A QoS Enabling Queuing Scheme for Fourth Generation Wireless Access Networks

Robert Otieno Achieng

This thesis is submitted in partial fulfillment of the academic requirements
for the degree of
Master of Science in Electrical Engineering
in the Faculty of Engineering and The Built Environment
University of Cape Town
November 2006

As the candidate's supervisor, I have approved this dissertation for submission.

Professor H. Anthony Chan

Signed: ______________________

Date: ________________________

Declaration

I declare that this thesis is my own work. Where collaboration with other people has taken place, or material generated by other researchers is included, the parties and/or materials are indicated in the acknowledgements or are explicitly stated with references as appropriate.

This work is being submitted for the Master of Science in Electrical Engineering at the University of Cape Town. It has not been submitted to any other university for any other degree or examination.

Robert Otieno Achieng

28th November 2006

Abstract

The rapid growth of wireless networking over the last decades has led to the emergence of a wireless system known as the Fourth Generation network (4G wireless networks). 4G networks will comprise heterogeneous wireless access technologies, e.g. Wireless Local Area Networks (WLAN), UMTS Terrestrial Radio Access Network (UTRAN), and WiMAX. Moreover, the networks will have to handle diverse types of multimedia traffic across the different wireless technologies. To realize these objectives, 4G systems will have to overcome some challenges. The challenges include support for QoS provisioning, mobility management and network security. Support for QoS provisioning entails network selection. Network selection is a key element of QoS support in 4G wireless networks; consequently, a lot of research effort is being spent in addressing it. Essentially, network selection is the process of choosing one network from a number of available networks in a heterogeneous system to serve a connection request by a mobile user.

This research proposes a scheme to accomplish the task of network selection; this is achieved by enhancing existing QoS provisioning approaches. The scheme models the radio access network as a network of queuing nodes. With the model, the link layer QoS statistics of user traffic in each available path through the network is determined. The author postulates that the statistics indicate the QoS capabilities of the network and can therefore be used to select the best network to serve the mobile user.

In the study, queuing theory is used to develop analytical expressions for the media access scheduling delay and throughput for real-time voice traffic and non real-time data traffic in both UTRAN and IEEE 802.11e WLAN. Performance evaluation experiments are performed using both Matlab and OPNET software. The results indicate that, notwithstanding some discrepancies between the analytical and simulation results, the proposed scheme represent a credible approach for solving the network selection problem in 4G networks.

Acknowledgements

Postgraduate research is a tremendous enterprise, and the final output of that effort can by no means be the product of one individual. Many people contributed to this thesis, in small and big ways, in subtle and obvious aspects. I might not be able to identify all of them, but I'm truly grateful to all.

Some people deserve special mention, though

Prof Chan, my supervisor. He always wanted the best in me, pushing me to the limits always. Of course, he himself also gave out all that he could to make sure that I successfully complete the program. Thanks a lot for all that you did

George, Dayo, and Eugene; fellow students in the CRG lab. You've been always ready with positive contributions and warm personal assistance when necessary.

All members of the CRG lab, family back home in Kenya and many more.

Finally, to God, for the gift of life.

Table of Contents

List of Figures

List of Tables

Glossary

Media Access Control (MAC) Delay. The channel access delay associated with the multiple access protocol of a channel. Normally, this is the component of channel access delay that is variable. It arises from the queuing and scheduling of packets by the node.

Medium Access Control (MAC) Throughput: The throughput associated with the multiple access protocol of a channel. MAC throughput is measure of the transmission efficiency of the wireless channel.

Nominal Data Rate (NDR): When determining the number of users in a UTRAN network, it is assumed that all users operate at the same data rate. This hypothetical data rate is what is known as nominal data rate.

Normalized Cell Load (NCL). This is the ratio of the amount of traffic in a network at a particular time to the maximum allowable traffic load in the network. NCL gives a measure of how much traffic a network can accept.

Pole Capacity (PC). Pole capacity is the maximum number of users in a wireless cell. This figured is arrived at after fixing a minimum performance level for a user.

QoS Enabled station (QSTA): A station in EDCA WLAN that can support QoS. A station becomes a QSTA after its channel access categories has been enabled.

1 Introduction.

There has been a phenomenal growth in wireless networking over the last two decades. The growth has been attributed to rapid advances in technology, competition in the business environment, and socio-economic developments that have changed how people live and work [27]. Consequent to this growth, a new paradigm in telecommunication systems known as the Next Generation Network (NGN) is emerging. A clear and fixed definition of NGN is yet to be developed, but there's universal consensus that it will comprise the following [38]:

i)A layered architecture, consisting of functionally differentiated planes, namely the Data Plane (or Transport plane), the Control plane, the Application plane, and the Management plane.

ii)Open and standardized interfaces to facilitate inter-working of systems across the planes.

iii)Ability to support multi media traffic for a variety of services

1.1 The Transport Plane of Next Generation Networks

The Data plane of the layered architecture is responsible for the transfer of user traffic across the network. Functions that are performed by the plane include data transmission, multiplexing, switching, and routing. The Control plane, on the other hand, is responsible for the control of data transfer in the Data plane. Elements of Control plane functions include definition/realization of bearer capabilities, signaling between network elements, and also the facilitation of inter-working between systems. The creation and deployment of services are handled by the Application plane. Finally, support for Quality of Service (QoS) management, provision of network security, and management of the network is to be accomplished by the Management plane [1].

NGN are expected to support multi-media services. Multimedia services combine two or more media components (e.g. voice, audio, data, video, pictures) within one call. Broadly speaking, the services to be supported are real-time (RT) services and non real-time services. Examples of real-time services are conversational voice and interactive applications like online gaming. Non real-time services include applications like internet access and email. Support of services in 4G networks will be characterized by the following attributes [32]

i)Tele-presence: This is a virtual reality situation where a user gets the illusion of being at a particular place, e.g. teleconference participants get the illusion of being together in the same room. Tele-presence requires high bandwidth and stringent delay constraints on the transport network.

ii)Information access: users of 4G services will expect to have instantaneous access of large volumes of data.

iii)Location based services: It is envisioned that 4G systems will be able to locate a mobile terminal with a high degree of accuracy (i.e. within 10m). This will enable service providers to offer services appropriate to the user's location e.g. notification of presence of a restaurant within the vicinity, emergency services etc.

iv)Security: Secure services will be an indispensable feature of 4G systems. Data integrity is a crucial factor in services like electronic banking, mobile commerce etc.

1.2 Requirements for the Data Plane NGN

The Data plane consists of three functional parts: the access, the edge, and the core networks. The access network in 4G systems will be a heterogeneous network with a variety of wireless and wire-line access technologies. Examples of wireless access technologies in 4G networks include Wireless Local Area Networks (WLAN), UMTS Terrestrial Radio Access Network (UTRAN) and Worldwide Interoperability for Microwave Access (WiMAX). However, in contrast to the present situation where the access technologies operate independently, in 4G systems the technologies will operate in concert. A wireless access network will typically consist of mobile terminals, a wireless access point (or base station), and an access router.

The edge network will be a system of routers that manage ingress into the core network through admission control, bandwidth management and QoS provisioning. Traditionally, each access network has been performing its admission control and QoS provisioning functions independently. However, because of the heterogeneity of the access network, these functions are to be centralized at the edge.

The core network will be a high-speed optical switching and transport system. The entire Data plane will be packet switched. Fig.1 shows network architecture of NGN.

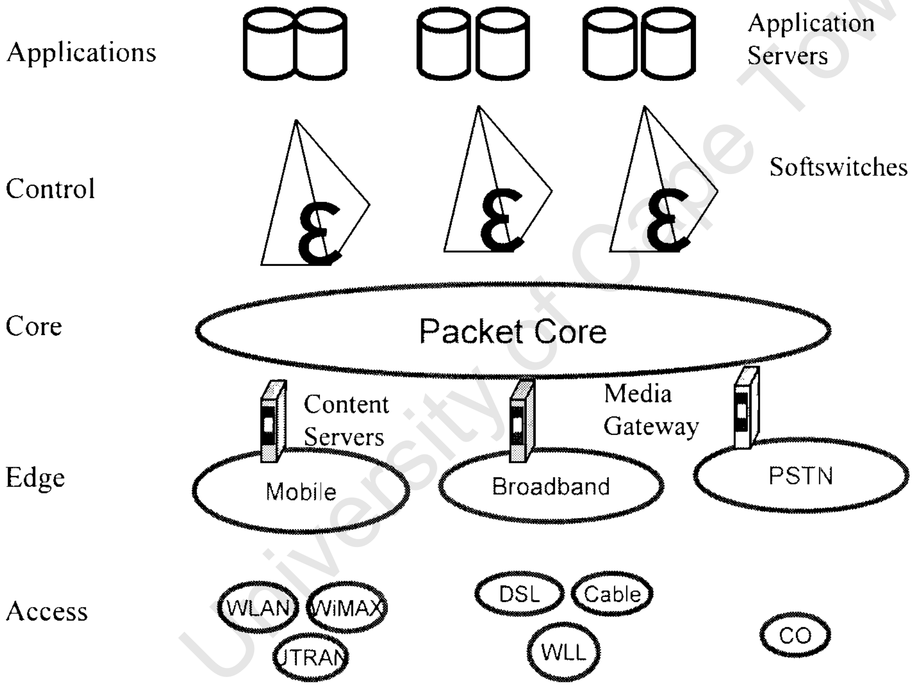

Fig. 1.1 NGN Architecture

In order that 4G systems deliver the proposed services satisfactorily, some challenges have to be overcome. From a user perspective, the challenges associated with the Data plane fall into three main categories, namely Mobility management, QoS management, and provision of Network Security [9]

1.2.1 Mobility Management

Mobility management is the process of keeping track of the location and identity of a mobile user for the purpose of delivering appropriate services to it. There are two broad Mobility Management objectives:

i)Location management (registration and paging for wireless access and user profile storage/lookup for wire-line access).

ii)Handoff management. Two types of handoff can be identified: horizontal handoff and vertical handoff. Horizontal handoff represents movement of a mobile user across wireless access points of the same access technology e.g. movement across cells of a UTRAN network. Vertical handoff, on the other hand, is handoff across access points of different access technologies e.g. movement from a UTRAN base station to a WLAN access point.

Below are some of the problems that have to be overcome with respect to Mobility management.

i)Traditional handoff management protocols assume homogeneous networks; consequently, vertical handoff protocols have to be developed.

ii)Current IP-based mobility protocols use the address to represent both the identity and location of the user. In 4G systems where there are such novel concepts as session and service mobility, there is need to use the IP address exclusively to represent the location of the mobile user.

1.2.2 QoS Management

Quality of Service (QoS) is defined as the "collective effect of service performance which determines the degree of satisfaction of a user of the particular service" [18, 17]. According to this definition, three types of QoS are identified

i)Intrinsic or Network Performance QoS. This is the QoS stemming from technical aspects; it is determined by transport network design and provisioning of network resources. This is achieved through the design and selection of appropriate transport protocols and QoS assurance mechanisms and related values of parameters [14]. Examples of intrinsic QoS performance parameters are delay, throughput, jitter, and packet loss. Intrinsic QoS is evaluated by considering measured values of the performance parameters against expected values.

ii)Perceived QoS. Perceived QoS reflects a customer's experience of using a service. Evaluation criteria for perceived QoS are observed performance against expected performance. In a network operator environment, the design and implementation of perceived QoS is done by the marketing department through product development and deployment.

iii)Assessed QoS. This depends on perceived QoS, service price (value for money!) and service provider customer care establishment. The evaluation criteria for Assessed QoS is the amount of demand for the service.

For purposes of this research, the author considers only QoS Management with respect to only Intrinsic or Network Performance QoS.

The objective of QoS Management is to ensure that the measured values of a user's intrinsic QoS performance parameters comply with expected values. The following are the mechanisms by which the network realizes the QoS management objectives.

i)Admission Control: Admission control is the procedure for admitting connections into the network. From a network operator perspective, admission control is a tool for resource management. For a user's perspective however, admission control is used to ensure that a user is only admitted if its QoS requirements are likely to be met.

ii)Network Selection. Network selection is the procedure for selecting the network to serve a mobile user; and it is a uniquely 4G system problem. In a heterogeneous network with more than access technology, each of which could be capable of serving a mobile user, network selection schemes resolve the problem by picking on one of the available networks. QoS parameters are valid bases for network selection schemes.

iii)Power control. The error performance of a shared wireless channel degrades with increase in the amount of radiofrequency power in the channel. Power control, which is the judicious and methodical control of radiofrequency power in a wireless channel, is often used as a QoS management tool in wireless networks.

There are a number of challenges that QoS management protocols in 4G systems have to overcome. Some of the challenges are mentioned below [28].

i)Minimize overhead (signaling traffic). Many network elements will be involved in service delivery; signaling among them should, however, be kept minimal.

ii)Differentiate among the many types of traffic that will characterize 4G systems.

iii)Network selection, i.e. be able to select the best network to serve a mobile user.

iv)User mobility (Handoff has to be QoS aware).

v)Scalability: protocols in 4G networks will be designed to accommodate increase in the number and diversity of network elements.

1.2.3 Provision of Network Security

Provision of network security will be mandatory in 4G systems to ensure that only the right users get authorized services without compromising on service delivery. Generally, network security has four components:

i)Authentication: This has to do with the correct mutual identification of all the parties to a connection.

ii)Confidentiality. Confidentiality ensures that only the genuine parties to a communication gain access to the content of the communication session.

iii)Integrity. Integrity ensures that information is not modified on transit. It also ensures that genuine information can not be repudiated by either the sender or the receiver.

iv)Authorization (QoS request authorization etc). Authorization procedures work in concert with Authentication to validate access to services.

v)Accounting. Access to and usage of services have to be recorded and charged accordingly.

Security challenges normally reflect themselves in the form of security threats/attacks. The availability of ubiquitous access and open interfaces increases the vulnerability of 4G systems. Thus, secure systems have to be built into 4G networks in order to safeguard against the attacks.

To recapitulate, one should note that, so far in this chapter the author has managed to present an overview of 4G networks by highlighting its functional architecture and the services that it is expected to deliver. In addition, the challenges that are to overcome in the Data plane have mentioned. From this point, therefore, we proceed to present the research problem statement, the objectives of the research and the scope of works.

1.3 Objective of Research: Developing a Queuing Theoretic Scheme for Network Selection in 4G Networks.

From the discussions in section 1.3, it is noted that Network Selection will be a key component of QoS provisioning in 4G networks. Recall that network selection in 4G networks is the process of selecting the access network to serve a mobile user. Network selection schemes/algorithms are used during connection request by a new user and during handoff. Considerable effort in research on wireless networking is spent on developing network selection algorithms. *The broad objective of the research is to develop a network selection scheme for 4G wireless access networks using queuing theory.*

The scope of the research will thus be as follows:

i)Present an exposition of the network selection problem and a literature survey of existing (i.e. state of art) approaches to providing a solution..

ii)Motivate and describe the proposed solution, i.e. the Queuing Theoretic Scheme.

iii)Use the proposed scheme to develop/derive analytical values for QoS performance parameters in two 4G networks (i.e. UTRAN and WLAN). A scheme for comparing the analytical values between the two networks will also be described.

iv)Obtain numerical results and design and perform experiments in an effort to verify and validate the analytical solutions. Numerical results are used to verify the model and are obtained using Matlab software. OPNET simulation software is used for the validation.

v) Make conclusions and recommendations from the results. However, the research might not declare that the analytical model has been validated by simulation results.

It is equally important to note what the research will NOT do. Specifically, the research will NOT do the following:

i)Develop radio link access protocols, e.g. propose a protocol to replace WCDMA protocol for UTRAN. Instead, the scheme will only apply Queuing Theory on the existing schemes.

ii)Calibrate the OPNET simulations; i.e. establish how well OPNET simulates the real network..

1.4 Thesis Outline.

The thesis document is organized as follows: Chapter II gives an exposition of the network selection problem in 4G networks. Both the limitations and opportunities that 4G presents to the network selection problem are discussed. In chapter III, a literature survey of existing approaches to the network selection problem is discussed. Chapter IV presents the proposed solution by motivating and describing a queuing model of the radio access network. In Chapter V, the queuing model is applied to traffic in UTRAN FDD (WCDMA network) and EDCA WLAN (CSMA/CA network) to derive the QoS performance parameters. Later in the chapter, a basis for comparing the analytical values between the two networks is described. The experimental framework for evaluating the performance of the scheme is given in Chapter VI;

the experimental scenarios, the results and analysis are presented in Chapter VII. Chapter VIII is a conclusion of the thesis and suggestions for further work.

2 QoS and Network Selection in 4G Networks

In Chapter I, it was established that QoS Management will be one of the major functional requirements in 4G systems. Network selection was identified as a core element of QoS management. Described as the process of selecting the best network to serve the mobile user based on the user's QoS requirements, network selection will be affected by very many factors. The nature of the access network, the different types of traffic, and the policies of network operators are some of these factors. The following sections discuss the factors that will determine the operation of network selection algorithms.

2.1 The Nature of the Access Network and Network Selection Process

The design of network selection algorithms for 4G networks will have to consider the scale, complexity and capabilities of the access network. [41].

The scale of the network is determined by the number of network operators, the available access technologies and the active user population. Network operators have different operational and business systems and the QoS management function has to factor these variables in the network selection process. For example, the peak and off-peak rates (even the time of day at which they occur) for services differ among network operators; thus the best network for the same service changes accordingly. Packet switched voice service might be cheaper in network (say) A than B during the day; during the evening, however, the reverse may be true. Again, the more the number of access technologies, the more challenging it becomes to select the best network. Finally, the number of active users affects the QoS management and radio resource management mechanisms like admission control, bandwidth allocation and power control. All these processes have been shown to have an effect on network selection [41].

Complexity of 4G networks will also affect the network selection problem. User mobility, a wide range of mobile terminal capabilities and a multiplicity of higher layer protocols greatly contributes to this complexity. In the following examples, scenarios in 4G networks illustrate how the complexity of 4G networks will affect network selection: [29].

i)Mobile terminal capability: the network selection algorithm should assign a high bandwidth access network to a user only if his mobile terminal can handle high data rates, otherwise a low data rate network can work just fine for him.

ii)Advanced protocols: The availability of Real-time Transport Protocol (RTP), a higher layer protocol which has excellent error control mechanisms, can allow the network selection function to relax the error rate criterion in the selection process.

iii)Speed of mobile users: High speed mobile users are readily assigned to wide-coverage networks like WiMAX, Mobile Broadband Service (MBS) etc.

On the bright side, the enhanced capabilities in 4G systems will leverage the difficulties that the massive scale and complexity of the network imposes on the network selection function. For example, content and rate adaptation mechanisms facilitate the support of a wide variety of mobile devices; consequently, the network is able to optimize its transport service to each mobile device. There are other enhanced capabilities in 4G networks that are envisaged to assist the network selection process. For example, advanced modulation and coding techniques will ensure reliable radio link performance; the network selection function will thus not be concerned with link performance. Traffic engineering technologies like MPLS will facilitate QoS routing and bandwidth management [32]. To manage this complexity, advanced QoS signaling is needed [40]

2.2 Type of Service

4G networks will be multi-service networks, i.e. they will support a variety of media components from voice, to text, to video. Generally, different services have different QoS requirements. Access networks, on the other hand, have different capabilities with regard to QoS support. The network selection function will thus have to choose the best network depending on the service that the user is running. Table 2.1 shows the various multimedia services that 4G systems will have to provide, and their QoS requirements. Table 2.2, on the other hand, shows future 4G access networks and their QoS capabilities.

4G networks will be multi-service networks, i.e. they will support a variety of media components from voice, to text, to video. Generally, different services have different QoS requirements. Access networks, on the other hand, have different capabilities with regard to QoS support. The network selection function will thus have to choose the best network depending on the service that the user is running.

Table 2.1 shows the various multimedia services that 4G systems will have to provide, and their QoS requirements.

Table 2.1: QoS Requirements of Main 4G Services [19]

Service	Traffic class	Real time?	Data rate (kbps)	BER	Delay	Jitter
Voice	Conversational	Yes	4 - 25	<10^-3	<150ms	<1ms
Videophone	Conversational	Yes	32 - 384		<150ms	
Gaming	Conversational	Yes	<32		<250ms	
Two-way telemetry	Interactive	Yes	<32		<250ms	
Voice messaging	Interactive	No	4 – 13		<2s	
Web browsing	Interactive	No			<4s	
E-commerce	Interactive	No			<4s	
Email access	Interactive	No			<4s	
Audio streaming	Streaming	Yes	32 - 128		<10s	
Video Streaming	Streaming	Yes	32 - 128		<10s	
SMS	Background	No			<30s	
Fax	Background	No			<30s	
Email (between servers)	Background	No			Variable	

Table.2.2. 4G Access Network Characteristics

Parameter	UTRAN	WiMAX	WLAN
Coverage area	Ubiquitous	Localized and high traffic density areas	Localized and quasi stationary mobile users.
Services	Voice, video and low demanding data applications	Voice, video and high demanding data applications	High bandwidth data applications
User data rate	Up to 2 Mbps for stationary user	Up to 150 Mbps	Upto 11 Mbps
Connections	Point to point and point to multi-to-multipoint with asymmetric transmission capabilities	Point to point and point to multi-to-multipoint with asymmetric transmission capabilities	Generally point to point and symmetric transmission capabilities; occasionally point to multi-to-multi-point with asymmetric transmission capabilities
Cell coverage range	Up to 3 km in outdoor environments. Indoor pico cells may also be used.	Up to 30 km	Up to 300 m

The information in Tables 2.1 and 2.2 shows that some services are better served by some access networks and not others. A network with low transfer delay (<150ms) and relatively low link quality (BER >10^-4) can work fairly well for voice services, while e-commerce applications would do with a network with a higher link quality (BER < 10^-6) and less stringent delay requirement (<4s). UTRAN can support low demanding data applications well, but applications with higher demands are better served by MBS.

2.3 Network Operator Policies

Different network operators have different QoS models and architectures [3]. This difference may lead to different QoS experiences by the user. Thus, in a multi-operator environment, the network selection process must consider the policies of the various network operators

3 Existing Approaches to Network Selection in 4G Networks

In the previous chapter (Chapter 2), it was shown that there is a challenge to provide a mechanism for network selection in 4G networks. Over the years, considerable research effort has been put to address this challenge. A number of approaches have thus emerged from these research initiatives. In this chapter, a survey of four major approaches is presented.

The approaches considered fall into two broad categories

1) Approaches that seek to develop mechanisms for determining the QoS parameters upon which the network selection decision is made. The approaches which belong to this category are

   i) Fuzzy Neural Based Approach for Joint Radio Resource Management.

   ii) The SMART 4G QoS Architecture

2) Approaches that formulate architectures for QoS provisioning and hence network selection. Only one approach in this category is presented: The QoS Broker Architecture for End-to-End QoS Support in 4G Networks

3.1 Fuzzy Neural based Approach for Joint Radio Resource Management

4G networks will be big and complex. Thus, according to [4], the quantity and diversity of the information about the network will be so overwhelming that decisions should not based on the numerical values of the network parameters. Instead, the information should be distilled and represented by some subjective linguistic variables e.g. good, fair, bad, low, high, medium etc. For example, instead of saying 'the expected media access delay in UTRAN for this service is 160 ms', one would better says 'the expected media access delay in UTRAN for this service is fair' The task of choosing the best network is then performed based on these linguistic variables.

This idea is the motivation for the fuzzy neural architecture proposed by [4] and other researchers before that [8] and [36].

Using fuzzy logic theory, a scheme is developed to map the numerical values of the network parameters into fuzzy variables. Thereafter, further processing and the eventual decision making process is accomplished using these fuzzy logic variables. Neural networks are also used in the process for re-enforcement learning. Fig. 3.1 depicts the structure and operation of the fuzzy neural system.

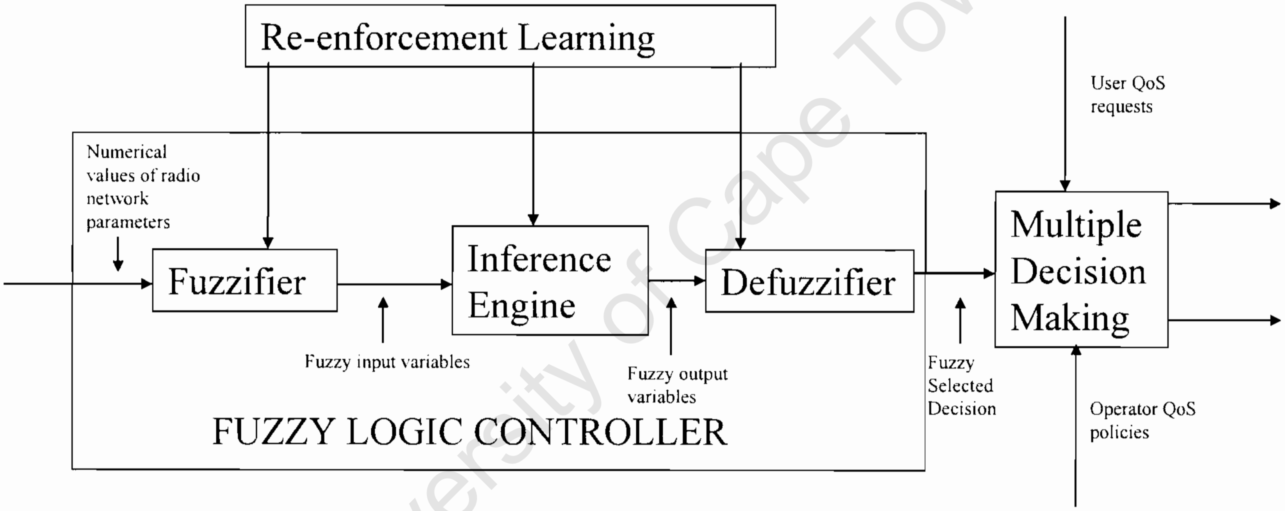

Fig. 3.1 Fuzzy Neural Architecture for Joint Radio Resource Management

The functions of the various component of the fuzzy neural system are given in Table 3.1

Table 3.1. Functions of components of the Fuzzy Neural Architecture

Component	**Function**	**Example**
Fuzzifier	Converts the numerical parameter values into fuzzy output variables	UTRAN RSS:-78 dBm? UTRAN RSS High
Inference Engine	Makes decisions based on fuzzy variables	UTRAN RSS delay? No WLAN delay? Probably Yes WiMAX delay? Yes
Defuzzifier	Processes the fuzzy variables to produce a fuzzy output indicating the relative ranks of the access network	UTRAN? 0.2 WLAN? 0.8 WiMAX? 0.5
Reinforcement learning	Imposes constraints on the fuzzy neural process	The selection criterion shall be that the probability of handoff failure to be less than 0.3
Multiple decision making	This is the final decision making unit. It incorporates user QoS requests and operator policies	Selected Network? UTRAN

In recapitulation, the following can be said about the fuzzy neural approach to network selection in 4G networks.

i)The algorithm accepts numerical values of the radio parameters and then converts them to fuzzy logic variables.

ii)Using fuzzy logic theory, the algorithm processes the fuzzy variables to select the radio network and grant QoS requests. Other factors that are considered are user QoS requirements, radio network parameters, mobile terminal capabilities, user profiles from the core network and operator QoS policy settings.

Results of experiments performed on the fuzzy neural approach are reported in literature [4]. However, this approach is unlikely to be adopted in the wireless networking industry. The adoption requires a radical change in industry practice. Fuzzy neural systems have traditionally been confined only to control engineering; adopting them in the wireless communications industry will require that relevant protocols be developed and incorporated in networking equipment. [36].

3.2 The SMART QoS Architecture

The SMART architecture for heterogeneous network seeks to achieve efficient radio resource management, user-friendly service provision, seamless mobility and security [15]. A distinguishing feature of this approach is the physical separation between the signaling and transport networks.

The architecture is hierarchical and consists of a common core network, a basic access network, and a number of radio access networks. Fig.3.2 is a diagrammatic representation of the architecture.

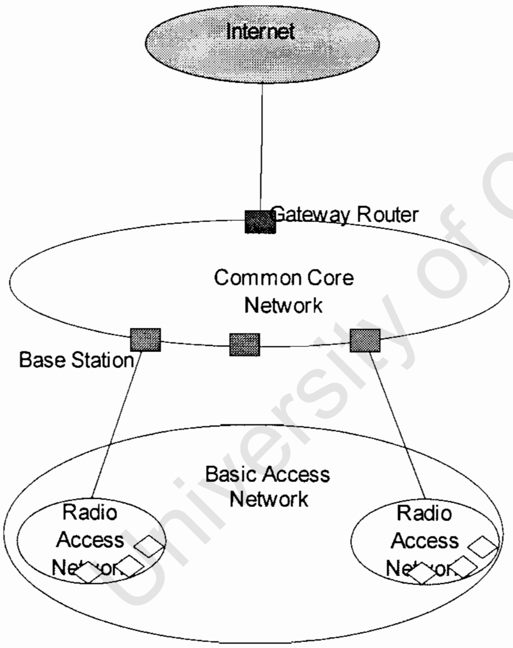

Fig.3.2 The SMART QoS Architecture

The architecture consists of four distinct components, namely the common core network, the basic access network, wireless access networks and the multi-service mobile terminal[15]. The structure and functions of the three components are outlined below.

3.2.1 Common Core Network

The common core network provides the common platform through which mobile terminals communicate with correspondent nodes in external networks. The chief functions of this system are mobility management (routing, location management and hand-offs), radio resource management (traffic distribution, access network selection). A key element in the core network for QoS provisioning is the Resource Manager.

3.2.2 Basic Access Network

The Basic Access Network (BAN) is the overlay QoS signaling network for the mobile terminal and the wireless access networks. Essentially, the BAN facilitates signaling, wireless system discovery, and location management for the mobile terminal and other wireless systems. The BAN provides paging, network discovery, and network synchronization services to the mobile terminal. The mobile terminal uses the common access component of its multi service interface to access this overlay network.

3.2.3 Wireless Access Networks

This comprises the access points (base stations) for the various wireless networks and constitutes the data plane (transport plane) elements in the wireless access system

3.2.4 Multi-service mobile terminal

The multi service mobile terminal is equipped with multiple system interfaces to enable communication with multiple radio access technologies. The common access component is the radio system interface used for signalling; it interfaces to the Basic Access Network. There will also be one or more radio subsystem interfaces for communicating with the available wireless access networks. Typically, these subsystem interfaces will based on software defined radio (SDR) technologies [25].

3.2.5 QoS Mechanisms in the architecture

QoS provisioning in the architecture is accomplished through the interaction between the mobile terminal, the basic access network and the Resource Manager in the core network. For this purpose, the mobile terminal has two modules, the Network Selector (NS) and the Local Resource Manager (LRM). In the core network, the Resource Manager provides the QoS functions of flow differentiation (i.e. service class identification), QoS mapping and filtering (converting service requirements to network requirements) and hence finally radio network selection. The Resource Manager is also responsible for cross-layer information management for QoS. The Basic Access Network provides enables QoS provisioning between the mobile terminal and the Resource Manager.

A desirable feature of the SMART architecture is that the mobile terminal is closely involved in QoS provisioning through the Local Resource Manager. This feature is consistent with the emerging paradigm of distributed intelligence in communication networks. Despite this strong feature, however, the architecture has some serious limitations. Firstly, the processes of flow differentiation, QoS mapping and network selection are not described in sufficient detail. Secondly, the use of an overlay signaling network is rather controversial. Many questions remain to be answered about the Basic Access Network: (i). which radio technology is to be used for the BAN? (ii) In a multi-operator environment, who owns the BAN? (iii) What are the capital and operational costs of the BAN?

3.3 The QoS Broker Architecture for End-to-End QoS Support in 4G Networks.

Providing mobility and QoS support across domains using different access technologies in a seamless way, with no perceived service degradation for the user, is a major requisite for the next generation networks [3]. The statement above sums up the chief challenge of QoS support systems in 4G networks. Moreover, QoS support systems have to be scalable to keep pace with growth in customer demand while at the same time conforming to the business models of several network operators. This composite requirement necessitates a common signaling framework for session negotiation, network resource reservation, and QoS renegotiation [3]. The Quality of Service (QoS) Broker architecture has been proposed.

3.3.1 Functional Structure of the architecture.

In this architecture, the heterogeneous radio access network is considered to comprise the mobile terminal, radio access networks and access routers. For QoS provisioning in this system, a QoS policy management framework is introduced [13]. The framework has the following elements:

i)Policy Repository (PR): This is a database system for policies. QoS policies that span all the layers of the network protocol stack are stored in this element.

ii)Policy Management Tool (PMT): PMT is an interface for creating and managing policies.

iii)Policy Decision Point (PDP): The PDP is the most important element of the QoS management framework. It retrieves policy rules from the PR, generates policy decisions, and oversees the enforcement of the policies.

iv)Policy Enforcement Point (PEP): Executes policy decisions generated by the PDP.

The PDP functionality is usually realized through a network element known as the QoS Broker. The QoS Broker contains a number of multi-media servers and processors that provide QoS support for the various 4G multi-media services. For example, there are streaming servers for streaming services, voice servers for voice services etc [3].

3.3.2 How the QoS Broker Works

Conceptually, the QoS Broker functionality can be described by Fig.3.3.

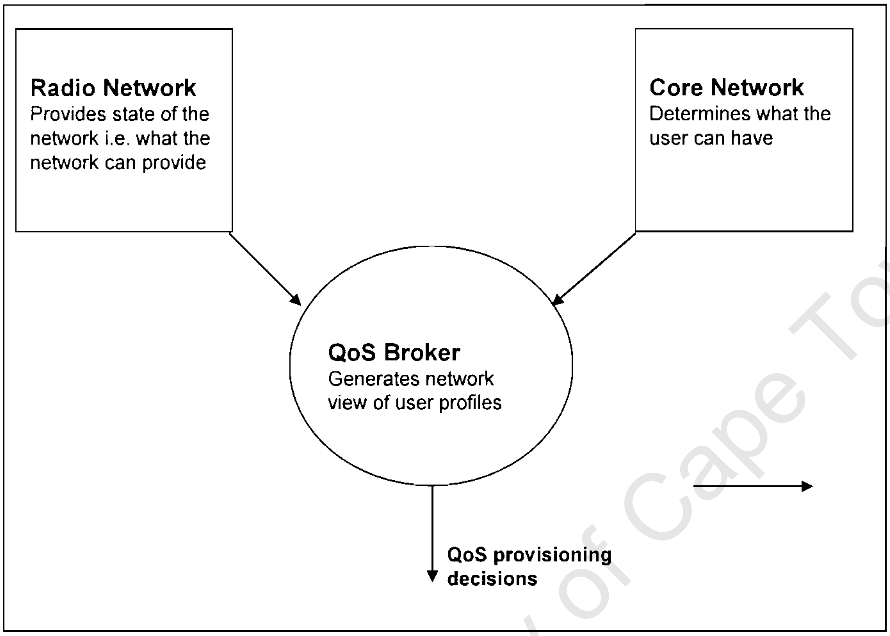

Fig.3.3 Conceptual diagram of the QoS Broker Architecture

The following observations can be made about Fig. 3.3 and hence the QoS Broker architecture.

i)QoS provisioning is accomplished via an integrated approach involving the heterogeneous access network, the core network and a QoS policy management framework. The access network provides the network state information; the core network provides user and service profiles while the management framework defines and executes the QoS provisioning function.

ii)The QoS Broker is at the heart of the QoS management framework. Within the constraints of state of the access network and profiles configured in the core network, the

QoS Broker provides the QoS admission control and network selection functions to the mobile user.

iii)The procedures in the QoS Broker will typically involve user authentication and authorization, service accounting and mapping of service requirements into network requirements.

The QoS Broker architecture provides a compelling QoS provisioning framework. As a result, it has been adopted in a number of research projects e.g., Daidalos [37]

However, the architecture does not specify the details of its functions. Specifically, for the case of network selection, it does not state how the state of the network and application profiles is used to select the best network. This flexibility permits various viable implementations; the objective of this research effort is to propose one such implementation.

4 The Proposed Solution: The Queuing Theoretic Scheme for 4G Wireless Access Networks

In the discussion about the QoS challenge in 4G networks, the author managed to describe the 4G network architecture, its services, and the challenges to be overcome in realizing its vision. Further, the challenge of best network selection was isolated and discussed in greater detail. Finally, three 4G QoS systems are described; these systems represent current initiatives to address the 4G QoS provisioning challenge in general, and network selection in particular.

From the discussions, two major limitations of the above architectures were identified:

i)Whenever a user requests to connect to the network, there should be a scheme for computing QoS performance statistics that each available access network ***is likely*** to give. Such a scheme is not available in each of the architectures mentioned.

ii)Once the QoS performance statistics have been determined as in (i) above, the relative QoS capabilities of the various access networks should be determined. This comparison should be based on the performance metrics determined as in (i) above.

Owing to these limitations, the author proposes a scheme that enhances existing approaches by providing a queuing model of the access network. The QoS performance metrics of the network can then be determined from the parameters of the queuing model.

4.1 Why the Queuing Model? The hypothesis and motivation for the model.

The heterogeneous radio access network is made up of service nodes e.g. UMTS cellular base stations and radio network controllers, WLAN wireless access points, and WiMaX base stations. These service nodes can be considered to be nodes in a queuing system. Traffic arrivals into the network are analogous to customer arrivals into the queuing system while network service is equivalent to queuing service. Therefore, if the radio access network is appropriately modeled by a queuing system, the performance parameters of the queuing system should be equivalent to the QoS performance of the radio access network.

Generally, the analogy between the wireless access system and a queuing system is reflected by the presence of queues, servers (channels), service constraints, and the dependence of performance on traffic distributions in both systems [24].

There are a number of instances in wireless communications research where an approach to treat the wireless network as a queuing system have been adopted. In these research efforts however, the researchers are investigating different aspects of the wireless channel. In [6], for example, the authors treated the IEEE 802.11e wireless channel as a queuing system when investigating the effect of contention window size on system throughput. [39] and [11] utilized principles of queuing theory to evaluate the Erlang capacity of a CDMA radio channel under various conditions.

4.2 Description of the Queuing Model

The author models the wireless access network as a network of queuing nodes. Wireless access nodes are considered to be queuing nodes and are duly replaced by queues in the model. Depending on the nature of the wireless channel, the rate and statistical distribution of traffic arrival into the network is determined. Thereafter, the performance metrics of the queuing system is computed. These metrics represent the QoS capabilities of the various access networks and are used to select the best network to serve the user.

In summary

i)Replace network nodes with queuing nodes..

ii)Determine the rate and statistical distribution of the traffic arrival into the network and service by the queuing nodes. Determine also the system capacity (available wireless channels) and, if necessary, the traffic scheduling disciplines.

iii)Compute the queuing performance metrics (i.e. blocking probabilities, access delays, system throughput etc).

iv)Develop a basis for comparing the networks based on the performance metrics computed.

v)Compare the networks.

vi)Choose the best network.

To characterize queuing systems, Kendall's notation is used, i.e. A/B/K/X/Y/Z

Where

A: traffic arrival distribution, e.g. Poisson, Deterministic etc

B: service distribution or call duration/packet length distribution

K: number of servers (or channels in a communications network)

X: system capacity or size of waiting room

Y: customer population size. In most systems this is normally taken to be infinite.

Z: queuing discipline, e.g. First In First Out (FIFO), Priority Queuing (PQ), Random etc

The most common probability distributions used in queuing theory are Poisson arrivals (denoted M for Markovian), Exponential (again denoted M), General (G), and Pareto (P).

The proposed model is based only on one dimensional queuing analysis; i.e. the time evolution of the queues of different traffic classes are independent. Multi dimensional queuing analysis is more appropriate for multi-service traffic typical of 4G networks; however, the mathematical analysis is exceedingly complex. Notwithstanding this difference, both analyses demonstrate the fundamental idea of using queuing analysis for network selection.

In 4G systems, there will be a number of paths for user traffic through the wireless access network. The user will have a choice of wireless access technologies like UMTS Terrestrial radio Access Network (UTRAN), WLAN (IEEE 802.11), or WiMAX (IEEE 802.16). Fig. 4.1

shows a typical 4G radio access network and the different paths that traffic can follow from the mobile user to the network edge.

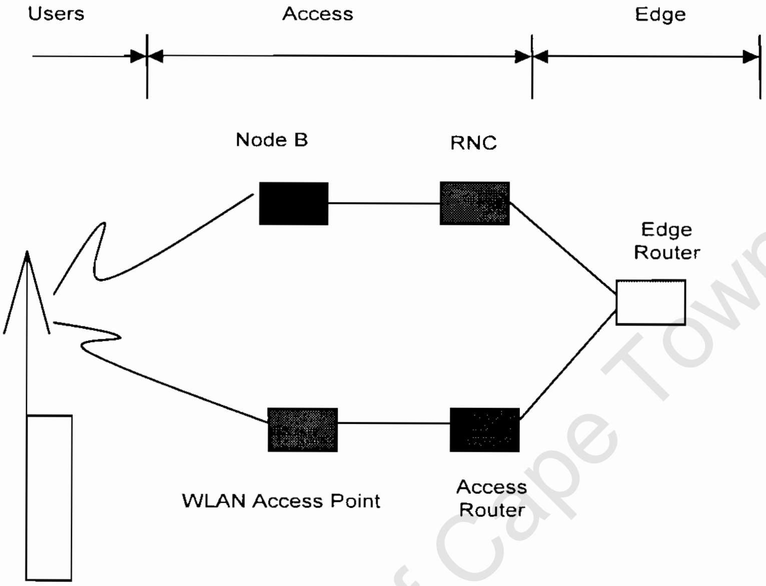

Fig. 4.1: Typical paths in 4G networks

From the diagram it can be seen that a traffic flow can traverse either of the two paths namely

1) UTRAN: Node B to RNC to Edge.

2) WLAN: Wireless Access Point, Access Router, Edge.

From the discussions, the wireless access system shown in Fig. 4.1 can be modeled by the queuing system shown in Fig. 4.2. In Fig. 4.2, service nodes (base stations, wireless access points, routers/switches etc) are represented by queues with service rates $\mu1$, $\mu2$, and $\mu3$. Traffic arrivals into the network are represented by traffic streams of arrival rates $\lambda1$, $\lambda2$, and $\lambda3$. The traffic streams correspond to different traffic/service classes.

Just as the QoS statistics for an arbitrary traffic in the access network depend on the path considered, so will the parameters of the queuing system depend on the branch of the queuing network. Generally, the statistics are determined by the rates and statistical distributions of the traffic arrivals (λ) and service (μ).

The rate and statistical distribution of traffic arrivals and service in a wireless access network depend on both the characteristics of the wireless channel and the type of traffic . In the following sections, we discuss the influence of wireless channel characteristics and type of traffic on the queuing model.

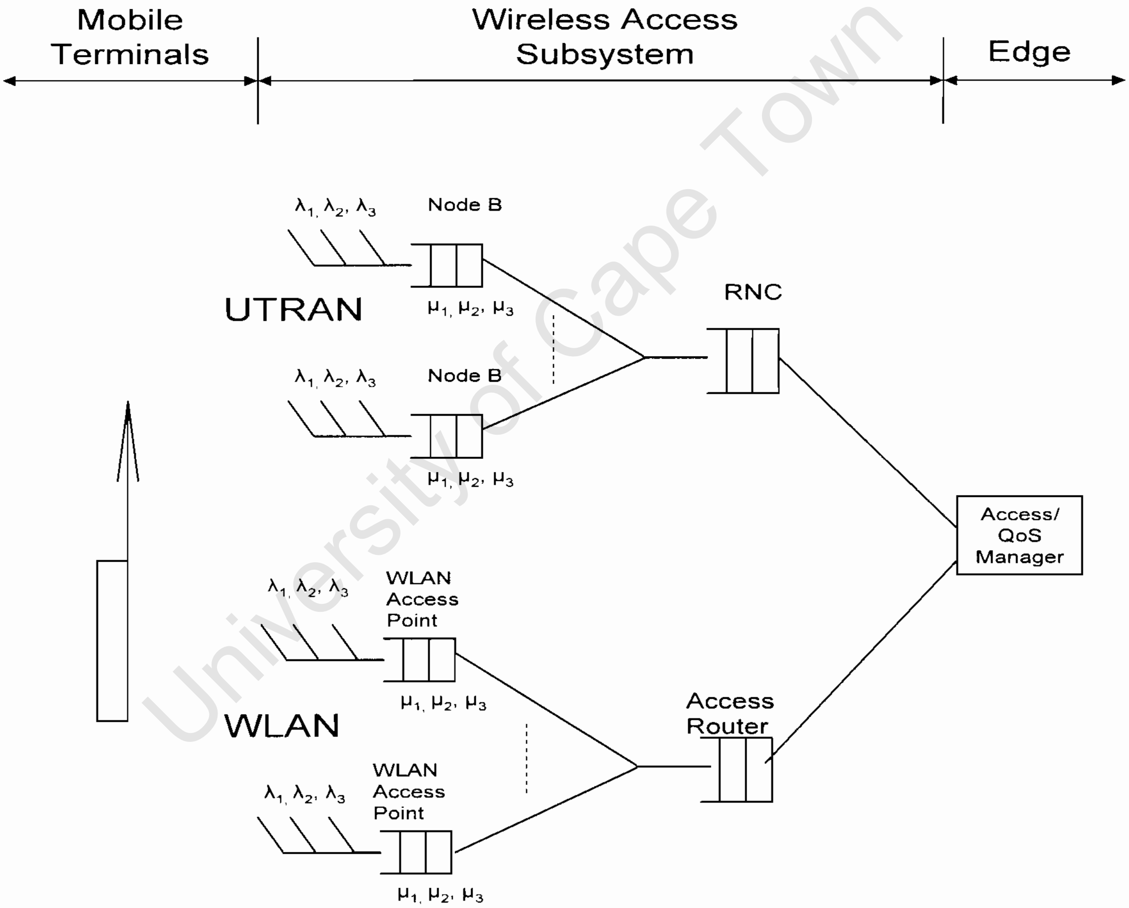

Fig.4.2 A Queuing model of the wireless access system

4.2.1 Influence of the Characteristics of the Wireless Channel on the Queuing Model

The rate and statistical distribution of traffic arrivals and service in the radio access network depend on the wireless channel characteristics. These characteristics, in turn, vary from one radio access technology to the other. Consequently, the queuing model also depends on the characteristics of the wireless channels of the radio technologies. Below are some of the ways in which the wireless channel characteristics influence the rate and statistical distribution of the traffic arrivals and service.

i)User mobility and load balancing: For radio access networks which operate within a cellular topology, user mobility and load balancing requirements necessitate handoff [19]. Handoff, in turn affects the traffic rate intensities (λ and μ) in a cell..

ii)Radio Resource Management Mechanisms. In UTRAN, the time varying channel noise level determines the wireless channel capacity and hence traffic service rate. Again, in UTRAN, the load control function determines the number of users that can be admitted and hence the arrival rate [35]. For WLAN, channel interference level affects the collision rate, which in turn affects the retransmission rate [20].

iii)Link Adaptation techniques. In a UTRAN channel, the coding and modulation scheme can be changed dynamically to adapt to time varying wireless channel conditions [33].This process of adaptive channel coding and modulation affects the data rate (arrival rate) and transmission success rate (service rate) in the wireless channel [30] and [34].

iv)Different Medium Access Protocols. The QoS performance of a wireless channel depends on the multiple access protocol used. Typically, random access protocols like Slotted Aloha and Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA used in IEEE 802.11e) have very large delays, but high throughput performances can be realized with such protocols. On the other hand, contention-less access protocols like CDMA used in UTRAN result in low medium access delays, but rather limited throughput [2]

4.2.2 Influence of Type of Traffic on queuing Model.

4G systems will support both real-time and non-real-time traffic. In a multi-service network, the network's QoS performance parameters vary from one type of traffic to another. The type of traffic affects the model in a number of ways:

i)Queuing Model to be used: Real-time services are modeled by the Erlang Loss model. On the other hand, the Erlang Delay model is used for non-real-time services. In the Erlang Loss Model, the queuing system has no waiting capacity; thus calls that arrive when all servers (i.e. traffic channels in a wireless network) are busy are cleared (declared lost). The Erlang Loss model is the most appropriate model for services that are delay sensitive. The delay model, on the other hand, represents a system with infinite buffer capacity. Thus, calls which arrive when all servers are busy are queued for delayed service. The consequence of adopting different queuing models for different types of traffic is that slightly different QoS parameters are considered. The Erlang loss model leads to blocking probabilities while the Erlang delay model leads to medium access delay.

ii)Priority Queuing: For WCDMA systems, the transport channel configuration and assignment depends on the type of traffic. Real-time services are always assigned dedicated transport channels while non real-time services are frequently assigned common and random access channels [22]. In the system's priority queuing discipline, dedicated transport channels, and hence real-time services are of higher priority than common transport channels. IEEE 802.11e MAC protocol is a priority based medium access scheme where access delay depends on the parameters of the class.

iii)Performance bounds: Different types of traffic have different performance bounds. For example, in a UTRAN channel, the SIR threshold for data services is typically much stringent (< 10^-4) than that for voice services (<10^-3). Given that the UTRAN cell outage depends on the error threshold [42], then the outage statistics of the queuing model will also depend on the type of traffic. A similar situation exists for the IEEE

802.11e channel; each traffic category has a fixed number of retransmission attempts. The number of retransmission attempts is a measure of the maximum allowable medium access delay for each traffic category

4.3 Selecting the Best Network from the QoS Performance Parameters

In the first two sections of this chapter, the author proposed a queuing model scheme as a solution to the 4G network selection problem. A hypothesis and motivation for the solution was presented; then, a detailed description of the model followed. Finally, how the characteristics of the various wireless channels and the different types of traffic affect the model was explained. This section (Section 4.3), discusses the relevant QoS parameters for a 4G access network, and hence the queuing model.

4.3.1 QoS Parameters of the Wireless Network

For each possible flow path through the access network, we determine the QoS performance metrics . The statistic will depend on the wireless access channel, and the type of traffic as discussed in the previous section. The following are some of the QoS metrics of the wireless network.

i)Blocking probability. This is the probability that either a new call or handoff call will be denied service in the wireless network because all servers are busy. This QoS parameter only applies to the Erlang Loss model for real-time services (voice).

ii)Mean channel access delay.

iii)Call dropping probability: This is the probability that a call will be dropped in the cell because of unsuccessful handoff attempt.

iv)Delay probability: The probability that a call will be queued because the system is busy.

v)Throughput: End-to-end throughput considering the fact that calls are delayed at the nodes.

vi)Jitter (Delay variation).

vii)Link BER

The total delay that traffic experiences in a wireless channel has two components i.e.

Total delay = Fixed delay + variable delay Fixed delay is further composed of processor delay, serialization time, and propagation delay [19]. Variable delay is exclusively due to queuing delay at nodes or links (typical values is 20ms). Queuing delay is sometimes called scheduling delay. Values of queuing delays can be obtained by adopting classical analytical queuing systems e.g. M/M/1, M/D/1, M/G/1. Given that scheduling delay is the most important component of channel access delay because it is the component that is variable.

4.3.2 Which then is the best network?

After the network QoS performance parameters have been determined for each network, the task then remains to rank and select the best network. To accomplish this, the author proposes a scheme based on the concept of normalized cell load. In chapter 5, the author describes how the model can be used to derive QoS performance parameters mentioned in Section 4.31 above. The scheme for selecting the best network is then presented in detail.

5 Using the Queuing Model to derive the QoS performance parameters

Chapter 4 presented the queuing model of the wireless network. In the model, analogies are drawn between the QoS performance parameters of a wireless network and the performance parameters of a queuing system. In this chapter, the performance parameters of the wireless channel in both UTRAN and WLAN are analytically derived. The queuing analysis is done for only UTRAN and WLAN networks. However, it should be noted that, although only UTRAN and WLAN are considered, the model can be used for any other wireless network, e.g. WiMAX. The derivation utilizes queuing theory, the model, and the characteristics of the wireless channel.

Only two categories of traffic are considered, i.e. real-time (RT) voice traffic and non real-time (NRT) packet data traffic. In each case, two QoS performance parameters are derived:

i)Media Access Control (MAC) Scheduling delay. This delay constitutes the variable delay in the access network and contributes to the difference in delay performance between networks.

ii)MAC throughput. MAC throughput is a measure of transmission efficiency of the channel access protocol.

5.1 QoS Performance Parameters in UTRAN.

The multiple access technology used in UMTS Terrestrial Access Network (UTRAN) is Wideband Code Division Multiple Access (WCDMA). In WCDMA , channels are identified by codes. Thus, for a user to gain access to the wireless channel, channel codes have to be assigned to it first. To facilitate the transport of a large variety of user traffic on a limited set of physical channels, three types of channels are defined in UTRAN [31].

i)Physical channels: Physical channels denote the physical existence of the Mu interface between the UE and the radio access domain of the network.

ii)Logical channels: specifies the use of data. Codes are assigned to logical channels which then provide bandwidth as needed.

iii)Transport channels: Transport channels define how different user channels are transferred within the physical channel. In UMTS there are very many types of user traffic and hence very many ways to manage them (i.e. very many logical channels). Thus, it is technically injudicious to define a physical channel for each type of logical channel as is the case in 2/2.5G CDMA systems. Thus, only a small repertoire of physical channels is defined and a system is provided to flexibly map the few physical channels into the very many logical channels. Transport channels are the vehicle for this flexible mapping. For example, conversational voice traffic typically requires low bandwidth but has stringent delay bounds. On the other hand, streaming video traffic requires high bandwidth but can tolerate moderate delay. In both cases, the same logical channel will be used (Dedicated Traffic Channel, DTCH). However, different transport channels will be used; dedicated channel for conversational voice and common packet channel for streaming video.

Real-time traffic in UTRAN uses the dedicated channel (DCH) as the transport channel; non real-time traffic uses the common packet channel (CPCH) instead. The DCH is a channel dedicated to one user in either the downlink or uplink. The CPCH is a contention based channel used for the transmission of packet data traffic

In the derivation of the QoS performance parameters , the DCH is used for real-time voice while the CPCH is used for non real-time packet data

5.1.1 QoS Performance parameters for real-time voice traffic.

The derivation proceeds in steps as follows:

Step 1: State and justify the queue model.

We use the M/M/m queue model for the DCH in each cell [39]. Thus the call initiation process for voice in each cell is modeled as independent Poisson streams. Moreover, call durations are exponentially distributed and independent of the arrival processes and other holding times. The author adopted these distributions because they are standard assumptions in tele-traffic engineering for voice calls.

Two assumptions in the adopted model need cautious justification. First, the M/M/m model assumes that the number of channels (i.e. DCH transport channels) in the CDMA cell is limited. This assumption contradicts the concept of 'soft Erlang capacity' of a CDMA cell. In a CDMA cell, the number of available channels depends on the level of interference in the cell, hence there is no hard limit. However, the author contends that if a 'maximum acceptable cell interference level' is agreed upon and fixed, then the channel capacity at this value of interference can be determined. This value would correspond to the number of available dedicated channels for voice and hence the number of servers (i.e. *m* in the queue model). Secondly, the queue model has infinite waiting capacity; hence no call blocking and calls can suffer delays. Why adopt this approach for delay intolerant voice? Okay, the idea is to queue the traffic, compute the expected delay and then compare this value against a threshold. If the expected delay is below the threshold, then accept the call, otherwise reject.

Mobility is not modeled; thus, the mobile user is associated with the cell of its call initiation for the duration of the call. Essentially, it is assumed that the cell size is large compared to the distance a mobile user will traverse during a typical call duration.

Step 2: Determine the number of channels in the cell (i.e. *m* in the M/M/m model)

In a WCDMA cell, the number of traffic channels is given by the soft traffic capacity of the cell. In [42], WCDMA soft traffic capacity C is shown to be

$$C = 1 + \frac{W/R}{E_b/I_o}\,\frac{\varphi_v \varphi_a}{1 + f} \tag{5.1}$$

where

W = system bandwidth (5 MHz for UTRAN)

R = user data rate

Eb/Io = Bit energy to interference ratio.

φv,φa = voice activity gain and antenna gain respectively

f = external cell interference factor (typically 0.6)

The number of servers (m) in the model is then set equal to C.

Step 3: Compute the queuing delay and throughput

The computation follows well known equations in queuing theory [5]
If there are *m* channels and the call arrival rate in the cell is λ and the service rate is μ, then

- Cell utilization is thus $\rho = \dfrac{\lambda}{\mu}$
- For an M/M/m queue the probability that a call arriving will find all channels busy and will thus be queued is

$$P(Queuing) = P_Q = \frac{p_0 (m\rho)^m}{m!(1-\rho)}$$

(5.2)

$$p_o = \left[\sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} + \frac{(m\rho)^m}{m(1-\rho)} \right]^{-1}$$

(5.3

p_0 is the probability of an idle cell.
The queuing delay is readily obtained as

$$T = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda}$$

(5.4)

And the throughput is obtained as [19]

$$\gamma = \lambda(1 - P_Q)$$

(5.5

5.1.2 QoS Performance parameters for non real-time packet data traffic.

Step 1: State and justify the queue model.

The queue model used for non real-time packet data is M/G/1.

Traffic is assumed to arrive in Poisson streams; this is a generally accepted assumption in network modeling. The service distribution is General distribution (i.e. arbitrary distribution). There is only one channel: the contention channel.

Packet data traffic in UTRAN is transported in the contention based common packet channel (CPCH). The transmission scheme used is the Hybrid Automatic Repeat reQuest (HARQ) [19]. A HARQ channel is a contention channel where transmitted data packets (or frames) are checked for errors at the receiver. If the packet is error-free, it is accepted, otherwise it is rejected and the transmitter notified and instructed to retransmit the packet until successful (i.e. error-free transmission.

Step 2: Determine the first and second moments ($\overline{X}\, and\ \overline{X^2}$) of transmission time

Assume general distribution (G) for packet length

P_r = P[Packet retransmission in a channel using HARQ]

P_i = P [Frame successfully transmitted after i retransmissions]

$$P_i = (1 - P_r)P_r^{\,i} \tag{5.6}$$

N_r = Mean number of retransmissions to achieve successful frame transmission

$$N_r = E(i) = \sum_{i=0}^{\infty} iP_i = \sum_{i=0}^{\infty} i(1 - P_r)P_r^{\,i} = \frac{P_r}{1 - P_r} \tag{5.7}$$

$$\overline{X} = 1 + \frac{P_r}{1 - P_r} = \frac{1}{1 - P_r} \tag{5.8}$$

$$\overline{X^2} = \frac{1+P_r}{(1-P_r)^2} \tag{5.9}$$

Step 3: Determine the queuing delay and the throughput in the HARQ channel

The Polaczek-Kintchine (PK) [5] formula gives the total delay in the channel as

$$T = \frac{\lambda \overline{X^2}}{2(1-\lambda \overline{X})} + \overline{X} \tag{5.10}$$

The throughput is given by [21]

$$\gamma = \frac{1-P_r}{\mu} \tag{5.11}$$

Tables 5.1 and 5.2 gives a summary of the formulae derived for both voice and data traffic in UTRAN.

Table 5.1 Analytical Formulae for Voice Parameters in UTRAN Cell

Performance Parameter	Formulae
Scheduling Delay	$T_V = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda}$
Throughput	$\gamma_V = \lambda(1-P_Q)$

Table 5.2 Analytical Formulae for Data Parameters in UTRAN Cell

Performance Parameter	Formulae
Scheduling Delay	$T_D = \frac{\lambda \overline{X^2}}{2(1-\lambda \overline{X})} + \overline{X}$
Throughput	$\gamma_D = \frac{1-P_r}{\mu}$

5.2 QoS Performance Parameters in WLAN

For QoS support in WLAN, the IEEE 802.11e standard has been developed. To support traffic with different QoS requirements, a channel access mechanism known as Enhanced Distributed Coordination Access (EDCA) is implemented [16]. The multiple access protocol used in IEEE 802.11e is the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA).

Traffic in EDCA WLAN cell is classified into Access Categories (AC). There are four ACs and each AC is characterized by the following parameters:

i)Minimum Contention Window (CWmin).

ii)Maximum Contention Window (CWmax).

iii)Arbitration Inter-Frame Space (AIFS). This is the time interval between the wireless medium becoming idle and the beginning of channel access negotiation. Put another way, this is the minimum idle period before channel contention.

iv)Transmit Opportunity Limit (TXOP limit). This is the Maximum duration (in ms) during which a QoS enabled station (QSTA) can transmit after a successful channel contention.

At each QSTA, a MAC function maintains four transmit queues, where each queue corresponds to each of the AC. Whenever there's a collision; a Virtual Contention Handler resolves the contention with the higher priority queue taking precedence. Fig. 5.1 illustrates the transmit queue operation of the EDCA MAC protocol

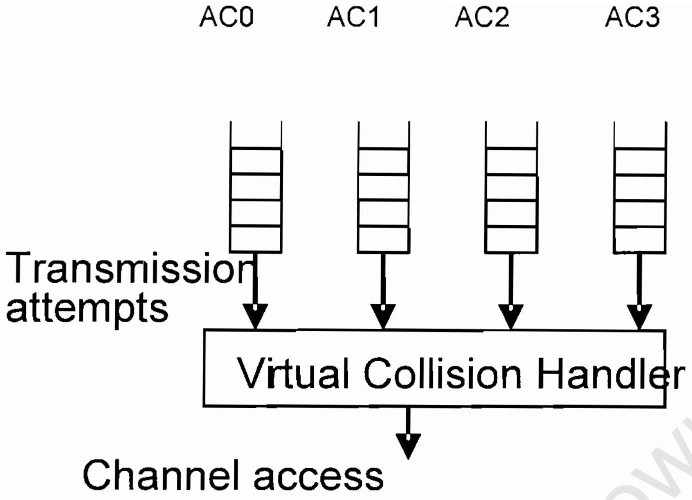

Fig. 5.1 AC transmit queues for EDCA

The CSMA/CA channel is a contention channel where the transmitting nodes strive to avoid collisions by backing off when the channel is sensed busy. Before a station transmits, it must first sense that the channel is idle for a time period known as the Arbitration Inter-frame Space (AIFS), and then restrains itself from transmitting for a random length of time known as the Back off time (b_n). The back-off time depends on the AC of the traffic to be transmitted. Channel Access Delay (CAD) is the delay before a transmission attempt. CAD has three components: an Arbitration Inter-Frame Space (AIFS), a Distributed Inter-Frame Space (DIFS) and a Contention Window (CW_n). DIFS has a fixed length. AIFS and CW_n have variable lengths that depend on the AC. The higher the priority of the traffic to be transmitted, the shorter the CAD. However, the total delay for successful frame transmission is the sum of a series of back-off times and the frame transmission time. CAD and total delay (X) are determined as follows [43]

$$AIFS = \eta DIFS$$

$0 \leq \eta \leq 1$ (η depends on traffic class)

$$CW_n = (2^{r+n-1} - 1) * AIFS \tag{5.12}$$

$n = n^{th}$ transmission attempt

r = a factor that depends on the traffic type. The higher the priority of the traffic, the lower the value of r

$$CAD = DIFS + b_n \; . \tag{5.13}$$

$$0 \leq b_n \leq CW_n \tag{5.14}$$

$$X = DIFS + T_b + T_f$$

In summary, the EDCA WLAN MAC protocol operates as follows [*23*]

1) A QSTA with data to transmit senses the channel for a DIFS period. If channel is idle, it transmits, otherwise it waits a random length of time b_n.

2) While waiting, a timer decrements for each AIFS time of channel being idle. The timer countdown is frozen whenever a transmission is sensed in the channel; and is resumed only after the channel has fallen idle for a DIFS.

3) Once the timer decrements to zero, the QSTA starts transmitting immediately. If the transmission is successful, an acknowledgement frame is sent to the sender; otherwise, a collision is presumed to have occurred.

4) In the event of a collision, the back-off event counter is incremented (i.e. n is incremented), a new contention window is calculated according to Eq (5.12) and a new contention event is started with bn set according to Eq. (5.14).

The background information presented above shows how the IEEE 802.11e channel operates. In the following sections, the mode of operation presented is used to derive the QoS performance parameters of voice and data traffic in IEEE 802.11e channel. It is worth noting that the MAC channel behaves the same way for both real-time voice and non real-time data traffic; only the parameters of the AC differentiate them.

5.2.1 Scheduling delay in EDCA WLAN

The process of deriving the queuing delay and throughput proceeds in steps.

Step 1: Determine the mean value of the total back-off time T_b in the EDCA channel. This value is a component of the total channel access delay and also contributes to the channel throughput.

T_b is the random sum of random variables. Here the individual random variables are the back off times (b_n) within a contention window. The number of back off events (n) before successful transmission is itself a random variable. Fig.5.2 shows the series of back-off events which together constitute the total back-off time T_b.

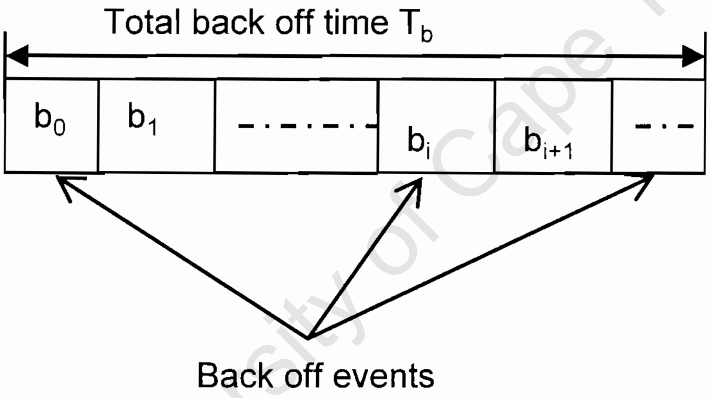

Fig. 5.2 Back off events in EDCA WLAN

From Fig.5.2

$$T_b = b_0 + b_1 + \ldots + b_n \tag{5.15}$$

- The back off time b_n of each of the back-off events is a random variable distributed over the interval [Cmin, Cmax] = $[2^r-1, 2^{r+n}-1] = [k_0-1, k_0 2^n-1]$ where $k_0=2^r$

Assuming that b_n follows a discrete uniform distribution over the contention window. Then the mean back-off time over a contention window is

$$E(b_n) = \sum_{k=a}^{b} k\left(\frac{1}{b-a+1}\right) = \frac{b+a}{2} = \frac{k_0 + k_0 2^n}{2} \tag{5.16}$$

For $a=k_0$ and $b=k_0 2^n$

- The mean number of back off events (n) before successful transmission follows a geometric distribution.

P_n = P[n back offs before successful transmission]

$$P_n = (1 - p)p^{n-1} \tag{5.17}$$

p = collision probability in the MAC channel.

$$E(n) = \sum_{n=0}^{\infty} nP_n = (1 - p)\sum_{n=0}^{\infty} np^{n-1} = \frac{p}{1 - p} \tag{5.18}$$

As has been mentioned in the preceding paragraph, the mean length of the back-off interval *Tb* is a random sum of random variables. Having determined the mean of the random variables (i.e. $E(b_n)$) and the expected number of such random variables ($E(n)$, the value *Tb* is obtained by the Law of Total Expectations as

$$\overline{T}_b = E(b_n)E(n) \tag{5.19}$$

Step 2: Obtain the mean value of effective frame transmission time X

The mean time between successful frame transmissions X is the sum of DIFS, mean value of total back off time (T_b) and frame Transmission time Tf. Fig. 5.3 is a diagram of the elements of the effective frame transmission time in EDCA channel.

From Fig.5.3

$$\overline{X} = DIFS + \overline{T_b} + T_f \tag{5.20}$$

T_f is a random variable that depends on the packet length distribution. In the analysis, RT packet lengths are taken to be exponentially distributed while NRT packet lengths are Pareto distributed.

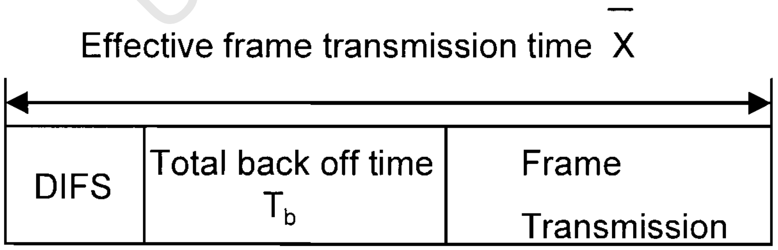

Fig. 5.3 Effective frame transmission time in EDCA WLAN

Step 3: Computing the queuing delay

The EDCA WLAN channel can be modelled by an M/G/1 queue. Contention channels are generally represented by the M/G/1 queue model [*23*].

After obtaining the effective frame transmission time $\overline{X}$ and the associated second moment $\overline{X^2}$, the Pollaczek-Kintchine (PK) formula for the M/G/1 queue model is applied to obtain the waiting time (W) and system time (T) for traffic in the channel.

$$W = \frac{\lambda \overline{X^2}}{2(1 - \lambda \overline{X})} \tag{5.21}$$

$$T = W + \frac{1}{\mu} \tag{5.22}$$

Where λ = traffic arrival rate

μ = service rate = 11 Mbps in an EDCA WLAN.

5.2.2 Throughput in EDCA WLAN

In the WLAN channel, when an arriving QSTA finds the channel idle, it starts transmitting immediately. When the channel is busy, the node becomes backlogged. If there is a collision, the nodes involved become backlogged. Backlogged nodes attempt retransmission with probability q. For persistent retransmission: $q = 1$, while for non-persistent retransmission: $q < 1$ [*5*]. Thus, the traffic arrival rate has two components: external arrivals (λ) and backlog retransmissions (nq).

For EDCA WLAN, retransmission is persistent (q = 1). A backlogged packet is transmitted with probability 1 after a random delay of mean 1/x; (n = backlog, 1/x = average time between retransmission attempts = $\overline{b_n}$ as given in Eq. 5.16

Let $g(n)$ be the attempt rate (the expected number of packets transmitted in a retransmission interval) when there are n backlogged nodes.

Then, the total arrival process is a time varying Poisson random process with rate

$$g(n) = \lambda + nx = \lambda + \frac{n}{b_n} \tag{5.23}$$

The number of attempted packets per interval in state n is approximately a Poisson random variable of mean $g(n)$

Probability that there are m transmission attempts

$$\text{P } (m \text{ attempts}) = g(n)^m \frac{e^{-g(n)}}{m!} \tag{5.24}$$

P (idle) = P(m=0), i.e. probability of no attempts in an interval = $e^{-g(n)}$

$$\text{P (success)} = \text{P}(m=1), \text{ i.e. probability of one attempt in an interval} = g(n)e^{-g(n)} \tag{5.25}$$

P (collision) = P (two or more attempts) = 1 -P(idle) -P(success)

In LAN systems, throughput is defined as the fraction of intervals with at least one successful transmission. Hence,

Throughput (γ)= P (success) = $g(n)e^{-g(n)}$

In recapitulation, the scheduling delay and throughput for the EDCA WLAN channel so far derived are shown in Table 5.3. Recall that the difference between real-time voice and non real-time data traffic lies in the values of the access category (AC) parameters. Table 5.4 gives typical parameter values for voice and data AC

Table 5.3 Analytical Formulae for QoS Parameters in EDCA WLAN Cell

QoS parameter	Formula
Queuing delay T	$W = \frac{\lambda \overline{X^2}}{2(1-\lambda \overline{X})}$ $T = W + \frac{1}{\mu}$
Throughput	$g(n)e^{-g(n)}$

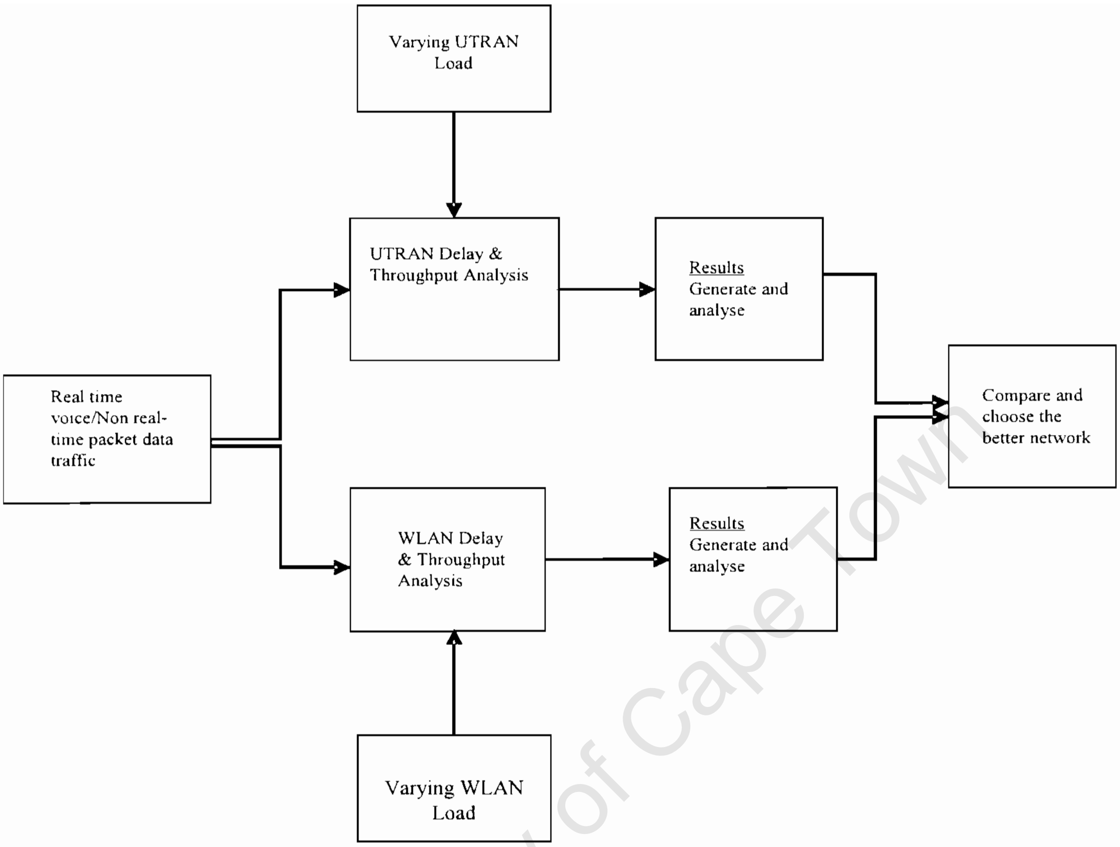

Fig. 6.1: Verification test-bed for proposed scheme

Verification tests were implemented using the Matlab software.

Reasons for using Matlab

i)The analytical algorithms developed during the research are largely mathematical. Thus, it was reckoned that software with strong computational capabilities would be the appropriate one for implementing the scheme. Matlab is such software.

ii)Matlab has built-in high-level language programming facilities (i.e. M-file programming). M-file programming is used for coding the analytical algorithms for the model.

Table 5.4 Access Category Parameters in EDCA WLAN Cell

AC parameter	Real Time Voice	Non real time data
AIFSN	2	3
CWmin	3	15
CWmax	7	1023
TXOP (ms)	3	5

5.3 Comparing the access networks.

After determining the expected queuing delays and throughput for the access networks as outlined above, the Access/QoS manager then selects the best network based on these parameters. Typically, network selection for real-time voice traffic will be based on the delay parameter. On the other hand, selection for data traffic is based on the throughput parameter. The QoS function compares the appropriate performance parameters of the two networks and then picks on the better of the two.

However, in comparing the performance of two networks based on delay and throughput, one also needs to specify the corresponding error performance value. For example, a queuing delay of say 20 milliseconds for voice in either UTRAN or WLAN should specify the bit error probability associated with that delay. Thus, to compare and select either networks one should (1) fix an error performance level for the service in the channel, (2) determine the expected delay or throughput for either networks, and finally (3) compare. In the sections that follow, we specify, how the objectives are achieved in the proposed scheme.

In all wireless access systems, the performance of the system degrades as the number of simultaneous users increases. Thus, if a minimum error performance level is set for a particular service, then, there will be a corresponding maximum number of such users of the system. The maximum number of users of a wireless system is known as **pole capacity** [7]. Moreover, wireless access systems can be divided into units called cells. In a UTRAN cell, the signal to interference ratio (SIR) is the performance measure, while the types of services that can be supported are conversational voice, interactive service, background service and Best Effort. However, for EDCA WLAN, the Basic Service Set (BSS) constitutes a WLAN cell and the collision rate is the performance measure. Four (Qty 4) different traffic categories are defined for EDCA service.

The number of users in a cell at a particular load is known as the cell load. Taking the pole capacity as the reference cell load, the cell load at any point in time can be normalized to the maximum cell load i.e.

$$Normalized\ Cell\,Load = \frac{Cell\,Load}{MaxCell\,Load}. \tag{5.26}$$

Normalized cell load figures can range from zero to one; with one corresponding to a full cell, i.e. a cell stretched to its full capacity and thus can no longer accept any additional user. The normalized cell load can be viewed as a measure of the capacity of a cell to accept more users, irrespective of the radio access technology. Thus a normalized cell load of, say 0.6, tells us that the cell is 60% full, be it a UTRAN cell, a WLAN cell, a WIMAX cell or indeed the cell of any other wireless access technology. Now, suppose we wish to determine the expected QoS performance of a particular service at a given normalized cell load. The author argues that the QoS performance values of different services at the same normalized cell loads indicate the QoS merits of the wireless systems.

In summary, the procedure for evaluating the QoS merits of different wireless technology consists of the following steps

1) Identify the wireless system unit (e.g UTRAN cell, WLAN BSS etc) and the system performance measure (e.g SIR for UTRAN and collision rate for WLAN)

2) For each supported service, determine the minimum allowable performance level and hence the pole capacity. For example, in UTRAN the minimum SIR for voice is 7 dB; the pole capacity corresponding to this value is then determined.

3) For system loads up to and including the pole capacity, determine the expected QoS performance parameters of the wireless access systems. Compare the figures to obtain the relative merits of the various wireless technologies.

5.3.1 System Load in UTRAN

Step 1

In UTRAN, the wireless system unit is the UTRAN cell and the measure of performance is the signal to interference ratio.

Step 2 : Determining the pole capacity, m_0

From [*42*], the number of users in UTRAN cell can be determined as

$$m = 1 + \frac{W/R}{E_b/I_o}\frac{\gamma_v\gamma_a}{1+f} - \frac{W/R}{E_b/N_o}\frac{\gamma_v\gamma_a}{1+f} \qquad (5.27)$$

Where m = Number of users

W = WCDMA system bandwidth = 5 MHz

R = nominal user data rate. Typically, this can be taken as 38.4 Kbps for voice and say 64 Kbps for data

E_b/I_o = signal Energy to interference ratio in dB

γ_v = Voice activity gain

γ_a = Antenna gain

f = Other cell interference factor

E_b/N_o = Signal energy to thermal noise ratio

Let $$\frac{(W/R)\gamma_v\gamma_a}{1+f} = \alpha \qquad (5.28)$$

Then $$m = 1 + \frac{\alpha}{E_b/I_o} - \frac{\alpha}{E_b/N_o} \qquad (5.29)$$

And $\frac{Io}{Eb} \approx \frac{m-1}{\alpha} + (E_b / N_o)^{-1}$ (5.30)

Or $\frac{Io}{Eb} \approx \frac{m-1}{\alpha}$ since $(E_b / N_o)^{-1} \prec\prec 1$

Finally $m = 1 + \frac{\alpha I_o}{E_b}$ (5.31)

m_0 *is* m at minimum acceptable SIR (i.e. $\frac{E_b}{I_0}$); 7 dB for voice and 10 dB for data

Step 3: Determining the uplink load

The Admission Control policy adopted is the throughput based uplink loading. Essentially, this policy establishes that calls are only admitted when the total uplink load is below a certain threshold. The threshold is the load at pole capacity.

The uplink load is the sum of bit rates of active connections divided by the throughput of the cell at pole capacity.

Thus, the uplink load at pole capacity is

Uplink load $\eta_0 = (1+f)\left(1+\frac{1}{m_0\beta}\right)^{-1}$ (5.32)

Where

m = Number of users

f = Other cell interference factor

$\beta = (E_b/I_o)R/W$

For real-time voice E_b/I_o = 7 dB while for non real-time packet data E_b/I_o = 10 dB

For $m = 1$ up to $m = m_0$, the uplink load is evaluated. The normalized cell load is obtained as η/η_0

Step 4

For each normalized cell load, determine, the MAC scheduling delay and throughput as appropriate

5.3.2 System Load in WLAN

Step 1

In IEEE 802.11e, the wireless system unit is the Basic Service Set (i.e. a set of wireless nodes under the domain of one access point) and the measure of performance is the collision rate (probability of collision among contending nodes). An important parameter in WLAN is the number of active nodes *n*.

Step 2

In WLAN, the collision rate p must satisfy the condition $p \leq 0.5$ for the cell to be stable (i.e. bounded delay) for both RT voice and NRT packet data traffic. In the Appendix, the collision rate p is determined as

$$p = 1 - 2\left(1 - \frac{1}{n}\right)^{n-1} \tag{5.33}$$

Where n = the number of active nodes

The load is obtained as $\gamma = g(n)e^{-g(n)}$ (5.34)

Maximum cell load is $\gamma_0 = g_0 e^{-g_0}$

Where g(n) is given by Eq.5.23

Step 3

For $n = 1$ to $n = n_0$, the normalized cell load is obtained as γ/γ_0

Step 4

For each normalized cell load, the MAC scheduling delay and throughput are obtained as given by equations in Tables 5.3

Finally, the values obtained in step 4 for both UTRAN and WLAN are compared, and a decision is made.

6 Experimental Framework for Evaluating the Queuing Model

In Chapter 5, the author developed procedures for the analytical derivation of the media access delay and throughput for real-time voice and non real-time data services in both UTRAN and IEEE 802.11e access networks. In this chapter, an experimental framework for the verification and validation of the proposed analytical scheme is discussed. According to [32], verification is the process of discovering how a model implements the assumptions used in constructing the model. Validation, on the other hand is used in determining the validity of the model; i.e. how the results from the model compare with that of real systems. The process of model validation consists of validating assumptions, input parameters and distributions, and output values and conclusions. Validation can be performed by a number of techniques, one of which is comparing numerical (analytic) results using queuing theory or other analytic methods with results from random simulations. The analytical results and simulations will attempt to answer the following questions.

i)How do varying traffic conditions affect the performance parameters (e.g. delay and throughput) obtained from the analytical model? How can this information be used to select the best network?

ii)How does the analytical results compare with simulation results from random data?

6.1 Numerical Results

To obtain the numerical results, the cell traffic load was varied and the corresponding scheduling delay and throughput obtained. In UTRAN, varying the traffic load was realised by varying the signal to interference ratio (SIR) [42]. In EDCA WLAN, however, the cell traffic load was varied by varying the number of active nodes. Different types of traffic (e.g. real-time voice and non real-time data) were injected into the network as the cell load was varied. The behaviour of the wireless channel was simulated by a code of the queuing model. The results obtained were analysed and appropriate conclusions made. Figures 6.1 depict the architecture of the verification experiment.

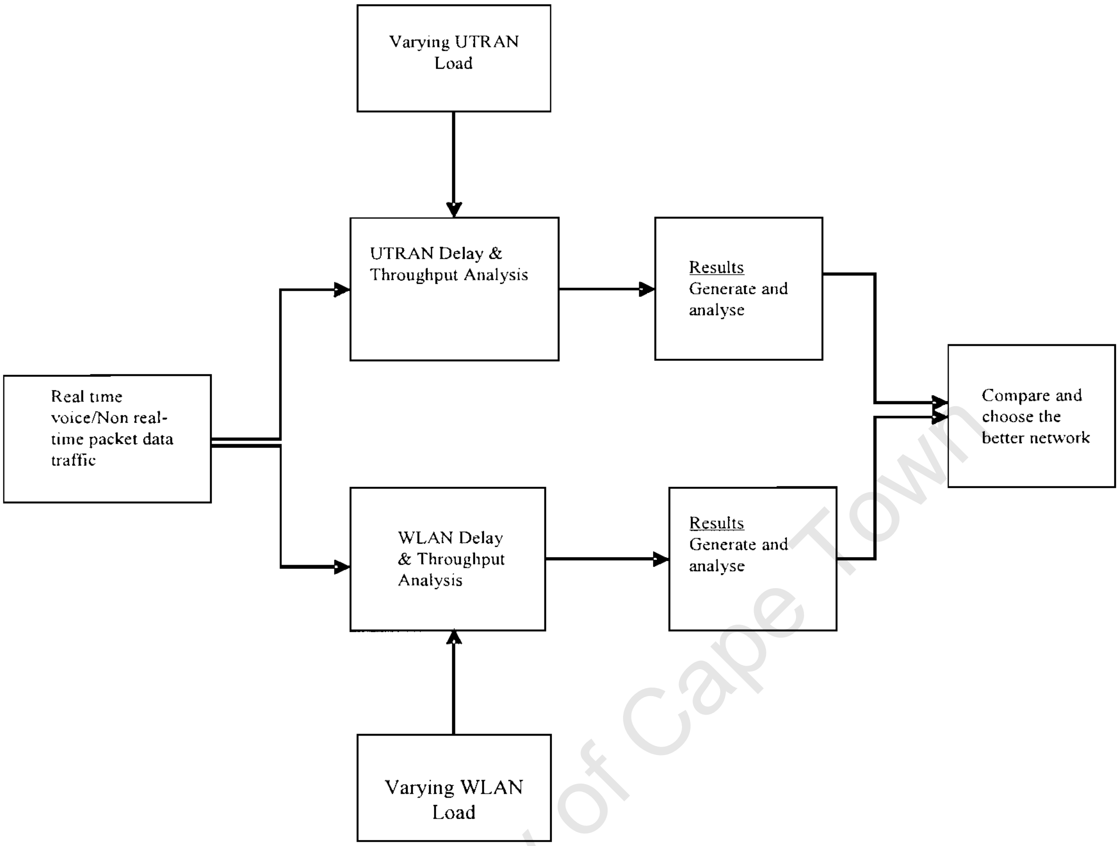

Fig. 6.1: Verification test-bed for proposed scheme

Verification tests were implemented using the Matlab software.

Reasons for using Matlab

i)The analytical algorithms developed during the research are largely mathematical. Thus, it was reckoned that software with strong computational capabilities would be the appropriate one for implementing the scheme. Matlab is such software.

ii)Matlab has built-in high-level language programming facilities (i.e. M-file programming). M-file programming is used for coding the analytical algorithms for the model.

iii)Matlab has a suite of appropriate utilities for analysis of results e.g. Graphics for plotting, Excelink for exchange with other applications and statistical analysis.

For the numerical results, the author investigated only two QoS performance parameters, i.e. scheduling delay and throughput. These are the major QoS parameters in a wireless network. Conclusions drawn from the results can be extended, with appropriate adjustments, to other parameters.

To simulate varying load in a UTRAN cell, the signal to interference ratio (SIR) was varied from a minimum value (7 dB for real-time voice and 10 dB for non real-time data) to a maximum value of 40 dB. The load was normalized to the maximum load (pole capacity) according to Eq. 5.26. Pole capacity corresponds to cell load at minimum acceptable SIR, e.g. 7 dB for voice.

In EDCA WLAN, cell load is determined by the number of active nodes. The load in the WLAN was varied as explained in Section 5.2, and the corresponding delays and throughputs obtained.

The algorithms were run under different scenarios. A scenario is characterized by a set of cell configuration parameter and traffic parameters. The MAC scheduling delay and throughput at each scenario was obtained as the normalized load was varied. The results from each of the scenarios were obtained and subsequently analysed. Tables 6.1 and 6.2 show the configuration parameters used for the UTRAN and WLAN cells respectively.

Table 6.1: UTRAN Cell Configuration Parameters

Parameter	*Real time voice*	*Non real time data*
WCDMA system bandwidth (MHz)	5	5
Nominal user data rate R (kbps)	Variable	Variable
Minimum SIR (dB)	7	10
Antenna gain	1.7	1.7
Voice activity gain	1.8	1
Other cell interference factor (f)	0.6	0.6

Table 6.2: EDCA WLAN Configuration

Parameter	Real time voice	Non real time data
IEEE 802.11e system bandwidth (Mbps)	11	11
CWmin	Variable	Variable
CWmax	63	255
AIFS	2	3
TXOP Limit (ms)	6	3

Table 6.3 has the rates and statistical distribution used for the traffic in both the UTRAN and WLAN cells.

Table 6.3: Traffic Distributions in the UTRAN and WLAN Cell

	Real time voice	Non real time data
Traffic arrival and service rates	Variable	Variable
Inter-arrival time distribution	Exponential	Exponential
Packet length distribution	Exponential	Pareto (2000,1.8)

6.2 Validation through simulations.

Simulations were carried out to establish how the results of the analytical model compare with results from random simulations The OPNET Modeler software was used as the random simulator.

Validation involved building a scenario and then running it in both Matlab and OPNET.Modeler. Matlab was used to realize the analytical model results while random simulation results were obtained from the OPNET experiments. The two sets of results were then compared and conclusions made about them.

Fig. 6.3 shows the procedure for the simulations.

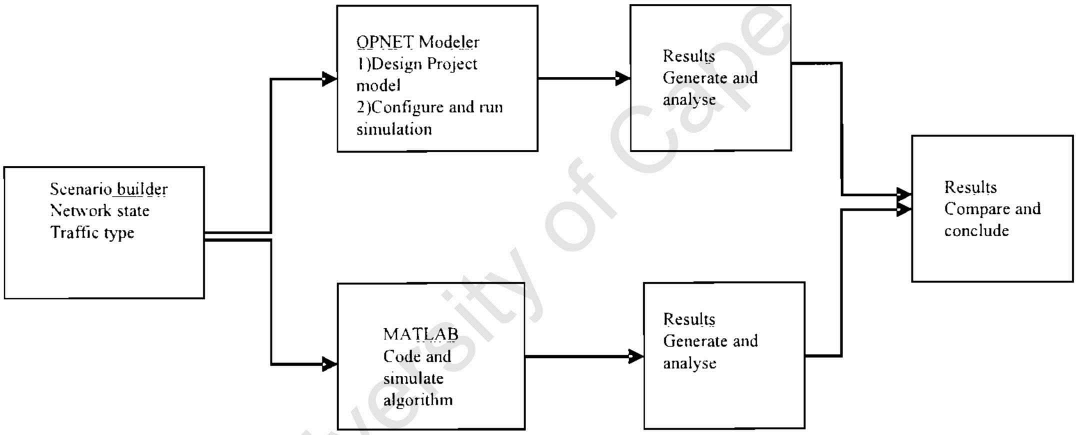

Fig. 6.2. Simulation Procedure

6.3 Scenarios

In order to investigate features of the model, five scenarios were evaluated and numerical results obtained. Scenarios are built by changing some parameters of either the user traffic or the network. The five scenarios are presented in Table 6.4

Table 6.4: Scenarios

Scenario Number	Purpose of scenario
1	To verify model performance for variable rate voice traffic with respect to scheduling delay
2	Comparative study of scheduling delay for constant rate voice traffic between UTRAN and WLAN
3	To verify model performance for variable rate data traffic with respect to throughput
4	Comparative study of throughput for constant rate data traffic between UTRAN and WLAN
5	Validation experiment: comparing the analytical model with a simulation

7 Results and Analysis

The results from the scenarios described in Chapter 6 are presented in the following sections. For each set of results, observations are made and possible explanations given. Finally, conclusions on the significance of the results are presented.

7.1 Scenario 1: Scheduling Delay for variable bit rate voice

Table 7.1.1: Scheduling Delay for RT Voice in UTRAN (Nominal Data Rate = 38.4 kb/s; traffic arrival rates λ are in kb/s)

Normalized Cell Load		0.00	0.14	0.63	0.77	0.84	0.88	0.89	0.91	0.92	0.93	0.94	0.95	0.96	0.97
Delay in msec.	λ = 16 kbps	0.41	0.41	0.81	1.18	5.21	5.23	5.25	5.26	5.28	5.30	5.32	5.34	5.39	5.56
	λ = 20 kbps	2.18	2.18	3.15	7.14	7.94	7.94	7.94	7.94	7.94	7.94	7.94	7.94	7.94	7.94
	λ = 38.4 kbps	5.21	5.21	6.13	9.52	9.52	9.52	9.52	9.52	9.52	9.52	9.52	9.52	9.52	9.52

Table 7.1.2: Scheduling Delay for RT Voice in WLAN (CWmin = 15, AIFS = 20μs, TXOP = 3ms; traffic arrival rates λ are in kb/s)

Normalized Cell Load		0	0.02	0.12	0.17	0.26	0.31	0.36	0.41	0.46	0.61	0.8	0.9	1
Delay in milliseconds	λ=16	0	21	25	25	26	26	26	26	26	26	26	26	26
	λ=20	0	26	31	31	31	31	32	32	32	32	32	32	32
	λ=38.4	0	46	56	57	58	58	58	58	59	59	59	59	59
	λ=64	0	75	93	94	96	96	96	96	97	97	97	97	97
	λ=128	0	147	187	190	193	194	195	195	196	197	197	197	197

Figs. 7.1.1 and 7.1.2 are graphical representations of the results in the table above.

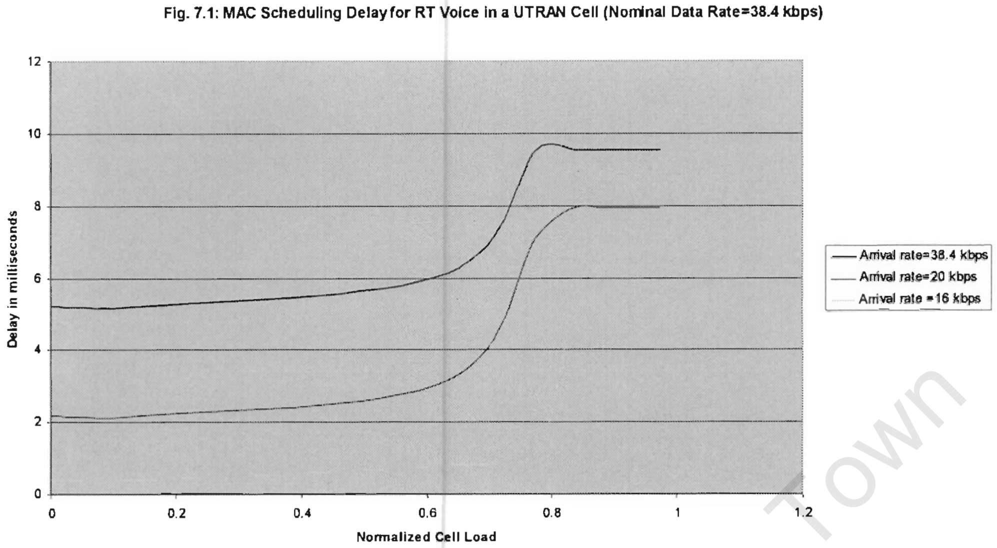

Fig 7.1.1 MAC Scheduling Delay for Voice in UTRAN

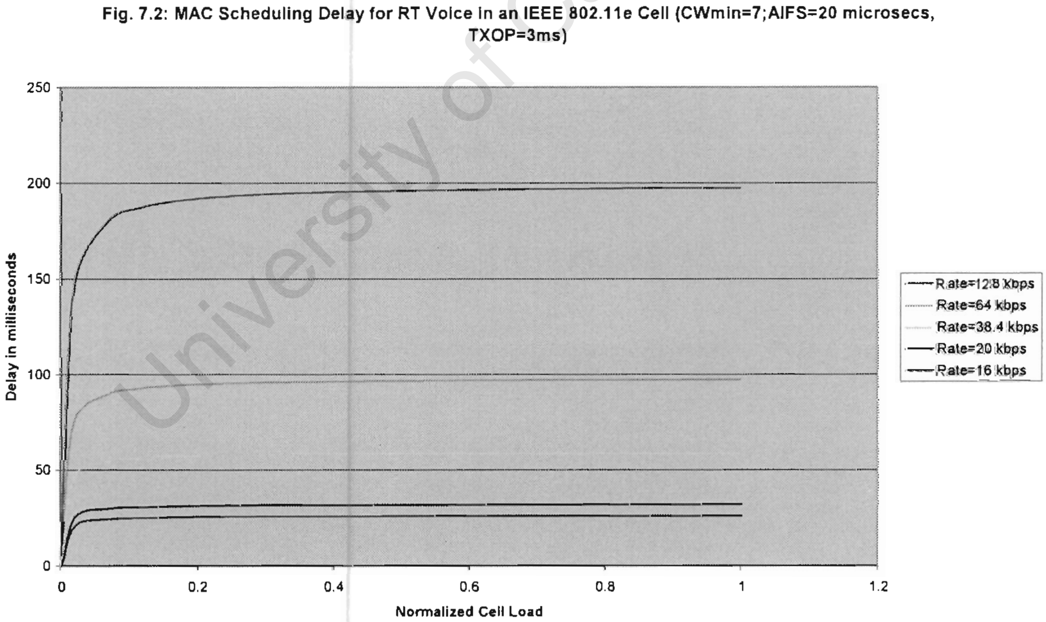

Fig 7.1.2 MAC Scheduling Delay for Voice in WLAN

Discussion of Results

For voice traffic in the UTRAN cell, the delay remains relatively constant for all values of traffic arrival rates below a certain threshold of Normalized Cell Load (e.g. at approximately 0.78 for data rate of 16 kbps). At the threshold, the delay spikes to a new value, where it remains constant for subsequent values of cell loading.

A possible explanation for this observation is that traffic arriving when the number of busy channels is equal to or less than the channel capacity of the UTRAN cell (i.e. number of busy channels $\leq$ m) experience the same low delay (approx 2ms for 16 kbps). However, above a threshold of cell loading, all channels are busy, and the queuing delay rises sharply to a higher value.

In WLAN, the delay settles at some constant value as soon as the network initializes (Fig. 7.1.2 at a Normalized Cell Load ~0). WLAN networks are rather prone to instabilities and are only stable within a narrow range of cell loadings. The MAC scheduling delay remains relatively constant within this stable region. This is so because, in order to maintain stability, the MAC protocol dynamically adjusts the packet retransmission rate to keep the collision rate within acceptable levels. The delay then depends only on the collision rate.

7.2 Scenario 2: Voice Delay in UTRAN and WLAN

This scenario is designed to generate comparative results for voice delay in both UTRAN and WLAN. The results are used in selecting which of the two networks can better meet the delay requirements for voice.

Table 7.2: RT Voice Delay in UTRAN and WLAN: A comparison (λ = 20 kb/s)

Normalized Cell Load	0.0	0.1	0.6	0.8	0.8	0.9	0.9	0.9	1.0		
Delay in ms: UTRAN	5.2	5.2	6.1	9.5	9.5	9.5	9.5	9.5	9.5		
Normalized Cell Load	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Delay in ms: WLAN	0.0	46.2	57.2	57.9	58.3	58.5	58.6	58.7	58.8	58.9	58.9

Fig 7.2 is a graph showing the relative voice delay performance of the proposed scheme in both UTRAN and WLAN networks.

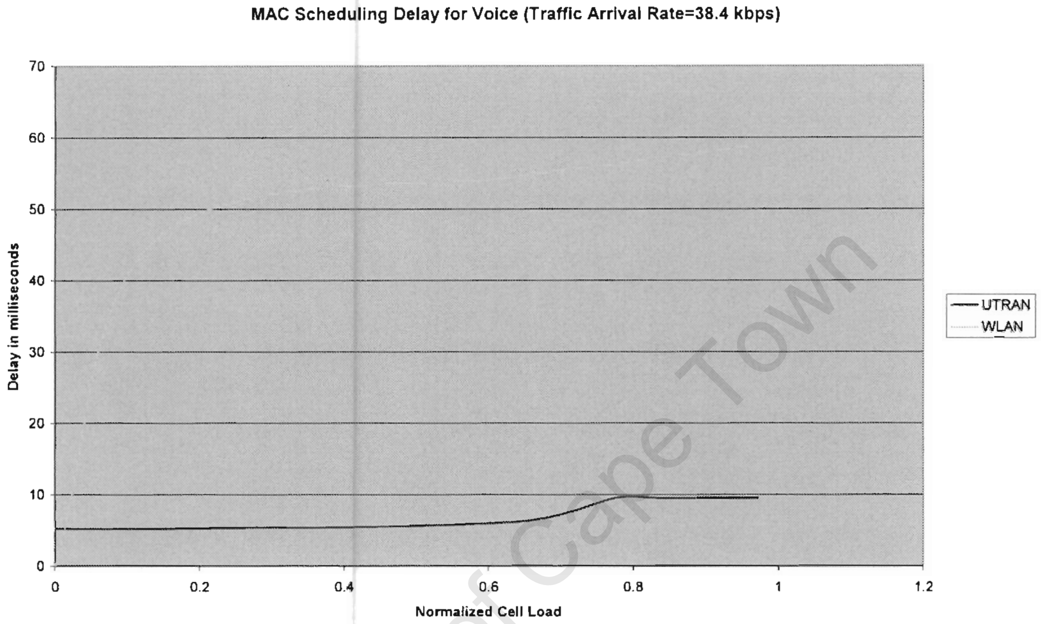

Fig 7.2 MAC Scheduling Delay for Voice in UTRAN and WLAN

Discussion of results

From Fig. 7.2, it is observed that the scheduling delay for voice in both UTRAN and WLAN is constant for most values of normalized cell load. However, the delay in UTRAN is always less than that of WLAN. The observations can be attributed to the fact in UTRAN, when voice traffic gets a free server (channel); the service will essentially be circuit switched (i.e. channel dedicated for entire duration of the session); hence bounded delay. The queuing delay obtained in WLAN is higher and this is consistent with the fact that the multiple access protocol in WLAN is contention based. Large medium access delays are usually associated with contention based channel access mechanisms.

The results obtained from the analytical results and simulation confirm that UTRAN is a better network for voice than WLAN, since the variable delay experienced is always less in

UTRAN than in WLAN. It can thus be concluded that the scheme can be used for network for selection.

7.3 Scenario 3: Throughput for variable bit rate data traffic

In scenario 1, real-time voice delay performance in both UTRAN and WLAN was investigated. Similarly, scenario 3 investigates the throughput performance of non real-time data traffic in both UTRAN and WLAN.

The throughput results for variable bit rate data are as in Tables 7.3.1 and 7.3.2 below.

Table 7.3.1: throughput for NRT Data in UTRAN (Nominal Data Rate = 38.4 kb/s; traffic arrival rates λ are in kb/s)

Normalized Cell Load		0.00	0.51	0.55	0.57	0.60	0.65	0.69	0.74	0.80	0.88	0.95
Throughput in kb/s	λ=16	16.0	15.8	15.6	15.3	15.0	14.5	13.8	13.0	11.9	10.4	8.5
	λ=20	20.0	19.8	19.5	19.2	18.7	18.1	17.3	16.3	14.9	13.0	10.6
	λ=38.4	38.4	38.0	37.5	36.8	35.9	34.8	33.2	31.2	28.6	25.0	20.4

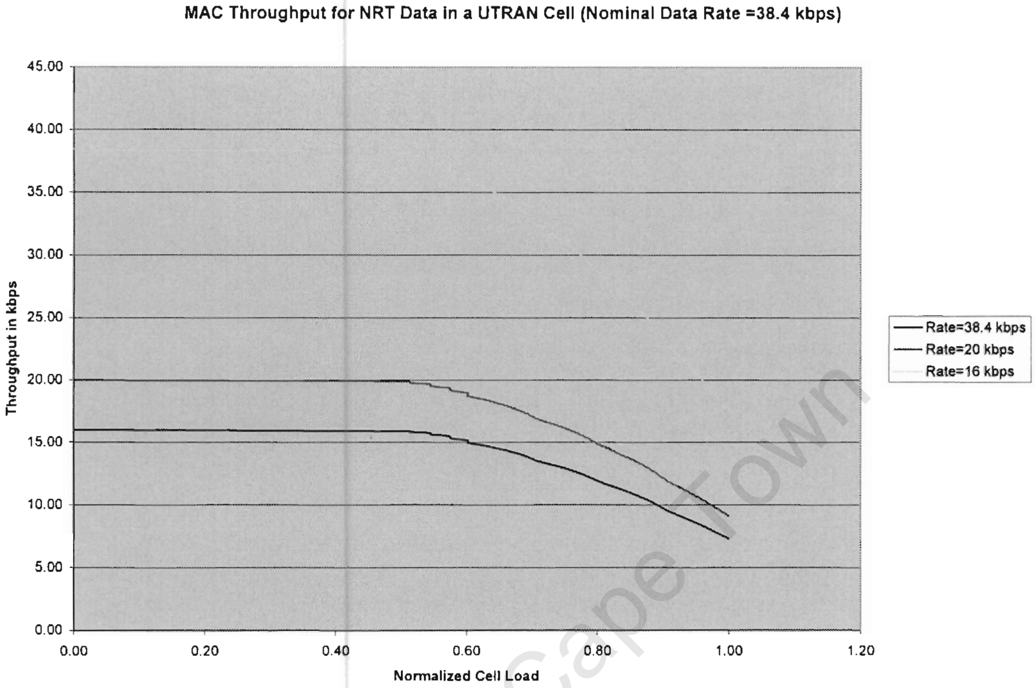

Fig 7.3.1 MAC Throughput for data in UTRAN

Table 7.3.2: Throughput for NRT Data in WLAN (CWmin = 15, AIFS = 20μs, TXOP = 3ms; traffic arrival rates λ are in kb/s)

Normalized Cell Load		0.02	0.12	0.22	0.3	0.41	0.5	0.61	0.7	0.8	0.9	1
Throughput in kb/s	λ=16	0.0	12.5	8.2	4.5	2.2	1.1	0.5	0.2	0.1	0.0	0.0
	λ=20	0.0	15.6	10.3	5.6	2.8	1.3	0.6	0.3	0.1	0.1	0.0
	λ=38.4	0.0	29.9	19.7	10.8	5.4	2.5	1.2	0.5	0.2	0.1	0.0
	λ=64	0.0	49.9	32.9	18.0	9.0	4.2	1.9	0.9	0.4	0.2	0.1
	λ=128	0.0	99.8	65.8	36.0	18.0	8.5	3.9	1.7	0.8	0.3	0.1

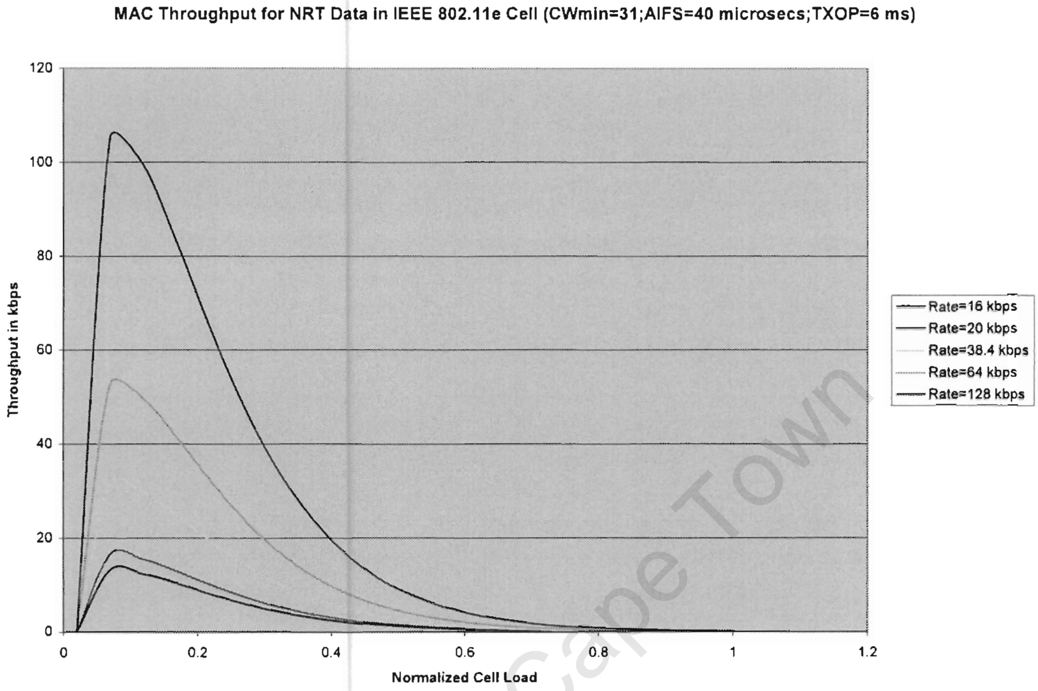

Fig 7.3.2 MAC Throughput for data in WLAN

Discussion of results

The results for UTRAN show that the throughput remains constant at a value close to the data traffic arrival rate when the cell loading is below a threshold of about 0.56. Above this threshold, the throughput drops precipitously to almost zero. To understand this observation, recall that data transmission in UTRAN follows the Hybrid Automatic Repeat reQuest (HARQ) scheme. A HARQ channel uses a contention based medium access protocol; consequently it has an associated stable operating point of cell loading. Above this stable point, the HARQ channel becomes unstable, retransmissions increase and the throughput decreases.

In WLAN, it is observed that throughput starts from zero, rises rapidly to a maximum value at a cell loading of about 0.1 and then drops precipitously to zero. It should be noted that in LAN networks (especially CSMA/CA/CD LANs), throughput is more a measure of transmission efficiency rather than quantity of data delivered [5]. At a cell loading of zero therefore, the transmission efficiency (read throughput) is zero because, accordingly, there is no data to transmit as there is no node transmitting. As the number of transmitting nodes increases (i.e. rising cell load), the throughput rises accordingly up to a peak value. After the peak, collisions become more prevalent, hence decreasing throughput.

7.4 Scenario 4: Data Throughput in UTRAN and WLAN

Scenario 4, like scenario 1 is a comparative study of UTRAN and WLAN with respect to throughput for non real-time data. The traffic arrival rate used in the experiment was 38.4 kb/s. The results of the experiments are presented below.

Table 7.4: Data Throughput in UTRAN and WLAN (λ = 38.4 kb/s)

Normalized Cell Load	0.00	0.51	0.55	0.57	0.60	0.65	0.69	0.74	0.80	0.88	0.95	1.00
Throughput: UTRAN in kb/s	36.5	36.1	35.6	35.0	34.1	33.0	31.6	29.7	27.1	23.8	19.4	16.6
Normalized Cell Load	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2
Throughput in kb/s: WLAN	n/a	n/a	37.54	37.38	37.51	37.62	37.76	38.04	38.14	37.96	37.54	26.39
Normalized Cell Load	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0				
Throughput in kb/s: WLAN	14.81	7.5	3.58	1.65	0.74	0.32	0.14	0.06				

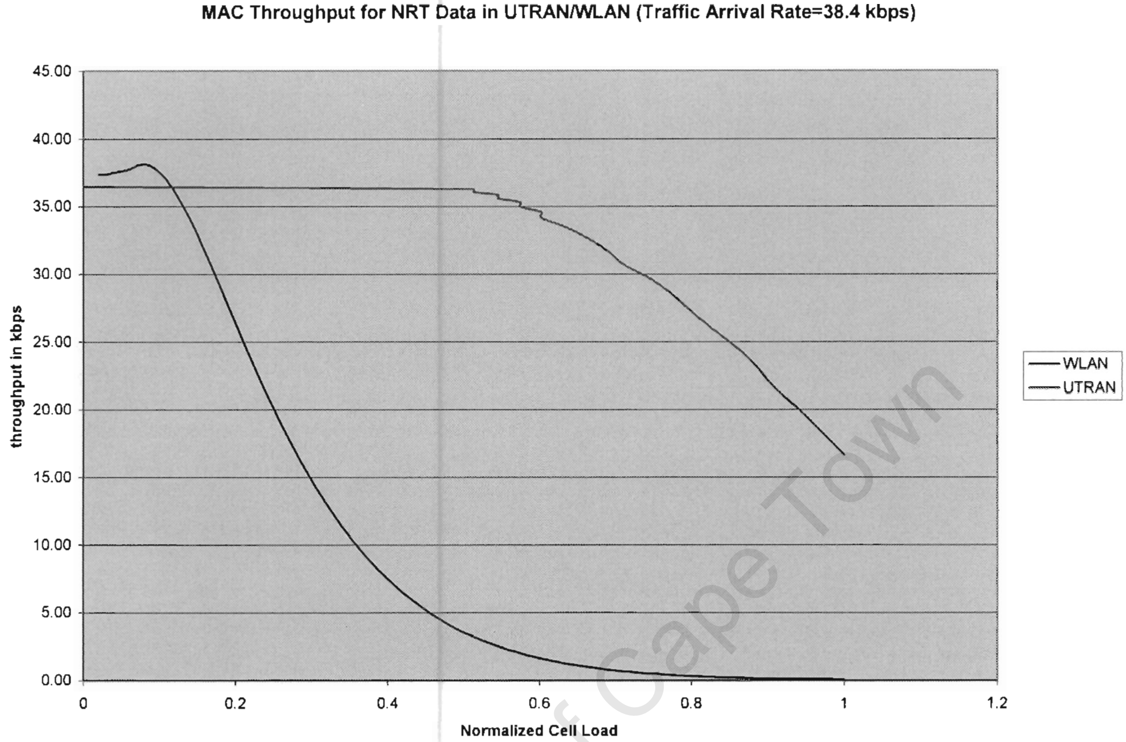

Fig 7.4 Comparison of MAC Throughput for data in UTRAN and WLAN

Discussion of results.

From figure 7.4, the data throughput in UTRAN for values of normalized cell load below 0.5 remains constant at approximately 36 kbps. The throughput drops rapidly towards zero as the normalized cell load increases. A normalized cell load of 0.5 corresponds to a threshold for interference level in the radio channel. Within this threshold, data throughput is constant because cell load is not a limiting factor to channel transmission efficiency. Above the threshold, however, the drop in throughput is due to increased HARQ retransmissions.

In the WLAN curve, the throughput at values of normalized cell load below 0.02 is undefined. According to Eq.5.33 in the model, the probability of successful transmission in an EDCA WLAN is obtained by optimizing the retransmission process; this process depends on the number of active nodes. This optimization procedure leads to a situation where the throughput obtained according to Eq.5.25 is undefined for small active node populations. An investigation to shed more light into this anomalous behaviour should be the subject of further work However, as the normalized cell load rises above 0.02, the throughput rises exponentially from 37.54 kbps to a peak throughput of 38 kbps at a normalized cell load of 0.08. The increase at this point can be attributed to the corresponding increase in the number of active nodes and high transmission success rate in WLAN cell. However, the throughput decreases to zero as the cell loading increases beyond 0.1. As discussed in [5], an increase in the number of active nodes beyond a threshold decreases the probability of successful packet transmission, hence a decrease in throughput.

From the results, it can be concluded that EDCA WLAN serves non real-time traffic better than UTRAN when the normalized cell loading is approximately 0.1. At other values though, UTRAN is gives better service.

7.5 Scenario 5: Validation through simulation: OPNET simulation vs. Analytical Model.

In an effort to validate the results from the analytical model, an OPNET simulation was carried out. The network and traffic configuration for the simulation scenario was identical to the configuration for the model. Table 7.5 shows the results for the experiment. Fig.7.5 is graphical presentation of the results.

Table 7.5: Delay for RT Voice in WLAN: Analytical vs. Simulation Results (CWmin=7, AIFS=2, TXOP=6ms; λ = 38.4 kb/s)

Analytical	Normalized Cell Load:	0.00	0.07	0.17	0.26	0.36	0.46	0.56	0.66	0.75	0.85	0.95	1.00
	Delay in ms:	0.0	8.8	9.1	9.2	9.2	9.2	9.2	9.2	9.2	9.3	9.3	9.3
Simulation	Normalized Cell Load:	0.1	0.1	0.2	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1.0
	Delay in ms:	0.0	2.0	4.4	6.0	6.8	6.8	6.8	7.0	6.9	6.9	7.3	9.9

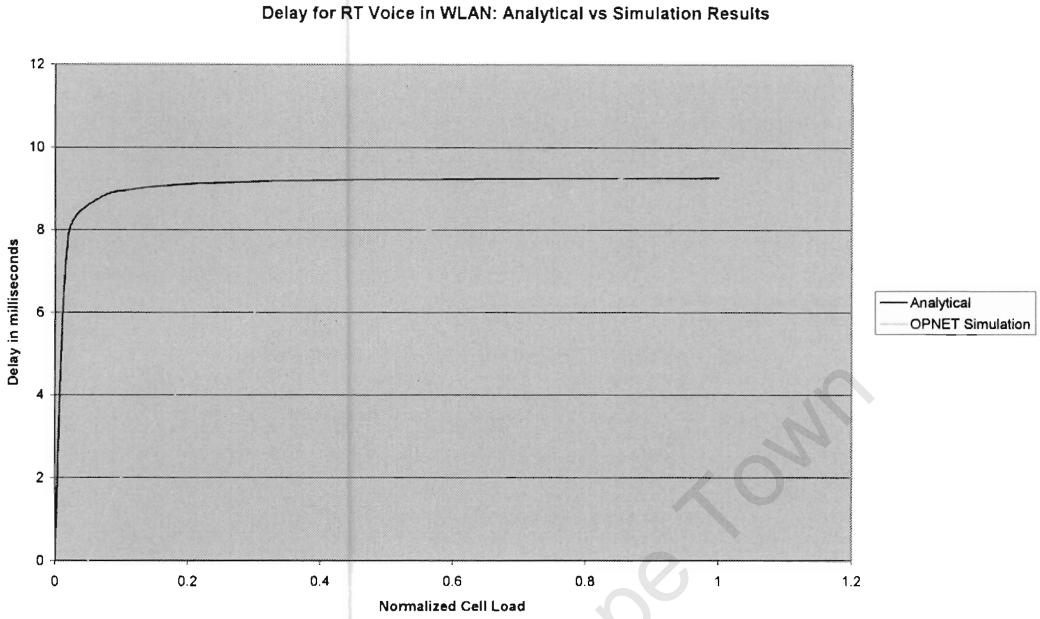

Fig 7.5 Delay for RT Voice in WLAN: Analytical vs Simulation Results

Discussion of Results

Both the analytical and random simulation results show that the MAC scheduling delay for real-time voice in WLAN increases with cell loading. The increase is faster at low values of cell loading, it tapers off at higher values. However, in the analytical scheme, the delay increases much faster and settles at a constant value earlier than the case for simulation experiments.

The model developed in the research utilised only one-dimensional queuing analysis; multi dimensional queuing analysis is more appropriate for multi-service traffic in 4G networks. This limitation of one dimensional analysis could explain the discrepancy between the analytical and random simulation results However, given that the general shape of the curves for both results are the same, one can conclude that both approaches are valid representations of the same phenomenon. Thus, despite discrepancies between the analytical and simulation results, the analytical scheme is a credible solution to the network selection problem. However, it can be improved further.

8 Conclusions and Suggestions

The objective of the research was to propose and evaluate a queuing model for QoS provisioning in 4G wireless access networks. The model would specifically be used for network selection. From the research accomplishments, the author makes a number of conclusions and suggestions.

8.1 Conclusions

There are many open issues in the development of 4G systems; QoS enabled network selection in the radio access network is one of them. Several research initiatives address this challenge. However, no standards have been developed and widely adopted.

The verification experiments reveal that the model can be used to predict the QoS performance of the wireless channel based on the type of traffic and traffic arrival rates. For example, it was observed that, as the traffic arrival rate increased, so was the channel access delay for both the real-time voice and non real-time data traffic. The variation depended also on the access network. This difference in performance between networks can be utilized for network selection.

Despite painstaking effort in developing a model, some anomalous behaviour can still be observed in analytical models. For example, in the model developed during the study, a situation arose where the throughput for very small values of normalized cell load in WLAN network is undefined. Such anomalies can be addressed in further investigations.

From the experiments, it was observed that in UTRAN there are large differences in the performance parameters between real-time voice and non real-time data traffic. In WLAN however, the differences are small. It can thus be concluded that UTRAN is optimized for real-time voice traffic while WLAN can support both traffic fairly well.

The curves of results from the analytical model and the random simulations exhibit a general trend. This shows that the assumptions made in constructing the model are credible. However, there are notable discrepancies between the two results. For example, the curve for the analytical result rises far too sharply before flattening out. The curve from the OPNET simulation rises more smoothly. This indicates significant difference between the two entities (i.e. the proposed and OPNET simulator). The difference may stem from either the theoretical basis of the models or the architecture of the experiments.

8.2 Recommendations

The analytical derivations were only performed for two QoS performance parameters, i.e. scheduling delay and throughput. Similar derivations can be attempted for the other QoS parameters of the wireless access network. Moreover, other wireless access networks other than UTRAN and WLAN can be considered in the studies.

4G systems will be multi-service networks in the sense that they will support multiple services. Therefore, a queuing model of a 4G network should be based on multi-dimensional queues, i.e. several simultaneous traffic queues with interdependent time evolution. Multi dimensional queue analysis was not performed in this research. In the development and subsequent testing of the analytical models, special attention should be paid to any apparent anomalous behaviour of the models.

Meticulous development of a theoretical model is just one component of a practical research endeavour. The evaluation of the model by well designed experiments is another very important component. It is thus necessary that such research enterprises be multi skilled, so that people bring their various skills and spin ideas off each other.

References

[1] 3GPP TS 22.105: "Services and Service Capabilities," "Technical Specifications Group".

[2] Akyilidiz, I.F. et al, "Medium Access Control Protocols for Multimedia Traffic in Wireless Networks," IEEE Trans. Networks, July/Aug. 1999, pp. 39-47.

[3] Aguiar, R.L., Sargento, S., Ruir, P., and Diogo Gomes. "Heterogeneous signaling framework for end-to-end QoS support in Next Generation Networks," Proceedings of the 38th Hawaii International Conference on System Sciences, Big Island, 2005.

[4] R. Agosti, O. Sallent, J. Perez-Romero, and L. Giupponi. "A fuzzy-neural based approach for Joint Radio Resource Management in a Beyond 3G framework," Proceedings of the First International Conference on Quality of Service in Heterogeneous Wireless/Wired Networks (QSHINE'04), Dallas, 2004.

[5] Bertsekas, D. and Gallager, R., Data Networks, Prentice Hall, New Jersey, 1987.

[6] Banchs, A., Perez-Costa, X., and Qiao, D., "Providing Throughput Guarantees in IEEE 802.11e Wireless LANs," in Proc.the 18th International Teletraffic Congress (ITC-18), Berlin, Germany, September,2003.

[7] Buehrer, M., Radio Resource Management Lecture, 12th Virginia Polytechnic Institute/MRPG Symposium on Wireless Personal Communications, June 5-7, 2002.

[8] Chan, P.M.L., Hu, Y.F., and Sheriff, R.E., "Implementation of Fuzzy Multiple Objective Decision Making Algorithm in a Heterogeneous Mobile Environment," Wireless Communications and Networking Conference, WCNC2002, pp. 332-336.

[9] Tianwei Chen, Sven Hermann, Gunter Shafer, "Secure QoS Enabled Mobility Support in All-IP Networks," Technical University of Berlin, TKN Technical Report TKN-04-013, Berlin, June 2003.

[10] ETSI TS 123 107, 'QoS Concepts and Architecture', V.4.3.0, 2001, http://etsi.org.

[11] Fapojuwo, A.O., "Radio Capacity of Direct Sequence Code Division Multiple Access Mobile Radio Systems," Proc. Inst. Elect. Eng.,vol.140, pp. 402–408, Oct. 1993.

[12] Granzow, W., '3G Mobile Communications Systems,' Ericsson Eurolab Mobile Communications Publications, Nordostpark, 2000.

[13] Gao, X., Wu, G., and Miki, T., 'End-to-End QoS Provisioning in Mobile Heterogeneous Networks,' IEEE Wireless Communications Magazine, June 2004.

[14] Hardy, W.C., "QoS Measurement and Evaluation of Telecommunications Networks," Wiley 2001.

[15] Havinga, P.J.M., Vognild, L., Wu, G., and Smit, G.J.M., "The SMART Project: Exploiting the heterogeneous mobile world,"

http://wwwhome.cs.utwente.nl/~havinga/papers/WI_IC2001.pdf .

[16] IEEE 802.11:1999 Part 11: "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," Reference Number ISO/IEC 8802-11:1999(E), 1999. Revised 2005.

[17] ITU-T Rec E.800, "Terms and Definitions Related to QoS and Network Performance including Dependability", Aug.1993.

[18] ITU-T Rec G.1000, "Communications Quality of Service: A Framework and Definitions," Nov.2001.

[19] Janevski, T., "Traffic Analysis and Design of Wireless IP Networks," Artech House, Norwood, 2003.

[20] Kurose, J. F.and Ross, K.W., "Computer Networking", Addison Wesley, New York, 2005.

[21] Lloyd-Evans, R., "QoS in integrated 3G Networks," Artech House, Norwood, 2002.

[22] Lundsjo, J. and Rinne, M. "UTRAN Transport Control Function," Third Generation Mobile Communication Systems, Artech House Publishers, Boston, 2000. eds. R. Prashad, W. Mohr, and W. Konhausser.

[23] Mangold, S, "Analysis of IEEE 802.11e and Application of Game Models for Support of Quality-of-Service in Coexisting Wireless Networks," Ph.D. thesis, ComNets Aachen University, June 2003.

[24] McKeown, N., "Network Algorithms Lecture 1," Stanford University, November, 2003.

[25] Mitola, J., "Software Radio Architectures," John Wiley and Sons, New York, 2000.

[26] Modante, E., "Data Networks Lecture 15," MIT Communications, November, 2002.

[27] Mohr, W., "Development of Mobile Communication Systems Beyond Third Generation," "Wireless Personal Communications", Vol.17, No.2, June 2001, pp 191-207.

[28] Muñoz, L., et al, "Empowering Next-Generation Wireless Personal Communication Networks," IEEE Wireless Communications Magazine, May 2004.

[29] Munoz, M. and Rubio, C.G., "A New Model for Service and Application Convergence in B3G/4G Networks," IEEE Wireless Communications Magazine, October 2004.

[30] van Nee, R. and Prasad, R., "OFDM for Wireless Multimedia Communications," Norwood, MA. Artech House, 2000.

[31] T. Ojanpera, "WCDMA" Third Generation Mobile Communication Systems," Artech House Publishers, Boston, 2000. eds R. Prashad, W. Mohr, and W. Konhausser.

[32] Papadimitriou, G. I., Pomportsis, A. S., Nicopolitidis, P., and Obaidat, M. S., "Wireless Networks," John Willey & Sons, West Sussex, 2003.

[33] Prasad, R. and Ruggieri, M., "Technology Trends in Wireless Communications," Norwood, MA., Artech House, 2003.

[34] Proakis, J.G., "Digital Communications," New York, McGraw Hill, 2001.

[35] Prasad, R. and Ruggieri, M., "Technology Trends in Wireless Communications", Norwood, MA., Artech House, 2003.

[36] Singh, M., Prakash, A., Anvekar, D.K., Kapoor, M., and Shorey, R. "Fuzzy Logic Based Handoff in Wireless Networks", Proceedings of the VTC 2000 Spring conference, pp. 2375-2379.

[37] Tipper, D., Dahlberg, T., Shin, H., and Charnsripinyo, C., "Providing Fault Tolerance in Wireless Access Networks", IEEE Communications Magazine, January 2002.

[38] Upkar Varshney and Radhika Jain, "Issues in emerging 4G wireless networks," Computer Communications Magazine, June 2001, pp. 94-96.

[39] Viterbi, A.M. and Viterbi, A.J., "Erlang Capacity of a Power Controlled CDMA system," IEEE J. Select. Areas Commun., vol.11, pp.892–900, Aug. 1993.

[40] Vrontis S et al "Enabling Distributed QoS management using active network technology," Seminar, National Technical University of Athens

[41] Yungsoo, K., et al, "Beyond 3G: Vision, Requirements, and Enabling Technologies," Topics in Wireless Communications", IEEE Communications Magazine, March 2003.

[42] Zigangirov, K., "Theory of Code Division Multiple Access Communication," Wiley-IEEE, New Jersey, 2004.

/43] Lindgren, A., Almquist, A., and Schelen, O., "Quality of service Schemes for IEEE 802.11 wireless LANs – An Evaluation," Mobile Networks and Applications Vol. 8(2003) No.3 pp.223-235, Kluwer Academic Publishers, Dordrecht 2003.

Appendix A: Analysis of Probability of Collision in systems using Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA).

Assume n active nodes and each node attempts transmission with probability p. Since transmission attempts can be considered as a series of Bernoulli trials, the attempts follow a geometric probability distribution.

Thus

Probability that k out of n active nodes attempt transmission is determined as

$$\text{P[k active nodes attempt]} = \binom{n}{k} p^k (1-p)^{n-k} \tag{A1}$$

Probability of successful transmission is equivalent to probability of exactly one transmission attempt, i.e.

$$\text{P[success]} = \text{P[k=1]} = np(1-p)^{n-1} \tag{A2}$$

In order to maximize success, differentiate A2 and equate the derivative to zero, i.e.

$$\frac{d}{dp}\left\{np(1-p)^{n-1}\ \right\} = n(1-p)^{n-1} - n(n-1)p(1-p)n - 2 = 0$$

$$\Rightarrow p_{optimal} = \frac{1}{n} \tag{A3}$$

$$\text{P[success]} = \text{P[k=1]} = \left(1-\frac{1}{n}\right)^{n-1} \approx \left(1-\frac{1}{n}\right)^{n} \tag{A4}$$

for $n << 1$

$$\text{P[idle]} = \text{P[k=0]} = \left(1-\frac{1}{n}\right)^{n} \tag{A5}$$

P[collision] = 1-P[idle] – P[success]

$$\Rightarrow p_{collision} = 1 - 2\left(1-\frac{1}{n}\right)^{n} \tag{A6}$$