less accurate (compared to reactive protocols) because it might be out of date.

5.2.2 Creating One- and Two-Hop Tables

One-hop tables are used for locating neighbouring nodes. These tables are created from information received from the neighbouring nodes. The source field in the network discovery message is used to identify each neighbouring node. The one-hop table (of a node) contains two sets of information, namely which neighbouring nodes are connected to the node and what type of link exists between these nodes. The latter can only be completed when the node receives two-hop information from its neighbours. However, all nodes can identify if links are unidirectional, by using one-hop information. A further distinction between types of unidirectional links can be made by detecting the allowed direction of communication. In this network model, type 1 refers to a unidirectional link that could be used to send information. A type 2 unidirectional link is a link on which information can be received.

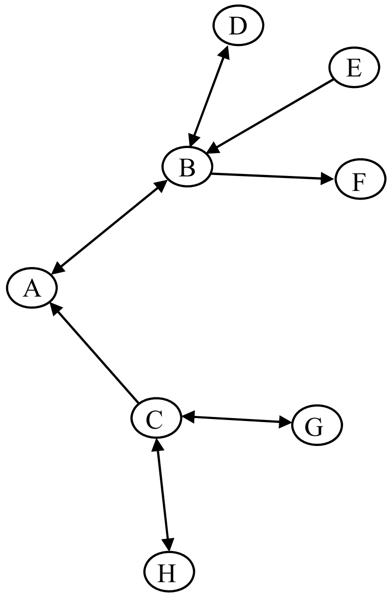
Two-hop tables are created from two-hop information and describe the connections of neighbouring nodes. Two-hop tables contain four sets of information, namely the neighbouring nodes (one-hop information); neighbour nodes of the neighbouring node (two-hop information), one-hop link type, and two-hop link type. Note that by assuming that all links connecting nodes are bidirectional can simplify both one- and two-hop tables. Figure 5.1 (a) depicts an example of a two-hop table constructed (at node A) from an example network, which is illustrated in Figure 5.1 (b). The table contains an entry stating that node A (which is also the node that constructed the table) is a two-hop neighbour. This is not completely true since node A is the node itself. This does, however, indicate that the link connecting these nodes must be bidirectional, as both nodes (node A and B) receive information regarding each other. The table does not contain the same self-information due to the connection formed between node A and C. This is due to the fact that the link connecting them is a unidirectional link (type 2) and node C is unaware of node A.

5.2.3 Link State Information

Topology control messages are used for distributing link state information. The purpose of link state information is to provide nodes with reliable information regarding their surrounding topology. All nodes are required to create their initial topology table, by making use of their two-hop tables [34]. Thus, at the first stage, each node would have a topology table containing destinations that are two hops away. Traditionally, all nodes are required to broadcast their topology table to

One-Hop Node	Two-Hop Node	One-Hop Link	Two-Hop Link
B	D	Bi	Bi
B	E	Bi	Uni-2
B	F	Bi	Uni-1
B	A	Bi	N/A
C	F	Uni-2	Bi
C	F	Uni-2	Bi

(a)



(b)

Figure 5.1: (a) An illustration of a two-hop table. (b) A network associated with the two-hop table.

all surrounding neighbours (although MPRs could be used for distributing this information, leading to lower overhead). Once neighbouring nodes receive the link state information, they update their own topology tables and when a change in the table is detected, the topology table is rebroadcast. This process is repeated until every node has received link state information that does not provide new information, and thus does not change their topology tables. The OLSR protocol provides another feature for overhead reduction; where MPRs could choose to rebroadcast selected link state information [35]. This selection is typically based on whether the link state information contains links between the MPR and the nodes that selected it as MPR. This link state information is referred to as partial link state information.

Figure 5.2 depicts a flow diagram illustrating the process of distributing link state information. Link state information typically contains three sets of information: the destination field, the next hop field, and the hop count. The destination field denotes the final destination to which a packet can be sent. The next hop field

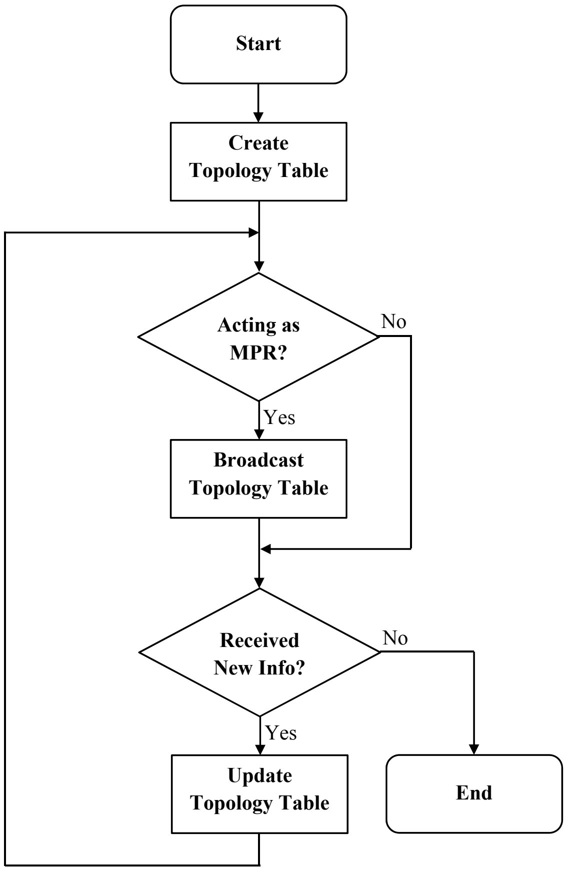


Figure 5.2: A flow diagram illustrating the process of distributing link state information.

indicates which neighbouring node should act as the first relay for the packet. The hop count denotes the amount of relays the packet needs to transverse before reaching its destination. Link state routing provides the advantage of faster network convergence. However, the algorithm is slightly more complex, requiring more memory and more processing compared to distance vector routing [90]. Furthermore, link state routing could lead to less accurate information (especially when partial link state information is used, as is the case in OLSR), which could, in turn, lead to the sub-optimal shortest path selection.

5.3 OLSR AUXILIARY FUNCTIONS

This section describes the auxiliary functions used in the network model for the OLSR protocol. These functions do not perform operations necessary for the network. Therefore these functions can be omitted, resulting in a functioning but sub-optimal network. The auxiliary functions are included as a method for optimising some aspects of the network. For MPRs, the optimisation is the reduction of overhead and the shortest path algorithm is used for minimising end-to-end delay.

5.3.1 Multi-point Relays

One problem regularly encountered in WMNs when considering routing protocols, is the large overhead created in implementing these protocols. Nodes are required to flood the network with control messages, as a method for successfully achieving routing functionality [16]. However, flooding leads to the duplication of information, as can easily be observed when considering two nodes with a mutual neighbour. These nodes might not be aware that they share a neighbour, which leads to the redundant transmission of information. In conventional protocols, flooding is performed without taking this mutuality into consideration. Protocols making use of MPRs can take advantage of this mutuality and effectively minimise redundant transmissions [35].

The OLSR protocol makes use of MPRs in two different ways, both of which result in an improvement compared to other protocols. First, MPRs are the only nodes that rebroadcast topology control messages. Secondly, MPRs are to only required to send messages to the nodes that selected them as MPRs. These requirements lead to nodes only receiving new information, as well as nodes not broadcasting redundant messages [35]. Furthermore, as the number of nodes in the network increases, the total amount of reduction that can be achieved increases. The use of MPRs in the OLSR protocol is applied only to topology control messages, because the complete two-hop neighbourhood information is required for calculating which nodes should be selected as MPRs.

Due to the distributed nature of wireless networks, each node is required to calculate their set of MPRs by themselves. The process starts by requiring every node in the network to select a set of nodes in its one-hop neighbourhood table as MPRs. The selection criterion is based on which neighbouring nodes cover the most two-hop neighbours. It is therefore necessary for each node to build a list of all two-hop neighbours. Nodes are removed from this list when a one-hop neighbour has been selected as an MPR which is a neighbour to these nodes. The one-hop neighbour which covers the most two-hop neighbours is selected first. The selection process ends when all two-hop nodes are removed from the list and results in each node storing a set of MPRs. Special cases that often arise include situations in which a node only has a single neighbour, and this neighbour automatically becomes the MPR for this node (these situations typically occur on the border of the network).

Once MPRs have been selected, it becomes the responsibility of each MPR to notify their neighbours of being nominated as MPR. Each MPR then builds a list of nodes that nominated it as MPR (known as selector nodes), which is used for identifying the nodes that selected it as MPR. MPRs can now decide to only address messages to the selector nodes, which leads to partial link state information as well as a possible reduction in the redundant reception of information. MPRs are reselected every time the proactive protocol performs routing functionality. This ensures that the protocol effectively adapts to changes in the network [33].

5.3.2 Shortest Path Algorithm

The goal of a shortest path algorithm is finding the shortest path between source and destination nodes. The shortest path algorithm can thus be seen as an optimization of the delay found between sending a message from a specific source node to a destination node [91]. The algorithm attempts to minimise these delays. A path is a set of connected links. In addition, as soon as a link is broken, the path is no longer considered valid, and should be removed from the routing table.

Due to the distributed nature of a wireless network, each node in the network is required to calculate their routing table by themselves. Routing tables are created by making use of topology tables and take on the same form as these tables. However, routing tables typically have fewer entries compared to topology tables, as only the entries containing the shortest paths should be listed. Routing tables have three entries, namely the destination, the next hop and the hop count field. Routing tables are calculated regularly; every time a change occurs in the topology table. Unlike topology tables, routing tables are not rebroadcast when they are updated.

The OLSR protocol makes use of the Dijkstra algorithm as a method for reducing the entries in the topology table and compiling the routing table [34]. It should be noted that if each node calculates the routing table by itself, the node should refer to itself as the source node. Table 5.2 summarises the steps used in the algorithm. First, the algorithm adds all one-hop neighbours into the routing table. All these nodes are one hop away from the source and are automatically the optimal choice in terms of end-to-end delay. Then the algorithm requires each node to duplicate the topology table into their routing table. The topology table is regarded as a tree containing all sources, destinations and the paths connecting them. Next, the algorithm inspects each destination and the paths connecting them to the source node. For each destination, the hop count field is inspected and only the path with the lowest number of hops is selected, the other paths are removed. The process is repeated until the routing table is left with only one entry for each destination. Although multiple shortest paths may exist for a single destination and could be left in the routing table, they are removed in an attempt to minimize memory use on each node. These entries are optimal in terms of the end-to-end delay between the source and destination nodes.

Table 5.2: Steps used for finding the shortest path and computing the routing table.

Steps	Description
1	Add one-hop neighbours into routing table as destination nodes.
2	Add topology table into routing table.
3	For each destination, find the shortest path.
4	Remove all entries that have not been selected as optimal.

5.4 OLSR FRAME STRUCTURE

Control messages are periodically distributed in the network when the nodes in the network perform routing functionality. Headers from the upper layer are not necessarily required for routing functionality. Control packets consist of seven general fields, as illustrated in Figure 5.3. The packet length defines the packet size and is normally defined in units of bytes. Although the reliable and in sequence reception of control messages is not necessary, OLSR provides such functionality with the sequence number field. The sequence number is typically used in situations where the information regarding the topology is too large to be represented in a single packet. The message type can take on values from 1 to 4 and denotes the type of information contained with the message. Message type 1 and 2 have been used in this model, where type 1 signifies that the message contains one- or two-hop information. A type 2 message signifies that the message contains topology control information. The Vtime field denotes the time that the information within the packet is valid information, and the time to live field denotes how many times the message may be rebroadcast before it is to be destroyed. The hop count field signifies the number of hops that the message has been through. The final field in the general structure is the source field, which represents the address of the node that sent the message.

It should be noted that the source field might not be the original creator of the message, but rather the relaying node. When the message type is set to 1, the payload consists of six fields, namely the Htime, Link Code, and Message Size fields, as well

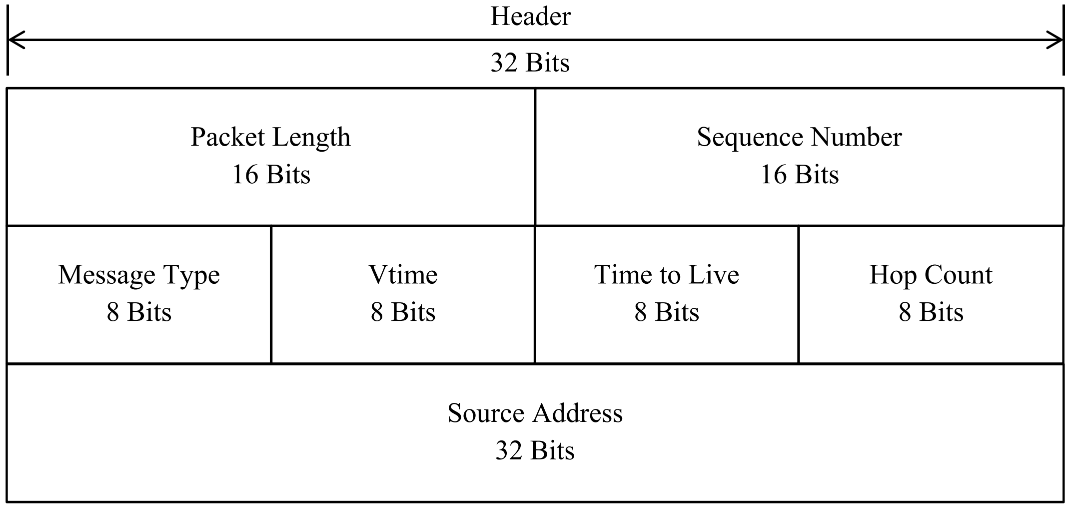


Figure 5.3: The header structure used in the OLSR protocol for control packets.

as three reserved fields. The Htime is used to notify all the other nodes of the period time, therefore this field defines the actual time of the periodic nature of the protocol. In this model, this value was set to 10 seconds. The Link Code defines the type of link existing between neighbouring nodes. The Message Size defines the size of the payload part of the message.

When the message type is set to 2, the payload consists of a varying number of fields. The topology control function requires the packets to contain the link state information and routing tables. These sets of information vary in size. This varying nature requires the message to specify important attributes, thus two fields are always required: the Sequence field and the Message Size field. These fields provide the same functionality as when the message type is set to 1.

Additional important features which require the distribution of vital information to the nodes in the network can be achieved by means of the reserved fields. The OLSR frame structure contains three reserved sections: the first is made up of 16 bits and the others each hold 8 bits. These reserved sections could be used to carry rate and power information. Furthermore, due to the proactive nature of the protocol, rate and power information is regularly provided to all nodes in the network, (see Chapter 6 for discussion).

5.5 CONCLUDING REMARKS

This chapter described the routing protocol used in this network model, namely the OLSR protocol, including its primary functions, its proactive nature, as well as how it creates routing tables and how link state information is distributed. The auxiliary functions, which are used for improving the performance of this protocol was described. Auxiliary functions of the OLSR protocol include MPR which are used to reduce the amount of overhead required for routing. It was shown how the OLSR protocol achieves shortest path routing, as well as how the protocol structures its headers. Finally, it was shown that the frame structure for this protocol contains fields that can be used by other algorithms (such as a rate adaptation heuristic) for passing messages between nodes in a WMN. Chapter 6 discusses rate adaptation in detail.

CHAPTER 6

RATE ADAPTATION

6.1 INTRODUCTION

Rate adaptation refers to the alteration of transmission rate whilst a device is actively communicating with another device. The transmission rate, also known as the data rate of a communication system, is the effective amount of information that can be transmitted at a given time [92]. Data rate is measured in units of bits per second (bps) and is dependent on the bandwidth of the channel. Consequently, higher data rates can be achieved by making use of more bandwidth. However, in 1948 Shannon defined the channel capacity theorem [93], which states that as long as the data rate is within the channel capacity, reliable communication can be achieved.

Throughput refers to the amount of information successfully received, and relates directly to the data rate and channel capacity [83]. In information theory, it is important to remember that a specific probability of error will exist for a specific data rate. The throughput will therefore be at a maximum when the data rate is set to the channel capacity, and the probability of error due to the noisy channel is at a minimum [74]. The probability of error is also referred to as the bit error rate (BER) and is a function of the signal-to-noise ratio (SNR), which in turn, is a function of the transmit power. Equation 6.1 illustrates the relationship between the signal power and the SNR:

$$SNR = \frac{P}{N_O BW} \tag{6.1}$$

where P is the signal power, N_O is Gaussian noise and BW is the bandwidth currently being used.

Although Equation 6.1 is applicable to wireless communication systems, it does not sufficiently capture the performance in multiple communication devices. In a WMN, the channel is a shared medium and nodes (or communication devices) are often selfish entities [66]. A communication device will increase its transmission power when it detects that the probability of error has increased. This is due to the fact that the device believes that increasing the transmission power will lead to an increase in SNR, which in turn leads to a decrease in BER. However, increasing the transmission power will lead to an increase in the interference experienced by other communication devices operating at the same frequency [83]. Equation 6.2 is an extension of Equation 6.1 and captures this interference; this relationship is known as the SINR:

$$SINR = \frac{P_i}{\sum_{\forall j \neq i} P_j + N_O BW} \tag{6.2}$$

where P_i is the signal power of user i.

Rate adaptation is typically achieved at the physical layer, but source coding allows for rate adaptation at the application layer [86]. Source coding is typically used for digitising and compressing an analogue signal. The compression leads to a communication device that operates at lower data rates, while the quality of the analogue single when uncompressed and decoded at the receiver is normally lower [87]. There is another trade-off at the physical layer, namely higher data rates lead to higher probability of error. Constellation modulation techniques, such as phase shift keying (PSK) and QAM, are used for mapping bits into symbols. These modulation techniques define the data rate by specifying the number of bits grouped into a single symbol. BPSK leads to a single bit being a symbol, whilst 4-QAM results in 2 bits for every symbol.

The most conventional method for achieving rate adaptation is by altering the FEC used by the communication device [74]. The simplest example of coding is a repeated code, where each bit is repeated a specific number of times. Repeated codes are sub-optimal since they pack codes in one dimension only. More advanced FEC techniques will pack codes in all available dimensions [92]. The two main types of FECs are block and convolutional types. Block codes make use of a generator matrix for encoding information. The same generator matrix can be used to derive a parity check matrix, which can be used for decoding block codes [94]. Convolutional codes rely on memory for encoding information and each bit depends on a combination of previous bits. Convolutional codes are typically decoded by making use of maximum likelihood decoders (most of which make use of trellis diagrams), such as the Viterbi decoder [95]. All types of FECs lead to coding rates less than equal to 1.

This chapter provides information regarding two requirements. Firstly, a description of the rate adaptation problem in the context of a WMN is required. Secondly, a possible solution for rate adaptation is required. This chapter provides a framework that addresses to these requirements, by describing the problem as a mathematical programme and using a heuristic to solve the problem.

6.2 OBJECTIVE FUNCTION

An objective function describes some real-world metric in terms of a mathematical equation. This function is used in mathematical programming as a method for performing optimisation [60]. The function should describe what a user is interested in, and thus what should be focused on. High throughput is a QoS requirement in several applications, and therefore becomes an important metric for optimisation [9]. Throughput is a function of two variables, namely, the probability of loss, and transmission rate. The probability of loss is a function of the SINR, which in turn is a function of the transmission power of all communication devices that utilise the channel. The transmission rate depends on the FEC and constellation modulation often referred to as the MCS [74]. Although other effects such as mobility, multipath fading and shadowing would also affect the channel capacity, the major influence is interference when considering an AWGN channel. The overall objective is maximising the throughput for all users in a wireless network. However, the throughput depends on the probability of error, therefore it is necessary to consider the probability of error in isolation [11].

6.2.1 Minimal Error Probability

The probability of error is related to the BER. The BER is determined by the specific MCS currently being used. Thus the probability of error should also be determined by the specific MCS [74]. However, the QoS requirements do not typically specify the probability of error as a metric. The packet loss ratio (PLR) is a metric used for quantifying the loss in a network [9]. The PLR is a relationship between the number of packets dropped versus the number of packets successfully received, and is given in Equation 6.3:

$$PLR = \frac{N_{LOSS}}{N_{LOSS} + N_{RX}}, \tag{6.3}$$

where N_{LOSS} is the number of packets lost and N_{RX} is the number of packets successfully received. The expected number of packets dropped relates to the probability of a packet being lost as illustrated in Equation 6.4:

$$E(N_{LOSS}) = Pack_{TOT} \times P_{LOSS}, \tag{6.4}$$

where $Pack_{TOT}$ is the total number of packets and P_{LOSS} is the probability of a packet being lost. The probability of a packet being lost can also be described in terms of the probability that no loss occurred (P_{NoLOSS}) as illustrated in Equation 6.5:

$$P_{LOSS} = 1 - P_{NoLOSS}. \tag{6.5}$$

The probability of a packet not being dropped can be modelled as a binomial sequence, if it is assumed that the FEC is applied before the probability is calculated. Equation 6.6 illustrates this probability:

$$P_{NOLOSS} = \sum_{i=0}^{k} \binom{n}{i} (p_e)^i (1-p_e)^{n-i}, \tag{6.6}$$

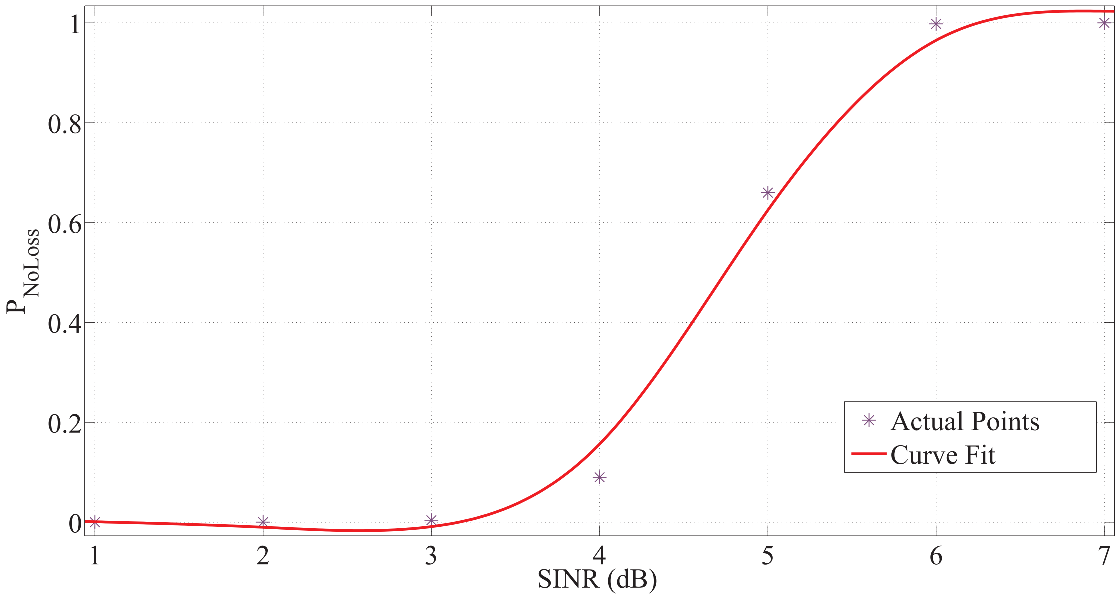


Figure 6.1: The probability of a packet being lost versus the SINR.

where n is the total number of bits in a packet; k is the allowed number of errors in a packet, and p_E is the probability of a bit being in error. If FEC is applied and one or more bits in a packet are incorrect, a CRC would detect this error and force the packet to be dropped and resent at a later stage [74]. If a packet amounts to 512 bits (or 64 bytes) and not a single bit is allowed to be in error, Equation 6.6 becomes

$$P_{NOLOSS} = (1 - p_e)^{512}. \tag{6.7}$$

Figure 6.1 depicts the probability of a packet not being lost versus the SINR, for a WiFi (IEEE 802.11n) communication system, with the MCS equal to 4. This scheme makes use of 16 QAM and an FEC of $\frac{3}{4}$ rate. The actual data point is obtained by using the BER performance curve (which is used to obtain the probability of error) of the IEEE 802.11n communication system in combination with Equation 6.7. Furthermore, it is assumed that there is no fading or shadowing and that the channel follows an AWGN channel.

6.2.2 Maximum Throughput

Apart from the end-to-end delay, the throughput of a communication system may be considered the most important QoS requirement [9]. These requirements are largely service specific, but in regard to data traffic, such as the type of traffic generated by peer-to-peer networks making use of HTTP or FTP, throughput is considered the most important QoS requirement [84]. Services provided by VoIP require communication with low end-to-end delay and high reliability. Services provided by HTTP and FTP do not require low end-to-end delay or high reliability, because when packet loss occurs, packets can be retransmitted without too much loss in QoS for these services.

When considering throughput, it is important to note that this quantity should be measured at the receiver, as the number of successfully received bits can only be determined once channel effects and interference have been applied [92]. The network load is often confused with throughput as both are measured in terms of bps. The network load, however, does not capture the loss of information, but instead the total resources used at that time instance [74].

Throughput refers to the amount of information successfully received (in terms of bytes) per specific time instance. The amount of information successfully received in a given time instance is in turn related to the probability that no error occurred and the rate at which packets can be transmitted [92]. The throughput is defined in mathematical form in Equation 6.8:

$$Throughput = (1 - p_E)^{512} R_D, \tag{6.8}$$

where R_D is the data rate typically measured in bps. Again, the BER versus SINR determines the actual probability of error p_E and data rate R_D of the communication system. Figure 6.2 depicts three probability of error curves used in the IEEE 802.11 standard. This indicates that for a single SINR, multiple probability of errors exist. Furthermore, multiple data rates can be obtained for a single SINR. For instance, when an SINR of 8 dB is selected, a probability of error equal to 0.001263, 0.005342

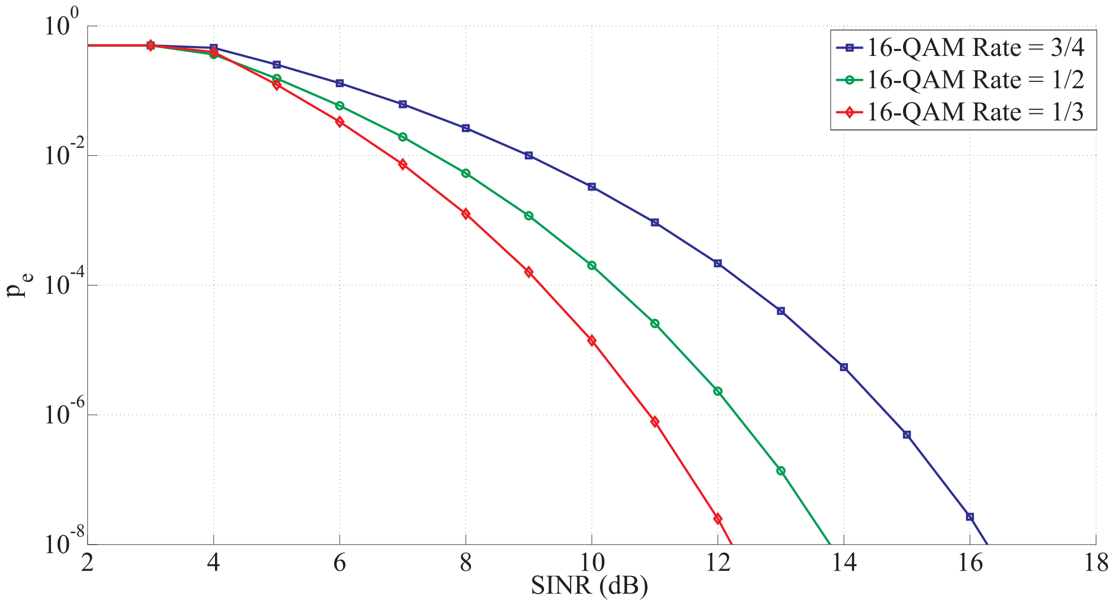


Figure 6.2: The probability of bit error versus the SINR for IEEE 802.11.

or 0.02653 with a rate of $\frac{1}{3}$, $\frac{1}{2}$ or $\frac{3}{4}$ can be obtained, respectively.

6.3 DECISION VARIABLES

The decision variables in an optimisation problem are those parameters that the communication device could alter in an attempt to operate optimally [14]. For communication systems, these values can take many forms; for example, a transmitter could change the number of antennas used when broadcasting a message in an attempt to optimise diversity or even capacity. The frequency band as well as the bandwidth can be altered in order to reduce interference in a specific band, although this would require both the transmitter and receiver to change synchronously. The selected nodes that should perform forwarding tasks in a wireless network may also be considered decision variables.

There are two decision variables in this optimisation problem. First, a variable is required to alter the SINR which would then change the probability of error. The transmission power becomes the first decision variable, as it is the only parameter that

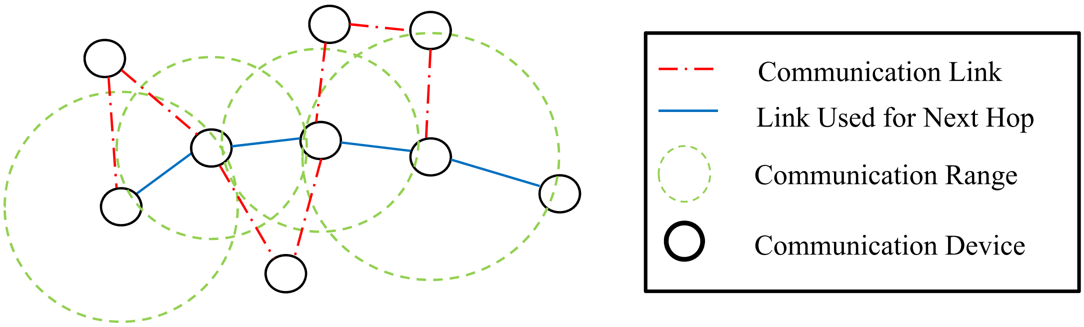


Figure 6.3: A network consisting of nodes that have decreased their power, whilst still successfully forming next hop links between nodes.

can influence the SINR. Although transmission power traditionally focused on the conservation and effective use of battery life, it forms the main focus in current power control problems. Varying transmission power proves challenging task as changing the transmission power of one communication device would affect surrounding devices [92]. This unique problem requires co-ordination between communication devices and in a wireless network this co-ordination should be distributive. Furthermore, it is important to note that varying the transmit power will alter the communication distance (see Chapter 4, Equation 4.2). Therefore, the power cannot be decreased below the threshold that results in a link forming between nodes, which is used for the next hop of a message [83]. Figure 6.3 illustrates a network consisting of nodes that have decreased their power, whilst still successfully forming next hop links between nodes.

The second decision variable is used for altering the transmission rate. Each MCS results in a different rate, as well as a different probability of error. The MCS currently selected is thus a decision variable for the rate adaptation problem. A limited number of MCSs exist for IEEE 802.11 (24 to be exact), making this decision variable an integer variable [31]. In fact, considering that only one MCS can be selected at a time, the decision variable can be simplified to a binary integer variable. This variable consists of a set of MCSs, of which only one can be selected at a time.

Although decision variables are not always mentioned, they are referred to here to clarify which parameters of the communication device can be changed. The decision variables selected for this optimisation problem were selected based on the influence they have on the objective function.

6.4 CONSTRAINTS

The constraints in terms of an optimisation problem are often considered the most vital part of the problem description, as they relate the physical limitations of the real-world to the mathematical model [14]. Several constraints can be defined in a single optimisation problem, each describing a different limitation. The constraints define the extent of change that can be applied to a decision variable [60]. For example, when designing a communication system on platforms such as FPGAs, the number of resources used as well as physical size could be important constraints. The amount of processing time would be a constraint when considering the required minimum delay of reception in a communication system. The number of sub-carriers used in an OFDM system would be a constraint when considering the allowed bandwidth of the channel allocated to the system. The specific amount of information transmitted in a specific time slot in TDMA systems is another example of a constraint found in communication systems. This section deals with the constraints applicable to the decision variables defined for this network model. The channel and physical layer limitations form the main focus of the constraints described in this section.

6.4.1 Channel Constraints

Shannons channel capacity theorem forms the basis of the constraints found due to the wireless channel. In the context of a rate adaptation problem, the channel capacity describes the maximum achievable transmission rate for a specific user [83]. Equation 6.9 illustrates the maximum transmission rate for a single user:

$$R^i_{max} = log(1 + \frac{P_i}{\sum_{\forall j \neq i} P_j + N_O}), \tag{6.9}$$

where, R^i_{max} is the maximum achievable transmission rate for user i. P_i is the transmission power of user i, $\sum_{\forall j \neq i} P_j$ is the sum of the transmission powers of all the other users and N_O is Gaussian noise. It should be noted that the units of Equation 6.9 are in bps only when the base of the logarithm is 2. Other units, such as Hartleys and Nats, can be obtained by using a base of 10 and exponent for the logarithm, respectively [92]. Furthermore, the second term in the logarithm is simply the SINR defined in Equation 6.2. The next constraint is captured in Equation 6.10:

$$\sum_i R_i < log(1 + \frac{\sum_i P_i}{N_O}). \tag{6.10}$$

This constraint captures the maximum capacity of a point-to-point AWGN channel. In this case, the capacity is the sum of the received powers of all the users. Therefore, Equation 6.10 can also be seen as the maximum total throughput achieved when all transmission rates are optimally chosen [83]. Equations 6.11 to 6.13 illustrate constraints 1 to 3 as defined in Chapter 4:

$$R_1 < log(1 + \frac{P_1}{P_2 + N_O}), \tag{6.11}$$

$$R_2 < log(1 + \frac{P_2}{P_1 + N_O}). \tag{6.12}$$

$$R_1 + R_2 < log(1 + \frac{P_1 + P_2}{N_O}). \tag{6.13}$$

The optimal fairness line (45° line) can be seen as the operation point where both users benefit equally by maintaining the same transmission rate. The relaxed fairness can be illustrated mathematically by Equation 6.14:

$$|R_1 - R_2| < \Delta f, \tag{6.14}$$

where Δf represents the allowed deviation from optimal fairness.

6.4.2 Physical Layer Constraints

The physical layer constraints are closely related to the channel constraints, but focus on the implementation of the communication system. There are several limitations in how the communication system is implemented. For instance, communication systems are typically developed by making use of digital systems, which leads to a limitation on how analogue signals can be represented (finite word length effects). This section focuses on the limitations encountered by making use of only a finite set of constellation modulations and FECs.

Typical communication systems make use of only one constellation modulation type. The most fundamental modulation type is BPSK, where all bits are mapped into only two different symbols [92]. These symbols are defined as -1 and 1 when a 0 and 1 are modulated, respectively. Differential BPSK is a more advanced scheme, improving performance by modulating the difference between two consecutive bits. The most popular modulation schemes are the QAM family, which include 4-QAM, 16-QAM and 64-QAM, and are used in several standards such as digital radio mondial (DRM), digital video broadcasting (DVB), WiMAX and WiFi [83]. 4-QAM maps two bits into a single symbol, and it is easy to see how the transmission rate could increase by using 4-QAM compared to BPSK. The disadvantage of choosing 4-QAM over BPSK is that the probability of error increases. Similar trends can be seen when increasing the modulation index further to, say, 64-QAM. Figure 6.4 illustrates the transmitted and received symbols for a 4-QAM communication link. Figure 6.5 depicts the same symbols for a 16-QAM communication link.

The probability of error can be reduced by making use of FEC. Again, typical communication systems only make use of one or two FECs. There are two main types of FECs exist, namely block and convolutional codes. Some communication systems make use of both types simultaneously [95]. The most popular block decoders are Reed-Solomon and low density parity check (LDPC) decoders, whilst the most popular convolutional decoders include the Viterbi and Bahl-Cocke-Jelinek-Raviv (BCJR) decoders [95]. Although FECs lead to lower probability of error by adding redundancy, they result in a decrease of transmission rate. The coding rate R_C defines this loss in rate and is illustrated in Equation 6.15:

$$R_C = \frac{k}{n}, \tag{6.15}$$

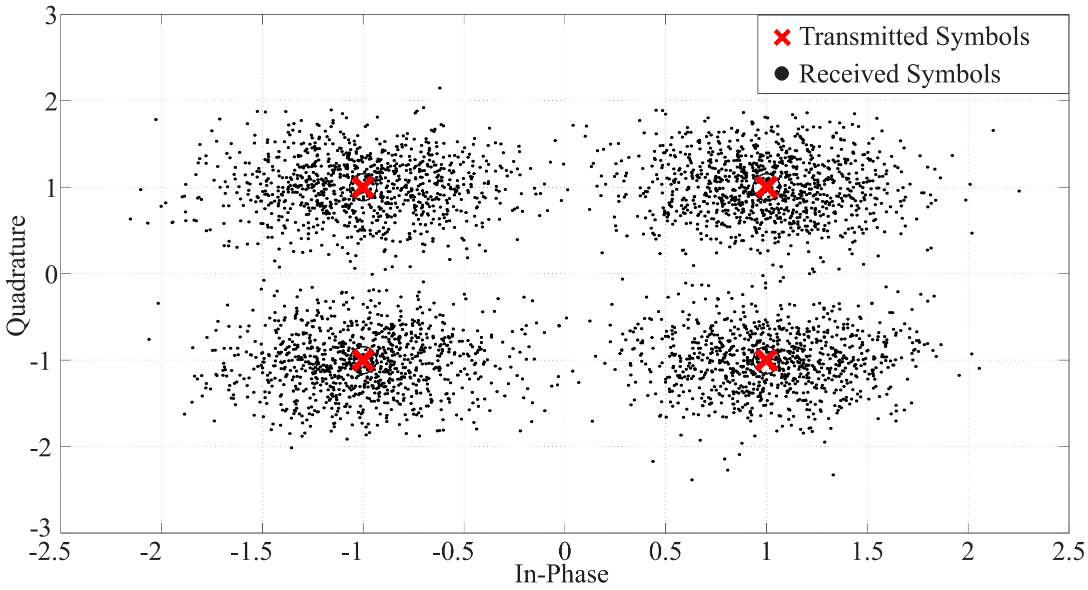


Figure 6.4: Transmitted and received symbols for a 4-QAM communication link.

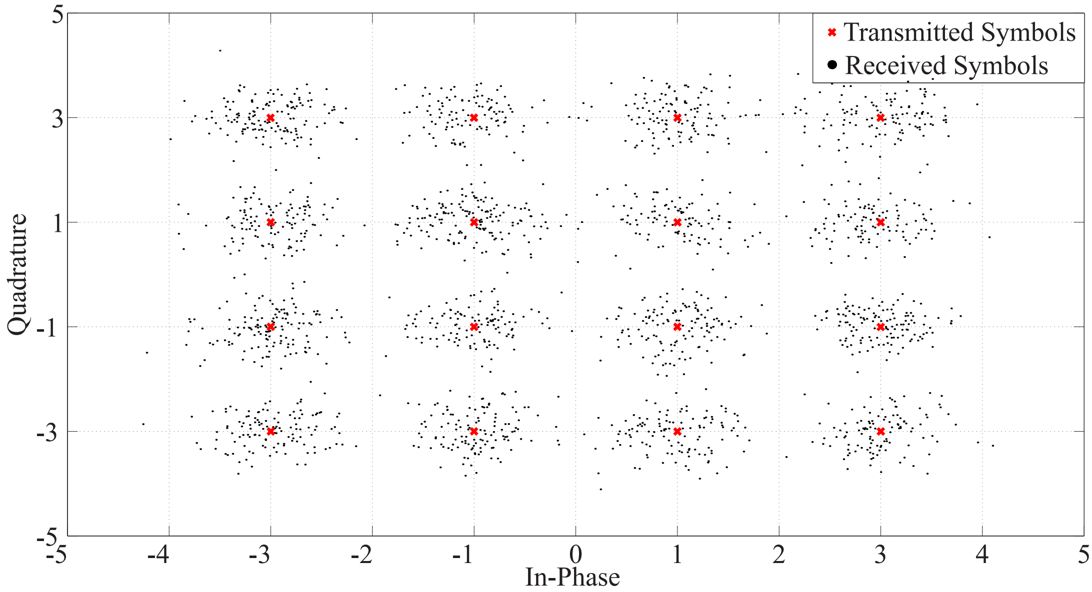


Figure 6.5: Transmitted and received symbols for a 16-QAM communication link.

where k is the number of information bits and n is the total number of transmitted bits. Typical communication systems use FEC at a fixed rate that does not adapt to channel effects.

An almost infinite number of MCSs can be created by varying the modulation type as well as the coding rate of the FEC. However, the IEEE 802.11 standard specifies a set of MCSs to be used in a communication system [31]. The decision variable representing which MCS is selected becomes a binary integer variable. Equations 6.16 and 6.17 describe the constraints placed on this variable:

$$\theta_j \in \{0, 1\}, \tag{6.16}$$

$$\sum_{j=1}^{m} \theta_j = 1, \tag{6.17}$$

where θ_j is a binary integer variable (limited to 0 or 1) indicating which one of the m MCSs has been selected. Equations 6.16 and 6.17 represent the MCS selection as an integer selection where only one scheme can be selected at a time.

6.5 MATHEMATICAL PROGRAMMING

Mathematical programming is used in operational research to find the optimal point of operation [13]. This is achieved by finding values for the decision variables that would result in a maximum or minimum of the objective function, whilst remaining within the constraints set forth by the system limitations. Problems solved by using mathematical programming include finding the maximum profit that could be obtained in a steel milling factory, taking into account the cost of the product as well as the cost of labour. Another example could be finding the minimum required resources for the successful implementation of a digital radio, by considering the required complexity, and number of adders and multipliers [13]. Mathematical programming requires a model describing three properties, namely the objective function, the decision variables, and the constraints.

6.5.1 A Non-Linear Mixed Integer Programme

The simplest form of mathematical programming is LP, but specific requirements have to be met before LP techniques can be applied [60]. In LP the objective function and constraints are linear functions, and once these requirements are met, the simplex algorithm can be applied to solve the problem. However, from Equations 6.8, 6.9 and 6.10 it is clear that these requirements are not met for the rate adaptation problem, and the problem now becomes an NLP problem. NLP problems are more difficult to solve, although in some special cases, when the objective function and constraints are a product of the decision variables, quadratic programming can be used to solve the problem [14]. Although the objective function might meet this requirement, the constraints in Equations 6.9 and 6.10 are not products of the decisions variables.

Equations 6.16 and 6.17 further complicate the problem by requiring certain decision variables to take on only integer values. In several cases, integer programmes lead to NP-hard problems, although in some cases they can be solved by relaxing the constraints and making use of column generation [14]. Equations 6.9 and 6.10 are not integer, and the combination of non-integer and integer decision variables changes the mathematical programme from an integer to a mixed integer programme. Moreover, due to the non-linear nature of the objective function and constraints, the problem can be considered a non-linear mixed integer programme. The problem can now be formulated mathematically as follows:

$$max. \quad Z = \sum_{k=1}^{n}(1 - \sum_{j=1}^{m} \theta_j p_e^{(k,j)}) R_k, \tag{6.18}$$

$$s.t. \quad R_k \leq log(1 + \frac{P_k}{\sum_{\forall l \neq 1}^{n} P_l + N_O}), \tag{6.19}$$

$$\sum_{k=1}^{n} R_k \leq log(1 + \frac{\sum_{k=1}^{n} P_k}{N_O}), \tag{6.20}$$

$$\theta_j \in \{0, 1\}, \tag{6.21}$$

$$\sum_{j=1}^{m} \theta_j = 1, \tag{6.22}$$

where $p_e^{(k,j)}$ is the probability of error for the k^{th} user when making use of the j^{th} MCS, θ_j is the integer variable, indicating which of the m MCSs has been selected. R_k is the effective rate of the k^{th} user. The probability of error $p_e^{(k,j)}$ is a function of the MCS and the SINR as experienced by the current user. Equation 6.23 captures this relationship:

$$p_e^{(k,j)} = f(Modulation, FEC, SINR), \tag{6.23}$$

The fact that the problem is a non-linear mixed integer problem, and that the exact function for $p_e^{(k,j)}$ is difficult to obtain, makes finding the exact solution to the problem challenging. A heuristic has been developed as a method for finding a fast and near-optimal solution to this problem.

6.5.2 Limitations

In this model the channel was assumed to be a channel with AWGN and no additional effects, such as multipath fading or shadowing, were incorporated. The constraints defined by Shannon's channel capacity theorems as given in Equations 6.9 and 6.10, would require alteration when the channel considered is no longer a point-to-point AWGN channel [83]. The model also assumes that the path selected for a set of packets (representing a message) remains fixed until at least a new set of packets is generated. Additional congestion avoidance techniques could be used to spread packets along multiple, equally distanced paths (if they existed) [1]. The model also assumes that the device will be operating by making use of only one frequency, although in certain cases, IEEE 802.11 allows communication on two separate frequencies (2.4 GHz and 5 GHz).

6.6 RATE HEURISTIC

A heuristic is an algorithm that attempts to find a specific operation point, based on the information it has gathered in the past [14]. Heuristics are used in situations where real-world problems cannot be solved easily by means of a mathematical technique and are thus considered NP-hard or NP-complete problems. The ant colony heuristic is an example of a heuristic that could be applied to problems such as the travelling salesman problem, which is known to be an NP-hard problem [13]. This heuristic was inspired by the way a colony of ants cooperates to achieve near-optimal operation. Ants would make a decision and notify other ants of the result of their decision (which is effectively information about the past) by making use of pheromones [96].

The heuristic described in this section attempts to find a near-optimal solution to the rate adaptation problem. Near-optimal in this case is seen as a value that falls within 20% of the optimal. The heuristic makes use of the network layer routing protocol for distributing information regarding rate adaptation among the nodes in the network. The messages include information on the current rate of the other nodes, as well as their power levels. This information is used to determine how far from the constraints the nodes are operating. The nodes can alter their decision variables in an attempt to find the highest rate whilst still keeping within the constraints.

The optimisation heuristic finds a feasible rate region that conforms to the set of constraints described in Section 6.5. The method, starts by setting the rate of the current user to its maximum (see Table 6.1). It then estimates the interference induced to other users using information regarding the power sent back from neighbouring users. If this interference causes a loss due to exceeding channel capacity, the method continues decreasing its rate and returns to Step 2. If there is no loss, the method goes on to estimate how effective the channel capacity has been used. When the channel capacity has not been used efficiently, the user increases its rate and returns to Step 2. If the maximum rate has been obtained according to the constraints, the method ends as it is now known that a near-optimal (or saddle) point has been found. It should be noted that this optimisation problem does not capture the fairness requirement. The fairness requirement is captured by making use of game theory (see Chapter 7 for details).

Table 6.1: Steps in the optimisation heuristic.

Step	**Description**
1	Set rate to maximum obtainable rate.
2	Estimate the amount of interference caused by selecting current rate (using information sent back from neighbouring nodes).
3	If channel capacity has been exceeded, decrease rate and return to step 2, otherwise continue.
4	Estimate how efficiently the channel capacity has been used, by again using information from neighbouring nodes.
5	If the channel capacity has been used optimally, the algorithm ends. Otherwise, increase the rate and return to step 2.

6.7 CONCLUDING REMARKS

This chapter described the rate adaptation problem as well as the limitations and important properties of a communication system, with regard to rate adaptation. A mathematical programming model was constructed making use of the objective function, decision variables and constraints identified in the communication system. The limitations of this mathematical model were discussed, focusing on the wireless channel and the IEEE 802.11 physical layer implementation. The rate adaptation problem was found to be a non-linear mixed integer problem. This along with the fact that the exact function for the error probability as a function of the modulation, FEC and SINR for each user is difficult to obtain, makes finding the exact solution to the problem challenging. Therefore, a heuristic was described that makes use of past information regarding the rates and powers of other nodes in the network. The heuristic quickly finds a near-optimal solution to the rate adaptation problem. Finally, it was noted that the heuristic does not capture the fair distribution of rates among all nodes in the network. Chapter 7 describes and applies game theory to find the fair operating point in terms of transmission rate.

CHAPTER 7

RATE FAIRNESS

7.1 INTRODUCTION

Game theory may be regarded as the multiple user equivalent of decision theory. Decision theory deals with a single user that attempts to maximise some utility [13]. In decision theory, a user attempts to find the optimal strategy in a specific situation with specific limitations. An optimal strategy is considered the objective function and the limitations are viewed as the constraints, which is similar to the mathematical model described in Chapter 6. Game theory differs from decision theory in the number of users participating in the decision process. A more important factor to consider, however, is the interactions between the users. Hence game theory may be regarded as a set of mathematical tools used for modelling the interactions between multiple users, and the interactions as the effect that one user's choice has on the other users [6].

Game theory is used in situations where users are forced to compete for a limited resource. This competitive environment leads to users acting as non-cooperative players, which in turn leads to sub-optimal performance across the network [7]. In game theory, non-cooperative users are modelled on the assumption of rationality. Assuming that a user acts rationally, thus presumes that the user will attempt to optimise its utility (minimising cost or maximising profit). When a communication device is considered as a user, rationality is justified by noting that devices are typically programmed to follow specifications and standards. The Nash equilibrium is the state at which no user can improve their utility by changing their action [44]. A Nash equilibrium often arises in situations where users act selfishly, a higher utility could be achieved for all users in the network. A Pareto optimal state refers to a state where no user could change their action without lowering the utility of other users [6].

A game theory model consists of three parts: users, actions, and utilities [7]. Users are the entities making decisions in an effort to optimise their utility (in the assumption that they are rational users.). Actions are the choice users make, thus the reduction in power level refers to an action by a certain user. These actions can be observed by other users, and would most likely affect their own actions. The utility of a user is the parameter of performance that a user would like to optimise, hence the utility function is not unlike the objective function (see Chapter 6, Section 6.2).

Game theory has been used to model several communication networks, ranging from packet forwarding to bandwidth auctioning. This has resulted in situations modelled as static non-cooperative games to dynamic cooperative games. This chapter deals with the application of game theory to the rate adaptation problem as a means of finding a more fair solution to rate adaptation. The model does not replace the heuristic described previously; but only forms an extension of the heuristic. The rate adaptation problem is modelled as a repeated non-cooperative game with an infinite horizon.

7.2 FAIRNESS

The aim of developing a game theory model for the rate adaptation problem is to find a fair solution. Examining the mathematical model, however, necessitates defining what is meant by a fair solution. Fairness has been a topic of interest in operational research for some time, resulting in several metrics for fairness being defined. Fairness is a difficult metric to define as it is based on the preference of the users [97]. One user might prefer higher throughput to lower throughput, even at the cost of added latency. Furthermore, a user could consider an equal distribution of throughput a fair distribution. Another user might argue that this is not fair and a more fair solution should be based on the demand of throughput. The three most popular measures of fairness, namely, the min-max fairness, Gini fairness index and Jains fairness index are briefly described next.

7.2.1 Min-Max Fairness Metric

Although the min-max fairness index is based on the notion of equality, the implementation of the metric does not always lead to an equal distribution of resources. Furthermore, the min-max fairness index is considered least complex method for measuring fairness [98]. The index consists of the minimum resource allocated to a user and the maximum resource allocated to that same user. Equation 7.1 illustrates this metric:

$$FI_{min-max} = ^{min}_{\forall r} \, ^{max}_{\forall r} \, (R_D^i), \tag{7.1}$$

where R_D^i is the amount of resources (in terms of data rate) allocated to user i. This index simply finds the minimum resource allocation between a set of maximum resource allocations for each user. Although this metric leads to a simple solution, it often results in leads to sub-optimal resource allocation [98].

7.2.2 Gini Fairness Metric

The Gini fairness index was originally applied in the field of economics, but has recently been extended to telecommunication and transportation networks [99]. As a measure of variability, the metric describes the distribution gap between resources allocated to specific users. Alternatively, the Gini fairness index may be said to describe the difference between the optimally fair distribution of resources and the actual distribution. The metric can be illustrated mathematically as in Equation 7.2:

$$FI_{Gini} = \frac{1}{2n^2 E(r)} \sum_i \sum_j |R_D^i - R_D^j|, \tag{7.2}$$

where n is the number of users, $E(r)$ is the expected value of the resource allocations and R_D^j is the amount of resources allocated to user j. Although the Gini fairness index finds application in communication areas, it is considered the least popular method of quantifying fairness [99].

7.2.3 Jain's Fairness Metric

Although many measures of fairness exists, several are considered to be too application-specific and inapplicable to more unique problems. Jain's fairness index provides a quantitative measure of fairness that is applicable to almost any resource allocation problem [100]. This is the most commonly used metric in the field of telecommunications. Equation 7.3 defines the index of fairness based on Jain's metric:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2}, \tag{7.3}$$

This equation results in an index that is bounded and scale independent. Due to the fact that it is most commonly used, this index provides a method of comparison. Furthermore, the index increases as the distribution of resources becomes fairer, making it monotonically increasing. Jain's fairness index was selected as fairness index for this model, because of these properties. It should be noted that an allocation is considered perfectly fair when Jain's fairness metric equates to 1.

7.3 GAME MODEL

Game theory has traditionally been applied in the field of economics, politics and sociology to model individuals' interactions. More recently, game theory has found significant application in the field of telecommunication. Here, users become communication devices and can be programmed according to standards and specifications [7]. Although this might appear to be a restriction, it is part of the justification of rationality. In the field of economics, politics or sociology, rationality is difficult to defend because individuals do not always act rationally. However, in the field of telecommunication, communication devices are assumed to act as standards and specifications require them to act. This section provides the necessary background regarding game theory as applied to this model.

7.3.1 Static Non-cooperative Game

The simplest form of game theory is a static game, because each user is only allowed a single action. These games imply that a user has no information regarding the actions taken by other users, and thus the user has no incentive to change his action [6]. Although this form of game theory could be considered simplistic, it provides a useful result; it describes the interaction of users based on what they assume the actions of the other users would be. Although a user might not know exactly what actions another users has taken, based on the assumption of rationality, the user could assume what action they would take [13]. Dynamic games are games where each user might change his action (at each stage of the game) based on the information gather from past actions. An example could be the game of chess, where users take turns in making moves (taking action) based on the past actions observed.

Non-cooperative game theory is the most commonly found type of game theory in the field of telecommunication [7]. These games are applied in situations where users typically compete for resources, due to some constraint placed on them (such as the channel capacity). Non-cooperative situations typically lead to nodes acting selfishly, which in turn leads to sub-optimal performance, system wide. Cooperative game theory, on the other hand, deals with situations where users agree to a set of rules. Traditionally, this state of cooperation requires an incentive for users to follow the rules, which is typically achieved by enforcing these rules (either by punishment or remuneration). Non-cooperative games can converge to a cooperative state, when enforcement is properly implemented, but this state of cooperation should be self-sustaining.

7.3.2 Repeated Game

Repeated games form an important part of game theory as they can be used in a non-cooperative environment and still lead to cooperation between users [6]. Repeated games are static non-cooperative games that are repeated over time. Every time a static game is repeated, it instigates a new stage of the game. When users receive information about other users' actions and utilities of at each stage, the repeated game is considered a dynamic game. When users receive information about all the actions and utilities of all other users, the information is regarded as complete. Complete information is not always realistic and often leads to higher complexity. Nevertheless, repeated game theory models can still be applied in situations where complete information is infeasible. Games with incomplete information are used in situations where users have partial information about others' actions and utilities. This information is gathered in a history set for each user, which could be used at each stage as a method of determining the next action for a user [7].

Repeated games have been applied to several problems in telecommunications [45, 51]. In routing problems, nodes traditionally select paths based on the shortest distance between source and destination [58]. When multiple nodes select the same path, however, congestion is induced. By modelling this situation as a repeated game, congestion can be reduced by informing all the nodes in the network of actions taken by the others. Again, providing complete information by notifying all the nodes would lead to high overhead and increased complexity. Incomplete information can likewise be used with similar results by only notifying nodes in a specific local area. The rate adaptation problem can also be modelled as a repeated game and users receive information about other users' the power level and current transmission rate.

7.3.3 Infinite Horizon

A non-cooperative environment could lead to cooperation between nodes in a network, if there is an incentive for cooperation [7]. In this section punishment of selfish behaviour is described the form of enforcement and leads to an incentive for cooperation. A repeated game with an infinite horizon is a game that would seem to continue indefinitely [6]. Although in a real situation no application would continue forever, the infinite horizon condition only requires the users in the network to believe that the game could continue forever. In other words, each user is required to act as if the following stage in the game is not the last stage [37]. This requirement leads to users not deviating from cooperation in the last round, and thus not being able to elude punishment.

The Folk theorem provides a method for ensuring cooperation between users in an infinite horizon repeated game. Cooperation is achieved when the loss in utility due to the punishment that a user receives outweighs the gain in terms of utility that the user receives. The loss or punishment received due to selfish behaviour relates to being banished from the network for a period. In this period the user will have no communication capabilities and its utility drops to zero.

The banishment is achieved by notifying all the other users in the network of the user that is to be banished. These notification messages are broadcast by making use of the reserved fields of the OLSR control packets (see Chapter 5, Section 5.4). The OLSR protocol was selected for this purpose as it provides overhead reduction by making use of MPR. The instances at which users receive notification messages are regarded as the stages in the game, and if the OLSR protocol has converged, these instances occur at the same time for all users. When users receive notification messages, they remove banished nodes as source and destination nodes from their routing tables. The period of banishment can be varied, but is set to a constant of four stages in this model.

7.4 RATE FAIRNESS MODEL

This section covers the mathematical background required in the rate fairness model; describes the game in strategic form; defines the utility, actions and stage gain in mathematical form. The section then continues by describing the fairness process by means of a flow diagram.

7.4.1 Game Theory Analysis

A repeated game Γ^r in strategic form can be described as illustrated in Equation 7.4:

$$\Gamma^r(\delta) = (N, \mathbf{A}, \mathbf{U}), \tag{7.4}$$

where δ represents the discount factor, N represents the number of users, $\mathbf{A}$ represents the action set and $\mathbf{U}$ represents the utility set. In this case the repeated game is regarded as a coordination game, because the utility function for each user is the same [7]. The utility set is illustrated by Equation 7.5:

$$\mathbf{U} = u_i, \tag{7.5}$$

where u_i is the utility function of user i. The utility function for an infinite horizon repeated game is defined as follows:

$$u_i = (1-\delta)\sum_{k=0}^{\infty}(\delta)^k g_i(a^k), \tag{7.6}$$

where k is the number of the current stage game and $g_i(a^k)$ is the kth stage utility or payoff for user i as a function of the action profile a^k. Here a homogeneous action set is chosen for all users which consist of increasing the rate ($\Uparrow$), decreasing the rate ($\Downarrow$) or keeping the rate the same ($\equiv$), as illustrated in Equation 7.7:

$$\mathbf{A} = A_i = \{\Uparrow, \Downarrow, \equiv\}. \tag{7.7}$$

The action profile a^k can now be described as the action taken by all users in the kth stage. The kth stage utility $g_i(a^k)$ now becomes a function of the rate as illustrated in Equation 7.8:

$$g_i(R_D^i) = (1-p_e)^{512} R_D^i. \tag{7.8}$$

It should be note that Equation 7.8 leads to the same throughput as described in objective function (see Chapter 6, Section 6.2). This is based on the assumption that a rational user would prefer a higher throughput compared to a lower throughput, or effectively, a rational user would attempt to maximise their throughput. The probability of error p_e in Equation 8 captures the loss due to the MCS selection. Here it is assumed that each user would prefer a lower probability of error compared to a higher probability of error.

Equation 7.8 describes an implicit conflict of interest between multiple users in a wireless network. This is due to the fact that as one user increases its rate, it would result in an increase in the utility function for that user, but most likely a decrease for all the other users (see Chapter 6). The next section describes the implementation of the rate fairness technique. This technique makes use of the Folk theorem which allows the stage utility to drop to zero when users are uncooperative, and results in a fairer rate allocation [6].

7.4.2 Fairness Model

The fairness model is added to this rate adaptation heuristic as a method for finding a fairer distribution of transmission rate in the network. The fairness model does not deal with finding the optimal rate, but instead redistributes resources fairly whilst still remaining close to the optimal point.

Figure 7.1 depicts the fairness process, which commences by the users broadcasting their own transmission rate information by making use of the structure provided by the OLSR protocol. Neighbouring users receive this information (referred to as Gamma) and continue by calculating an average rate. The rate is calculated from the transmission power and the MCS. The maximum attainable rate can be described by Shannon's channel capacity theorem, which is dependent on the transmission power of all users (a decision variable used in the rate heuristic). The MCS also affects the rate by varying the number of bits per symbol, as well as the error correction code. The MCS also forms a decision variable used in the rate heuristic. If each node keeps track of these decision variables, they can be used in combination with the transmission power of other nodes (received in messages sent back from neighbouring nodes as described in step 2 of the rate heuristic) to estimate the effective transmission rate.

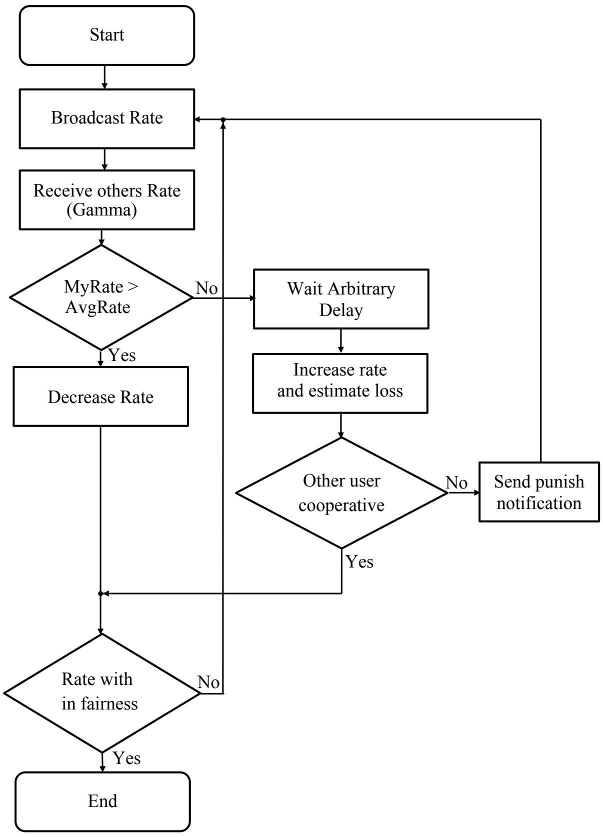


Figure 7.1: Flow diagram describing the second section of optimisation using game theory.

The process continues by comparing the user rate to the average rate and if the user rate is greater than the average rate, the user is forced to decrease its rate. If the resulting rate is within the fairness bound, the process ends for this user, otherwise it re-broadcasts its rate information. If the user rate is not greater than the average rate, the user is forced to wait an arbitrary short time. The user then estimates the loss that would occur if it increases its rate, using Equation 7.8. The user continues determining if the other users were cooperative. If the other users were cooperative, it continues by checking that the resulting rate is within the fairness bound and the process ends for this user. Otherwise it re-broadcasts its rate information. If the other user was not cooperative, a punishment notification is sent, by again making use of the structure provided by the OLSR protocol. Finally, the user re-broadcasts its rate information and restarts the fairness process.

It is assumed that the network consists of at least some cooperative nodes and that they will act rationally. This assumption is justified by considering that nodes act as they are programmed, which in turn, is based on the standards governing the development of these devices [7].

7.5 COMBINED MODEL

This section describes how the rate heuristic and fairness model are combined to provide a solution that is both an optimal and fair solution to the rate adaptation problem. Figure 7.2 describes how these methods are combined and integrated into the network model. Furthermore, the figure illustrates the network model as described in Chapter 4, as well as a flow diagram illustrating how the rate heuristic and fairness model are combined. It should be noted that the rate heuristic and fairness model are implemented at the physical layer. The rate heuristic and fairness model is executed in two phases, the first phase makes use of the rate heuristic, when this phase has been completed, a near optimal rate allocation results. The combined model then starts executing the second phase, which makes use of the fairness model. When this phase ends, a more fair rate allocation results.

The process starts by executing the rate heuristic and thus follows the steps explained in Table 6.1. The heuristic makes use of the network layer control signal which delivers information used to estimate the amount of interference, as well as the usage of the channel. The rate heuristic then uses the physical layer control signal to adjust the transmission rate (by adjusting the two decision variables). If the channel capacity has been used optimally, the algorithm ends and the next phase can commence. The next phase of the process starts by executing the fairness model as

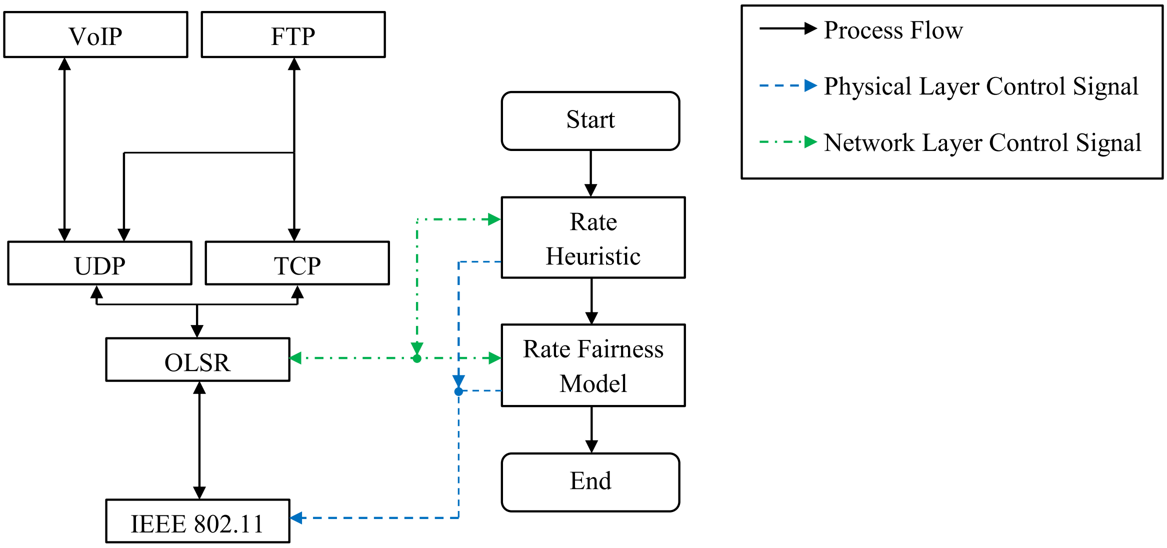


Figure 7.2: Flow diagram describing the combination of the rate heuristic and fairness model alongside the network model.

described in Figure 7.1. The fairness model starts by broadcasting rate information using the network layer control signal. This is followed by making use of the physical layer control signal to adjust the transmission rate. The network layer control signal could be used to notify other users of possible punishment. The phase ends when an acceptable rate (in terms of fairness) has been achieved.

7.6 CONCLUDING REMARKS

This chapter defined the concept required for understanding the fairness model. The chapter described the concept of game theory, where and why it is used as well as the components in a game theoretic model. Fairness is described along with three metrics that could be used. Each metric was defined mathematically and their individual advantages and disadvantages were identified. Jain's fairness index was selected because it can be applied to several resource-allocation problems and used for comparison of the rate adaptation problem. The rate adaptation problem was described in terms of an infinite repeated game, using the concepts of non-cooperative and cooperative games. The mathematical formulation of an infinite repeated game tailored to the rate adaptation problem was provided. Finally, the rate fairness process and how it achieves punishment and information passing was described. Chapter 8 presents and discusses the findings of the study.

CHAPTER 8

RESULTS AND DISCUSSION

8.1 INTRODUCTION

This chapter presents the performance evaluation of the rate adaptation heuristic and rate fairness model, described in Chapters 6 and 7. The network model described in Chapter 4, was used for the evaluation. Open modular network (OMNeT)++ 4.2.2 simulator was selected as the simulation platform and the stack, half diamond, full diamond and random topologies was used as a frame of comparison. Each section deals with a specific topology, with the exception of one section dealing with a comparison of the fixed topologies. Section 8.2 deals with the stack topology, whilst Section 8.3 and 8.4 deal with half diamond and full diamond topologies, respectively. Although these topologies do not represent real-world situations, they offer an ideal structure for comparison, as illustrated in Section 8.5. Section 8.6 deals with the more realistic situation commonly found in WMNs, namely a randomized network. This topology investigates the rate performance as well as the scalability of the rate adaptation heuristic.

It was assumed that each user had been modelled using the layer structure illustrated in Chapter 4. The results are presented according to the type of traffic generated: traffic making use of the VoIP protocol and traffic making use of the FTP protocol. Traffic handled by these protocols differ significantly, due to the fact that the origin of the traffic types differs. Voice traffic, which is associated with VoIP, consists of active speech and silence, whilst data traffic (associated with FTP) consists of a more uniform payload. The different traffic types illustrate how the rate adaptation heuristic performs when subjected to erratic or bursty traffic (voice) and constant or uniform traffic (data). The traffic type in the network is assumed to be homogeneous, thus voice and data traffic are not used simultaneously.

Furthermore, it was assumed that all nodes had an infinite buffer size and thus the only factor contributing to the loss of packets was the interference and effectively the SINR. This assumption isolated the performance of the rate adaptation heuristic, which in turn provided a structure of comparison between topologies. In the fixed topology, it was assumed that only the worst case paths would be selected. This ensured the performance obtained by the rate adaptation heuristic represented the worst case performance, and further techniques such as congestion avoidance could be used to improve the performance. It was assumed that all nodes were IEEE 802.11 communication devices. This ensured that all nodes followed the same standard and consisted of the same physical layer (type of modulation schemes, FEC, number of antennas, antenna gain, frequency band and bandwidth). It was also assumed that all nodes were in the same plane, thus all nodes had the same antenna height. This assumption allowed the use of Friis law for the propagation of electromagnetic waves in free space. Finally, the channel was assumed to be an AWGN channel.

The performance of the rate adaptation heuristic was measured in terms of throughput, end-to-end delay and rate fairness. These terms have been defined in Chapters 6 and 7 and have been selected due to the QoS requirement placed on wireless communication devices in a WMN for voice and data traffic.

8.2 STACK TOPOLOGY RESULTS

This section discusses the performance analysis of the rate adaptation heuristic combined with the fairness model for the stack topology. The results are divided into two sections; firstly the rate adaptation model was applied to a network making use of data traffic, followed by a network making use of voice traffic. In both cases the PLR and data rate (throughput) were used as performance metrics for the rate adaptation heuristic. The rate fairness model was evaluated by means of Jain's fairness metric (Chapter 7 contains more detail regarding fairness metrics). The simulation was conducted for 90 seconds, since the network converges easily within this time.

8.2.1 Data Traffic Analysis

Figure 8.1 illustrates the PLR of each flow in the network, where a flow was regarded as the motion of information from source to destination node. It should be noted that all flows in the network were regarded as users; all attempting to communicate from source to destination simultaneously. Furthermore, the PLR was only dependent on

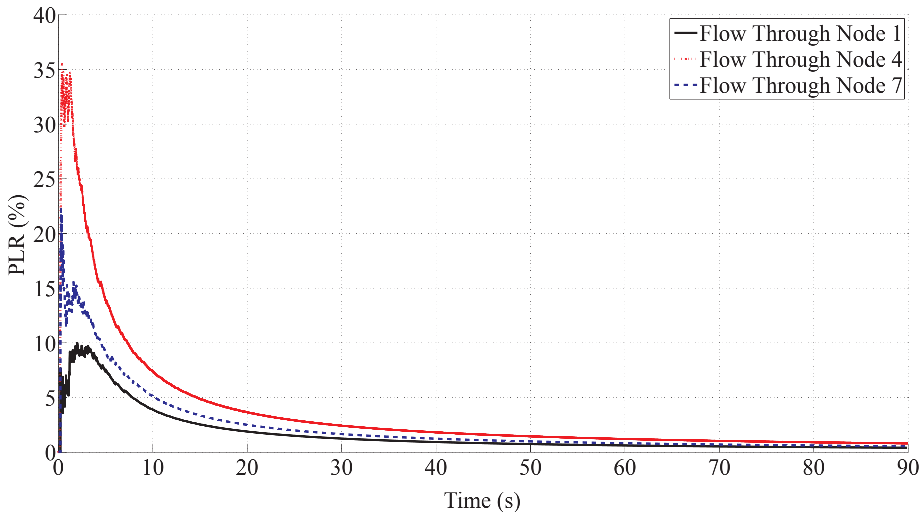


Figure 8.1: PLR obtained by applying the rate adaptation heuristic to the stack topology.

the interference and the buffer size on each node was assumed to be infinite. From Figure 8.1, it is clear that the inner flow started out with the highest PLR, as this flow experiences the most interference and consequently had the highest probability of error. After about 15 seconds, all flows were reduced to a PLR of 5% or less. Finally, after just 80 seconds the PLR dropped below 1% and became almost completely insignificant. The time required for the PLR to drop below 5% and 1% is considered a measurement of the effectiveness of the rate heuristic combined with the fairness model.

Figure 8.2 illustrates the data rate achieved by the inner flow and outer flows in the stack topology. Due to the symmetric nature of this topology, the two outer flows obtained the same data rate. After 15 seconds, the inner flow enjoyed a higher

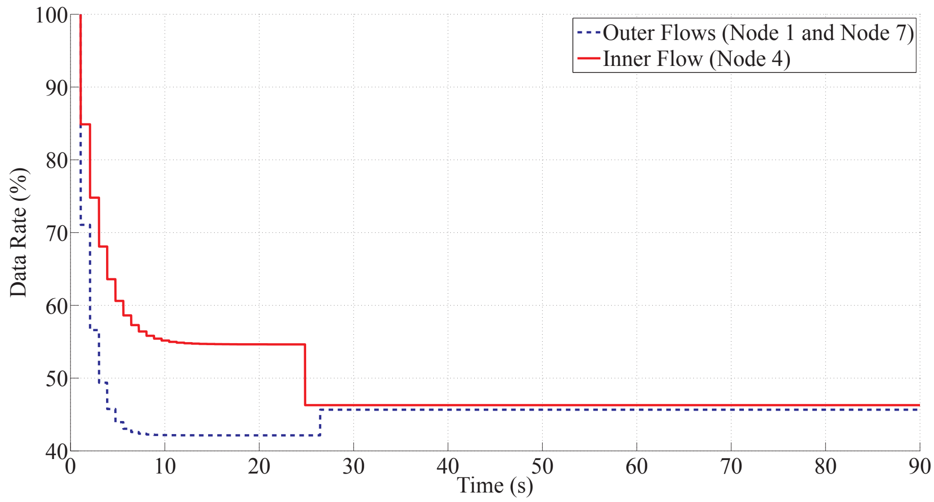


Figure 8.2: Data rate obtained by applying the rate adaptation heuristic to the stack topology.

data rate compared to the two outer flows. At this point a near-optimal data rate had been achieved, corresponding to a saddle point (see Chapter 6, Section 6.4.1). It should be noted that at this point the allocation data rate among flows was not fair. Equation 8.1 illustrates the fairness index at this point:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2} = 0.9832. \tag{8.1}$$

After about 25 seconds the rate fairness stage of optimisation started (see Chapter 7, Section 7.4.2). Initially, a decrease in the data rate of the inner flow was seen and for this brief period the channel capacity was not being utilised optimally. About 2 seconds later, both outer flows simultaneously increased their rates and the channel capacity was once again optimally utilised. Equation 8.2 illustrates the fairness index at this point:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n \sum_{i=1}^{n} (R_D^i)^2} = 1.0000. \tag{8.2}$$

8.2.2 Voice Traffic Analysis

Voice traffic consists of two types of packets, namely, voice and silence packets (see Chapter 4, Section 4.2.2). A user moves from a voice to a silent state with a probability of 0.4 to 0.6. Similarly, a user moves from a silent to a voice state with the same probability. This transition led to an erratic allocation of data rates among flows. Table 8.1 depicts the PLR, data rate and fairness allocations for all flows in the stack topology, where a single flow was forced to stay in a silent state for the whole simulation time. The other flows remained in the voice state for the simulation time. Force F13 denotes that F13 was forced to remain in the silent state, and the same convention was applied to all flows in the network. It should be noted that when a specific flow was currently operating at a silent state, other users could take advantage of the reduced rate (see Chapter 4, Section 4.2.2).

Table 8.1: Performance of rate adaptation heuristic when subjected to forced voice traffic in a stack topology.

	PLR below 5%	**Data Rate**	**Fairness Index**
Force F13	13.50 s	58.49%	1.0000
Force F46	11.45 s	68.96%	1.0000
Force F79	13.50 s	58.49%	1.0000

Table 8.1 indicates that the PLR dropped below 5% earlier when the flow contributing to the most interference was forced into the silent state. Moreover, the longest time taken for the PLR to drop below 5% (in this case obtained by forcing F13) was still lower compared to the case when using data traffic. This is due to the fact that lower source rates were allowed for silence packets, leading to a lower probability of error. The lower source rates also led to lower audio quality, but this

reduction in quality was not noticed when considering silence.

The actual data rate was most when the flow inducing the most interference (F46) was forced into the silent state. The lowest data rate occurred when forcing either F13 or F79 into a silent state. It was noted that this data rate was still higher compared to the case when utilising data traffic. This is also due to the fact that lower source coding led to higher data rates. To see this consider the physical bits required to transmit high quality audio, now consider that the high quality audio only consisted of silence. Silence owns very little entropy, and thus requires very few physical bits. This leads to higher channel coding with similar data rates, or effectively similar channel coding with higher data rates.

The fairness index was similar to the case when using data traffic. All situations resulted in a fair rate allocation because the second stage of optimisation was independent of the type of traffic being utilised. Although this method of forcing certain flows into a silent or voice state does not represent a real-world situation, it did provide insightful results. For example, the PLR dropped below 5% quicker and higher data rates could be achieved compared to data traffic, whilst still resulting in a fair solution for the allocation. The comparison of fixed topologies in Section 8.5 presents results that are more closely related to a real- world situation.

8.3 HALF DIAMOND TOPOLOGY RESULTS

This section provides a performance analysis of the rate adaptation heuristic combined with the fairness model for the half diamond topology. The half diamond topology differs from the stack topology by adding two additional links forming a half diamond structure between nodes 2, 4, 5 and 6. Although these extra links provided an additional path for the inner flow, the flow was forced through the worst case path, resulting in an accurate evaluation of the rate adaptation model only. Again, both data and voice traffic were applied to the network, but not simultaneously. The simulation was conducted for 90 seconds and the performance metrics were the throughput, PLR and fairness index.

8.3.1 Data Traffic Analysis

Figure 8.3 illustrates the PLR of each flow in the network. The inner flow started out with the highest PLR, but in this case the PLR of the flow moving through node 1 was significantly higher than the flow through node 7. This can be considered intuitive as

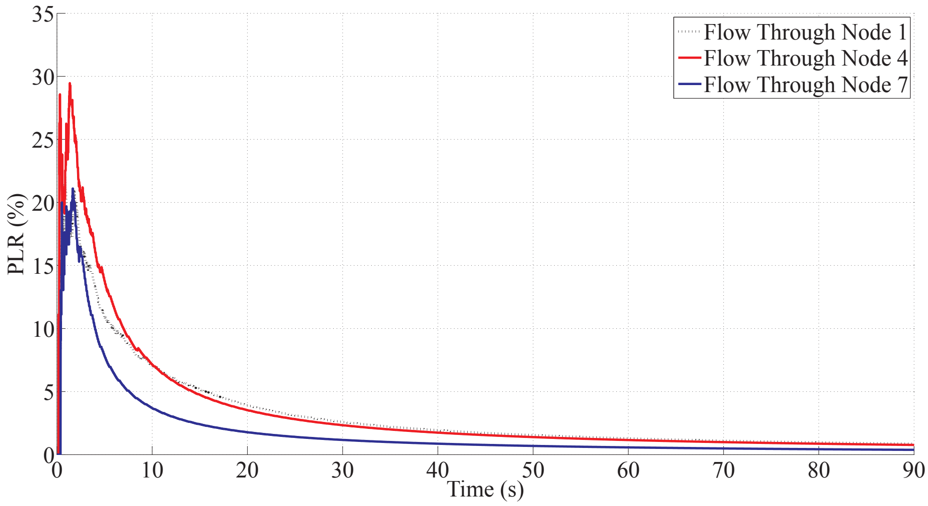


Figure 8.3: PLR obtained by applying the rate adaptation heuristic to the half diamond topology.

the amount of interference experienced by the flow through node 1 was significantly higher than that experienced by the flow through node 7. However, the inner flow still experienced the most interference and thus resulted in the highest probability of error. After about 15 seconds, all flows were reduced to a PLR of 5% or less, although the PLR of the flow through node 1 was higher compared to its counterpart in the stack topology. After just 80 seconds the PLR dropped below 1% and became almost completely insignificant.

Figure 8.4 illustrates the data rate achieved by the inner flow and outer flows in the half diamond topology. It should be noted that the outer flows differed in this case, due to the asymmetrical nature of the half diamond topology. After about 15

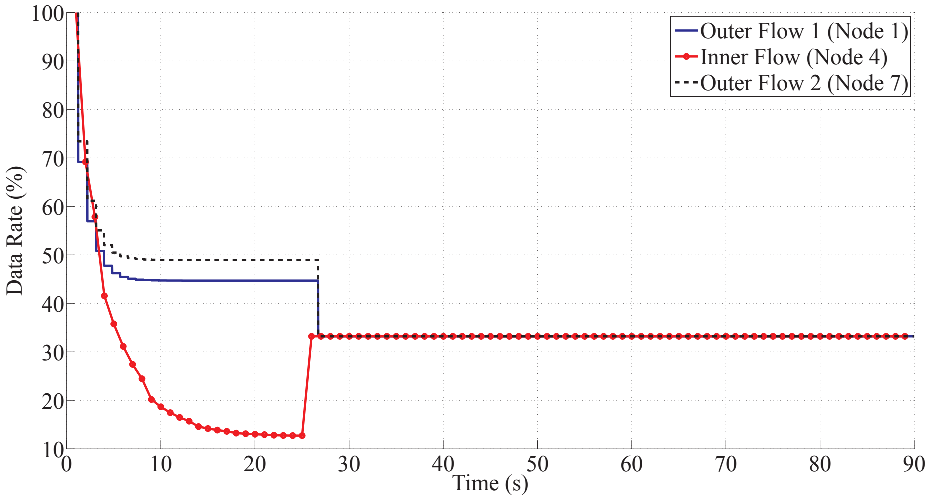


Figure 8.4: Data rate obtained by applying the rate adaptation heuristic to the half diamond topology.

seconds, the flow through node 7 obtained a constant rate, higher than the other two flows. The inner flow suffered the lowest rate, due to the amount of interfering channels. At this stage a near-optimal rate allocation (corresponding to a saddle point) was found, which was also an unfair allocation. The fairness index is illustrated in Equation 8.3 at this point:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2} = 0.8237. \tag{8.3}$$

After about 25 seconds the rate fairness stage of optimisation started (see Chapter 7, Section 7.4.2). A brief increase in the data rate was seen for the inner flow; for this brief period the allocation of data rates was considered sub-optimal. However, this was corrected within 2 seconds, where both outer flows decreased their data rates. From this point onwards the data rate allocation was once more near-optimal and more fair. The fairness index of the rate allocation at this point is illustrated by Equation 8.4:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2} = 1.0000. \tag{8.4}$$

8.3.2 Voice Traffic Analysis

Table 8.2 depicts the average PLR, average data rate and fairness allocations for all flows in the half diamond topology, where a single flow was forced to stay in a silent state for the whole simulation. The other flows remained in the voice state for the simulation. The first column in Table 8.2 depicts the time it took for the PLR of all the flows in the network to drop below 5%. It was noted that when the flow contributing to the most interference (F46) was forced into the silent state, the PLR dropped below 5% sooner compared to the situation of forcing the other flows. The PLR for all situations of forcing flows dropped below 5% sooner compared to the network utilising data traffic. The PLR dropped below 5% sooner when utilising voice traffic in the half diamond topology compared to the stack topology. Nevertheless it did not drop significantly enough to merit a performance improvement in this topology.

Table 8.2: Performance of rate adaptation heuristic when subjected to forced voice traffic in a half diamond topology.

	PLR below 5%	**Data Rate**	**Fairness Index**
Force F13	12.50 s	49.96%	1.0000
Force F46	10.15 s	48.38%	0.9995
Force F79	13.40 s	45.94%	1.0000

The second column of Table 8.2 represents the average data rate achieved by the flow in the half diamond topology. In all situations the data rate was higher than that achieved when compared to the data traffic case. The lowest data rate when utilising voice traffic was about 10% higher than that achieved when utilising data traffic. This is again due to the fact that the rate adaptation heuristic could take advantage of the lower source encoding required for silence packets. Finally, the fairness index was similar to the case when using data traffic and all situations resulted in a fair rate allocation.

8.4 FULL DIAMOND TOPOLOGY RESULTS

The full diamond topology, was used for evaluating the performance of the rate adaptation heuristic combined with the fairness model. This topology adds an additional two links to the half diamond topology, forming a full diamond structure along nodes 2, 4, 8 and 6. These links provide an additional path compared to the half diamond topology, and again the worst case path was forced as a selection, which led to the accurate evaluation of the rate adaptation model (by isolating the interference mitigation mechanism). Data and voice traffic were applied to the network and the simulation was conducted for 90 seconds. The performance was measured in terms of throughput, PLR and fairness index.

8.4.1 Data Traffic Analysis

Figure 8.5 illustrates the PLR of the inner and both outer flows the network. The inner flow initially obtained the highest PLR, but after only 10 seconds all the flows in the network obtained similar PLRs. All other topologies had significantly lower

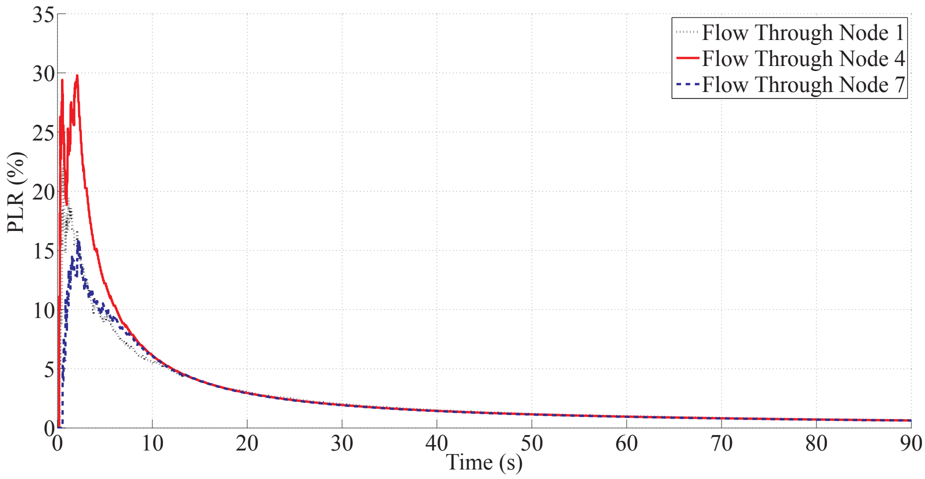


Figure 8.5: PLR obtained by applying the rate adaptation heuristic to the full diamond topology.

PLR at this point; this is due to the fact that flows in the full diamond topology experience the most interference. The inner flow in this topology experienced the most interference compared to all other flows, and thus obtained the highest probability of error. Although this topology consists of the most interference, the PLR for all flows still dropped to 5% or less within just 15 seconds. The PLR also became almost completely insignificant when it dropped below 1% after about 80 seconds. Figure 8.5 illustrates that even when faced with significantly more interference, the PLR could still be forced below a certain threshold by adapting the power levels.

Figure 8.6 illustrates the data rate achieved by the inner and both outer flows in the full diamond topology. Similar to the stack topology, this topology is symmetrical around the inner flow, therefore the two outer flows obtained similar data rates. After about 15 seconds, the flow through nodes 1 and 7 obtained a constant rate, which was higher compared to the flow through node 4. Again, a near-optimal

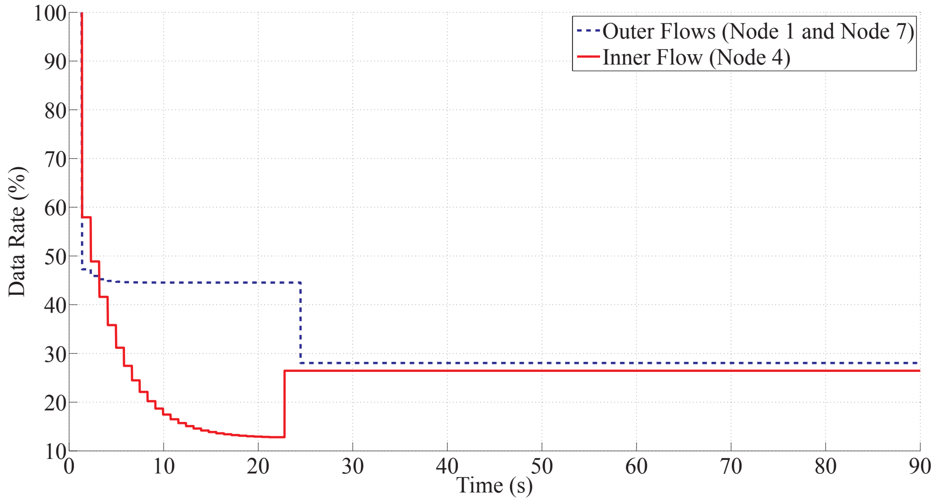


Figure 8.6: Data rate obtained by applying the rate adaptation heuristic to the full diamond topology.

rate allocation corresponding to a saddle point was found. The rate allocation at this point was again an unfair allocation, as illustrated in Equation 8.5:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2} = 0.8251. \tag{8.5}$$

The rate fairness stage of optimisation started after about 23 seconds. Initially, an increase in the data rate of the inner flow was observed and for this brief period the channel capacity was not utilised optimally. The two outer flows reacted to the change in capacity in just 2 seconds by decreasing their rates simultaneously. This resulted in a more fair and near optimal allocation of data rates for all flows in the network. The fairness index of the rate allocation at this point is reflected by Equation 8.6:

$$FI_{Jain's} = \frac{(\sum_{i=1}^{n} R_D^i)^2}{n\sum_{i=1}^{n}(R_D^i)^2} = 0.9985. \tag{8.6}$$

8.4.2 Voice Traffic Analysis

Table 8.3 depicts the average PLR, average data rate and fairness allocations for all flows in the full diamond topology, where a single flow was forced to stay in a silent state throughout the simulation. The PLR dropped below 5% within just 10.75 seconds when the flow inducing the most interference was forced into the silent state. In this case, the situation resulting in the longest time taken to drop below 5% was also longer compared to the data traffic case, whereas the shortest time was shorter than the data traffic case. The longest time was longer than that of the half diamond topology and shorter than that of the stack topology. Again, this difference was not significant compared to the other the performance measures. The data rate was significantly more

Table 8.3: Performance of rate adaptation heuristic when subjected to forced voice traffic in a full diamond topology.

	PLR below 5%	**Data Rate**	**Fairness Index**
Force F13	13.30 s	46.11%	0.9998
Force F46	10.75 s	36.35%	0.9999
Force F79	12.80 s	46.39%	0.9998

than that obtained in the data traffic case. The lowest data rate obtained occurred when forcing F46 into the silent state and resulted in a data rate of 36.35%, which was about 8% more than that achieved when utilising data traffic. The data rate allocation in this case led to a fair solution, and thus each flow obtained a rate almost equal to the other flows (or at least within a minimum relation bound).

8.5 COMPARISON OF FIXED TOPOLOGIES

This section discusses the differences observed between the fixed topologies. The section is divided into two subsections, each presenting comparisons in different types of traffic. The comparisons provide insights into the amount of interference experienced and the effective near-optimal rate that can be achieved. It should be noted that only the worst case paths were used as a method for isolating the effectiveness of the rate adaptation heuristic.

8.5.1 Data Traffic Analysis

Table 8.4 illustrates the PLR, data rate and fairness achieved after the second state of optimisation (after the game theory model was applied) for the stack, half diamond and full diamond topologies when utilising data traffic. The PLR was documented for the time taken to drop below 5% as well as the time taken to drop below 1%. It was observed that the PLR dropped below 5% sooner in the full diamond topology compared to the other topologies. This is due to the fact that information is passed between nodes more quickly due to more available links, which ultimately leads to quicker cooperation. The 5% drop in PLR for the half diamond topology takes longer (even though it consists of more links) due to the asymmetrical structure of the topology. Similar trends were observed when the PLR dropped below 1%.

Table 8.4: Performance of the rate adaptation heuristic when subjected to data traffic for all fixed topologies.

	PLR below 5%	**PLR below 1%**	**Data Rate**	**Fairness Index**
Stack	14.67 s	72.50 s	47.27%	1.0000
Half Diamond	15.76 s	77.20 s	33.20%	1.0000
Full Diamond	11.95 s	56.60 s	28.03%	0.9985

The data rate obtained by the flows in the stack topology were higher than those achieved in the half and full diamond topologies. This is intuitive when considering the amount of interference induced by this topology compared to the other topologies. Moreover, as the number of interfering links increased, the data rate decreased, as was observed when comparing the half diamond with the full diamond topology. A trade-off can be illustrated when considering the time required to reduce the PLR to 5% or less and the data rate achieved. For example, the full diamond topology obtained a lower PLR sooner by utilising more links at the cost of a lower data rate. The fairness index illustrated the effectiveness of the second stage of the optimisation process. Nevertheless, all topologies obtained similar fairness indices, illustrating no significant differences.

8.5.2 Voice Traffic Analysis

The previous sections illustrated the performance achieved when forcing certain flows into a silent state. However, such situations do not occur often in real-world applications. This section focuses on the comparisons made when considering voice traffic when flows switch between states with a 50% likelihood (see Chapter 4, Section 4.2.2). Figure 8.7 illustrates the data rate achieved by the stack, half diamond and full diamond topologies utilising randomised voice traffic. From Figure 8.7 it is clear that the data rates were more erratic due to the frequent changes between states. In contrast to the previous sections describing voice traffic, the results presented here were obtained by allowing all flows to change states at any point (thereby contributing to the erratic nature). It should be noted, however, that on average a higher data rate can be achieved by utilising voice traffic compared to data traffic. The data rate again increased as the number of interfering links decreased.

8.6 RANDOMISED NETWORK RESULTS

The previous sections presented results based on topologies that do not occur naturally in real world applications. Although these topologies presented insightful results based on the PLR and data rates when considering the amount of interfering links. This

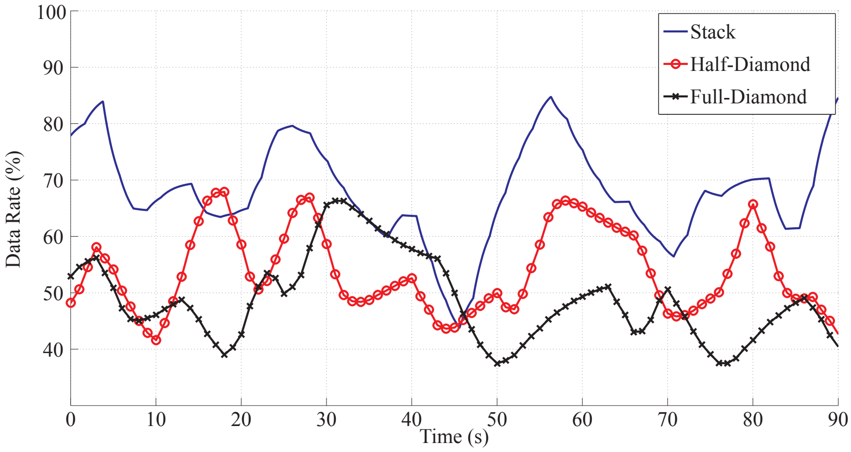


Figure 8.7: Data rate obtained in all topologies when utilising voice.

section provides results focused on the application of the rate adaptation heuristic to a more real-world scenario. The topology used for performance analysis in this section is randomised (see Chapter 3, Section 3.5). Furthermore, the simulation was repeated 200 times to ensure that the results were significant. The area in which nodes could be generated was fixed and the number of nodes was increased, which resulted in different levels of network density and, in turn, led to more interference. The network was populated with 50, 75 and 100 nodes. In a random model, the actual data rate varies according to the topology. The effectiveness of the rate adaptation heuristic was thus measured according to the average PLR and average data rate of an average node in such a network.

Figure 8.8 illustrates the PLR of a randomised network, repeated and averaged 200 times for 50, 75 and 100 nodes. It was observed that as the number of nodes in the network increased, the time it took for the PLR in this network to drop below 5% increased. This is due fact that to more nodes required information on the power and rates of other nodes in the same local area. It was also noted that the PLR converged on a specific point (about 800 seconds), illustrating that the heuristic could be scaled to large networks and still effectively reduce the PLR.

Figure 8.9 illustrates the data rate achieved in a randomised network, repeated

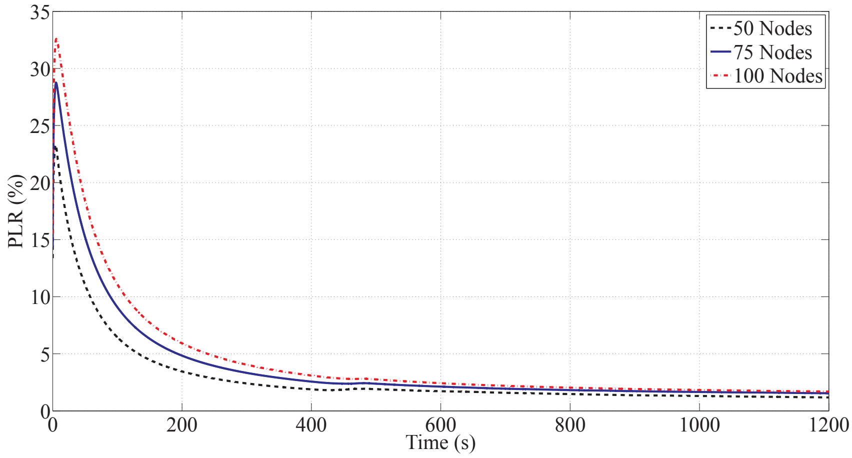


Figure 8.8: PLR obtained by applying the rate adaptation heuristic to 50, 75 and 100 node network.

and averaged 200 times for 50, 75 and 100 nodes. A trend was observed when considering the steady state data rate and the number of nodes. As the number of nodes increased, the achieved data rate decreased. This was consistent with the

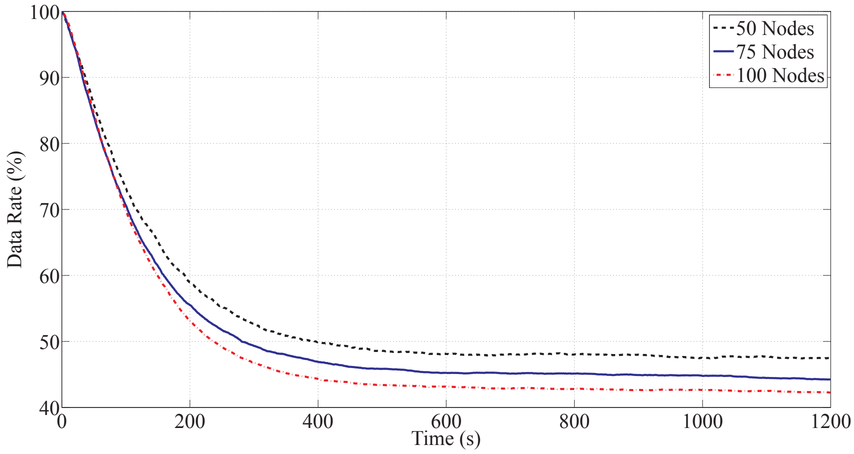


Figure 8.9: Data rate obtained by applying the rate adaptation heuristic to 50, 75 and 100 node networks.

notion that increasing the amount of interference decreases the achievable data rate. Figure 8.9 illustrates that a near-optimal data rate allocation could be obtained for large networks. Table 8.5 summarises the time it took for a reduction of 5% in PLR, the data rate achieved and the statistical significance achieved (illustrated by the variance) of the data rate at a steady state.

Table 8.5: Summary of results obtained in randomised networks.

	PLR below 5%	**Data Rate**	**Data Rate Variance**
50 Nodes	133 s	48.02 %	12.843e-4
75 Nodes	193 s	45.13 %	4.0367e-4
100 Nodes	239 s	42.80 %	6.0162e-4

8.7 COMPARISON OF RATE ADAPTATION ALGORITHMS

Several rate adaptation techniques have been proposed in an attempt to counter the poor performance observed when employing the traditional TCP congestion avoidance technique in wireless networks. The wireless control protocol (WCP) addresses this problem by making use of acknowledgement and congestion notifications [10]. When a node that makes use of WCP receives an acknowledgement message, the node increase its rate. However, when this node receives a congestion notification it decreases its rate. It is thus easy to see that WCP recognizes congestion and attempts to alleviate it.

Although WCP provides a technique for rate adaptation, it can not adjust the rate by an optimal amount. The wireless control protocol with capacity estimation (WCPcap) on the other hand extends WCP by taking the amount of interference into account and can thus estimate the channel capacity [10]. This leads to a method that adjusts the rate by a specific amount. Although this method could lead to a fair allocation of rates across a specific network, it does not achieve optimal rate allocation.

Relative fairness and optimized throughput (REFOT) is another rate adaptation technique used in ad hoc networks [101]. This technique computes the channel quality by keeping track of the number of consecutive successful and failed transmissions [102]. Although the REFOT method and the rate heuristic and fairness model

(RHF), share a common goal; to find the optimal and fair allocation of data rates in wireless networks, they differ significantly. Firstly, the REFOT method is modelled using a 3-dimensional Markov chain, whereas the RHF method makes use of the rate heuristic and fairness model presented in Chapter 6 and 7, respectively. Secondly, the REFOT method makes use of acknowledgement and congestion notifications at the transport layer to estimate the channel quality, whereas RHF makes use of the OLSR routing protocol to distribute transmission power and MCS information to calculate the amount of interference. The REFOT method was not used for comparison in this chapter.

In this section the WCP and WCPcap rate adaptation algorithms are compared to the RHF model. Figure 8.10 illustrates the RHF model, that is presented in this study, compared to the WCP and WCPcap algorithms in terms of data rate.

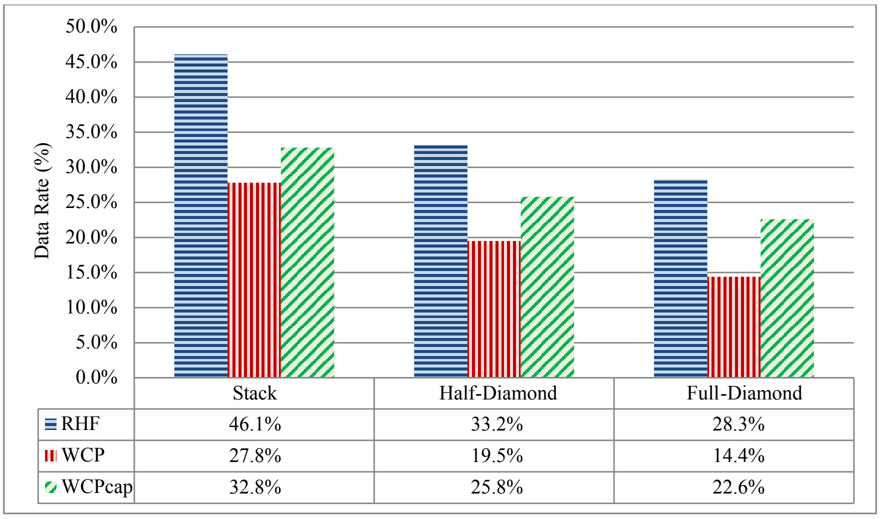
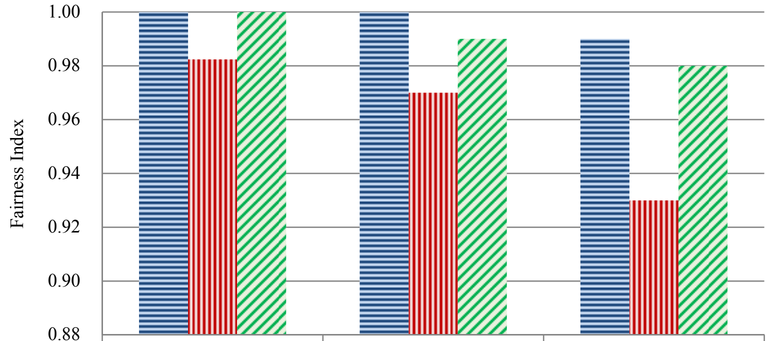


Figure 8.10: A comparison of the RHF model to the WCP and WCPcap methods, in terms of data rate.

The figure illustrates a comparison making use of the stack, half diamond and full diamond topologies. It should be noted, however, that this comparison presents performance when the worst case path was selected for each flow. Thus each flow can only convey information along the single worst case path, as explained in Chapter 3. It is seen from Figure 8.10 that the RHF model outperforms the other two algorithms by about 13% for the stack topology. It can be seen that the RHF model also outperforms the other methods for the other topologies, although a smaller difference is observed. The RHF model outperforms the other algorithms by about 8% and 5% for the half diamond and full diamond topologies.

The fairness indexes of all three methods are compared in Figure 8.11. The fairness index was calculated using Jains fairness metric, as defined in Equation 7.3. The RHF model obtains a higher fairness index compared to the WCP method and matches the fairness of the WCPcap method when applied to the stack topology. The fairness index of the RHF model is higher than both the other methods when applied to the half diamond topology. Although the fairness index for the RHF model when applied to the full diamond topology is slightly lower than the fairness found for other topologies, it still achieves a higher fairness metric compared to the other rate adaptation algorithms.



	Stack	Half-Diamond	Full-Diamond
RHF	1.00	1.00	0.99
WCP	0.98	0.97	0.93
WCPcap	1.00	0.99	0.98

Figure 8.11: A comparison of the RHF model to the WCP and WCPcap methods, in terms of fairness.

8.8 CONCLUDING REMARKS

This chapter discussed the performance of the rate adaptation heuristic and rate fairness model when subject to the stack, half diamond and full diamond topologies. The performance was measured in terms of PLR, data rate and fairness index. It was illustrated how the PLR decreased more rapidly when more links were present. This is due to the fact that information is passed to neighbouring nodes more efficiently. The data rate was found to increase as the number of links decrease, due to the decrease in interfering nodes. Finally, it was shown that fairness can be achieved irrespective of the number of links.

Both data and voice traffic were applied to the fixed topologies in order to illustrate the advantage of lowering the source encoding rate when generating silence packets. In most cases, a gain of about 10% in data rate (whilst remaining within similar PLR reduction) was observed when making use of this technique. Finally, it was shown how a real-world application would generate voice traffic, which would ultimately lead to a more erratic allocation of data rate. This erratic allocation of data rates is seen to be about 10% higher on average compared to data traffic.

The final section in this chapter presented a randomised network, which closely relates to a real-world scenario. In this case it was shown that the rate adaptation heuristic could be applied to large networks, proving the scalability of the model. It was seen that as the number of nodes in the network increases, the time it takes for the PLR to drop below 5% increases and the data rate decreases. This is due to the fact that more nodes require information regarding the power and rates of other nodes, and the interference increases as the number of nodes increase.

CHAPTER 9

CONCLUSION

This dissertation presented a method that provides the users in a wireless network additional performance benefits. It was illustrated that when users cooperate, the channel could be used more effectively. Furthermore, this dissertation presented framework illustrated in the form of an optimisation problem, which could be used to develop several models (not only limited to rate adaptation). It was shown how a routing protocol (such as the OLSR protocol) could be used to aid rate adaptation in a network. A game theory model based on this approach was developed, which led to the cooperation between users. It was shown that a more fair allocation of data rate could be achieved whilst the data rate remains near-optimal, by making use of this cooperation. Finally, a performance analysis has shown that the method effectively and fairly allocates data rate to each user in the network. It was shown how the method reacted to different types of traffic and that the method is scalable to large networks.

9.1 SUMMARY

The dissertation was laid out using a top-down approach, where the top most level consisted of the network topology and lowest level consisted of physical layer techniques for rate adaptation. The dissertation included all relevant layers as defined by the OSI layering model. This dissertation also illustrated two different types of traffic and their related application layer protocols, additionally; it has been shown how these types of traffic could be applied to a WMN. Details on the network layer, specifically the routing protocol has been discussed, as well as, how these protocols could be used to aid rate adaptation. The physical layer has been presented with specific reference to rate adaptation and how game theory plays a role in this technique. Finally results have been presented, illustrating the effectiveness of the rate adaptation heuristic and game theory model.

Chapter 2 discussed the background information required for WMNs, game theory, resource allocation and IEEE communication standards. The advantages and challenges of WMNs were stated along with the intended applications for these types of networks. Game theory has been discussed in terms of the applicableness to wireless communication networks and the challenges encountered in such networks. Different types of techniques and the feasibility and advantages of these techniques were presented in the section entitled resource allocation, it was also illustrated that heuristics provide feasible solutions to NP-hard and NP-complete problems (although they often lead to sub-optimal solutions). Finally, the physical layer standards for communication devices were presented with specific reference to the standard most commonly used in WMNs; the IEEE 802.11 standard.

Chapter 3 presented the topologies used for evaluating networks and form the top most level in the top-down approach. Three fixed topologies, as well as, a randomised topology were presented. Fixed topologies follow a particular structure and provide significant advantages in terms of evaluating the performance of rate adaptation in an interfering environment. These fixed topologies did not reflect real world situations; the random topology on the other hand, better represented a real world environment. The random topology was presented in three forms, namely, a network consisting of 50 nodes, 75 nodes and 100 nodes all of which were randomly generated in a fixed area. The randomised topology provided advantages in terms of the performance of the rate adaption technique when subjected to large and dense networks.

Chapter 4 illustrated the layers used to model the communication device. This model closely followed the seven-layer OSI model which traditionally consists of an application, presentation, session, transport, network, data link and physical layer. The application layer consisted of two protocols, namely FTP and VoIP. The protocols handle different types of traffic; data traffic is normally used in conjunction with the FTP protocol, where VoIP is typically used for voice traffic. The next significant layer discussed was the network layer, responsible for the effective routing of information from destination to source. This layer made use of the OLSR protocol which was explained in much greater detail in Chapter 5. The physical layer in its most simplistic form consisted of an encoder, modulator, channel, demodulator and decoder. The physical layer was illustrated in terms of the IEEE 802.11 standard.

Chapter 5 described the routing protocol used in this network model, namely, the OLSR protocol. The primary functions of the OLSR protocol was discussed with specific reference to the proactive nature of the protocol, as well as how the protocol created routing tables and how link state information was distributed. The auxiliary functions, which were used for improving the performance of this protocol was described. Auxiliary functions of the OLSR protocol included MPR which were used to reduce the amount of overhead required for routing. It was shown how the OLSR protocol achieves shortest path routing, as well as how the protocol structures its headers. Finally, it was shown that the frame structure for this protocol contains fields that could be used by other algorithms, (such as a rate adaptation heuristic) for passing messages between nodes in a WMN.

Chapter 6 defined the rate adaptation heuristic that takes place at the physical layer. This chapter defined the rate adaptation problem in terms of the objective function, the decision variables and the constraints found in a wireless communication system. A formal mathematical problem formulation was derived and the problem was identified as a non-linear mixed integer programming problem. Furthermore, it was shown that an exact solution to the problem was infeasible and a heuristic would be more appropriate. Finally, the rate adaptation heuristic found a near optimal solution to the problem was described. It was illustrated that this heuristic required messages passing between nodes and how the OLSR protocol could aid in this requirement. Lastly, it was noted that the heuristic did not attempt to find the fair allocation of resources among nodes, and that game theory would be required to achieve this.

Chapter 7 presented the game theory model that achieved a fair allocation of data rates among all nodes in a network. This chapter illustrated the different measures of fairness most commonly found in resource allocation problems. The advantages and disadvantages of each metric were given, and it was shown why Jain's fairness metric was the most feasible metric for the rate adaptation problem. The game theory modelled was then defined by illustrating how it follows a static non-cooperative model. Furthermore, it was shown that the problem can be model as a repeated game with an infinite horizon. Finally, a mathematical analysis of the model was given and it was shown how cooperation could be achieved by making use of the Folk theorem.

Chapter 8 illustrated the performance of the rate adaptation heuristic and rate fairness model. In this chapter it was shown how the heuristic and fairness model performed when subject to different amounts of interference. This was achieved by making use of the different fixed topologies as was described in Chapter 3. In addition to the amount of interference, the rate adaptation heuristic and fairness model was evaluated by making use of both data and voice traffic (discussed in Chapter 4). For all 3 topologies, using both data and voice traffic, it was shown that a fair and near-optimal rate allocation can be achieved. Furthermore, it was shown that by taking advantage of lower source coding, an increase of about 10% in terms of data rate can be achieved (compared to data traffic). This chapter also illustrated that the rate adaptation heuristic and fairness model can be applied to large networks.

9.2 FUTURE RESEARCH

Based on the findings of this study, the author recommends that further research be conducted on the following topics:

- An investigation into the performance of WMNs utilising a joint rate adaptation and path selection model. This model should adapt the transmission rate of each node, as well as select a path that would lead to the highest data rate and lowest interference for other nodes in the network. This would require alteration of the routing protocol (Chapter 5), as well as the mathematical problem formulation (Chapter 6). The biggest challenge would probably arise when considering the trade-off between shortest path and path with least interference.

- The next challenge would be to relax the assumption of an AWGN channel. This would allow for the evaluation of the rate adaptation heuristic in a network that contains multipath fading and shadowing. The most significant challenge with this improvement would be to solve the hidden terminal problem in a distributed network overcoming shadowing. Furthermore, developing a distributed channel estimation technique would probably become a necessity when attempting to mitigate the effects of multipath fading. These techniques would likely result in added complexity and more overhead.

- Achieving a general WMN by extending the rate adaptation heuristic and fairness model to other IEEE standards, a more general WMN would be obtained. A heterogeneous network, where desktop computers communicate with mobile phones and even television sets could result from extending the heuristic to other devices. One challenge that would probably require attention would be the different constraints defined by each standard and how these constraints should be dealt with in terms of fairness. Other issues that might arise would be the requirement of additional hardware for communicating in different frequency bands, encoding and decoding different FECs.

- An exact mathematical solution could lead to improved performance, as well as an improved mathematical framework. However, this solution would require a significant amount of information regarding the exact probability of loss and SINR for each node. An exact equation for each BER curve for all FECs and modulation schemes would be required, as well as, the added processing power required for finding the optimal solution in the set of integer MCSs. Additional overhead would be required for passing this added set of information to all nodes in the network.

- Passing information between nodes in a network leads to additional security concerns. Providing a method for passing information between sources that the node trusts, is another area of research not considered in this work. It is however required that this improved security method is decentralised as there is no central authority governing a WMN. Security techniques often rely on a technique making use of public and private keys, this in turn leads to added overhead.

- Finally, the effects of nodes that do not accept punishment due to non-cooperative behaviour could be investigated. Nodes that do not accept punishment could rebel, by attempting to jam other nodes in the network. The effective countering of jamming in this network could lead to a significant area of research. However this would require the investigation of different jamming techniques, such as denial of service (DoS) and null frequency jamming (NFJ).

REFERENCES

[1] M. Chiang, "Balancing transport and physical Layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 104 – 116, Jan. 2005.

[2] S. D. Barnes, P. A. Jansen van Vuuren, and B. T. Maharaj, "Spectrum occupancy investigation: Measurements in South Africa," *Measurement*, vol. 46, no. 9, pp. 3098 – 3112, 2013.

[3] A. A. Aziz, Y. A. Sekercioglu, P. Fitzpatrick, and M. Ivanovich, "A Survey on Distributed Topology Control Techniques for Extending the Lifetime of Battery Powered Wireless Sensor Networks," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 121–144, 2013.

[4] R. Bruno, M. Conti, and E. Gregori, "Mesh networks: commodity multihop ad hoc networks," *IEEE Communications Magazine*, vol. 43, no. 3, pp. 123–131, 2005.

[5] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23–S30, 2005.

[6] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjorungnes, *Game Theory in wireless and communication networks.* Cambridge Unversity Press, 2012.

[7] A. B. MacKenzie and L. A. DaSilva, "Game theory for wireless engineers," *Synthesis Lectures on Communications*, vol. 1, no. 1, pp. 1–86, 2006.

[8] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452 –1463, Aug. 2006.

[9] T.-S. Kim, Y. Yang, J. C. Hou, and S. V. Krishnamurthy, "Resource Allocation for QoS Support in Wireless Mesh Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2046–2054, 2013.

[10] S. Rangwala, A. Jindal, K.-Y. Jang, K. Psounis, and R. Govindan, "Neighborhood-Centric Congestion Control for Multihop Wireless Mesh Networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1797 –1810, Dec. 2011.

[11] L.-P. Tung, W.-K. Shih, T.-C. Cho, Y. S. Sun, and M. C. Chen, "TCP Throughput Enhancement over Wireless Mesh Networks," *IEEE Communications Magazine*, vol. 45, no. 11, pp. 64–70, 2007.

[12] D. Liu, B. Wang, and S. Guo, "Distributed cross-layer resource allocation in wireless ad hoc networks," in *Proc. 8th International Conference Expo on Emerging Technologies for a Smarter World (CEWIT)*, Nov. 2011, pp. 1 –6.

[13] P. R. Thie and G. E. Keough, *An introduction to linear programming and game theory*. Wiley-Interscience, 2011.

[14] R. J. Vanderbei, *Linear programming: foundations and extensions*. Springer, 2008, vol. 114.

[15] R. C. Dixon, "Lore of the token ring," *IEEE Network*, vol. 1, no. 1, pp. 11–18, 1987.

[16] L. Le, "Multipath Routing Design for Wireless Mesh Networks," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2011, pp. 1 –6.

[17] J.-H. Huang, L.-C. Wang, and C.-J. Chang, "Power Fairness in a Scalable Ring-

Based Wireless Mesh Network," in *Proc. IEEE 66th Vehicular Technology Conference Fall.*, 2007, pp. 341–346.

[18] S. Jayaprakasam, T.-C. Chuah, and S. W. Tan, "Collaborative mesh networking for low cost wireless coverage in rural areas," in *Proc. IEEE Symposium on Industrial Electronics Applications*, vol. 1, 2009, pp. 313–318.

[19] J. Ishmael, S. Bury, D. Pezaros, and N. Race, "Deploying Rural Community Wireless Mesh Networks," *IEEE Internet Computing*, vol. 12, no. 4, pp. 22–29, 2008.

[20] P. Dutta, S. Jaiswal, D. Panigrahi, and R. Rastogi, "A New Channel Assignment Mechanism for Rural Wireless Mesh Networks," in *Proc. 27th IEEE Conference on Computer Communications*, 2008, pp. 2261–2269.

[21] S. Vural, D. Wei, and K. Moessner, "Survey of Experimental Evaluation Studies for Wireless Mesh Network Deployments in Urban Areas Towards Ubiquitous Internet," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 223–239, 2013.

[22] K. R. Chowdhury and I. F. Akyildiz, "Cognitive Wireless Mesh Networks with Dynamic Spectrum Access," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 168–181, 2008.

[23] W. Kim, A. J. Kassler, M. Di Felice, and M. Gerla, "Urban-X: Towards distributed channel assignment in cognitive multi-radio mesh networks," in *Wireless Days (WD), 2010 IFIP*, 2010, pp. 1–5.

[24] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Usage Patterns in an Urban WiFi Network," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1359–1372, 2010.

[25] Z. Wan, K. Ren, B. Zhu, B. Preneel, and M. Gu, "Anonymous User Communi-

cation for Privacy Protection in Wireless Metropolitan Mesh Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 519–532, 2010.

[26] F. Jan, B. Mathieu, and D.-E. Meddour, "A monitoring tool for Wireless Multi-hop Mesh Networks," in *Proc. IEEE Network Operations and Management Symposium*, 2008, pp. 587–601.

[27] W. Liu, H. Nishiyama, N. Kato, Y. Shimizu, and T. Kumagai, "A novel gateway selection method to maximize the system throughput of Wireless Mesh Network deployed in disaster areas," in *Proc. IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012, pp. 771–776.

[28] A. Blair, T. Brown, K. M. Chugg, and M. Johnson, "Tactical Mobile Mesh Network System Design," in *Proc. IEEE Military Communications Conference (MILCOM)*, 2007, pp. 1–7.

[29] S. Glass, M. Portmann, and V. Muthukkumarasamy, "Securing Wireless Mesh Networks," *IEEE Internet Computing*, vol. 12, no. 4, pp. 30–36, 2008.

[30] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of the 10th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2004, pp. 114–128.

[31] X. Wang and A. O. Lim, "IEEE 802.11s wireless mesh networks: Framework and challenges," *Ad Hoc Networks*, vol. 6, no. 6, pp. 970 – 984, 2008.

[32] J. Luo, C. Rosenberg, and A. Girard, "Engineering Wireless Mesh Networks: Joint Scheduling, Routing, Power Control, and Rate Adaptation," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1387–1400, 2010.

[33] U. T. Nguyen and J. Xu, "Multicast Routing in Wireless Mesh Networks: Minimum Cost Trees or Shortest Path Trees?" *IEEE Communications Magazine*,

vol. 45, no. 11, pp. 72–77, 2007.

[34] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot, "Optimized link state routing protocol for ad hoc networks," in *Proc. IEEE International Multi Topic Conference on Technology for the 21st Century*, 2001, pp. 62–68.

[35] D. Nguyen and P. Minet, "Analysis of MPR Selection in the OLSR Protocol," in *Proc. IEEE 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW)*, vol. 2, 2007, pp. 887–892.

[36] A. B. MacKenzie and S. B. Wicker, "Game theory and the design of self-configuring, adaptive wireless networks," *IEEE Communications Magazine*, vol. 39, no. 11, pp. 126–131, 2001.

[37] M. Wooldridge, "Does Game Theory Work?" *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 76–80, 2012.

[38] A. B. MacKenzie and S. B. Wicker, "Game theory in communications: motivation, explanation, and application to power control," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, 2001, pp. 821–826 vol.2.

[39] K. Leyton-Brown and Y. Shoham, "Essentials of game theory: A concise multidisciplinary introduction," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 2, no. 1, pp. 1–88, 2008.

[40] I. Ha, Y. Yoon, and M. Choi, "Determinants of adoption of mobile games under mobile broadband wireless access environment," *Information and Management*, vol. 44, no. 3, pp. 276–286, 2007.

[41] J. J. Jaramillo and R. Srikant, "A game theory based reputation mechanism to incentivize cooperation in wireless ad hoc networks," *Ad Hoc Networks*, vol. 8, no. 4, pp. 416–429, 2010.

[42] H. Yaiche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.

[43] F. Meshkati, M. Chiang, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 6, pp. 1115–1129, 2006.

[44] S. Koskie and Z. Gajic, "A nash game algorithm for SIR-based power control in 3G wireless CDMA networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1017–1026, 2005.

[45] R. J. La and V. Anantharam, "Optimal routing control: game theoretic approach," in *Proc. IEEE 36th Conference on Decision and Control*, vol. 3, Dec 1997, pp. 2910 –2915 vol.3.

[46] J. M. Zolezzi and H. Rudnick, "Transmission cost allocation by cooperative games and coalition formation," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1008–1015, 2002.

[47] L. Zhao, J. Zhang, and H. Zhang, "Using Incompletely Cooperative Game Theory in Wireless Mesh Networks," *IEEE Network*, vol. 22, no. 1, pp. 39–44, 2008.

[48] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[49] C. J. Tomlin, J. Lygeros, and S. S. Sastry, "A game theoretic approach to controller design for hybrid systems," *Proceedings of the IEEE*, vol. 88, no. 7, pp. 949–970, 2000.

[50] T. Li and M. Shahidehpour, "Strategic bidding of transmission-constrained GENCOs with incomplete information," *IEEE Transactions on Power Systems*,

vol. 20, no. 1, pp. 437–447, 2005.

[51] B. Xu and L. Fan, "Information disclosing and cooperation in public goods game with punishment," in *Proc. 2nd International Conference on Information Science and Engineering (ICISE)*, 2010, pp. 2896–2899.

[52] D. K. Petraki, M. P. Anastasopoulos, H.-H. Chen, and P. G. Cottis, "Distributed Resource Allocation for Delay-Sensitive Services in Satellite Networks Using Game Theory," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 2, pp. 134–144, 2009.

[53] Z. Dziong and L. G. Mason, "Fair-efficient call admission control policies for broadband networks-a game theoretic framework," *IEEE/ACM Transactions on Networking*, vol. 4, no. 1, pp. 123–136, 1996.

[54] J. Virapanicharoen and W. Benjapolakul, "Fair-efficient threshold parameters selection in call admission control for CDMA mobile multimedia communications using game theoretic framework," in *Proc. IEEE Consumer Communications and Networking Conference (CCNC)*, 2005, pp. 439–444.

[55] Y.-L. Kuo, E.-K. Wu, and G.-H. Chen, "Noncooperative admission control for differentiated services in IEEE 802.11 WLANs," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 5, 2004, pp. 2981–2986 Vol.5.

[56] A. Arsie, K. Savla, and E. Frazzoli, "Efficient Routing Algorithms for Multiple Vehicles With no Explicit Communications," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2302–2317, 2009.

[57] R. Kannan and S. S. Iyengar, "Game-theoretic models for reliable path-length and energy-constrained routing with data aggregation in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1141–1150, 2004.

[58] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker, "On selfish routing in Internet-like environments," *IEEE/ACM Transactions on Networking*, vol. 14, no. 4, pp. 725–738, 2006.

[59] H.-J. Lee and J.-T. Lim, "Fair congestion control over wireless multihop networks," *IET Communications*, vol. 6, no. 11, pp. 1475 –1482, 24 2012.

[60] G. B. Dantzig, *Linear programming and extensions.* Princeton university press, 1998.

[61] P. Bjorklund, P. Varbrand, and D. Yuan, "Resource optimization of spatial TDMA in ad hoc radio networks: a column generation approach," in *Proc. IEEE Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, vol. 2, 2003, pp. 818–824 vol.2.

[62] P. Mitran, L. B. Le, and C. Rosenberg, "Queue-Aware Resource Allocation for Downlink OFDMA Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3100–3111, 2010.

[63] R. Madan, S. Cui, S. Lall, and A. Goldsmith, "Cross-Layer Design for Lifetime Maximization in Interference-Limited Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3142–3152, 2006.

[64] D. Nandiraju, L. Santhanam, N. Nandiraju, and D. P. Agrawal, "Achieving Load Balancing in Wireless Mesh Networks Through Multiple Gateways," in *Proc. IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2006, pp. 807–812.

[65] K.-D. Lee and V. C. M. Leung, "Fair Allocation of Subcarrier and Power in an OFDMA Wireless Mesh Network," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2051–2060, 2006.

[66] V. Ramamurthi, A. Reaz, and B. Mukherjee, "Optimal Capacity Allocation in

Wireless Mesh Networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–5.

[67] K. Papadaki and V. Friderikos, "Approximate dynamic programming for link scheduling in wireless mesh networks," *Computers & Operations Research*, vol. 35, no. 12, pp. 3848 – 3859, 2008, ¡ce:title¿Part Special Issue: Telecommunications Network Engineering¡/ce:title¿. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054807000470

[68] "IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications," *IEEE Std 802.3-2008 (Revision of IEEE Std 802.3-2005)*, pp. 1–2977, 2008.

[69] "ISO/IEC Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (Includes IEEE Std 802.11, 1999 Edition; IEEE Std 802.11A.-1999; IEEE Std 802.11B.-1999; IEEE Std 802.11B.-1999/Cor 1-2001; and IEEE Std 802.11D.-2001)," *ISO/IEC 8802-11 IEEE Std 802.11 Second edition 2005-08-01 ISO/IEC 8802 11:2005(E) IEEE Std 802.11i-2003 Edition*, pp. 1–721, 2005.

[70] "Approved Draft IEEE Standard for Local and metropolitan area networks Corrigendum to IEEE Standard for Local and Metropolitan Area Networks-Part 16: Air Interface for Fixed Broadband Wireless Access Systems (Incorporated into IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 E)," *IEEE Std P802.16/Cor1/D5*, pp. –, 2005.

[71] "Approved Amendment to Standard for Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks Specific

Requirements - Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPAN): Amendment to MAC Sublayer (Amendment to IEEE Std 802.15.3-2003) Replaced by IEEE 802.15.3b-2005," *IEEE Std P802.15.3b/D04*, pp. –, 2005.

[72] S. Xu and T. Saadawi, "Does the IEEE 802.11 MAC protocol work well in multi-hop wireless ad hoc networks?" *IEEE Communications Magazine*, vol. 39, no. 6, pp. 130–137, 2001.

[73] S. Chakraborty and S. Nandi, "MAC layer fairness in IEEE 802.11 DCF based Wireless Mesh Networks," in *Proc. International Conference on Computing, Networking and Communications (ICNC)*, 2012, pp. 1138–1143.

[74] D. Qiao, S. Choi, and K. G. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 1, no. 4, pp. 278–292, 2002.

[75] A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 3, 2005, pp. 2223–2234 vol. 3.

[76] O. Tickoo and B. Sikdar, "Modeling Queueing and Channel Access Delay in Unsaturated IEEE 802.11 Random Access MAC Based Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 4, pp. 878–891, 2008.

[77] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, 2004.

[78] I. Vukovic, "Analysis of the hidden node problem in HIPERLAN," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 2, 1999, pp. 1234–1238 vol.2.

[79] R. Van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Communications Magazine*, vol. 37, no. 12, pp. 82–88, 1999.

[80] X. Wang and S. Li, "Scalable Routing Modeling for Wireless Ad Hoc Networks by Using Polychromatic Sets," *IEEE Systems Journal*, vol. 7, no. 1, pp. 50–58, 2013.

[81] B. Boskovic and M. Markovic, "On spread spectrum modulation techniques applied in IEEE 802.11 wireless LAN standard," in *IEEE/AFCEA Information Systems for Enhanced Public Safety and Security (EUROCOMM)*, 2000, pp. 238–241.

[82] J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Communications*, vol. 10, no. 5, pp. 8–14, 2003.

[83] A. F. Molisch, *Wireless Communications*, ser. Wiley - IEEE. Wiley, 2010.

[84] V. Corvino, V. Tralli, and R. Verdone, "Cross-Layer Resource Allocation for MC-CDMA," in *Proc. 4th International Symposium on Wireless Communication Systems (ISWCS)*, 2007, pp. 267–271.

[85] B. Goode, "Voice over Internet protocol (VoIP)," *Proceedings of the IEEE*, vol. 90, no. 9, pp. 1495–1517, 2002.

[86] M.-H. Fong, R. Novak, S. Mcbeath, and R. Srinivasan, "Improved VoIP capacity in mobile WiMAX systems using persistent resource allocation," *IEEE Communications Magazine*, vol. 46, no. 10, pp. 50–57, 2008.

[87] J. Puttonen, T. Henttonen, N. Kolehmainen, K. Aschan, M. Moisio, and P. Kela, "Voice-Over-IP Performance in UTRA Long Term Evolution Downlink," in *Proc. IEEE Vehicular Technology Conference Spring (VTC)*, 2008, pp. 2502–2506.

[88] A. Sfairopoulou, C. Macian, and B. Bellalta, "QoS Adaptation in SIP-based VoIP

calls in Multi-rate IEEE 802.11 Environments," in *3rd International Symposium on Wireless Communication Systems (ISWCS)*, 2006, pp. 475–479.

[89] M. Slavik and I. Mahgoub, "Spatial Distribution and Channel Quality Adaptive Protocol for Multihop Wireless Broadcast Routing in VANET," *IEEE Transactions on Mobile Computing*, vol. 12, no. 4, pp. 722–734, 2013.

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

[91] S. Waharte, R. Boutaba, Y. Iraqi, and B. Ishibashi, "Routing protocols in wireless mesh networks: challenges and design considerations," *Multimedia Tools and Applications*, vol. 29, pp. 285–303, 2006, 10.1007/s11042-006-0012-8.

[92] J. G. Proakis and M. Salehi, *Digital Communications*, ser. McGraw-Hill International Edition. McGraw-Hill Higher Education, 2008.

[93] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[94] S. Lin and D. J. Costello, *Error control coding: fundamentals and applications*, ser. Prentice-Hall computer applications in electrical engineering series. Prentice-Hall, 1983.

[95] C. Berrou, "Convolutional turbo codes," in *Codes and Turbo Codes*, ser. Collection IRIS. Springer Paris, 2010, pp. 213–270.

[96] L. Ma, Y. Xu, W. Meng, and C. Li, "An Ant Colony Based Congestion Elusion Routing Scheme for MANET," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, dec. 2011, pp. 1 –5.

[97] H. Shi and H. Sethu, "An evaluation of timestamp-based packet schedulers using a novel measure of instantaneous fairness," in *Proc. IEEE International Confer-*

ence on Performance, Computing, and Communications Conference, 2003, pp. 443–450.

[98] M. Dianati, X. Shen, and S. Naik, "A new fairness index for radio resource allocation in wireless networks," in *Proc. IEEE Conference on Wireless Communications and Networking*, vol. 2, 2005, pp. 712–717 Vol. 2.

[99] H. Shi, R. Prasad, E. Onur, and I. Niemegeers, "Fairness in Wireless Networks:Issues, Measures and Challenges," *IEEE Communications Surveys Tutorials*, vol. 99, no. 99, pp. 1–20, 2013.

[100] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system.* Eastern Research Laboratory, Digital Equipment Corporation, 1984.

[101] A. Benslimane, A. Rachedi, and D. Diwakar, "Relative Fairness and Optimized throughput for Mobile Ad Hoc Networks," in *Proc. IEEE International Conference on Communications (ICC)*, May 2008, pp. 2233–2237.

[102] A. Benslimane and A. Rachedi, "Rate adaptation scheme for IEEE 802.11-based MANETs," *Journal of Network and Computer Applications*, vol. 39, pp. 126 – 139, 2014.