2014-5

A Remote Capacity Utilization Estimator for WLANs

Yi Ding
Technological University Dublin

A Remote Capacity Utilization Estimator for WLANs

By

Yi Ding

A thesis submitted to the Dublin Institute of Technology

for the degree of

Doctor of Philosophy

Supervisor: Dr. Mark Davis

Communications Network Research Institute (CNRI)

School of Electronic and Communications Engineering,

Dublin Institute of Technology (DIT),

Dublin, Ireland.

May, 2014

Declaration

I certify that this thesis which I now submit for examination for the award of _____, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology (DIT) and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the Dublin Institute of Technology's guidelines for ethics in research.

DIT has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis is duly acknowledged.

Signature _____ Date _____

Acknowledgements

This four years PhD study in DIT has become the most significant period of time in my life. I owe a great appreciation to many people who have helped me in many different ways during my PhD study period and the completion of this thesis.

First I would like to express my deepest gratitude my supervisor Dr. Mark Davis, for his invaluable guidance and support in all these years, his encouragement, advice, inspiration and infinite patience on the direction of my research, and for working so hard in helping this thesis to be written and guiding me through a lot of difficulties.

I also would like to give a huge amount thanks to Professor Gerald Farrell, Professor Bin Wu, and Professor Zhiguang Qin from UESTC, and Chinese Scholarship Council who provided this study opportunity in DIT.

A special thanks to my colleagues in Communication Network Research Institute (CNRI), Dr Jianhua Deng, Dr Fuhu Deng, Mr Yin Chen, Mr Chenzhe Zhang, Dr Tanmoy Debnath, Dr Mirek Narbutt, Dr Mustafa Ramadhan, and Mr Tony Grennan who help so many things on my research. I also thank to my friends who gives me useful technique suggestions: Dr Erqiang Zhou, Dr Yi Ding, Mr Wenliang Ao, and Mr Jianfeng Wu. A number of other people also deserve to be thanked here: my roommates and best friends in Ireland, Dr Rong Hu, Dr Qiaohuan Chen, Dr Lin Chen and his wife Mrs Jiemei Zhan, Dr Shipeng Wen, Mr Liang Jiang and Mrs Yanfen Zhou, Mr Jiajun Li, Ms Wanyu He, Mr Heliang Sun, and Mr Zhiqiang Yu who provided many supports for my life in Ireland in last five years. Without all these people, this thesis can never be finished that easy.

Finally, I also want to thank my parents, who give me the unconditional and endless love, a constant source of encouragement and support throughout my study and indeed throughout my whole life.

Abstract

In WLANs, the capacity of a node is not fixed and can vary dramatically due to the shared nature of the medium under the IEEE 802.11 MAC mechanism. There are two main methods of capacity estimation in WLANs: Active methods based upon probing packets that consume the bandwidth of the channel and do not scale well. Passive methods based upon analyzing the transmitted packets that avoid the overhead of transmitting probe packets and perform with greater accuracy. Furthermore, passive methods can be implemented locally or remotely. Local passive methods require an additional dissemination mechanism in order to communicate the capacity information to other network nodes which adds complexity and can be unreliable under adverse network conditions. On the other hand, remote passive methods do not require a dissemination mechanism and so can be simpler to implement and also do not suffer from communication reliability issues. Many applications (e.g. ANDSF etc) can benefit from utilizing this capacity information. Therefore, in this thesis we propose a new remote passive *Capacity Utilization* estimator performed by neighbour nodes. However, there will be an error associated with the measurements owing to the differences in the wireless medium as observed by the different nodes' location. The main undertaking of this thesis is to address this issue. An error model is developed to analyse the main sources of error and to determine their impact on the accuracy of the estimator. Arising from this model, a number of modifications are implemented to improve the accuracy of the estimator. The network simulator ns2 is used to investigate the performance of the estimator and the results from a range of different test scenarios indicate its feasibility and accuracy as a passive remote method. Finally, the estimator is deployed in a node saturation detection scheme where it is shown to outperform two other similar schemes based upon queue observation and probing with ping packets.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

AB	Available Bandwidth
AC	Access Category
ACK	Acknowledgement
AEF	Access Efficiency Factor
AIFSN	Arbitration Interframe Spacing Number
ANDSF	Access Network Discovery and Selection Function
AODV	Ad Hoc On-Demand Distance Vector Routing
AP	Access Point
ARE	Absolute Relative Error
BC	Backoff Counter
BER	Bit Error Rate
BPSK	Binary Phase-Shift Keying
BSA	Basic Set Area
BSS	Basic Service Set
BSSID	Basic Service Set Identification
CCK	Complementary Code Keying
CDF	Cumulative Distribution Function
CQ	Connection Quality
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance

CSMA/CD	Carrier Sense Multiple Access/Collision Detection
CTS	Clear to Send
CU	Capacity Utilization
CUAR	Capacity Utilization Aware Routing
CW	Contention Window
CW_{max}	Maximum Contention Window Size
CW_{min}	Minimum Contention Window Size
DBPSK	Differential Binary Phase-Shift Keying
DCF	Distributed Coordination Function
DDoS	Distributed Denial-of-Service
DIFS	DCF Interframe Space
DSDV	Destination-Sequenced Distance-Vector Routing
DSR	Dynamic Source Routing
DSSS	Direct-Sequence Spread-Spectrum
DQPSK	Differential Quadrature Phase-Shift Keying
ECW_{max}	Exponent Form of CW_{max}
ECW_{min}	Exponent Form of CW_{min}
EDCA	Enhanced Distributed Channel Access
EIFS	Extended Interframe Space
ESS	Extended Service Networks

FAR	False Alarm Ratio
FCFS	First-Come First Served
FDR	Failed Detection Ratio
FIFO	First In First Out
FHSS	Frequency-Hopping Spread-Spectrum
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
HR/DSSS	High-Rate Direct Sequence Spread-Spectrum
IBSS	Independent Basic Service Set
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
ISM	Industrial, Scientific, and Medical
LAN	Local Area Network
LLC	Logical Link Control
MAC	Medium Access Control
MAN	Metropolitan Area Network
ModCU	Modified Capacity Utilization Estimator
MIMO	Multiple-Input Multiple-Output
NAV	Network Allocation Vector
ID	Identity

IDS	Intrusion Detection System
ISMP	Inter-System Mobility Policy
ISRP	Inter-System Route Policy
NAV	Network Allocation Vector
NS2	Network Simulator Version 2
NonModCU	Non-Modified Capacity Utilization Estimator
OFDM	Orthogonal Frequency Division Multiplexing
OLSR	Optimized Link State Routing
OSI	Open Systems Interconnection
OTcl	Object Tcl
OWD	One-Way Delay
PCF	Point Coordination Function
PDM	Packet Delay Method
PDF	Probability Distribution Function
PGM	Packet Gap Method
PHY Rate	Physical Layer Bit Rate
PIFS	PCF Interframe Space
PLCP	Physical Layer Convergence Protocol
PRM	Packet Rate Method
PSDU	PLCP Service Data Unit

QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
RSSI	Received Signal Strength Indication
RTS	Request to Send
RTT	Round-Trip Delay Time
RV	Random Variable
SIFS	Short Interframe Space
SIR	Signal to Interference Ratio
SLoPS	Self-Loading Periodic Streams
SNR	Signal to Noise Ratio
SSID	Service Set Identifier
3GPP	Third Generation Partnership Project
Tcl	Tool Command Language
TclCL	Tcl with classes
TCP	Transmission Control Protocol
TOPP	Trains of Packet Pairs
TTL	Time-to-Live
UDP	User Datagram Protocol

UE	User Equipment
UNII	Unlicensed National Information Infrastructure
VoIP	Voice over Internet Protocol
VPS	Variable Packet Size
Wi-Fi	Wireless Fidelity
WLANs	Wireless Local Area Networks
WMN	Wireless Mesh Network
WIMAX	Worldwide Interoperability for Microwave Access

Chapter 1 Introduction

The wireless local area network (WLAN) based on the IEEE 802.11 standard is a popular data transmission system that provides wireless communications for users operating in the 2.4 GHz or 5 GHz ISM (Industrial, Scientific, and Medical) bands [1]. The accurate measurement of throughput-related concepts [2] such as capacity, available bandwidth and other metrics can be used to more effectively achieve the optimization of wireless network services for many applications. In wired networks, the definition of capacity that is widely accepted is the maximum possible transmission rate can be achieved on a link [2]. However, this definition and many of the proposed estimation techniques cannot be applied directly to WLANs due to the shared nature of the medium under the IEEE 802.11 MAC mechanism, fading and interference, and varying link quality. Consequently, the capacity of a WLANs node will not be fixed and depends on what the node and other nodes that it shares the medium with are doing.

1.1 Motivation

Currently, the various schemes proposed for capacity estimation in WLANs can be divided into two categories. One category is active approaches based upon the transmission of probe packets. This active probing method uses a series of probe packets transmitted at a number of different traffic rates [3, 4] to estimate the capacity of the channel. However, this approach consumes the bandwidth of the channel which can have a negative impact on the performance of a network due to the increased contention on the medium. Moreover, it does not scale well due to the extra network traffic generated which can affect the accuracy of the estimation. The other category includes passive techniques based upon analyzing the transmitted packets on the medium to determine the available capacity. Passive approaches perform with a higher accuracy than active approaches [5-7]

and have no overhead. However, many factors that influence the accuracy of the estimation in this approach need to be taken into account such as contention, collision probability, retransmission, and hidden nodes etc.

There are two main measurement methods adopted in the passive approaches: local measurement and remote observation. In local measurement, a node monitors the channel and estimates the available capacity and then broadcasts this information to its neighbour nodes to support various wireless applications (e.g. Quality of Service (QoS) aware routing [8] and admission control [9]). This mechanism increases the overhead of networks and makes the applications more complex. In remote observation, a node captures and analyses the transmitted packets within its reception range to estimate its neighbours' available capacity directly. Moreover, remote observation does not require an additional dissemination method and is more reliable compared to local measurement approaches.

In this thesis, we combine the advantages of the passive technique and remote measurement in order to propose a *Capacity Utilization* estimator based upon remote observations performed by neighbour nodes. The *Capacity Utilization* is defined as the ratio of a node's traffic load and its node capacity. This *Capacity Utilization* metric reflects the usage of the node capacity during a specified measurement interval.

Once the *Capacity Utilization* of a WLAN node can be estimated, many wireless applications can benefit from utilizing this information. An important application is to support the access point (AP) selection mechanism in an access network discovery and selection function (ANDSF) [10], e.g. where a mobile user enters a Wi-Fi hotspot zone where there are multiple APs present. The traditional metric for the user is to select an AP based upon the received signal strength indication (RSSI) which is dependent only on the relative locations of the user and the APs and does not provide any AP performance

information. However, this may lead to a substantial load imbalance [11] which results in high contention in the medium, an overload of the AP with a possible resulting AP saturation or congestion; and hence a deterioration in network performance. Our *Capacity Utilization* estimator is a more useful metric on which to base the ANDSF rather than RSSI. It provides more relevant AP information such as the number of clients associated with the AP, the traffic load of the AP and its neighbour nodes, the contention experienced by the AP and the *Capacity Utilization* of the AP.

In order to ensure fair load balancing, improve the user's throughput, and enhance the utilization of network resources, the IEEE 802.11k standard [12] employs the local measurement of channel information (e.g. utilization) and reports the AP's information in a special management frame after receiving the user's request. An intelligent AP selection mechanism for handoff incorporating our remote *Capacity Utilization* estimator does not require any message exchange mechanism and considers hidden nodes. Moreover, it can estimate the *Capacity Utilization* of all APs within its reception range directly and promptly before the association phase takes place. Furthermore, our estimator can provide better available AP selection and handoff decisions owing to its awareness of contention, traffic loads of the network nodes, the available capacity and *Capacity Utilization* of the AP.

In addition, the measurements provided by the *Capacity Utilization* estimator could be employed as a route metric to support routing decisions (i.e. to realise resource aware routing) thereby allowing the network nodes to select their next hop directly by using the neighbours' *Capacity Utilization* measurement to guarantee the QoS performance and avoid saturation or congestion. The network nodes also can monitor their neighbour's *Capacity Utilization* to autonomously select the appropriate operating channel when the neighbour node is operating under heavy load conditions in multi-radio or multi-channel

networks. Moreover, estimating each neighbour node's *Capacity Utilization* within an admission control scheme can ensure that sufficient bandwidth can be allocated to each node to satisfy its QoS requirements.

Compared with active probing and local measurement methods, our remote *Capacity Utilization* estimator is more reliable and can support multiple wireless applications. However, there will be an error associated with this estimation owing to the differences in the wireless medium as observed by the different nodes. As the observer node and the observed node do not experience the same medium, the estimation of the *Capacity Utilization* value performed by the observer node will differ from that experienced by the observed node itself. The errors associated with our *Capacity Utilization* estimator are unavoidable but they can be minimized after the modifications based on a number of reasonable assumptions.

1.2 Framework of the Thesis

This thesis is concerned with developing a remote estimator for *Capacity Utilization* based upon remote observations performed by neighbour nodes in WLANs. This scheme is based upon a temporal analysis framework that models the way in which the IEEE 802.11 MAC mechanism wins transmission opportunities, passively analyses the transmitted packets and extends the existing MAC bandwidth components model [13, 14] to determine the neighbours' *Capacity Utilization*.

This thesis is essentially a study of the performance of the estimator, i.e. how accurately the estimator can measure the actual *Capacity Utilization* experienced by a node. An error model is proposed to analyse the error associated with the *Capacity Utilization* estimation and the impact of this error on the accuracy of the estimator. Therefore, this analysis will necessarily involve a statistical characterisation of the error associated with the estimate

of *Capacity Utilization* produced by the remote estimator. The main sources of error in the estimation of *Capacity Utilization* are the error in the calculation of the neighbour load and the error in the calculation of the contention experienced by the observed neighbour nodes. The main factors that affect the accuracy of estimations are the number of unobservable neighbours of the observed node and the traffic load of the unobservable neighbour nodes.

In order to minimize the error associated with the *Capacity Utilization* estimation, three simple and reasonable assumptions related to the aspects of the traffic load of the unobserved nodes, the underestimation of contention and failed transmissions are introduced to modify the estimator and improve its accuracy.

Detecting node saturation is an application used to illustrate the usage and accuracy of the *Capacity Utilization* estimator. This thesis compares its performance with two other detection algorithms: a queue observation method and a regularly pinging method.

1.3 Contributions

The main contributions of this thesis are the development of WLAN *Capacity Utilization* estimator based upon remote observations performed by neighbour nodes. The specific contributions are listed as follows:

- Introduces the concepts of node capacity and *Capacity Utilization*, the challenges, benefits, and potential applications of developing a node *Capacity Utilization* estimator based upon remote observations by neighbour nodes. Presents the specific methodology of the *Capacity Utilization* estimator based upon remote observations by neighbour nodes involving passively monitoring and analyzing the transmitted frames.
- Proposes a model of the error associated with the *Capacity Utilization* estimation,

analyses the sources of errors, investigates and identifies the factors that influence the accuracy of the *Capacity Utilization* estimator such as the number of neighbours, the number of observable neighbours, network traffic load and traffic type.

- Modifies the *Capacity Utilization* estimator to improve the performance and validate the feasibility and accuracy of the modifications under different simulation scenarios.

1.4 Thesis Outline

The remainder of this thesis is organized as follows:

Chapter 2 provides the background knowledge to the work such as the basic concepts of WLANs, the IEEE 802.11 MAC mechanism, the concepts of node capacity and *Capacity Utilization*, the challenges in estimating the node *Capacity Utilization*, the benefits and possible applications arising from being able to estimate node *Capacity Utilization.* It also describes the ns2 simulator used in this thesis.

Chapter 3 discusses the related research in the areas of the measurement of node capacity and other throughput-related metrics, other proposed algorithms, performance evaluation of these estimation techniques, and some methods proposed for utilizing capacity information in different wireless applications.

Chapter 4 presents a detailed description and explanation of the remote node *Capacity Utilization* estimator. The analysis of the error associated with the *Capacity Utilization* estimator and the modifications to the estimator to minimize the error is also presented in this chapter. Moreover, the algorithm for node saturation that combines a Bayesian decision process is also described here.

Chapter 5 investigates the performance evaluation of our *Capacity Utilization* estimator

and identifies the factors that determine the accuracy of the estimator and implements modifications to improve the accuracy. A comparison of the simulation results is presented to evaluate the performance of the remote *Capacity Utilization* estimator compared to other two algorithms (i.e. queue observation method and regularly pinging method) in detecting node saturation.

Chapter 6 summarizes the conclusions of this thesis and outlines possible future research work in the area of capacity estimation in IEEE 802.11 WLANs.

Chapter 2 Technical Background

This chapter introduces the relevant technical background material for this thesis. This work is concerned with developing an estimator for the node *Capacity Utilization* using remote observations performed by neighbour nodes in WLANs. This estimator has been designed to operate with the IEEE 802.11 DCF method and is intended for network nodes with multiple-neighbours, i.e. infrastructure networks, wireless mesh networks or wireless ad-hoc networks operating under the IEEE 802.11b/a/g/n standards. This *Capacity Utilization* estimator can be employed in many applications such as AP selection, routing, node saturation detection, channel selection, admission control, and QoS provision. This thesis uses the detection of node saturation as one of its potential applications in order to investigate the performance of the estimator. The neighbour nodes performing the remote estimation can broadcast the saturation information to other network nodes in beacon frames so that they may take preventative actions to avoid further deterioration in the node's saturation condition.

2.1 Wireless Local Area Networks

With the rapid development of information technology, especially with the widespread use of portable computers, smartphones, tablets and other wireless products, the traditional fixed Ethernet cannot satisfy the users who need more and more communication services anytime and anywhere.

A WLAN is a convenient data transmission system which provides connectivity and communications to wireless devices through employing radio frequency (RF) techniques rather than traditional wired networks based upon cables. It has been globally adopted due to its mobility, flexibility, low cost, ease of deployment [15]. The IEEE 802.11 WLAN standard is a member of the IEEE 802 family which is a series of specifications for local

area network (LAN) and metropolitan area network (MAN) technologies. The IEEE 802 standards define a MAC and a PHY component. The MAC represents a set of rules to determine how to send data and access the medium, and the PHY is responsible for transmission and reception of the digital signals. The IEEE 802.3 defines the Carrier Sense Multiple Access/Collision Detection (CSMA/CD) protocol [16] and is related to Ethernet networks, IEEE 802.5 is the specification of Token Ring access method [17] and IEEE 802.2 defines Logical Link Control (LLC) protocol [18]. The IEEE 802.11 group of standards developed by the IEEE LAN/MAN Standards Committee (IEEE 802.11) specifies the technologies for wireless LANs and the original version of the IEEE 802.11 standard was published in 1997 [1] and has had numerous amendments since. The base IEEE 802.11 standard comprises the IEEE 802.11 MAC and two physical layers operating in the 2.4 GHz ISM band: a frequency-hopping spread-spectrum (FHSS) physical layer and a direct-sequence spread-spectrum (DSSS) link layer.

2.1.1 The IEEE 802.11 Family

The IEEE 802.11 working group has many family members some of which are listed below:

Table 2.1 Some Family Members of the IEEE 802.11 Standard

IEEE standard	Notes
802.11a (1999)	Operates in the 5 GHz ISM band with a maximum PHY rate 54 Mbps
802.11b (1999)	Operates in the 2.4 GHz ISM band with a maximum PHY rate 11 Mbps
802.11ac (Still under development)	High-throughput WLAN operated in the 5 GHz ISM band
802.11ad (Published in 2010)	Supports a maximum date rate of 7 Gbps and operates in 60 GHz ISM band
802.11d (2001)	Additional regulatory domains
802.11e (2005)	Quality of service (QoS) enhancements for the MAC
802.11f (withdrawn 2006)	Inter-access point protocol
802.11g (2003)	Operates in the 2.4 GHz ISM band with a maximum PHY rate of 54Mbps
802.11h (2004)	Standard to make IEEE 802.11a compatible with European regulations in the 5 GHz ISM band
802.11i (2004)	Improvements to security
802.11k (2008)	Radio resource measurement enhancements
802.11n (2009)	Higher throughput improvement using MIMO and packet aggregation
802.11p (2010)	Adopting IEEE 802.11 for use in vehicular environment
802.11s (2011)	Enhancing IEEE 802.11 for use in mesh networks
802.11v (2011)	Standard to support wireless network management
802.11w (2009)	Protecting management frames

IEEE 802.11b: IEEE 802.11b was ratified in 1999. It amends the original IEEE 802.11 wireless network specification and extends the maximum physical layer bit rate (PHY rate) to 11 Mbps. It expands the application areas of wireless LAN and allows for wireless network functions comparable to Ethernet in the 2.4 GHz ISM band [19]. It can also drop its PHY rate to 5.5 Mbps, then to 2 Mbps, then to 1 Mbps under poor RF conditions. The IEEE 802.11b specifies a high-rate direct sequence spread-spectrum (HR/DSSS) physical layer and uses complementary code keying (CCK) modulation scheme. The CSMA/CA mechanism was introduced in the original IEEE 802.11 (1997) standard.

IEEE 802.11a: IEEE 802.11a is an amendment to the IEEE 802.11 standard ratified in 1999 [20] and operates in the 5 GHz unlicensed national information infrastructure (UNII) band. It specifies a physical layer based on orthogonal frequency division multiplexing (OFDM) with 52 subcarriers as the modulation scheme which can be BPSK, QPSK, 16-QAM or 64-QAM scheme. The maximum PHY rate is 54 Mbps and can be reduced to 48, 36, 24, 18, 12, 9 and 6 Mbps. It cannot interoperate with IEEE 802.11b due to separate frequency bands unless the device has a dual band capacity.

IEEE 802.11g: As WLAN devices became more widely used and users demanded higher data transmission rate and appropriate modulation techniques to avoid multi-path effect, fading. IEEE 802.11g was introduced as a further higher PHY rate extension to the original IEEE 802.11 specification in June 2003 [21]. It operates in the same 2.4 GHz ISM band to maintain backward compatibility with IEEE 802.11b standard with a PHY transmission rate of up to 54 Mbps. IEEE 802.11g uses an OFDM modulation scheme to support transmission rates of 6, 9, 12, 18, 24, 36, 48 and 54 Mbps, a CCK modulation scheme for 5.5 and 11 Mbps, a DQPSK/DSSS modulation scheme for 2 Mbps and a DBPSK/DSSS modulation scheme for 1 Mbps.

IEEE 802.11n: The MIMO-OFDM technique is widely used in various kinds of wireless products. The multiple-input multiple-output (MIMO) technique employs multiple antennas to transmit synchronously at the sender, and the receiver also provides multiple antennas to address the problem of multi-path fading. IEEE 802.11n was officially ratified in 2009. It is an amendment to IEEE 802.11 which adds the MIMO technique, 40 MHz channels to transmit data in PHY layer and block acknowledgement, packet aggregation to enhance the efficiency in MAC layer. It also increases the maximum PHY rate from 54 Mbps in IEEE 802.11a to 600 Mbps and supports the use of smart antennas technology [22].

IEEE 802.11e: IEEE 802.11e is a wireless standard that defines a set of Quality of service supports [23]. IEEE 802.11e specifies QoS components and multimedia support that are backward compatible with IEEE 802.11b and IEEE 802.11a standards. IEEE 802.11e also supports two new channel access mechanisms: enhanced distributed channel access (EDCA) which extends DCF and HCF (hybrid coordination function) controlled channel access (HCCA) which extends Point coordination function (PCF). EDCA defines different parameters: the arbitration interframe spacing number (*AIFSN*[AC]), ECW_{min} (exponent form of CW_{min}) and ECW_{max} (exponent form of CW_{max}) [24] in order to improve the access to the medium and reduce the delay of high priority communications. There are four access categories (ACs) defined that correspond to four different traffic classes (based upon different priorities). The priority level of ACs from low to high is: background traffic (AC_BK), best-effort traffic (AC_BE), video traffic (AC_VI) and voice traffic (AC_VO).

In this thesis, the estimator of node *Capacity Utilization* can be used in networks operating under the IEEE 802.11 b/a/g/n standards using the standard DCF mechanism. Our estimator will be extended for IEEE 802.11e operation in the future work.

2.1.2 WLAN Components

An IEEE 802.11 network consists of several principal components [15] as follows:

Access Points (APs): Access points are essentially the base stations of WLANs and perform the wireless-to-wired bridging function. They provide wireless connections by using a radio frequency link for wireless enabled devices to communicate with and convert the IEEE 802.11 frame to another type for delivery to other networks.

Clients: Clients are computing devices which contain a wireless adapter card to provide the wireless connectivity, such as a laptop, tablet computer, smartphone and other wireless products. In certain circumstance (i.e. in order to avoid pulling new cable lines), desktops or other fixed devices can be equipped with a wireless interface to connect to a WLAN.

Essentially, all APs and Clients devices are referred to as network nodes in this thesis.

Wireless Medium: The standard uses a wireless medium to transmit frames from node to node and defines different physical layers that include two RF physical layers and one infrared physical layer.

Distributed System: The distributed system which consists of a distributed system medium and a bridging engine is the logical component of IEEE 802.11 networks. It is also called a backbone network (also uses Ethernet) used to forward frames to their destination or between access points.

Basic Service Set (BSS): BSS is the basic building block in an IEEE 802.11 wireless network which consists of a set of nodes that can communicate with each other in its coverage area, called a Basic Set Area (BSA) and is limited by the propagation characteristics of the wireless medium. Each BSS has its own unique 48-bit binary

identifier called the Basic Service Set Identification (BSSID) to identify different BSSs. IEEE 802.11 defines two operating modes: Infrastructure mode and ad hoc mode.

In the ad hoc mode, also called the peer to peer mode as shown in Figure 2.1, clients can communicate with each other directly without the need of an AP. At least two wireless clients need to be configured to use ad hoc mode in order to form an Independent BSS (IBSS), one of them can play a master role and take over some of the responsibility of an AP. The infrastructure mode is distinguished from the ad hoc mode by using an AP as a central controller which is in charge of all communications and includes the functions of relaying and connecting to a wired Ethernet within its BSS. A client can be associated with only one AP at any given time.

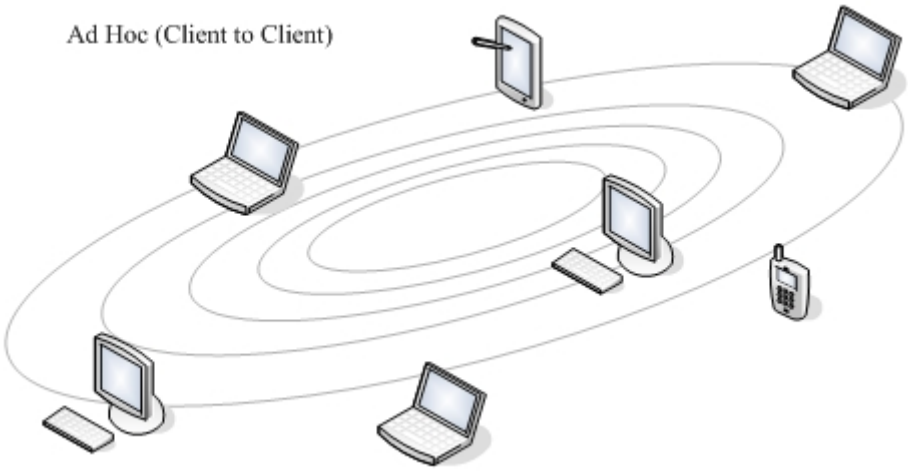

Figure 2.1: An Example of an Ad Hoc Network

Extended Service Networks (ESS): IEEE 802.11 allows a set of two or more BSSs to be interconnected to form an extended service networks (ESS) by a backbone network in order to extend the coverage of a wireless network. All the APs use the same service set identifier (SSID) in an ESS. The nodes in different BSSs but within the same ESS can communicate with each other and even move between different BSSs.

2.1.3 Wireless Mesh Networks

IEEE 802.11 WLANs still rely on wired networks to provide the backhaul connection to the network. Unfortunately this dependency is costly and inflexible. A centralized structure and fixed topologies also limit the performance of some network applications [25] which need the peer-to-peer connectivity or require choosing a better path for communication. A Wireless Mesh Network (WMN) is an alternative network topology with multiple-neighbours which resolves this problem and provides broadband wireless Internet services to a large community of users, i.e., community networks, campus networks, high speed metropolitan area networks, and enterprise networks. It resolves the limitations and significantly improves the performance of ad hoc WLAN networks. These features brings many advantages for the client, such as robustness, low cost, easy to deploy, flexible wireless service and higher bandwidth [26] to mobile users.

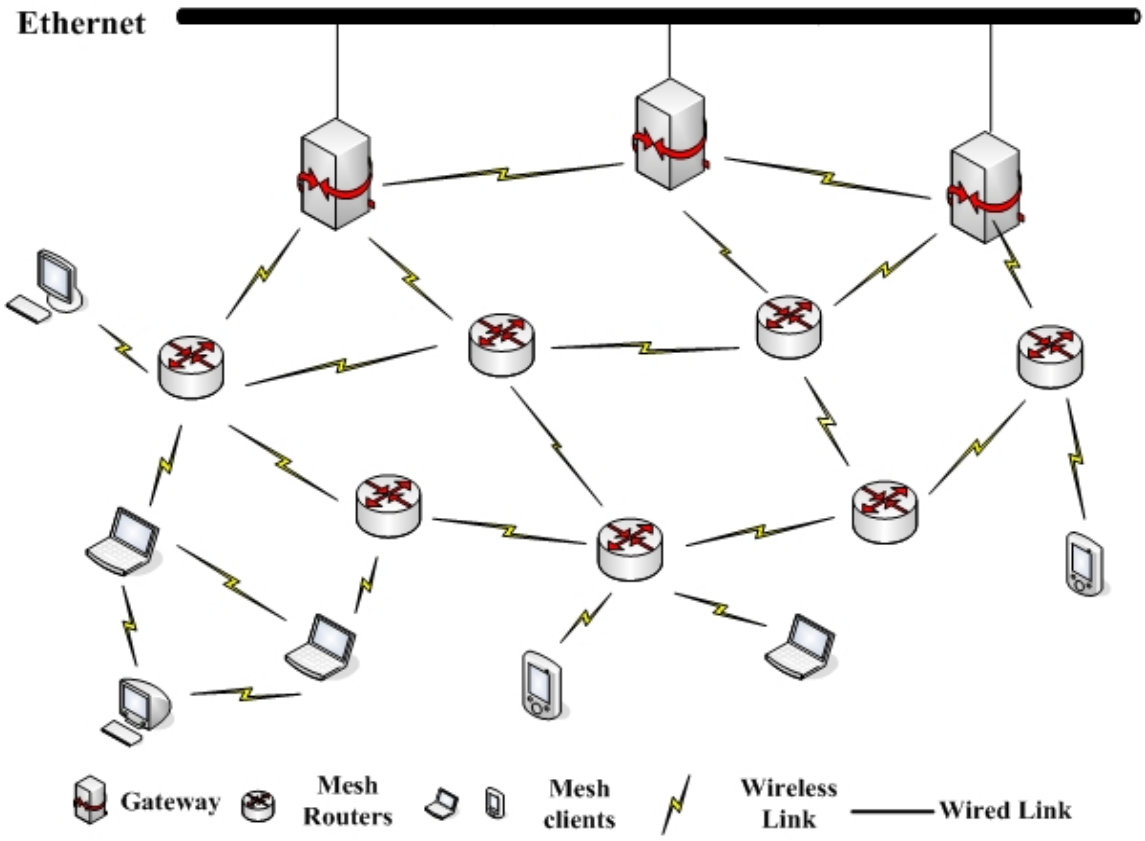

Figure 2.2: Architecture of a Wireless Mesh Network

A typical WMN consists of mesh gateways, mesh routers (mesh nodes) and mesh clients as shown in Figure 2.2. Mesh gateways are special wireless routers with a high-bandwidth wired connection to the Internet. The fixed mesh routers are equipped with power lines, multiple wireless interfaces, high processing and memory capability [27] to form the wireless backbone. Mesh clients access the network through mesh routers as well as directly communicating with each other. Mesh clients can be mobile nodes. In this thesis, the estimator of the node *Capacity Utilization* is suitable for networks which have multiple neighbours, i.e. infrastructure networks, ad hoc networks, and wireless mesh networks.

2.2 Fundamentals of the IEEE 802.11 MAC Mechanism

The WLAN data link layer is divided into two parts: A Logical Link Control (LLC) sub-layer and a Medium Access Control (MAC) sub-layer. The LLC defined by the IEEE 802.2 is the upper layer of the data link layer and acts as the unified interface to the network layer, and the MAC is the lower layer of data link layer and acts as the interface to the physical layer. The MAC is the key mechanism of the IEEE 802.11 specification that controls the transmission of user data and the core framing operations. This section introduces the IEEE 802.11 MAC Mechanism. Coordination functions control the node access to the wireless medium to avoid collisions. The Distributed Coordination Function (DCF) is the basis of the CSMA/CA mechanism which checks that the wireless link is clear before transmitting the frame, similar to Ethernet. The network nodes use a random backoff mechanism to avoid a collision. The *Capacity Utilization* estimator is based upon the DCF access mode where it captures and processes the frames transmitted by all nodes within its reception range.

2.2.1 Hidden Nodes Problem

IEEE 802.11 provides the addressing and the channel access mechanism to allow different nodes communicate with each other and to avoid collisions by using the CSMA/CA mechanism. The CSMA/CA mechanism is similar to the CSMA/CD protocol in Ethernet defined by the IEEE 802.3 standard. Both mechanisms support multiple users in sharing the medium. CSMA/CD uses a carrier sense scheme where the node waits until the medium becomes idle. If a collision occurs while transmitting the frame, the node uses a collision detection (CD) procedure to stop transmitting and send a jam signal in order to let all the nodes on the shared medium be aware of the occurrence of a collision. It then waits for a random time interval to re-transmit the frame. However, collision detection cannot be realised in WLANs because the radios operate in half-duplex mode.

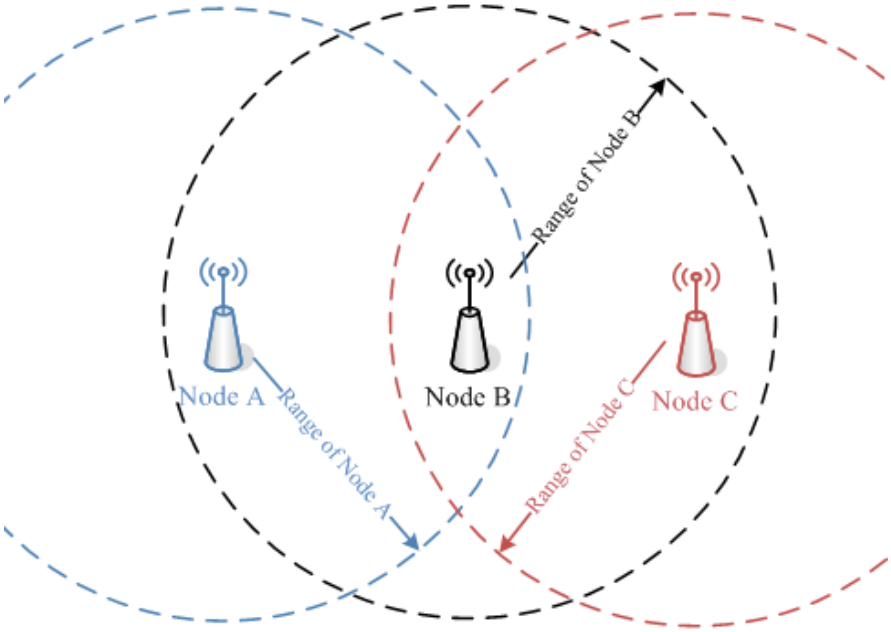

Figure 2.3: Node A and Node C are “Hidden” from each other

The hidden nodes problem is a big problem that affects the performance of WLANs. If two or more nodes are within the communication range of an AP but they cannot sense and communicate with each other due to different transmit powers, distance or locations, this gives rise to the “hidden node” problem. In Figure 2.3, node *B* can communicate with

node *A* and node *C*, but node *A* and node *C* cannot communicate with each other directly. For node *A*, node *C* is a "hidden node". If node *A* and node *C* simultaneously transmit a frame then a collision will occur at receiver node *B*. It is difficult to detect the collisions resulting from hidden nodes in WLANs because wireless transceivers usually operate in half-duplex mode and so cannot receive and transmit at the same time.

In order to avoid the collisions that can arise from hidden nodes, the CSMA/CA mechanism supports the use of Request to Send (RTS) and Clear to Send (CTS) frames to clear out the transmission medium before transmitting a frame. However, this mechanism has been proposed under the assumption that the hidden nodes are within the reception range of the receivers [28] (i.e. node *B*) and causes extra overhead to the network. Therefore, the RTS/CTS mechanism is not enabled unless the frame size exceeds a predefined threshold.

2.2.2 Interframe Spacing

The interframe spacing plays a significant role in the coordination function in order to generate different priority levels for different types of traffic. The relationship between the different interframe spaces and medium access method is shown in Figure 2.4.

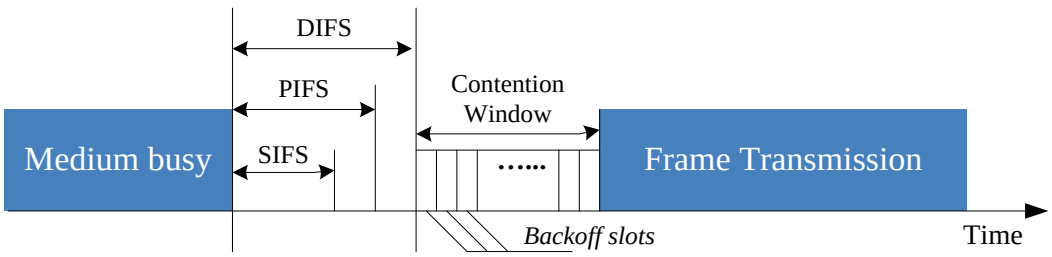

Figure 2.4: Basic Interframe Spaces and Medium Access Method

Short Interframe Space (SIFS): SIFS is the shortest interframe spacing and provides the highest-priority transmission for some control frames such as ACK frame, CTS response

frame, fragmentation frames and their acknowledgements to ensure that the node can occupy the channel during the fragmentation burst.

PCF Interframe Space (PIFS): PIFS is used to support contention-free services and has a higher priority than other contention-based traffic. Nodes can transmit data after the PIFS has elapsed once the medium is detected as idle.

$$PIFS = SIFS + SlotTime \tag{2.1}$$

DCF Interframe Space (DIFS): DIFS is used to support contention-based services. If the medium is continuously idle for a period of DIFs or longer, the node can access the medium immediately. DIFS is calculated from the following formula:

$$DIFS = SIFS + 2 \times SlotTime \tag{2.2}$$

Extended Interframe Space (EIFS): EIFS is not a fixed time interval and is only used following a corrupted frame transmission.

All the interframe spaces except for EIFS are constant values and have different values for the different standards shown in Table 2.2. It is to be noted here that IEEE 802.11g uses a *SlotTime* of 20 μs for backward compatibility with IEEE 802.11b [29] and uses a *SlotTime* of 9 μs in an IEEE 802.11a or pure IEEE 802.11g mode only.

Table 2.2 Interframe Spaces in the Different IEEE 802.11 Standards

Standard	*SlotTime*(μs)	SIFS(μs)	PIFS(μs)	DIFS(μs)
802.11b	20	10	30	50
802.11a	9	16	25	34
802.11g	9 or 20	10	19 or 30	28 or 50

Arbitration Interframe Space (AIFS): In order to support QoS, IEEE 802.11e introduces new interframe space AIFS to differentiate the traffic priority as shown in Figure 2.5 where AC_i, AC_j and AC_k represent the different access categories. The value of AIFS is given by equation (2.3) where *AIFSN* is the arbitration interframe space number.

$$AIFS = SIFS + AIFSN \times SlotTime \tag{2.3}$$

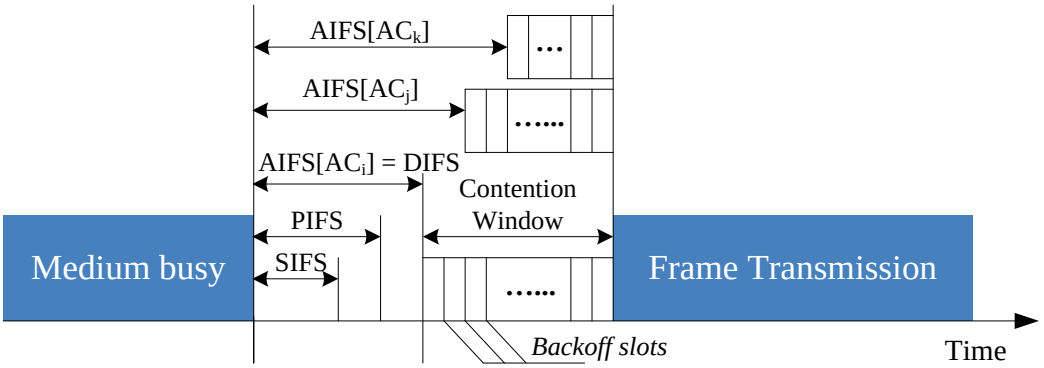

Figure 2.5: Arbitration Interframe Spaces under the IEEE 802.11e Standard

2.2.3 Contention-Based Access Using the Distributed Coordination Function

Most transmitting operations employ the DCF and provide contention-based access to the medium for contending nodes without the need for a central controller. It can be used in both IBSS and infrastructure networks. When a node attempts to transmit its data, it must first check whether the wireless medium is idle or not. If the medium is busy, the node must defer (i.e. wait) and use a binary exponential backoff algorithm to win a transmission opportunity. At this point, the node generates a random backoff interval before transmitting in order to minimize the probability of a collision [30] with frames being transmitting by other nodes.

This backoff interval is known as the contention window (*CW*) and is divided into fixed time slots whose value depends on the physical layer used.

$$Backoff\ Interval = Random() \times SlotTime \tag{2.4}$$

Where *Random()* is a random integer selected from [0, *CW*] and acts as a backoff interval counter and *CW* is the size of the contention window. When several nodes wish to transmit frames, if the medium is idle, all the nodes begin to decrement their backoff interval, the node whose backoff interval reaches zero first (i.e. has the smallest backoff interval) wins the opportunity to transmit. When the other nodes sense the medium becoming busy, they must stop decrementing their backoff interval and wait until the medium becomes idle again.

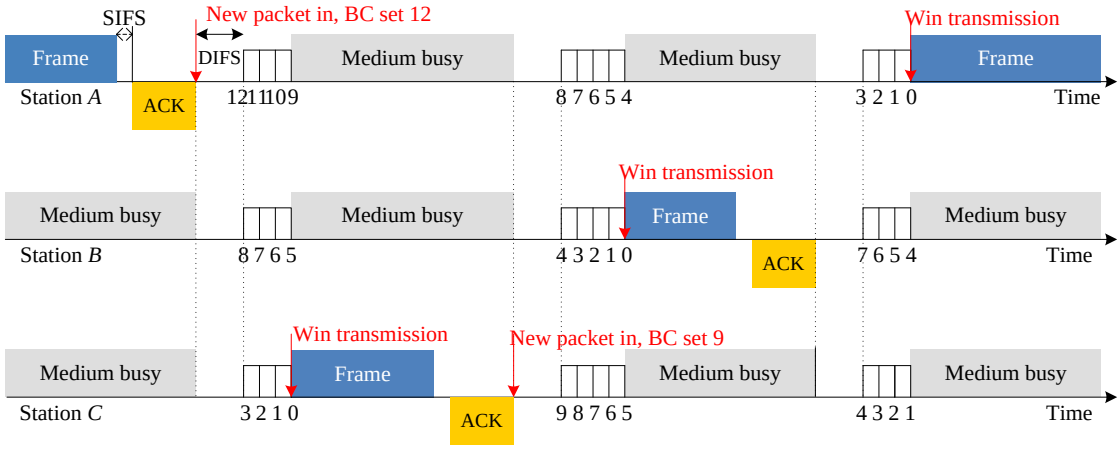

Figure 2.6: Contention-based Access Operations

Figure 2.6 shows three nodes *A*, *B* and *C* competing for access to the medium. When the channel becomes idle, nodes *A*, *B* and *C* must wait for an interval of DIFS and then generate a random backoff interval. Node *C* picks the smallest backoff interval to transmit its frame. When node *C* begins transmitting, the nodes *A* and *B* stop decrementing at values 8 and 4 respectively. After this frame transmission, node C resets a new backoff interval after a DIFS for the next transmission. Node *A* and *B* must wait for an idle interval of DIFS to elapse before resuming the decrementing of their backoff counters.

After the transmission of a frame, the receiver will return an ACK frame if it successfully receives this frame. If the frame is not received or received with an error, there is no ACK frame transmitted. If the sender does not receive an ACK, it will continue to retransmit

the frame until it is successful or else it will drop the frame if it exceeds the maximum number of retransmission permitted.

At the first transmission attempt, *CW* is set to the minimum contention window size (CW_{min}), after each unsuccessful retransmission *CW* size is doubled up to its maximum value (CW_{max}) as the number of unsuccessful retransmissions increases. If the frame is retransmitted successfully, the *CW* is reset to CW_{min}. If the retransmission counter reaches its permitted maximum value, the frame is dropped. Different physical layer protocols employ different CW_{min} and CW_{max} values. Figure 2.7 shows how the contention window increases under the DCF mechanism where CW_{min} is 31, and CW_{max} is 1023.

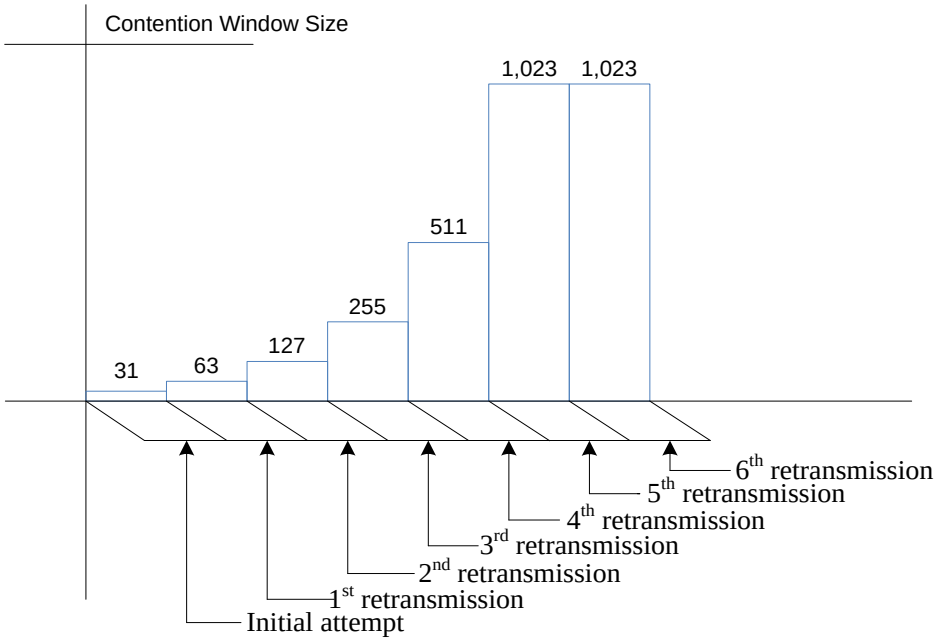

Figure 2.7: Contention Window Size under Multiple Retransmission Attempts

To expedite the transmission of higher-priority data in the IEEE 802.11 QoS enhancement scheme ECW_{min} and ECW_{max} can be set according to the traffic expected in each access category [23]. The contention window can be expressed as $2^{ECW} - 1$, where *ECW* is equal to ECW_{min} at the initial transmission. IEEE 802.11e employs four queues with different parameters (which can be found in Table 2.3) for the different access

categories. The network nodes must wait for an interval of AIFS before decrementing their backoff counter when the medium is busy, the AC whose backoff counter first reaches zero wins the transmission opportunity. Different ACs within the same AP will contend for access to the medium and well as competing with ACs in other APs. Currently, our *Capacity Utilization* estimator only considers the contention among nodes based upon the original DCF mechanism where *AIFSN* = 2, ECW_{min} = 5, ECW_{max} = 10. In the future work, the consideration of medium access among different ACs will be investigated and the estimator will be modified for the IEEE 802.11e protocol.

Table 2.3 The Default EDCA Parameters for Different ACs

Access Categories	CW_{min}	CW_{max}	AIFSN
AC_BK	CW_{min}	CW_{max}	7
AC_BE	CW_{min}	CW_{max}	3
AC_VI	$\frac{CW_{min}+1}{2}-1$	CW_{min}	2
AC_VO	$\frac{CW_{min}+1}{4}-1$	$\frac{CW_{min}+1}{2}-1$	2

2.2.4 IEEE MAC frame

The IEEE 802.11 standard defines various frame types for communications, managing and controlling the wireless link. All frames have a frame control field to describe the IEEE 802.11 protocol version, frame type and other indicators; and contain MAC addresses for the source address (except for some control frames) and destination address, frame sequence number, frame body and frame check sequence for error detection. There are three main types of frame specified [15]: Data frames which are used for data delivery, control frames which provide the necessary services such as area clearing, channel

acquisition, carrier-sensing maintenance and positive acknowledgement of received data to enhance the reliability in data transmission, and finally Management frames which enable nodes to establish and maintain communications such as authentication frames, association frames and re-association frames and so on. In this thesis, the neighbour nodes only capture the frames which contend for access to the medium which includes all data frames, all management frames, RTS frames, PS-POLL frames and CTS-to-self frames [29] (developed in IEEE 802.11g). The generic IEEE 802.11 MAC frame format and the frame control field components are shown in Figure 2.8. Moreover, the retransmission flag (i.e. the Retry Field in Figure 2.8) is important to calculate the load of failed transmitted frames due to a collision or transmission error. The retransmitted frames set this field to 1. This will be described further in Chapter 4.

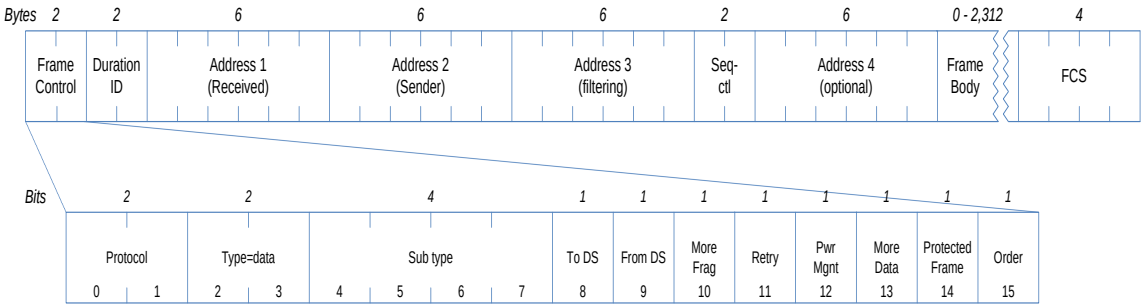

Figure 2.8: IEEE 802.11 MAC Frame Format and Frame Control Field

Beacon frames are transmitted at regular intervals by an AP in an infrastructure BSS to announce the presence of a network. In ad hoc networks, nodes also transmit beacon frames. Information elements which have their own element ID, length and variable-length component are variable-length components of management frames (Figure 2.9). An element ID (0-10, 16, 32-42, 48, 50, 221) is used for different information elements, other ID values are reserved and some can be extended for vendor specific applications. In this thesis, the *Capacity Utilization* information can be included as a metric in the

extended management information element in a beacon frame to support further network management actions.

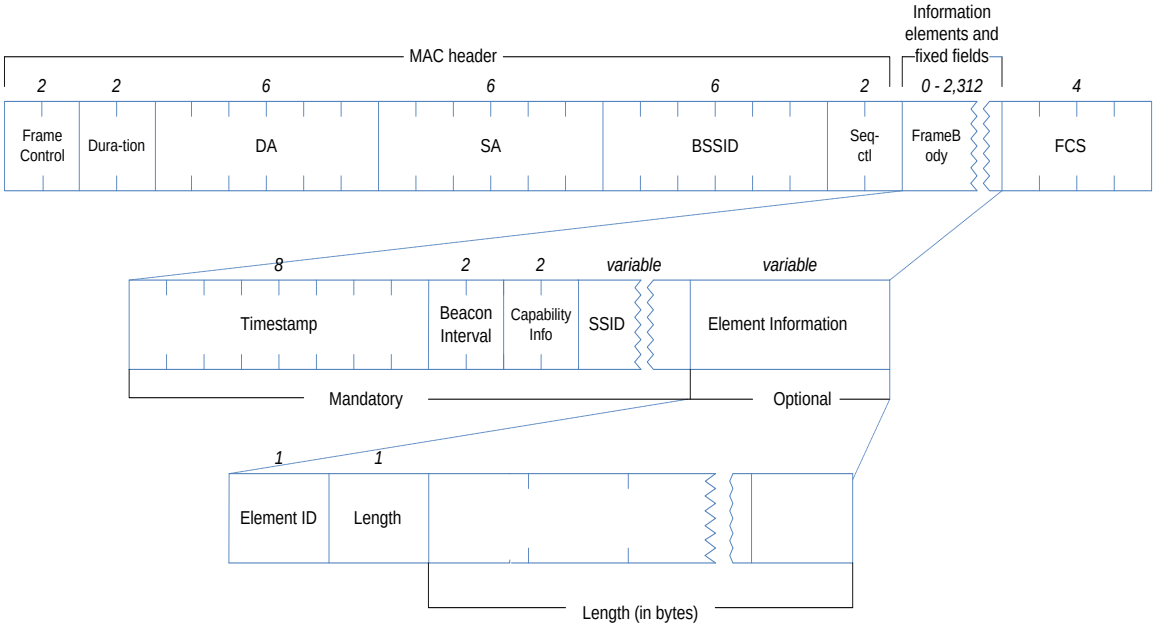

Figure 2.9: Structure of an IEEE 802.11 Beacon Frame

2.3 The Concept of Node Capacity and *Capacity Utilization* in WLANs

In digital and analogue communications, the bandwidth measured in hertz (Hz) refers to the range of frequencies used for transmitting a signal. In data networks, bandwidth is a fundamental resource and is quantified as the amount of data transferred per unit of time (usually one second). It is typically measured in bits per seconds (bps). The bandwidth available to the application directly impacts on the application performance. For example, multimedia streaming is often more sensitive to latency than throughput [2] and produces a better performance from the lower delay associated with a high-link bandwidth. The accurate estimation of the node capacity and available bandwidth can be used to more effectively achieve the optimization of wireless network services for many applications. The term bandwidth can be applied to a variety of throughput-related concepts such as capacity, available bandwidth, and achievable bandwidth. It should be mentioned here

that there is no universally accepted definitions for the concepts of capacity and available bandwidth, as different researchers tend to adopt different definitions for the analysis of their estimation methods (this will be discussed further in Chapter 3).

2.3.1 Capacity in Wired Networks

The most widely accepted definition of the capacity C_i of a wired link *i* between two nodes is the maximum possible transmission rate can be achieved on that link. The capacity of an end-to-end path *C* [2] is defined as the maximum transfer rate that a path can transfer data from a source to a sink. It can be defined according to (2.5), where *H* is the number of hops in the path. The capacity C_i of a wired link depends on the underlying transmission technology.

$$C = min_{i=1,\dots,H}\, C_i \tag{2.5}$$

Another metric is the available bandwidth A_i of a link that relates to the unused or spare capacity of the link during a specified time. This metric is a time-varying metric which depends on both the capacity C_i and traffic load of the link. The basic available bandwidth estimation formula can be written as [2]:

$$A_i = (1 - u_i) \times C_i, \text{ where } \bar{u}_i = \frac{1}{\tau}\int_{t-\tau}^{t} u_i(x)dx \tag{2.6}$$

Where $u_i(x)$ is the instantaneous utilization of the network link at time *x*, $\bar{u}_i$ is the average utilization for a time period *(t- τ, t)* and value τ is the measurement interval of interest for the available bandwidth. Once the available bandwidth can be estimated, many applications can benefit from this information, i.e. the sender node can adjust its outgoing traffic rate to avoid congestion, the network can configure its routing table for path selection to obtain an optimal routing or the traffic flow can re-route to satisfy its requirements, and the network can also control the admission of new traffic flows.

2.3.2 Capacity in Wireless Networks

Unlike a wired network where the link capacity is a constant value and can be referred to as the node capacity, the term “link” is ambiguous and is difficult to define in WLANs due to the shared nature of medium. Wired nodes use cables to connect to each other and the capacity can be assumed to be the maximum throughput that the link can deliver. In contrast, the WLANs nodes share the medium and use a collision avoidance method (i.e. the DCF scheme) which leads to a variability in the capacity of a WLAN node.

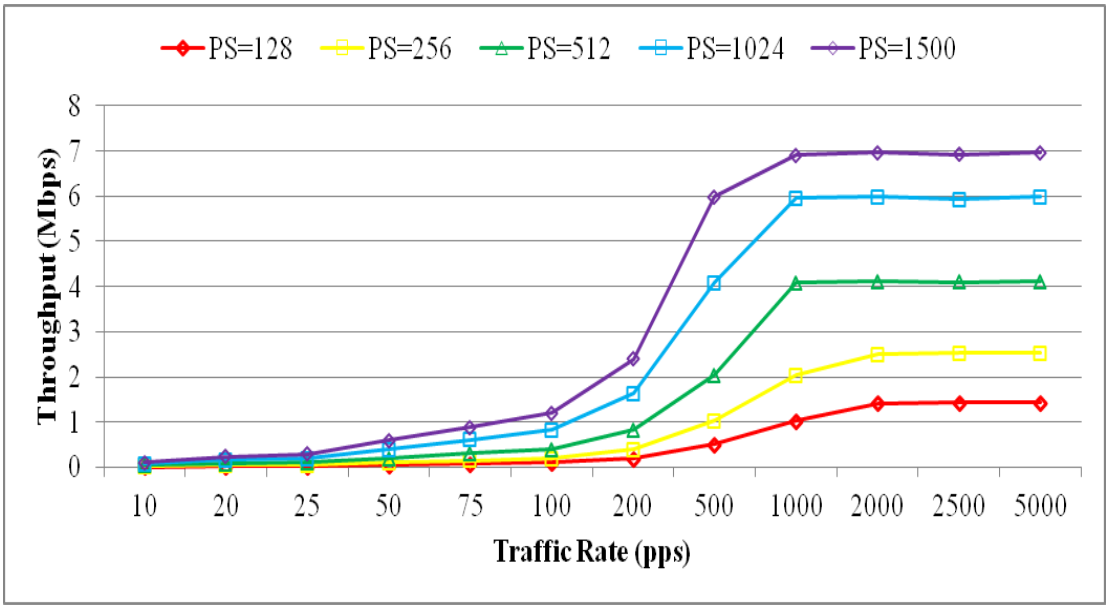

Figure 2.10: The Maximum Throughput for a Single Node WLAN

For example, consider Figure 2.10 where there is a network node sending traffic to a destination node in a WLAN. As the offered load increases, the throughput increases up to a maximum value which can be referred to as the “Saturation Throughput”. The result from this simple IEEE 802.11b simulation (the network environment parameters are detailed in Appendix A) indicates that the saturation throughput for a single node is related only to the frame size. The single network node sends UDP traffic with a frame size of 512 bytes using an 11 Mbps PHY rate and results in a saturation throughput of approximately 4 Mbps. Winning a transmission opportunity consumes bandwidth – in the

sense that the medium is required to be idle while the backoff counter decrements to zero. In transmitting a frame, a node has to first win a transmission opportunity (which consumes bandwidth) and then transmit the frame (which also consumes bandwidth). With the growth in the traffic frame size, the saturation throughput is increased. Once the frame size exceeds the predefined maximum payload, fragmentation occurs and splits the frame into smaller frames which decreases the maximum throughput. The simulation result shows that the maximum bandwidth depends on the frame size when there is only one node operating on the medium.

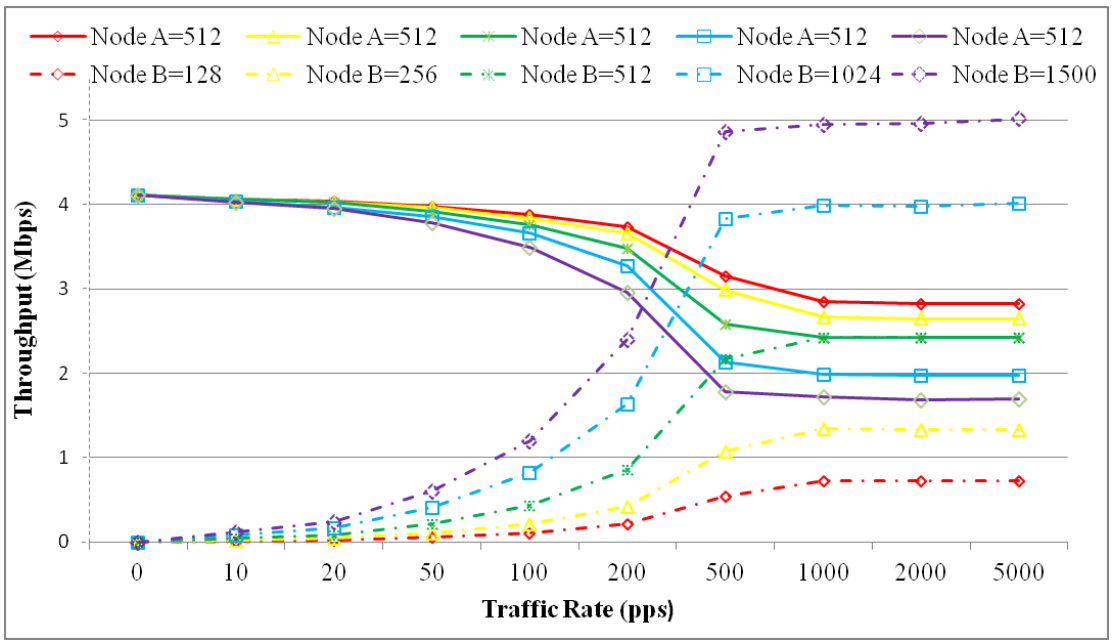

Figure 2.11: The Maximum Throughput for a Two Node WLAN

Figure 2.11 shows the maximum throughput for a WLAN node when there is another node also contending for the medium with different packet sizes and packet rates. The solid line represents the maximum throughput for node *A* which transmits saturated traffic with a fixed packet size of 512 bytes on the medium. The dashed line represents the throughput for the other contending node *B* which sends a traffic load with a gradually increasing rate and a different packet size. It can be seen that the maximum throughput for node *A* decreases with an increase in the packet rate and packet size of the competing

traffic load from node *B*, i.e. the maximum throughput of node *A* is 3.9 Mbps when the competing traffic load of node *B* is 100 pps with a 128 bytes packet size, and is only 1.8 Mbps when the competing traffic load of node *B* is 1000 pps with a 1500 bytes packet size.

In a wired network, if the transmission rate is 10 Mbps, we can say that the node capacity is approximately 10 Mbps. However, this definition cannot be employed directly in wireless networks. The two simulation results above shows the node capacity, i.e. the maximum achievable bandwidth for a node not only depends on its own sending traffic load but also depends on the traffic transmitted by its neighbour nodes which also contend for the medium. In this thesis, a WLAN node capacity is defined as:

The bandwidth available under the current load conditions and represents the maximum load that can be achieved by the node provided that the other network nodes maintain their current load.

In WLANs, the nodes share the medium and contend for transmission opportunities and consequently the node capacity varies dramatically and is difficult to measure due to contention, changing channel conditions and network traffic loads, interference, retransmissions and other reasons (as will be discussed further in section 2.4.1). It should be noted here rate adaption is not considered in this definition.

2.3.3 Node *Capacity Utilization* in Wireless Networks

The node *Capacity Utilization* is defined as the ratio of the bandwidth utilized by a node in transmitting its load and the node capacity. This term reflects the usage of the node capacity in a given time interval.

In Figure 2.12, it can be seen that the *Capacity Utilization* is essentially a measure of how close to saturation that a node is operating. If the *Capacity Utilization* is equal to 100%, it

means that the current load is consuming all of the available capacity and cannot win any more transmission opportunities, i.e. the node is said to be saturated. If the *Capacity Utilization* is smaller, it means that the node can transmit more traffic. Moreover, the development of an estimator for *Capacity Utilization* represents a more generic solution for numerous applications as discussed in section 2.4. The specific measurement method will be described in Chapter 4.

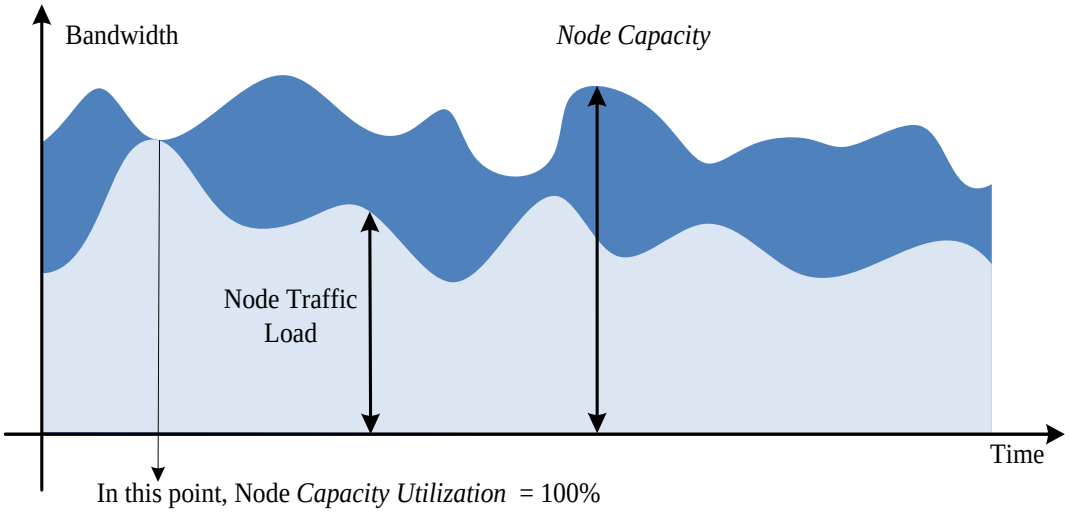

Figure 2.12: The Node Capacity and Node Traffic Load

2.4 Developing a Node *Capacity Utilization* Estimator

This section introduces the challenges, benefits, and applications of developing a node *Capacity Utilization* estimator based upon remote observations by neighbour nodes.

2.4.1 The Challenges in Developing a Node *Capacity Utilization* Estimator

Shared Medium and Contention: Most current research regarding bandwidth estimation has been developed for wired networks with the assumption of a First-In First-Out (FIFO) queue [31], constant link capacity, and individual links for data transmission. However, under the IEEE 802.11 DCF MAC mechanism, nodes share the medium and compete for access to the channel with a binary backoff method to avoid collisions (i.e. the CSMA/CA

mechanism). Thus, the node capacity and *Capacity Utilization* depends on the traffic behaviour of the node itself and the other nodes also contending for access to the medium.

Rate Adaption: In the IEEE 802.11 standard every node operates at a PHY rate selected from the list of PHY rates defined for the specific PHY mechanism used. The PHY rate selection is based upon the channel quality and is usual related to the number of successful and unsuccessful frames transmitted. As the data transmission time depends on the modulation scheme, PHY rate and packet size, a varying PHY rate may lead to a dramatically varying node capacity.

Fading and Interference: Unavoidable and unpredictable fading and interference can cause high levels of transmission errors. The transmission rate is determined by the rate adaptation mechanism in the PHY layer based upon the RF conditions. The loss ratio, signal-to-noise ratio, and signal strength can be used as indicators [32] in the rate adaption mechanism to make decisions regarding the appropriate transmission rate to use.

Retransmissions: High frame loss on a wireless link due to collisions, fading and interference gives rise to retransmissions. The IEEE 802.11 defines a retransmission scheme to ensure the reliability of frame delivery. The exponential binary backoff mechanism contention window impacts the time for a packet to win a transmission opportunity can in turn affect the capacity estimation performed under passive techniques. Moreover, corrupted transmitted frames cannot be analysed by the monitoring node which leads to an underestimation in estimating the network traffic load.

Physical Layer Thresholds: Two thresholds are specified at the PHY layer in IEEE 802.11. The transmission threshold is the minimum received power required to receive a frame successfully. The carrier sense threshold indicates the receiver power that is required to determine whether the medium is busy or not. Essentially, nodes located out

of the transmission range but within the carrier sense range cannot communicate with each other directly, but may still contend for the medium [33]. For example, nodes *m* and *n* in Figure 2.13 are outside of the transmission range but within the carrier sense range of node *i*. In this thesis, the terms carrier sense range and interference range are used interchangeably.

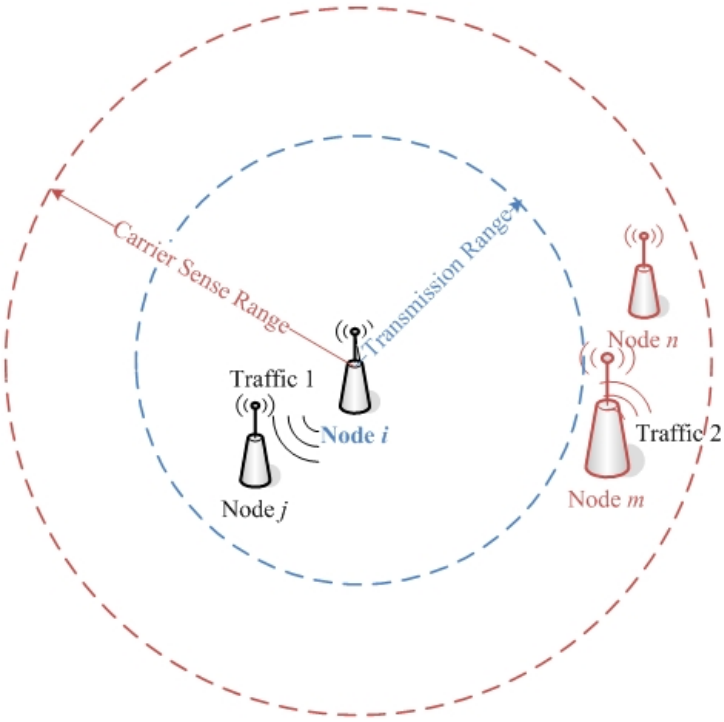

Figure 2.13: Transmission Range and Carrier Sense Range

Hidden Nodes: Due to the physical limitation of wireless hardware (i.e. reception range), the hidden node problem is unavoidable. The hidden nodes have an impact on the estimation of the node capacity using remote observation methods. The remote neighbour nodes cannot capture all the relevant information to analyse the observed node's behavior which results in an underestimation of node capacity.

Overhead: A number of methods use active probe packets to estimate the capacity and available bandwidth. However, these periodic probe packets in wireless networks

consume the bandwidth and increase the contention between network nodes and can have a detrimental impact on network performance.

2.4.2 The Challenges of Remote Measurement

The estimation of *Capacity Utilization* can be divided in two approaches: local measurement and remote measurement. The node needs to broadcast its locally measured *Capacity Utilization* information to its neighbours to support wireless applications. This additional dissemination mechanism increases the overhead of channel and makes the application more complex. Moreover, the communication of this measurement information to its neighbour nodes may be unreliable owing to the packet losses or delays due to collisions, transmission errors caused by fading or interference arising from hidden nodes.

The remote measurement of a neighbour's *Capacity Utilization* value does not rely upon any inter-node communication or information broadcasting mechanism. Moreover, many wireless applications can benefit from utilizing this information directly through remote measurement (this will be discussed in next sub-section). However, there will be an error associated with this measurement owing to the differences in the wireless medium as observed by the different nodes. For example, errors can arise due to the neighbour node having neighbours that cannot be observed by the node performing the measurement. As the observer node and the observed node do not experience the same medium, the estimation of *Capacity Utilization* value performed by a remote node will differ from that experienced by the observed node itself. In this thesis, we have developed a *Capacity Utilization* estimator based upon remote neighbour observation by (1) passively monitoring the network, (2) collecting the observed node's traffic and its neighbours' traffic, (3) measuring the observed node's contention, (4) using three reasonable and simple assumptions (to address the hidden node problem) to minimize the error associated

with this estimation.

2.4.3 The Applications of Remote *Capacity Utilization* Estimator

Once the node *Capacity Utilization* can be estimated, many network management applications can benefit from utilizing this information. However, the node needs to broadcast its *Capacity Utilization* measurements to its neighbour nodes. The neighbour node can make decisions promptly to obtain a better performance for the wireless service as follows:

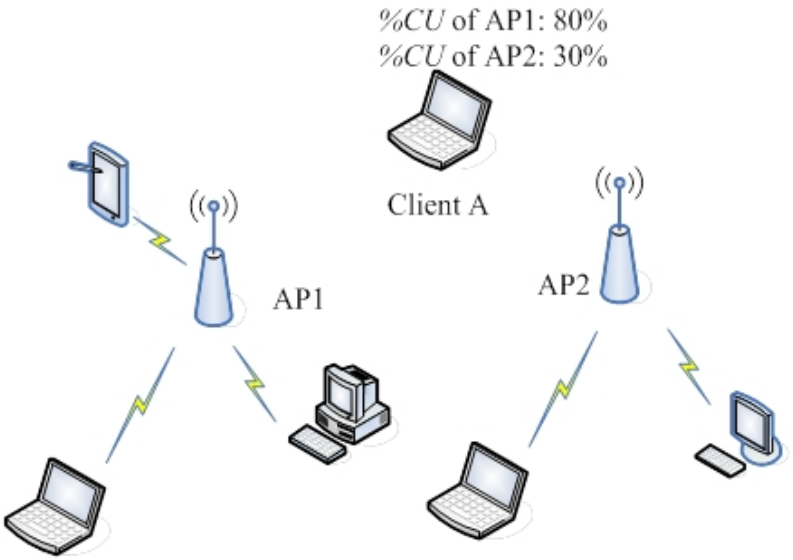

Figure 2.14: An Application of Node *Capacity Utilization (%CU)* in AP Selection

AP Selection: An important use of *Capacity Utilization* information is to support AP selection mechanism in an access network discovery and selection function (ANDSF) [10]. Once a client attempts to associate with an AP in order to access the network, the client can scan the APs within its reception range and use the *Capacity Utilization* estimator to select an appropriate AP which can provide a better service. By comparing each AP's *Capacity Utilization*, the client can select the AP with the lower *Capacity Utilization* value as illustrated in Figure 2.14.

Currently, the client select the AP based upon RSSI, a widely used metric that only provides an indication of how close the client is to an AP. However, it does not provide any information regarding the usage of the AP or the availability of bandwidth at the AP.

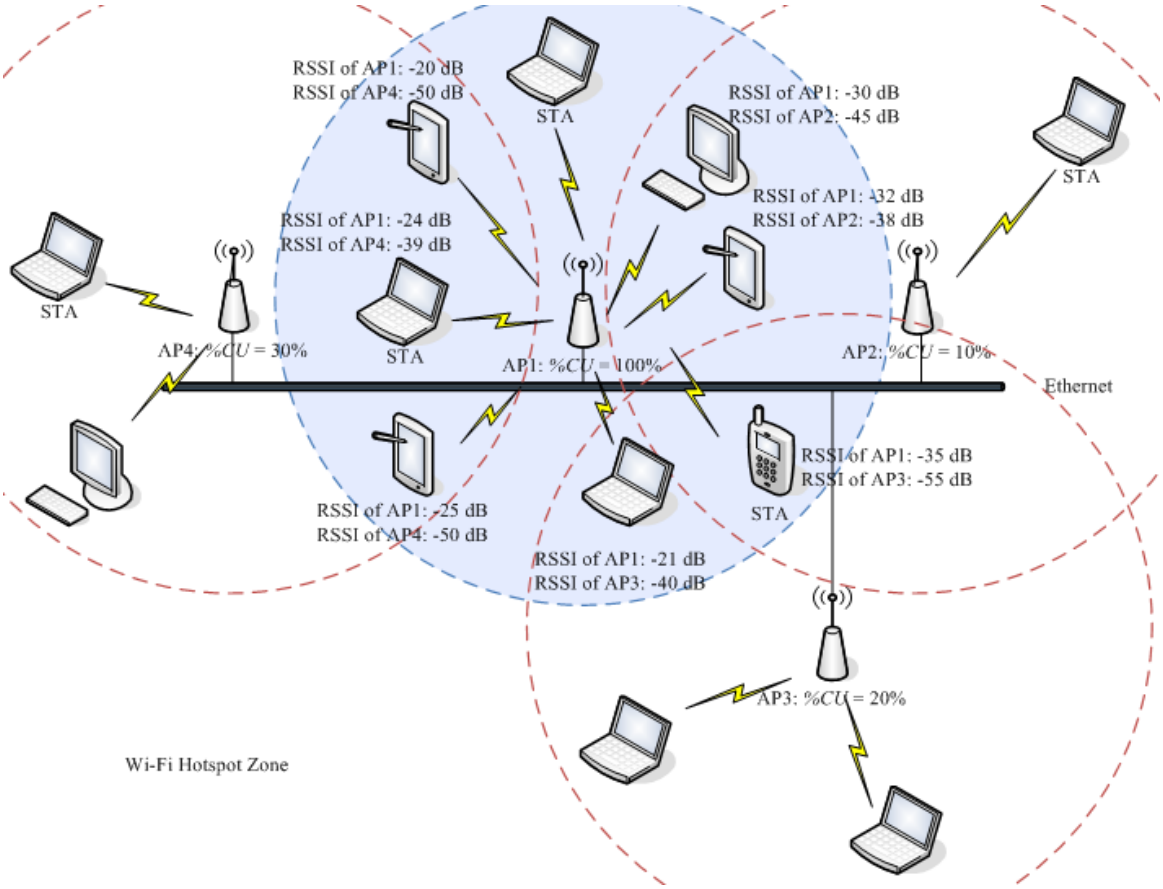

Figure 2.15: An AP Selection Scenario based upon the Use of RSSI

Figure 2.15 illustrates an example where the mobile users enter a Wi-Fi hotspot zone containing multiple APs and associates with the AP based upon the strongest signal. This may lead to a load imbalance [11], an overload of the AP resulting in possible AP saturation or congestion, while other neighbour APs still remain under used. This behaviour may have a detrimental impact on the service for wireless users such as unacceptable packet delay and losses and network performance degradation [11, 34]. Therefore, an alternative metric for AP selection is required to analyse the channel

condition, maintain load balancing, improve the user's throughput, and enhance the utilization of network resources in WLANs.

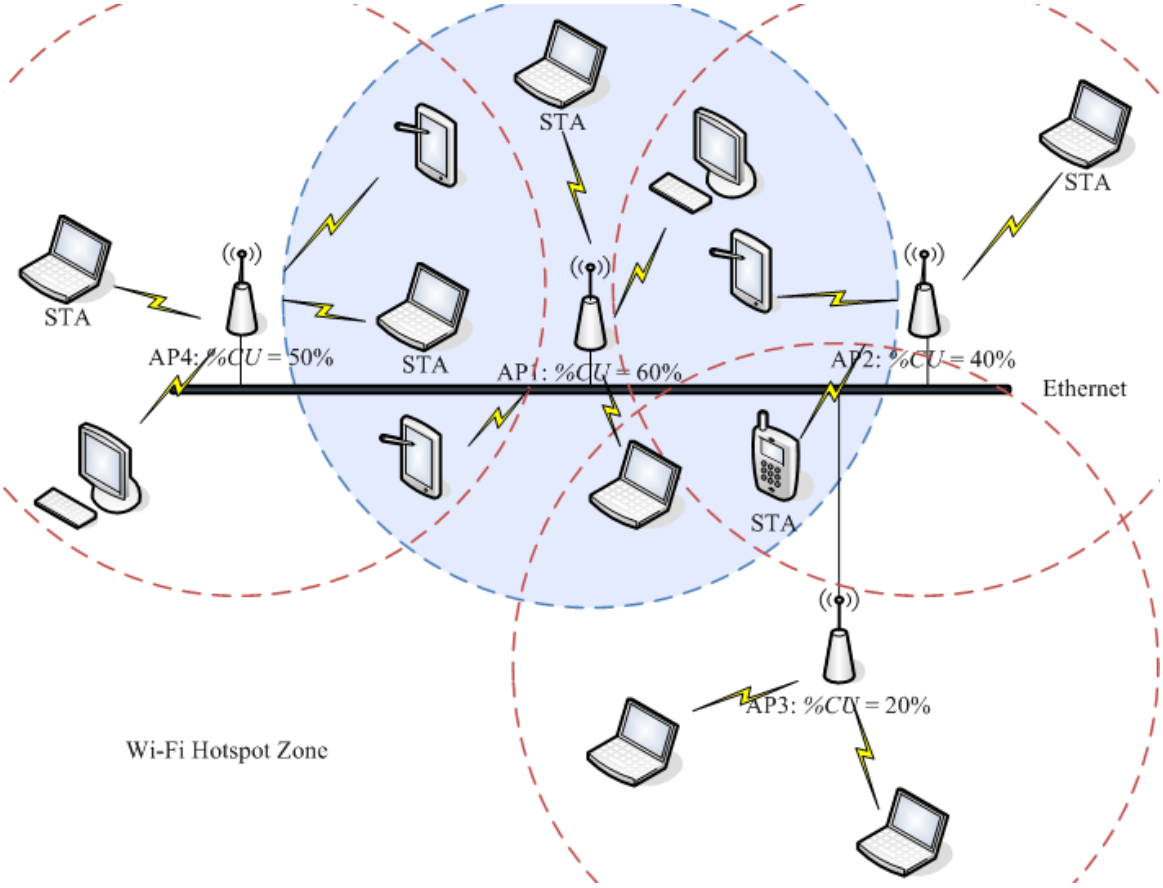

Figure 2.16: An AP Selection Scenario based upon the Use of *Capacity Utilization*

Our *Capacity Utilization* estimator can establish the number of clients associating with the AP, the traffic load of the AP and its neighbour nodes, the contention experienced by the AP, the capacity and *Capacity Utilization* of the AP, and determine whether the AP is saturated or not. Our remote estimator also considers hidden nodes and does not rely upon message exchange between nodes. The remote *Capacity Utilization* metric can also be used to make better handoff decisions as shown in Figure 2.16. Once a mobile user finds that the associated AP cannot satisfy its traffic load demand, it can (1) scan the APs within its reception range, (2) re-calculate the *Capacity Utilization* of all the APs after the

connection in order to avoid saturation in another AP, and (3) make a decision for selecting a new AP as shown in Figure 2.16.

IEEE 802.11k [12] uses a special management frame to exchange information between an AP and a client in order to maintain a load balance and increase the network throughput. However, the AP needs to be equipped with another adapter card or switch to monitor mode for monitoring the networks. Moreover, the message sent from APs may increase the channel overhead and make the application more complex. The dissemination of this information may not be reliable due to packet loss and delay. In particular, packet loss and delays are more likely to occur under heavy or saturated operating conditions. On the other hand, our *Capacity Utilization* estimator has none of the shortcomings mentioned above. It can passively monitor the network, analyse the *Capacity Utilization* of all APs under current network condition and is feasible for applications involving AP selection and handoff.

Most recently, ANDSF developed by the 3rd Generation Partnership Project (3GPP) is used to discover target access points in order to maintain load balancing [35] across a network. It supports the user equipment (UE) in discovering the non-3GPP data access networks (e.g. IEEE 802.11 (Wi-Fi) networks and IEEE 802.16 (WIMAX) networks) in Release 8 [10] in 2008. The ANDSF contains data management and control functionality [36] and consists of three components: inter-system mobility policy (ISMP), inter-system route policy (ISRP), and access network discovery information. Firstly, ISMP decides whether the inter-system mobility is allowed or restricted. ISMP also selects the most preferable access network type and identifier for UE, e.g. whether a WLAN is preferable to WIMAX, or WLAN-SSID1 is preferable to WLAN-SSID2. Secondly, ISRP can support the routing of IP traffic simultaneously over multiple access networks. Finally, Access network discovery information provides a list of available access networks

(including type, identifier and other relevant information) in the vicinity of the UE. Our estimator can also be employed in ANDSF to discover the preferred access network.

Apart from AP selection, our *Capacity Utilization* estimator can be utilized in other application area such as:

Routing Decisions: Currently, two main routing protocols have been proposed in WLANs: proactive protocols (e.g. DSDV [37], OLSR [38]) that use routing tables and probing technique to maintain routes for all nodes. However, the overhead associated with updating the routing table is high. When a route is broken, the reaction may be slow. Reactive protocols (e.g. AODV [39], DSR [40]) broadcast *Route Request* packets to select a route on demand. However, high overhead, latency and low throughput [41] associated with a discovered route may influence the network performance. Moreover, once the nodes can become aware of the capacity of network nodes, they can select a route with a high throughput to improve the performance of applications.

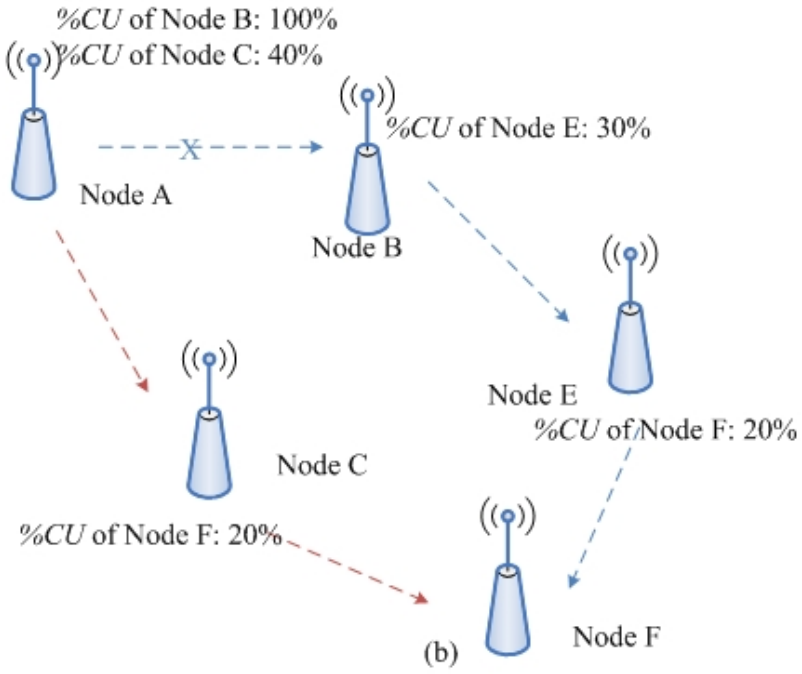

Figure 2.17: An Application of Node *Capacity Utilization (%CU)* in Route Selection

Capacity estimation can be used to support routing decisions (e.g. QoS routing [42], QoS aware routing [8] and resource aware routing [43] etc) where the node *Capacity Utilization* can be used as a route metric to find a path from the source node to the destination node to satisfy the application's requirements. For instance, in Figure 2.17, once the node A becomes aware that node B's *Capacity Utilization* is 100%, it can limit its traffic rate. It also can re-route and select the node B (*Capacity Utilization* of 40%) to re-direct its traffic.

Channel Assignment and Selection: Channel allocation in multi-radio or multi-channel networks is used to allocate wireless bandwidth and assign the communication channels to the interfaces of each network node in order to improve the throughput of networks or reduce the interference, i.e. interference mitigation. The traditional channel selection and assignment methods are divided into fixed channel assignment [44], dynamic channel assignment [45] and hybrid channel assignment [46]. However, none of them takes the *Capacity Utilization* of the channels into account. The node can monitor its own *Capacity Utilization* or its neighbours' to select the appropriate operating channel when the node is operating under heavy load conditions in order to alleviate avoid the saturation or congestion.

Admission Control: The node *Capacity Utilization* can be used to assess the availability of the network capacity and to support an admission control function [47] to either admit or reject the incoming requests to join a network. For example, once an AP's *Capacity Utilization* is 100%, it indicates that this AP cannot win any more transmission opportunities and should reject any new incoming traffic. Similarly, the neighbour nodes can use the estimator to measure the AP's *Capacity Utilization* and then stop transmitting new traffic to the node if they find it to be saturated.

Radio Resource Management: In [14], radio resource management is used for bandwidth provision that can prevent a node's bandwidth from being seized by other nodes in VoIP and video streaming services. Estimating the neighbour nodes' *Capacity Utilization* within an admission control scheme can ensure that sufficient bandwidth can be allocated to each node to satisfy its QoS requirements or to provide different priorities for different applications or users.

Node Saturation Detection: In Bianchi's model [30], node saturation is defined as where there is one packet always waiting for transmission or the transmission queue is nonempty. In this thesis, we define node saturation as a situation that can arise where the node cannot win a sufficient number of transmission opportunities to satisfy its traffic load. It is the most serious problem in managing WLANs as it can give rise to node congestion which in turn can give rise to unacceptable packet delay and losses on a network resulting in a poor performance for most applications. Once the onset of node saturation is detected by neighbour nodes, the neighbour nodes can take timely action in order to prevent the node from becoming congested, thereby minimizing packet delay and losses. It can also be used to determine the reason for the increased packet delay and losses. The potential benefit of the *Capacity Utilization* estimator will be illustrated later in section 5.4 by deploying it in an application for detecting node saturation in order to evaluate the performance of our estimator.

2.5 Network Simulation

In order to simulate a real network environment, network simulator version 2 (ns2) was developed by UC Berkeley in 1997 and is now available as an open source and free software simulation platform [48]. The ns2 simulator is one of most popular source open source network simulators with variety of modules that runs on Unix-based systems. It is a discrete event simulator (where timing of events is determined by a scheduler) targeted

at networking research. The architecture of ns2 follows closely that of the Open Systems Interconnection (OSI) network model. The ns2 also provides substantial support for the simulation of TCP/UDP/RTP/SRM, routing, queuing, and multicast protocols over wired and wireless (local and satellite) networks. Compared to ns1 (no longer developed or maintained) and ns3 (is being actively developed), ns2 is more mature and fully developed, thus we choose ns2 as the simulation tool in this thesis. Apart from these simulators, other simulators such as OPNET, OMNET++, and GloMoSim have been developed for academic research. Moreover, a real-system/testbed can be costly, complex, and difficult to deploy. The simulators can quickly configure different parameter settings, is easy to extend and can quickly produce results.

2.5.1 The Structure of ns2

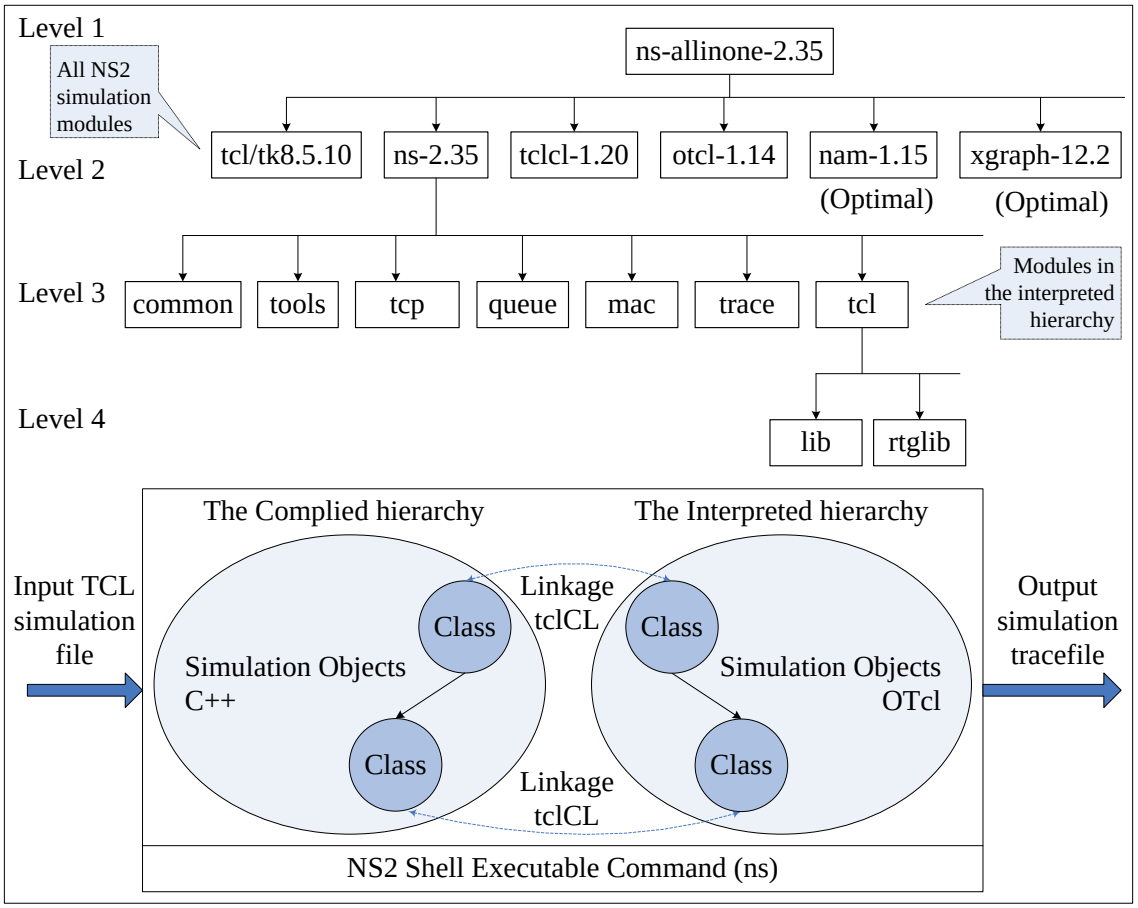

Figure 2.18: Basic Architecture and Components of ns2 Simulator

The basic structure and main components of ns2 are shown in Figure 2.18. It shows that the ns2 simulator is based on two key languages [49]: an object oriented simulator, written in C++ which defines the internal mechanism of simulation objects, and an OTcl (an object oriented extension of Tcl (Tool Command Language)) which sets up simulation, configures the relative parameters and executes a user's command scripts. The C++ and OTcl classes are referred to as a Complied hierarchy and an Interpreted hierarchy. The linkage between the C++ and OTcl is TclCL (Tcl with classes). Mapped to a C++ object, variables in the OTcl domains are sometimes referred to as handles. For example, a handle is just a string in the OTcl domain and the functionality of this handle is defined in the mapped C++ object.

2.5.2 The Advantages and Benefits of ns2

Rich libraries and modules: ns2 also has a rich library of network and protocol objects. Different modules and classes are set up in different levels of dictionaries. The modules in the Complied hierarchy are classified at level 3, i.e., directory *Common* contains basic modules such as *Node* class, *Packet* class, *Session* class and *Encapsulator* class etc. The directory *Tools* contains various helper classes including a random variable generator *Random* class and traffic generator such as *CBR_traffic* class, *EXPOO_traffic* class (exponential traffic) and *PARETO_traffic* class. The directory *mac* module defines the *MAC* classes, *mac_802_3* classes, *mac_802_11* classes, *mac-csma* classes and other classes concerned with the MAC layer.

Easy to use: In ns2, C++ is used to implement the detailed protocol as it provides sufficient execution speed. New agents, packet types, protocols can be introduced by modifying or adding to the libraries and modules. OTcl is used to assemble different components and configure network parameters in order to control the simulation scenario

and schedule the events. The combination of OTcl and C++ provides users with both convenience and usability.

Evaluation of ns2: In the simulation studies of [50], ns2 is the most popular simulation tool used in network research. Research using ns2 includes the numerous WLANs applications and protocols in different layer such as resource allocation, real-time communication, energy issues, transport protocols and control strategies [51]. Comparing with other simulators such as OMNeT++ and QualNet, the author [52] tests different network models and concludes that the results from ns2 come closest to reality. Some research presents a validation of ns2 wireless model and concludes with some recommendations. In [53, 54], the authors suggest a series of recommendations for researchers when using ns2 simulator such as to list the assumptions about environment clearly, to explore a range of model parameters, and to develop a range of propagation models that suit different environments. The results in [51] show that the packet delivery ratios, the connectivity graphs, and the packet latencies comparing the characteristics of a real network and the corresponding model present an average error of 0.3%, 10%, and 57% respectively.

2.6 Chapter Summary

This chapter introduces the relevant basic concepts of WLANs, the IEEE 802.11 MAC mechanism, the concepts of node capacity and *Capacity Utilization*, the benefits and possible applications of being able to estimate node *Capacity Utilization*, the challenges in estimating the node *Capacity Utilization* and the ns2 simulator. This basic knowledge is required for an understanding of the characteristics, challenges, benefits, applications in designing an estimator for *Capacity Utilization* based upon passive observations by remote neighbour nodes.

In order to provide for the better performance of WLANs, we developed an estimator for the *Capacity Utilization* based upon the IEEE 802.11 DCF mechanism which can be deployed on IEEE 802.11 b/a/g/n networks. The estimator can be employed in multi-neighbour network topologies such as mesh and ad hoc networks. However, the hidden node problem is a major challenge for this type of network. Our estimator handles the hidden node problem by passively monitoring the transmitted frames and processing them under a number of assumptions. Once it becomes possible to measure node *Capacity Utilization*, it can be employed in many applications to enhance the network performance. We implemented our *Capacity Utilization* estimator in a passive node saturation detection method and compare it to two other methods: queue observation method based upon Bianchi's [28] definition and regularly pinging method. We use the ns2 simulator to validate our error model and identify the factors influencing the accuracy of the estimator. The details and descriptions of our *Capacity Utilization* estimator will be presented in Chapter 4.

Chapter 3 Literature Review

In wired networks, two nodes use a cable to connect to each other and the capacity is widely defined as the maximum transmission rate that can be achieved between the two nodes. Most estimation schemes assumed that the capacity on wired networks is a constant value. However, owing to the shared nature of medium and the IEEE 802.11 MAC mechanism, the term “link” is ambiguous and difficult to define in WLANs. The WLAN nodes contend for the medium and interact with their neighbour nodes by employing the CSMA/CA mechanism. Consequently, the capacity of a WLAN node is not fixed and can vary dramatically due to the characteristics of WLANs operating environment.

Different researchers have tended to define their own capacity definition and other throughput-related metrics (e.g. node capacity and available bandwidth) and consequently there is no universally accepted definition of capacity. Therefore, in this thesis we will use the two terms node capacity and bandwidth availability interchangeably.

The ability to measure these metrics can be used in many wireless applications to support improved services such as AP selection [34], QoS provisioning [55], resource aware routing protocols [8, 43], channel selection [56], admission control [57], radio resource management [13], and node saturation detection. Therefore, the analysis and discussion of the literature in this chapter describes the different concepts, definitions, methodologies and application areas of capacity estimation techniques in WLANs.

Currently, there are three main methods to estimate the capacity: active probing methods, analytical and mathematical methods, and passive analysis methods as shown in Figure 3.1. However, these methods exhibit a number of disadvantages with regard to accuracy, reliability and overhead in the estimation of capacity. A number of active probing

techniques for capacity estimation are addressed here that show low accuracy and high overhead for the network. Analytical and mathematical methods are based upon certain assumptions that are not suitable for the wireless applications in real networks. Local passive measurements based upon analyzing the transmitted packets performed at a node tend to produce more accurate results [5-7]. The proposed algorithms, measurement metrics, and performance evaluation of these estimation techniques are investigated and discussed in section 3.1, 3.2, 3.3 and 3.4 respectively.

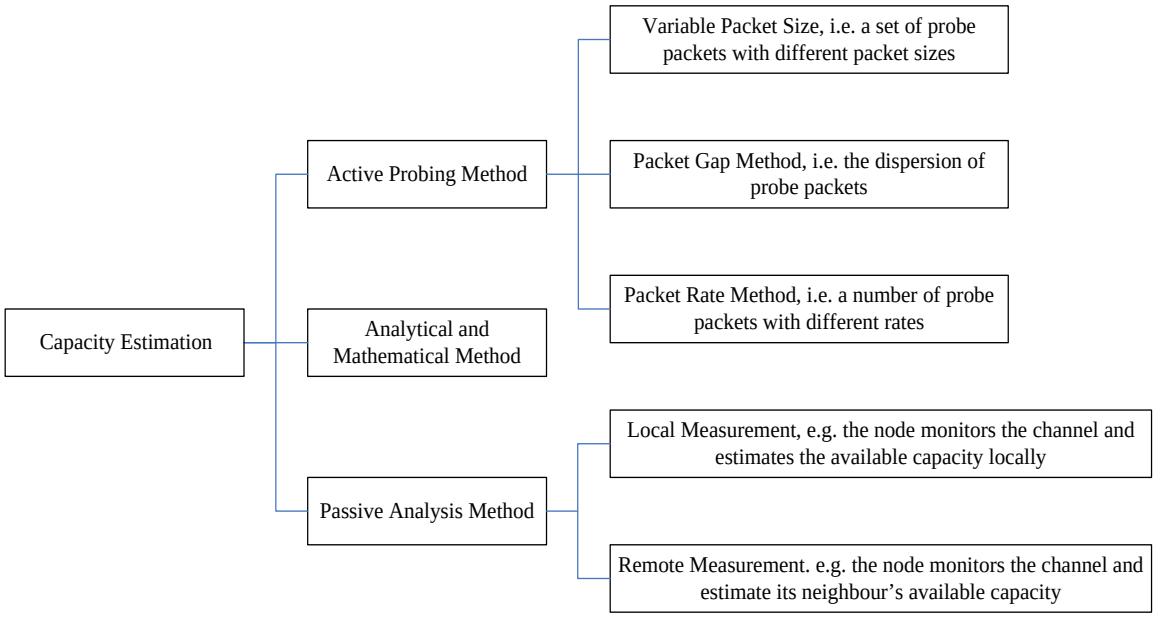

Figure 3.1: The Main Techniques Used for Capacity Estimation

Finally, in order to demonstrate the benefits and potential applications of using capacity estimation techniques, section 3.5 introduces and evaluates some of the methods proposed for utilizing capacity information in different wireless application areas.

3.1 Active Probing Approaches in Capacity Estimation

In a wired network, the fundamental definition [2] of capacity C_i is the maximum possible throughput that a wired link or a network path can deliver. Most researches also differentiate between capacity and another important metric called available bandwidth A_i defined as the maximum unused or spare throughput that a wired link or a network hop

can deliver as introduced in section 2.3.1. For a wired end-to-end path, the available bandwidth A is the minimum available bandwidth of all H hops [2] and is defined as:

$$A = \min_{i=1,\cdots\cdots,H} A_i \tag{3.1}$$

A number of estimation tools have been proposed that transmit a sequence of packets called probe packets with pre-defined inter-packet time intervals to estimate the capacity or available bandwidth in wired networks. These estimation tools can be divided into three categories: variable packet size (VPS) methods, packet gap methods (PGM), and packet rate methods (PRM) which are described in the following sections.

3.1.1 Active Probing Approaches in Wired Networks

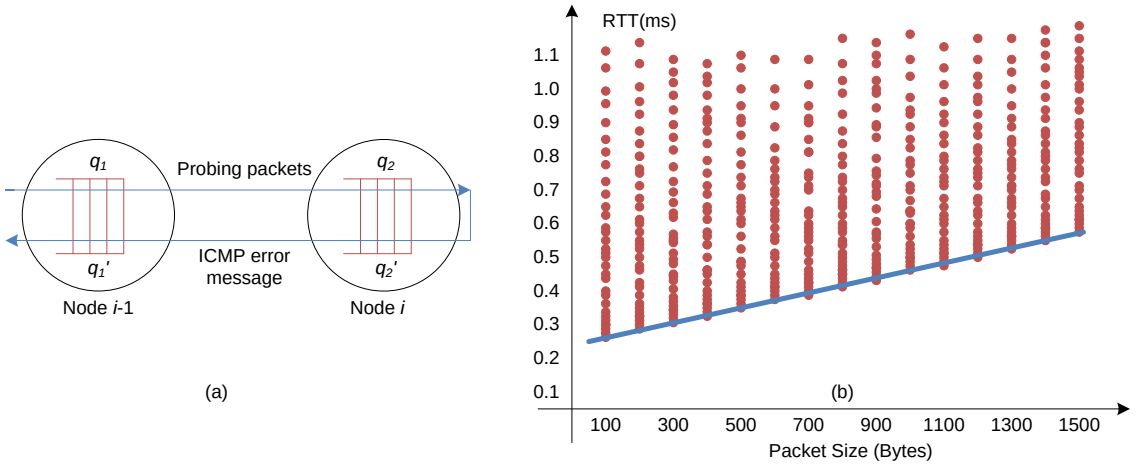

Figure 3.2: (a) A VPS Network Model (b) An Example of the Relationship between *RTT* and Packet Size

The VPS employs a set of probe packets with a variable packet size [58, 59] which measures the round-trip delay time (*RTT*) of a single hop or an end-to-end path as shown in Figure 3.2(a). The sending node i-1 forces the probe packets to expire by setting the Time-To-Live (TTL) field. When node i receives the probe packets, it drops the packets and returns an ICMP error message (i.e. a “time exceeded” response). The *RTT* on hop i

is considered to contain four types of delay for a given packet size L and capacity C between two nodes as follows:

$$RTT_i = \frac{L}{C} + d_{prop} + d_{proc} + q_1 + q_2 + \frac{L'}{C} + d'_{prop} + d'_{proc} + q'_1 + q'_2$$

$$= \alpha_i + \frac{L}{C} \tag{3.2}$$

In the equation (3.2), d_{prop} represents the propagation delay that is independent of the packet size (assumed to be a constant value), d_{proc} is the delay of processing the packet on the router, q_1 and q_2 is the queuing delay, $\frac{L}{C}$ is the delay in transmitting a probe packet, and $\frac{L'}{C}$, $d'_{prop}, d'_{proc}, q'_1$ and q'_2 are the corresponding delay parameters of the return path. The source node selects the minimum *RTT* (under the assumption that the queue delay and process delay is zero) of the sending packets, then computes a linear regression among the minimum *RTTs* for each packet length to measure the delay α_i and capacity C as shown in Figure 3.2(b). Such a method can cause a high overhead. Therefore, much work has been carried out to improve this method by implementing several techniques such as using two pairs of probe packets [60] and a multi-probe packets model that does not send back ICMP error messages [61].

Another capacity estimation method is the packet gap method (PGM) that uses the time gap or difference between the arrival time of two successive probe packets [62, 63] (*Δin, Δout*) to measure the capacity in a time interval *(t, t+ Δin)* as shown in Figure 3.3.

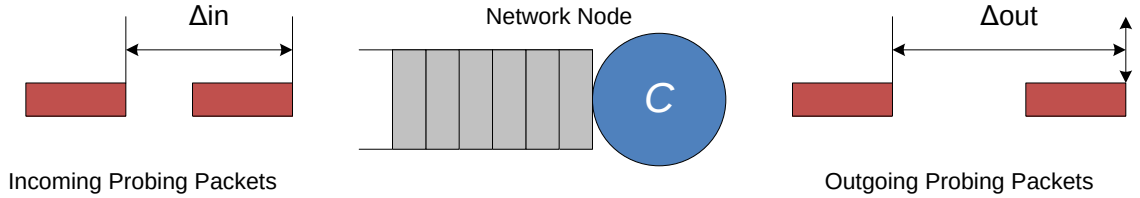

Figure 3.3: Active Packet Gap Model

As a result of using the same packet sizes *L*, the pair of probe packets experiences a similar transmission delay, propagation delay, and process delay etc. If the separation between the pair of incoming probe packets *Δin* is smaller than the transmission delay of the first packet, the separation of two outgoing probe packets *Δout* will be increased. Otherwise, the *Δout* remains unchanged. Thus the relationship of the separation of two probe packets and capacity is:

$$\Delta out = \max\left(\Delta in, \frac{L}{C}\right) \tag{3.3}$$

Pathrate [64] studies the separation of long probing trains that contains *N* probe packets of the same packet size *L*, then sums the time gap of all probe packets $\Delta(N)$ in the train. The capacity *C* of the path without any traffic is given by:

$$C = \frac{(N-1) \times L}{\Delta(N)} \tag{3.4}$$

During the time interval *Δin* between the two probe packets, there are a total of *(Δout-Δin)×C* traffic bits that have passed along the hop. Therefore the available bandwidth defined as the unused throughput that a link can achieve under the presence of network traffic can be derived as:

$$Available\ Bandwith(t, t + \Delta in) = \left(1 - \frac{\Delta out - \Delta in}{\Delta in}\right) \times C \tag{3.5}$$

Moreover, other methodologies use the PGM approach in available bandwidth estimation such as Spruce [65] employing equation (3.5) directly by using a Poisson process of probing pairs, IGI [66] using probing trains with increasing gaps and Delphi [67] employing an exponentially spaced probe packet train.

The packet rate method (PRM) is an iterative method that induces congestion at some bottleneck link of a path with different probe packet rates and makes a comparison between the input probing and output probing packet rates. With an increase in the input probing packet rate, the output probing packet rate also increases. However, once the

input rate of probe packets exceeds the available capacity, the output probing packet rate will not increase. The receiver seeks the turning point in the rate response curve and determines the available capacity. Generally, there are two types of probe packets rate used: a periodic packet stream at a certain rate (i.e. self-loading periodic streams (SLoPS) [68]) and a number of packets with a gradually increasing rate from sender to receiver (i.e. trains of packet pairs (TOPP) [69]). Other estimation techniques include Pathload [70], PathChirp [71], PathMon [72], BART [73], MR-BART [74] that employ various probing schemes and extend the above approaches to estimate capacity or available bandwidth in wired networks.

3.1.2 Active Probing Approaches in WLANs

Previous definitions and capacity estimation methods and tools have been primarily directed at the investigation of wired networks. However, the majority of these definitions and schemes cannot be directly applied to wireless networks. The capacity of a WLAN node depends largely on the characteristics of the wireless shared medium and coordination functions which may result in a variability of available medium resources [5]. This makes accurate measurement-based estimation of capacity more complex in wireless networks. Moreover, the accurate capacity estimation is a major challenge as the channel experiences various unpredictable factors such as interference, fading, path loss [75] which affects the performance of estimation. For example, in IEEE 802.11b WLANs, the physical rate may change dramatically and rapidly from 11 Mbps down to 1 Mbps due to the rate adaptation mechanism used. Thus the concept of capacity in wired networks is not suitable for wireless networks. Thirdly, collisions arising from hidden nodes in WLANs are hard to predict and this leads to an inaccuracy in the capacity estimation. Currently researchers improving these active probing models in WLAN networks will be discussed in the following.

Variable Packet Size

WBD [76] also uses a series of probe packets with different packet sizes to estimate the link bandwidth *C* defined as the maximum transmission PHY rate by calculating the minimum or mean *RTT* for different probe packet sizes *L*. The *RTT* is derived as:

$$RTT = t_{wait} + \frac{L}{C} + t_d \tag{3.6}$$

Where t_{wait} represents the time interval associated with backoff and deferral before a packet transmission which can be regarded as a random value, and t_d is a constant value including SIFS, DIFS, the time duration for transmitting RTS/CTS and ACK, and the propagation delay. In [77], the author defines the end-to-end bandwidth as the maximum transmission rate that a source node can achieve without competing traffic from neighbour nodes. The author also assumes that the time required to seize and release the channel in a one hop transmission is a random variable (whose mean is a function of the packet length).

However in the above method, the source node needs to send a large number of probe packets and compares all the *RTT*s of the sending packets in order to obtain the minimum *RTT* to measure the maximum transmission rate. It has a high overhead and requires a long time to produce an accurate result. Moreover, the VPS focuses on the estimation of maximum transmission rate (referred to as link capacity in wired networks). It does not take into consideration the achievable bandwidth for a WLAN node under the existence of competing traffic from neighbour nodes. Finally, it is different from the definition of a WLAN node capacity in this thesis.

ProbeGap [31] tool sends a series of Poisson-spaced probe packets and collects the one-way delays (OWD) of these probe packets. It picks the maximum OWD of the probe packets, estimates the fraction of time that the channel is idle by probing for “gaps” in the busy periods, and then multiplies this by the channel rate to obtain an estimate of the

available bandwidth. However, the result produces an overestimation of the available bandwidth when the network traffic rate is high.

Packet Gap Model

An Adhoc Probe [78], also employs a pair of packets with fixed packet sizes based on the CapProbe [79] technique which measures the capacity defined as the maximum achievable rate over the wireless path in the absence of any competing traffic, i.e. it is simply the packet size divided by the separation. The authors in [80, 81] proposed an alternative and more accurate approach by extending the Adhoc Probe for throughput estimation of a WMN network. The results show that the capacity measurement is dependent on the probe packet length. However, it does not consider the estimation under the conditions that the neighbour nodes contend for the medium.

In a packet gap method based tool WBest [82], the author defines the available bandwidth as the maximum amount of capacity that a newly arriving traffic flow can acquire at the bottleneck router under the existence of competing traffic. Firstly, using a probing pair separation method the effective capacity C_e can be written as equation (3.7), which represents the maximum capability of the wireless network to deliver network layer traffic.

$$C_e = \frac{\int_{t0}^{t1} \frac{L}{T(t)} dt}{t_1 - t_0} \tag{3.7}$$

Where L is the packet size and $T(t)$ is the packet separation at time t in a observation interval (t_0, t_1). Secondly, it employs a probe packet train technique at a rate C_e and measures the average separation rate R at the receiver to estimate the available bandwidth using:

$$A = C_e \times \left(2 - \frac{C_e}{R}\right) \tag{3.8}$$

However, this method needs an initial estimate to measure the C_e value by sending a number of probe packet pairs. Then it is required to send another probe packet train to estimate the available bandwidth which increases the network overhead. The results show that this method has a higher accuracy, but the author only compared this WBest tool with other wired networks capacity estimation tools.

An empirical approach [4] employs a stream of probe packets in an uniformly distributed time interval to estimate the channel utilization and residual bandwidth in IEEE 802.11 ad hoc networks. The author defines the residual bandwidth as the 'spare' portion of the total bandwidth that is not used by neighbour traffic. The sender node estimates the residual bandwidth before initiating any data traffic. It measures the time period during which the channel is occupied by a neighbour's traffic packet through observing a train of probe packets and measuring the delays of probe packets. The residual bandwidth is denoted as $\eta_{res} = (1 - \hat{u}_c) \times \eta_{max}$, where $\hat{u}_c$ is the measured channel utilization and η_{max} is the maximum throughput when the channel is saturated. Simulation results show that the measurements are considerably more accurate when the sum of neighbour traffic and probe traffic is not close to congestion. However under a congestion condition, this method provides an overestimation of the channel utilization. Moreover, the accuracy of estimation is dependent on the number of probe packets *K* sent and neighbour traffic load (i.e. the estimation performs with high accuracy when using a smaller *K* under lower network traffic or using a larger *K* under heavier network traffic respectively). The author suggested an adaptive probing rate, however the traffic load from the hidden nodes are impossible to accurately predict in a WLAN.

Packet Rate Method

In [83], the author compared both SLoPS and TOPP algorithms based upon PRM through mathematical analysis and simulation. The SLoPS technique performs with a faster

estimation time and a poorer accuracy than that of TOPP technique. In order to increase the accuracy and minimize the measurement time, the author also proposed the SLOT algorithm to estimate available bandwidth in IEEE 802.11 networks. SLOT uses the SLoPS strategy by sending a stream of probe packets pairs at a constant rate as a first step to seek the measurable range of available bandwidth. In the second stage, it employs a TOPP strategy through sending trains of probe packets rate with an increasing rate to obtain the accurate available bandwidth from the measurable range acquired in first step.

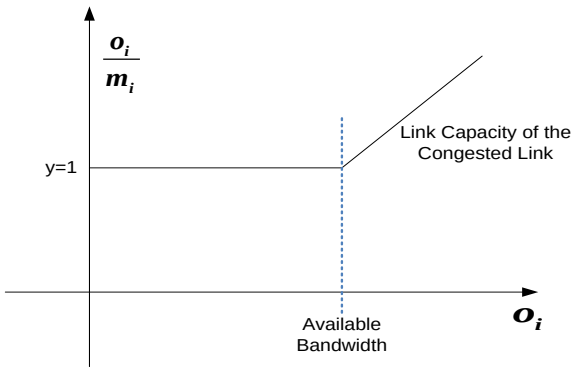

Figure 3.4: Basic Model of DietTOPP [3]

Based on the trains of probe packet pairs method, DietTOPP [3] defines a wireless link as a traffic flow path from source node to destinations. It has been developed to measure the available bandwidth of an end-to-end link in a wireless environment. DietTOPP is a packet rate based method that injects probe packets trains with identical sizes and with an increasing rate along the path towards the receiver. The relationship between available bandwidth and probe packets rate is shown in Figure 3.4 where the offered rate o_i of the probe packets gradually increases until congestion occurs. At the point at which the measured probe packets rate m_i becomes constant, this can be used to provide a measure of the available bandwidth. A BART tool [84] also used by the PRM and employs a Kalman filter to estimate how much bandwidth an application or a WLAN node can expect to use when sending or receiving network traffic.

In contrast to the wired networks, the result indicates that the accuracy of available bandwidth estimation based probe packet techniques in wireless networks is not only dependent on the probe packet size but also on the neighbour traffic intensity due to the contention between the network nodes in accessing the medium.

3.1.3 Discussion

The concept of capacity estimation was first proposed in wired networks to provide a better service for applications such as routing, flow control, congestion detection etc. A number of estimation tools based upon active probe packets have been studied to provide accurate capacity estimation in wired networks. Most estimation schemes assume the capacity of the wired link is constant while the link is assumed to be point-to-point [31] and connected by a cable. However, these assumptions and definitions are not applicable to wireless networks due to the characteristics described in section 2.4.1. The results from the experiments [5-7] which compare the performance of some of the existing tools for available bandwidth estimation suggest that probe-based tools are not the best choice for wireless networks. The authors conclude that those probe-based tools that perform well in wired networks cannot provide accurate and consistent results in wireless networks. The active measurement has an effect on the observed channel of introducing latency and jitter [85]. The passive scheme is more accurate and less costly. Moreover, the basic model employs the common assumption of First-Come First Served (FCFS) in capacity estimation which may not be suitable for WLANs with contending traffic due to the IEEE 802.11 DCF mechanism which seeks to achieve a fair access allocation [5, 31]. When employing this approach, some important considerations need to be borne in mind. The trade-off between overhead and accuracy cannot be ignored because probe packets consume the shared medium and channel bandwidth. This approach impacts negatively on the network performance under node congestion conditions. Moreover, the results

from more recent researchers show that the accuracy of the estimation also depends on the network traffic intensity and probe packet size. Thirdly, active probing methods rely on the successful transmission of probe packets, the accuracy will be reduced when a number of significant probe packets are lost due to collisions or transmission errors. Currently, most researchers focus on passive methods to obtain the node capacity.

3.2 Analytical and Mathematical Approaches

Some of the capacity estimation methods use a purely analytical approach to model and evaluate the performance of a wireless network. Bianchi [30] presents an analytical model to compute the saturation throughput (defined as the maximum load that a single cell can deliver) performance of the IEEE 802.11 DCF with a bi-dimensional Markov Chain model to model the backoff counter and the number of collisions of each node under saturation conditions. The saturated throughput TP_{sat} in a randomly chosen slot time T (i.e. time interval between the beginnings of two successive decrementing slot times) is:

$$TP_{sat} = \frac{E[Successful\ transmitted\ payload\ bits\ in\ T]}{E[T]} \tag{3.9}$$

Bianchi's model produces accurate measurements under the assumptions of:

- Finite number of terminals and every node is saturated
- Ideal channel conditions (i.e. no hidden nodes)
- Constant and independent collision probability of a transmitted by each node

However, this saturation assumption is a rare situation in real IEEE 802.11 networks, i.e. nodes are far from being all saturated at the same time period. Duffy [86] developed an extension of Bianchi's model to consider a non-saturated environment. However the extended model still assumes an ideal channel without transmission errors and performs with high accuracy under the assumptions of constant packet arrival probability and small buffers. Garetto [87] proposed an analytical model to predict each flow's throughput and

extended Bianchi's model in CSMA/CA multi-hop wireless networks. Kwak [88] analyses the performance of DCF binary exponential backoff with a maximum retry limit to obtain the saturation throughput and the medium access delay with *N* network nodes under the similar assumptions of Bianchi's model. In [89], the author defined the end-to-end throughput capacity as the maximum achievable end-to-end throughput of a new flow in the presence of the existing traffic flows. The author also proposed an analytical methodology in multi-hop wireless networks and assumed a fixed packet size for the UDP traffic streams, fixed signal to interference ratio (*SIR*) value, and non-empty queue.

Gupta and Kumar [90] considered two possible models for the successful reception of a frame transmission over one hop. The protocol model prevents a neighbouring node from transmitting on the same sub-channel at the same time and in a physical model a successful transmission depends on a minimum *SIR*. The capacity of wireless networks with *n* identical random located nodes which are capable of transmitting at *W* bits per second with fixed range is $\theta(\frac{W}{\sqrt{n\log n}})$ bits per second under a protocol model. The model was then extended to compute an approximation of the maximum throughput for arbitrary wireless networks under several interference models [91]. However, the fundamental limitation of this model is that it requires an ideal scheduling algorithm which knows all the nodes' locations and all their traffic demands. Other approaches [92] [93] also proposed an analytic method to model the wireless characteristics for capacity estimation. In [92], the author assumes every node has the same arrival rate in order to compute the average delay and normalized throughput for the standard packet size.

The analytical models contribute significantly to modeling the network behaviour and analyzing the theoretical saturated throughput, capacity or achievable capacity boundary in WLANs. However all the analytical models discussed above have their limitations and have been operated under some specified assumptions. Some assumptions are not suitable

for real IEEE 802.11 networks, i.e. where every node is saturated and an ideal channel with no transmission errors [30, 87], constant packet arrival probability, small buffer [86], fixed packet size of traffic load, or under a given interference model [90] [91] [94]. Thus these results are suitable for theoretical analysis, not for the performance analysis of wireless applications in a real environment.

3.3 Passive Approaches for Capacity Estimation

Passive approaches are techniques that only analyse the transmitted frames and use local information to estimate the capacity of a single node or a pair of nodes (denoted as a wireless link by some researchers and introduced in section 3.3.4) in WLANs. Most researches use the term "available bandwidth" to denote node capacity in order to describe the bandwidth available for a WLAN node under the current channel conditions. Different researchers have adopted their own definitions and throughput-related measurement metrics to optimize the service performance of wireless applications that will be discussed in the following. The WLAN node passively monitors the channel usage and estimates the throughput-related metric (i.e. available bandwidth) and then broadcasts the information to its neighbour nodes for the wireless applications such as routing selection or admission control.

3.3.1 Factors Influencing the Accuracy of Estimation

In the passive approaches for available bandwidth estimation, a node *i* locally monitors the channel utilization ratio $\overline{U}_{chan}$ by sensing the channel status then multiplying it by the maximum transmission rate C_i to obtain the available bandwidth *AB(i)* in a given measurement interval of interest. The basic formula [55] [2] is:

$$AB(i) = (1 - \overline{U}_{chan}) \times C_i \tag{3.10}$$

However, once the network nodes attempt to transmit packets, they need to wait for an interval of DIFS before decrementing their backoff counter. In another words, this time

period required to actually win a transmission opportunity on the medium does not belong to "available" time period which can be utilized for packet transmission. In [95], the author considered all the factors affecting the available bandwidth (*AB*) estimation and takes into account the formula (3.11) where T_{idle} represents the channel idle time and T represents a given measurement interval of interest in sensing the channel.

$$AB \leq \frac{T_{idle}}{T} \times C_i \tag{3.11}$$

As the simulation results in section 2.3.2 have shown, an 11 Mbps implementation of IEEE 802.11b cannot deliver throughputs higher than 7 Mbps under UDP traffic. Thus simply sensing whether the channel is idle or busy may not accurately estimate the available bandwidth. For a node, the available bandwidth estimation obtained by measuring channel utilization should be considered by using a factor *"Access Time Interval"* (referred to as T_{access}) in a total time period as shown in Figure 3.5.

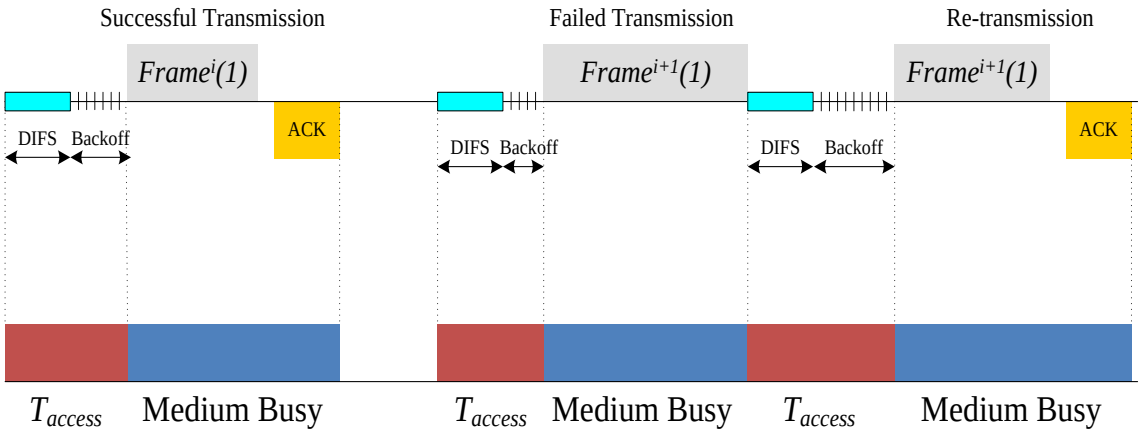

Figure 3.5: The Factor "Access Time Interval"

The term T_{access} represents the proportion of extra time introduced by DIFS and the backoff counter. Suppose there is only one packet being transmitted on the medium, this access time for one packet should be DIFS plus a random backoff counter. For two nodes competing for the medium, this access time is shared by the nodes and is more

complicated. However, it is a challenge to measure T_{access}, as it is related to the contention and cannot be calculated directly due to the random backoff counter. It is also affected by the number of neighbours competing for the medium, collision probabilities, and the network traffic rate. Moreover, Figure 3.5 shows that the failed packet transmission also occupies the channel. However, it is difficult to capture the failed packets through a passive monitoring approach. This section introduces a number of studies that use passive approaches to develop the available bandwidth in WLANs. Different methods have been proposed that use different definitions, methods and consider different factors influencing the accuracy of estimation. In order to better organize the proposed references, the estimation methods are divided into two categories: Local estimation with non-interfering nodes and interfering nodes respectively.

3.3.2 Local Estimation with Non-Interfering Nodes

The author in [96] defines the maximum achievable bandwidth as the maximum throughput that can be achieved by locally observing the ratio of successfully transmitted payload and the channel occupancy time at the sender nodes. An estimator called *EVA* [97] estimates the available bandwidth (defined as maximum achievable bandwidth) as the number of successfully transmitted frames multiplied by the payload size, divided by the expected frame transmission time. However, the above two methods do not consider the available bandwidth in the presence of neighbour traffic.

In [55], each wireless node continually monitors the channel status (idle or busy states) and estimates its available bandwidth for each measurement interval *T* by computing the channel utilization ratio U_{chan} as equation (3.12) and using a exponentially weighted moving average filter to measure the current channel utilization ratio $U_{chan}(t)$ at time *t*.

$$U_{chan} = \frac{T_{channel\_busy}}{T} \tag{3.12}$$

The available bandwidth at time t of a node is $AB(t) = (1 - U_{chan}(t)) \times C(t)$ where $C(t)$ is the line rate. In [98], the author also estimates the available bandwidth by computing the channel utilization ratio at each node by a continual monitoring of the channel status (idle or busy) and its queue delay. However, the above methods and other techniques [99, 100] simply estimate the available bandwidth by using equation (3.10) which results in an unavoidable overestimation of the available bandwidth. Speedo [101] estimates the available bandwidth at the IP layer by measuring channel utilization U_{chan} and the maximum bandwidth at the IP layer B_{IP} that a node can obtain. The IP level available bandwidth AB_{IP} is defined as $(1 - U_{chan}) \times B_{IP}$. The author considers the time interval required to win a transmission opportunity on the medium in estimating B_{IP}, but the author simply uses the average contention window size ($\frac{CW_{min}}{2}$) to estimate the interval of backoff counter. This may lead to an underestimation of T_{access} when the collision probability is higher. Similarly in other reports [102, 103], the authors use similar passive monitoring methods and definitions. In [104], a node gathers the information by monitoring its environment and employing a neural network technique. However, most of the above algorithms in section 3.3.2 simply estimate the available bandwidth by using equation (3.10) and a few of them take into account the interference of hidden nodes. This results in an unavoidable overestimation of the available bandwidth.

3.3.3 Local Estimation in the Presence of Interfering Nodes

In order to improve the accuracy of available bandwidth estimation, various research efforts have attempted to improve the estimate of available bandwidth by employing different methods as follows.

Modifying the Reception Range/Interference Range

The reception range and interference range (also denoted as carrier sense range in some works) are another challenge for estimating bandwidth. It is difficult to measure the channel utilization by monitoring neighbours within the reception range. Moreover, the bandwidth consumption of a traffic flow and available bandwidth estimation of a node is not only related to a local node but also to the neighbour nodes traffic within its interference range.

BRulT [9] provides information on their neighbours to the nodes. Each node periodically broadcasts a message (i.e. a *hello* packet) propagated within two hops to every other node that can hear it in its communication range in order to obtain a balance between overhead and accuracy. The *hello* packet contains the address of the transmitter and the total bandwidth that it will require for QoS provision.

CACP [33] estimates the available bandwidth of itself and its neighbour nodes by proposing three approaches to propagate node information to the nodes within its interference range:

1. Using two-hop neighbourhood broadcasting via *hello* packets
2. Using a larger transmit power for queries packets
3. Introducing a larger threshold called the *neighbour-interference threshold* which can cover all of the interfering node's interference ranges

The local available bandwidth B_{local} and neighbour available bandwidth B_{nbor} are written as:

$$B_{local} = \alpha \times (1 - B_{local}) + (1 - \alpha) \times \frac{T_{idle}}{T} \times B_{chan} \tag{3.13}$$

and

$$B_{nbor} = \alpha \times (1 - B_{nbor}) + (1 - \alpha) \times \frac{T_{idle}^{nbor}}{T} \times B_{chan} \tag{3.14}$$

Where T_{idle} and T_{idle}^{nbor} are the amount of idle channel time and the amount of time that the channel is idle for all neighbours respectively during every period of time *T*, α is a smoothing factor and B_{chan} is the channel capacity defined as the maximum transmission rate. In order to eliminate the interference in the interference range, a perceptive admission control (PAC) [105] defines the *receiver interference distance* (*RID*) which increases the interference range between two senders to ensure uncorrupted packet reception at a receiver by avoiding interference. The formula is:

$$Interference\ Range\ =\ 2 \times Reception\ Range\ +\ RID \tag{3.15}$$

AAC [106] defines the serviceable bandwidth as the smallest available bandwidth within the interference range IR_i of node *i* by using *hello* packets for exchanging the information $B(j)$ to its neighbours:

$$B_{serv}(i) = \min_{\forall j \in IR_i} B(j) \tag{3.16}$$

Where *B(j)* is the available bandwidth of the node located within the interference range IR_i by simply summing the size of sent and sensed packets over a fixed period of time.

In conclusion, using additional hops to exchange information is relatively easy to implement however it increases the overhead. Moreover, the two-hop propagation range may not be enough to cover all the nodes within the interference range in some high node density networks. It may lead to a poor estimation. Increasing the transmission power or the interference range enhances the accuracy but it consumes more of the node energy. Some of the above schemes consider that the interval time of deferral and backoff is negligible. The available bandwidth is still estimated by using equation (3.10) which leads to an overestimation of the actual available bandwidth. However these approaches suggest some methods to solve the problem of interference from hidden nodes within the interference range.

Monitoring Based upon NAV with CTS/RTS mechanism

DCF defines a virtual interference mechanism provided by the Network Allocation Vector (NAV). In this way all the nodes hearing a RTS or CTS frame learn of the medium reservation by setting their NAV to the value obtained from the duration field [107] shown in Figure 3.6. The NAV prevents other nodes from accessing the medium until the transmission is completed.

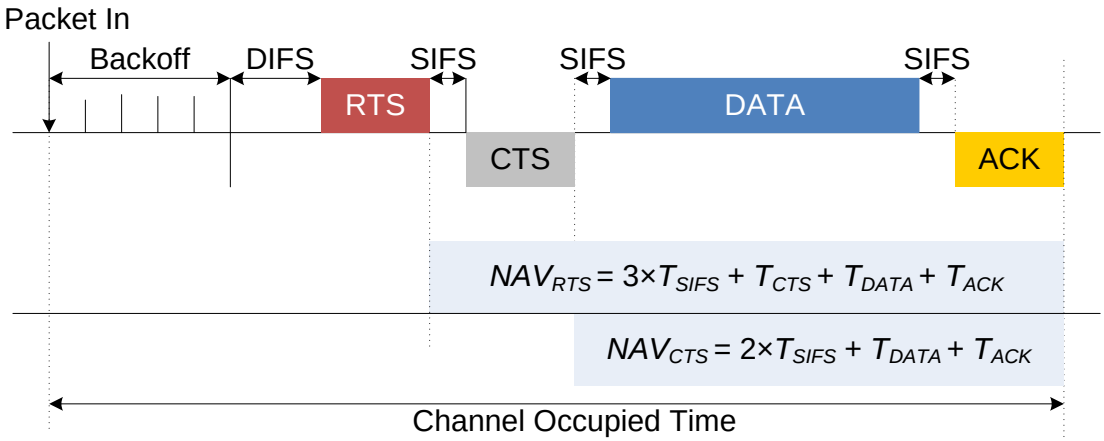

Figure 3.6: The IEEE 802.11 NAV with CTS/RTS Mechanism

The references [108-110] monitor the channel based upon the use of the NAV scheme to measure the channel occupancy and estimate the available bandwidth from the perspective of the network node.

The channel busy time for a node that can receive RTS frames is T_{busy}, where

$$T_{busy} = NAV_{RTS} + T_{DIFS} + T_{RTS} + T_{backoff} \tag{3.17}$$

The channel busy time for a node that can receive CTS frames is T_{busy}, where

$$T_{busy} = NAV_{CTS} + T_{DIFS} + T_{RTS} + T_{CTS} + T_{SIFS} + T_{backoff} \tag{3.18}$$

Where NAV_{RTS} and NAV_{CTS} are the NAV timer of the RTS frame and CTS frame respectively as shown in Figure 3.6. T_{DIFS} and T_{SIFS} represent the distributed inter-frame interval and short inter-frame interval. T_{RTS}, T_{CTS}, T_{DATA} and T_{ACK} denotes the

transmission time of a RTS frame, CTS frame, data frame, acknowledgement frame respectively, $T_{backoff}$ is the duration of the backoff timer.

A passive available bandwidth estimation strategy [109] was proposed to predict the probability of a collision during a frame transmission by measuring the average contention window at each node. It also collects the *hello* packets periodically sent by their neighbour nodes to obtain their NAV (i.e. NAV_j) to estimate the available bandwidth estimation of node i with N number of neighbour nodes in a time interval T and the formula used is given by:

$$AB(i) = \frac{T - \sum_{j=N}(NAV_j + T_{DIFS}) - T_{backoff}(i)}{T} \times C_i \tag{3.19}$$

Based upon the NAV mechanism, the channel status can be measured accurately. However when a route breaks or a collision occurs, the reserved resources corresponding to the control frame cannot be released immediately [8, 111] which affects the accuracy of the estimation. Moreover, this kind of mechanism cannot be implemented in a network supporting multiple priorities. Finally, the periodic *hello* packet used for requesting NAV consumes bandwidth.

3.3.4 Available Bandwidth on a Pair of Nodes

Some authors define the available bandwidth between a pair of nodes as the maximum throughput that can be transmitted between these two nodes in the presence of network traffic in the network in order to provide an accurate estimation to WLAN applications such as QoS aware routing or admission control.

ABE [112] argues that the AAC [106] scheme does not consider synchronization between the sender *s* and the receiver *r* which overestimates the real available bandwidth on a pair of nodes. The author estimates the available bandwidth $AB_{(s,r)}$ between the two peers *(s,r)* as:

$$AB_{(s,r)} \leq min(AB_s, AB_r) \tag{3.20}$$

Where AB_s and AB_r are the available bandwidth of node *s* and *r* respectively. Considering a random uniform distribution of the medium busy time over the observation period, the expected estimated available bandwidth $E(ab_{(s,r)})$ can be evaluated at a time unit *i* as the equation (3.21) without RTS/CTS where n_T is the number of time units in a period or measurement, n_s and n_r is the number of time units during which the medium is available for node *s* and *r* respectively.

$$E(ab_{(s,r)}) = \sum_{i=0}^{min(n_s,n_r)} i.P(ab_{(s,r)} = i) = n_s \times n_r \tag{3.21}$$

where

$$P\left(b_{(s,r)} = i\right) = \frac{\binom{n_s}{i}\binom{n_m - n_s}{n_r - i}}{\binom{n_m}{n_r}}$$

In order to enhance the accuracy of this estimation of available bandwidth, the author not only estimates the locally used bandwidth by monitoring the channel utilization, but also combines a probabilistic evaluation of the overlap of the silence periods (referred to as access time) of the two end-points of a link and an estimation of the collision probability via *hello* messages to exchange the information [113, 114]. After this improvement, the final available bandwidth *AB* is:

$$AB = \left(1 - \frac{T_{DIFS} + \overline{T_{backoff}}}{\Delta T}\right) \times (1 - P_{hello}) \times E(ab_{(s,r)}) \tag{3.22}$$

Where ΔT is the inter-frame time between two consecutive frames, P_{hello} is the collision probability of *hello* messages, T_{DIFS} and $\overline{T_{backoff}}$ are the time interval of deferral and backoff. ABE_MM [115] proposed as an extension of the ABE scheme to manage mobility specifically in bandwidth measurement on dynamic topologies.

In order to further improve the accuracy of the estimation of the available bandwidth $AB_{(s,r)}$, IAB [116, 117] considers the synchronization between the sender and receiver, and differentiates the nodes' *BUSY* (in the state of transmitting or receiving) and *SENSE BUSY* states (in the state of sensing).

A cognitive passive estimation of the available bandwidth (cPEAB) [118] measures the proportion of waiting and backoff delay, packet collision probability, acknowledgement delay, and channel idle time. The author analyses the error of available bandwidth estimation and the impact of hidden or exposed nodes. An monitor tool [119] and network management which employs traffic flow collection and exchange technique among neighbour nodes are proposed to improve the cPEAB algorithm. APBE [120] considers the RTS/CTS overhead in order to improve the previous work.

An novel passive estimation of available bandwidth for multi-hop wireless networks called RABE [121] includes the average number of retransmission attempts in the available bandwidth estimation. It also estimates the collision probability, the average number of retransmission attempts, the extra backoff times due to retransmissions and the packet loss ratio for available bandwidth per link.

In PABE [122], the effective channel capacity is estimated by considering the backoff and retransmission of frames. In order to estimate the available channel idle time ratio, a new interference threshold *NCSR-threshold* is adopted. A CARC [123] scheme defines the channel busyness ratio, channel idleness ratio, and channel utilization to support QoS provision.

The author [124] estimated the available bandwidth for a pair of nodes and investigated the relationship between the estimation scheme's accuracy and the SNR ratio. The result shows that a low SNR ratio leads to inaccurate estimation.

Finally, Shen *et al* [57] estimated the available bandwidth according to the information collected by a node itself during the MAC operation without any overhead and considers the impact of the existence of hidden nodes on the available bandwidth estimation. The author calculated the probabilities of idle, packets successful transmitted and collision with the number of nodes in the interference range of the observed node n and the number of hidden nodes n_h for the normalized bandwidth estimation based upon the assumption that every node has a similar average transmission probability.

3.3.5 Discussion

This section has described and discussed the passive methods that do not rely on in the transmission of probe packets for capacity estimation in WLANs. Due to the open nature and hardware limitations (i.e. reception range and lack of centralized management) of the wireless medium in WLANs, the errors of capacity estimation cannot be completely eliminated but can be minimized in order to obtain a higher accuracy. The node locally monitors the channel, measures the channel utilization and estimates the node capacity or available bandwidth and then broadcasts this to neighbour nodes to enhance the performance of IEEE 802.11 WLANs. However, this approach is required to take into account many factors as follows:

- **Collision Probabilities:** Frame collision essentially has a random behavior. It is related to the varying contention and is hard to observe. For example, if there are two nodes that send their packets simultaneously, collisions occur at the receiver node and the packets cannot be decoded correctly, but nevertheless they have occupied the channel for an interval of time. Thus the transmission time of the colliding packets needs to be considered. The proposed method like BRuIT, CACP, and AAC fail to take collisions into account which leads to an overestimation of the available bandwidth. Some method use analytical models or send *hello* packets regularly to

measure the collision probabilities.

- **Deferral and Backoff Time:** How to accurately measure the time interval required to win a transmission opportunity on the medium before transmitting a packet is another challenge. The duration of this time interval in a given time period is related to the contention and the packet size. When transmitting small frames under high contention, ignoring the influence of the deferral and backoff time increases the error into the estimated available bandwidth.
- **Retransmissions:** Firstly, a failed packet transmission still occupies the channel. It cannot be captured by passive monitoring and has an impact on the channel utilization measurement. Secondly, once the node has failed to successfully transmit a packet, the size of contention window is doubled which affects the backoff time measurement.
- **Interference Range:** The node interference within both the reception range and interference range impacts on the accuracy of estimation. Changing the transmission power or interference range or employing *hello* packets or NAV method to obtain the information can solve this problem. However, these methods still have their disadvantages such as an increase in the overhead of sending extra packets and increased node energy consumption. Therefore, we set the interference range equal to the reception range in this thesis in order to eliminate the node interference within the interference range.
- **The Existence of Hidden Nodes:** This leads to collisions and also impacts on the accuracy of algorithm when using a neighbour observation approach.
- **Poor communication reliability:** From the aspect of WLAN applications, most schemes in capacity estimation are designed to improve the performance of applications, e.g. QoS aware routing. For example, a number of estimation tools require estimating the capacity of the local neighbour and its neighbours to obtain an

appropriate path from source to destination. Therefore, an additional dissemination mechanism is employed for exchanging the message. However, this communication may not be reliable due to packet losses and delay.

The literature review in [6] compares four active probing tools Pathload, Spruce, PathChirp, PTR, and a passive method based upon equation (3.12). The result shows that passive methods are more accurate than active methods. However, ensuring reliable communication is another problem for passive methods. In order to overcome this problem, we have designed an estimator of node *Capacity Utilization* based upon neighbour node observations that does not rely on reliable communication between nodes and has no overhead. We also consider the collision, retransmission, hidden nodes and estimate the contention by estimating average backoff and deferral time to enhance the accuracy of estimation.

3.4 Capacity Estimation: Measurement Metrics and Evaluation Criteria

This section summarizes the classifications of the different measurement metrics and the methodologies used by the different capacity estimation tools. It also summarizes the existing research results for wireless estimation algorithms, introduces the evaluation criteria and then investigates their advantages and drawbacks.

3.4.1 Performance Evaluation of Capacity Estimation

A performance evaluation is important to investigate the accuracy of the proposed scheme and its corresponding overhead in capacity estimation. Caution is required in the comparison of the different schemes because they use different measurement metrics and assumptions. Some important evaluation criteria and questions are introduced as follows:

What is the overhead of the estimation tools?

There are two types of overhead, the network overhead and the processing cost of the algorithm. All of the methods that involve the nodes sending wireless packets consume the bandwidth and increase the contention in WLANs. These approaches require sending extra packets for information exchange that may reduce the network performance under heavy traffic conditions, i.e. probe packets in active approaches or *hello* packets in passive approaches.

The other overhead component consists of the processing and measurement time, occupied memory and energy consumption. For example, the estimation tool using *RTT* measurement in Packet Delay Methods needs to pick the minimum *RTT* from all the sending probe packets [76, 77]. It requires a longer processing time than the passive monitor approaches.

Moreover, in sensor and ad hoc networks, a node would normally use its energy for transmission and reception. Once a node uses more complicated methods, the power consumption will increase. Thus power consumption should be selected as a performance metric for evaluating how efficiently the proposed algorithm can reduce the power consumed in a battery powered node [125].

How accurate are the capacity estimation tools?

The most important criteria for capacity estimation is the accuracy of the estimation. A high accuracy can prevent unnecessary actions that may have an adverse impact on the network performance. The accuracy can be represented by the error, absolute error, relative error and absolute relative error that are shown in Table 3.1.

Absolute error describes the specific value of difference between estimation and actual value. It is more suitable for an estimation system which requires a given specific error limitation. Relative error is a ratio that indicates how accurate a measurement is relative to its actual value. In WLAN networks, the node capacity is time-varying and the error

associated with the estimations depends on the many factors. Thus the relative error can better show the accuracy of the capacity estimation techniques.

Table 3.1 The Types of Estimation Error

Error Calculator	Formula
Error	$Error = Estimate\ Value - Actual\ Value$
Absolute Error (*AE*)	$AE = \lvert Estimate\ Value - Actual\ Value \rvert$
Relative Error(*RE*)	$RE = \frac{Estimate\ Value - Actual\ Value}{Actual\ Value}$
Absolute Relative Error(*ARE*)	$ARE = \left\lvert \frac{Estimate\ Value - Actual\ Value}{Actual\ Value} \right\rvert$

How to reliably communicate the information to other users?

Reliability relates to how this capacity estimation information is guaranteed to be delivered to the recipients in order that they may take further actions. Once the algorithm can successfully estimate the capacity of a WLAN node, the estimated node should have the ability to broadcast this information to its neighbour nodes. Most proposed algorithms use an additional dissemination mechanism for taking further actions by applications. However, the communication of this information may be not reliable due to packet losses and delay. Under certain network conditions (i.e. once a node is saturated or congested), the local estimated node may fail to broadcast its capacity information to its neighbours. Most proposed methods employ continuous broadcasting of bandwidth estimate information to ensure reliability. However, this method further increases the network overhead.

3.4.2 Comparison and Classification of Proposed Literatures

This section selects some typical algorithms from the above discussed literature involving different approaches and target measurement metrics in order to highlight their contributions and drawbacks.

Table 3.2 Comparison of Different Methods of Capacity Estimation

References	Metrics and Methodologies	Contributions and Drawbacks
Active Probing Approaches in Wired Networks		
[58, 59, 64-66, 68, 69, 71]	Capacity and available bandwidth of a wired link or an end-to-end network path	• Sophisticated estimation technique with high accuracy • Neither the definition nor methodology can be applied to wireless networks directly
Active Probing Approaches in Wireless Networks		
[76, 77]	• Maximum transmission rate • Packet Delay Model • RTT delay	• Does not estimate the available bandwidth for the node • Requires a long time to produce an accurate result • High overhead
OWD[31]	• Available bandwidth • Packet Delay Model • One-way-delay	• Overestimation when network traffic rate is high • High overhead
Adhoc probe[78]	• Maximum achievable rate • Packet Gap Model	• The result depends on the packet size • Does not estimate the available bandwidth • High overhead

WBest[82]	• Available bandwidth • Packet Gap Model	• A trade-off between accuracy, number of probe packets and processing time • Two stages in estimation that leads to high overhead
Empirical[4]	• Residual bandwidth • Packet Gap Model	• An adaptive probing rate to reduce the overhead • High accuracy under non-congested channel but poor accuracy under high network traffic • Use of probing packets has an impact on the existing traffic
DietTOPP[3] **BART[84]**	• Available bandwidth of an end-to-end path • Packet Rate Method	• The accuracy of estimation depends on probe packet size and network traffic intensity • High overhead • Use of probing packets has an impact on the existing traffic
Analytical and Mathematical Models		
[30, 87] **[90, 94]**	• Network capacity • Analytic model	• Accurate under the specified assumptions (i.e. under idle channel conditions) • No overhead • They are not suitable for wireless applications
Passive Approaches in Wireless Networks		
[55, 96, 98]	• Node available bandwidth • Monitor the channel utilization	• Lower complexity and no overhead • Not reliable • Does not consider the deferral and backoff times • Does not consider the interference of hidden nodes
BrulT[9]	• Node available bandwidth	• High accuracy due to collecting of all the

CAAP[33]	• Changing Tx_Power or CS range • Broadcast *hello* packets	information • High overhead (using *hello* packets) • High energy consumption • Not reliable
[108-110]	• Node available bandwidth • Using NAV scheme	• High accuracy due to collecting of all the neighbour nodes' NAV information • High overhead • Not reliable • Inaccuracy when a route breaks
AAC[106], ABE[112], IAB[117], cPEAB [118]	• Available bandwidth of a pair nodes • Model DCF mechanism and retransmission mechanism • Considers synchronization of idle periods	• Considers the possible operating states of WLANs • Considers the interference from network nodes • High accuracy • Under the specified assumption (i.e. random uniform distribution of the medium busy time) • Need to broadcast information for taking further actions

3.5 Capacity Estimation: Potential Wireless Applications Area

The previous sections discussed the definition, methodology, performance, comparison of the proposed capacity estimation algorithms. Accurate, fast, and low overhead capacity estimation can be used to more effectively achieve the optimization of wireless network services for many applications. This section analyses the main application areas such as AP selection, ANDSF, resource aware routing and admission control that can benefit by employing capacity estimation techniques. The basic approaches, benefits and drawbacks will be described in the following sub-sections. In this thesis, we will implement a node

saturation detection scheme to illustrate the potential application of our *Capacity Utilization* estimator. Therefore, some proposed saturation or congestion detection methods will also be introduced in this section.

3.5.1 AP selection and ANDSF

In IEEE 802.11 networks, a client simply associates with an AP based upon RSSI which in reality is only a measure of the distance between the AP and the client. However, using RSSI as the association metric may lead to an overload of AP, load imbalance, a saturated or congested AP, the underutilized capacity of other APs and a degradation of the whole network performance [11]. Therefore, how to select the "best" AP for a client in the presence of multiple APs has become a popular topic in WLANs. The main proposed AP selection schemes can be divided into three categories: (1) passively listening to the beacon frames which are periodically transmitted by the AP, (2) Using an active probe request/response method to measure some parameters (e.g. available bandwidth and probing delay) for making association decisions, and (3) select the AP from the list of specific information provided by candidate APs.

EVA [97] calculated the maximum achievable bandwidth on all operating channels through estimating the contention time interval, the time interval of successful and failed transmissions. However, this time interval is measured through the successful transmission probability of a packet. It is difficult to calculate and the author obtained this value from the predefined frame error rate (estimated from *BER* and *SNR*). *APM* [126] modeled the IEEE 802.11 MAC mechanism through passively monitoring the channels, capturing all the transmitted frames and estimating the available bandwidth of the APs. Then the clients select the AP which has the highest available bandwidth. However, this approach does not consider the hidden nodes problem.

In [34], the authors defined the potential bandwidth B_P as the maximum bandwidth that a client can obtain after associating with the AP. The potential bandwidth is measured by passively monitoring the beacon delay of AP. B_P is equal to the ratio of the beacon packet size and the beacon delay included the time interval of deferral, transmission delay and ACK delay. An improved work [127] combined RSSI and B_p as an indicator for making a decision. However, these approaches have been tested only under contention free networks.

The authors [11] considered the effective throughput that the client can attain by associating with the AP and its impact on the other clients associated with the AP. However, this approach needs channel utilization information provided by the AP through modifying the beacon frame or sending special probe request frame.

The IEEE 802.11k standard [12] focused on two key WLAN components: clients and APs. It allows the client to connect to the APs with lower utilizations to improve network performance. Although these APs may have poor signal strength, they can provide higher throughput. IEEE 802.11k specified the extension of management frames which provides the information about AP channel report, antenna information, the number of associated APs, average access delay and channel utilization. Once a client attempts to connect to an AP, the client can request a list of available candidate APs with a channel report. Then the client can select the “best” AP for association and access to the network.

Today, the access network discovery and selection function (ANDSF) [10] from the 3GPP is used to discover the target access point is attracting significant attention. ANDSF assists in selecting the “best” (e.g. lowest expense, best QoS, or best experience [128]) access network and provides a list of available access network and selection rules. Some proposed works focus on the extension of the ANDSF, e.g. congestion aware function [128], energy efficiency improvement [129] or QoS control mechanism [130]. Our

Capacity Utilization estimator can extend the ANDSF so that a client can be aware of the *Capacity Utilization* of every AP within its reception range of the client.

3.5.2 Resource Aware Routing

The current popular routing protocols such as DSDV [37], AODV [39] etc are mainly concerned with finding a feasible path from a source to a destination. However, these routing protocols are not suitable for real-time multimedia traffic [131] as they cannot satisfy the users' QoS demand. In [132] the author summarized the proposed QoS aware routing by employing metrics such as bandwidth (resource), delay, energy and hybrid metrics. The last three metrics are out of the scope of this thesis. Resource aware routing can be divided into three steps: bandwidth estimation, routing discovery and routing maintenance. RARE [43] measured the available bandwidth though passively monitoring the packet transmission on the medium and the signal strength as the cost function of a link to find a good path between nodes. In [133], the author estimated the residual bandwidth locally based upon the bandwidth dissemination from neighbour nodes within two hops. The source node sends a *RREQ* packet based upon AODV protocol with a requirement for a minimum bandwidth and then the intermediate node compares the required bandwidth of source node and its residual bandwidth after receiving the *RREQ* packet. If the intermediate node can satisfy the requirement of bandwidth, it forwards the *RREQ* packet until it discovers an end-to-end path. Conversely, it discards the *RREQ* packet. However, this local measurement mechanism relies upon the information dissemination to neighbour nodes. Our *Capacity Utilization* allows a node to become aware of its neighbours' *Capacity Utilization* directly and promptly as it does not incur any further delays such as those associated with employing an additional dissemination mechanism. Other bandwidth techniques [42, 134] use different throughput metrics and methodologies but similar routing discovery methods to support resource aware routing.

3.5.3 Admission Control

In [57] the author estimated the normalized maximum bandwidth, the used bandwidth, and the available bandwidth and established an admission control algorithm (ACA) on each network node. The node only accepts and forwards traffic if its available bandwidth can accommodate the new flow. So in order to predict the available bandwidth and bandwidth consumption of a new flow, CACP [33] proposed an admission control scheme which considers neighbour nodes within the interference range but out of the reception range (called *c-neighbour*). The bandwidth consumption $B_{csp}(i)$ of a new flow *i* is:

$$B_{csp}(i) = N_{pkt} \times T_{data} \times B_{chan} \times N_{cc} \tag{3.23}$$

Where N_{pkt} is the number of packets of the new flow, T_{data} is the transmission time of a packet which includes the time interval of deferral, interframe space and the transmission time of RTS/CTS, data packet, and ACK frame, B_{chan} is the channel capacity (maximum transmission rate), and N_{cc} is the *contention count* which represents the number of nodes on a traffic flow that may contend for the medium within the interference range of the local estimate node. Unlike ACA, the local node needs to compare its available bandwidth, *c-neighbour* available bandwidth (calculated from equations (3.13) and (3.14)) and the bandwidth consumption of a new flow to make a decision for accepting or rejecting the flow. Other algorithms [105, 135] employ a similar admission control scheme with different available bandwidth estimation methods. Once our *Capacity Utilization* estimator determines that an AP's *Capacity Utilization* is 100%, it indicates that this AP cannot accommodate the new flow and should reject any new incoming traffic.

3.5.4 Node Saturation Detection

Node saturation is a condition that can arise where a node cannot win a sufficient number of transmission opportunities to satisfy its traffic load. It can give rise to node congestion which in turn can give rise to packet delay and losses on a network resulting in a poor performance for most applications. Generally, there are 3 basic ways to detect the presence of saturated nodes on a network.

Once a WLAN node is saturated, there will be frames waiting for transmission in the transmission buffer. The node can detect whether it is has become saturated or not by checking the occupancy of its transmission buffer periodically to determine if it is full or overflowing. Although the measurement of saturation is accurate, the communication of this measurement to neighbour nodes is unreliable as the saturated node may not be able to announce its saturated condition to its neighbours.

The second approach is to regularly ping all the network nodes with probe packets to determine if they are operating normally. Measuring the end-to-end loss or delay by sending probe packets is widely used in detecting a congested or saturated node [136]. For example, DiffProbe [137] employed a pair of low priority and high priority packets to measure one way delay. Apart from the method based upon measuring packet delay, the authors [138] use this packet-pair probe to measure the available bandwidth and estimate the accurate target rate to avoid congestion [139]. Recursive Packet Trains [140] applied a probe packet train passing through the routers along a network path, then calculates the changes in packet train length due to the change in available bandwidth on each hop to find the congestion or saturation position. However, this approach also suffers from the lack of reliable communication with saturated nodes and moreover it does not scale well. It also consumes the bandwidth of the channel which can have a negative impact on the performance of a network due to increased node contention on the medium.

In order to overcome the unreliable communication of saturated or congested node, using the nodes to passively monitor the performance of their neighbours in order to detect if saturation is occurring is proposed. In [141], intrusion detection system (IDS) nodes are selected among neighbour nodes that do not suffer from congestion or saturation by using an optimized method in order to passively monitor the neighbourhood. A distributed cluster-based mechanism for supporting multiple classes of traffic is proposed in [142]. The network nodes are divided into different clusters. Each cluster is managed by a pre-selected sentinel node. Each network node within the cluster estimates the traffic load and sends the estimates to the sentinel node which takes charge of processing and determination with congestion level. The above methods need to select a sentinel node from the neighbourhood nodes under non-congested and non-saturated status. How to select this node and guarantee its reliability for communication is still a major challenge because the network behaviour is difficult to predict. Thus in this thesis, we choose a neighbour-based observation method. As neighbour nodes do not experience the same channel or medium conditions as the node being observed, there will be an error associated with this observation. However, it has no overhead as it is completely passive and does not rely upon the need for reliable communication with the saturated node.

3.6 Summary

This chapter presents the current research into the measurement of node capacity and the relevant throughput-related metrics used to evaluate the performance of the methods. The concepts of node capacity estimation are first discussed for wired networks.

Active probe packets techniques employ a series of probe packets sent from a source to a destination at a specific rate to determine the capacity of a node or an end-to-end path. However, this active approach is not appropriate for wireless networks due to the variability of the wireless medium, shared medium, contention, hidden nodes, and

retransmission schemes. Although some researchers have attempted to improve this approach, the overhead of probe packets has an impact of network performance and can even exacerbate the congestion under heavy traffic conditions. Moreover, the accuracy of active probing method is also a problem because it does not scale well under heavy traffic. Other methods based upon the analytical and mathematical models have been proposed to estimate the node capacity. However, in general the analytic models are proposed under specific assumptions and are not suitable for real networks.

Passive approaches analyse the network traffic on the wireless medium and record the idle interval to estimate the available resource for the node. However, the reliability and accuracy is a big challenge when using this method. Many factors such as collision probabilities, deferral and backoff times, retransmissions, interference range and hidden nodes still affect the accuracy of estimation. Moreover, the reliability of the broadcasting of this information due to packet losses and delay is another problem. Most approaches use a periodical broadcasting scheme that gives a rise to a consumption of channel bandwidth. Thus choosing the appropriate estimate approaches should depend on the particular requirements of the wireless application.

This thesis is concerned with developing an estimator for capacity utilization through remote observation by neighbour nodes in WLANs with a high reliability and no overhead. Although there is an error associated with estimation, our remote *Capacity Utilization* estimator has no overhead. Our estimator also considers the main factors that affect the accuracy of estimation and then uses three assumptions to improve the accuracy. This estimator can be developed for many wireless applications such as AP selection, routing aware protocol, admission control, radio resource management, channel selection and node saturation detection. This thesis uses saturation detection as a potential WLAN application to evaluate the performance of the estimator.

Chapter 4 The *Capacity Utilization* Estimator

This chapter describes the remote *Capacity Utilization* estimator based upon neighbour node observations. This scheme is based upon a temporal analysis framework that models the way in which the IEEE 802.11 MAC mechanism wins transmission opportunities. The *Capacity Utilization* estimator has been developed for the DCF mechanism only. However, there will be an error associated with this measurement owing to the differences in the wireless medium resulting from the different locations of the nodes. This chapter analyses the error associated with the *Capacity Utilization* estimator and employs three simple and reasonable assumptions to minimize this error. Finally, a node saturation detection mechanism as a potential application of our estimator will be described in order to show the feasibility and accuracy of the *Capacity Utilization* estimator.

4.1 The Node *Capacity Utilization* Estimation

The concept of MAC bandwidth components to characterize and quantify the usage of and availability of the radio resource for IEEE 802.11 WLANs was proposed in 2004 [13, 14]. These MAC bandwidth components are related to the IEEE 802.11 MAC access mechanism and a number of different time intervals.

4.1.1 MAC Bandwidth Components

Figure 4.1 shows that the wireless medium can be defined in terms of a number of different time intervals under the CSMA/CA MAC mechanism. The busy time interval T_{busy} represents the time on the medium consumed with the transmission of frames. The complementary idle time interval T_{idle} represents the medium idle time which can be used by a node to win transmission opportunities. From the perspective of a contending

node T_{idle} consists of two parts: the time interval T_{access} spent by a node in contending for access to the medium and the time interval T_{free} which corresponds to the unused idle time on the medium.

Figure 4.1: Illustration of the Various Time Intervals involved in Accessing the Medium under the IEEE 802.11 MAC Mechanism

The busy and idle time intervals are summed in a measurement interval of interest as follows:

$$T_{idle} = \sum_{i=1}^{\#_{idle}} T_{idle}^{(i)} \tag{4.1}$$

And

$$T_{busy} = \sum_{i=1}^{\#_{busy}} T_{busy}^{(i)} \tag{4.2}$$

Where $T_{busy}^{(i)}$ and $T_{idle}^{(i)}$ are the durations of the i^{th} busy and idle intervals respectively

within the measurement interval of interest, $\#_{idle}$ and $\#_{busy}$ is the number of idle and busy time intervals within the measurement interval respectively. We normalize them by converting them to a fraction of the medium bandwidth as follows:

$$BW_{busy} = \frac{T_{busy}}{T_{busy} + T_{idle}} \tag{4.3}$$

And

$$BW_{idle} = \frac{T_{idle}}{T_{busy} + T_{idle}} \tag{4.4}$$

The relationship between BW_{busy} and BW_{idle} can be written as:

$$BW_{idle} = 1 - BW_{busy} \tag{4.5}$$

Here BW_{busy} represents the portion of the medium bandwidth used in the transport of the total traffic load. Similarly, BW_{idle} represents the portion of the medium bandwidth that is idle and may be used by any node to win access opportunities for its load.

Three bandwidth components are identified [13] that describe the MAC mechanism operations in the utilization of the wireless medium. Specifically, the definitions of the MAC bandwidth components are:

- $\boldsymbol{BW_{load}}$ is the load bandwidth that represents the portion of the medium bandwidth utilized by a WLAN node in transmitting its load.
- $\boldsymbol{BW_{access}}$ is the access bandwidth that represents the portion of the medium bandwidth required by a WLAN node to win transmission opportunities for its load.
- $\boldsymbol{BW_{free}}$ is the free bandwidth that represents the portion of the medium bandwidth unused by a WLAN node and serves to define the capacity of the node.

By examining the address fields of the IEEE 802.11 MAC frame header, it is possible to

determine the sender address of a frame and monitor all the frame transmissions from a particular node *k*. The load bandwidth is directly related to the throughput of the node *k*. The busy time used by node *k* on the medium to transmit its load is:

$$T_{load}(k) = \sum_{i=1}^{\#_{frm}} T_{load}^{(i)}(k) \tag{4.6}$$

Where $T_{load}^{(i)}(k)$ is the duration of the i^{th} load interval and $\#_{frm}$ is the numbers of frames transmitted by node *k* within the measurement interval and the following formula can be used to convert it to a normalized bandwidth:

$$BW_{load}(k) = \frac{T_{load}(k)}{T_{busy} + T_{idle}} \tag{4.7}$$

When there is only one node transmitting frames on the medium, its BW_{load} will be equal to BW_{busy}. However, BW_{load} represents the successfully transmitted frames, while $BW_{collision}$ represents the unsuccessfully transmitted frames. In the case where two or more nodes are contending for access, some bandwidth will inevitably be lost because of collisions between the nodes attempting to access the medium at the same time. All frames, irrespective of whether they were successful or not, contribute to BW_{busy}. Suppose there is *N* number of nodes competing with node k for accessing the medium, thus BW_{busy} is expressed as:

$$BW_{busy} = \sum_{k=1}^{N+1} BW_{load}(k) + BW_{collision} \tag{4.8}$$

According to the basic MAC access mechanism, the nodes share the idle intervals and use them to decrement their backoff counters to win their transmission opportunities if they have frames to transmit. Therefore, the access bandwidth BW_{access} is used to contend for

the medium and the unused idle bandwidth BW_{free} for any node k can be stated as the following:

$$BW_{access}(k) + BW_{free}(k) = BW_{idle} = 1 - BW_{busy} \tag{4.9}$$

The access time has two components [13]: the time spent deferring (e.g. waiting for an interval of DIFS to elapse) and the time spent backing off (e.g. decrementing the backoff counter to zero). A node may experience several intervals of deferral (i.e. waiting for DIFS or EIFS to elapse) which is dependent on a number of factors, i.e. the number of nodes contending for access to the medium, its own initial backoff counter value, and the transmissions from other nodes. These two components of the access time are considered to be random, so it makes sense to consider them in terms of their average values. Assuming that the traffic packets sent from every node are independent, we also assume that the idle inter-packet time intervals are independent. Furthermore, we assume that the random idle inter-packet time intervals constitute a stationary process. Finally, assuming that this process is ergodic, we can use the time average instead of the ensemble average to estimate the mean of this process. In this case we use the sample mean (or arithmetic mean) of the process. The average time spent deferring is $\bar{T}_{defer}$ and the average time interval for the backoff counter to reach zero is $\bar{T}_{backoff}$. Hence, the average access time $\bar{T}_{access}$ required to win a transmission opportunity is:

$$\bar{T}_{access} = \bar{T}_{defer} + \bar{T}_{backoff} \tag{4.10}$$

where

$$\bar{T}_{defer} = \bar{\#}_{defer} \times \bar{T}_{IFS} \tag{4.11}$$

and

$$\bar{T}_{backoff} = \overline{BC} \times Slot\ Time \tag{4.12}$$

In the above equations, $\bar{\#}_{defer}$ is the average number of times that a node needs to defer

to a busy medium condition. In other words, if there is only one active node on the medium $\overline{\#}_{defer}$ is equal to 1. The parameter $\bar{T}_{IFS}$ represents the average duration of the interframe idle time and $\overline{BC}$ is the average initial *BC* value. The average access time $\bar{T}_{access}$ multiplied by the total numbers of frames transmitted in the measurement interval of interest ($\#_{frm}$) is the access time interval T_{access} which can be written as follows:

$$T_{access} = \bar{T}_{access} \times \#_{frm} \tag{4.13}$$

Normalizing and converting the access time T_{access} to a fraction of the medium bandwidth as follows:

$$BW_{access}(k) = \frac{T_{access}(k)}{T_{busy} + T_{idle}} \tag{4.14}$$

It is difficult to measure the parameters $\overline{\#}_{defer}$, $\bar{T}_{IFS}$ and $\overline{BC}$ directly from the medium due to the random nature of the backoff counter initial value and the deferral time. However, it is possible to obtain an estimate for these parameters from another parameter called the *average contention* which represents the average number of nodes contending for access to the medium in a measurement interval of interest. The detailed measurement method is described in section 4.2.3.4.2.

4.1.2 Access Efficiency Factor and Node Capacity

Based upon how the IEEE 802.11 MAC mechanism wins transmission opportunities under the MAC components framework, the *Access Efficiency Factor (AEF)* represents how efficiently a node *k* contends for access to the wireless medium and can be found in [13]:

$$AEF(k) = \frac{BW_{load}(k)}{BW_{load}(k) + BW_{access}(k)} \tag{4.15}$$

The wireless node capacity is defined as the bandwidth available under the current load

conditions and represents the maximum load that can be achieved by a node provided that the other nodes maintain their present load. It is assumed that $BW_{collision}$ is negligible because only successful transmitted frames are considered in this analysis. The capacity of node *k* can be derived as follows:

$$C_{avail}(k) = BW_{load}^{Sat}(k) = BW_{load}(k) + \Delta BW_{load}(k) \tag{4.16}$$

Where

$$\Delta BW_{load}(k) = AEF(k) \times BW_{free}(k) = AEF(k) \times \left(BW_{idle} - BW_{access}(k)\right) \tag{4.17}$$

It can be rewritten as:

$$\Delta BW_{load}(k) = AEF(k) \times \left[1 - \sum_{j=1}^{N} BW_{load}\,(j) - BW_{access}(k)\right] \tag{4.18}$$

Substituting equations (4.17) and (4.18) into (4.15) gives:

$$C_{avail}(k) = BW_{load}(k) + AEF(k) \times \left[1 - \sum_{j=1}^{N} BW_{load}(j) - BW_{access}(k)\right] \tag{4.19}$$

It can be shown that:

$$C_{avail}(k) = AEF(k) \times \left[1 - \sum_{j=1}^{N} BW_{load}(j) - BW_{access}(k) + \frac{BW_{load}(k) + BW_{access}(k)}{BW_{load}(k)} \times BW_{load}(k)\right] \tag{4.20}$$

This can be re-written as:

$$C_{avail}(k) = AEF(k)$$

$$\times \left[1 - \sum_{j=1}^{N} BW_{load}(j) - BW_{access}(k) + BW_{load}(k) + BW_{access}(k)\right] \quad (4.21)$$

Then:

$$C_{avail}(k) = AEF(k) \times \left[1 - \sum_{j \neq k}^{N} BW_{load}(k)\right] \quad (4.22)$$

4.1.3 Node *Capacity Utilization*

In this thesis, a new metric *Capacity Utilization* based upon *Access Efficiency Factor* and node capacity is proposed. The critical performance metric produced by the model is the *Capacity Utilization* (*%CU*) which is defined as the ratio of the bandwidth utilized by a node in transmitting its load and the node capacity. For a node *k*, its *Capacity Utilization* is expressed as:

$$\%CU(k) = \frac{BW_{load}(k)}{C_{avail}(k)} \times 100\% \quad (4.23)$$

This metric reflects the usage of the node capacity in a measurement interval of interest described in section 2.3.3. When this metric is equal to 100%, it means the node *k* is saturated.

4.2 A *Capacity Utilization* Estimator

Section 4.1 introduced the theory behind the *Capacity Utilization* estimation method. In this section, we describe how this *Capacity Utilization* estimator is implemented in practice.

4.2.1 Impact of Network Topology

This method is intended to employ passive remote observations by neighbour nodes to estimate the *Capacity Utilization* of the observed node as shown in Figure 4.2.

The node *k* is an observed node and within its reception range are five neighbour nodes labeled 1, 2, 3, 4, and 5 who transmit their traffic loads to node *k*. Node *k* forwards the traffic to the gateway node. Within its reception range, node *k* competes with other five neighbour nodes to win access opportunities on the medium. If node 1 wishes to estimate the *Capacity Utilization* of node *k*, it can only monitor three of the neighbour nodes of *k*, i.e. node 2, node 3 and itself as they are within its reception range. This partial observation of the neighbours of *k* means that node 1 lacks sufficient information to make an accurate estimate of node *k*'s *Capacity Utilization*. Essentially, as far as node 1 is concerned, nodes 4 and 5 are "hidden" nodes.

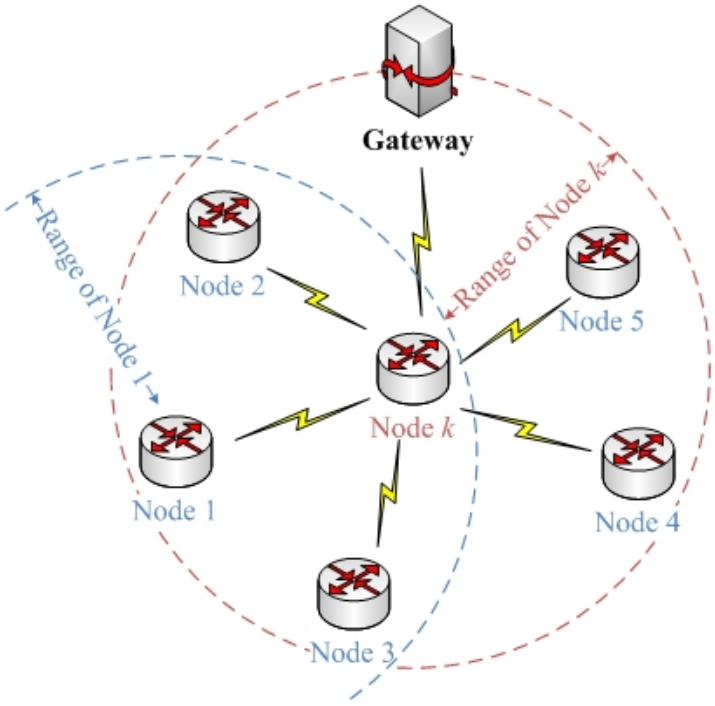

Figure 4.2: A Network Topology for Remote Observations by Neighbour Nodes

4.2.2 Terms and Definitions

In order to better understand the terms used in the thesis, some important terms and their definitions are presented in Table 4.1.

Table 4.1 Some Relevant Terms and Definitions

Term Name	Definitions
Range	The range is defined as the reception range and the interference range where all these ranges are assumed to have the same value.
Reception Range	A node can receive and decode a frame successfully within this range.
Interference Range	The frames sent from another node cause the medium to be sensed as busy within this range.
Neighbour	A neighbour node is defined as a network node located within the reception range of the observed node.
Node *k*	The observed node
Node *j*	The neighbour node monitoring and wishing to generate an estimate of the *%CU* value of the observed node *k*.
N	The **N**umber of neighbour nodes of the observed node, e.g. *N*=5 in Figure 4.2.
M	The number of neighbour nodes of the observed node *k* that can be **M**onitored by observer node *j*, e.g. *M*=3 in Figure 4-2.
$\boldsymbol{BW_{load}\ (k)}$	The load bandwidth of node *k*
$\boldsymbol{BW_{load}\ (j)}$	The load bandwidth of neighbour node *j*
$\sum_{j\neq k}^{N} BW_{load}(j)$	The total neighbour load bandwidth of neighbour nodes observed at node *k*.
$\overline{cont(k)}$	The *average contention (*which represents the average number of nodes contending for access to the medium in a

	measurement interval of interest*)* experienced by node *k* during a measurement interval, see section 4.2.3.4.2.
$\boldsymbol{BW_{access}(k)}$	The access bandwidth of node *k*

4.2.3 Calculation and Measurement of the *Capacity Utilization* Estimator

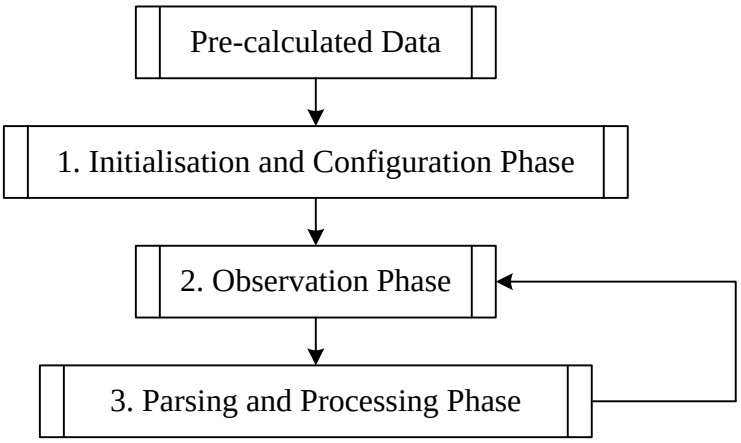

Figure 4.3: Four Phases Involved in the Operation of the *Capacity Utilization* Estimator

The algorithm of the *Capacity Utilization* estimator is divided into several steps as shown in Figure 4.3 and these are described in the following sections:

4.2.3.1 Pre-calculated Data

Once the *average contention* experienced by a node is measured, the average number of deferrals and the average initial backoff counter value can be determined to estimate the access time T_{access}. A C++ program was developed to simulate the contention for access under saturation conditions instead of ns2 simulator because this program was specifically developed to focus on the deferral times and initial backoff counter values. Moreover, the process of each packet transmission in the medium (including the modulation scheme used and packet size) is not relevant to the calculation. Thirdly, this program involves calculating a large number of test simulations and was designed to run

much faster than the ns2 simulator.

Firstly, we assume a fixed number of WLANs nodes (referred to as *#DEV*) where each node always has a frame awaiting transmission, i.e. it is assumed that every node is saturated [30]. We use a C++ program to simulate and model the DCF mechanism, deferral and backoff procedure, re-transmission scheme, and the binary exponential back-off contention window. A pseudo random number generator provided by the Mersenne Twister [143] algorithm is employed to generate the initial backoff counter value. Then we iterate this process one million times to calculate the average backoff counter value *BC*, the average number of times that a node needs to defer $\overline{\#}_{defer}$ and the average number of collisions $\overline{\#}_{collision}$.

Table 4.2 Computer Simulation Results

#DEV	$\overline{BC}$	$\overline{\#}_{defer}$	$\overline{\#}_{collision}$	*#DEV*	$\overline{BC}$	$\overline{\#}_{defer}$	$\overline{\#}_{collision}$
1	15.50004	1	0	**11**	26.90998	9.252916	169007
2	16.52372	1.941183	30300	**12**	28.05749	10.00291	177450
3	17.61385	2.838401	55799	**13**	29.16558	10.74528	185325
4	18.74555	3.702782	77382	**14**	30.2661	11.48171	192454
5	19.91028	4.541039	95944	**15**	31.37181	12.20946	199492
6	21.08136	5.359451	112041	**16**	32.45013	12.93329	205672
7	22.25546	6.161746	126072	**17**	33.50976	13.65177	211870
8	23.44629	6.945296	138617	**18**	34.57446	14.36579	217493
9	24.61254	7.727979	149676	**19**	35.60316	15.0705	222937
10	25.79027	8.494361	159852	**20**	36.61676	15.77516	228240

The results are presented in Table 4.2. For example, if there is only one node present on the channel, the average initial backoff counter value is 15.5 and average deferral number is 1. If there are two nodes always competing for the medium, the average initial backoff counter value is approximately 16.5 and average deferral number is approximately 1.94. As the number of network nodes grows, the average initial backoff counter value and the average number of deferrals also increases due to:

- The occurrence of collisions which doubles the size of the contention window.
- When a node cannot win the transmission opportunity, it needs to defer for a time of DIFS and keeps decrementing its backoff counter until either the frame is successfully transmitted or the frame is dropped by having exceeded its maximum number of retries.
- A large number of network nodes competing for access the medium which leads to a high probability of collision.

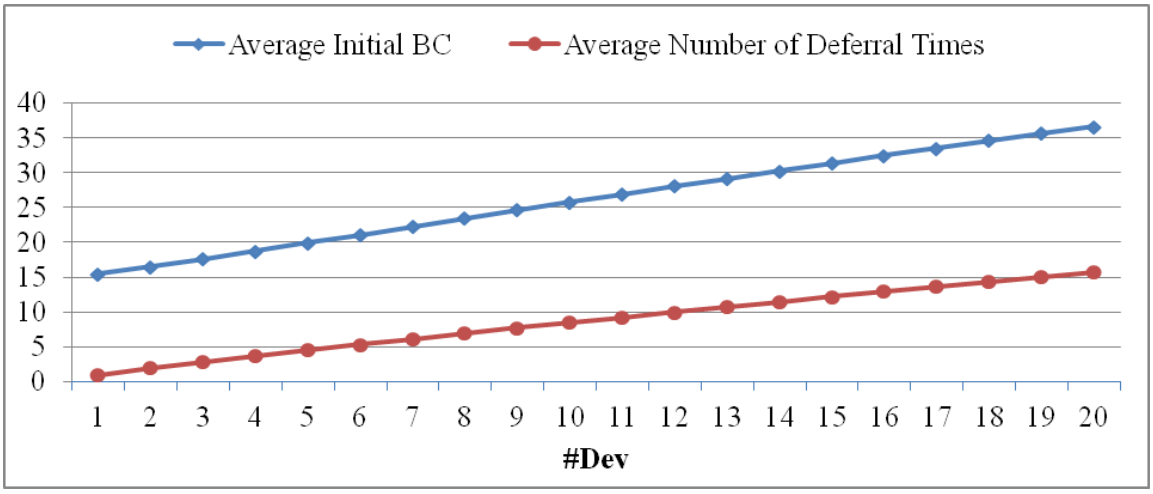

Figure 4.4: The Curves Fitted to the Average Initial BC and Deferral Number Results

The results from this computer simulation are plotted in Figure 4.4. We fitted quadratic curves to the data which are described by the following functions:

$$\overline{BC} = 14.105671 + 1.1986482x - 0.0034232245x^2 \qquad (4.24)$$

And

$$\overline{\#}_{defer} = 0.2197168 + 0.88068625x - 0.0052725504x^2 \quad (4.25)$$

Where the variable *x* represents the average contention, i.e. the $\overline{cont(k)}$ value estimated in a given measurement interval for node *k*.

4.2.3.2 Phase 1: Initialisation and Configuration Phase

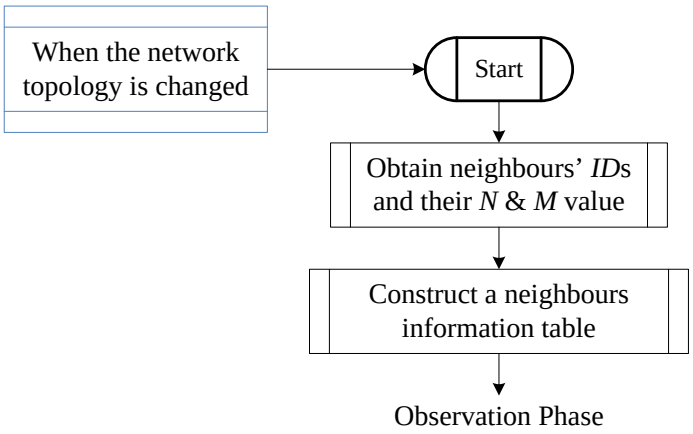

Figure 4.5: Flow Chart of the Initialisation and Configuration Phase

The values of *N* and *M* are determined before an observation phase as shown in Figure 4.5. Every node constructs a neighbour table containing the node identities (*ID*) with their corresponding *N* and *M* values. This table is updated whenever the network topology changes. Figure 4.6 is an example of the neighbour table corresponding to the network topology in Figure 4.2. Once a new node joins or a node leaves the network, the Neighbour Information Tables of the nodes affected by the change on the medium is updated.

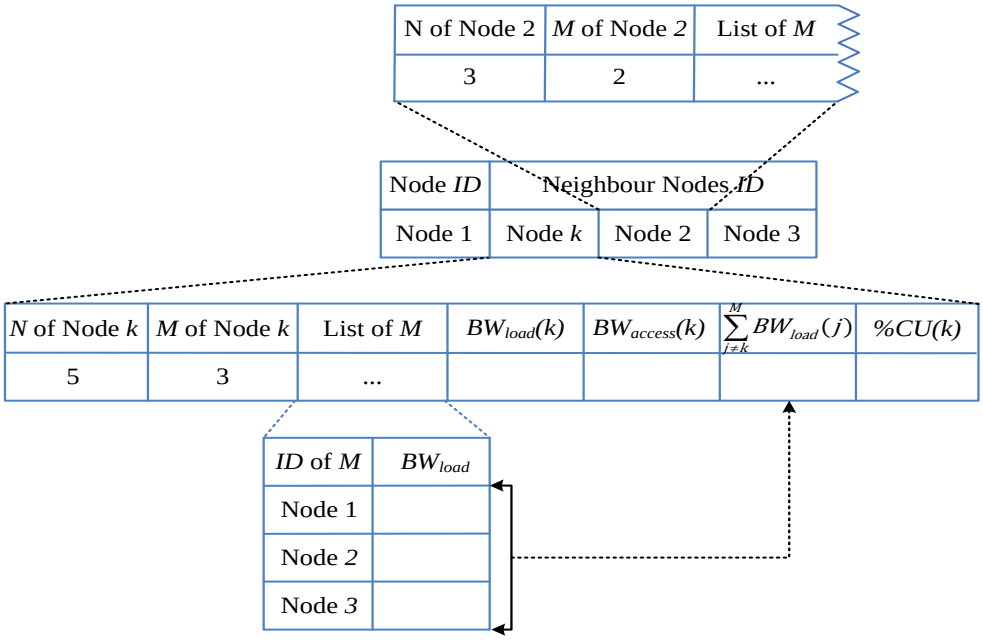

Figure 4.6: Neighbour Information Table at the Remote Observer Node

4.2.3.3 Phase 2: Observation Phase

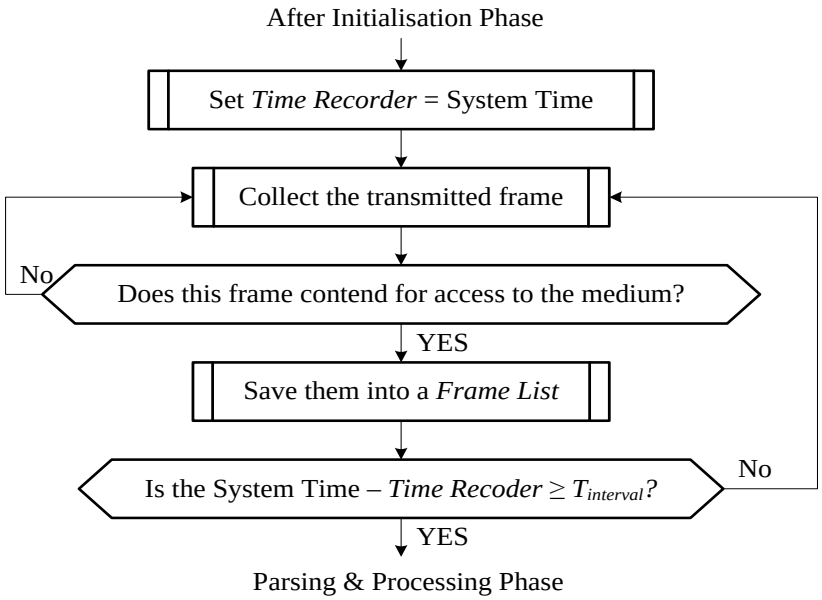

Figure 4.7: Flow Chart of the Observation Phase

The observer node captures frame transmissions from all the neighbour nodes within its reception range. It only captures the frames that contend for access to the medium (i.e. data frames and management frames) and then processes the frames transmissions within

a specified measurement time interval ($T_{interval}$). In our algorithm as shown in Figure 4.7, the variable *Time Recorder* is used to record the local system time and to calculate the observation interval. The observer node collects the transmitted frames within its own reception range on the medium and saves them sequentially in a buffer called the *Frames List*.

Choosing a longer $T_{interval}$ to observe the neighbours' transmissions leads to higher accuracy in the estimation but also leads to an increased estimation delay which may be unacceptable. For example, taking a long time to detect saturation may lead to a higher probability of buffer overflow. Using a shorter $T_{interval}$ for observing neighbours' transmissions has a poorer accuracy, but produces a faster saturation detection result. Thus an appropriate time to minimize the cost while forming an accurate estimate is required [144]. In this thesis, for convenience, the value of $T_{interval}$ is set to 1 second.

4.2.3.4 Phase 3: Parsing and Processing Phase

After the observation phase, the node estimates the *Capacity Utilization* value of its neighbour nodes by using a series of operations that includes calculating the neighbour nodes traffic load, measuring their contention and calculating their MAC bandwidth components.

The flow chart in Figure 4.8 shows the different steps in estimating the *Capacity Utilization* in the parsing and processing phase. This phase consists of two parts: BW_{load} estimation and BW_{access} estimation. The former can be divided into the observed node *k*'s load bandwidth $BW_{load}(k)$ estimation and its neighbours' load bandwidth $\sum_{j\neq k}^{M} BW_{load}(j)$ estimation. The $BW_{access}(k)$ estimation will be described in section 4.2.3.4.2.

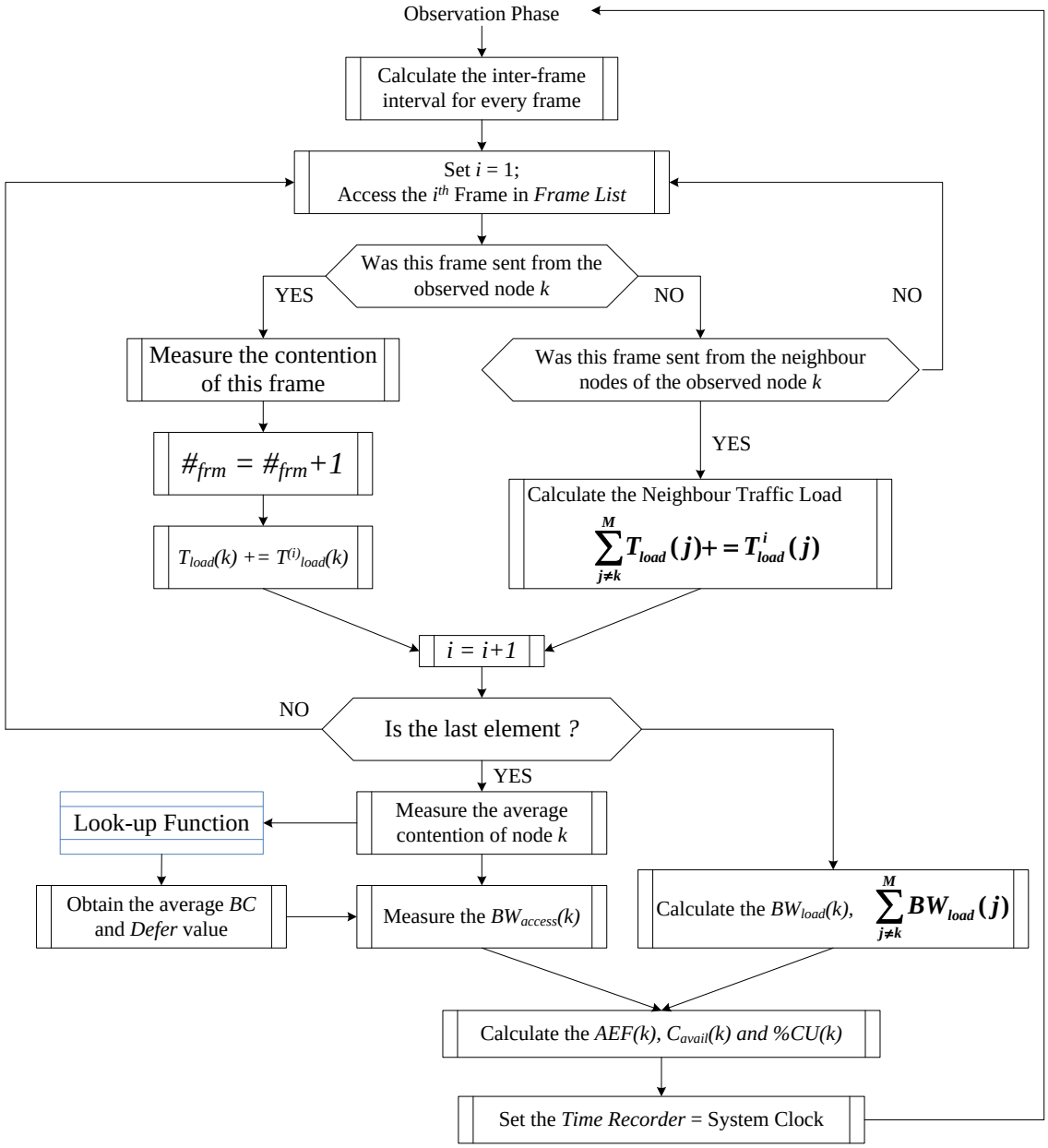

Figure 4.8: Flow Chart of the Parsing and Processing Phase

4.2.3.4.1 Load Bandwidth Measurement

The load time of the i^{th} frame $T_{load}^{(i)}(k)$ transmitted by node k is calculated using (4.26), where *FrameSize* (also denoted as PLCP Service Data Unit (PSDU)) consists of a MAC header, frame body and FCS (frame check sequence) as shown in Figure 4.9.

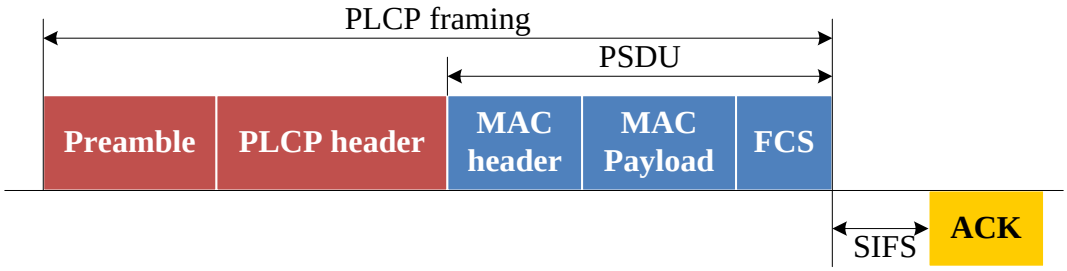

Figure 4.9: The DSSS PLCP Framing Format in a Successful Transmission

$T_{Preamble}$ and T_{SIFS} represent the Physical Layer Convergence Protocol (PLCP) preamble and header transmission time, and the short inter-frame interval (SIFS) respectively, T_{ACK} is the acknowledgement frame transmission time ($T_{ACK} = \frac{FrameSize*8}{Basic\_Rate}$). *Line_Rate* is the PHY transmission rate and *Basic_Rate* depends on the physical layer used.

$$T_{load}^{(i)}(k) = \frac{FrameSize * 8}{Line\_Rate} + T_{Preamble} + T_{SIFS} + T_{ACK} \quad (4.26)$$

Then we substitute this equation into equation (4.7) to obtain the load bandwidth of the observed node and its neighbour nodes. It should be noted that this expression only holds for the DSSS modulation scheme used in the IEEE 802.11b standard. Different standards (e.g. OFDM and MIMO) define different frame durations which are well defined in their respective standards.

4.2.3.4.2 Access Bandwidth Measurement

Under the IEEE 802.11 DCF mechanism, the initial backoff counter value and the number of deferral times are random and cannot be predicted but their average values can be estimated. Only the frames that wait for at least DIFS and contend for the medium are included in the contention measurement. The flow chart of the access bandwidth measurement process is shown in Figure 4.10. It consists of five steps: (1) select a sliding window size, (2) inter-frame interval calculation, (3) contention measurement for a single frame, (4) average contention for a node, and (5) BW_{access} measurement for a node.

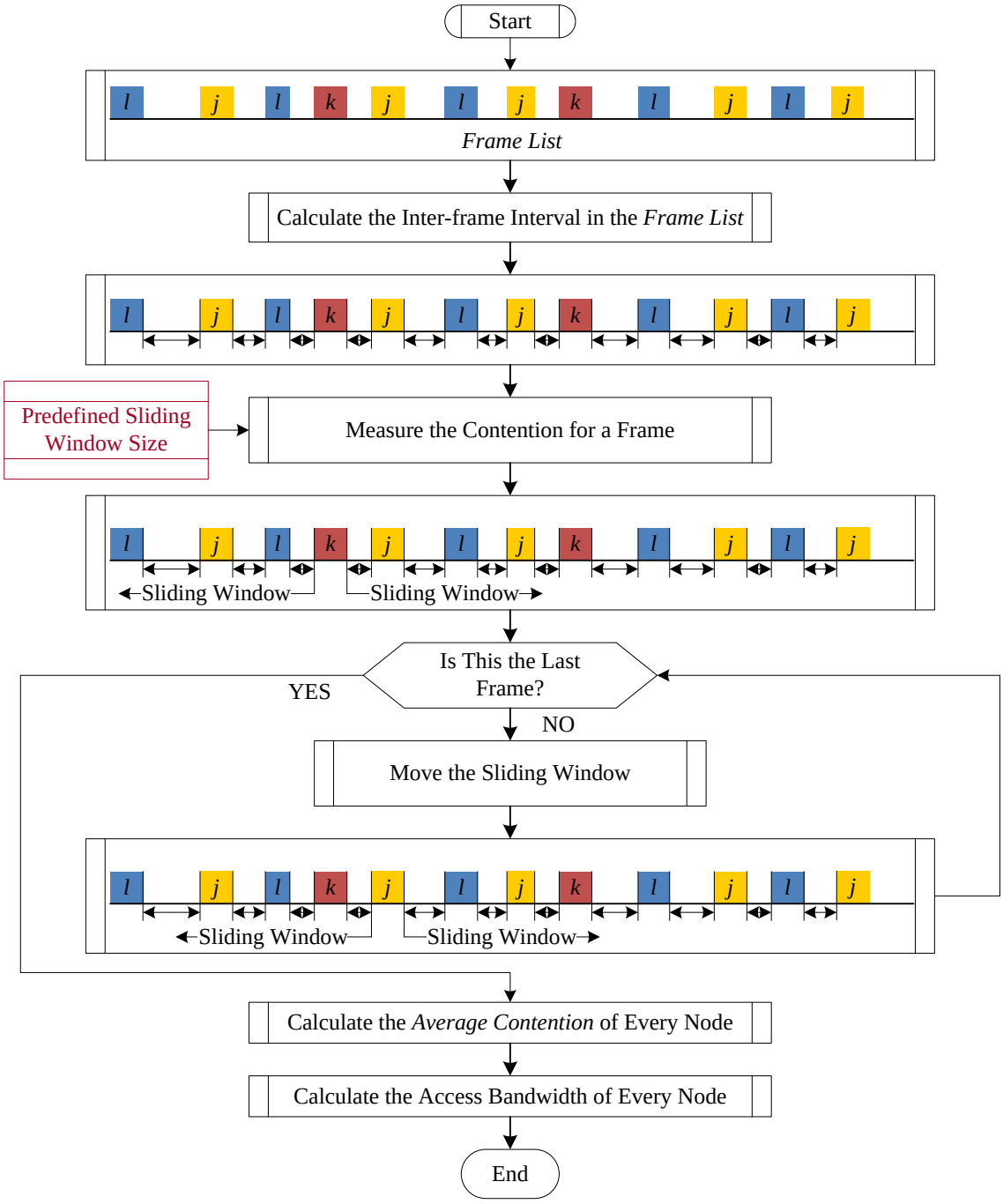

Figure 4.10: Flow Chart of the Contention Measurement

1. A Sliding Window Scheme

It is difficult to measure how many nodes are contending for the medium at any given time, but we can estimate how many frames are contending with the current transmitted frame to estimate the *average contention* of a node. We use a sliding window scheme to measure the inter-frame idle time intervals between the two successive frames

transmissions and estimate the contention of the transmitted frames. Once a frame sent from other neighbour nodes are within the scope of the sliding window of the frame sent from the observed node, we can assume that this frame contends with the frame transmitted by the observer node. The size of the sliding window is given by the average backoff counter value plus a DIFS. Therefore, the size of the sliding window is determined using:

$$Sliding\ Window = \frac{CW}{2} \times SlotTime + DIFS \tag{4.27}$$

In this thesis, $\frac{CW}{2}$ represents the average backoff counter value where *CW* is the contention window size which can be set as 31, 63 or 127 etc depending on whether frame retransmissions are being attempted. Selecting different *CW* values may cause a different impact on the accuracy of the *average contention* estimation. For example, using 31 as the *CW* value may lead to an underestimation of the *average contention* estimation when the retransmission probability is higher. Based upon the test scenarios investigated in this thesis and owing to high collision probabilities and resultant retransmissions in our scheme (due to "hidden nodes") that results in a doubling of the contention window size, we have selected *CW* as 63. Thus we set the width of sliding window as 680 μs. The *average contention* estimation will be overestimated when the retransmission probability is lower. Therefore, the window width should be adaptively adjusted, but this is beyond the scope of this thesis and can be investigated in future work.

2. Inter-frame Interval Calculation

The inter-frame interval for the i^{th} frame experienced by a node is $\Delta T^{(i)}(id)$ and is calculated as:

$$\Delta T^{(i)}(id) = T_s^{(i)}(id) - T_e^{(i-1)} \tag{4.28}$$

Where *id* is the source address of the frame and *i* is the capture sequence number, $T_s^{(i)}(id)$ represents the start time of i^{th} frame sent in the medium and $T_e^{(i-1)}$ is the end time of the $(i\text{-}1)^{th}$ frame transmitted on the medium (including the ACK time) as shown in Figure 4.11.

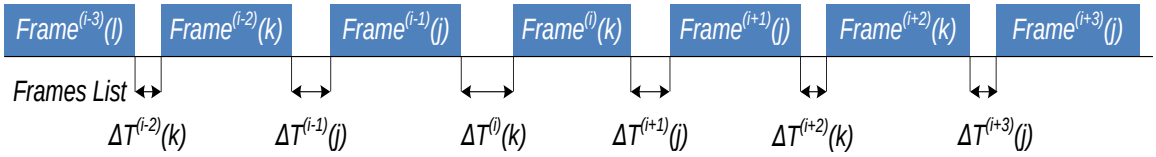

Figure 4.11: The Inter-Frame Intervals for the Transmitted Frames on the Medium

3. Contention Experienced by a Frame

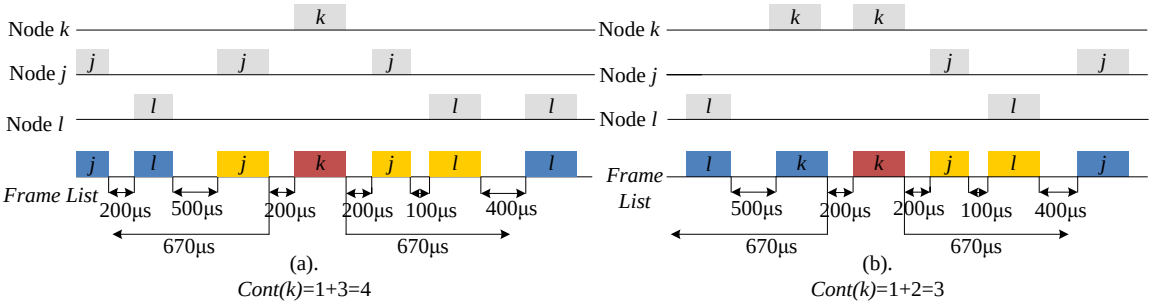

Figure 4.12: Contention Measurement for a Frame

The computational scheme for measuring the contention of a frame is described as follows:

1) Calculate the inter-frame interval for all frames in a *Frame List*.

2) For the frame in "red" as shown in Figure 4.12 (a) and (b), initialize its contention value (also referred to as *cont*) to 1 which represents the case where there is one node contending for the medium.

3) Calculate the number of contending frames within the width of the sliding window. There are 3 "yellow" frames within the width of sliding window in Figure 4.12 (a) which contend with the "red" frame, thus *cont(k)* = 4.

4) If the previous or the subsequent frames in the *Frame List* have the same *id* as the "red" *frame(k)* within the width of the sliding window as shown in Figure 4.12 (b), these frames will not be included. All the previous or the subsequent frames in blue are also excluded as they have contended with the previous or subsequent frame sent from node *k*, thus *cont(k)* = 3.

4. *Average Contention* Measurement

The *average contention* can be measured from the sum of the *cont(k)* values experienced by every frame in the measurement interval. For example, the node *k* transmits a total number of frames $\#_{frm}$ over a measurement interval. The i^{th} frame measures its contention $cont^{(i)}(k)$ with other frames. The *average contention* over a measurement interval can be written as:

$$\overline{cont}(k) = \frac{\sum_i cont^{(i)}(k)}{\#_{frm}} \tag{4.29}$$

Where $cont^{(i)}(k)$ is measured using the sliding window based contention process as described in the previous section.

5. $BW_{access}(k)$ Measurement

In a measurement interval of interest, all the contention of the frames from the same source address are summed as in equation (4.29) and then substituted into equations (4.24) and (4.25) to produce the $\overline{BC}^{mesr}$ and $\overline{\#}_{defer}{}^{mesr}$ value of the observed node *k* measured by the neighbour node. Then substituting these two parameters into equations (4.13) and (4.14) respectively, the access bandwidth $BW_{access}(k)^{mesr}$ of node *k* observed and measured by the neighbour node is given by:

$$BW_{access}(k)^{mesr} = \left(\overline{BC}^{mesr} \times SlotTime + \overline{\#}_{defer}{}^{mesr} \times DIFS\right) \times \#_{frm} \tag{4.30}$$

4.3 Model and Error Analysis

This section presents an analysis of our *Capacity Utilization* estimation model. In making a remote measurement of a neighbour's *Capacity Utilization* value, there will be an error associated with this measurement owing to the differences in the wireless medium as observed at the different locations of the nodes. For example, errors can arise due to the neighbour node having neighbours that cannot be observed by the node performing the measurement. As the observer node and the observed node do not experience the same medium, the estimation of *Capacity Utilization* value performed by the observer node will differ from that experienced by the observed node itself.

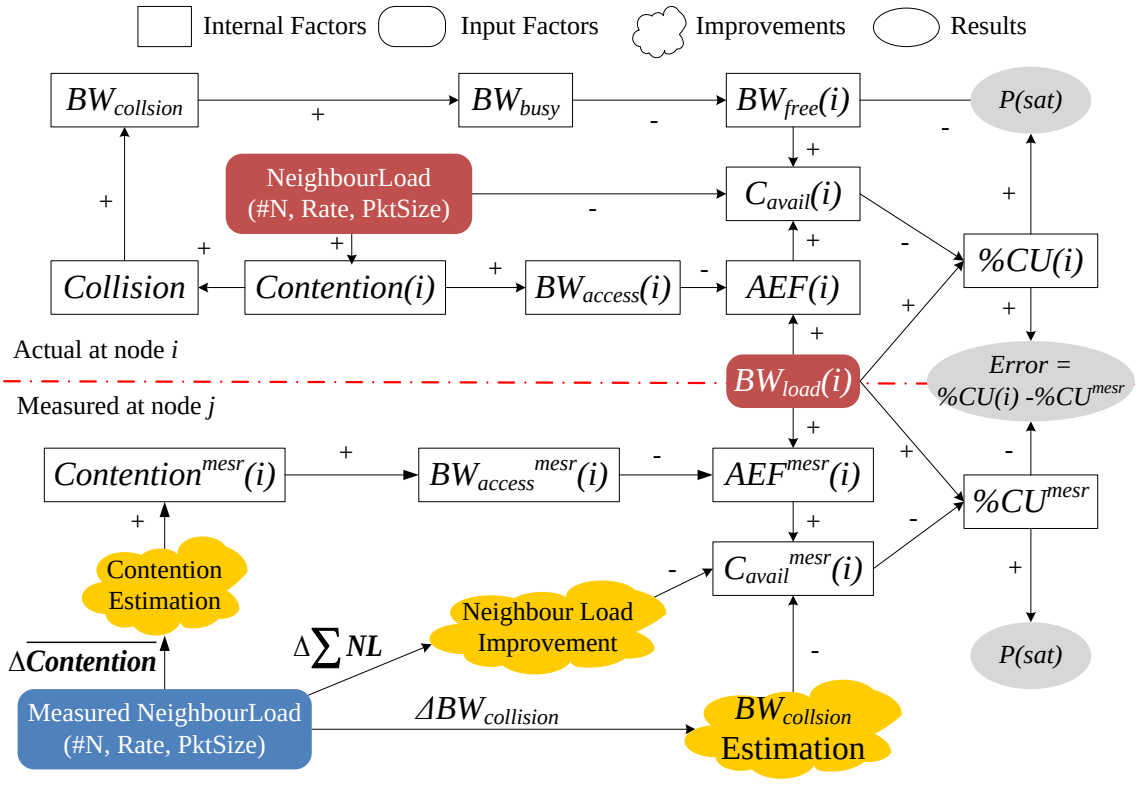

Figure 4.13: The Interaction Model of the Factors Affecting the Error associated with the Remote *Capacity Utilization* Estimator

Figure 4.13 shows the sources of error associated with our *Capacity Utilization* estimator where the symbols "+" and "-" represent the in-phase and anti-phase dependencies

between parameters respectively. From the perspective of the observed node, the process above the red dashed line in Figure 4.13 shows the interactions among the different parameters such as network traffic and topology, MAC components, node capacity and *Capacity Utilization* measured locally at the observed node (referred to as the actual value). A higher neighbour load (that may arise from a large number of neighbours, a high packet rate and a large packet size of traffic load) leads to an reduction in the node *k's* capacity $C_{avail}(k)$. More neighbours contending for the medium or sending a traffic load with higher packet rate increases the contention. If the contention of a node increases, then $BW_{access}(k)$ will increase, which means that the Access Efficiency Factor (*AEF*) will decrease. This will also reduce the $C_{avail}(k)$ value. The reduction of $C_{avail}(k)$ in turn produces an increase in *%CU(k*) that leads to high *Capacity Utilization* for node *k*. Moreover, a high contention causes high collision probabilities which gives rise to an increase in failed transmitted frames, $BW_{collision}$ and BW_{busy}. This will lead to a reduction in $BW_{free}(k)$ which decreases $C_{avail}(k)$ based upon equation (4.16) that result in a high *Capacity Utilization* value. The higher the *Capacity Utilization* experienced by the observed node, the larger the error associated with the remote measurement through neighbour observations.

From the perspective of a neighbour node that measures the *Capacity Utilization* of the observed node remotely as shown in the chart below the red dashed line in Figure 4.13 (referred to as the measured value), the "missing" information regarding the neighbour load causes an underestimation of the contention of node *k* and of the load bandwidth of the neighbours. The lack of information regarding failed transmitted frames leads to an overestimation of $BW_{free}(k)$. All of these underestimations will produce an overestimated $C_{avail}(k)^{mesr}$ value and an underestimated load bandwidth $BW_{load}(k)$ for the observed the node *k* which in turn reduces the measured *Capacity Utilization*

value $\%CU(k)^{mesr}$. This leads to an underestimation of the actual *Capacity Utilization* value. The greater the difference in the state of the medium, the larger the error involved in the estimation of the *Capacity Utilization* value. For example, the larger the number of unobservable neighbour nodes of the observed node or the higher the traffic load (consisting of a high packet rate and large packet size), the greater the error associated with our estimator. However, the "missing" information cannot be obtained directly but estimated through measurements on a set of observable variables [145, 146] such as the number of known neighbours, the number of unobservable neighbours of the observed node, and the captured transmitted frames on the medium, thus the error associated with the remote *Capacity Utilization* estimation can be minimized.

Based upon the error model analysis in Figure 4.13, some predictions regarding the performance of the estimator are listed below:

- The partially observable traffic load due to the number of unobservable neighbours causes an underestimation of the neighbour load and contention where the greater the number of unobservable neighbours, the larger the error associated with the *Capacity Utilization* estimation.
- With an increase in the traffic load sent from neighbour nodes, the error associated with the *Capacity Utilization* estimation will be increased due to an increase in the unobserved neighbour traffic load. A large packet size, a high packet rate for the neighbour load and a large number of neighbour nodes will lead to high traffic load in the network.
- High collision probability due to a large number of neighbours, a high packet rate associated with the traffic load or a smaller interference range without the CTS/RTS mechanism may cause a high failed transmission ratio which leads to an

underestimation of the neighbour load and *Capacity Utilization*.

- Both a large number of neighbour nodes and a high packet rate causes high contention for access to the network. From the perspective of measured nodes, this will lead to an overestimation of the access efficiency factor (*AEF*) and an underestimation of the *Capacity Utilization*.
- The traffic load of the observed node can be monitored by its neighbours, thus it does not affect the accuracy of the estimator.

In the chapter 5, the simulation results will be analyzed to validate these predictions. Generally, the main sources of error in the estimation of *%CU* are the error in the calculation of the neighbour load bandwidth and the error in the measurement of the contention experienced by the observed neighbour nodes as shown by a comparison of equation (4.31) and equation (4.32), where $C_{avail}(k)^{actual}$ is the actual capacity of node *k* and $C_{avail}(k)^{mesr}$ is the remote measured capacity of node *k* by the neighbour node. $BW_{access}(k)^{Unobsr}$ and $BW_{access}(k)^{Obsr}$ are the unobserved and observed access bandwidth of the observed node measured by its neighbour nodes respectively. $BW_{load}(k)^{Unobsr}$ and $BW_{load}(k)^{Obsr}$ are the unobserved and observed load bandwidth of any node *k* measured by its neighbour node respectively.

The actual $C_{avail}(k)^{actual}$ measured at the observed node is:

$$C_{avail}(k)^{actual} = \frac{BW_{load}(k)}{BW_{load}(k) + (BW_{access}(k)^{Obsr} + BW_{access}(k)^{Unobsr})} \times \left[1 - (\sum_{j \neq k}^{Obsr} BW_{load}(j) + \sum_{j \neq k}^{Unobsr} BW_{load}(j)) - BW_{collision}\right] \quad (4.31)$$

Where

$$BW_{load}(k) = BW_{load}(k)^{Obsr} + BW_{load}(k)^{Unobsr}$$

The $C_{avail}(k)^{mesr}$ measured at the remote node is:

$$C_{avail}(k)^{mesr} = \frac{BW_{load}(k)^{Obsr}}{BW_{load}(k)^{Obsr} + BW_{access}(k)^{Obsr}} \times \left[1 - \sum_{j \neq k}^{Obsr} BW_{load}(j)\right] \quad (4.32)$$

Based on a comparison between the actual and measured *Capacity Utilization* value, it is easy to identify the key parameters that determine the magnitude of the error. Once the values of $BW_{access}(k)^{Unobsr}$, $\sum_{j \neq k}^{Unobsr} BW_{load}(j)$ and $BW_{collision}$ can be estimated, the error associated with *Capacity Utilization* estimations can be minimized.

4.4 Improving the Accuracy of the Remote *Capacity Utilization* Estimator

As a consequence of using a passive remote observation technique to detect node saturation in neighbour nodes, the errors associated with the *Capacity Utilization* value estimation are unavoidable but they can be minimized.

4.4.1 Assumptions

Three simple and reasonable assumptions are now introduced to minimize the error associated with this estimation.

***Assumption 1*: *The mean load of the unobservable neighbour nodes is equal to the mean load of the observable nodes*.** This assumption has been shown to reduce the error associated with the estimation of *Capacity Utilization* for the majority of the test cases considered. This will be discussed later in section 5.1.2.

***Assumption 2*:** One can correct the contention calculation of the observed neighbour node measured by the monitor node by assuming that *all the frames from the unobservable neighbours of the observed node are transmitted with a uniform time interval within the measurement interval of interest.* This assumption is related to the first assumption as the

observed node and its neighbour nodes share the same medium.

***Assumption 3*:** In order to avoid a double counting of the failed bandwidth, halving the load bandwidth of failed transmissions under the assumption that *most collisions involve just two nodes.*

4.4.2 An Improved Remote Node *Capacity Utilization* Estimator

This section describes how to improve the estimate of the neighbours' *Capacity Utilization* by employing the three assumptions. The flow chart of the model modifications is shown in Figure 4.14. The "Yellow" box employs *Assumption 1* to improve the calculation of the neighbour load bandwidth. The "Blue" box represents the correction of the contention by using *Assumption 2* in order to enhance the accuracy of *average contention* and BW_{access} estimation, and the "Red" box shows the consideration of failed transmissions on the medium. This modified algorithm will be described in the following subsections.

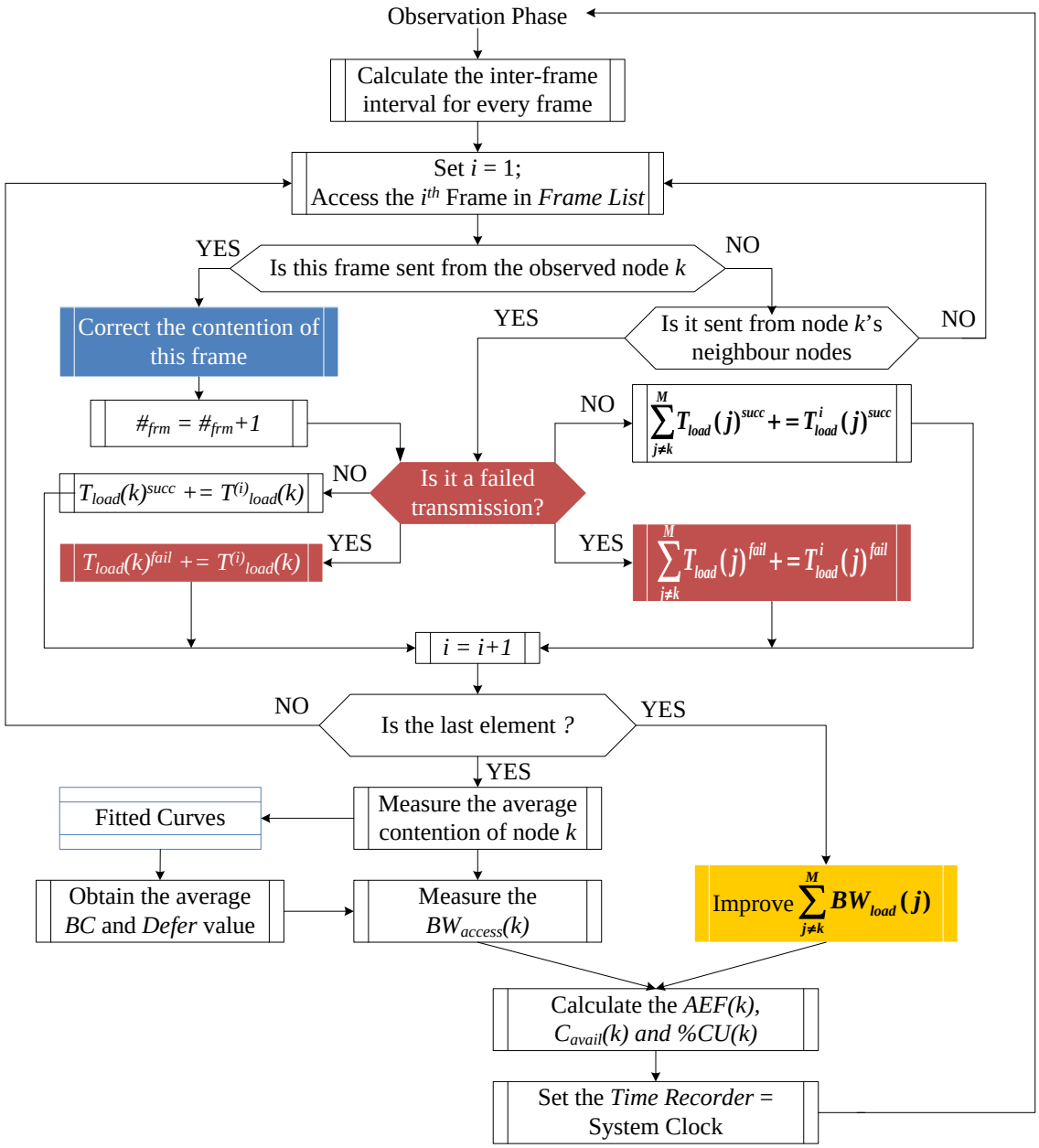

Figure 4.14: Flow Chart Showing the Modifications to the *Capacity Utilization* Estimation

4.4.2.1 Neighbour Load Improvement

Suppose that the observed node k has N neighbours and the observer node can monitor M neighbours of node k. From the perspective of the observer node, the total neighbour traffic load of node k that can be observed is $\sum_{j\neq k}^{Obsr} BW_{load}(j)$. Then using *Assumption 1* to estimate the neighbour load $\sum_{j\neq k}^{Est} BW_{load}(j)$ which is given by:

$$\sum_{j\neq k}^{Est} BW_{load}(j) = \sum_{j\neq k}^{Obsr} BW_{load}(j) + \sum_{j\neq k}^{Obsr} BW_{load}(j) \times \frac{N-M}{M} \tag{4.33}$$

4.4.2.2 Contention Correction

Suppose the total number of frames observed by neighbour node in the reception range of node *k* that contends for access to the medium is $\#_{frm}{}^{Obsr}$, and the average frame size is *AvgFS*. According to *Assumption 1* and *Assumption 2*, the steps for correcting the contention measurement are as follows:

1) Estimate the number of unobserved frames $\#_{frm}{}^{Unobsr}$ sent from the neighbour nodes of the observed node by using the number of frames than can be monitored $\#_{frm}{}^{Obsr}$. The estimate is provided by the following expression:

$$\#_{frm}{}^{Unobsr} = \#_{frm}{}^{Obsr} \times \frac{N-M}{M} \tag{4.34}$$

2) In the measurement interval of interest, the normalized uniform time interval of the unobserved frames is assumed to be $\frac{1}{\#_{frm}{}^{Unobsr}}$.

3) "Inserting" the $\#_{frm}{}^{Unobsr}$ number of estimated unobserved frames uniformly into the measurement interval.

4) Then the *average contention* of observed node is measured as described in section 4.2.3.4.2.

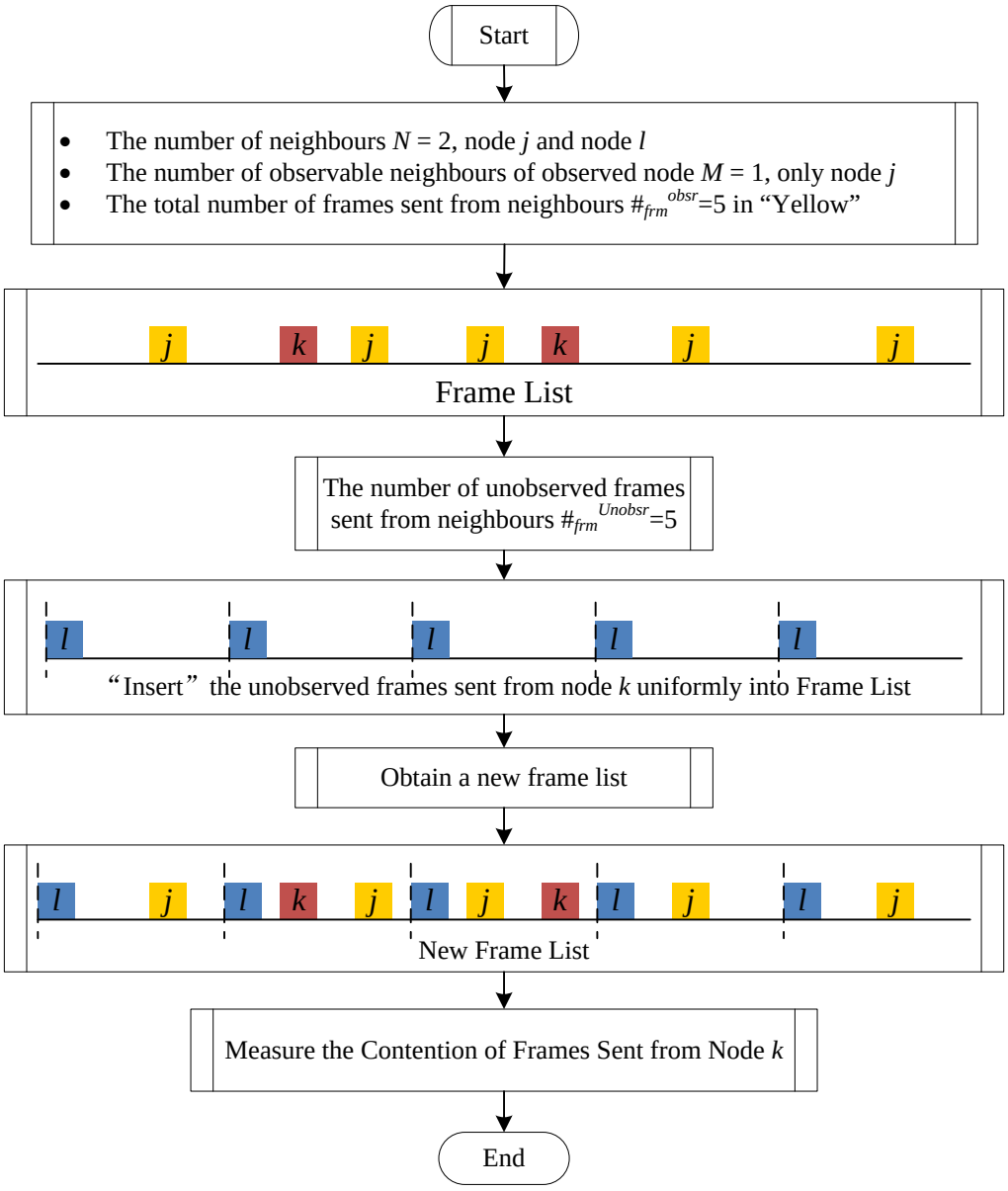

Figure 4.15: Flow Chart of the Contention Correction Calculation

The flow chart shown in Figure 4.15 describes an example of the contention correction scheme.

4.4.2.3 Halving the Failed Retransmission Bandwidth

In order to improve the accuracy of our *Capacity Utilization* estimator, the failed transmitted frames due to collisions or transmission errors need to be considered. Similar to successfully transmitted frames, the failed frames also consume the channel bandwidth and contribute to BW_{busy}. The load time of the i^{th} frame $T_{load}^{(i)}(k)$ sent from node k can be

divided into two categories: a successful transmission $T_{load}^{(i)}(k)^{succ}$ calculated using (4.26) or a failed transmission $T_{load}^{(i)}(k)^{fail}$. Once a collision occurs at the receiver node, the node cannot observe this failed transmitted frame. However, the node can detect the failed transmission through the next successfully transmitted frame where the re-transmission flag has been set to 1 as illustrated in Figure 4.16.

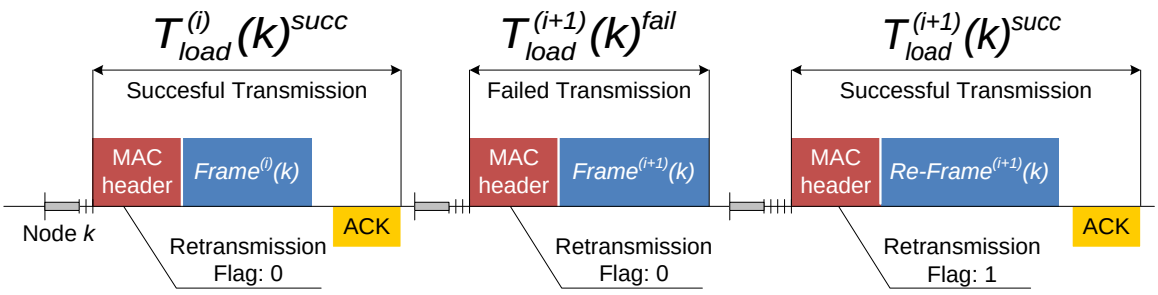

Figure 4.16: T_{load} Measurement of Node *k* for its Successful and Failed Transmissions

If the sender node *k* does not receive an ACK, $T_{load}^{(i)}(k)^{fail}$ is measured as:

$$T_{load}^{(i)}(k)^{fail} = \frac{FrameSize * 8}{Line\_Rate} + T_{Preamble} \tag{4.35}$$

Then converting this to a fraction of the medium bandwidth as follows:

$$BW_{load}(k)^{succ} = \frac{\sum_i T_{load}^{(i)}(k)^{succ}}{T_{busy} + T_{idle}} \tag{4.36}$$

and

$$BW_{load}(k)^{fail} = \frac{\sum_i T_{load}^{(i)}(k)^{fail}}{T_{busy} + T_{idle}} \tag{4.37}$$

Based upon the IEEE 802.11 MAC frame format, it is difficult to estimate how many failed transmissions have occurred before a successful re-transmission. Generally, the probability of second retransmission (i.e. with probability p^2) will be far lower than the probability of the first retransmission p. Thus we assume that all re-transmitted frames

arise from a single failed transmission in this thesis. However, it can be shown that the probability of a second retransmission can be higher than the probability of the first retransmission under certain conditions [147]. This assumption may not be appropriate under heavy load conditions where the probability of a second or more retransmission attempt may not be negligible. This is a challenge for the *Capacity Utilization* estimation performed by remote observation and the improvement of accuracy will be investigated in the future works.

Then substituting equations (4.36) and (4.37) into (4.38) gives:

$$BW_{load}(k) = BW_{load}(k)^{succ} + BW_{load}(k)^{fail} \tag{4.38}$$

The frames transmitted by node k and its neighbour node j at the same time cause a collision to occur at the received node. The received node cannot collect the information from the two frames involved in the collision. However, the collision frames occupy the same channel period and the occupied time interval depends on the frame with the larger transmitted time as shown in Figure 4.17.

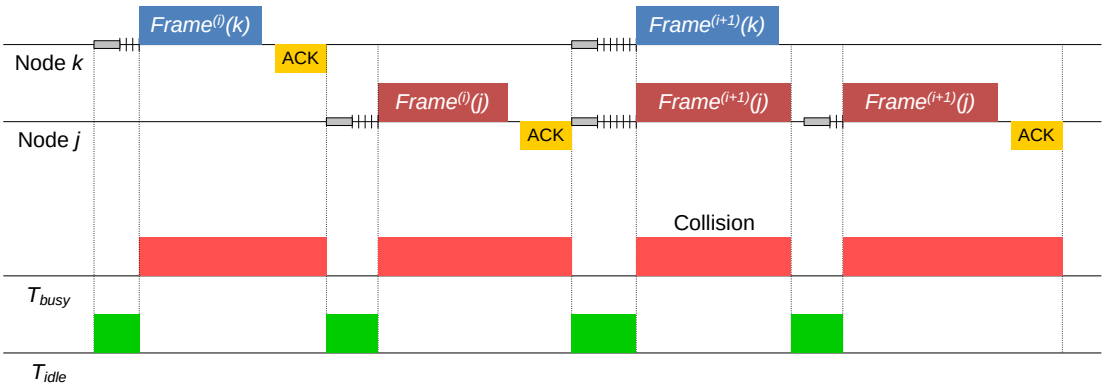

Figure 4.17: The “Double Counting” Problem arising from Collisions

As is shown in Figure 4.17, by simply summing all the failed bandwidth of neighbour nodes of observed node, the bandwidth $BW_{collision}$ will be double counted or even counted multiple times. This will lead to an overestimation of the *Capacity Utilization* for

the observed node. Thus under *Assumption 3*, simply halving the failed bandwidth can increase the accuracy of the remote estimator as:

$$\sum_{j\neq k}^{Est} BW_{load}(j)^{mesr} = \sum_{j\neq k}^{Est} BW_{load}(j)^{succ} + \frac{1}{2}\sum_{j\neq k}^{Est} BW_{load}(j)^{fail} \tag{4.39}$$

4.4.2.4 *Capacity Utilization* Improvement

Finally, substituting the equations (4.33) and (4.39) into equation (4.22), and using the new contention measurement as described in section 4.4.2.2, then the node capacity of *k* estimated by the neighbour node $C_{avail}(k)^{est}$ can be derived as:

$$C_{avail}(k)^{est} = AEF(k)^{est} \times \left[1 - (\sum_{j\neq k}^{M} BW_{load}(j)^{succ} + \sum_{j\neq k}^{M} BW_{load}(j)^{succ} \times \left(\frac{N-M}{M}\right)) - \frac{1}{2}\left(\sum_{j\neq k}^{M} BW_{load}(j)^{fail} + \sum_{j\neq k}^{M} BW_{load}(j)^{fail} \times \frac{N-M}{M}\right)\right] \tag{4.40}$$

Then substituting the equation (4.40) and equation (4.33) into equation (4.23), the final form of the remote estimator of the node *Capacity Utilization* ($\%CU^{est}(k)$) is given as:

$$\%CU(k)^{est} = \frac{BW_{load}(k)^{succ} + BW_{load}(k)^{fail}}{C_{avail}(k)^{est}} \times 100\% \tag{4.41}$$

4.5 Statistical Characterization of the Estimator Error

This thesis is essentially a study of the performance of the estimator, i.e. how accurately can the estimator measure the actual *Capacity Utilization* experienced by a node. The main source of error in the estimator is the traffic load experienced by the node being observed which may not be the same as that experienced by the node performing the estimation, i.e. the monitor mode. Specifically, the error arises from the traffic load of

those nodes that cannot be observed by the monitor node.

For the purposes of the analysis, this source of error is considered to be a random variable. Consequently, the objective of the analysis is to ascertain the impact of this unobservable traffic load on the accuracy of the estimator. Therefore, this analysis will necessarily involve a statistical characterisation of the error associated with the estimate of *Capacity Utilization* produced by the estimator. We use the absolute relative error (*ARE*) to measure the accuracy of our remote *Capacity Utilization* estimator which is defined as:

$$Absolute\ Relative\ Error = \left| \frac{\%CU(k)^{actual} - \%CU(k)^{est}}{\%CU(k)^{actual}} \right| \qquad (4.42)$$

Where $\%CU(k)^{actual}$ is the actual *Capacity Utilization* of node *k*, and $\%CU(k)^{est}$ is the remote estimated *Capacity Utilization* of node *k* by its neighbour node. Then we use the ns2 simulator to generate a large numbers of topologies with random node locations and traffics. The analysis of the results will then be used to validate our model and assumptions. The simulation scenarios and results are discussed in Chapter 5.

4.6 Node Saturation Detection

4.6.1 A New Algorithm in Detecting Node Saturation

A new algorithm to detect node saturation that combines the remote *Capacity Utilization* estimator and a simple Bayesian decision process based upon a *Capacity Utilization* threshold parameter (referred to as CU^{TH}) is proposed as a potential application to illustrate the usage and show the accuracy of the *Capacity Utilization* estimator. This threshold parameter can be optimized by selecting an appropriate value for the saturation probability of the node. After the decision, the observing node can broadcast the saturation information to its neighbours using its beacon frames so that they may take preventative actions to avoid further deterioration in the node's saturation condition. The

detection algorithm is shown as Figure 4.18.

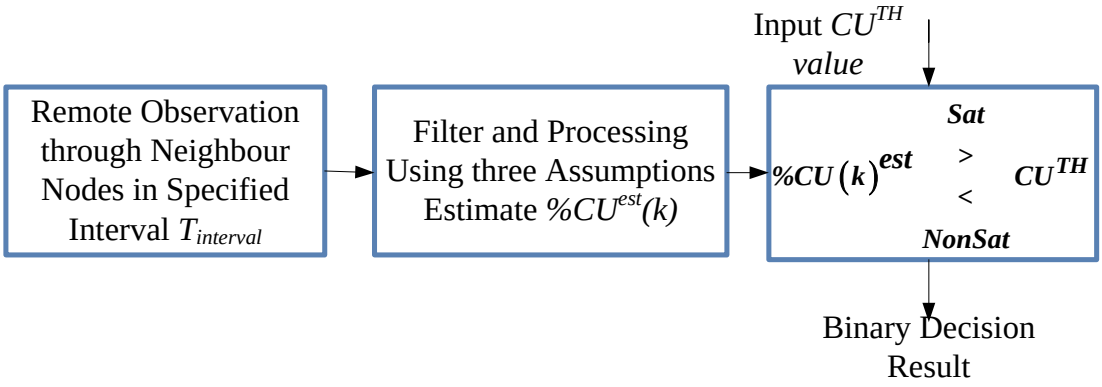

Figure 4.18: Algorithm for Node Saturation Detection

Two metrics *Failed Detection Ratio (FDR)* and *False Alarm Ratio (FAR)* are used to assess the performance of this method. *FDR* defined in equation (4.43) corresponds to the case where a saturated network node is detected as being non-saturated and indicates the accuracy with which this method can accurately and correctly detect node saturation.

$$FDR = P(\%CU(k)^{est} < CU^{TH} | Saturation) \tag{4.43}$$

FAR defined in equation (4.44) corresponds to the case where a non-saturated node is determined to be a saturated node, the consequence of this false alarm is that the resulting unnecessary preventive actions may have an adverse impact on the network performance.

$$FAR = P(\%CU(k)^{est} > CU^{TH} | NonSaturation) \tag{4.44}$$

The metrics *FDR* and *FAR* are related to the CU^{TH} value used in this method where increasing the CU^{TH} threshold increases the *FDR* and reduces the *FAR*. Similarly, reducing the CU^{TH} threshold reduces the *FDR*, but increases the *FAR*. Both *FDR* and *FAR* are unavoidable aspects of the detection algorithm and need to be balanced against each other. For example, a high *FDR* can give rise to packet losses and unacceptable delay. However, a high *FAR* is also serious due to its potential to unnecessarily waste the network resources.

A minimum error-rate classifier based upon Bayesian decision theory [148, 149] is used to perform a trade-off between *FDR* and *FAR* in finding the optimal value of CU^{TH}. The optimal CU^{TH} corresponds to the probability of node saturation in the network, i.e. the higher probability of node saturation, the lower the optimal *Capacity Utilization* threshold that can be chosen (discussed in Chapter 5).

4.6.2 The Performance of the Saturation Detection Algorithms

In order to analyse the accuracy of our remote *Capacity Utilization* estimator, we select two other simple algorithms, queue monitoring and regularly pinging (discussed in Chapter 3) to compare the *FDR* and *FAR*.

Once a node is saturated, the node cannot win a sufficient number of transmission opportunities which leads to an increase in queue size that causes the delay to increase and may lead to packet loss. Local queue monitoring has a number of advantages such as low complexity and high accuracy of detection. Most proposed detection schemes use different queue size thresholds to make a decision regarding node saturation. The threshold selection is a key problem to evaluate the detection performance. Larger threshold may cause an increased detection delay that may give a rise to a congestion condition that leads to packet losses. Owing to the dynamically changing queue size, we use the average queue size over a measurement interval as the threshold to detect node saturation for simplicity. However, it is recognised that in some cases, this approach can produce a poor measurement of saturation owing to the large variance often observed in queue lengths [150]. The average queue size threshold for determining the saturation is pre-defined as 1 (using the widely used definition adopted in Bianchi's model [30]) and make the observed node monitor its buffer during a measurement interval (we also define this interval as 1 second for comparison with our algorithm) to detect node saturation in this thesis. If this average value is greater than a predefined queue size threshold then the

node is determined to be saturated. It is to be noted the predefined queue size threshold can be defined as some other value (or may even be adaptively adjusted). However, this is beyond the scope of this thesis.

Another simple saturation detection mechanism is to have the observer neighbour nodes regularly ping the observed node to determine if they are operating normally. The observer node sends ping packets to the observed node and records the timestamp, and then the receiver node returns it immediately in order to measure the round-trip time (*RTT*). In this thesis, if the sender node cannot receive the return packets from observed node within a measurement interval (again defined as 1 second to permit a meaningful comparison of the algorithms) for measuring the *RTT*, we assume that the observed node is saturated. In order to improve the accuracy, a predefined *RTT* threshold could also be used to detect saturation. However, this is beyond the scope of this thesis. The performance of these two methods will be discussed in section 5.4.1.

The version 2.35 of the ns2 simulator used includes both queue management and the ping protocol and thus does not require any modifications to the ns2 source files.

4.7 Summary

This chapter gives a detailed description of the remote node capacity utilization estimator with a description of the MAC bandwidth components framework. It introduces the *average contention* measurement, load bandwidth measurement, access bandwidth measurement and neighbour load bandwidth measurement. Then it describes the calculation method for the *AEF*, a factor that measures how efficiently a node contends for access to the wireless medium, the node capacity and the *Capacity Utilization*. However, there will be an error associated with this measurement by the neighbour nodes owing to the differences in the wireless medium as observed at their different locations. A

model of the error associated with the remote *Capacity Utilization* estimator is used to analyse the sources of error associated with the estimator. The simulation results used for validating the predictions will be presented in Chapter 5. In order to minimize this error, three assumptions are employed to improve the remote estimator. The detection of node saturation is selected as one of the wireless applications of this estimator and two parameters *FAR* and *FDR* are employed as the performance criteria to evaluate the performance of our remote estimator.

Chapter 5 Simulation Results and Performance Evaluation

This chapter presents and analyses the simulation results and the performance evaluation of our remote *Capacity Utilization* estimator. A large number of random topologies are generated with random traffic by the ns2 simulator to validate and test the model. The absolute relative error (*ARE*) is employed to measure the accuracy of the *Capacity Utilization* estimator. Chapter 4 presented an error model associated with the estimates and made some performance predictions. This chapter analyses the accuracy of the estimator and discusses the factors that determine the accuracy of our estimator. The modifications implemented to improve the accuracy by using three assumptions are also presented in this chapter. Finally, node saturation detection is selected as an application to show the usage and accuracy of the remote *Capacity Utilization* estimator.

5.1 Simulation Set Up and Scenarios

The error associated with the remote *Capacity Utilization* estimator can be considered as a random variable. Therefore, this analysis will necessarily involve a statistical characterisation of the error associated with the estimate of the *Capacity Utilization* produced by the estimator. Different scenarios with random parameters are generated to test the accuracy of the *Capacity Utilization* estimator and to identify the factors that impact on the estimator accuracy as discussed in chapter 4.

5.1.1 Simulation Set Up

The simulation parameters pertain to IEEE 802.11b operation without the RTS/CTS mechanism and using an omnidirectional antenna. In order to avoid the interference from out of reception range nodes but within the interference range, both the reception range

and the interference range are set to 50 meters (in practice this range depends on the transmit power and operating environment). The buffer size of every node is 50.

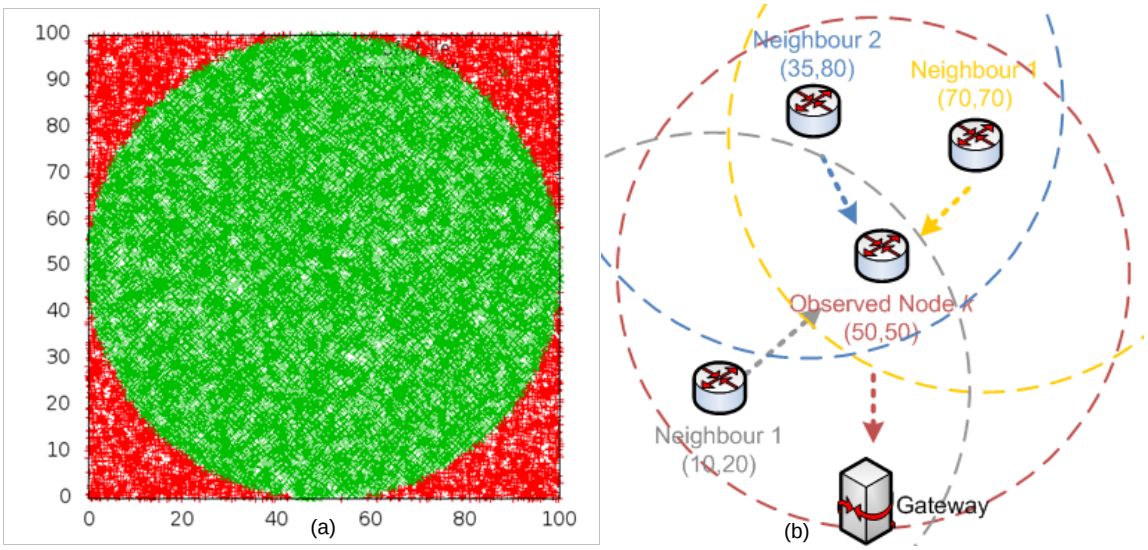

Figure 5.1: (a) Coordinate Generation of Nodes (b) An Example Topology with 3 Neighbours

In this thesis, we use the Mersenne Twister [143] algorithm as the pseudo random number generator where the seed used is the system time. In order to obtain the uniform random distribution of the nodes' positions, we first randomly generate a node's coordinates *(x, y)*, where the *x* and *y* values are random real numbers selected from [0.00, 100.00]. If the generated node is located outside of the circle in red area as Figure 5.1 (a), we drop this coordinate of the node and generate a new set of random coordinates.

Figure 5.1 (b) shows one of the generated topologies. It can be seen that the observed node has three neighbours (i.e. $N = 3$), where neighbours 1 and 2 can hear two of the observed node's neighbours transmissions (i.e. $M = 2$), and neighbour 3 only can hear one of the observed node's neighbours transmissions (i.e. $M = 1$). In order to avoid the impact of the routing protocol used, a single hop only is considered in this thesis. The neighbour nodes only send traffic to the observed node and the observed node sends its traffic load

to the gateway node. In all test scenarios, the position of the observed node and the gateway is fixed, however its neighbours' position will be randomly determined.

5.1.2 Scenarios Test

In our *Capacity Utilization* estimator, the partial observed load traffic produces an unavoidable error associated with the estimation of *Capacity Utilization* (%*CU*). We established a set of different categories of scenarios to investigate the accuracy of the *Capacity Utilization* estimator and to validate the error model discussed in chapter 4. All the following scenarios comprise 1,000 random topologies of 30 seconds simulation which results in a total of 30,000 separate %*CU* estimates for every neighbour of the observed node. The network traffic load is divided into two parts: the traffic load sent from the neighbour nodes and from the observed node. This is because they have different impacts on the accuracy of the *Capacity Utilization* estimator. All the scenarios use Poisson traffic except for scenario B-7 and C-7, as it explores the performance of our estimator under Exponential On-Off traffic. Poisson traffic is one of the most widely used traffic models employed to investigate the network performance. At a time *t*, the total number of packets *n(t)* follows a Poisson distribution with the parameters λt and can be written as:

$$P(t) = \frac{e^{-\lambda t} \times (\lambda t)^n}{n!} \tag{5.1}$$

The inter-arrival time period *T* of two consecutive packets with fixed packet size follows an exponential distribution as $F(t) = 1 - e^{-\lambda T}$. The symbol λ represents the expectation of packet arrival numbers per unit time (average packet arrival rate) and $\lambda = \frac{1}{E(t)}$. In Exponential On-Off traffic model, packets are generated with a fixed size and at a constant rate. The durations of the "On" and "Off" intervals follow a Gaussian distribution with mean values equal to predefined value. The different traffic and

topology scenarios are presented in Table 5.1. The observed node's position is fixed and its neighbour nodes are located randomly within its reception range. The parameters are uniformly randomly selected from a specific range of traffic load which is denoted as x_1~x_2, where x_1 and x_2 represent the minimum value and maximum value respectively. In this chapter, *NonModCU* and *ModCU* represent the *Capacity Utilization* estimator before and after the improvement modifications respectively.

Scenarios of Group A: These scenarios present two general cases to analyse the performance of the estimator before and after the modifications have been implemented to improve the accuracy.

Scenarios of Group B: The scenarios in this group investigate the factors that influence the accuracy of the *Capacity Utilization* estimator before the modifications are implemented.

Scenarios of Group C: The factors that have an impact on the accuracy of the *Capacity Utilization* estimator after the modifications are studied in this group.

Scenarios B-1 and C-1: The effect of the number of neighbours *(N)* located within the reception range of the observed node on the accuracy of the estimator before and after the modifications. In order to avoid an increase in the traffic load with the growth of *N*, the total traffic load of neighbour nodes for every scenario is set within a fixed range, i.e. the packet size is uniformly randomly selected between 50 bytes and 1500 bytes and the total packet rates of the neighbour nodes (referred to as NL_{pr}) are the same for all scenarios, thus the packet rate for every neighbour is $\frac{NL_{pr}}{N}$. In this thesis, in a lower traffic range case, the NL_{pr} is set to 200~300 pps. In a higher traffic range case, the total traffic load is double (i.e. 400~600 pps) that of the lower traffic range.

Scenarios B-2 and C-2: These two scenarios assess the impact of the number of neighbours of the observed node that can be monitored (*M*) on the accuracy of the estimator before and after the modifications. The results for $N = 3$ and $N = 5$ with the number of observable neighbours *M* are chosen for discussion in this chapter. The results for other scenarios are presented in Appendix C and D.

Scenarios B-3 and C-3: An investigation of the impact of the packet size of the observed node's traffic load on the accuracy of the estimator before and after the modifications is performed here. The packet rate of the observed node is uniformly randomly selected between 10 and 100 pps and the packet sizes used are: 128, 256, 512, 1024, and 1500 bytes.

Scenarios B-4 and C-4: A further investigation of the influence of the packet rate of the observed node on the accuracy of the estimator before and after the modifications. The packet size of the observed node's traffic load is random and the packet rates of the observed node's traffic load used are: 10, 25, 50, 100, 200 and 500 pps.

Scenarios B-5 and C-5: The influence of the variations of the packet size of the neighbour nodes' traffic load on the accuracy of the estimator before and after the modifications. The neighbours' traffic loads are similar to the scenarios A-3 and B-3.

Scenarios B-6 and C-6: These two scenarios investigate the influence of the traffic packet rate of the neighbour nodes on the accuracy of the estimator before and after the modifications. The neighbours' traffic loads are similar to the scenarios A-4 and B-4.

Scenarios B-7 and C-7: This scenario uses Exponential On-Off traffic to investigate the accuracy of the estimator before and after the modifications. During "on" periods, packets are generated at a constant rate. During "off" periods, no traffic is generated. Burst times and idle times are taken from exponential distributions. This traffic is more representative

of “real world” traffic than Poisson traffic [86, 151]. In order to test the impact of “on-off” period on the accuracy of the estimator, the average “on” period is set equal to average “off” period with different values (i.e. where the average “on” time = 0.5 s, 2 s and 5 s). The simulation time is 100 seconds in this scenario in order to obtain more *%CU* estimates values. The packet size and rate of traffic load for all nodes are uniformly randomly selected between 50~1500 bytes and 10~100 pps respectively.

Scenarios of Group D: The detection of node saturation is used as one of the potential applications to validate the feasibility and the accuracy of the *Capacity Utilization* estimator after the modifications compared with other two simple node saturation mechanism: queue monitoring and regularly pinging. The *FDR* and *FAR* of the node saturation are the two criteria used to evaluate the performance of the estimator in a practical application. Nine different scenarios of N = 2, 3, 4, 5, 6, 7, 8, 9, and 10, have been used to examine the *FDR* and *FAR* of this group.

Scenarios D-1: All the nodes send high traffic levels within a random selection range in these scenarios in order to force the observed node into saturation. Then the estimator is used to investigate the *Capacity Utilization* of the observed node through neighbour nodes and to calculate the *FDR*.

Scenarios D-2: In order to ensure the same saturation probability under different N cases, the total traffic load for all neighbour nodes is fixed. 1,000 topologies under non-saturated operation observed for every case are selected to calculate the *FAR*.

Table 5.1 Classification of Simulation Test Scenarios

Scenarios	*N*	Traffic load of Observed Node		Traffic load of Neighbour node		Metric	Traffic Type	Test Target
		Packet Size (*PS*)	Packet Rate (*PR*)	Packet Size (*PS*)	Packet Rate (*PR*)			
A-1	2	50~1500	10~25	50~1500	10~25	*NonModCU*	Poisson	General Case
A-2	5	50~1500	100~250	50~1500	100~250	*NonModCU*	Poisson	General Case
A-3	2	50~1500	10~25	50~1500	10~25	*ModCU*	Poisson	General Case
A-4	5	50~1500	100~250	50~1500	100~250	*ModCU*	Poisson	General Case
B-1-1	2-10	50~1500	100~100	50~1500	200~300 for total neighbour load	*NonModCU*	Poisson	Variation of *N*
C-1-1						*ModCU*		
B-1-2	2-10	50~1500	100~100	50~1500	400~600 for total neighbour load	*NonModCU*	Poisson	
C-1-2						*ModCU*		
B-2	2-10	2-10	50~1500	100~100	200~300 for total neighbour load	*NonModCU*	Poisson	Variation of *M*
C-2						*ModCU*		
B-3	5	128,156,512, 1024, 1500	10~100	50~1500	10~100	*NonModCU*	Poisson	Variation of $BW_{load}(k)$
C-3						*ModCU*		
B-4	5	50~1500	10,25,50,100, 200,500	50~1500	10~100	*NonModCU*	Poisson	Variation of $BW_{load}(k)$
C-4						*ModCU*		

Scenarios	*N*	Traffic load of Observed Node		Traffic load of Neighbour node		Metric	Traffic Type	Test Target
		Packet Size (*PS*)	Packet Rate (*PR*)	Packet Size (*PS*)	Packet Rate (*PR*)			
B-5	5	50~1500	10~100	128,156,512, 1024,1500	10~100	*NonModCU*	Poisson	Variation of $\sum_{j\neq k} BW_{load}(j)$
C-5						*ModCU*		
B-6	5	50~1500	10~100	50~1500	10,25,50,100,200, 500	*NonModCU*	Poisson	
C-6						*ModCU*		
B-7	5	50~1500	10~100	50~1500	10~100	*NonModCU*	On-OFF	Different Traffic Type
C-7						*ModCU*		
D-1	3-10	50~1500	500~1000	50~1500	500~1000	*ModCU*	Poisson	*FDR*
D-2	3-10	50~1500	10~200	50~1500	400~600 for all nodes	*ModCU*	Poisson	*FAR*

5.2 Analysis of the Accuracy of the *Capacity Utilization* Estimator without the Modifications

This section presents the simulation results for Group A and B scenarios for the *Capacity Utilization* estimator before the modifications are implemented. The factors influencing the performance of *Capacity Utilization* estimator are identified. The ground truth data (i.e. the actual MAC bandwidth components) are directly calculated from the ns2 simulator trace file. These are then compared with the results obtained indirectly from the average contention measurement and quadratic fitted curves (i.e. the *Capacity Utilization* estimator process) as shown in Figure 5.2. In this thesis, we use the empirical Cumulative Distribution Function based upon the assumptions that each error associated with the *Capacity Utilization* estimation is independent (for simplicity we use the term CDF to refer to the empirical CDF).

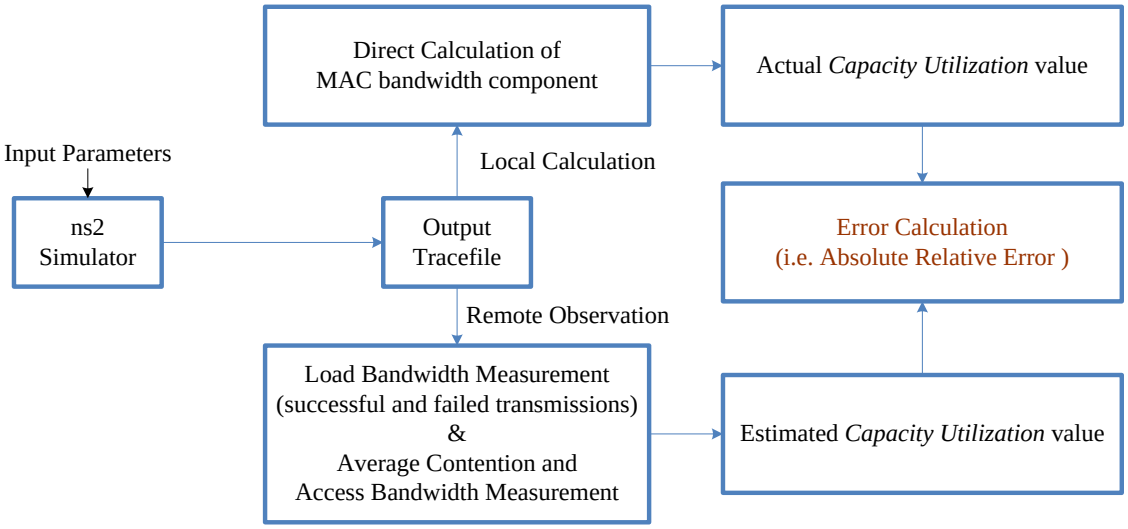

Figure 5.2: The Data Collection Process

The results for scenarios A-1 and A-2 are shown in Figure 5.3 which presents the CDF of the absolute relative error (*ARE*) of our *Capacity Utilization* (*%CU*) estimator before the modifications. The corresponding PDFs (probability density function) of all the scenarios are given in Appendix C. It can be seen here that 94% of the %*CU* estimates and 1% of

the %*CU* estimates have an *ARE* less than 10% under scenarios A-1 and A-2 respectively. Similarly, 90% of the %*CU* estimates have an *ARE* less than 9% and 64% of the *ARE* under scenarios A-1 and A-2 respectively.

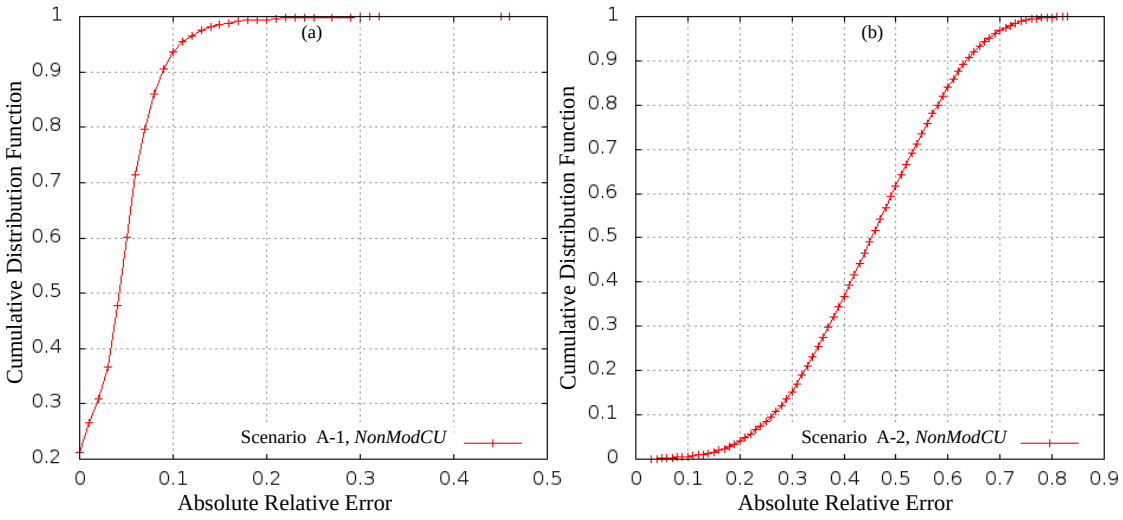

Figure 5.3: The CDF of the *ARE* for *NonModCU* in (a) Scenario A-1 and (b) Scenario A-2

This shows that the accuracy of our estimator under scenario A-1 is much higher than that under scenario A-2 due to a number of different factors such as a different number of neighbours, traffic load of the observed node and the neighbour nodes. We next investigate the impact of various factors on the accuracy of the *Capacity Utilization* (*%CU*) estimator in order to identify the sources of error from the predictions provided by the error model before the modifications are applied. The six main factors that are considered here are listed below:

- The number of neighbour nodes (N)
- The number of observable neighbours of the observed node (M)
- The traffic load of the observed node consisting of various combinations of different packet sizes and packet rates

- The traffic load of the neighbour nodes consisting of various combinations of different packet sizes and packet rates
- Traffic type

5.2.1 Different Number of Neighbour Nodes (*N*)

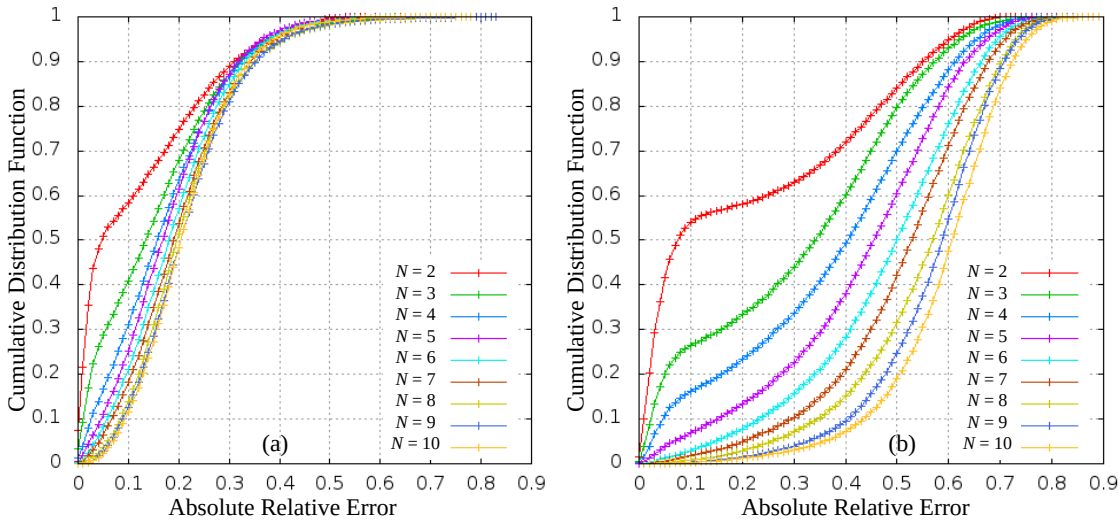

Figure 5.4: The CDF of the *ARE* of *NonModCU* under Different *N* Scenarios with (a) Lower Traffic Load (b) Higher Traffic Load

Scenario B-1-1 and B-1-2 describes the impact of the different numbers of neighbour nodes (*N* = 2, 3, 4, 5, 6, 7, 8, 9 and 10) on the accuracy of the *Capacity Utilization* estimator under different levels of network traffic loads. The traffic load of the neighbour nodes in Figure 5.4 (b) is twice as high as that in Figure 5.4 (a). It can be seen in Figure 5.3 (a) that 90% of the *%CU* estimates have an *ARE* less than 35%. In Figure 5.4 (b), as the number of neighbours increases, *ARE* becomes greater, i.e. 90% of the *%CU* estimates where *N* = 3, 6, and 10 have an *ARE* less than 55%, 63% and 72% respectively.

In order to further analyse the performance of the estimator, the relationship between the fraction of the *%CU* estimates that have an *ARE* less than 10% and the number of neighbours is shown in Figure 5.5.

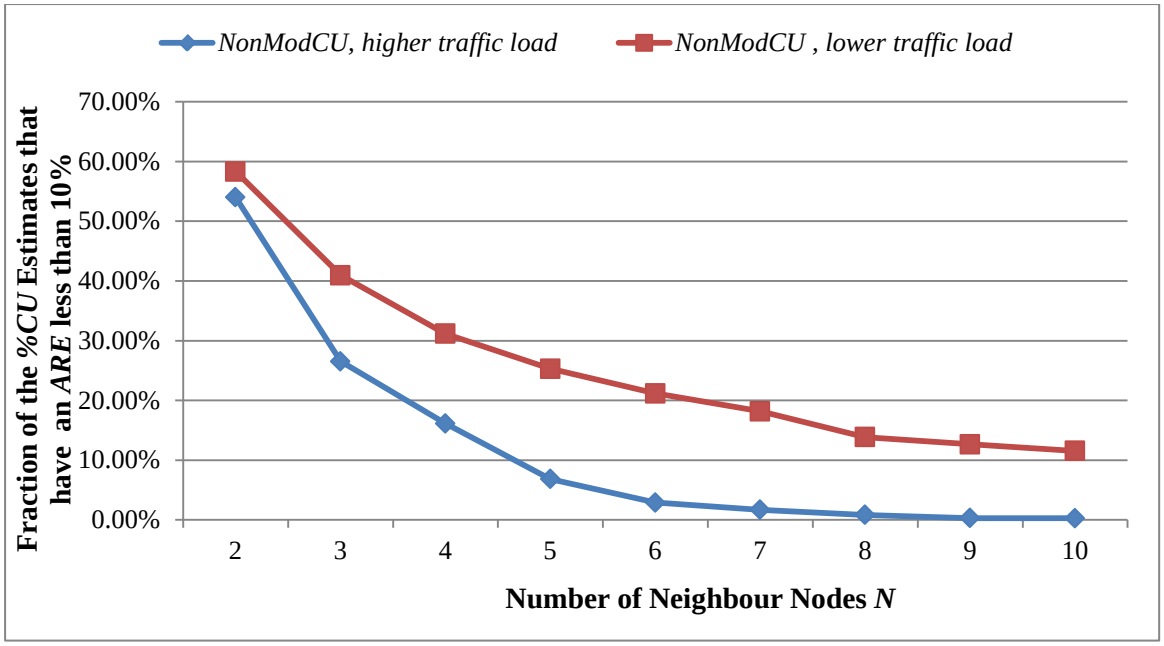

Figure 5.5: Fraction of the *%CU* Estimates that Have an *ARE* less than 10% as a Function of *N*

Figure 5.5 shows that the fraction of the *%CU* estimates before the modifications that have an *ARE* less than 10% decreases with an increase in the number of neighbours *N*. When the traffic load is lower, there are 59% of the *%CU* estimates experiencing an *ARE* less than 10% where $N = 2$, then this fraction drops to 20% of the *%CU* estimates where $N = 6$. After that, with the growth of *N* the fraction of topologies decreases gradually, i.e. 11% of the *%CU* estimates have an *ARE* less than 10% where $N = 10$. When the traffic load is higher, the fraction of *%CU* estimates shows a similar trend and drops even more drastically with an increase of *N*.

The reasons for the results presented in Figure 5.4 and 5.5 can be explained as follows: First of all, the number of unobservable neighbours of the observed node increases with the number of neighbours. This causes an increasing difference in the state of the medium between the observing and observed node which leads to an increased underestimation of the *Capacity Utilization* value. More importantly, the probability of a large number of unobservable neighbours falls as the number of neighbours increases as shown in Figure

5.5 (this will be discussed later in section 5.2.2). Thus with an increase in the number of neighbours, the accuracy of the *%CU* estimator falls dramatically at first and then reduces more gradually. Thirdly, a large number of neighbour nodes with a high packet rate causes a high contention as shown in Figure 5.4 (b) and leads to a high actual *Capacity Utilization*. The error associated with the estimates depends on both the actual *Capacity Utilization* value and the measured *Capacity Utilization* value. Thus the accuracy of the *%CU* estimator in Figure 5.4 (a) is higher than that in Figure 5.4 (b).

5.2.2 Different Number of Observable Neighbours of the Observed Node (*M*)

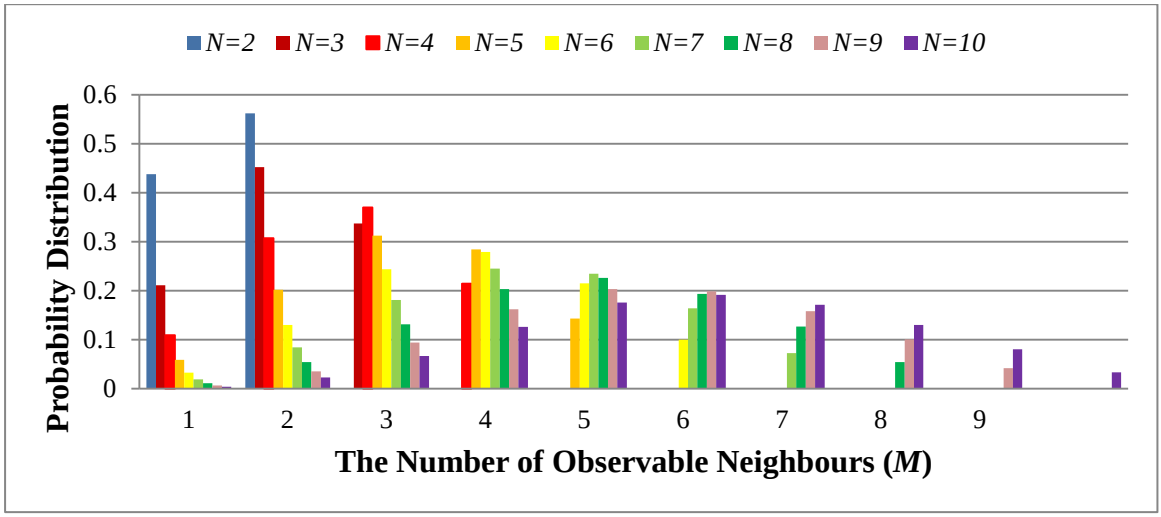

Figure 5.6: Probability Distribution of the Number of Observable Neighbours *M*

The chart in Figure 5.6 shows the probability distribution of *M* under *N* different scenarios generated by a Monte-Carlo method (i.e. it generates the node locations randomly within a predefined reception range of the observed node and calculates the ratio of *M* and *N*, and aggregates the results [152]). With the growth in the number of observable neighbours of the observed node (*M*), the PDF of *M* increases at first, climbs to maximum when $M = \frac{N+1}{2}$ (where *N* is odd) or $M = \frac{N}{2} + 1$ (where *N* is even), and then begins to decrease. For example, if the number of neighbours *N* is 5, for one of its

neighbours, the probability of observing 1, 2, 3, 4, and all of the observed node's neighbours is approximately 6%, 20%, 31%, 29% and 14% respectively.

This chart is used to explain the results in section 5.2.1, i.e. the accuracy of our estimator decreases dramatically at first and then reduces more gradually with an increase in the *N* value. The specific data values of all probabilities can be found in Appendix B.

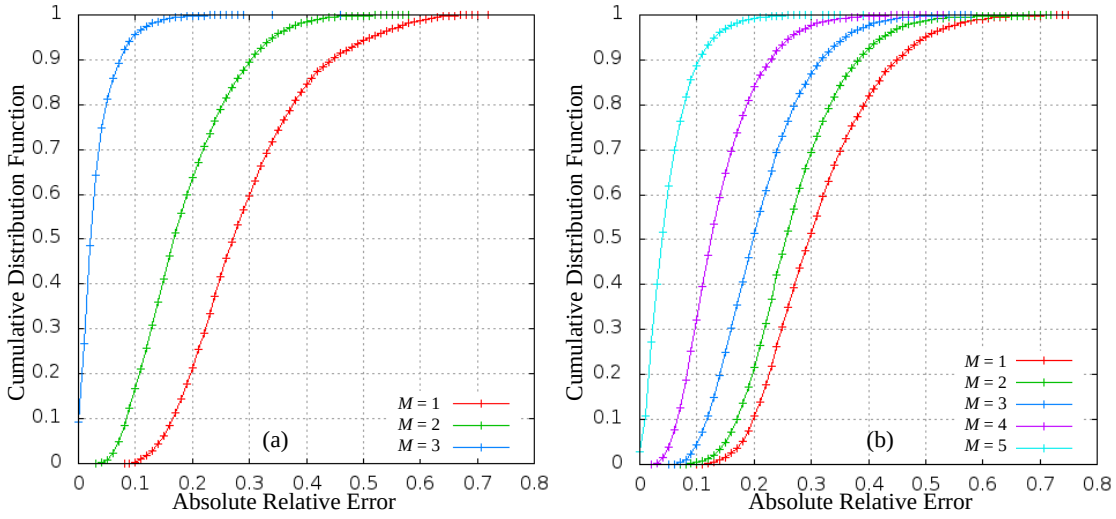

Figure 5.7: The CDF of the *ARE* of *NonModCU* where (a) $N = 3$ (b) $N = 5$

In order to test the accuracy of our *%CU* estimator (before the modifications) under the influence of different values of *M*, the CDF of the *ARE* where $N = 3$ and $N = 5$ under scenarios B-2 is plotted as Figure 5.7. As can be seen from Figure 5.7 (a), 90% of the *%CU* estimates experience an *ARE* less than 7%, 30% and 45% where $M = 1$, 2, and 3 respectively. Figure 5.7 shows that the fraction of the *%CU* estimates that have an *ARE* less than 10% have been accompanied by a corresponding increase in the number of observable neighbours, i.e. 95%, 17% and 2% of the *%CU* estimates have an *ARE* less than 10% where M=1, 2, and 3 respectively. This is because as the number of observable neighbours increases, the difference in the state of the medium becomes smaller which leads to a higher accuracy in the *%CU* estimates.

5.2.3 Different Traffic Load of the Observed Node

This section investigates the impact of the traffic load sent from the observed node on the accuracy of the *%CU* estimator. The simulation results in Figure 5.8 from scenario B-3 and B-4 consist of different packet sizes and packet rates for the traffic loads.

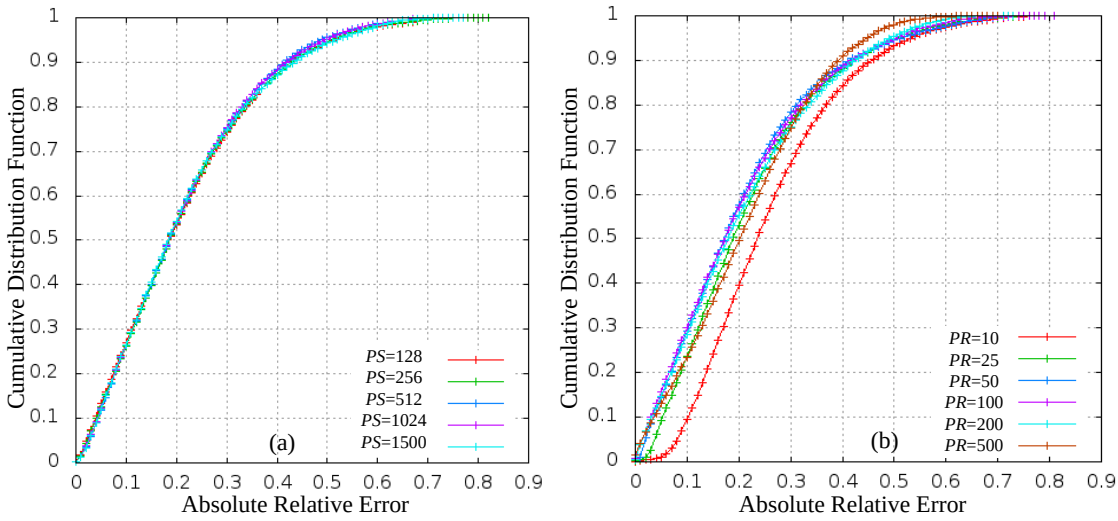

Figure 5.8: The CDF of the *ARE* of *NonModCU* for Different (a) Packet Sizes (b) Packet Rates of Traffic Load of the Observed Node

It can be seen here that 90% of the *%CU* estimates have an *ARE* less than 40%. There is no significant difference in changing the packet size of the observed node's traffic load as shown in Figure 5.8 (a) due to the fact that most transmitted frames sent from the observed node (except for failed transmissions) can be observed by its neighbour node. In Figure 5.8 (b), there is a small decrease when the packet rate is small (10 pps) due to a higher probability of a failed transmission. This will lead to an underestimation of the traffic load of the observed node that in turn gives rise to an underestimation of *Capacity Utilization*.

5.2.4 Different Traffic Load of Neighbour Nodes of the Observed Node

The traffic load of the observed node does not influence the accuracy of our estimator. However, the traffic load of neighbour nodes has a significant impact on the *%CU*

estimator due to the partially observed traffic load. The greater the difference in the state of the medium, the larger the error associated with the *%CU* estimate.

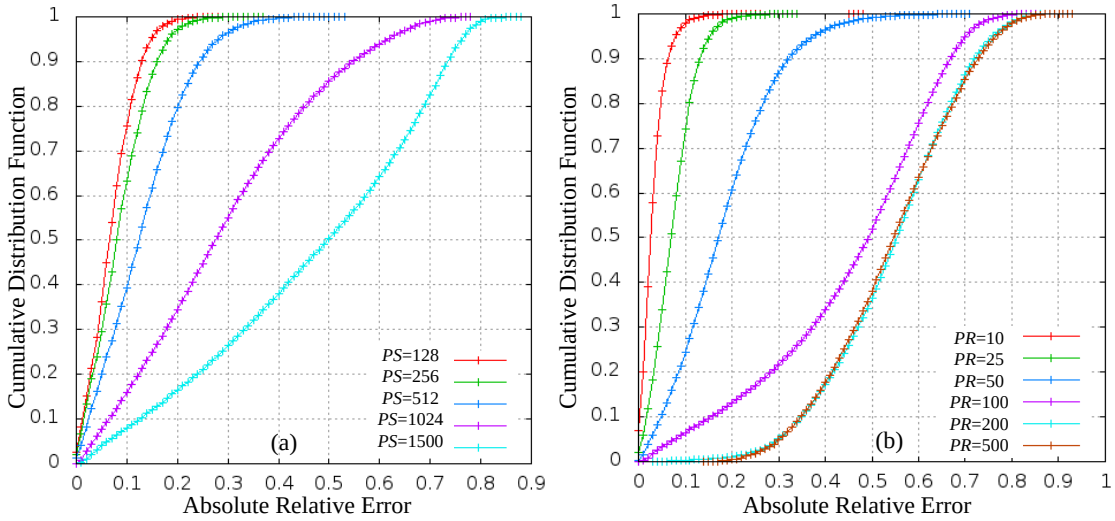

Figure 5.9: The CDF of the ARE of NonModCU for Different (a) Packet Sizes (b) Packet Rates of Neighbour Traffic Load

Scenarios B-5 and B-6 with $N = 5$ for different packet sizes and packet rates of neighbour traffic have been generated and plotted in a CDF graph as shown in Figure 5.9. Here 90% of the *%CU* estimates where the packet size of the neighbour traffic load is equal to 128, 512, 1024, and 1500 bytes have an *ARE* less than 13%, 25%, 55% and 74% respectively in Figure 5.9 (a). Also 90% of the *%CU* estimates where the packet rates of neighbour traffic load is equal to 10, 100, and 500 pps have an *ARE* less than 7%, 68%, and 73% respectively in Figure 5.9 (b).

As shown in the results, less than 20% of the *%CU* estimates have an *ARE* less than 10% under heavier traffic loads, such as a large packet size (i.e. ≥ 1024 bytes) or a higher packet rate (i.e. >100ps). Several reasons can account for the above results. Firstly, a high neighbour load gives rise to a high *Capacity Utilization* value and a greater difference in the state of the medium that leads to a large error associated with the estimation, as predicted by the error model in Chapter 4. Also, the high traffic rate increases the

collision probability among the network nodes that leads to a higher probability of failed transmissions. This will lead to an underestimation of the neighbour traffic load and the *Capacity Utilization* value. Finally, a large packet size under a high traffic rate may also cause a high collision probability among hidden nodes that overestimates the nodes *Capacity Utilization* value. Consequently, a high traffic rate in the neighbour load has the largest impact compared to the other factors on the accuracy of the *Capacity Utilization* estimator. In the worst case scenario, such as a large number of neighbours with saturated loads, the *Capacity Utilization* estimator has a lower accuracy. However, this case is an unusual situation to encounter in real IEEE 802.11 networks [86].

5.2.5 Different Traffic Types

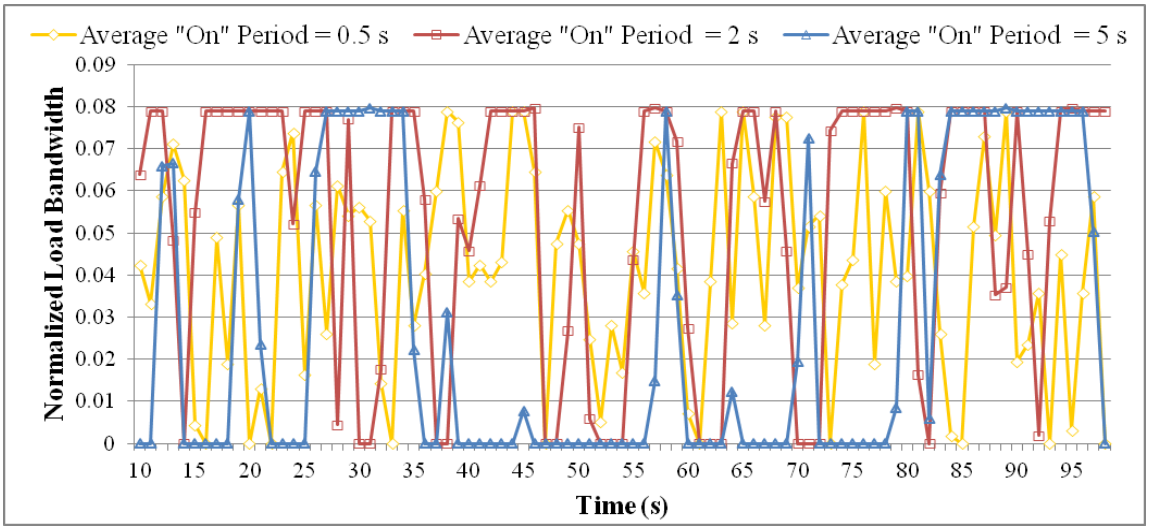

Figure 5.10: The Normalized Load Bandwidth of Exponential On-Off Traffic with Different Average “On” Periods

In scenario B-7, we set different average “On” time periods where the average “On” is equal to 0.5 second, 2 seconds and 5 seconds as shown in Figure 5.10, the traffic type of the observed node is still Poisson traffic in order to obtain a continuous traffic load being sent from the observed node for measurement. The result in Figure 5.11 shows that the estimator performance under On-Off traffic has a higher accuracy compared to that under

Poisson traffic (i.e. 70% and 28% of the *%CU* estimates have an *ARE* less than 10% respectively) due to a reduction in the average number of unobservable neighbours. This will lead to the total neighbour traffic load being smaller under On-Off traffic type than that under continuous traffic. In other words, the average value of the aggregated neighbour traffic load is reduced under On-Off traffic. Therefore, the estimator would be expected to observe an improved performance compared to Poisson traffic under scenario B-7. For example, suppose an observed node has 5 neighbours (N = 5), but there are only 3 of them sending traffic at some time interval which reduces the total traffic load of neighbours and hence the actual *Capacity Utilization* value. From the perspective of one of its neighbours, the number of unobservable neighbours of the observed node is 4 (N = 5, M = 1) and only two of them are sending traffic, thus the estimate is much closer to the actual traffic load of neighbours than that under continuous traffic.

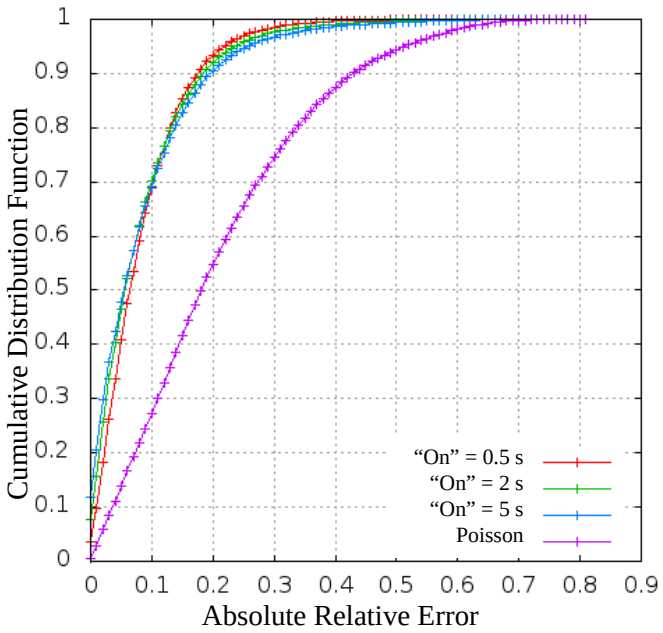

Figure 5.11: The CDF of the ARE of *NonModCU* under On-Off traffic

5.2.6 Conclusions

In this section, six factors have been investigated to determine the accuracy of the remote *Capacity Utilization* estimator before the modifications to improve the accuracy were implemented under the various test scenarios. These conclusions are as follows:

- With the growth in the number of neighbour nodes, the accuracy of the estimator decreases gradually.
- When the number of observable neighbours increases, the accuracy of the estimator will increase.
- The traffic load sent from the observed node has a minor impact on the estimator.
- The accuracy of the estimator decreases dramatically with an increase in the neighbour traffic load.
- The worst case for the estimator is where all neighbours send saturated traffic as shown in scenario B-6 under a heavy neighbour packet rate.
- Under On-Off traffic, the average number of unobservable neighbours and average traffic load sent from all neighbours is lower than that under continuous traffic. This leads to a higher accuracy in the estimator.
- Two key factors determine the performance of the remote *Capacity Utilization* estimator, namely the number of unobservable nodes and the traffic load of the unobservable nodes.
- In conclusion, the remote *Capacity Utilization* estimator has a higher accuracy under a smaller number of neighbours, a larger number of observable neighbours or under any light traffic load sent from neighbour nodes.

5.3 Performance Evaluation of the *Capacity Utilization* Estimator after the Modifications

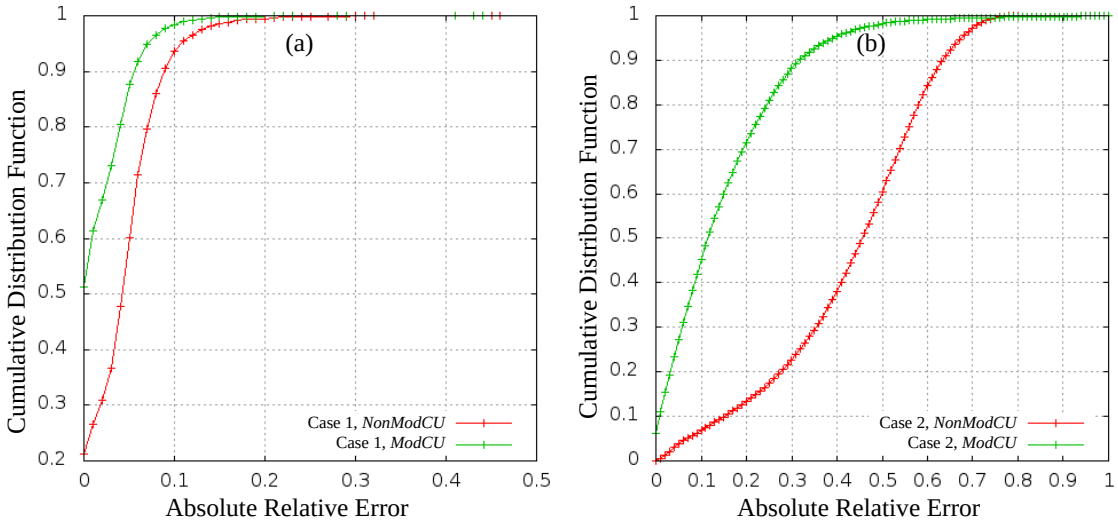

Figure 5.12: CDF of *ARE* for *ModCU* in (a) Scenario A-3 and (b) Scenario A-4

From the simulation results presented in Figure 5.12(a) and 5.12(b), the *Capacity Utilization* estimator in "red" shows a poor performance under certain conditions such as a large number of neighbours, a small number of observable neighbours or a high neighbour traffic load. Thus we have proposed some modifications to improve the accuracy of the estimator. The impact on these modifications on the accuracy of the estimator will be investigated in this section. It should be noted here that our estimator always produces an underestimation of the *Capacity Utilization* before the modifications and therefore the *ARE* will be less than 100%. However, when the modifications are applied, the *Capacity Utilization* produced by the estimator can be either an underestimation or an overestimation. Consequently, it is possible to have *ARE* values greater than 100%, however these values were found to occur infrequently in the majority of the scenarios investigated, see the PDFs results in Appendix D. Therefore, for convenience we will only consider values of ARE between 0 and 100% in the graphs presented in this section.

The comparison of the *ARE* of the estimator before *(NonModCU)* and after the modifications *(ModCU)* for two general cases in scenarios A-3 and A-4 is shown in Figure 5.12. The estimator after the modifications exhibits a significant improvement in the accuracy, i.e. 98% of the *%CU* estimates have an *ARE* less than 10% in Figure 5.12(a), 90% of the *%CU* estimates after the modifications have an *ARE* less than 30% compared to 90% of the *%CU* estimates before the modifications have an *ARE* less than 65% as shown in Figure 5.12(b).

5.3.1 The Impact of Factors on the Accuracy of the Estimator after the Modifications

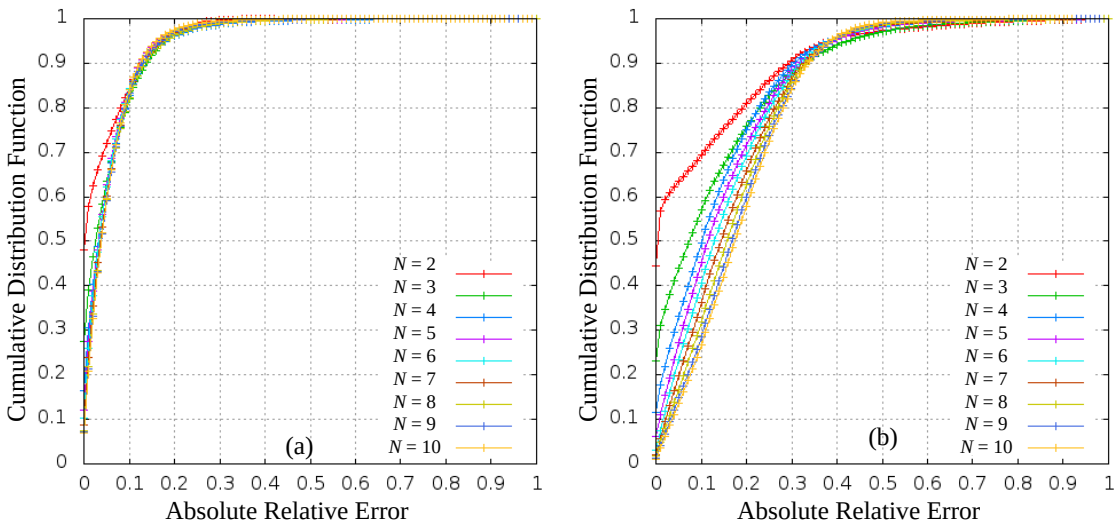

Figure 5.13: The CDF of the *ARE* of *ModCU* under Different *N* Scenarios with (a) Lower Traffic Load (b) Higher Traffic Load

There is a clear improvement in the accuracy of the *%CU* estimator after the modifications as shown in Figure 5.13 (scenario C-1-1 and C-1-2 respectively) where 90% of the *%CU* estimates for all *N* cases have an *ARE* less than 13% and 33% under low and high traffic loads respectively. In Figure 5.14, the dashed line and the solid line is the estimate before and after the modifications respectively. Under lower traffic loads, approximately 84% of the *%CU* estimates after the modifications have an *ARE* less than

10%. The accuracy of the estimator is independent of the number of neighbours due to the unobserved neighbour load and the contention can be corrected by the modifications. However, under a heavy traffic load, the accuracy of the estimator after the modifications is reduced with an increase in the number of neighbour nodes. This is because most collisions involve three or more nodes when the number of neighbours is large but our *%CU* estimator assumes the *most collisions involve just 2 nodes* (*Assumption 3*). The further improvement of *Assumption 3* under some topologies with many neighbours under heavy traffic will be investigated in the future work.

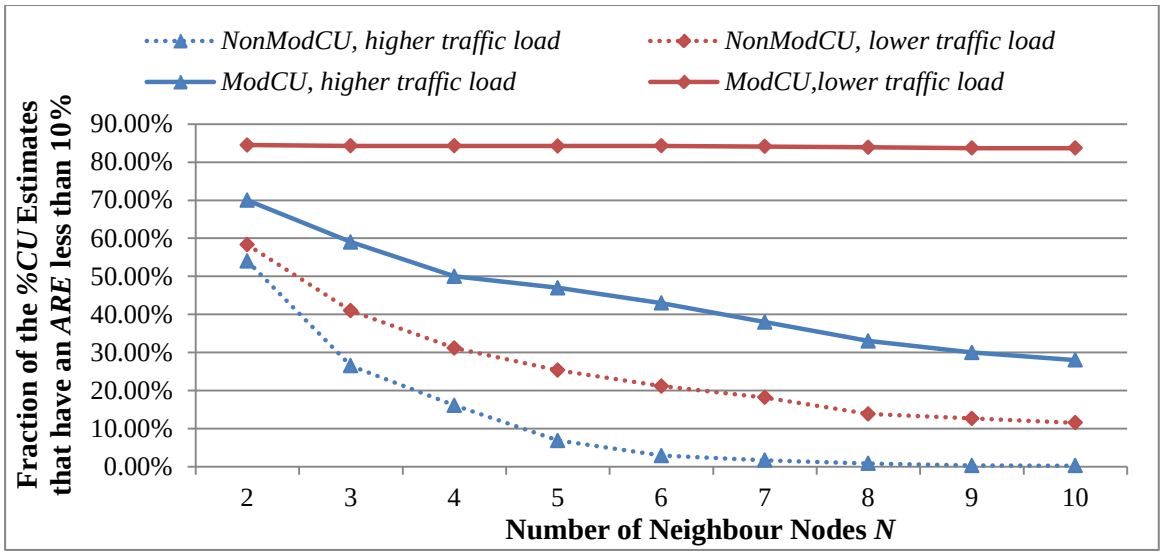

Figure 5.14: Fraction of the %CU Estimates after the Modifications that Have an ARE less than 10% as a Function of *N*

After the modifications, the *Capacity Utilization* estimator exhibits an improvement in the accuracy under different numbers of observable neighbours (*M*) as shown in Figure 5.15, i.e. 90% of the *%CU* estimates have an *ARE* less than 25% and 33% where $N = 3$ and 5 respectively.

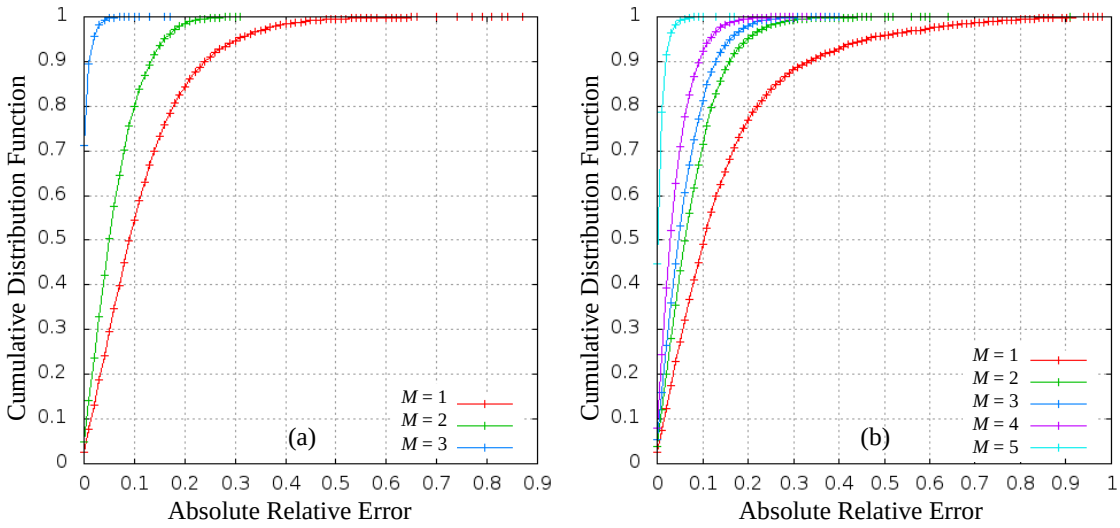

Figure 5.15: The CDF of the *ARE* of *ModCU* where (a) $N = 3$ (b) $N = 5$

When the number of observable neighbours is larger, i.e. $M \geq 3$, more than 75% of the *%CU* estimates have an *ARE* less than 10% where $N = 5$ and $N = 7$ as shown in Figure 5.16.

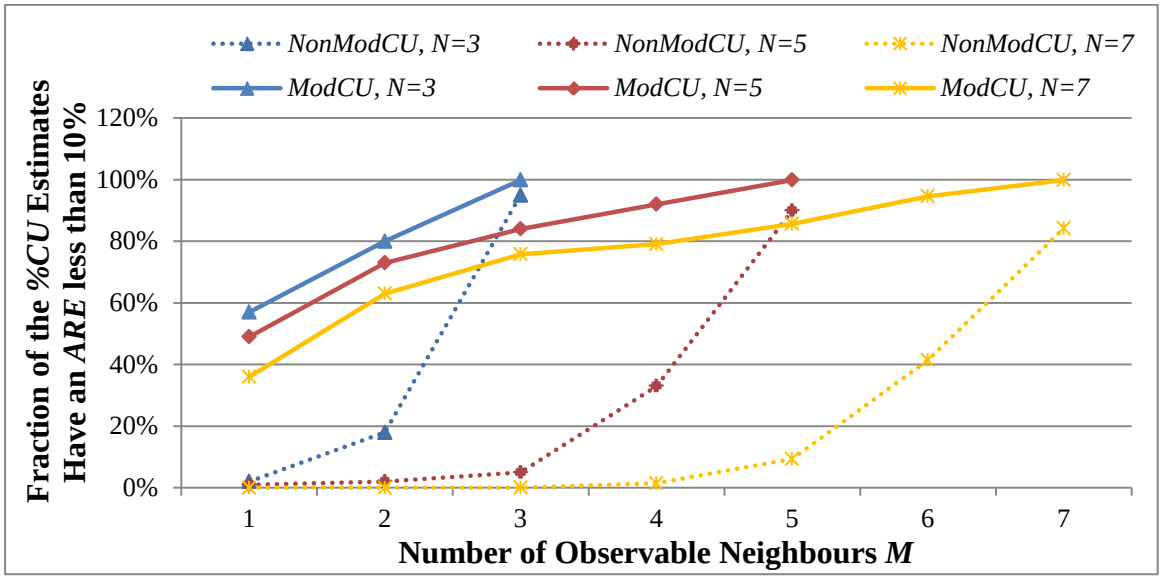

Figure 5.16: Fraction of the %CU Estimates after the Modifications that Have an ARE less than 10% as a function of *M*

Our estimator has been modified by using *Assumption 1* where *the mean load of the unobservable neighbour nodes is equal to the mean load of the observable nodes* to

estimate the traffic load of the unobserved neighbour nodes. Thus the estimator produces a higher accuracy when the number of observable neighbours is large since more information on the traffic load can be obtained.

As shown in Figure 5.17, it can be seen that different packet sizes and packet rates sent from the observed node do not have any significant impact on the accuracy of the *%CU* estimator after the modifications. This is because the failed transmitted frames can be also measured by the remote neighbour nodes by utilizing the retransmission flag. However, our estimator minimizes the error associated with the *Capacity Utilization* estimate, i.e. only 30% of the *%CU* estimates before the modifications have an *ARE* less than 10% compared to approximately 90% of the *%CU* estimates after the modifications have an *ARE* less than 20% and 70% of the *%CU* estimates have an *ARE* less than 10%.

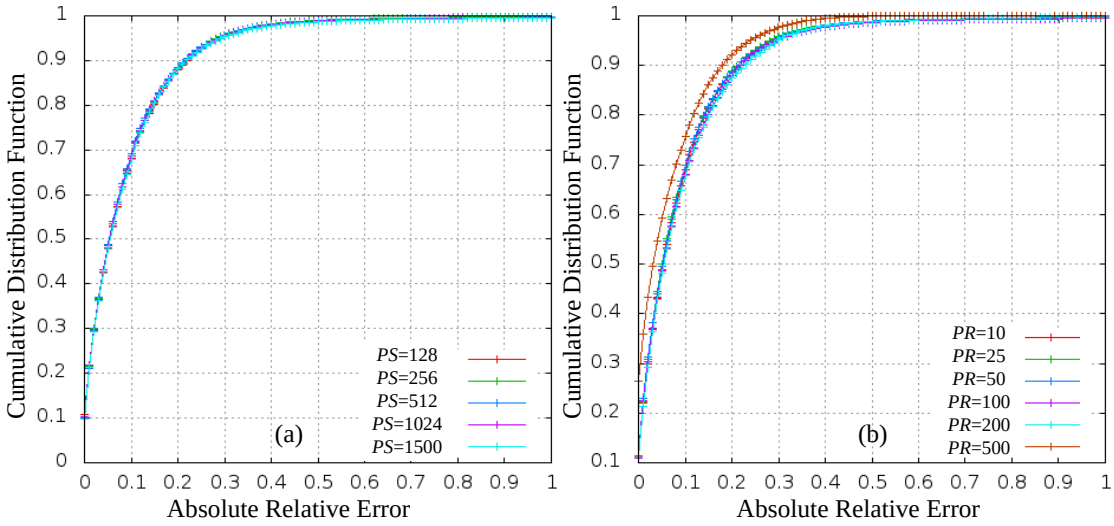

Figure 5.17: The CDF of the *ARE* of *ModCU* for Different (a) Packet Sizes (b) Packet Rates of Traffic Load of the Observed Node

When the traffic load of the neighbour nodes increases, our estimator after the modifications also produces a decrease in the accuracy as shown in Figure 5.18 and Figure 5.19. More than 80% of the *%CU* estimates have an *ARE* less than 10% under lower neighbour traffic loads (packet size ≤ 512 bytes or packet rate ≤ 50 pps) where N =

5. Under heavier traffic loads (packet size =1500 bytes or packet rate > 100 pps), only 40% of the *%CU* estimates have an *ARE* less than 10%.

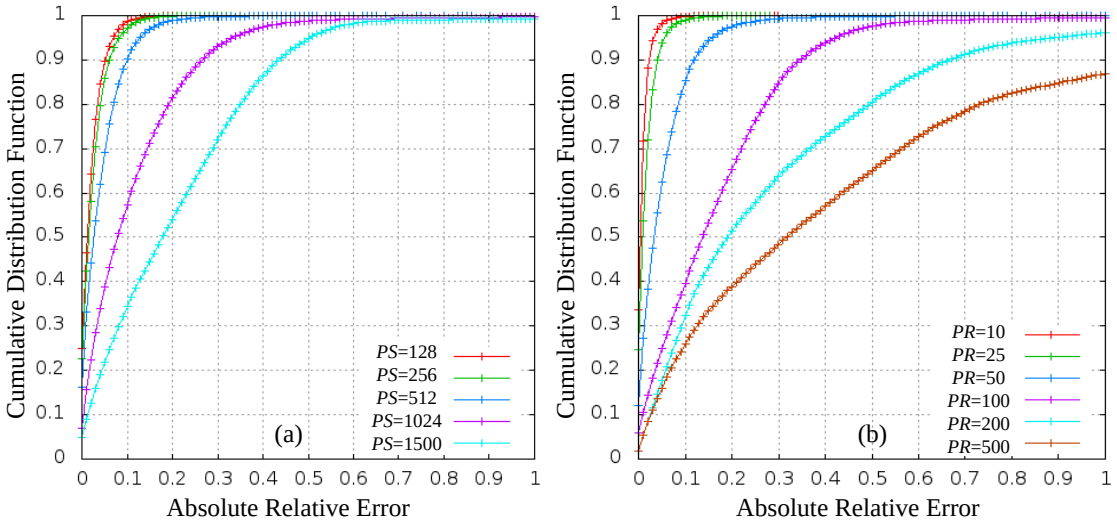

Figure 5.18: The CDF of the *ARE* of *ModCU* for Different (a) Packet Sizes (b) Packet Rates of Neighbour Traffic Load

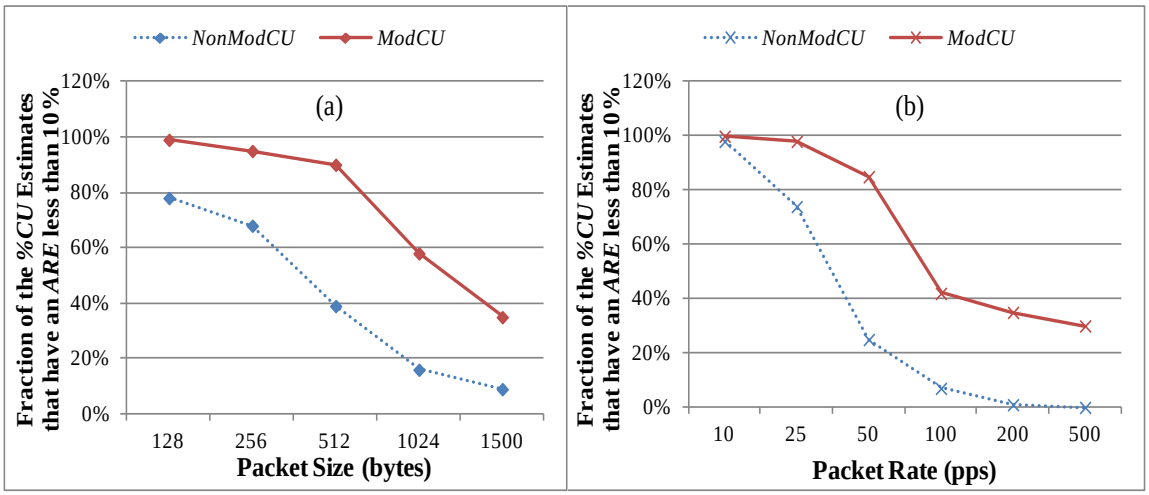

Figure 5.19: Fraction of the *%CU* Estimates after the Modifications that Have an *ARE* less than 10% as a Function of (a) Packet Size and (b) Packet Rate

The main reasons for the lower accuracy of the estimator under heavier traffic loads have been discussed in section 5.2.4. In addition, the high traffic rate increases the collision probability among the network nodes, more collisions involving 3 or more nodes and

halving the total failed bandwidth of all neighbour nodes may cause an overestimation of neighbour load that leads to an overestimate of nodes *Capacity Utilization*. However the accuracy of the estimator is still improved by the modifications, i.e. 0% of the *%CU* estimates before the modifications have an *ARE* less than 10% when all the neighbours are saturated (i.e. the packet rate of every neighbour node is greater than 500 pps).

To further explain these results, two scenarios under an increased interference range (in order to avoid the “hidden nodes” problem under different neighbour traffic loads) are presented here. We can see that the *ARE* of the neighbour load estimates are similar to Figure 5.20 (a) under both low and high traffic loads of neighbour nodes, the *ARE* of the *Capacity Utilization* estimator still decreases under heavier traffic loads as shown in Figure 5.20 (b). In order to improve the accuracy of the estimates under heavy neighbour loads, the total neighbour load estimates should be more accurate than those under a low neighbour load. However, the network traffic load of unobservable neighbours is unpredictable which makes the estimation less accurate.

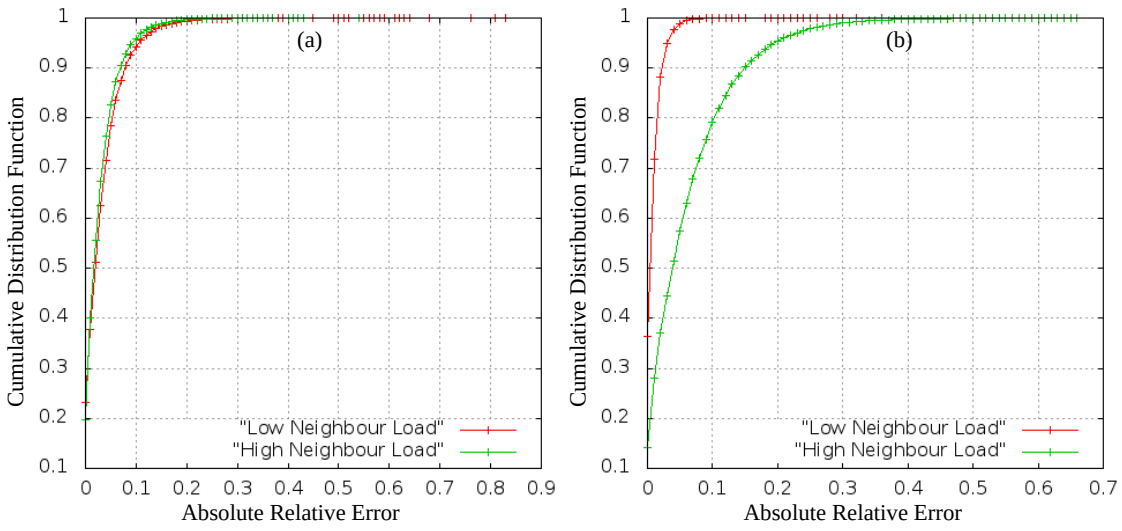

Figure 5.20: The CDF of the *ARE* of (a) Improved Estimated Neighbour Load (b) *ModCU*

Under Exponential On-Off traffic, the *Capacity Utilization* estimator after the modifications still produces a higher accuracy than that under Poisson traffic, i.e. 90% of

the *%CU* estimates have an *ARE* less than 10%, see Figure 5.21. This is because the average number of unobservable neighbours and the average traffic load is reduced which produces a lower contention and hence collision probability. This will lead to a higher accuracy in the estimator.

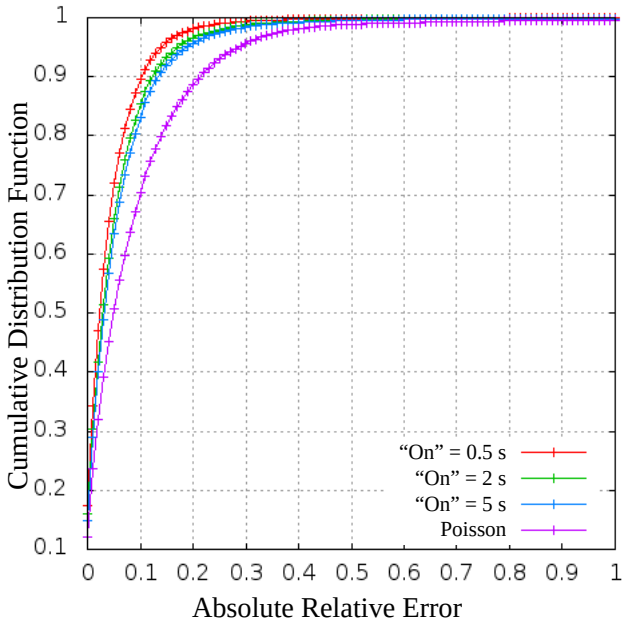

Figure 5.21: CDF of the *ARE* of *ModCU* under Exponential On-Off traffic

5.3.2 Conclusions

In this section, the accuracy of the *Capacity Utilization* estimator after the modifications have been implemented has been investigated under different scenarios. The following conclusions can be drawn from these results under the test scenarios:

- The modifications significantly improve the accuracy of the estimator under all scenarios.
- When the traffic load of neighbour nodes is low, the accuracy of the *Capacity Utilization* estimator will be independent of the number of neighbours. However, the accuracy of the *Capacity Utilization* estimator decreases with a reduction in the number of observable neighbours.

- The traffic load of the observed node has no impact on the accuracy of the accuracy of the *Capacity Utilization* estimator. The accuracy of the estimator after the modifications decreases with an increase in the traffic load of neighbour nodes.
- Under On-Off traffic, the *Capacity Utilization* estimator after the modifications exhibits a higher accuracy than that under continuous traffic due to the lower average number of unobservable nodes and average traffic load.
- The benefits arising from the improvement modifications proposed in chapter 4 have been validated in that they have been shown to significantly improve the accuracy of the remote contention estimation, the neighbour load estimation and the *Capacity Utilization* estimator.

5.4 Saturation Detection

As described in Chapter 2, once node saturation can be detected, the remote clients can select and associate with another non-saturated AP in order to maintain the network connectivity and alleviate the saturation situation. Also, in multi-channel networks, once a node finds that its neighbour node is saturated in the current channel, it can re-select another channel to avoid congestion and possible packet losses. The node can also re-route to find another path when there is a saturated neighbour node in its next hop in order to guarantee its QoS requirement. Therefore, this section investigates our *Capacity Utilization* estimator after the modifications in an application to detect node saturation in WLANs. Two other algorithms, queue observation (local measurement) and regularly pinging (active approach) are used to compare and evaluate the accuracy in node saturation detection by using two criteria, namely the *FDR* and *FAR*.

5.4.1 A Comparison of the Three Methods

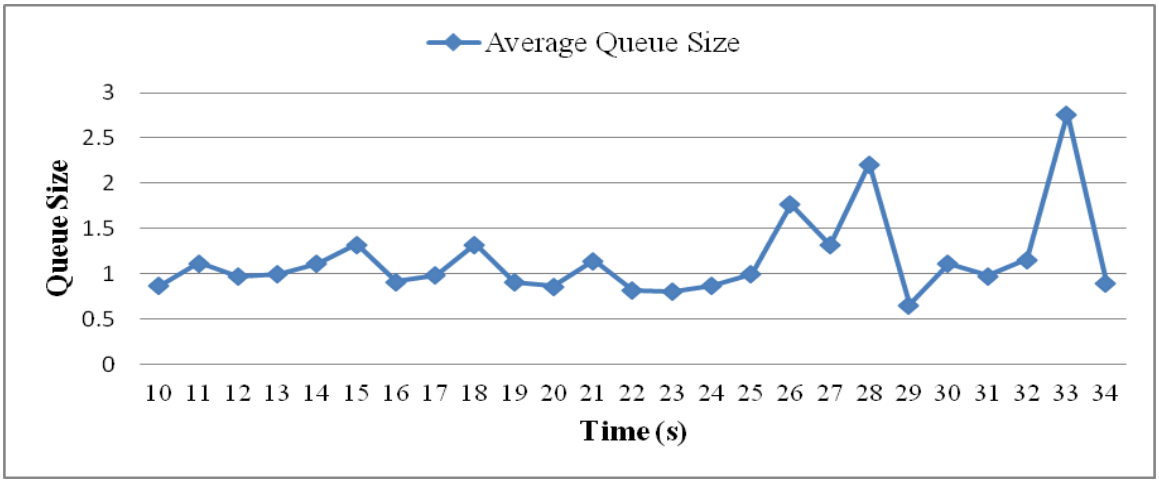

Figure 5.22: Average Queue Size in the Example Topology

We randomly selected one of 1,000 topologies from scenario D-1 where N = 5 as an example topology to show the use of the three detection algorithms. The specific packet size, packet rate, and topology can be found in Appendix E. Figure 5.22 shows the average queue size from t = 10 seconds to t = 34 seconds in a simulation. At t = 13, 23, and 33 seconds, the average queue size is approximately 1, 0.8 and 2.7 respectively. Under the saturation case, once the average queue size is lower than 1, we assume that it is a false alarm. Otherwise, if the average queue size is higher than or equal to 1 under non-saturated case, we assume that it is a failed detection.

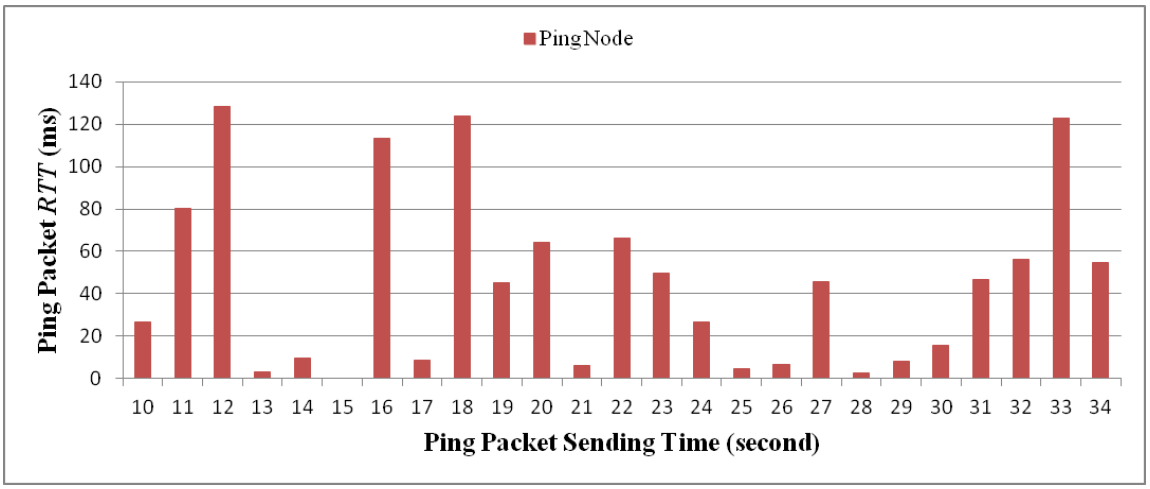

Figure 5.23: *RTT*s of Ping Packets in the Example Topology

If the traffic load of the observer nodes is heavy, the ping packets may be dropped before being transmitted due to buffer overflow. Therefore, we set an extra neighbour node, called the ping node with a random position to send ping packets in the simulation. The ping node can send one or more ping packets every measurement interval to measure the round-trip time (*RTT*) or average *RTT* respectively. However, sending extra ping packets will consume the channel bandwidth which can have a negative impact on performance of a network due to the increased contention on the medium. In order to facilitate a valid comparison between three methods, they all use a standard time interval of 1 second. We send 1 ping packet per second where a packet size of 50 bytes has been used. The *RTTs* of the ping packets from the ping node in the example topology are shown in Figure 5.23. It can be seen that there is no ping echo packet returned at t = 15 seconds, and hence we assume that the observed node is saturated at this time.

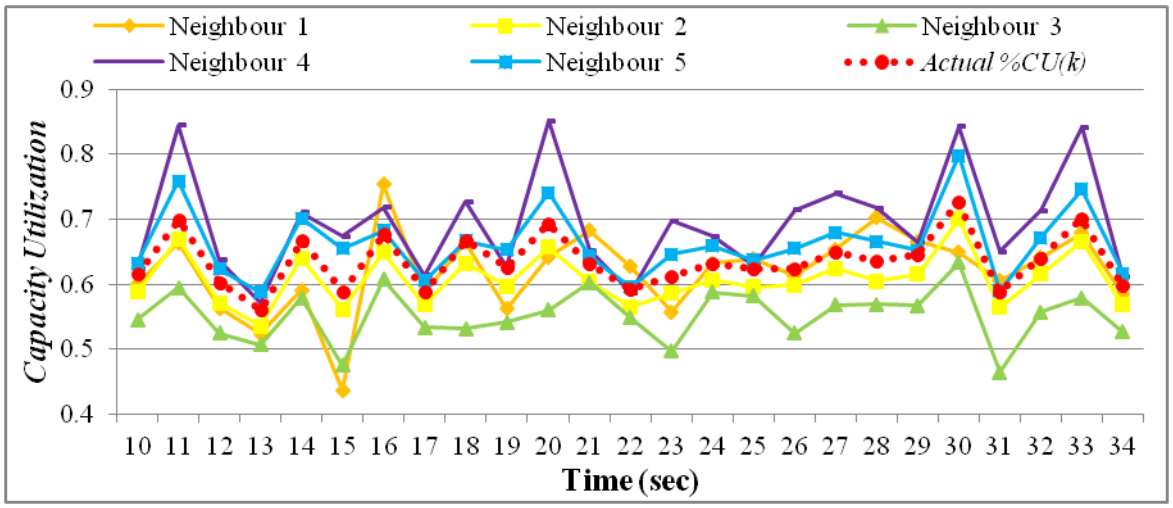

Figure 5.24: The Remote *Capacity Utilization* Estimation in the Example Topology

In Figure 5.24, the red dashed line represents the actual *Capacity Utilization* value of the observed node, and the solid lines are the *Capacity Utilization* estimation performed by the different neighbour nodes of the observed node. At t = 20 seconds, it can be seen that *Neighbour* 4 performs an overestimation of *%CU*, and *Neighbour* 3 shows an underestimation of *%CU*. The *%CU* values estimated by *Neighbour* 1, 2, and 5 are close

to the actual *Capacity Utilization* value of node *k*. If the *%CU* estimation is equal to 100%, we assume that the observed node is saturated.

This example topology only illustrates how these three detection algorithms operate and does not represent any comparison of results. In the next section, 1,000 topologies with random parameters where N = 2, 3, 4, 5, 6, 7, 8, 9, and 10 described in scenario D-1 and D-2 are generated to analyse the accuracy of detecting node saturation by using these three algorithms.

5.4.2 The *Capacity Utilization* Estimator in Node Saturation Detection

Two scenarios D-1 and D-2 are presented to investigate the *FDR* and *FAR* respectively. The original PDFs results can be found in Appendix E. The simulation results in Figure 5.25 and 5.26 show that the relationship between *FDR*, *FAR* and the *Capacity Utilization* threshold parameter CU^{TH}.

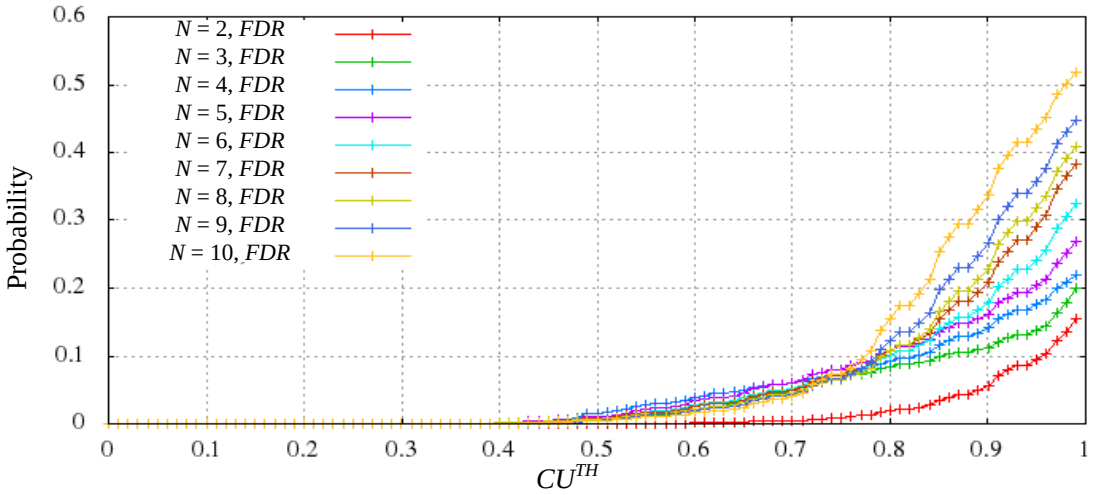

Figure 5.25: Relationship between *FDR* and CU^{TH}

It can be seen that there is an increase in both *FDR* and *FAR* with an increase of the number of neighbours *N*. This is because that our *Capacity Utilization* estimator performs with a higher accuracy under the smaller number of neighbours and light network traffic load. If the CU^{TH} is set as 1 which represents the situation when the estimator measures

its neighbour's *Capacity Utilization* value to be 100%, it decides that this neighbour is saturated. The *FDR* decreases with the growth in the number of neighbour nodes as shown in Figure 5.25. The *FDR* is 15% where $N = 2$ and is approximately 50% where $N = 10$. However, by using $CU^{TH} = 1$ as shown in Figure 5.26, the *FAR* is less than 2% for all scenarios.

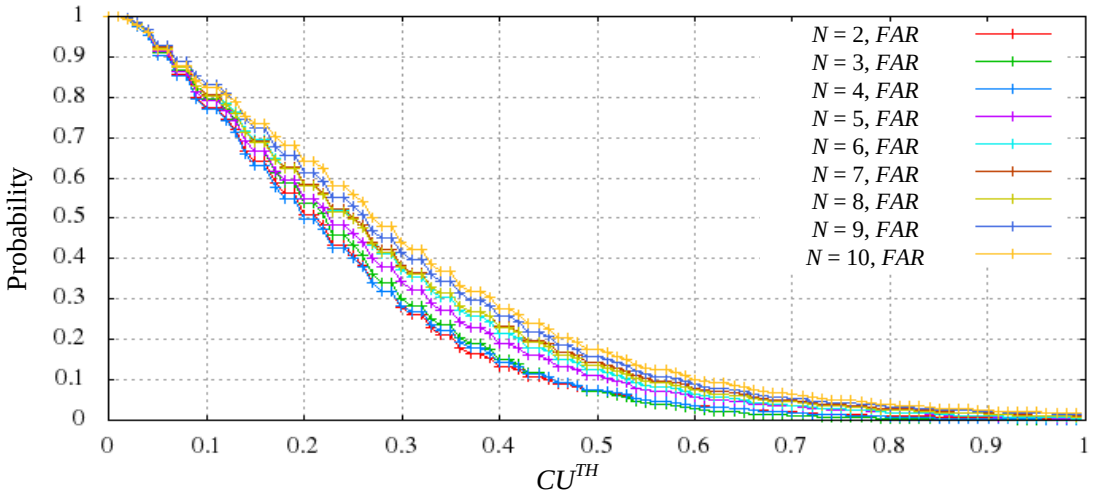

Figure 5.26: Relationship between *FAR* and CU^{TH}

Using a CU^{TH} less than 100% will decrease the *FDR* but increase the *FAR*. In order to find an optimal threshold CU^{TH} that achieves a trade-off between *FDR* and *FAR*, a minimum error-rate classifier based upon Bayesian decision theory is employed here as shown in Figure 5.27.

The optimal CU^{TH} is selected based upon the saturation probability, i.e. the value of the optimal CU^{TH} is reduced by an increase in the saturation probability. For example, if the network saturation probability is 10%, the optimal CU^{TH} can be set as 0.8. If the saturation probability is 50%, the *FDR* is less than 5% and *FAR* is 20% where an optimal $CU^{TH} = 0.6$ as shown in Figure 5.27. The saturation probability depends upon the number of neighbour nodes and their traffic load. Thus the CU^{TH} can be adaptively adjusted. For

example, when the number of neighbours is larger or during the peak period of network traffic, the CU^{TH} can be decreased to a smaller value automatically.

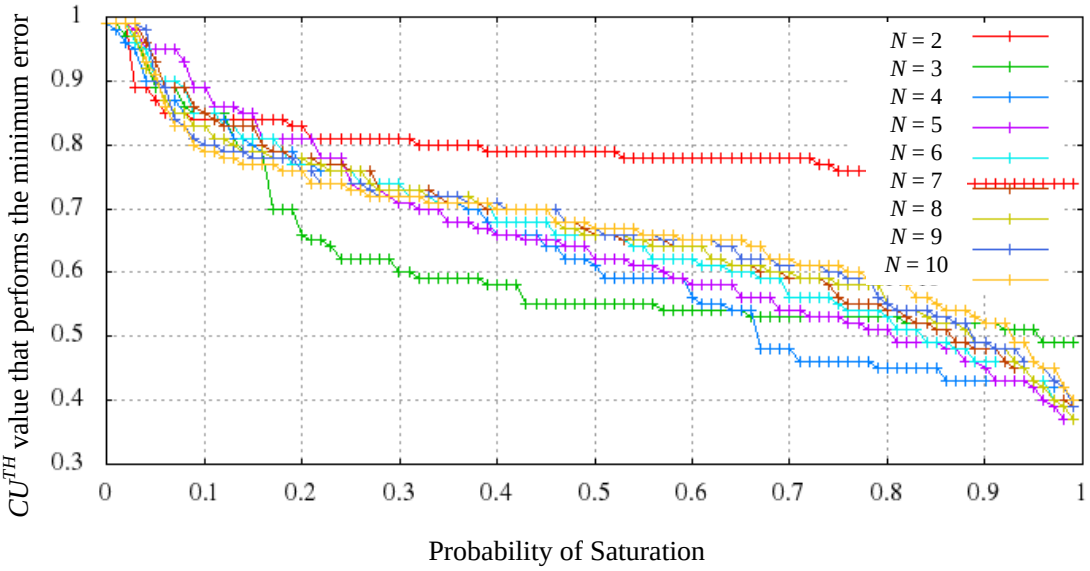

Figure 5.27: Relationship between Probability of Saturation and Optimal CU^{TH}

5.4.3 Comparison of Three Node Saturation Detection Algorithms

The *FDR* and *FAR* results generated by three algorithms and using a CU^{TH} = 0.8 are shown in Table 5.2 and Figure 5.28.

It can be seen that the queue observation method has the lowest *FDR* due to the nature of node saturation. Once a node is saturated, there is at least one packet that cannot win a transmission opportunity. However, the *FAR* of this method increases with an increase of the number of neighbours. This is because the waiting time of the packets in the queue not only depends on the maximum buffer size but is also related to the contention. The larger the contention, the longer the deferral time that will be required.

The regularly pinging method produces both higher *FDR* and *FAR* due to the *RTT* dependence on the contention on the medium. Firstly, the node could return ping echo packets because it may still win the transmission opportunities when it is saturated. This leads to a higher *FDR*. Moreover, the collision probability of ping packets increases with

an increase in contention. Therefore, the observed node may not receive the ping packets or fail to return the packets due to unavoidable collisions when the node is not saturated. Both *FDR* and *FAR* of the regularly pinging method indicate that this method does not scale well.

Table 5.2 *FDR* and *FAR* of the Three Algorithms

	Queue Monitoring		Regularly Ping		*%CU* estimator (Using CU^{TH} = 1)		*%CU* estimator (Using CU^{TH} = 0.8)	
	FDR	*FAR*	*FDR*	*FAR*	*FDR*	*FAR*	*FDR*	*FAR*
***N* = 2**	0%	4.1%	44.9%	2.1%	15.6%	0.34%	1.9%	1.1%
***N* = 3**	0%	4.9%	38.4%	3.6%	20.1%	0.11%	8.5%	0.5%
***N* = 4**	0%	5.6%	35.5%	5.1%	21.9%	0.11%	9.2%	0.8%
***N* = 5**	0%	6.7%	32.1%	5.9%	26.8%	0.38%	10.8%	1.9%
***N* = 6**	0%	7.4%	29.3%	6.9%	32.5%	0.52%	10.2%	2%
***N* = 7**	0%	8.2%	30.1%	8.7%	38.4%	0.9%	10.8%	2.8%
***N* = 8**	0%	9.2%	26.9%	10.3%	40.8%	0.86%	10.7%	2.5%
***N* = 9**	0%	10.9%	25.3%	12.3%	44.8%	1.4%	12.2%	3.3%
***N* = 10**	0%	11.4%	23.9%	14.3%	51.9%	1.7%	15.4%	3.8%

Our *%CU* estimator produces the lowest *FAR* among the three detection algorithms due to its high accuracy under the light traffic which has been discussed in previous section. However, our estimator has a poor *FDR* in scenario D-2 because the heavy traffic loads have the biggest impact on the accuracy of the estimation. After using a CU^{TH} = 0.8, our *FDR* is reduced from 15.6% to 1.9%, from 26.8% to 10.8% and from 51.9% to 15.4%

where N =2, 5, and 10 respectively, as shown by the grey solid line and yellow solid line in Figure 5.28.

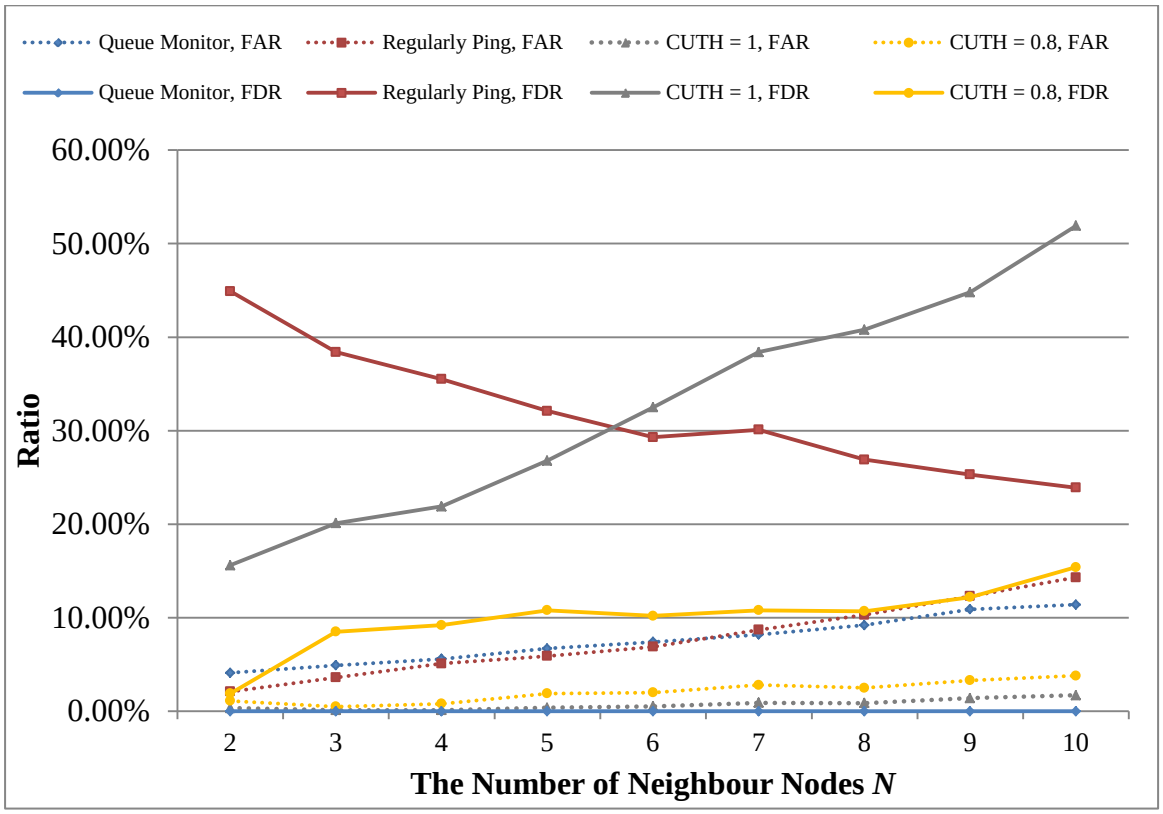

Figure 5.28: The Comparison of *FDR* and *FAR* among Three Algorithms

The above results show the accuracy for our *Capacity Utilization* estimator by using an optimal CU^{TH} = 0.8 in detecting saturated nodes that produces a *FDR* less than 15% and a *FAR* less than 5%. Compared to other two algorithms, we have a lower *FDR* than that of the regularly pinging method, and the lowest *FAR* among three algorithms. Moreover, our *Capacity Utilization* estimator is more reliable and has no overhead in detecting node saturation. In summary, the simulation results under the tested scenarios indicate the feasibility and accuracy of our *Capacity Utilization* estimator in detecting node saturation. Once the node saturation can be detected accurately, other applications such as AP selection, channel selection and resource aware routing can benefit by providing a better performance for WLANs users.

5.5 Summary

The simulation results and performance evaluation of our *Capacity Utilization* estimator have been presented in this chapter. Four different scenarios of groups are investigated by using the ns2 simulator to investigate the accuracy of our *Capacity Utilization* estimator before and after applying the modifications, to compare the improvement in the accuracy, and to validate the effectiveness of the modifications. Finally, we investigate node saturation detection in an application to evaluate the performance of our estimator.

The effect of varying various network parameters such as the number of neighbours, the number of observable neighbours of the observed node, the traffic load of the observed node, traffic load, and traffic type on the accuracy of our *Capacity Utilization* estimator before and after the modifications has been investigated in this chapter.

The error associated with the *Capacity Utilization* estimation through remote observation method is unavoidable but can be minimized. Our modified estimator performs with high accuracy and significantly minimizes the error of the original estimation on average. We thus conclude that the modifications implemented to improve the estimate of neighbour's *Capacity Utilization* value are feasible. A simple node saturation detection scheme using the *Capacity Utilization* estimator is investigated where its performance is compared with two other schemes and the results show that the feasibility, usage and accuracy of our estimator.

Chapter 6 Conclusions and Future Work

In wired networks, the *capacity* of a link is generally defined as the maximum transmission rate that can be achieved on the link. Two wired nodes use some form of a cable to connect them, thus the capacity of a link can be considered as a constant value that depends only the communications protocols used. However, due to the shared nature of the wireless medium, the IEEE 802.11 MAC layer mechanism, dynamic rate adaption, fading and interference the estimation of wireless capacity is far more challenging in WLANs. Therefore, the definitions and methodology of capacity estimation used in wired networks are not appropriate for WLANs.

The work presented in this thesis is a significant contribution to developing an estimator based upon remote observations performed by neighbour nodes, designed to provide an accurate estimation of node *Capacity Utilization* in WLANs, which reflects the usage of the WLAN node *capacity*. By passively analyzing the transmitted packets, the node *Capacity Utilization* estimator uses remote observations performed by neighbour nodes instead of local measurements at the node itself. Our *Capacity Utilization* estimator described in Chapter 4 extends the MAC bandwidth components model [13] and presents the specific methodology used to measure traffic load, contention, node capacity and the *Capacity Utilization* value. However, hidden nodes present a challenge for neighbour nodes to estimate the available capacity of observed nodes accurately. The error associated with the estimation is unavoidable due to the differences in the wireless medium as observed by the different locations of the nodes. Three simple and reasonable assumptions were introduced to minimize the error associated with this *Capacity Utilization* estimation. We assume that (1) *the mean load of the unobservable neighbour nodes is equal to the mean load of the observable nodes*; (2) *all the frames from the*

unobservable neighbours of the observed node are transmitted with a uniform time interval within measurement interval of interest; (3) *most collisions involve just two nodes* to estimate the load bandwidth due to failed transmissions. These modifications have been shown to significantly reduce the error associated with the estimation of *Capacity Utilization* for all the test scenarios considered in Chapter 5. Assumptions 1 and 3 are used to correct the error associated with the neighbour load estimation, and *Assumption 2* is employed to correct the contention estimation.

The main source of error in the *Capacity Utilization* estimator is the traffic load experienced by the node being observed which may not be the same as that experienced by the node performing the estimation, i.e. the monitor mode. An error model is introduced in Chapter 4 to analyse the performance of our estimator and make some performance predictions. This source of error is considered to be random. A performance evaluation of our *Capacity Utilization* estimator is essentially a study of the accuracy of our estimator. Therefore, an absolute relative error (*ARE*) metric is used to compare the actual *Capacity Utilization* experienced by a node with our remote *Capacity Utilization* estimator and to investigate the impact of unobserved traffic load on the accuracy. The investigation of factors influencing the accuracy of our estimator such as the number of neighbours, the number of observable neighbours of observed node, the traffic load of observed node and its neighbour nodes, and different traffic types are introduced and examined in Chapter 5. From the simulation results, we can determine the impacted factors and conclude that our modified estimator is capable of producing accurate measurements.

Node saturation is a situation that can arise where the node cannot win a sufficient number of transmission opportunities to satisfy its traffic load under heavy network load conditions. If the *%CU* is equal to 100%, it means that the current load is consuming all

of the available capacity and the node cannot transmit any more frames. Detecting node saturation is one of the applications of this estimator used to demonstrate the usage of our *Capacity Utilization* estimator. A minimum error-rate classifier based upon Bayesian decision theory is introduced to perform a trade-off between *FDR* and *FAR* in finding the optimal value of *Capacity Utilization* threshold. The results are compared with other two saturation detection algorithms (i.e. a queue observation method and a regularly pinging method) in order to demonstrate the feasibility and accuracy of our *Capacity Utilization* estimator in detecting node saturation.

In this chapter, a brief review of the work will be conducted in the following and some future research topics are proposed.

6.1 Conclusions

From what has been discussed above, the following conclusions can be summarized as:

- Current capacity estimation tools exhibit many shortcomings in the aspects of accuracy, reliability and overhead in the estimation of capacity and available bandwidth. The accuracy of active probing methods is significantly impacted by the traffic load in wireless networks which reduces the accuracy of the estimation. Transmitting a number of probing packets also has an impact on the existing traffic and network performance, i.e. it can result in a high overhead. Compared to active approaches, the passive approaches which monitor the channel locally have a lower overhead. However, many factors such as collision probabilities, the number of deferral intervals and backoff time, retransmissions, interference range and hidden nodes can affect the accuracy of the estimation. The passive approaches can be divided into two methods: local measurement and remote observation. However, the local measurement requires an additional dissemination mechanism to broadcast the

estimation of capacity information to its neighbour nodes for taking further actions. This may increase the overhead, make the applications more complex and the communications may not be completely reliable due to the packet losses or delay.

- Our *Capacity Utilization* estimator measures the node capacity and *Capacity Utilization* of a WLAN node. The node capacity is defined as *the bandwidth available under the current load conditions and represents the maximum load that can be achieved by the node provided that the other network nodes maintain their current load.* The *Capacity Utilization* is the ratio of the bandwidth utilized by a node in transmitting its load and the node capacity.
- Compared to other proposed passive local measurement or active bandwidth estimation algorithms, our *Capacity Utilization* estimator based upon neighbour observations eliminates the shortcomings of both of them. The remote *Capacity Utilization* estimator is more reliable, accurate, and completely passive and has no impact on the existing network traffic and network performance. From the perspective of wireless applications, our estimator can be used to directly optimize and improve the performance of those applications, such as the AP selection mechanism in ANDSF, resource aware routing, channel selection and admission control.
- Our *Capacity Utilization* estimator considers both successful and failed transmissions, models the IEEE 802.11 DCF exponential backoff mechanism and takes explicit account of hidden nodes.
- The main shortcoming is the unavoidable error associated with the *Capacity Utilization* estimation because the monitor node and the observed node do not necessarily see the same neighbour nodes, i.e. they do not experience the same medium. The main factors that affect the accuracy of estimations are as follows: the number of unobservable neighbours of the observed node and the traffic load of

unobservable neighbour nodes.

- The factors influencing the accuracy of the *Capacity Utilization* estimator before and after the modifications are investigated in chapter 5. The remote *Capacity Utilization* estimator has a higher accuracy under a smaller number of neighbours, a larger number of observable neighbours, and a light traffic load sent from neighbour nodes.
- In order to minimize the error associated with the estimation, the *Capacity Utilization* estimator uses three reasonable assumptions. The accuracy of the *Capacity Utilization* estimation has been shown to be significantly enhanced as a result of the modifications.
- Our *Capacity Utilization* estimator is evaluated by two metrics, namely the failed detection ratio (*FDR*) and false alarm ratio (*FAR*) in a node saturation detection application. An adjustable threshold CU^{TH} selected through different saturation probability allows for a trade-off between *FDR* and *FAR* in order to enhance the accuracy. The simulation results show the feasibility, usage and accuracy of our *Capacity Utilization* estimator in detecting node saturation compared with queue monitoring method and a regularly pinging method.
- Based on the analysis presented in this thesis, it has been shown that our estimator which uses remote observations performed by neighbour nodes is feasible, applicable and accurate in measuring neighbours' *Capacity Utilization*. It also can be utilized in wireless networks with multiple-neighbours and is applicable in many wireless applications.

6.2 Suggestions for Future Work

This section presents some suggestions for possible future work that could extend the work of this thesis.

6.2.1 Validate, Improve and Extend the Performance of the *Capacity Utilization* Estimator

In this section, some additional future work is proposed that could address some of the weaknesses and limitations of our estimator:

- *Validation and implementation on a testbed experiment.* In this thesis, we have used the ns2 simulator to test the feasibility and evaluate the performance under the test scenarios. Our *Capacity Utilization* estimator should be validated and implemented in a testbed experiment in both indoor and outdoor environments in order to examine the performance of the estimator under channel errors due to interference or fading caused by physical objects such as walls and doors. An IEEE 802.11 a/b/g wireless adapter card on a PC based on Linux platform installed with the modern Atheros driver (due to the development of *madwifi* [153] has ceased in 2008) such as ath5k [154] (for IEEE 802.11a/b/g), ath9k [155] (for IEEE 802.11a/b/g/n) or ath10k [156] (for IEEE 802.11ac) can be used to configure an monitor mode interface. Using this interface, a *Libpcap* [157] program can be employed to monitor the network, capture the packets, analyse their transmissions and estimate the *Capacity Utilization* of neighbour nodes within its reception range.
- *Further improving the accuracy of the Capacity Utilization estimator.* The size of sliding window has an impact on the accuracy of the average contention estimation. A smaller sliding window size may cause an underestimation of average contention under a high level of frame retransmissions. Conversely, a larger sliding window size may lead to an overestimation of average contention when the level of retransmissions is lower. Therefore, the size of sliding window (described in Chapter 4) should be adaptively adjusted where the higher the collision probability or the level of retransmissions within the reception range of the observed node, the larger the size of

the sliding window that should be used. Moreover, when the number of neighbours is large and there is a high traffic rate among the network nodes, many collisions will involve three or more nodes. Thus the *Assumption 3* will need to be modified to adjust the collision probabilities dynamically. The probability distribution of collisions can be calculated by an analytical method or modeled and measured through simulation, i.e. the probability of the collisions involving two P_{2n}, three P_{3n} or more *Hn* nodes P_{Hn}. When the collision probability exceeds a predefined collision probability threshold, the estimator should assume that *most collisions involve in three nodes or more*. The neighbour load bandwidth should be:

$$\sum_{j\neq k}^{Est} BW_{load}(j)^{mesr} = \sum_{j\neq k}^{Est} BW_{load}(j)^{succ} + \frac{1}{2}\sum_{j\neq k}^{Est} BW_{load}(j)^{fail} \times P_{2n}$$

$$+\frac{1}{3}\sum_{j\neq k}^{Est} BW_{load}(j)^{fail} \times P_{3n} + \cdots + \frac{1}{H}\sum_{j\neq k}^{Est} BW_{load}(j)^{fail} \times P_{Hn} \qquad (6.1)$$

- *Extending the Capacity Utilization estimator to include EDCA operation under the IEEE 802.11e standard.* Currently, our *Capacity Utilization* estimator is designed for the original IEEE 802.11 DCF MAC mechanism where each node competes for access under the same MAC mechanism conditions. Each network nodes measures *average contention* and access bandwidth by using a fixed contention window size (e.g. CW_{min} = 31, CW_{max} = 1023 in IEEE 802.11b networks), interframe space (a fixed DIFS value) and retry counters in the present scheme. However, DCF does not support QoS and priority traffic categories for real-time streaming multimedia applications in WLANs. The prioritised access mechanism EDCA specified by IEEE 802.11e [23] defines four access categories (ACs) with different *AIFSN*, ECW_{min}, and ECW_{max} parameters permitted (which have been discussed in Chapter 2) to differentiate the service types in order to support QoS guarantees for real-time

applications. Therefore, our *Capacity Utilization* estimator will be further extended to include the EDCA mechanism in IEEE 802.11e. However, the contention is difficult to measure because the deferral time interval and backoff interval of a frame depends on the traffic priority. Our sliding window scheme which employs a fixed sliding window size calculated to measure *average contention* will not be appropriate for EDCA operation. Moreover, our *Assumption 2* will not be suitable due to the unpredicted interval of AIFS and backoff of unobserved transmitted load (i.e. unknown traffic priority). A new mechanism to estimate the *average contention* and BW_{access} bandwidth based upon EDCA mechanism instead of using sliding windows needs to be developed and designed in the future.

- *Including support for rate adaption mechanism.* Currently, the *Capacity Utilization* estimator does not include the impact of the line rate adaption mechanism. However, the line rate is not constant and is dynamically adjusted according to the communication conditions. In IEEE 802.11b, the line rate can be reduced from 11 Mbps to 5.5 Mbps, then 2 Mbps and 1 Mbps due to poor RF condition. In this thesis, the line rate is assumed to be fixed (the maximum value of 11 Mbps) which will lead to errors being associated with the *Capacity Utilization* estimation in real environments. When the WLAN adapter receives a packet, the driver (i.e. ath5k) can decode the frame and obtain the transmission rate through the modulation scheme used for the frames. Then we can calculate the load bandwidth of the observed node and neighbours and estimate the load bandwidth of the unobserved neighbours by *Assumption 1*. The impact of this rate adaption mechanism under IEEE 802.11a/b/g/n/ac on the accuracy of estimator will be investigated in the future, i.e. a lower line rate will produce larger traffic loads that will give rise to a higher actual *Capacity Utilization* value and larger errors associated with the estimators.

6.2.2 Wireless Application Areas for the *Capacity Utilization* Estimator

In this section, the challenges and potential applications for our estimator in other WLAN areas will be introduced:

- *Our Capacity Utilization estimator will be investigated to support AP selection and handoff in ANDSF.* In traditional networks, the client selects an AP based upon RSSI. However, it is not suitable for large scale wireless networks such as public hotspots (i.e. an airport or a university campus) because it may cause higher contention in the medium, overload of an AP, AP saturation or congestion due to an asymmetric flow of traffic, i.e. numerous clients downloading via a single AP, and the underutilization of the other APs' resources which may degrade the performance of the whole network. Using RSSI as an indicator to support AP selection also cannot maintain the load balancing that leads to the unsatisfied requirements of users. Therefore, before a client attempts to associate with an AP, the client could check both the *Capacity Utilization* information of APs and RSSI before making an association decision. Compared to the IEEE 802.11k mechanism [12], the remote *Capacity Utilization* estimator does not need any information exchange before the association phase. A hybrid AP selection mechanism that combines our *Capacity Utilization* estimator and RSSI of two or more APs can be designed to provide a better performance to wireless uses. Three approaches can be employed to solve this problem for AP selection mechanism in the future:

  I. Firstly, the clients can simply choose an AP with the lower *Capacity Utilization* whose RSSI can maintain the connection.

  II. Secondly, this algorithm classifies the traffic into different priorities to select different APs, i.e., for time-sensitive traffic types such as Voice or Video traffic types which have smaller packet sizes, it should choose the AP with the stronger

signal in order to avoid rate adaption due to poor link quality. For some non-time sensitive traffic types which have larger packet sizes such as Email, Web request or offline download, it should choose the AP with the lower *Capacity Utilization* which indicates a lower contention that may have a lower collision probability, lower retransmission probability, and even lower packet losses.

III. Thirdly, a new metric which considers both *Capacity Utilization* and RSSI can be developed to evaluate the performance the APs within the reception range. This metric called the *Connection Quality (CQ)* of AP *k* can be defined as:

$$CQ(k) = \frac{\%CU(k)}{\%RSSI(k)} \tag{6.2}$$

Where $RSSI$ is the actual signal strength value (dB), $RSSI_{min}$ is the minimum signal strength required to maintain a reliable connection (e.g. -95 dBm [158] for most chipsets), $\%RSSI(k)$ reflects the ratio of real signal strength that can be derived as:

$$\%RSSI(k) = \frac{RSSI - RSSI_{min}}{RSSI_{min}} \tag{6.3}$$

Both smaller $\%CU(k)$ and larger $\%RSSI(k)$ value lead to smaller *CQ* values which would indicate that an AP *k* is capable of providing a more reliable connection and lower *Capacity Utilization*. Thus the client should associate with the AP with the smaller *CQ* value. This *CQ* metric could be used to support information discovery in ANDSF.

Moreover, an intelligent seamless handover mechanism for WLANs through the use of our *Capacity Utilization* estimator can also be developed in the future in order to guarantee the QoS of the mobile users' applications. When the mobile users' traffic loads requirement cannot be satisfied, the users can switch to the AP which has the lower *Capacity Utilization*.

- *A channel selection and assignment mechanism by employing our remote Capacity*

Utilization estimator in order to avoid the node saturation or congestion in multi-channel networks can be investigated in the future. The traditional channel selection and assignment methods are divided into static channel assignment [44] which assigns the channel to the WLANs nodes permanently, dynamic channel assignment [45] which allocates the channel dynamically after receiving the request, and hybrid channel assignment [46]. However, none of them considers the *Capacity Utilization* of the channels used. Our estimator can be used as a simple metric to select the appropriate channel for the users to satisfy their traffic load requirement. Once the user detects that its neighbour nodes in the current channel are saturated, then the user can scan the other channels in WLANs to find another appropriate channel to avoid or alleviate node saturation and congestion.

- *Our Capacity Utilization estimator will be used as a route metric to find paths from a source node to destination node in the future.* Generally, there are two parts in finding a good path in WLANs: routing metrics and routing information dissemination [159]. The metric hop counter [160], average *RTT* [161], expected transmission count [41] and effective number of transmissions [162] have been proposed as a routing metric to provide better performance of user service. However, most of them need to measure the metric of a "link" between a pair of nodes and require information dissemination. Moreover, some routing metric and routing protocols have a high overhead (e.g. DSDV [37]) and latency (e.g. AODV [39]) in finding a path which does not consider the resource utilization and the user's traffic load requirement. Therefore, we can use our *Capacity Utilization* estimator to design and support routing protocols in the future, i.e. *Capacity Utilization* aware routing (*CUAR*). The node monitors its neighbours' *Capacity Utilization* value periodically or under poor QoS link conditions, then re-routes by selecting a new path using the neighbours' *Capacity Utilization*

information to satisfy its QoS requirement. The network node checks its next-hop neighbour nodes' *Capacity Utilization* and combines it with the hop counter metric [160] which is a popular shortest-path metric used to minimize the end-to-end delay. Taking the *Capacity Utilization* metric into account in routing protocols may effectively and promptly avoid the node saturation condition that can gave rise to packet delay and losses.

References

1. IEEE 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. *IEEE Std 802.11-1997*. pp. i-445. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=654749&isnumber=14251. (Last Accessed: 30/4/2014).
2. Prasad, R., et al., Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. *IEEE Network*, 2003. 17(6): pp. 27-35.
3. Johnsson, A., B. Melander, and M. Björkman. Bandwidth Measurement in Wireless Networks. in *Fourth Annual Mediterranean Ad Hoc Networking Workshop*. Île de Porquerolles, France. 21 - 24 June, 2005.
4. Li, F.Y., et al. Estimating Residual Bandwidth in 802.11-based Ad Hoc Networks: An Empirical Approach. in *WPMC 2004, Seventh International Symposium on Wireless Personal Multimedia Communications*. Abano Terme, Italy. 12 - 15 Septemper, 2004
5. Bredel, M. and M. Fidler. A Measurement Study of Bandwidth Estimation in IEEE 802.11g Wireless LANs using the DCF. in *Seventh International IFIP-TC6 Networking Conference*. Singapore. 5 - 9 May, 2008.
6. Gupta, D., et al. Experimental Comparison of Bandwidth Estimation Tools for Wireless Mesh Networks. in *IEEE INFOCOM 2009*. Rio de Janeiro, Brazil. 19 - 25 April, 2009.

7. Angrisani, L., et al. Measuring Wireless Links Capacity. in *ISWPC, First International Symposium on Wireless Pervasive Computing*. Phuket, Thailand. 16 - 18 January, 2006

8. Chen, L. and W.B. Heinzelman, QoS-Aware Routing based on Bandwidth Estimation for Mobile Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications*, 2005. 23(3): pp. 561-572.

9. Chaudet, C. and I.G. Lassous. BRuIT: Bandwidth Reservation under Interferences Influence. in *European Wireless 2002*. Florence, Italy. 25 - 28 February, 2002.

10. 3GPP TS 24.312 version8.0.0: Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Release 8). 2008. pp. 1-91. [Online]: Available at: http://www.3gpp.org/DynaReport/24312.htm. (Last Accessed: 30/4/2014).

11. Abusubaih, M., et al., A New Access Point Selection Policy for Multi-Rate IEEE 802.11 WLANs. *International Journal of Parallel, Emergent and Distributed Systems*, 2008. 23(4): pp. 291-307.

12. IEEE Standard for Information Technology-- Local and Metropolitan Area Networks-- Specific Requirements-- Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs. *IEEE Std 802.11k-2008 (Amendment to IEEE Std 802.11-2007)*. 2008. pp. 1-244. [Online]: Available at: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4544755&sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A4544754%29. (Last Accessed: 30/4/2014).

13. Davis, M. A Wireless Traffic Probe for Radio Resource Management and QoS Provisioning in IEEE 802.11 WLANs. in *Seventh ACM International Symposium*

on Modeling, Analysis and Simulation of Wireless and Mobile Systems. Venice, Italy. 4 - 6 October, 2004

14. Davis, M. and T. Raimondi. A Novel Framework for Radio Resource Management in IEEE 802.11 Wireless LANs. in *WIOPT 2005, Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*. Riva del Garda, Trentino, Italy. 3 - 7 April, 2005.
15. Gast, M., *802.11 Wireless Networks: The Definitive Guide*, Second Edition, 2005. O'Reilly Media Inc., ISBN: 7-5641-0316-7.
16. IEEE Standards for Local Area Networks: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications. *ANSI/IEEE Std 802.3-1985*. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=30707&isnumber=1314. (Last Accessed: 30/4/2014).
17. IEEE Standards for Local Area Networks: Token Ring Access Method and Physical Layer Specifications. *IEEE Std 802.5-1989*. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=193467&isnumber=4988. (Last Accessed: 30/4/2014).
18. Logical Link Control. *ANSI/IEEE Std 802.2-1985*. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=30705&isnumber=1313. (Last Accessed: 30/4/2014).
19. IEEE 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: Higher-Speed Physical Layer Extension in the 2.4 GHz Band. *IEEE Std 802.11b-1999*. pp. i-90. [Online]: Available at:

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=817038&isnumber=17714. (Last Accessed: 30/4/2014).

20. 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 Amendment 1: High-Speed Physical Layer in the 5 GHz Band. *IEEE Std 802.11a-1999*. pp. i-83. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=887490&isnumber=19177. (Last Accessed: 30/4/2014).

21. IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks- Specific Requirements Part Ii: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications-Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band. *IEEE Std 802.11g-2003* pp. i-67. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1210624&isnumber=27242. (Last Accessed: 30/4/2014).

22. IEEE Standard for Information Technology-- Local and Metropolitan Area Networks-- Specific Requirements-- Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput. *IEEE Std 802.11n-2009* pp. 1-565. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5307322&isnumber=5307317. (Last Accessed: 30/4/2014).

23. IEEE 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 Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements. *IEEE Std 802.11e-2005*. pp. 1-212. [Online]: Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1541572&isnumber=32891. (Last Accessed: 30/4/2014).

24. Choi, S., J. Del Prado, and S. Mangold. IEEE 802.11e Contention-based Channel Access (EDCF) Performance Evaluation. in *ICC'03, IEEE International Conference on Communications*. Anchorage, Alaska, U.S.A. 11 - 15 May, 2003

25. Hiertz, G.R., et al., IEEE 802.11s: the WLAN Mesh Standard. *IEEE Wireless Communications*, 2010. 17(1): pp. 104-111.

26. Akyildiz, I.F. and X. Wang, A Survey on Wireless Mesh Networks. *IEEE Communications Magazine*, 2005. 43(9): pp. S23-S30.

27. Gungor, V., et al., *Challenges and Issues in Designing Architectures and Protocols for Wireless Mesh Networks*, in *Wireless Mesh Networks*, 2007, Springer Science+Business Media, LLC, ISBN: 978-0-387-68838-1. pp. 1-27.

28. Xu, K., M. Gerla, and S. Bae, Effectiveness of RTS/CTS Handshake in IEEE 802.11 Based Ad Hoc Networks. *Ad Hoc Networks*, 2003. 1(1): pp. 107-123.

29. Vassis, D., et al., The IEEE 802.11g Standard for High Data Rate WLANs. *IEEE Network*, 2005. 19(3): pp. 21-26.

30. Bianchi, G., Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, 2000. 18(3): pp. 535-547.

31. Lakshminarayanan, K., V.N. Padmanabhan, and J. Padhye. Bandwidth Estimation in Broadband Access Networks. in *Fourth ACM SIGCOMM Conference on Internet Measurement*. Taormina, Sicily, Italy. 25 - 27 October, 2004

32. Biaz, S. and S. Wu. Rate Adaptation Algorithms for IEEE 802.11 Networks: A Survey and Comparison. in *ISCC'08, IEEE Symposium on Computers and Communications*. Marrakech, Morocco. 6 - 9 July, 2008.

33. Yang, Y. and R. Kravets, Contention-Aware Admission Control for Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 2005. 4(4): pp. 363-377.

34. Vasudevan, S., et al. Facilitating Access Point Selection in IEEE 802.11 Wireless Networks. in *Fifth ACM SIGCOMM Conference on Internet Measurement*. Berkeley, California, U.S.A. 19 - 21 October, 2005.

35. Garcia Villegas, E., R. Vidal Ferre, and J. Paradells Aspas. Load Balancing in WLANs through IEEE 802.11k Mechanisms. in *Eleventh IEEE Symposium on Computers and Communications*. Cagliari, Sardinia, Italy. 26 - 29 June, 2006.

36. 3GPP TS 23.402 version11.8.0: Architecture Enhancements for Non-3GPP Accesses (Release 11). 2013. pp. 1-254. [Online]: Available at: http://www.3gpp.org/DynaReport/23402.htm. (Last Accessed: 30/4/2014).

37. Perkins, C.E. and P. Bhagwat, Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. *ACM SIGCOMM Computer Communication Review*, 1994. 24: pp. 234-244.

38. Jacquet, P., et al. Optimized Link State Routing Protocol for Ad Hoc Networks. in *IEEE INMIC 2001, Technology for the 21st Century. IEEE International Multi Topic Conference*. Lahore, Pakistan. 30 - 30 December, 2001.

39. Perkins, C.E. and E.M. Royer. Ad-Hoc On-Demand Distance Vector Routing. in *WMCSA '99, Second IEEE Workshop on Mobile Computing Systems and Applications*. New Orleans, LA, U.S.A. 25 - 26 February,1999.

40. Johnson, D.B. and D.A. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks*, in *Mobile Computing*, 1996, Springer US, ISBN: 978-0-7923-9697-0. pp. 153-181.

41. De Couto, D.S., et al., A High-Throughput Path Metric for Multi-Hop Wireless Routing. *Wireless Networks*, 2005. 11(4): pp. 419-434.

42. Lin, C.R. and J.-S. Liu, QoS Routing in Ad Hoc Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 1999. 17(8): pp. 1426-1438.

43. Kowalik, K., B. Keegan, and M. Davis. Rare-Resource Aware Routing for Mesh. in *ICC'07, IEEE International Conference on Communications*. Glasgow, Scotland. 24 - 28 June, 2007.

44. Raniwala, A., K. Gopalan, and T.-c. Chiueh, Centralized Channel Assignment and Routing Algorithms for Multi-Channel Wireless Mesh Networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2004. 8(2): pp. 50-65.

45. Bahl, P., R. Chandra, and J. Dunagan. SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad-Hoc Wireless Networks. in *ACM MobiCom 2004, Tenth Annual International Conference on Mobile Computing and Networking*. Philadelphia, Pennsylvania, U.S.A. 26 September - 1 October, 2004.

46. Kyasanur, P. and N.H. Vaidya, Routing and Link-Layer Protocols for Multi-Channel Multi-Interface Ad Hoc Wireless Networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2006. 10(1): pp. 31-43.

47. Ahmed, M.H., Call Admission Control in Wireless Networks: A Comprehensive Survey. *IEEE communications Surveys and Tutorials*, 2005. 7(1-4): pp. 50-69.

48. Fall, K. and K. Varadhan. "*The Network Simulator NS-2"*. 2007. [Online]: Available at: http://www.isi.edu/nsnam/ns. (Last Accessed: 30/4/2014).

49. Issariyakul, T. and E. Hossain, *An Introduction to Network Simulator NS2*, Second Edition, 2012. Springer, ISBN: 978-1-4614-1405-6.

50. Kurkowski, S., T. Camp, and M. Colagrosso, MANET Simulation Studies: the Incredibles. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2005. 9(4): pp. 50-61.

51. Ivanov, S., A. Herms, and G. Lukas. Experimental Validation of the ns-2 Wireless Model using Simulation, Emulation, and Real Network. in *Communication in Distributed Systems (KiVS), 2007 ITG-GI Conference*. Bern, Switzerland. 26 February - 2 March 2007

52. Reineck, K.M., MSc Thesis: *Evaluation and Comparison of Network Simulation Tools*, Department of Computer Sciences, University of Applied Sciences Bonn-Rhein-Sieg, Bonn, Germany, 2008

53. Kotz, D., et al. Experimental Evaluation of Wireless Simulation Assumptions. in *Seventh ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. Venice, Italy. 4 - 6 October, 2004

54. Liu, J., et al. Simulation Validation using Direct Execution of Wireless Ad Hoc Routing Protocols. in *PADS 2004, Eighteenth Workshop on Parallel and Distributed Simulation*. Kufstein, Austria. 16 - 19 May, 2004.

55. Kaixin, X., et al. Adaptive Bandwidth Management and QoS Provisioning in Large Scale Ad Hoc Networks. in *MILCOM '03, Military Communications Conference*. Boston, MA, U.S.A. 13 - 16 October, 2003.

56. Wu, C., S. Ohzahata, and T. Kato. A Routing Protocol for Cognitive Radio Ad Hoc Networks Giving Consideration to Future Channel Assignment. in *CANDAR 2013, First International Symposium on Computing and Networking*. Matsuyama, Japan. 4 - 6 December, 2013

57. Shen, Q., et al., Admission Control based on Available Bandwidth Estimation for Wireless Mesh Networks. *IEEE Transactions on Vehicular Technology*, 2009. 58(5): pp. 2519-2528.

58. Downey, A.B. Using Pathchar to Estimate Internet Link Characteristics. in *SIGCOMM '99, Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. Cambridge, Massachusetts, U.S.A. 30 August - 3 September, 1999.

59. Jacobson, V. "*Pathchar: A Tool to Infer Characteristics of Internet Paths*". 1997. [Online]: Available at: ftp://ftp.kfki.hu/pub/packages/security/COAST/netutils/pathchar/msri-talk.pdf. (Last Accessed: 30/4/2014).

60. Pásztor, A. and D. Veitch. Active Probing using Packet Quartets. in *Second ACM SIGCOMM Workshop on Internet Measurment*. Marseille, France. 6 - 8 November, 2002

61. Lai, K. and M. Baker, Measuring Link Bandwidths using a Deterministic Model of Packet Delay. *ACM SIGCOMM Computer Communication Review*, 2000. 30(4): pp. 283-294.

62. Jacobson, V., Congestion Avoidance and Control. *ACM SIGCOMM Computer Communication Review*, 1988. 18(4): pp. 314-329.

63. Bolot, J.-C., End-to-End Packet Delay and Loss Behavior in the Internet. *ACM SIGCOMM Computer Communication Review*, 1993. 23(4): pp. 289-298.

64. Dovrolis, C., P. Ramanathan, and D. Moore. What Do Packet Dispersion Techniques Measure? in *INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*. Anchorage, Alaska, U.S.A. 22 - 26 April, 2001

65. Strauss, J., D. Katabi, and F. Kaashoek. A Measurement Study of Available Bandwidth Estimation Tools. in *Third ACM SIGCOMM Conference on Internet Measurement*. Miami Beach, FL, U.S.A. 27 - 29 October, 2003

66. Hu, N. and P. Steenkiste, Evaluation and Characterization of Available Bandwidth Probing Techniques. *IEEE Journal on Selected Areas in Communications*, 2003. 21(6): pp. 879-894.

67. Ribeiro, V., et al. Multifractal Cross-Traffic Estimation. in *ITC specialist seminar on IP traffic Measurement*. Monterey, CA, U.S.A. 18 - 20 September, 2000.

68. Jain, M. and C. Dovrolis, End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput. *ACM SIGCOMM Computer Communication Review*, 2002. 32(4): pp. 295-308.

69. Melander, B., M. Bjorkman, and P. Gunningberg. A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks. in *GLOBECOM'00, IEEE Global Telecommunications Conference*. San Francisco, U.S.A. 27 November - 1 December, 2000.

70. Jain, M. and C. Dovrolis. Pathload: A Measurement Tool for End-to-End Available Bandwidth. in *Passive and Active Measurements Workshop*. Collins, Colorado, U.S.A. 25 - 27 March, 2002.

71. Ribeiro, V., et al. Pathchirp: Efficient Available Bandwidth Estimation for Network Paths. in *Passive and Active Measurement Workshop*. La Jolla, California, U.S.A. 6 - 8 April, 2003.

72. Kiwior, D., J. Kingston, and A. Spratt. PathMon, a Methodology for Determining Available Bandwidth over an Unknown Network. in *2004 IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*. Princeton, NJ, U.S.A. 26 - 27 April, 2004

73. Ekelin, S., et al. Real-Time Measurement of End-to-End Available Bandwidth using Kalman Filtering. in *NOMS 2006, Tenth IEEE/IFIP Network Operations and Management Symposium*. 3 - 7 April, 2006.

74. Sedighizad, M., B. Seyfe, and K. Navaie, MR-BART: Multi-Rate Available Bandwidth Estimation in Real-Time. *Journal of Network and Computer Applications*, 2012. 35(2): pp. 731-742.

75. Wang, S., et al. Providing Statistical Delay Guarantees in Wireless Networks. in *Twenty-fourth International Conference on Distributed Computing Systems*. Hachioji, Tokyo, Japan. 24 - 26 March, 2004.

76. Fu, H., L. Lin, and W. Jia. Efficient Wireless Link Bandwidth Detection for IEEE 802.11 Networks. in *ICC 2005, IEEE International Conference on Communications*. Seoul, Korea. 16 - 20 May, 2005

77. Alzate, M.A., et al. End-to-end Mean Bandwidth Estimation as a Function of Packet Length in Mobile Ad Hoc Networks. in *ISCC 2007, Twelfth IEEE Symposium on Computers and Communications*. Aveiro, Portugal. 1 - 4 July, 2007

78. Chen, L.-J., et al. Ad Hoc Probe: Path Capacity Probing in Wireless Ad Hoc Networks. in *WICON 2005, First IEEE International Conference on Wireless Internet*. Budapest, Hungary. 10 - 15 July, 2005.

79. Kapoor, R., et al. CapProbe: A Simple and Accurate Capacity Estimation Technique. in *ACM SIGCOMM 2004 Conference on Applications, Technologies,*

Architectures, and Protocols for Computer Communication. Portland, Oregon, U.S.A. 30 August -3 September, 2004.

80. Venkatesh, G. and K.-C. Wang. Estimation of Maximum Achievable End-to-End Throughput in IEEE 802.11 based Wireless Mesh Networks. in *Thirty-fourth IEEE Conference on Local Computer Networks*. Swissôtel Zürich, Switzerland. 20 - 23 October, 2009

81. Ergin, M.A., et al., Available Bandwidth Estimation and Admission Control for QoS Routing in Wireless Mesh Networks. *Computer Communications*, 2008. 31(7): pp. 1301-1317.

82. Li, M., M. Claypool, and R. Kinicki. WBest: A Bandwidth Estimation Tool for IEEE 802.11 Wireless Networks. in *IEEE Thirty-third Conference on Local Computer Networks*. Montreal, AB, Canada. 14 - 17 October, 2008.

83. Amamra, A. and K.M. Hou. SLOT: A Fast and Accurate Technique to Estimate Available Bandwidth in Wireless IEEE 802.11. in *Tenth International Conference on Computer Modeling and Simulation*. Cambridge, UK. 1 - 3 April, 2008.

84. Johnsson, A. and M. Bjorkman. On Measuring Available Bandwidth in Wireless Networks. in *IEEE Thirty-third Conference on Local Computer Networks*. Montreal, AB, Canada. 14 - 17 October, 2008

85. Nielsen, M.N., K. Ovsthus, and L. Landmark. Field Trials of Two 802.11 Residual Bandwidth Estimation Methods. in *IEEE International Conference on Mobile Adhoc and Sensor Systems*. Vancouver, BC, Canada. 9 - 12 October, 2006

86. Duffy, K., D. Malone, and D.J. Leith, Modeling the 802.11 Distributed Coordination Function in Non-Saturated Conditions. *IEEE Communications Letters*, 2005. 9(8): pp. 715-717.

87. Garetto, M., T. Salonidis, and E.W. Knightly. Modeling Per-Flow Throughput and Capturing Starvation in CSMA Multi-Hop Wireless Networks. in *INFOCOM 2006*. Barcelona, Catalunya, Spain. 23 - 29 April, 2006.

88. Kwak, B.-J., N.-O. Song, and M. Miller, Performance Analysis of Exponential Backoff. *IEEE/ACM Transactions on Networking*, 2005. 13(2): pp. 343-355.

89. Gao, Y., D.-M. Chiu, and J. Lui. Determining the End-to-End Throughput Capacity in Multi-Hop Networks: Methodology and Applications. in *SIGMETRICS/Performance 2006, Joint International Conference on Measurement and Modeling of Computer Systems*. Saint Malo, France. 26 - 30 June, 2006.

90. Gupta, P. and P.R. Kumar, The Capacity of Wireless Networks. *IEEE Transactions on Information Theory*, 2000. 46(2): pp. 388-404.

91. Kumar, V., et al. Algorithmic Aspects of Capacity in Wireless Networks. in *SIGMETRICS 2005, International Conference on Measurements and Modeling of Computer Systems*. Banff, Alberta, Canada. 6 - 10 June, 2005.

92. Babu, A., L. Jacob, and S. Karimuddin. Nonintrusive Estimation of Link Characteristics in IEEE 802.11 Wireless LANs. in *Thirteenth IEEE International Conference on Networks jointly held with the 2005 Seventh IEEE Malaysia International Conference on Communications*. Kuala Lumpur, Malaysia. 16 - 18 November, 2005.

93. Anna, K. and M. Bassiouni. Analysis of the Saturation Throughput and Node Contention Levels in the 802.11 MAC Protocol. in *2010 IEEE Radio and Wireless Symposium*. New Orleans, LA, U.S.A. 10 - 14 January, 2010

94. Li, J., et al. Capacity of Ad Hoc Wireless Networks. in *Seventh Annual International Conference on Mobile Computing and Networking*. Rome, Italy. 16 - 21 July, 2001

95. Wu, D.-p., R.-y. Wang, and S. Huang. The Research on the Factors Affecting Network Bandwidth in Mobile Ad Hoc Networks. in *WiCOM 2010, Sixth International Conference on Wireless Communications, Networking and Mobile Computing*. Chengdu City, Sichuan Province, P.R.China. 23 - 25 September, 2010.

96. Guerin, J., et al. Time-based and Low-Cost Bandwidth Estimation for IEEE 802.11 Links. in *Eighth International Wireless Communications and Mobile Computing Conference*. Limassol, Cyprus. 27 - 31 August, 2012.

97. Lee, H., et al. Available Bandwidth-based Association in IEEE 802.11 Wireless LANs. in *Eleventh International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. Vancouver, British Columbia, Canada. 27 - 31 October, 2008.

98. Wang, Q., A. Feng, and J. Cao. Available Bandwidth Estimation in IEEE 802.11 Ad Hoc Networks. in *Ninth International Conference on Hybrid Intelligent Systems*. Shenyang, Liaoning Province, P.R.China. 12 - 14 August, 2009

99. Peng, Y., et al., An Efficient Joint Channel Assignment and QoS Routing Protocol for IEEE 802.11 Multi-Radio Multi-Channel Wireless Mesh Networks. *Journal of Network and Computer Applications*, 2013. 36(2): pp. 843-857.

100. Deng, X., et al. Channel Quality and Load Aware Routing in Wireless Mesh Network. in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. Shanghai, P.R.China. 7 - 10 April, 2013.

101. Khan, M.A.Y. and D. Veitch. Speedo: Realistic Achievable Bandwidth in 802.11 through Passive Monitoring. in *LCN 2008, Thirty-third Conference on Local Computer Networks*. Montreal, AB, Canada. 14-17 October, 2008

102. Wang, H., L. Gao, and Z. Li. Node-to-Node Available Bandwidth Estimation in Ad Hoc Networks. in *2008 International Conference on Computer and Electrical Engineering*. Phuket, Thailand. 20 - 22 December, 2008

103. Garcia, E., et al. Achievable Bandwidth Estimation for Stations in Multi-Rate IEEE 802.11 WLAN Cells. in *WoWMoM 2007, IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*. Espoo, Finland. 18 - 21 June, 2007.

104. Oriol, D., N. Thi Mai Trang, and G. Pujolle. Available Bandwidth Estimation in Wireless Mesh Networks using Neural Networks. in *NoF 2012, Third International Conference on the Network of the Future*. Tunis, Tunisia. 21 - 23 November, 2012.

105. Chakeres, I.D. and E.M. Belding-Royer. PAC: Perceptive Admission Control for Mobile Wireless Networks. in *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*. Dallas, TX, U.S.A. 18 - 20 October, 2004.

106. de Renesse, R., V. Friderikos, and H. Aghvami, Cross-Layer Cooperation for Accurate Admission Control Decisions in Mobile Ad Hoc Networks. *IET communications*, 2007. 1(4): pp. 577-586.

107. Du, L. and L. Chen. Receiver Initiated Network Allocation Vector Clearing Method in WLANs. in *Eleventh Asia-Pacific Conference on Communications*. Perth, Western Australia. 3 - 5 October, 2005.

108. Lee, H.K., et al., Bandwidth estimation in wireless lans for multimedia streaming services. *Advances in Multimedia*, 2007. 2007(1): pp. 9-9.

109. Liu, H. and L. Cheng, Available Bandwidth Estimation Strategy Based on the Network Allocation Vector. *Journal of Networks*, 2012. 7(12): pp. 2089-2095.

110. Chen, C., C. Pei, and L. An. Available Bandwidth Estimation in IEEE802. 11b Network Based on Non-Intrusive Measurement. in *Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies*. Taipei, Taiwan. 4 - 7 December, 2006.

111. Ru-yan, W., D. Qi, and W. Da-peng. The Research on Available Bandwidth Estimating in Mobile Ad Hoc Networks. in *Second International Conference on Future Computer and Communication*. Wuhan City, Hubei Province, P.R.China. 21 - 24 May, 2010.

112. Sarr, C., et al., A Node-based Available Bandwidth Evaluation in IEEE 802.11 Ad Hoc Networks. *The International Journal of Parallel, Emergent and Distributed Systems*, 2006. 21(6): pp. 423-440.

113. Sarr, C., et al., Bandwidth Estimation for IEEE 802.11-based Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 2008. 7(10): pp. 1228-1241.

114. Sarr, C., et al. Improving Accuracy in Available Bandwidth Estimation for 802.11-based Ad Hoc Networks. in *2006 IEEE International Conference on Mobile Adhoc and Sensor Systems*. Vancouver, BC, Canada. 9 - 12 October, 2006.

115. Belbachir, R., Z.M. Mekkakia, and A. Kies, Towards a New Approach in Available Bandwidth Measures on Mobile Ad Hoc Networks. *Acta Polytechnica Hungarica*, 2011. 8(4).

116. Zhao, H. and E. Garcia-Palacios, Rethinking Available Bandwidth Estimation in IEEE 802.11-based Ad Hoc Networks. *Electronics Letters*, 2009. 45(4): pp. 211-213.

117. Zhao, H., et al., Accurate Available Bandwidth Estimation in IEEE 802.11-based Ad Hoc Networks. *Computer Communications*, 2009. 32(6): pp. 1050-1057.

118. Tursunova, S., K. Inoyatov, and Y.-T. Kim. Cognitive Passive Estimation of Available Bandwidth (cPEAB) in Overlapped IEEE 802.11 WiFi WLANs. in *NOMS 2010, Twelfth IEEE/IFIP Network Operations and Management Symposium*. Osaka, Japan. 19 - 23 April, 2010.

119. Wei, A. "*Routing Algorithm based on the Flow Sensing Parameter"*. 2009. [Online]: Available at: http://tools.ietf.org/html/draft-wei-manet-rafsp-00. (Last Accessed: 30/4/2014).

120. Park, H.J. and B.-h. Roh. Accurate Passive Bandwidth Estimation (APBE) in IEEE 802.11 Wireless LANs. in *Fifth International Conference on Ubiquitous Information Technologies and Applications*. Sanya, Henan Province, P.R.China. 16 - 18 December, 2010.

121. Nguyen, N.V., et al. Retransmission-based Available Bandwidth Estimation in IEEE 802.11-based Multihop Wireless Networks. in *Fourteenth ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. Miami, FL, U.S.A. 31 October - 4 November, 2011.

122. Peng, Z., et al. On an Efficient Estimation of Available Bandwidth for IEEE 802.11-based Wireless Networks. in *GLOBECOM 2011, IEEE Global Telecommunications Conference*. Houston, Texas, U.S.A. 5 - 9 December, 2011.

123. Zhai, H., X. Chen, and Y. Fang, A Call Admission and Rate Control Scheme for Multimedia Support over IEEE 802.11 Wireless LANs. *Wireless Networks*, 2006. 12(4): pp. 451-463.

124. Xiuchao, W. and A.L. Ananda. Link Characteristics Estimation for IEEE 802.11 DCF based WLAN. in *Twenty-ninth Annual IEEE Conference on Local Computer Networks*. Tampa, FL, U.S.A. 16 - 18 November, 2004.

125. Sohrabi, K. and G.J. Pottie. Performance of a Novel Self-Organization Protocol for Wireless Ad-Hoc Sensor Networks. in *VTC 1999-Fall, IEEE VTS Fiftieth Vehicular Technology Conference*. Amsterdam, Netherland. 19 - 22 September, 1999.

126. Ghini, V., S. Ferretti, and F. Panzieri. A Strategy for Best Access Point Selection. in *Third IFIP Wireless Days Conference*. Venice, Italy. 20 - 22 October, 2010.

127. Papaoulakis, N. and C.Z. Patrikakis. A Proactive, Terminal based Best Access Point Selection Mechanism for Wireless LANs. in *2008 IEEE GLOBECOM Workshops*. New Orleans, LA, U.S.A. 30 November - 4 December, 2008.

128. Kwon, Y.M., et al. Andsf-based Congestion Control Procedure in Heterogeneous Networks. in *2013 International Conference on Information Networking*. 28 - 30 January, 2013. Bangkok, Thailand.

129. Triantafyllopoulou, D., T. Guo, and K. Moessner. Energy Efficient ANDSF-Assisted Network Discovery for Non-3GPP Access Networks. in *CAMAD 2012, Seventeenth IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks*. Barcelona, Spain. 17 - 19 September, 2012.

130. Mueller, J., et al. UE & Network Initiated QoS Reservation in NGN and Beyond. in *NoF 2011, Second International Conference on the Network of the Future*. Paris, France. 28 - 30 November, 2011.

131. Wang, N.-C. and C.-Y. Lee, A Reliable QoS Aware Routing Protocol with Slot Assignment for Mobile Ad Hoc Networks. *Journal of Network and Computer Applications*, 2009. 32(6): pp. 1153-1166.

132. Asokan, R. A Review of Quality of Service (QoS) Routing Protocols for Mobile Ad Hoc Networks. in *ICWCSC 2010, International Conference on Wireless Communication and Sensor Computing*. Chennai, Tamil Nadu, India. 2 - 4 January, 2010

133. Park, Y. and E.-S. Jung. Resource-Aware Routing Algorithms for Multi-Hop Cellular Networks. in *MUE'07, International Conference on Multimedia and Ubiquitous Engineering*. Seoul, Korea. 26 - 28 April, 2007.

134. Zhu, C. and M.S. Corson. QoS Routing for Mobile Ad Hoc Networks. in *INFOCOM 2002, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*. New York, NY, U.S.A. 23 - 27 June, 2002.

135. Pong, D. and T. Moors. Call Admission Control for IEEE 802.11 Contention Access Mechanism. in *GLOBECOM'03, IEEE Global Telecommunications Conference*. San Francisco, U.S.A. 1 - 5 December, 2003

136. Duffield, N.G., et al., Multicast Topology Inference from Measured End-to-End Loss. *IEEE Transactions on Information Theory*, 2002. 48(1): pp. 26-45.

137. Song, J., S.S. Lee, and Y.S. Kim. DiffProbe: One Way Delay Measurement for Asynchronous Network and Control Mechanism in BcN Architecture. in *ICACT 2006. The 8th International Conference Advanced Communication Technology*. February, Korea. 20 - 22 February, 2006.

138. Legout, A. and E.W. Biersack, PLM: Fast Convergence for Cumulative Layered Multicast Transmisson Schemes. *ACM SIGMETRICS Performance Evaluation Review*, 2000. 28(1): pp. 13-22.

139. Puangpronpitag, S. and S. Hassan. Explicit rate Adjustment for Multirate Multicast Congestion Control using TCP Throughput Equation and Packet-Pair Probe. in *APCC 2003, The Ninth Asia-Pacific Conference on Communications*. Penang, Malaysia. 21 - 24 September, 2003.

140. Hu, N. and P. Steenkiste. Report: *RPT: A Low Overhead Single-End Probing Tool for Detecting Network Congestion Positions*. Pittsburgh Pa Dept. of Computer Science, Carnegie-Mellon University, 2003. [Online]: Available at: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA461141. (Last Accessed: 30/4/2014).

141. Choi, J., G. Kim, and S. Kim, A Congestion-Aware IDS Node Selection Method for Wireless Sensor Networks. *International Journal of Distributed Sensor Networks*, 2012. 2012.

142. Karenos, K., V. Kalogeraki, and S.V. Krishnamurthy, PhD Thesis: *Cluster-based Congestion Control for Supporting Multiple Classes of Traffic in Sensor Networks*, University of California, Riverside, 2005

143. Matsumoto, M. and T. Nishimura, Mersenne Twister: a 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 1998. 8(1): pp. 3-30.

144. Wan, C.-Y., S.B. Eisenman, and A.T. Campbell. CODA: Congestion Detection and Avoidance in Sensor Networks. in *Proceedings of the 1st international conference on Embedded networked sensor systems*. Los Angeles, California, USA. 5 - 7 November, 2003.

145. Vardi, Y., Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data. *Journal of the American Statistical Association*, 1996. 91(433): pp. 365-377.

146. Ji, C. and A. Elwalid, Measurement-Based Network Monitoring and Inference: Scalability and Missing Information. *IEEE Journal on Selected Areas in Communications*, 2002. 20(4): pp. 714-725.

147. Huang, K., K.R. Duffy, and D. Malone, On the validity of IEEE 802.11 MAC modeling hypotheses. *IEEE/ACM Transactions on Networking*, 2010. 18(6): pp. 1935-1948.

148. Berger, J.O., *Statistical Decision Theory and Bayesian Analysis*, Second Edition, 1985. Springer, ISBN: 0-387-96098-8.

149. Duda, R.O. and P.E. Hart, *Pattern classification and scene analysis*, First Edition. Vol. 3. 1973. Jon Wiley & Sons, Inc., ISBN: 0-471-22361-1.

150. Meyn, S.P., *Control techniques for complex networks*, Edition, 2008. Cambridge University Press, ISBN: 0521884411.

151. Wang, P., H. Jiang, and W. Zhuang, Capacity improvement and analysis for voice/data traffic over WLANs. *Wireless Communications, IEEE Transactions on*, 2007. 6(4): pp. 1530-1541.

152. Doucet, A., N. De Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, First Edition, 2001. Springer New York, ISBN: 978-1-4419-2887-0.

153. Leffler, S. "*The MadWifi Project*". 2009. [Online]: Available at: http://madwifi-project.org/. (Last Accessed: 30/4/2014).

154. "*The Ath5k Project*". 2010. [Online]: Available at: http://wireless.kernel.org/en/users/Drivers/ath5k#A2010-04-07. (Last Accessed: 30/4/2014).

155. "*The Ath9k Project*". 2013. [Online]: Available at: http://wireless.kernel.org/en/users/Drivers/ath9k. (Last Accessed: 30/4/2014).

156. "*The Ath10k Project*". 2013. [Online]: Available at: http://wireless.kernel.org/en/users/Drivers/ath10k. (Last Accessed: 30/4/2014).

157. Jacobson, V. and S. McCanne. "*Libpcap: Packet Capture Library*". Lawrence Berkeley Laboratory, Berkeley, CA 2009. [Online]: Available at: http://www.tcpdump.org/. (Last Accessed: 30/4/2014).

158. Benkic, K., et al. Using RSSI Value for Distance Estimation in Wireless Sensor Networks based on ZigBee. in *IWSSIP 2008, Fifteenth International Conference on Systems, Signals and Image Processing*. Bratislava, Slovak Republic. 25 - 28 June, 2008.

159. Koksal, C.E., *Quality-Aware Routing Metrics in Wireless Mesh Networks*, in *Wireless Mesh Networks*, 2007, Springer Science+Business Media, LLC, ISBN: 978-0-387-68838-1. pp. 227-243.

160. Park, V.D. and M.S. Corson. A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks. in *INFOCOM 1997, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies*. Kobe, Japan. 7 - 11 April, 1997.

161. Adya, A., et al. A Multi-Radio Unification Protocol for IEEE 802.11 Wireless Networks. in *BROADNETS 2004, First International Conference on Broadband Networks* San Jose, CA, U.S.A. 25 - 29 October, 2004.

162. Koksal, C.E. and H. Balakrishnan, Quality-Aware Routing Metrics for Time-varying Wireless Mesh Networks. *IEEE Journal on Selected Areas in Communications*, 2006. 24(11): pp. 1984-1994.

Appendices

Appendix A

Table A.1 The Parameters Set Up of IEEE 802.11b Networks

Parameters	Set Up
Radio Propagation Model	Two-Ray Ground
Antenna Model	Omni-Directional Antenna
Maximum Packet in Buffer	50
Routing Protocol	DSDV
Maximum UDP Packet Size	1500 Bytes
CW_{min}	31
CW_{max}	1023
Short Retry Limit	7
Long Retry Limit	4
Slot Time	20 µs
SIFS	10 µs
DIFS	50 µs
Short Preamble	72 Bits
PLCP Header Length	48 Bits
PLCP Data Rate	2Mbps (for short preamble)
Line Rate	11Mbps
Basic Rate	1 Mbps

RTS Threshold	3000 Bytes
Transmit Power	1.561611e-1 W
Reception Range	50 meters
Interference Range	50 meters
Channel Frequency	2.472 GHz

Appendix B

Table B.1 The Data Values of all Probabilities of the Number of Observable Neighbours *M* under Different *N*

M \ N	2	3	4	5	6	7	8	9	10
1	0.438	0.211	0.1089	0.0584	0.03281	0.018769	0.010989	0.006462	0.0039
2	0.562	0.452	0.307	0.2019	0.1302	0.084134	0.054397	0.035067	0.022763
3		0.337	0.37	0.3121	0.2436	0.181058	0.130919	0.094185	0.066567
4			0.214	0.2842	0.2787	0.245023	0.203155	0.161912	0.126128
5				0.1432	0.2151	0.234452	0.225894	0.203451	0.175407
6					0.0997	0.164196	0.193161	0.198452	0.191057
7						0.072368	0.126733	0.157944	0.170786
8							0.054156	0.100758	0.129869
9								0.041769	0.08047
10									0.032963

Appendix C

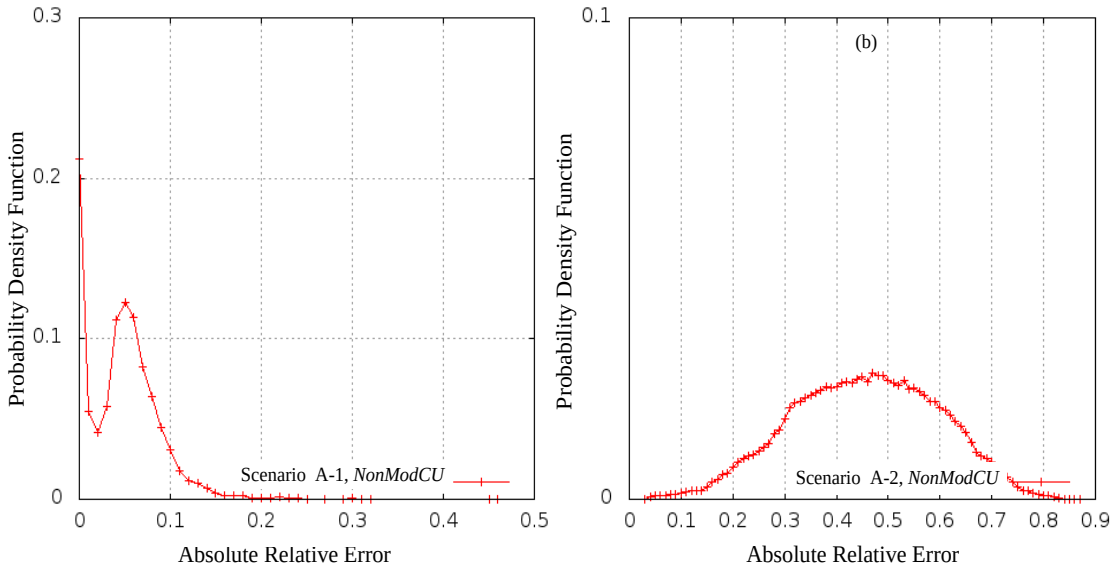

Figure C.1: The PDF of the ARE for *NonModCU* in (a) Scenario A-1 and (b) Scenario A-2

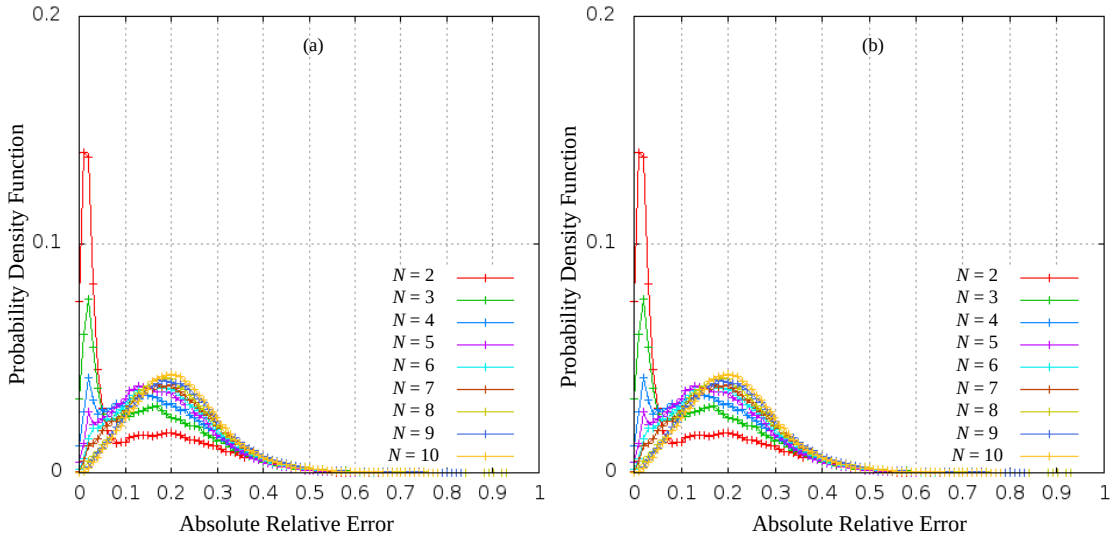

Figure C.2: The PDF of the *ARE* of *NonModCU* under Different *N* Scenarios with (a) Lower Traffic Load (a) Higher Traffic Load

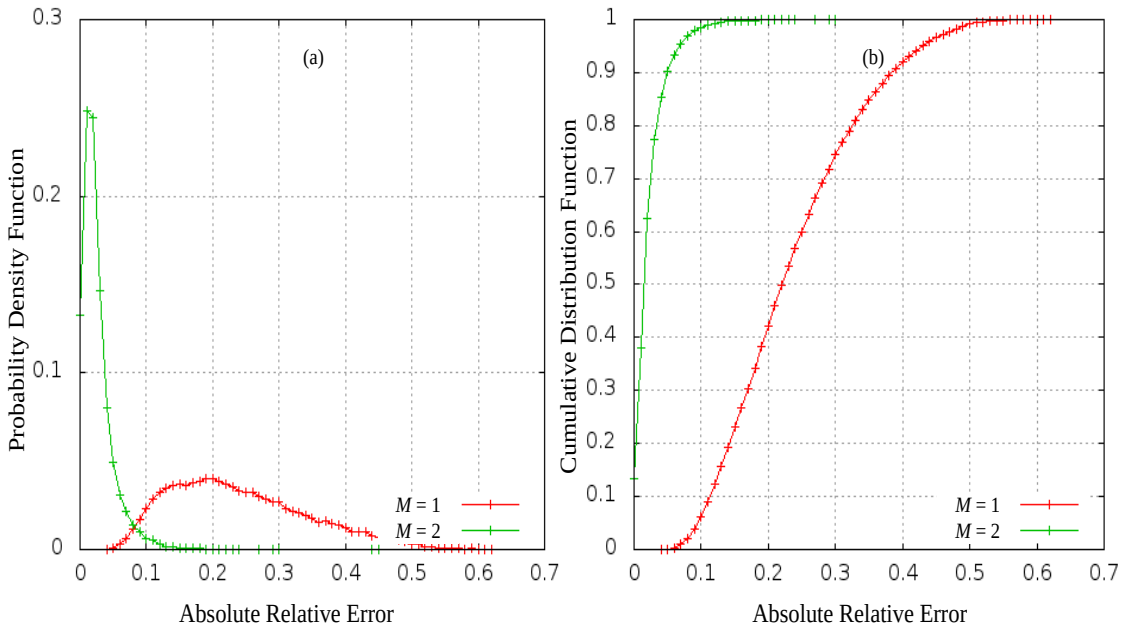

Figure C.3: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 2$

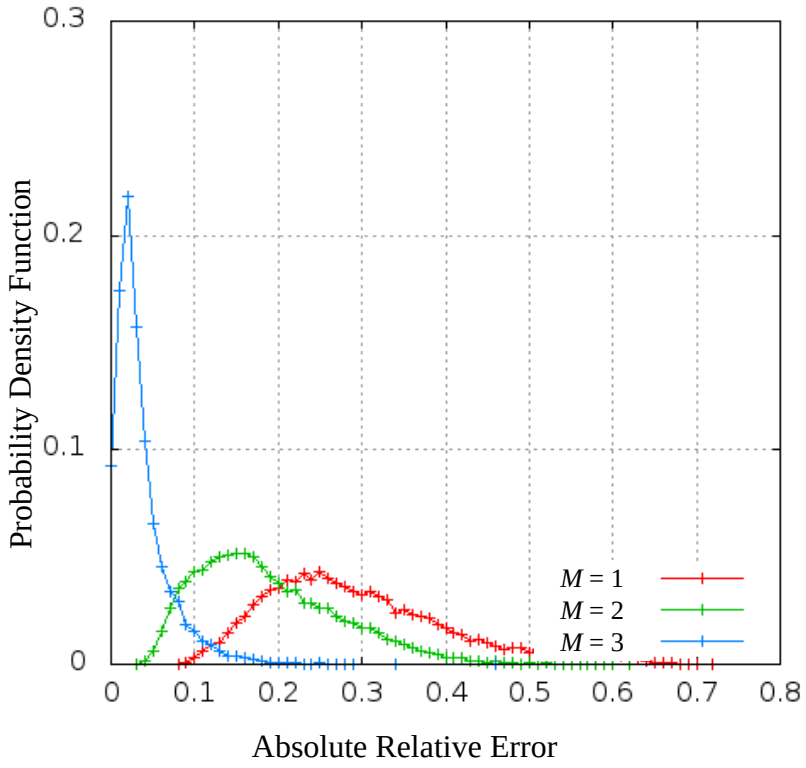

Figure C.4: The PDF of the *ARE* of *NonModCU* where $N = 3$

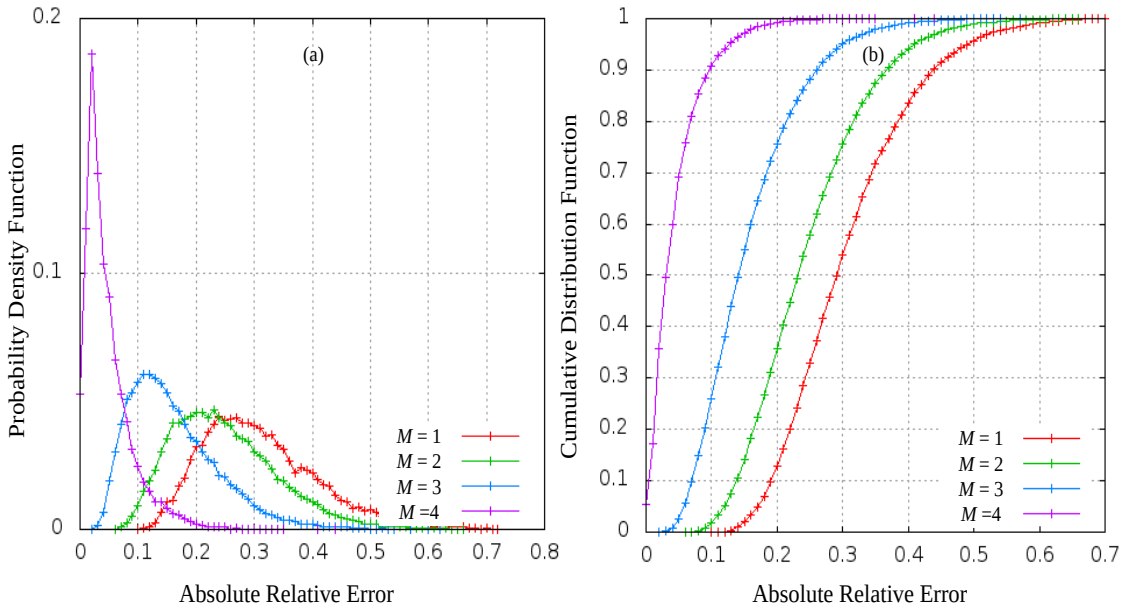

Figure C.5: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 4$

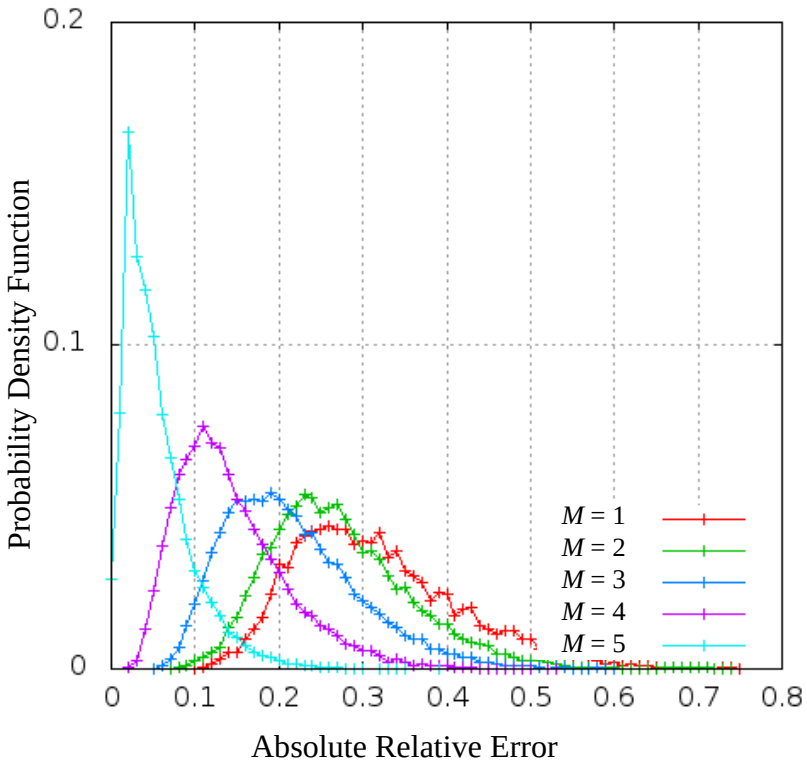

Figure C.6: The PDF of the *ARE* of *NonModCU* where $N = 5$

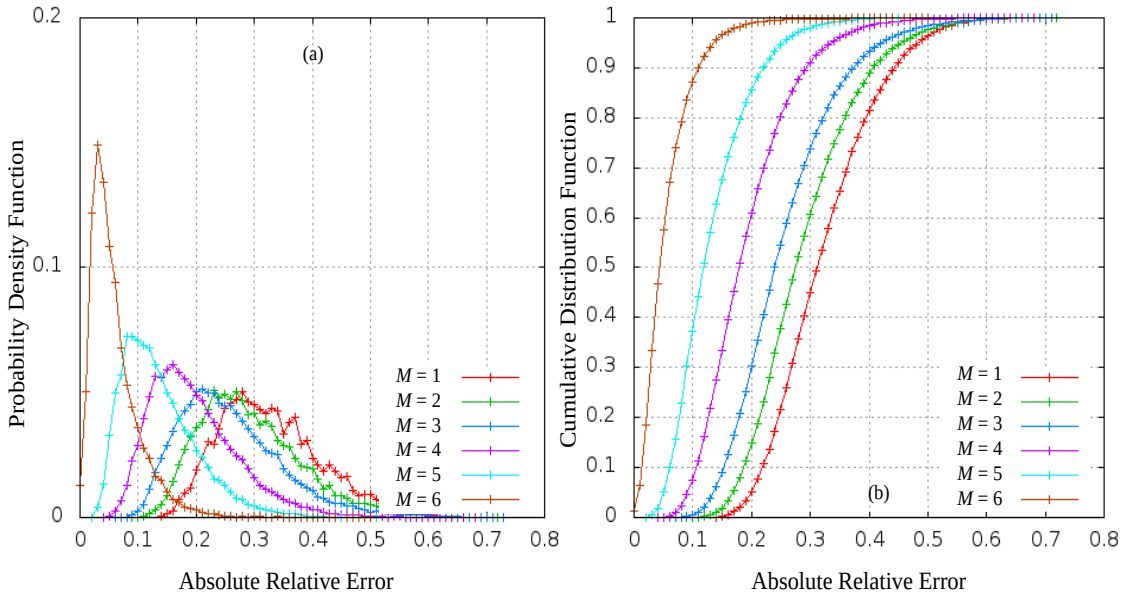

Figure C.7: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 6$

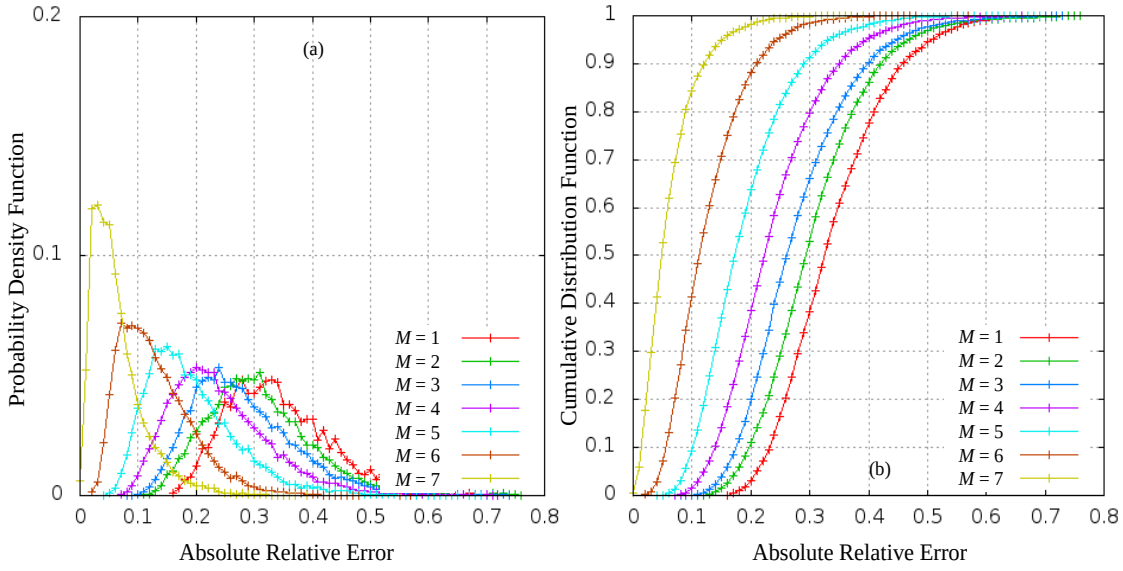

Figure C.8: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 7$

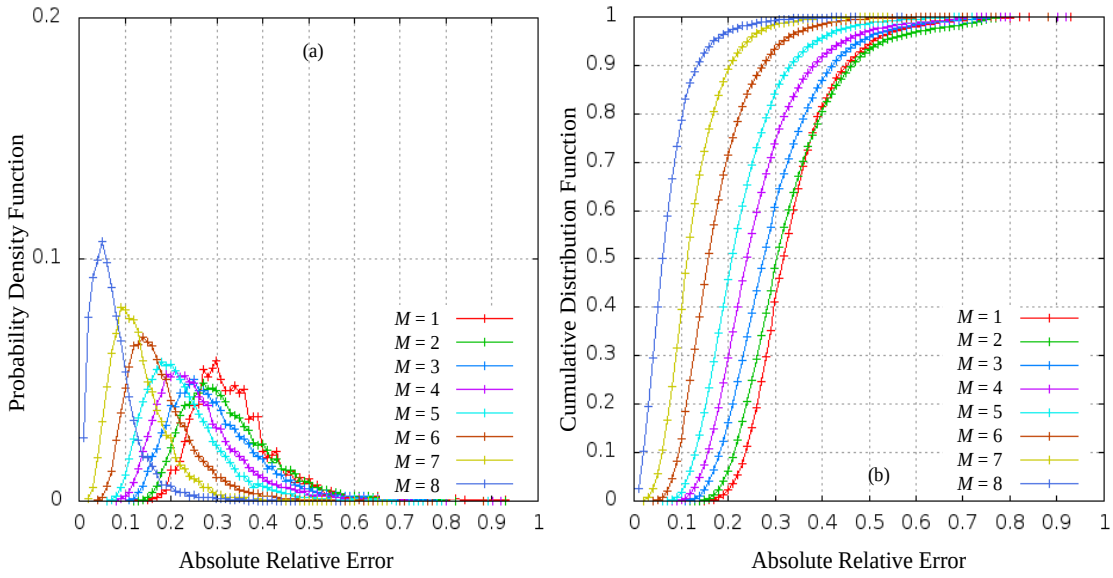

Figure C.9: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 8$

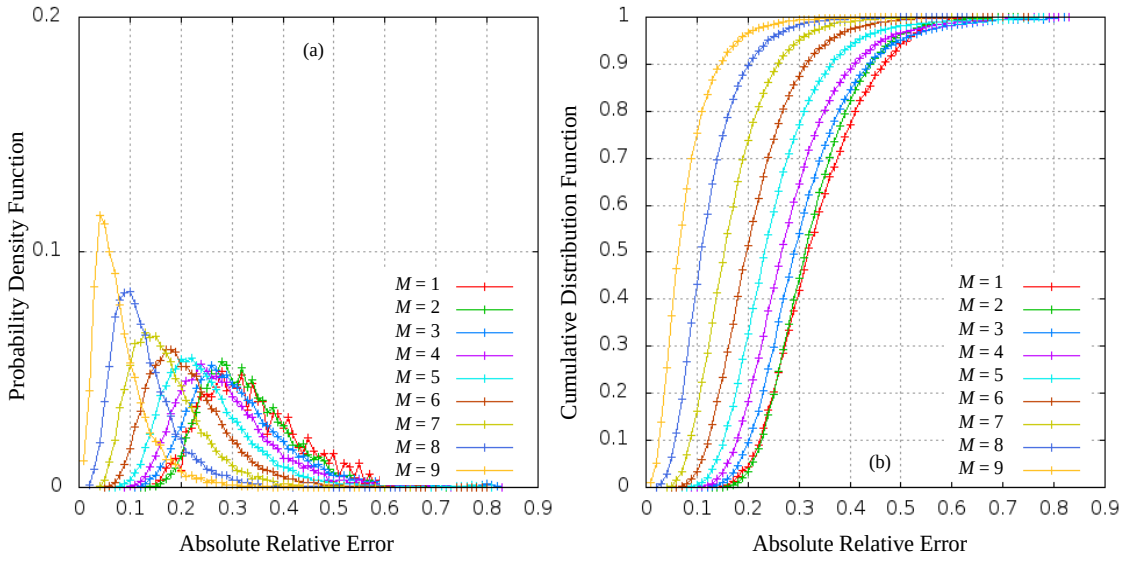

Figure C.10: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 9$

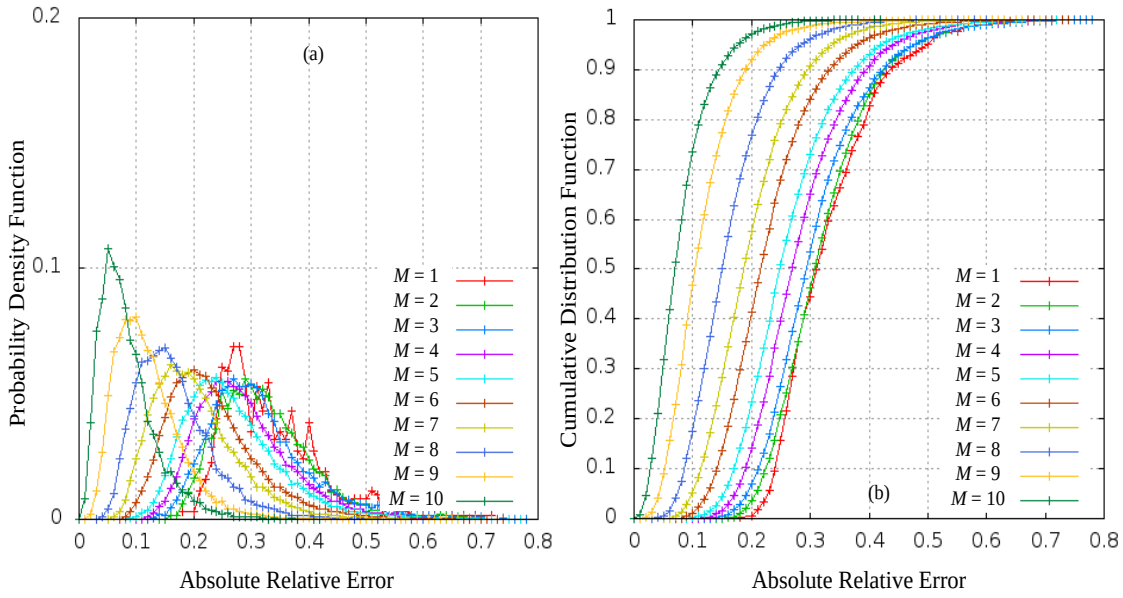

Figure C.11: The (a) PDF and (b) CDF of the *ARE* of *NonModCU* where $N = 10$

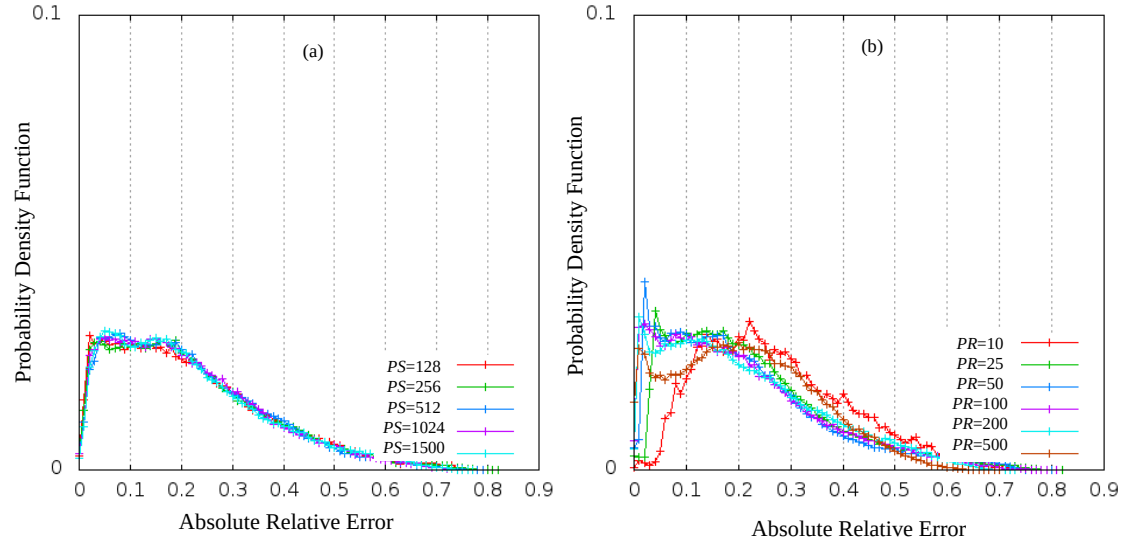

Figure C.12: The PDF of the *ARE* of *NonModCU* for Different (a) Packet Sizes (b) Packet Rates of Traffic Load of the Observed Node

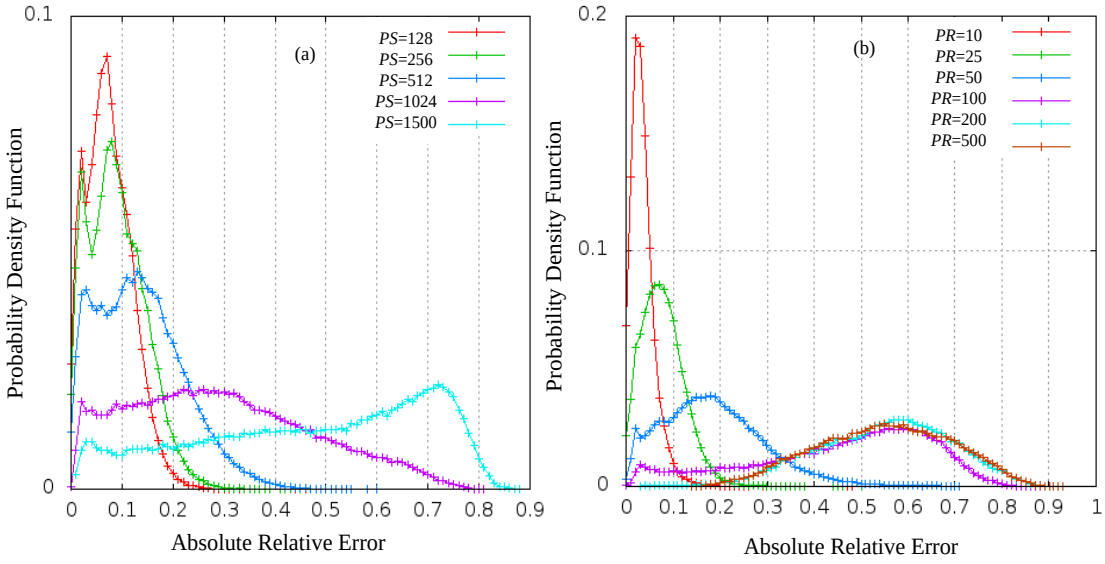

Figure C.13: The PDF of the *ARE* of *NonModCU* for Different (a) Packet Sizes (b) Packet Rates of Neighbour Traffic Load

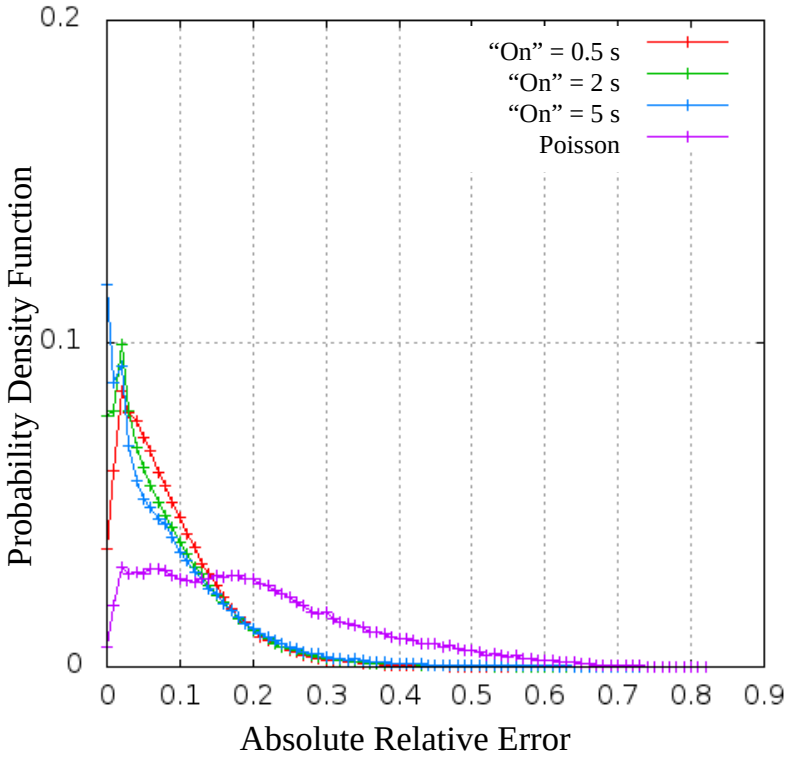

Figure C.14: The PDF of the *ARE* of *NonModCU* under On-Off traffic

Appendix D

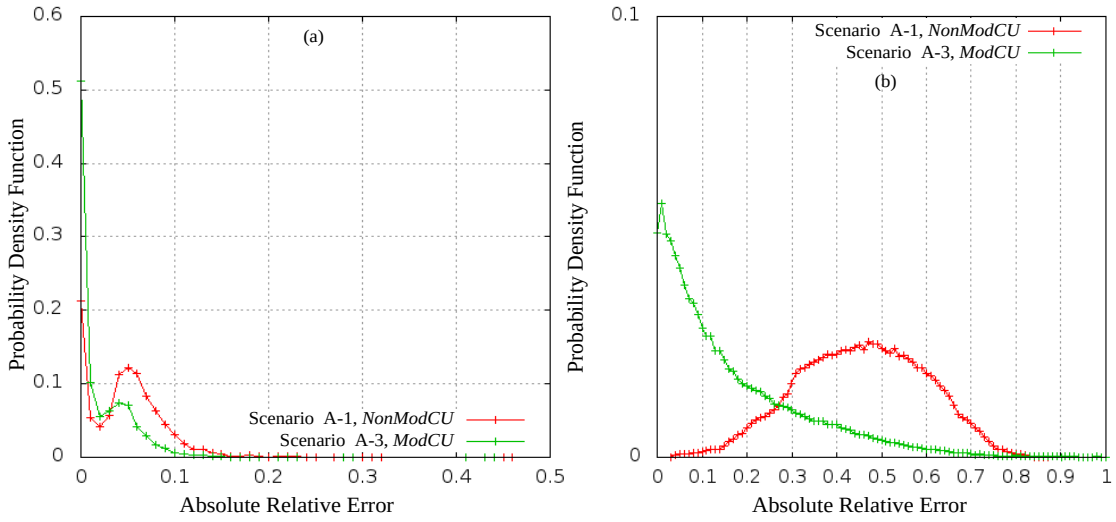

Figure D.1: PDF of *ARE* for *ModCU* in (a) Scenario A-3 and (b) Scenario A-4

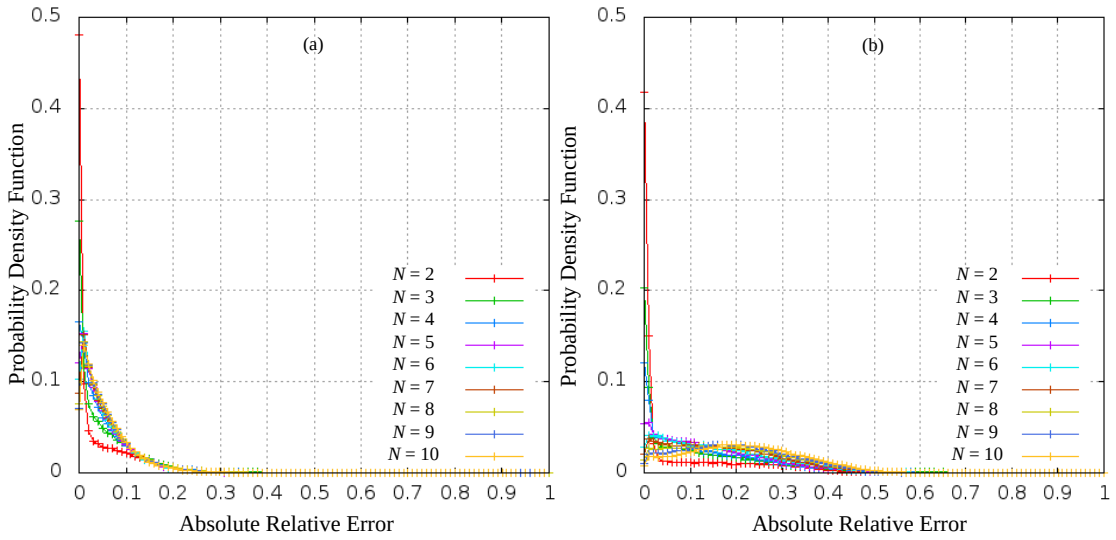

Figure D.2: The PDF of the *ARE* of *ModCU* under Different *N* Scenarios with (a) Lower Traffic Load (a) Higher Traffic Load

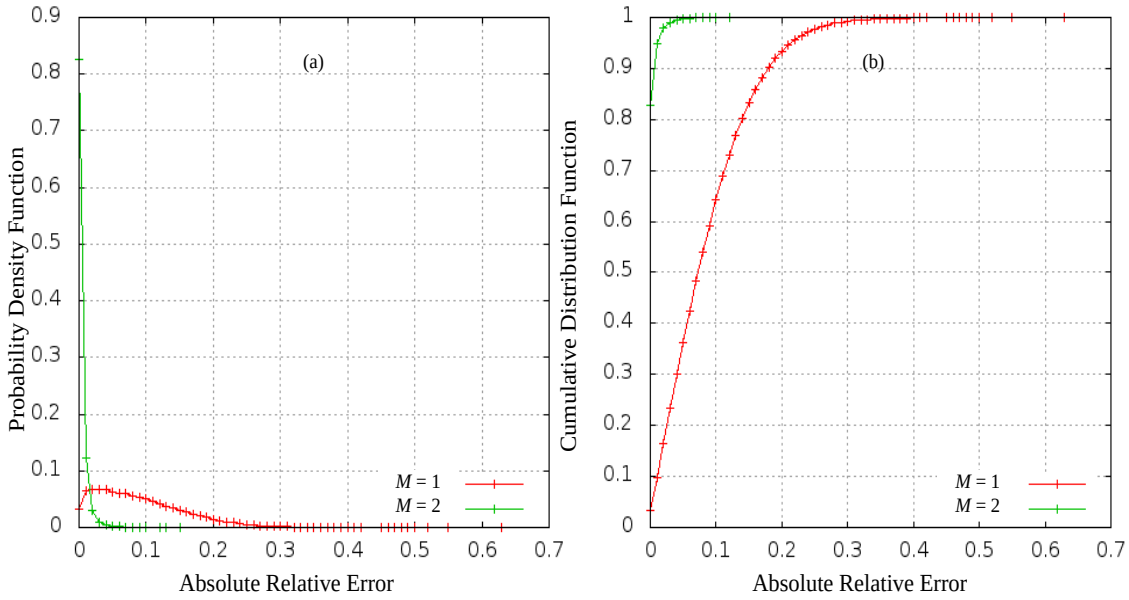

Figure D.3: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 2$

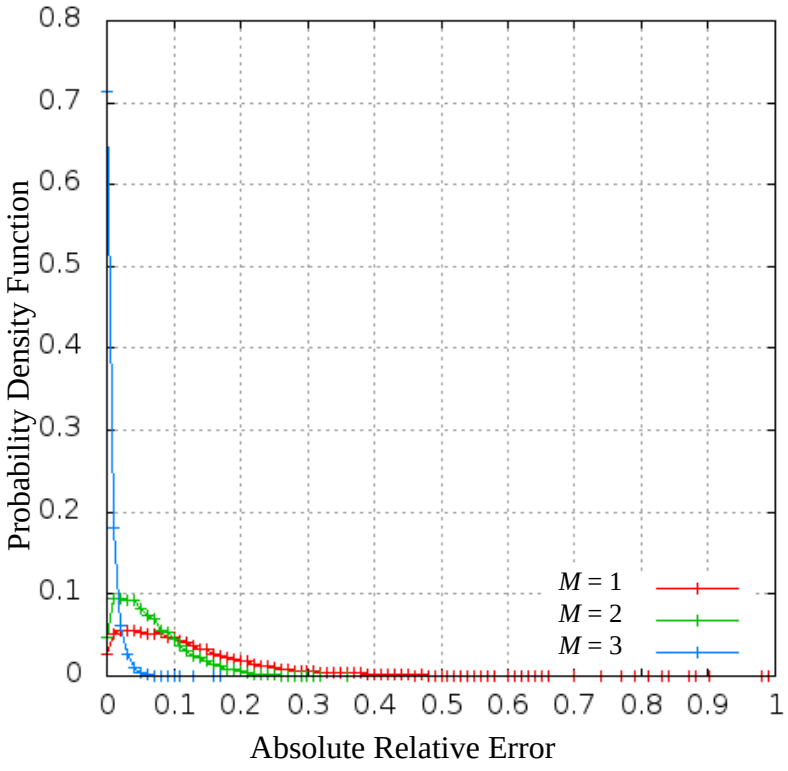

Figure D.4: The PDF of the *ARE* of *ModCU* where $N = 3$

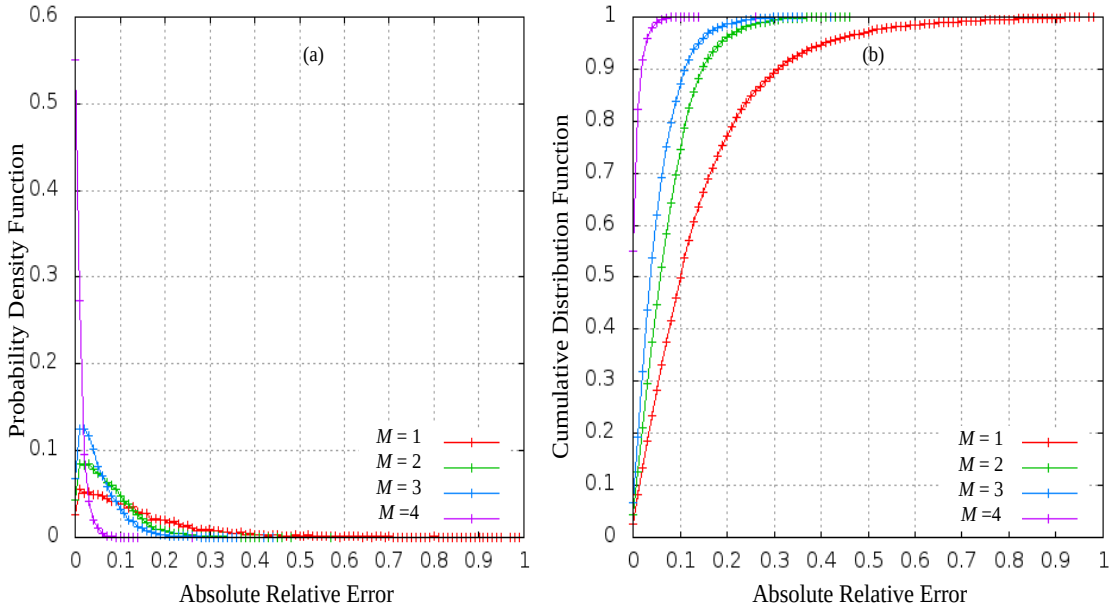

Figure D.5: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 4$

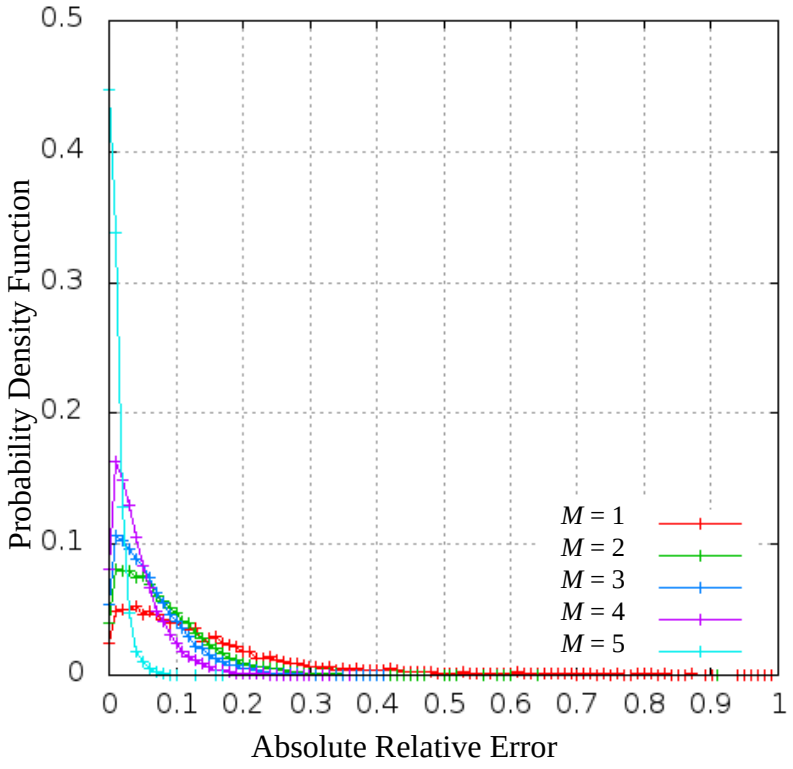

Figure D.6: The PDF of the *ARE* of *ModCU* where $N = 5$

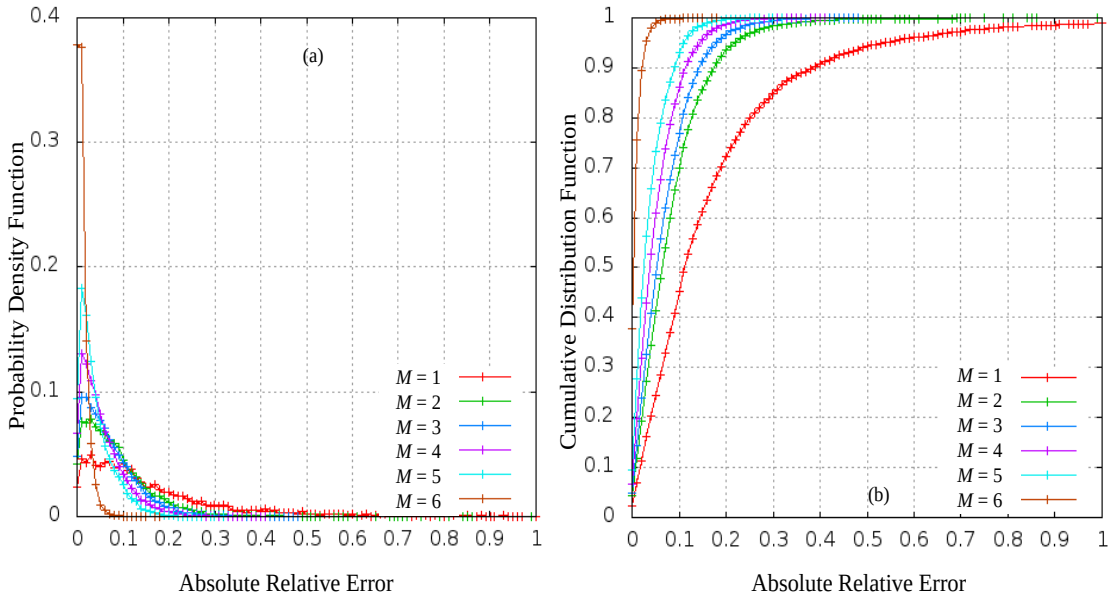

Figure D.7: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 6$

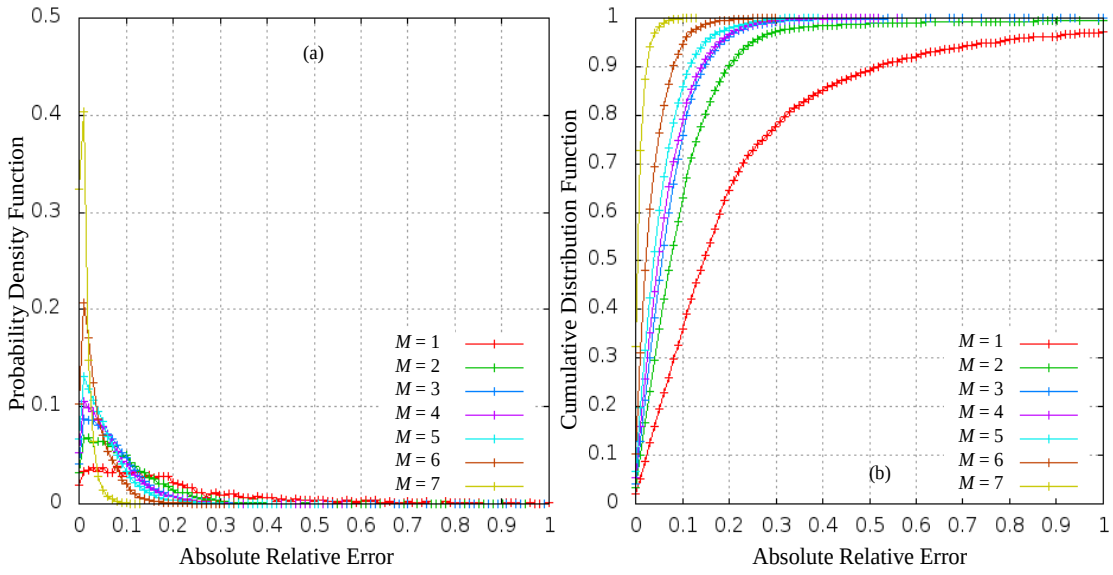

Figure D.8: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 7$

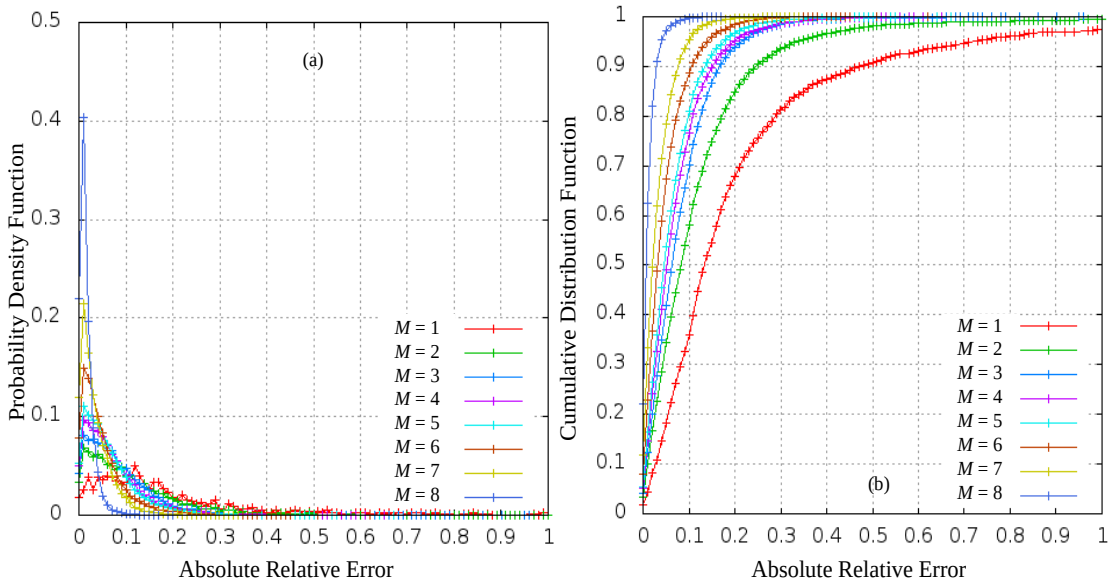

Figure D.9: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 8$

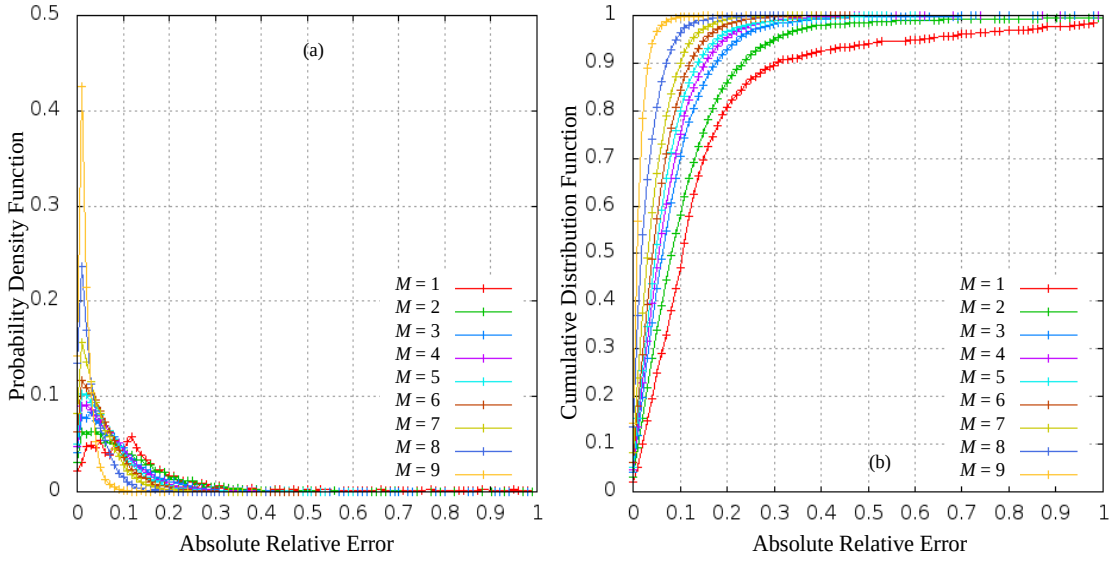

Figure D.10: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 9$

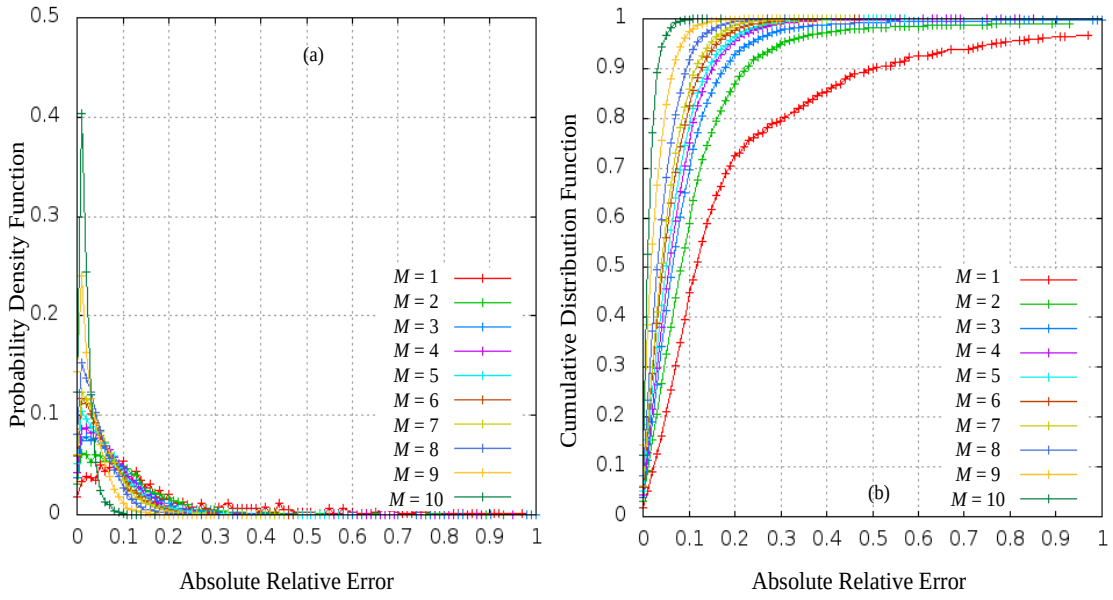

Figure D.11: The (a) PDF and (b) CDF of the *ARE* of *ModCU* where $N = 10$

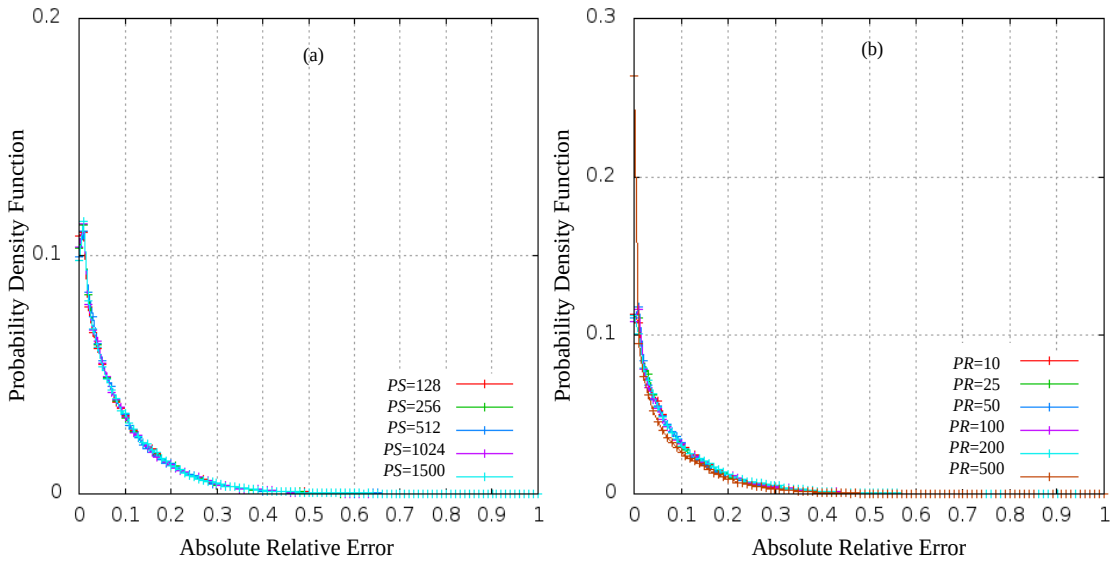

Figure D.12: The PDF of the *ARE* of *ModCU* for Different (a) Packet Sizes (b) Packet Rates of Traffic Load of the Observed Node

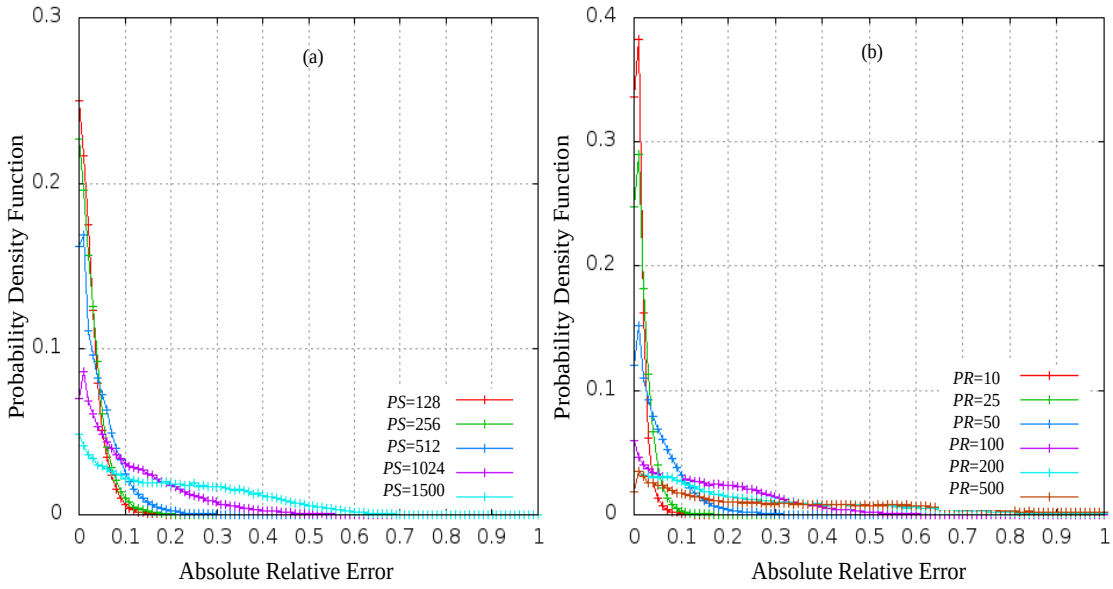

Figure D.13: The PDF of the *ARE* of *ModCU* for Different (a) Packet Sizes (b) Packet Rates of Neighbour Traffic Load

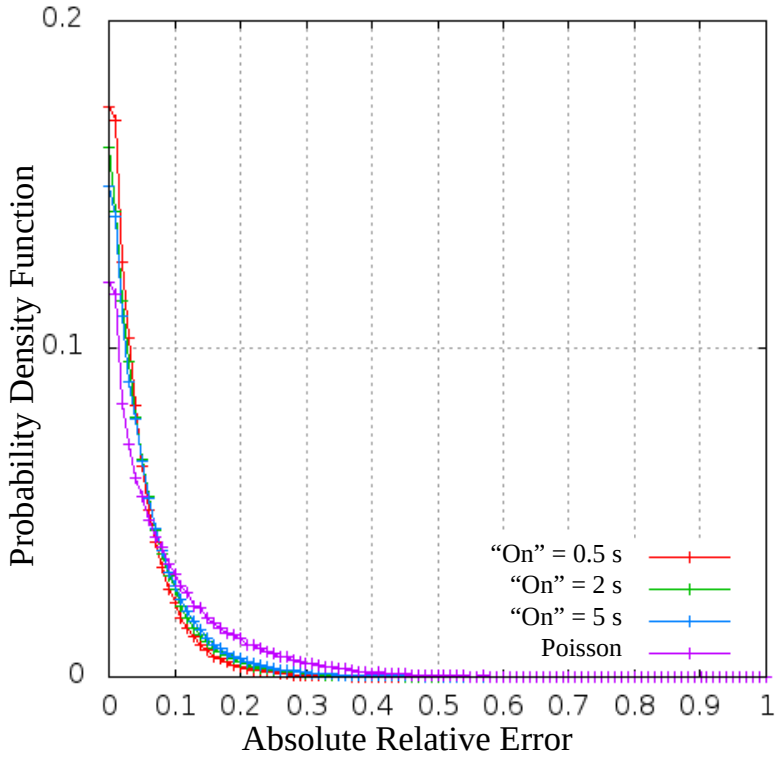

Figure D.14: The PDF of the *ARE* of *ModCU* under On-Off traffic

Appendix E

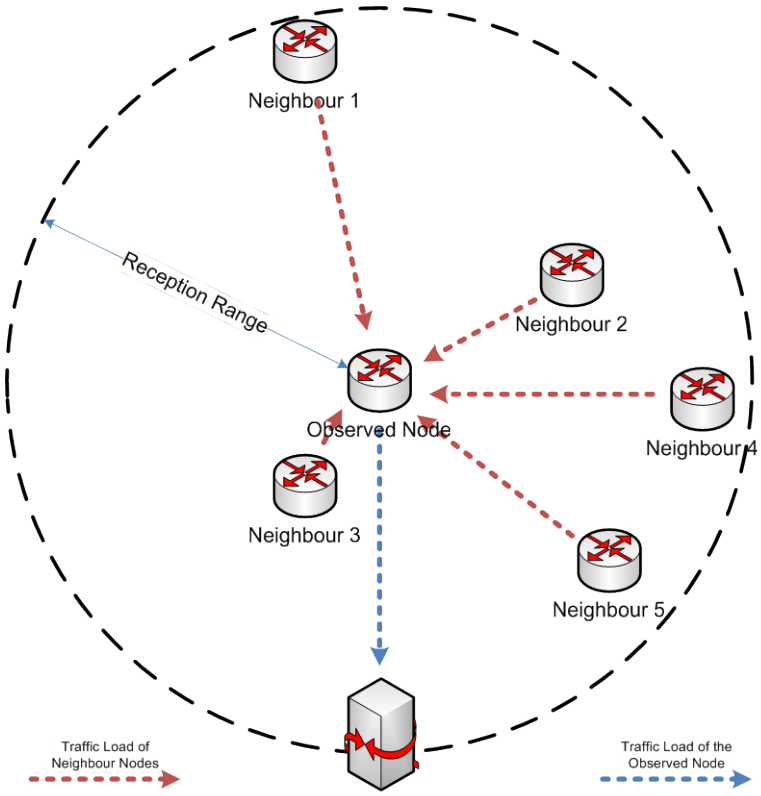

Figure E.1: The Example Topology

Table E.1 The Parameters of Traffic Load of the Example Topology

	N	*M*	Packet Size (bytes)	Packet Rate (pps)	Traffic Type
The Observed Node	5	5	1101	191	Poisson
Neighbour 1	5	2	239	94	Poisson
Neighbour 2	5	5	1209	115	Poisson
Neighbour 3	5	3	542	97	Poisson
Neighbour 4	5	3	1097	114	Poisson
Neighbour 5	5	4	85	84	Poisson

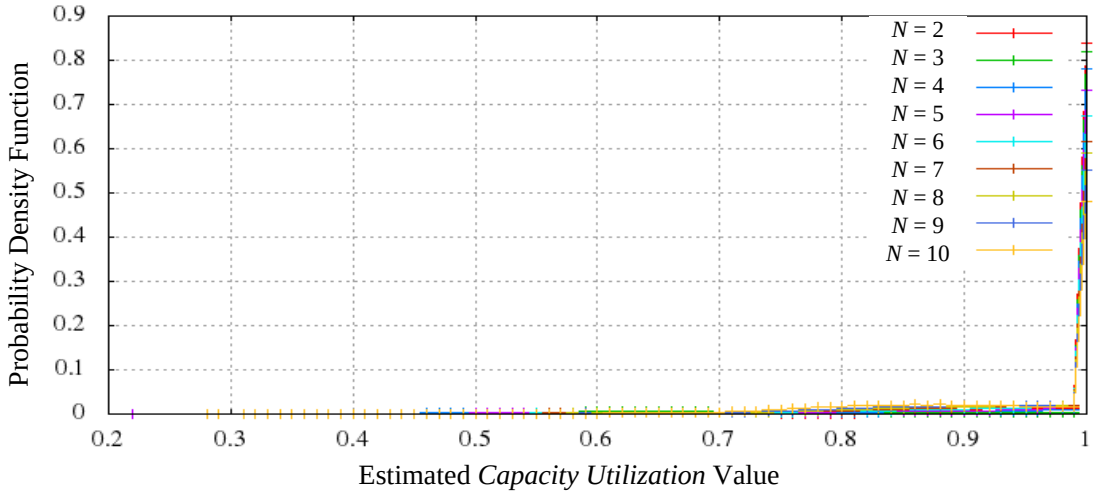

Figure E.2: The PDF of *ModCU* Measurement under Scenario D-1

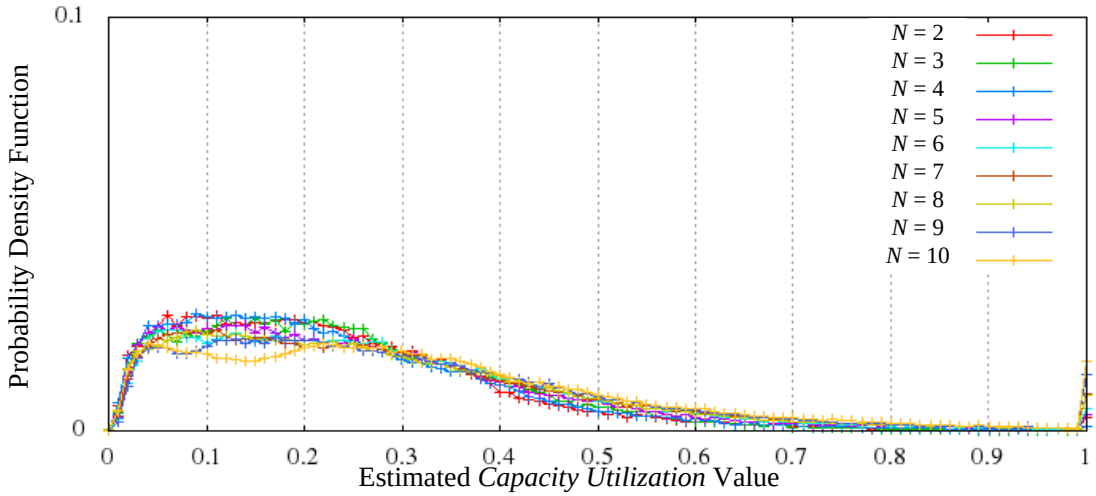

Figure E.3: The PDF of *ModCU* Measurement under Scenario D-2