QUALITY OF SERVICE AWARE TRAFFIC SCHEDULING ALGORITHM FOR HETEROGENEOUS WIRELESS NETWORKS

A Thesis submitted for the award of the degree of

Doctorate of Philosophy

By

Sonia

(950906038)

Under the Supervision of

Dr. Rajesh Khanna
Supervisor
Professor, ECED
TIET, Patiala

Dr. Neeraj Kumar
Supervisor
Associate Professor, CSED
TIET, Patiala

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Established under the section 3 of UGC Act, 1956)

PATIALA-147004 (PUNJAB)

CERTIFICATE

I, **Sonia,** hereby certify that the work which is being presented in this thesis entitled **"Quality of Service Aware Traffic Scheduling Algorithm for Heterogeneous Wireless Networks"** being submitted by me to Electronics and Communication Engineering Department, Thapar Institute of Engineering and Technology, Patiala in fulfillment of the requirement for the award of the degree of 'Doctor of Philosophy' is an authentic record of my own work carried under the supervision of **Dr. Rajesh Khanna** and **Dr. Neeraj Kumar.**

The matter embodied in this thesis to the best of my knowledge and belief has not been submitted to any other University/Institute for the award of any degree.

Date: 30-05-19

Sonia
(950906038)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

(Dr. Rajesh Khanna)
Professor
Department of Electronics & Communication Engineering
TIET, Patiala
Punjab. India
Dated: 30-05-19

(Dr. Neeraj Kumar)
Associate professor
Department of Computer Science & Engineering
TIET, Patiala
Punjab. India
Dated: 30-05-19

ACKNOWLEDGEMENTS

During my research work, many people affected and supported me personally and scientifically. I dedicate the following humble words to express my gratitude and to acknowledge their help each in their own way.

I would like to start by thanking my supervisor **Dr. Rajesh Khanna** for his expert advice and constant encouragements during the entire period of research work and for being available whenever I needed a meeting with him despite his busy schedule. It was a real honor to work under his supervision and to absorb some of his research methodology. He has provided a good balance of freedom and interest and has been a constant source of ideas, suggestions and recommendations.

Many thanks to my supervisor **Dr. Neeraj Kumar** for his precious guidance, advice and support regarding my research. He taught me the proper research methodology and how to present the work. I thank him for his kind guidance, criticism and advice throughout my research work.

I would like to thank the administration and staff of the University for their constant help. I express my thanks to all my friends and colleagues for their encouragement and support.

I am fortunate to have a loving family who supports me all the time and has faith in me. I dedicate this work to my parents-in-law who have taught me how to face challenges with patience and persistence. Finally, I take this opportunity to express my profound gratitude to my beloved parents for their invaluable love and support throughout the years. I cannot thank you enough for your prayers, unwavering support, encouragement, and for always believing in me.

The most importantly, I would like to thank my husband Mr. Amit Khullar for his love, encouragement and patience. His presence in my life is a great blessing. The love and affection of my son Shivansh is a great inspiration.

Finally, I am thankful to Lord Shiva, my success can only come from Him.

Sonia

ABSTRACT

Recent developments in multiple technologies evolving these days provide users with broadband services of high quality and seamless mobility. Wireless networks comprising of third generation (3G) and fourth generation (4G) provide wide coverage and good mobility capabilities. All these high quality wireless networks have been deployed within one region overlapping each other, hence generating a heterogeneous network for wireless access which can be represented as 4G heterogeneous network. The architecture of the system is such that it should support bulk data transfer. Although the Worldwide Interpretability for Microwave Access (WiMax) supports huge data transfer but still there is problem with the node request and service management. It becomes a sophisticated task for the nodes to handle number of requests at the same time. This work focuses on different aspects of the WiMax network and scheduling algorithms with hybrid architecture which is a combination of the First Come First Serve (FCFS) and Round Robin (RR) Algorithm. An analytical study of the number of jobs scheduled and completed along with accuracy for FCFS, RR and Hybrid algorithm is represented. The work is preceded by designing architecture of Long Term Evolution (LTE) which is a recently evolving technology which ensures the reliable delivery of the heterogeneous traffic services with high speed data rate and lower delays through their mobile and other hand held devices. The key feature of LTE is its traffic engineering which is used for effectively managing the network resources for efficient utilization. When LTE expertise has arisen there are some prevailing problems to be taken care of i.e. load balancing and traffic scheduling. Even LTE is the debauched technology, but it is also been anguishing from these problems. The work represents a Load Balancing Strategy that has been espoused based on traffic scheduling. Various Traffic scheduling algorithms are designed earlier to assign shared resources among users to optimize the performance of LTE systems in an efficient manner. The performance of three types of scheduling algorithms is compared in this work namely FCFS, RR and Hybrid algorithm which is the combination of FCFS and RR in extremes conditions. The discussed scheduling algorithms performance is measured in terms of Symbol error rate and Signal to noise ratio. The 4G heterogeneous network overlapping LTE and WiMax has been studied and a hybrid scheduling algorithm is designed and

implemented for the heterogeneous network in this research work. The three types of scheduling algorithms have been implemented on the heterogeneous network and compared. A new load balancing approach is proposed which is combination of FCFS, RR including priority scheduling. The proposed scheduling algorithm's performance is compared with other mentioned algorithms and measured in terms of bit error rate, signal to noise ratio and throughput analysis.

LTE and WiMax both suffer from load management and congestion as algorithms like FCFS, RR and priority scheduling are not sufficient enough to cope up with high demand network traffic. The memory buffer of serving architecture is not so big that it can compensate for all the node requests at the same time. As the requests are increasing, the scheduler is unable to manage the resources and providing services to all the users. Hence, task scheduling has been implemented to overcome the increasing traffic demands. The research work proposes a parallel scheduling algorithm inspired by a parallel structure which enhances the performance of scheduling in terms of certain QoS parameters as energy consumed, throughput and packet delivery ratio (PDR).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronyms	**Meaning**
1G	First Generation
2G	Second Generation
3G	Third Generations
3GPP	Third Generation Partnership Project
4G	Fourth Generations
5G	Fifth Generations
AAI	Advanced Air Interface
ABC	Always Best Connected
AMC	Adaptive Modulation and Coding
AMPS	Advance Mobile Phone System
AP	Access Point
aPS	Adaptive Polling Service
AWGN	Additive White Gaussian Noise
BE	Best Effort
BER	Bit Error Rate
BPSK	Binary PSK
BS	Base Station
BWA	Broadband Wireless Access
CAC	Call Admission Control
CCAC	Coordinated CAC
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CP	Cyclic Prefix
CPE	Customer Premises Equipment
CPU	Central Processing Unit
CQI	Channel Quality Indicator
DAG	Direct Acyclic Graph
DL	Downlink
EDCF	Enhanced Distributed Coordination Function
EDF	Earliest Deadline First

EU	Evolved User
EDGE	Enhanced Data rates for GSM Evolution
eNB	Evolved Node
ertPS	Extended Real Time services
E-UTRA	Evolved UMTS Terrestrial Radio Access
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
EPC	Evolved Packet Core
FCFS	First Come First Serve
FDMA	frequency Division Multiple Access
FTP	File Transfer Protocol
GO	Group Orthogonal
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
HSPA	High Speed Packet Access
HTTP	Hyper Text Transfer Protocol
IMT	International Mobile Telecommunications
IoT	Internet of Things
IP	Internet Protocol
ITU	International Telecommunications Union
LAN	Local Area network
LOS	Line of Sight
LTE	Long Term Evolution
LTE-A	LTE Advanced
MAC	Medium Access Control
MAP	Markovian Arrival process
MIMO	Multi Input Multi Output
MoM	Method of Moments
MVA	Mean Value Algorithm
MS	Mobile Station
NGN	Next Generation Network
NLOS	Non-line of Sight
nrtPS	Non-real Time Polling Service
NWG	Network Working Group

OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OSTBC	Orthogonal Space Time Block Codes
PAPR	Peak to Average Power Ratio
PDCCH	Physical Downlink Control Channel
PDN	Packet Data Network
PQ	Priority Queue
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoS	Quality of service
QPSK	Quadrature PSK
RAN	Radio Access Network
RR	Round Robin
rtPS	Real Time Polling Service
SAE	System Architecture Evolution
SC-FDMA	Single Carrier FDMA
SDU	Service Data Unit
SER	Symbol Error Rate
SINR	Signal to Interference Ratio
SJF	Shortest Job First
SNR	Signal to Noise Ratio
SOFDMA	Scalable OFDMA
SP	Strict Priority
SS	Subscriber Station
STC	Space Time Coding
TCP	Transmission Control Protocol
UDP	User Defined Protocol
UE	User Equipment
UGS	Unsolicited Grant Services
UMTS	Universal Mobile Telecommunications System
UL	Uplink
UTRAN	Universal Terrestrial RAN
VBR	Variable Bit rate

VHO	Vertical Handover
VoIP	Voice over IP
WAN	Wide Area Network
WFQ	Weighted Fair queue
WiMax	Worldwide interoperability for microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPAN	Wireless Personal Area Network
WRR	Weighted Round Robin
WWW	World Wide Web

Chapter 1

INTRODUCTION

1.1 Introduction

The communication world is greatly revolutionised by the growth of wireless technologies on a global scale and the broadband communication services have become the major necessity in this growing world of technology. Since the last two decades conventional cable infrastructure is out of date or simply saturated, recent wireless access technologies fill the gap exquisitely and faultlessly, providing systematic and cost effective services to millions of users. To provide efficient broadband services globally, the need of the hour is to integrate different technologies to achieve the desired results with good Quality of service (QoS) as per users' requirement as all the demands cannot be fulfilled by single access network or technology. The integration of more than one number of networks or technologies is represented as heterogeneous networks. One of today's issues is to design traffic scheduling algorithms for these integrated networks. Significant amount of work has been done for proposing number of scheduling algorithms for different type of integrated networks involving technologies such as GSM, UMTS, WLAN and WiMax, 3GPP and LTE [1]. Although existing cellular data services provide excellent connectivity in wide area, these service do not provide QoS to satisfy users.

As the communication industry is growing, the need and use of mobile devices is growing instantaneously. The access of high speed services is the requirement of users no matter how and where they want the access. The requirements are somewhat more than as provided by the traditional copper or optical cable technology. So a high speed broadband wireless access (BWA) is required to fulfill the user's demands. The WiFi has good compatibility with local area networks (LANs) but to cater with the wide area networks (WANs), WiMax can adjust very well. The licensed frequency band is allocated to WiMax to reduce the interference from other devices. The mobile WiMax is supported with features as: higher data rates, advanced QoS, scalable transmission and coding, security, mobility/ mobile IP, non line of sight (NOS) and smart antennas.

LTE framework helps an incredible number of portable social contact administrations. An essential gimmick of interpersonal organizations is the need of the same activity substance by more than one terminal. Along these requirements, remote multicast can possibly help portable social contacts in view of its remarkable asset productivity for transmitting bundles from a solitary sender to numerous beneficiaries at practically the same remote asset as unicast. Notwithstanding, remote multicast execution is limited by the terminal with the most exceedingly awful channel condition. Be mindful of the rare range asset, asset designation for remote multicast is still open for more research.

LTE framework uses gadget to gadget innovation which has the capacity to enhance productivity in the cell arranged by building immediate connections between terminals without sending in the other cell. In the interim, wireless radio is an alternate great innovation in enhancing range asset use, which gives more potential range asset by optional client with legitimately sensing the range conditions and looking to overlay its flag with those of the essential clients without meddling with them.

The future cellular radio networks will empower all wireless networks to commune in a domain constituted by a number of devices and technologies using several radio interfaces. For this environment, one essential task is to generate common control functions for all required applications. To increase the coverage and capacity of cellular networks, the concept of heterogeneous networks came into existence. When the load of one network is shared by other networks available with in that region and fulfills all the requirements to satisfy QoS needs is represented as heterogeneous network. Various networks differ mainly in terms of capacity, coverage area, transmit power, signal strength and cost of deployment. Besides, heterogeneous networks play an important role in WiMax and LTE networks and certainly will do so in the upcoming 5G networks as shown in Fig. 1.1. In this research work, the main emphasis

Figure 1.1: Concept of 5G Heterogeneous Networks [2]

is given on traffic scheduling algorithm in heterogeneous wireless network environment by considering some QoS attributes. There are many scheduling algorithms present for achieving the QoS in 4G heterogeneous networks. The main objective of this work is to develop hybrid approach for scheduling in a 4G heterogeneous network.

1.2 Motivation for Research

BWA systems and devices have been utilized globally since last so many years. These devices communicate based upon different standards for communication developed by IEEE. The standards may be categorized on the basis of either the technology used or the sequence of generations with which these standards are available for access.

In around 1980s, analog mobile phone networks were introduced represented as first generation (1G) of cellular networks and were known as Advance Mobile Phone System (AMPS). Second generation (2G) for cellular networks came into existence in early 1990s designed as Global System for Mobile Communication (GSM) was a great revolution towards digital system design. Later on the 2G networks were extended in data rates and services, introduced the 2.5G as Enhanced Data rates for GSM Evolution (EDGE) and General Packet radio Service (GPRS).

The third generation (3G) of networks were able to support higher data rate services like wireless broadband connectivity, wireless digital television, video conferencing and many more services with the introduction of Universal Mobile Telecommunication System (UMTS) and CDMA 2000. The goal of global coverage is almost achieved with 3G cellular technology.

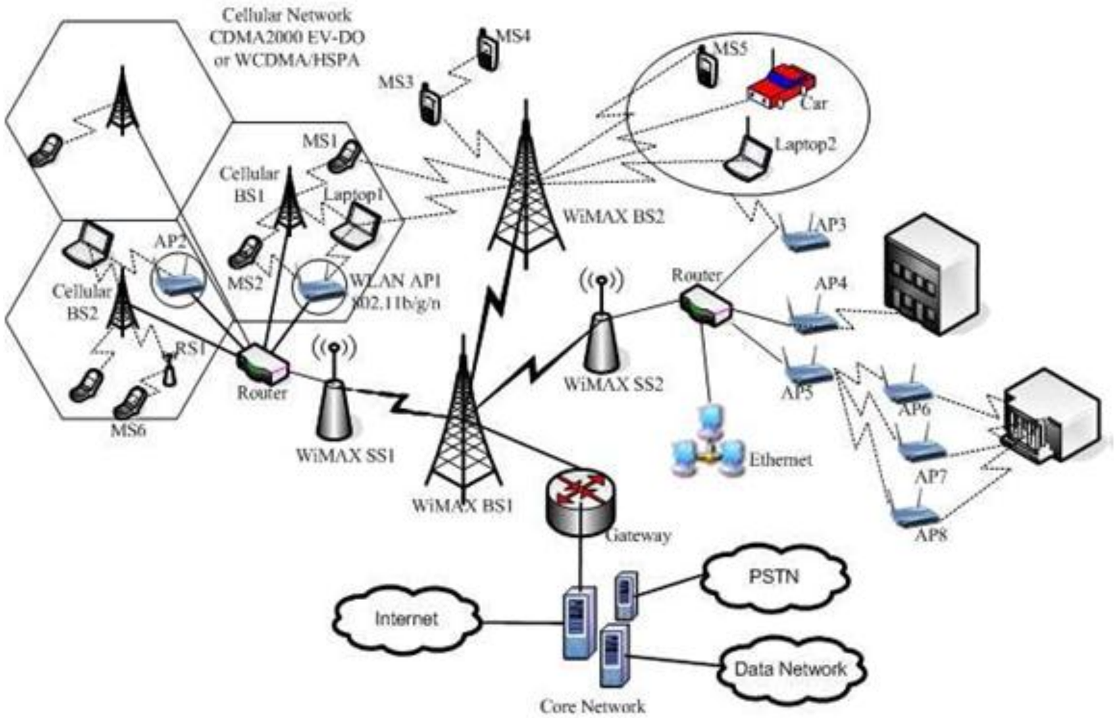

Figure 1.2: An Example of Heterogeneous Network [3]

The next generation of cellular networks currently being deployed is fourth generation (4G) and able to fill the gap among various technologies with higher data rates. 4G is an all IP based technology working with multiple air interfaces, fixed and mobile heterogeneous broadband networks and multiple number of devices simultaneously providing users and networks with an Always Best Connected (ABC) facility [4], high QoS and lower delays with great broadband experience as shown in Fig. 1.2. 4G is considered to be a merging platform provisioning upper hands in terms of coverage area, bandwidth availability and energy consumed. 4G will provide global roaming and seamless mobility with various access technologies such as GSM, WiFi, UMTS, WiMax, and Digital Video Broadcasting.

WiMax and LTE are considered as 4G wireless technologies chosen for IMT-Advanced certification [5]. IEEE 802.16 is one of the wireless broadband standards is entitled as WiMax is a single network has the ability to provide voice, data, video along with mobility using an individual approach. In Next Generation Networks (NGNs), WiMax has been proposed to be a promising technology for fixed and mobile connectivity for a great number of real time and non-real time applications. Same way, as approved by International Telecommunications Union (ITU), LTE is another candidate for 4G system for wider mobile applications. The technology is supported by most of the operators and the invention of LTE has revolutionized the goal of telecommunication industry. The major facilities provided by LTE are flexible bandwidth deployments, higher data rate, reduced latency and packet optimization. The support of packet switched traffic with high mobility and great QoS is one of the main goals of LTE network architecture.

1.3 Goals and Challenges of Research

With the development of standards, the new technology and specifications have been generated to make the standard accessible with new applications developed by the enterprises with the expansion of market trends and demand analysis. Now days, WiMAX technology has converged the market trends with air interface specifications, making the seamless mobility a truth with great supportive applications and devices. To provide high speed communication services in rural area, the WiMAX has been proposed as a cost effective alternate as compared to wired connections (e.g. cable or DSL). The last mile connectivity with higher speed is not the only purpose of WiMAX systems instead it allows the user to move freely during transmission with no data loss when handover occurs from one base station (BS) to another. The handover is easily achieved in WiMAX due to high bandwidth. Generally, WiMAX BS antennas are installed at a great height to support wide range as there are no barriers between the transmitting and receiving points. The line of sight (LOS) links in WiMAX can support a broader range of approximately 30 to 50 kilometres and 3 to

10 kilometres for non-line of sight (NLOS). Fig. 1.3 represents the common devices like notebook, internet connection at residences and Wi-Fi hotspots connected with the help of WiMAX network.

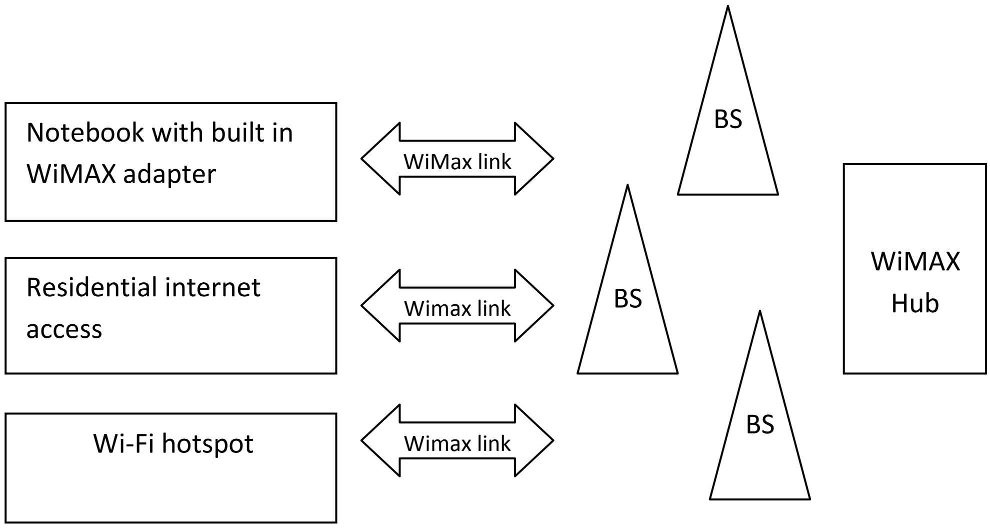

Figure 1.3: Network using WiMAX

The LTE technology is based on the methodology of EDGE/GSM and HSPA/UMTS. It is the core network improvement which enables to have an increased speed and capacity using a different radio interface. Orthogonal frequency division multiplexing (OFDM) is deployed in LTE for better network performance [6]. Long-term evolution is a standardized system which is used for high-speed wireless communication for various types of mobile devices and terminals which are based on the UMTS and GSM technologies. The revolutionary phases of LTE are represented in Fig. 1.4.

LTE mainly has its peak download rates up to 326.4 Mbit/s and its upload rates as for 86.4 Mbit/s but its bit rate depending on the user equipment category. On the other hand, it uses 4x4 antennas in the case of download rates, but it utilizes a single antenna for uploading. LTE support the services at low cost. LTE has five different classes which have been defined by low data rate voice service to high data rate end terminals to support all types of services. Due to this process, all terminals are able to process 20 MHz bandwidth.

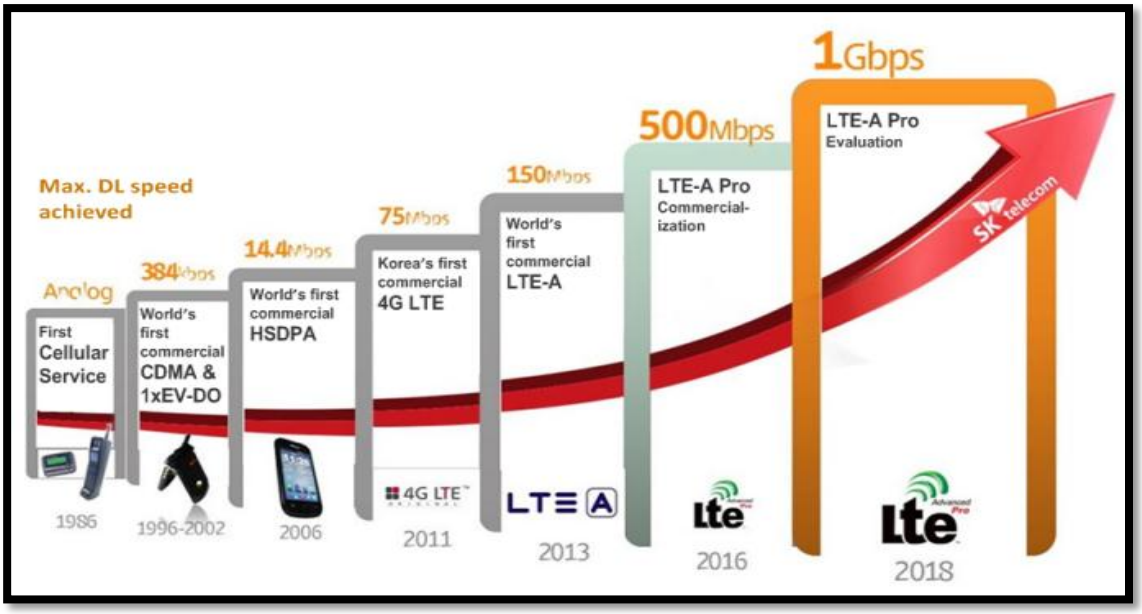

Figure 1.4: Long Term Evolution [6]

LTE network uses IP architecture due which it is dedicated to packet switched operations as well as it also supports a high bandwidth to deliver the good voice signal. This signal can be transported by using voice over LTE protocols and then it falls back to legacy networks that are within 3G and 2G. LTE also supports multi input multi output (MIMO) technology which always gives higher data rate for multiple connections of devices. The main advantage of LTE is that it is commonly linked with OFDMA in the downlink which automatically helps to utilize the channel resources effectively. It also increases the total number of user capacity for different channels to access the system.

LTE handoff is smooth that allows users to connect to devices in other regions. LTE always supports higher versions of LTE standard based products due to which they give a high range of facilities to its users [7]. It also results in improving the performance of the system. With the help of LTE, users are also able to access faster connection speeds and a greater level of coverage related to the internet. It provides enhancements regarding the speed, capacity and reliability which will automatically help to make mobile broadband too feasible for its users.

The basic idea to develop a standard technology was not only to design a new network to generate new services in multi hop environment but to impart a composition to enhance ongoing single-hop standards to multi-hop. Although the two technologies have been installed in many countries yet there are lots of improvements to be done in some scenarios on performance basis. The fixed location of BS with LOS and NLOS coverage is one of the challenges as it hurdles the throughput and latency requirements. The performance is poor inside the building, tunnels and subways. Hence, some service solutions must be provided by the end users to manage the issues. In case of large traffic in some areas, additional equipments may be provided to route the traffic towards destination. The vehicular network support is again one of the challenges for telecommunication companies.

1.4 Quality of Service (QoS)

The quality of service describes the quality of a network or in other words, it shows how effectively network manages the resources with minimum waiting time and maximum outcomes. The main goal of any network is to provide available resources among all the users, with minimum delays to get the appropriate results. The standard defined for WiMax is differentiated into five traffic classes categorised as Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non real-time Polling Service (nrtPS), Extended Real-time Polling Service (ertPS) and Best Effort (BE). The application of each class is different [8]. The working of each class in detail is as under:

UGS (Unsolicited Grant Service): It is the first class of IEEE802.16 QoS standard. It supports the application of VoIP (voice over internet protocol). All the incoming packets of VoIP can be given to this class by the classifier. It generates fixed size data packets. The advantage of allocating the fixed data size from the base station is that if there is any request from SS given to the BS, the base station will reject this request.

rtPS(real-time Polling Service): It is the second class of IEEE802.16 QoS standard. It supports the application of MPEG video. All the incoming packets of MPEG video can be given to this class by classifier. It supports the real time application. It generates the variable size of packets at different time periods. These packets are the MPEG packets with less suppression. In this class if SS require bandwidth grant from the base station, it will grant the request and allocate the required bandwidth.

nrtPS(non real-time Polling Service): It is the third class of IEEE802.16 standard. It supports the application of FTP (file transfer Protocol). All the incoming packets of FTP can be given to this class by the classifier. It supports non real time applications which need variable grants of different data size on regular basis. The throughput requirement of this class is high but delay can be accepted by this class. In this class if subscriber station (SS) require bandwidth grant from BS, the BS can grant the request of bandwidth for longer interval. The base station can allocate the bandwidth even in the high traffic.

ertPS (Extended Real-time Polling Service): It is the fourth class of IEEE802.16 standard. It supports the application of MPEG and VoIP, In other words we can say that it is combination of the UGS and rtPS. All the incoming packets of VoIP and MPEG Video can be given to this class by the classifier. As it is combination of UGS and rtPS it offers real time service. It also generates the packets of variable size on regular time period. In this case any latency introduced by bandwidth can be removed easily.

BE (Best effort): It is fifth class of IEEE802.16 standard. It supports the application of web browsing that is WWW (World Wide Web). All incoming packets of WWW can be given to this class. If there is much traffic then the packets of best effort will take long time for transmission.

1.5 Scheduling Process

Scheduling is a process to schedule many tasks. The scheduling helps a person to do many tasks so that time is efficiently utilized. The advantage of scheduling is that it saves time, increase efficiency, task ends faster with less energy consumption. The similar concept is launched in broadband internet when huge traffic of packets is to be scheduled in a manner that the speed of jobs can be faster. So in other words we can say that scheduling is basically a process of management. An example of scheduling is that brain is ultimate process manager. Nature gives a power to the brain to manage a task in a proper way. The Fig. 1.5 shows that brain is an ultimate process manager.

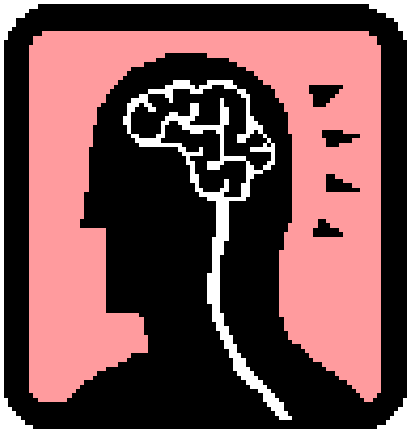

Figure 1.5: A Brain is a Ultimate Scheduler

As in previous years the demand of high speed broadband wireless internet increased day by day, it is necessary to create such a system which helps to meet the demand of high speed internet. So there is a need of implementing such a scheduling algorithm which helps to increase the speed of internet or in other words we can say that scheduling of packets will help to adjust the execution time of jobs. There are certain parameters which help to examine the scheduling process are explained as follows:

Throughput: The throughput is a parameter which gives the information for how many processes complete their execution per unit time. In case of NGNs, there are number of jobs that are present in a ready queue. So throughput is represented as how many ready jobs complete their execution per unit time. A Scheduling Algorithm said to be efficient if the throughput should be as high as possible.

Turnaround time: In general the time taken for submission of program plus execution and getting the output. But if can we take turnaround time in scheduling then it the time between when a process is submitted as a starting time and when the process is completed as a end time. So the turnaround time is the time between the process is submitted and its completion. It should be minimized.

Fairness: In general fairness as the name indicates equality or a true value. The fairness in scheduling is the equal time distribution among all incoming packets. For a scheduling algorithm it is necessary that equal time is distributed among all priority queues. So in other words we can say that fairness is the time equality among all priority queues. A scheduling algorithm is said to be efficient if the fairness is as high as possible.

Efficiency: The application of scheduler comes into the use when there is more than one process in the ready queue. The scheduler efficiency is the parameter which shows the involvement of processor as much as possible when traffic is high. A scheduling algorithm said to be efficient if the efficiency is as high as possible.

Waiting time: In general waiting time is the time that a job has to wait for a time or in other words we can say that the total time for which a job can wait. In case of scheduling, it is time for which processes or jobs can be waiting in the ready queue. In the process of Scheduling, all incoming packets are given to the classifier where they can be prioritized. After the classifier all packets are placed in the queues. A scheduling algorithm is said to be efficient if the waiting time is as low as possible.

Response time: It is the time that a job is responded by the connected device. In case of scheduling, it is the time taken by the scheduler to response to the request given by the packet. It should be as low as possible. Or in other words we can say that it the time calculated from when the request is submitted and first response produced by the scheduler. A scheduling algorithm is said to be efficient if the response time is as low as possible.

Latency: The latency is basically refers to delay. The delay occurs for transmission of packets from one place to other. The latency is also referred to as packet is transmitted from sender and how much time it takes to return back to the sender is called round trip latency. There are number of parameter by which latency can occur. These parameters are propagation delay, transmission delay, storage delay and processing delay. A scheduling algorithm is said to be efficient if the latency is as low as possible. Ideally it should be zero.

The incoming traffic is categorized into five service classes as described earlier. Every service class is accountable for different application. The application of UGS is to support Voice over IP packets, the application of Real-time Polling Services (rtPS) is to support MPEG video, the application of Non Real-time Polling Services (nrtPS) is to support packets of File Transfer Protocol (FTP), the application of Best Effort (BE) is to support the packets of Web browsing, data transfer and the application of Extended Real-time Polling Service (ertPS) is to support the packets of Voice with activity detection (VOIP).Once the packets have been classified according to the standard, the packets are set down in priority queue. The queue will arrange these packets in high priority to low priority. The general consideration of queues defined is high priority queue, medium priority queue, normal priority queue and low priority queue. Then the scheduler is connected next to schedule the services according to priority assigned to each process. Fig. 1.6 shows the functional block diagram of scheduler. All the incoming traffic packets are given to the classifier to assign the priority to each service class. The classified traffic is forwarded to multi-priority queue and then to scheduler [9]. The scheduler accomplishes all incoming traffic packets as per priority settings.

1.5.1 Applications of Scheduling

Broadband internet application: The scheduling can be possible only if there is scheduling algorithm present in the scheduler. To control the overall traffic on IP network, the scheduling plays an important role.

CPU application: The scheduling algorithms are also implemented in the CPU process to manage the commands given to the processor. By implementing the scheduling algorithm the speed of processor can be increased and waiting time for jobs can be reduced.

Managing traffic Packets: A Scheduling Algorithm can manage the traffic of different forms of packets with different priority. It executes the jobs present in the ready queue in a managed way.

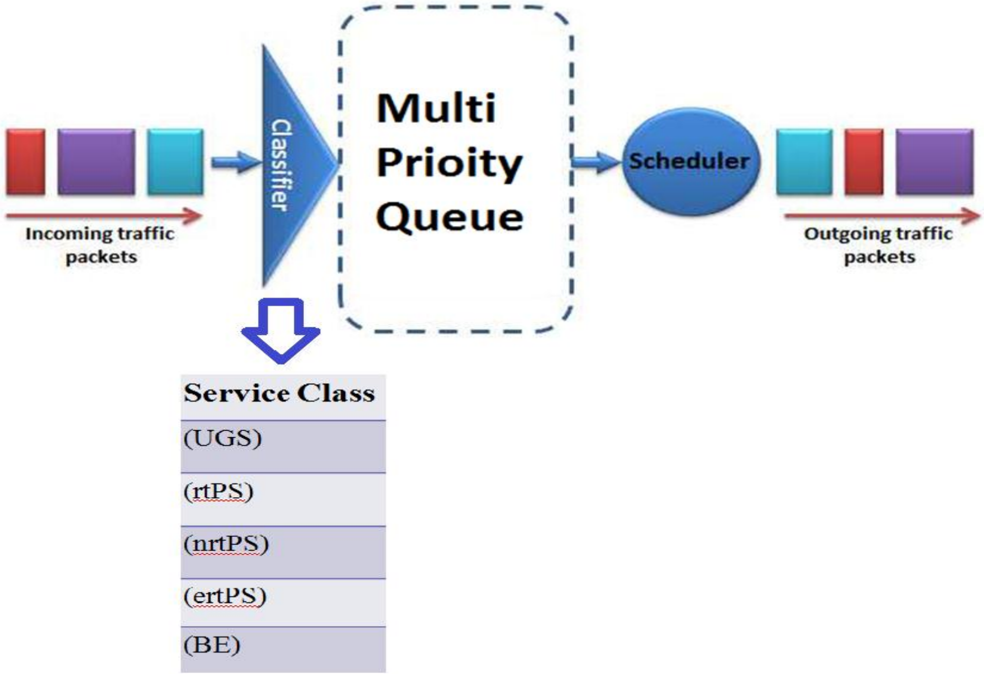

Figure 1.6: Functional Block Diagram of Scheduler [51]

Desktop grids: The scheduling algorithms are also implemented on the desktop grids to manage the BoT (bag of task) applications. On desktop grids they can manage the traffic of various task applications. By implementing the scheduling algorithms the turnaround time can be increased up to some extent.

Grid computing: A scheduling algorithms can also be implemented in the area of grid computing. Grid computing is basically the bundle of computers connected in different location is connected to a common node to achieve a common goal. In this process a number of inputs are coming from different location computers. These computers are connected to a common node. A scheduling algorithm is implemented in that node to manage the inputs from different computers so that a goal can be achieved with fast speed.

Liner workflow optimization: A scheduling algorithms can also be implemented in liner workflow optimization. In this workflow the computing, mapping and then scheduling can be done so that execution of workflow can be done on parallel machines.

Multi programming for processor: A Scheduling Algorithm can also be implemented for multiprogramming for processor. An advantage of multiprogramming is that it helps to full utilization of processor and also it helps to schedule the various tasks of processor.

1.5.2 Scheduling Algorithm Objective

Most scheduling algorithms are application specific and have different applications requirements. The following main objectives are considered in the implementation of the scheduling algorithm.

Maximum Throughput: A throughput is a parameter which gives the information for how many processes complete their execution per unit time. In case of heterogeneous networks, there are number of jobs that are present in a ready queue. So throughput in that case is how many ready jobs complete their execution per unit time. Throughput should be maximized. A scheduling algorithm said to be efficient if the throughput is as high as possible.

Minimum Turnaround Time: In general the time taken for submission of program plus execution and getting the output and transferring it to the user. The turnaround time considered in scheduling is the time between when a process is submitted as a starting time and when the process is completed as end time. So the turnaround time is the time between the process is submitted and its completion. It should be minimized.

Maximum Fairness: In general fairness as the name indicates equality or a true value. In case of Scheduling, the fairness is the equal time distribution among all incoming packets. For a scheduling algorithm it is necessary that equal time is distributed among all priority queues. So in other words we can say that fairness is the time equality among all priority queues. A Scheduling Algorithm said to be efficient if the Fairness should be as high as possible Fairness should be maximized.

Maximum Efficiency: Efficiency in general is the efficient working of a device. The scheduler efficiency is the parameter which shows the involvement of processor as much as possible when traffic of packets is more. This is because the application of scheduler comes into the use when there are more than one processes in the ready queue. A scheduling algorithm is said to be efficient if the efficiency is as high as possible. Efficiency should be maximized.

Minimum Waiting Time: In general waiting time is the time that a job has to wait. In other words it is the total time for which a job can wait. Considering waiting time in case of scheduling, it is the duration for which processes or jobs can be waiting in the ready queue. All in coming packets are given to the classifier where they can be prioritized. After the classifier all packets are placed in the queues to wait for processing. A scheduling algorithm is said to be efficient if the waiting time is as low as possible.

Maximum Response Time: In general the response time is the time that a job is responded by the connected device. It is the time taken by the scheduler to respond to the request sent by the packet. It should be as low as possible. It the time calculated from when the request is submitted and first response is produced by the scheduler. A scheduling algorithm is said to be efficient if the response time is as low as possible.

Minimum Latency: The latency is basically a delay. In scheduling it is referred to as the delay occurring for transmission of packets from one place to other. The latency is also referred to as the time taken by a packet to be transmitted from sender and time it takes to return back to the sender is called round trip latency. There are number of parameter for latency to occur. These are propagation delay, transmission delay, storage delay and processing delay. A scheduling algorithm said to be efficient if the latency is as low as possible. Ideally it should be zero.

1.5.3 Challenges for design of Scheduling Algorithm

The design of Scheduling Algorithm for WiMax is challenging because of several constraints. The Scheduling Algorithm suffers from throughput, turnaround time, waiting time, fairness, response time, latency, efficiency and QoS. The design challenges for scheduling algorithm in networks involve the following main aspects:

Large Traffic of Nodes: As the demand of internet services is increased, the traffic on nodes has also increased. So it is a challenge to manage the traffic and provide high QoS to all nodes. While designing the scheduling algorithm, it is necessary that it may able to manage all traffic on nodes.

Demand of BWA: As the demand of BWA increased with high speed, designing an algorithm must include the speed of execution of packets as high as possible.

Bandwidth Management: As the request of allocation of bandwidth can be send by any of node within the network so it is also a very challenging issue while designing a scheduling algorithm that equal bandwidth can be provided to each queue.

Deadline: Scheduling schemes can be classified based on the deadline of arrival of data packets to BS. So it is also a very challenging issue while designing a scheduling algorithm.

Packet Latency: The packet latency is often referred to as packet delay or the time from the source sending a packet to the destination receiving it. So it is also a very challenging issue while designing a scheduling algorithm that the packet latency should as low as possible.

Starvation Problem: In scheduling, the tasks can be divided into queue. The high priority queue served first and lower priority queue will suffer the problem of starvation. So it also a very challenging issue while designing a scheduling algorithm that starvation problem can be minimized.

Waiting Time: As there is a problem of starvation the jobs or processes need to wait for long time for execution. Or in other words the time at which the process remain in ready queue. The required waiting time should as low as possible.

1.5.4 Categories of Scheduling Algorithms

Scheduling algorithms can be executed in two ways. The first case represents the execution process without any interruption. The second case can interrupt the ongoing execution process if some high priority queue is placed in the ready queue [10]. Based on the two execution ways, there are two types of scheduling algorithms:

- Non Preemptive Scheduling Algorithms
- Preemptive Scheduling Algorithms

The algorithms can be examined on the basis of different parameters. These parameters are waiting time, response time, throughput, Fairness, efficiency and latency. The aim of studying all the algorithms is to design new approach which will overcome the problems faced by the previous algorithms. Table 1.1 describes the comparison of two types of algorithms.

Table 1.1: Comparison of Preemptive Vs Non preemptive Scheduling Algorithms

Non preemptive Scheduling Algorithms	**Preemptive Scheduling Algorithms**
In the scheduler, the tasks are assigned with priorities. Some of tasks are with higher priorities and they are executed first and other which are with less priority are executed later. In Non preemptive scheduling algorithms the higher priority task cannot disrupt the ongoing process or in other words we can say that in non-preemptive Scheduling Algorithms the running task cannot be interrupted until the task which is going on executed first.	In Preemptive scheduling algorithms, if a task of higher priority comes in between the ongoing process, the task with higher priority is executed firstly by interrupting the ongoing process. In other words we can say that in preemptive scheduling algorithms the running task can be interrupted so that high priority cab is executed first.
Types: • First come First Serve (FCFS) • Shortest Job First (SJF) • Strict priority (SP)	Types: • Round Robin (RR) • Weighted Round Robin (WRR)
Waiting time: The waiting time will be more for shorter jobs	Waiting time: The waiting time will be less for shorter jobs
Efficiency: The efficiency of this algorithm is more.	Efficiency: The efficiency of this algorithm is less.

1.6 Problem Statement

Different wireless networks and technologies are available now days to fulfill the demands of users as per requirement. Based upon different standards, the technologies may able to communicate among various devices to provide required QoS. But when the users are moving at high speeds or the data rates required are higher, the present
standards are not compatible with the requirements. The IEEE 802.16 standard is supposed to provide a wide-range broadband wireless service for both LOS and NLOS services for fixed and mobile scenarios. The IEEE 802.11e standard contains Enhanced Distributed Coordination Function (EDCF) which provides QoS at lower data rates. To enhance the QoS functioning, the 802.16e standard is defined to coordinate the devices moving at high speed and with high data rates. A more flexible and efficient QoS frame work has been described by 802.16m for supporting high data rate mobile applications along with a advances air interface (AAI) for QoS scheduling, less delays and priority access [11]. A number of scheduling schemes have been addressed in different research works to meet QoS requirements in WiMax networks.

With the investigation of UMTS, Universal Terrestrial Radio Access Network (UTRAN), the research is followed by LTE based on 3GPP to identify the requirements from different users. The aim of LTE is to generate high performance BWA technology to offer high mobility and high data rates as well as coexistence with the already existing networks. With the invent of LTE, the concept of heterogeneity is developed and extended. The LTE framework has been extended to include higher order MIMO/OFDMA in heterogeneous networks. The WiMax and LTE are the two most promising 4G technologies to provide support for service competitiveness, cost satisfaction and variety of devices to be supportive with. Careful planning is required to combine two technologies to design a heterogeneous network.

Scheduling in heterogeneous networks is also the current topic of research as the demand of high speed, high bandwidth for multiple and simultaneous transmissions are growing day by day. The requirement of scheduling algorithm is to support different types of traffic streams in heterogeneous networks. Various researches proposed different scheduling algorithms to support the QoS guaranteed by the User. None of them is able to address the QoS scheduling requirements for different service types and no analytical work has been developed to improve the performance of the scheduling algorithms.

To make compatibility with the increasing demand of users, wireless communication technology companies are forced to review their standards, strategies and products. To provide needed capacity, coverage as well as higher QoS, the standard developing companies are continuing to work towards developing new technologies or to enhance the features of existing technologies individually or in hybrid architecture form.

1.7 Research Objectives

The main objective of this thesis was to study, design and simulate various scheduling algorithms for heterogeneous wireless communication networks. In this thesis work, a hybrid scheduling algorithm separately for WiMax and LTE networks have been proposed. A heterogeneous wireless network is considered consisting of WiMax and LTE networks. The thesis focuses on the design of hybrid scheduling algorithm suitable for heterogeneous wireless communication networks and some of its QoS parameters have been discussed. Based upon the literature review and identified research gaps, the following objectives were proposed:

1. To study the existing traffic scheduling algorithms for heterogeneous wireless networks and studying various parameters for given QoS.
2. Development of a new scheduling algorithm for heterogeneous wireless networks to obtain required QoS.
3. Performance evaluation of the proposed algorithm.
4. Comparison of proposed algorithm with existing algorithms and also comparing various QoS attributes.

1.8 Methodology

Unlike earlier networks which are designed using either homogeneous or heterogeneous approach, may able to serve a number of users with a series of BSs planned in a layout to cover the maximum area. The advances in wireless communication motivated the concept of heterogeneity to achieve optimal performance by single base station. The increasing demand of smart devices around the world is forcing the telecommunication companies to deploy LTE or Mobile WiMax as well as heterogeneous networks. To satisfy growing demands of users, the suitable choice of technology supported by MIMO/OFDM is either WiMax or LTE or the combined approach for heterogeneous networks for faster, reliable and cost effective connections. An effective combination of the two technologies may able to complement each other and provide high QoS. The research work methodology is presented with the help of Fig. 1.7.

The work flow is started with the study of queuing theory which can solve a number of questions regarding service flows in queues, how to schedule the flows with QoS constrains is the major concern of networking today. The communication networks can be easily modeled in queuing theory and the behavior of various queue length moments is described. The new models can be designed depending upon job size distribution. The models study may enhance the performance of a network in terms of cost minimization and throughput maximization. The work is extended with bulk queuing service model and may be applied to design efficient traffic scheduling model for various networks.

WiMax systems is one of the major analysis platform for high speed data transfer but when the load increases over the nodes there must be a suitable mechanism which can handle the number of nodes and its architecture system . There might be several algorithms which can work with scheduling. But in this thesis three algorithms namely FCFS/RR, priority and hybrid algorithm, a combination of FCFS, RR and Priority algorithm, have been implemented. The hybrid algorithm shows a vital increase in the performance of the jobs completed and error rate reduction.

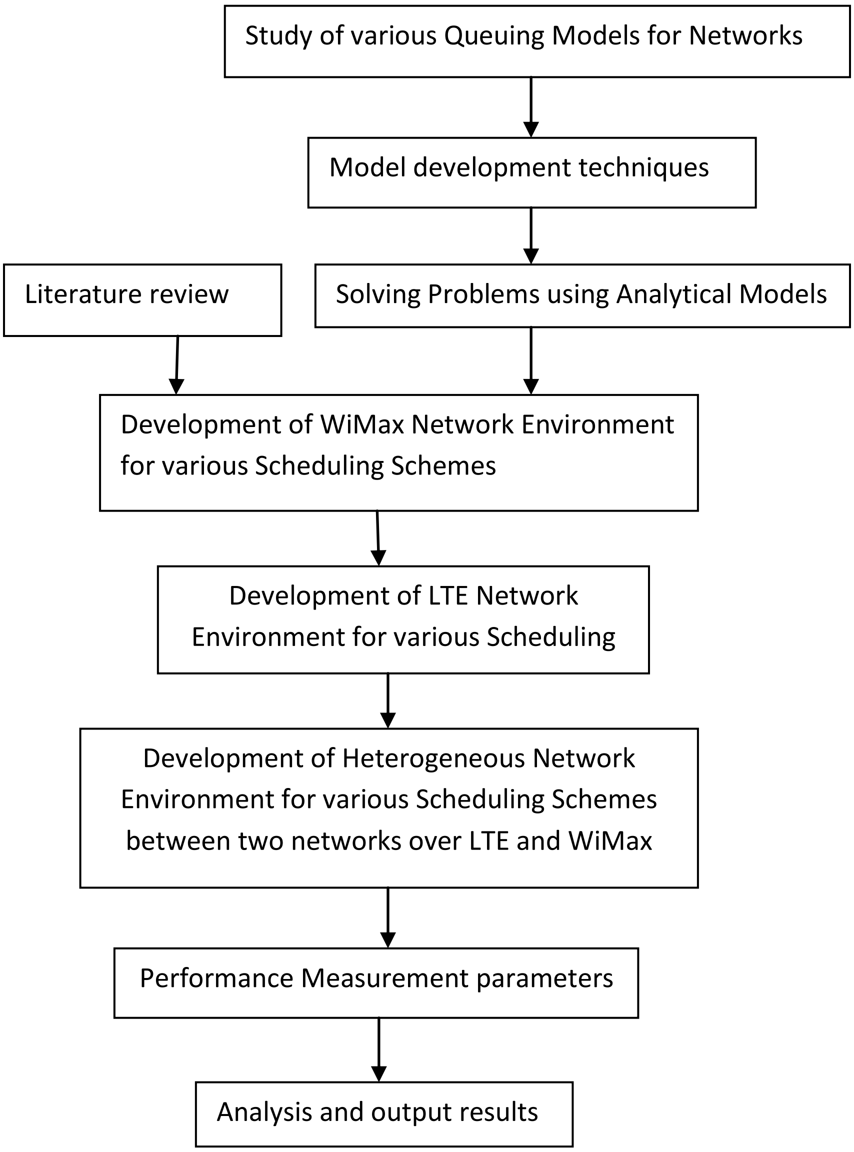

Figure 1.7: Methodology for Work Flow

The load balancing structure in LTE networks is discussed with fixed channel allocation or dynamic channel allocation strategy. The bandwidth is divided among different channels to manage all the resources for proper utilization. The radio resource scheduling includes resource management, power management and mobility. The hybrid structure is designed by combining more than one number of scheduling algorithms to achieve best of QoS as well as energy efficiency during load balancing.

The network performance is greatly affected by scheduling algorithm and has a direct impact on jobs requested and completed successfully. In this research work, a hybrid scheduling algorithm is proposed for heterogeneous networks. The heterogeneous network is comprised of overlapped LTE and WiMax network implemented in same environment by considering common parameters for both networks. The three scheduling algorithms are implemented separately and in hybrid manner over the simulation environment and the performance of these algorithms is compared. The obtained simulation results demonstrated that the proposed algorithm can select the best available network in heterogeneous environment based on jobs scheduled, load balancing and throughput and error rate are improved.

1.9 Contributions of the Research Work

As research is a never-ending process, there are several implications to explore the research. According to requirements of enterprises and demand of users, BWA networks have been provisioning operators and enterprises for many years up to a great satisfaction. The system specifications are continuously updated to modify the network operation to achieve high QoS. So, the main contributions of this work are to propose a Hybrid scheduling approach to deal with the demands as well as requirements of the broadband users.

1. To accomplish the first objective, wide range of available literature has been studied. After cavernous mining from the various research articles, it has been found that there is lot of space where researchers need to concentrate. Few of such gaps have been mentioned in Chapter 2.
2. From the available literature, some research gaps have been formulated which laid down the foundation of this research work. The challenging issues of wireless networks have been taken into consideration to design the new scheduling algorithm for heterogeneous networks. The QoS parameters for WiMax and LTE have been studied and their dependencies are also taken into consideration while designing the new approach. Some state of art work consisting of queuing schemes and their utilization factors along with the comparison is discussed in Chapter 3.
3. To achieve the second objective, traffic scheduling algorithms for WiMax and LTE are discussed in Chapter 4 & 5. The system model is designed for the WiMax and QoS parameters and constraints have been discussed. Then the network is

design by random selection of source and destination nodes and applying a random generated traffic over the specified area. The scheduling algorithms are then implemented by varying the amount of traffic and obtained results are compared. The performance evaluation is done on the basis of energy consumed, throughput, BER and SER. All the simulation setup is repeated for LTE network configuration and obtained results are compared.

4. The third objective is fulfilled by designing a heterogeneous network represented as the combination of WiMax and LTE. The heterogeneous network model is designed and random generated traffic is applied on the network. The hybrid scheduling and the existing scheduling algorithms are then implemented by varying the amount of traffic and obtained results are compared. The performance evaluation is done on the basis of energy consumed, throughput, BER and SER.

The concept of task scheduling with parallel processing is also implemented in heterogeneous environment to fulfill the fourth objective as represented in Chapter 6. The performance of parallel scheduling algorithm is better as compared the state of the art algorithms. The results obtained show that the proposed algorithms outperform the other algorithms with respect to various selected parameters.

1.10 Thesis Organization

Depending upon the various wireless standards available in the market nowadays, Wireless Access Networks are widely spread and in a stage of constant development. As the scenario is moving towards providing the heterogeneity among the various wired and wireless networks and analyzing benefits of LTE and WiMax, it can be observed that combination of different technologies will produce a network with competent features to satisfy market trends.

Based on the above research objectives, the main emphasis of this thesis is to develop a new scheduling algorithm for heterogeneous networks to obtain required QoS. The various scheduling techniques have been implemented separately on WiMax, LTE and heterogeneous Network. A new approach named load balancing hybrid Traffic Scheduling has been designed and implemented on WiMax, LTE and heterogeneous network. The results for the various scheduling schemes have been combined and compared. The systematic approach for the work is represented in next sections.

Chapter 1 gives an introduction about the existing fixed and wireless networks, cellular and non-cellular technologies, the different services and standards provided by various networks and technologies. This chapter also discusses the role of quality of service parameters on different traffic classes. It also explains the basic principle of traffic engineering and scheduling algorithms to be used for different traffic classes, their advantages and disadvantages, techniques and its applications and specifications.

Chapter 2 presents the comprehensive literature survey carried out in the area of traffic scheduling problems in wireless networks. A number of scheduling schemes for the heterogeneous networks have already been implemented and some new suggested algorithms are in implementation phase. It presents the results obtained for various hybrid or multilevel scheduling algorithms to satisfy the requirement of proper load distribution along with resource scheduling. It also represents the need of shifting from uniprocessor to multiprocessor systems to fulfill the QoS demands of users.

Chapter 3 discusses the queuing theory which is adequate to solve optimization problems, decision tasks and may able to perform large number of evaluations. The queue length and queue production theories are studied to satisfy QoS demands of different systems. The recursive formulas for queue length distribution are analyzed to generate higher moments of queue length distribution. In this thesis, queue length moment distribution of various queues are discussed. Analysis of moments of queue length in M/M/1, M/M/1/B, M/M/c, M/Er/1 and Er/M/1 queue models is represented. Also the recursive formula of $M/M^{[K]}/1$ queue is derived and compared with other queue models discussed in the chapter.

Chapter 4 represents the WiMax support of huge data transfer but shows that still there is problem with the node request and service management. It becomes a sophisticated task for the nodes to handle number of requests at the same time. The work focuses on different aspects of the WiMax network and the scheduling algorithm with hybrid architecture which is a combination of the First Come First

Serve (FCFS) and Round Robin (RR) Algorithm. An analytical study of the number of jobs scheduled and completed along with accuracy for FCFS, RR and Hybrid algorithm is represented.

In **Chapter 5** The work is preceded by designing architecture of LTE which is a progressing technology to provide guaranteed delivery of services in heterogeneous networks with high data rate and low latency along with effective utilization of resources. The work presents a load balancing strategy to schedule the traffic in an efficient manner. The scheduling algorithms namely FCFS, RR and a hybrid algorithm have been implemented with load balancing scheme in the LTE network. The performance of the system is measured in terms of throughput, SER and signal to noise ratio.

In **Chapter 6** the 4G heterogeneous network overlapping LTE and WiMax has been studied and a hybrid scheduling algorithm is designed and implemented for the heterogeneous network. The three types of scheduling algorithms have been implemented on the heterogeneous network and compared. A new load balancing approach is proposed which is combination of FCFS, RR including priority scheduling. The proposed scheduling algorithm's performance is compared with other mentioned algorithms and measured in terms of BER, signal to noise ratio and throughput. The proposed work also presented a parallel scheduling architecture to overcome the problem of overloading in WiMax and LTE network to reduce the waiting queue length. The evaluation is made on the base of throughput and Packet Delivery ratio. Due to its simple yet classic nature, it reduces the packet drop and enhances the PDR.

Chapter 7 Finally, this chapter presents the summary of conclusions drawn from research work reported in this thesis. The proposed solution has lot of future possibilities. Energy consumption could have been a critical parameter to find. There are various aspects for energy optimization which requires a good hand in this area of research.

1.11 Summary

The characteristics of wireless communication network make it quite difficult task to schedule the traffic with lesser delays and efficient utilization. As the type of traffic and volume of traffic is increasing, the QoS becomes quite difficult to achieve. The future cellular radio networks will empower all wireless networks to commune in a domain constituted by a number of devices and technologies using several radio interfaces. For this environment, one essential task is to generate common control functions for all required applications. The chapter presents the overview of heterogeneous networks taken as the combination of WiMax and LTE networks. Along with that the concept of scheduling algorithms has been discussed which is the problem taken as an issue in this work for heterogeneous network. The objective, methodology and thesis organization has also been designed.

Chapter 2

LITERATURE SURVEY

2.1 Introduction

To provide last mile solution to enterprises and business industries, the idea of deploying fixed broadband networks to deliver internet access has been generated by telecommunication network providers in 1990. The focus was to assemble a network with compatibility, flexibility and low cost, while maintaining the speed and reliability. The invented technology would require working well with the existing infrastructure and must also balance with the new researches. The fixed broadband services showed cooperative coexistence with existing telecommunication networks and race towards the research of wireless networking started. The internet services became demanding both for wired and wireless access. Then the LAN, MAN and WAN were described for local and global point to multipoint access. The IEEE designed 802.11 standards to work for LAN and 802.16 standards for MAN. In 2001, the 802.16 standard, named as WiMax was released for point to point and point to multipoint radio links, with frequency range varying between 11 to 66GHz for line of sight transmission. In 2005, LTE technology was developed for global networking in coexistence with a concept of heterogeneous networks. The heterogeneous networks represent the combination of more than one network working with high data rates, greater capacity and interference avoidance.

2.2 Related work for WIMAX Networks

To promote the new standard IEEE 802.16 in the market, the WiMax forum was established in 2001 to set it as technology for wireless connectivity in metropolitan area networks. The modifications in the standards were started and IEEE 802.16a came out in 2004 with NLOS radio channels supported a frequency range of below 11GHz with omni-directional antennas [12]. Further changes in standard were released as 802.16b, 802.16c and 802.16d and the updates were combined to extend the model as IEEE 802.16e, in 2005 as mobile WiMax [13]. The aim was to achieve high QoS with good level of interoperability, handover management and more number of subcarriers. The amendments in the standard are further done as 802.16f, 802.16g, 802.16h, 802.16j, 802.16k and 802.16m represented as AAI. The system is operable with MIMO/OFDM/SOFDMA to support lesser delays, low packet loss, optimized horizontal and vertical handovers with more than 2000 number of subcarriers. The system specifications are continuously updated to modify the network operation to achieve high QoS. The research review in this field is preceded in next section of this chapter.

2.3 QoS Based WIMAX Design Considerations

Guaranteed QoS in the wireless medium is a challenging issue for upcoming different wireless systems by using the scarce available resources [14]. The various IP protocols have been specified to fulfill different QoS metric demands defined at different layers. The QoS at physical layer is represented by the acceptable signal strength level, bit error rate (BER) at the receiver, AMC or symbol rate applied based upon type of communication traffic. The QoS at higher layers is defined in terms of jitter, delay and throughput. Different procedures have been suggested for guaranteed QoS requirements for different layers.

2.3.1 QoS Based Upon Queuing Theory

Queuing theory is one of important aspects to be considered for wireless communication networks. Queuing theory can be utilized to direct numerous problems in achieving QoS, which is supposed to be major challenge in telecommunication world today. In many communication technologies and IP services, the packets are transmitted, processed and received in the form of queues. So, delays and waiting time are major concern in packet transmission [15]. A bulk arrival queuing model has been represented to schedule bulk packets in various traffic conditions to reduce delays. Bulk arrivals, bulk queues and vacation queues are becoming great topic of interest for researchers due to large requirement in high speed telecommunication networks. In many real time applications, the server processing is done in group of packets for server active time and server is relaxed for vacation time. Mostly, queues of this type have finite buffer. Sufficient buffer space is required to be provided to these queues so that the packets may not have to wait or delays and losses of a packet are minimized [16]. In future, to increase the service capacity of the queuing system, a long range dependent queuing model has been designed to work upon the important QoS parameters as packet drop and delays and queuing length [17].

A bulk queuing single server model is discussed to describe different characteristics of the process systems. 'N' numbers of customers arrive in a batch, served in a batch of size k if number of customers is less than the batch size. The time of service is considered the same as applicable to packets already available in the system [18-19]. The behavior of the system is analyzed by the characteristics like number of customers present in system, number of customers in queue, queue size of server, waiting time in queue, variability property of the system.

If multiple packet streams from different devices is arriving the single server with a queue distribution of first in first out class, the service time distribution will be different among all streams. It is difficult to analyze the service time distribution of this class to complete the queue for customer service satisfaction. The problem is solved by defining a queue dynamics in other way [20-21]. The research provides recursion formulas to compute the stationary joint queue length distribution, the stationary distribution of the total number of customers and the marginal distribution of the number of customers from a particular stream.

To describe the node to node traffic in a decay based queuing system, Markov arrival processes (MAP) is applied to calculate the inter-arrival times for joint number of tasks in the inter node traffic. The markovian queuing network is utilized for traffic division and addition among the various nodes [22]. Even for large queuing networks, the traffic model designed will remain small in size is one of the salient attributes of proposed model. An additional analytical model is proposed for input queue operation during switching from one node to another [23]. The queuing model considered is M/G/1 which has general processor sharing property and works the same way as M/M/1 queue model for first and second moments of queue length. The arrival rate and service time are considered the same as M/M/1.

A new algorithm based upon multiclass queuing networks is discussed for closed queue structures having large queues to find the solution for infeasibility problem in multiclass networks. Again a recursive method is proposed to calculate the higher order moments for queue length in lieu of Mean value algorithms (MVA) [24]. The moments of queue length are fixed and do not depend upon the tasks allocated to the model. A method of moments (MoM) is one of the easiest approaches to calculate the moments of queue length for higher order queues. Based on recursion, the system is represented as set of linear equations is easier to solve as compared to MVA. The capacity planning in communication systems and networks is mostly represented by multiclass queuing model [25]. With a generalized algorithm proposed for capacity planning, the solution to this problem is revealed for multiclass queuing networks. The time and space required will get reduced along with low computational costs and less memory consumption as compared to MVA.

To cater with rising demand of bandwidth, various dynamic resource allocation algorithms are developed. One among these is being developed to make cognitive radio a reality based upon queuing theory [26]. A mathematical model is designed for two Markovian distributed queues modeled as M/M/1 and M/G/S/N. The centralized architecture of MAC has been implemented for the queues and mathematical analysis is targeted to calculate bandwidth required and latency of an unlicensed user.

Although huge research attempts have been made to model and analyze the queuing networks for varying type of traffic as well as varying conditions of traffic, the exact performance analysis are hard to achieve using queuing theory individually. The discussed research is primarily based upon fixed capacity queuing networks to simplify the scenarios. The study will be more elaborated by considering some additional parameters for QoS to be discussed in next phases.

2.3.2 QoS Based Upon OFDM-MIMO Physical layer Design

The service capacity of a system is varied depending upon the varying channel capacity due to fading and interference impact on the signal strength [27]. Various fading channels like Rayleigh, Rician and AWGN are categorized to study the strength of a signal using more number of transmit and receive antennas in MIMO system and to analyze the capacity of channel. An extensive performance queuing model has been proposed to get a deep understanding of capacity of a system under real time conditions with variable capacity [28].

OFDM is combined with MIMO structure to boost communication features and large quantity of data. The paper displays channel enrichment utilizing MIMO-OFDM mechanics [29]. The consequences of pilot-symbol-aided channel evaluation on the lower bound of the capacity of the system are discussed. The author [30] proposed simplified estimation of channel for OFDM systems with Multiple Transmit Antennas. In this paper multiple transmitting and receiving antennas are employed in Orthogonal Frequency Division Multiplexing (OFDM) structures to boost characteristic and capacity. In this paper, two techniques are utilized to enhance the performance and decrease the complication of channel parameter prediction: excellent training-sequence method and easy channel estimation. The optimal training series not only simplify the primary channel estimation, but also achieve the best computation performance. The complexity of the channel estimation can be reduced at the expense of insignificant performance degradation [30]. The new technique is determined through the simulation of an OFDM system. For transmit diversity, the space-time coding with 240 data bits per code word is depleted. From the simulation, the necessary signal-to-noise proportion is about 9 dB for 10% information error rate for a channel with the conventional urban delay profile and Doppler frequency is 40 Hz.

By using multiple transmitting and receiving antennas to arrange MIMO mediums can boost the capacity [31]. MIMO-OFDM is treated for wide band communication to reduce inter-symbol interference and improving capacity of system. In this system, two space time codes considered separately for two transmitting antennas. At the receiver side, these space–time codes are decipher separately adopting pre-whitening, pursued by minimal Euclidean-distance decoding depend on successive interference reduction effects. For example, computer simulation reveals that the data transmitted at 4 Mb/s for a 1.25MHz channel for 4 X4 input-output systems, the prescribed SNRs i.e., signal to noise proportion considering 10% and 1% Word Error Rates (WER) are 10.5 dB to 13.8 dB, appropriately, when every code word consists of 500 information bits and the Doppler frequency of channel is 40 Hz. Growing the total number of the receive antennas enhances the performance of system. Although the receiving antennas are raised from four to eight, the SNRs needed are decreased to 4 dB to 6dB respectively. Therefore, for highly spectral efficient wideband transmission the MIMO-OFDM is a promising technique [31].

For wideband transmissions, the channels have been divided into sub channels using OFDM for a frequency selective mode to reduce the complexity on the receiver side. Space-time coded (STC) system is extended as MIMO/OFDM STC system [32]. The concept of multi diversity is utilized for diversity modulation to reduce the BER in AWGN channel. The performance is compared with and without using convolution coding. The issues regarding MIMO/OFDM STC are also discussed.

An additional algorithm to enhance the capacity of n channel OFDM system is proposed using binary signaling methods. The information transmitted and hence the BER can be reduced with optimal choice of parameters for link design [33]. The errors can further be reduced by sending k channels out of total channels. A subcarrier changing allocation method is developed and designed in order to enhance the capacity of an OFDM scheme in the downlink situation [34]. Based on this allocation design, adaptive modulation is selected for every user on the assigned partitions. The results acquired depict that the proposed dynamic allocation strategy defeats a static allocation strategy in terms of a decreased BER and achieve a higher system capacity for the similar signal to noise ratio (SNR). The capacity enhancement is investigated for enormous performance in WLAN and can be extended for 4G broadband encouraging results for improving the data rates of wireless transmission systems for next generation [35]. In this, the main focus is on Internet consumers in hotspots such as airport etc. demanding huge data rates.

Nowadays for research in mobile communication systems, the MIMO-OFDM system is modeled using Rayleigh Fading Channel [36]. By using adaptive modulation and Space-Time Coded encoder, the capacity of MIMO-OFDM systems is increased. To overcome the performance problem, these codes are used for deceasing Peak to Average Power Ratio (PAPR) and increasing spectral efficiency of the system. With the rapid enhancement in wireless communication systems, spectrum utilization is a major problem. Using different antenna arrangements over Rayleigh, Rician and Nakagami fading channels, the BER performance of MIMO-OFDM system has been compared over different modulation schemes in [37]. The performance improvement in MIMO-OFDM-based air interface can be done by applying channel modeling, modifying transceiver design and reducing the BER ratio by increasing the Maximum Likelihood function through channel estimation has also been discussed [38].

The major disadvantages in OFDM system is high Peak to average power ratio (PAPR). The subcarrier waveform in OFDM is a form of pre-coding technique, where every OFDM block is linearly constructed by shaping matrix since modulation and transmission. This path has the possibility of decreasing the PAPR of the OFDM without touching the bandwidth adaptability of the system and BER accomplishment [39]. The PAPR decline of MIMO-OFDM systems are acquired using the pre-coding pattern. The simulations display that the pre-coding pattern executes better with MIMO-OFDM schemes. An analysis on uplink resource assigning in OFDMA wireless systems is proposed in [40]. OFDMA has been chosen as the multiple approach schemes for state-of-art wireless communication systems. In addition to this, effective capability allocation in OFDMA wireless systems is essential in consideration of gathering the quality of service necessities of emerging values. In this paper, an overview of resource assigning and organizing designs in OFDMA wireless system is conferred. The target is on the uplink control. Resource distribution is evaluated in numerous designs: centralized and distributed, immediate plus ergodic, optimal and sub-optimal, single unit and multiple unit, coordinated and non-coordinated, coupled with distinctive consolidations of these variants. Directions for upcoming analysis are marked.

In centralized and distributed data communication systems, high speed has become the major requirement with great flexibility in connectivity among different wired and wireless networks. The issues related to RF and IF wireless networks have been discussed. Advantages and disadvantages of these networks, causes of interference in wireless links, modulation and transmission schemes are also discussed [41]. The speed of system can be increased by implementing more number of transmitters and receivers. The number of transmitters can help to achieve maximum capacity with less outage probability. Based upon acceptable performance and reduced outage probability, the number of transmitters can be chosen by adopting special diversity methods [42].

The special multiplexing and transmit diversity are the two parameters analyzed for MIMO communication systems based upon the special mapping to reduce the errors at receiver signal [43]. Each cluster will choose its own group of subcarriers to make a choice of multiplexing and channel diversity. The proposed scheme works with STC to province choice between multiplexing and diversity.

The WiMax physical layer model is designed for BER evaluation of the real time data by comparing different modulation and channel encoding schemes [44-45]. The parameters are chosen based upon the standard for transmitter and receiver section physical layer and performance evaluation is done. In an indoor OFDM based wireless communication system, the subcarriers are effected by bit errors in the center of frequency selective fading and due to the slow varying nature of the indoor channel, the fading effects the same subcarrier for duration of several OFDM symbols. After demodulation, errors in the number of bits in an OFDM symbol are observed. An adaptive physical layer uplink model suitable for characterization of indoor OFDM based systems is proposed [46]. The proposed model is validated by calculating error count, error length and error free intervals length in a frame.

The applications of OFDM consist of multicarrier CDMA which is strictly based upon the concept of OFDM. Variable rate services are managed by multicarrier techniques by allocating a range of subcarriers to each user depending upon bit rate request. A joint PHY-MAC solution is proposed for controlling access of large number of carriers in the wireless medium [47]. As the performance degradation occurs due to interference at Physical layer level, the continuous observation has been achieved in order to provide better signaling along with better transmission parameters. To decrease the data rate of users and channel impairment conditions, a set of parameters is proposed to attain a condition of required user QoS. An additional multi-code Group Orthogonal (GO) OFDMA-CDMA system is presented to fix the orthogonal subcarrier groups in the available subcarriers [48]. According to data rate, the user data stream is selectively multiplexed into a variable number of sub streams. Then, these sub streams are transmitted over an OFDM subcarrier group, randomly allocated to the users. The results obtained by adopting this method outperform the multicarrier CDMA for higher data rates whereas BER performance of lowest rate users is slightly degraded. So, the tradeoff is managed between performance achieved and computational complexity.

2.3.3 QoS Based upon WiMax Scheduling

The requirement of QoS is highly desirable in IP based multimedia, communication and internet services accessed with wireless networks. Almost all the sub-urban and rural areas are required to get connected with the wireless networks where the cost of implementing wired network is high. WiMax is considered as the feasible solution to the problem for covering larger areas with high speed and QoS service provisioning. The proposed solution covers the whole unused bandwidth by presenting QoS architecture for different service classes [49]. A number of key issues have been discussed regarding scheduling QoS architecture: resource utilization, channel conditioning, scheduling survey and comparison between existing scheduling schemes [50]. The scheduling scheme is proposed for bandwidth utilization, to increase the throughput and decreasing the power consumption. Another scheduling architecture is proposed to support constant bit rate (CBR), variable bit rate (VBR) and BE types of traffic [51]. The efficiency and performance was studied and analyzed to check the performance of proposed methodology.

The service mechanisms in WiMax can be represented in integrated or differentiated form for mapping and signaling in Time division mode. A highly efficient and fastest signaling control mechanism is proposed for QoS architecture to support guaranteed throughput and delay requirements [52]. The work is focused on Always Best Connected (ABC) mode to support QoS handovers. The handovers are initiated between WLAN and WiMax for multimedia services and simulation results are represented in form of throughput, delay and packet losses [53]. The CAC criterion is proposed for ABC mode to extend the model for wireless personal area network (WPAN) and 3G services.

Different service classes are having different parameters to represent the QoS characteristics of these services for various wireless networks. The services are scheduled according to the priority assigned to them. Each class is having its own set of QoS parameters to be defined in MAC layer. The class based scheduling algorithm is proposed for UGS, rtPS and nrtPS services [54]. The QoS is proposed based upon channel conditions, transmission delays, and received data rate using priority traffic conditions.

Scheduling may be categorized as uplink or downlink scheduling based upon applications to be served during UL and DL for large amount of traffic streams. Depending upon the bandwidth available the channels may be allocated to busty traffic during UL and DL which however can't fulfill the requirements of QoS. An adaptive polling service mechanism (aPS) is proposed for bandwidth requests generated by busty traffic which reduces the signaling overhead and bandwidth requirements are also reduced [55]. Hence the efficiency of polling is increased during UL bandwidth allocation. IEEE 802.16 is able to support both fixed and mobile services and able to provide significant QoS performance if the choice of parameters will be suitable as per network conditions. QoS performance has been studied for various traffic streams in accordance with the suggested characteristics of WiMax standard and performance parameters like throughput, delay and packet loss are analyzed [56] by comparing different service standards.

An adjustable dynamic call admission control (CAC) algorithm is discussed to support guaranteed QoS for different service classes in mobile WiMax environment for maximum network utilization. A bandwidth allocation algorithm is also proposed to evaluate amount the channel utilization with respect to traffic load distribution. CAC and bandwidth allocation schemes work together to maintain network utilization in SS up to 100% while dropping probability of mixed on-going services (MogS) and blocking probability of UGS are kept well below 1%; also lowers in Blocking probability of rtPS and nrtPS. This CAC algorithm has not included the queuing delay of each service class which will be considered as future work [57]. As the BE request is considered low level request, no proposed algorithm is defined to improve the CAC in BE traffic. Hence, the author proposed a token bucket algorithm to penalize high level traffic and to make BE traffic accessible with priority. A Markov chain model has been considered to improve the blocking probability of BE traffic with and without token bucket algorithm. The throughput has also been improved for low level traffic [58]. Additionally, Markov chain model has been designed for UGS, rtPS, nrtPS type of traffic to increase the connection probability and reduce the average connection delay [59]. A bandwidth request mechanism has been generated to analyze the performance of system in TDD uplink sub frame duration. Priorities are assigned to real time services so that these may get the opportunity to get the time slots in the contention period in order to send their bandwidth provisions.

Most of the proposed scheduling algorithms in IEEE802.16 are deployed as UL or DL scheduling for real time services. The focus is mainly on the bandwidth allocation for UL and DL flows. An attempt has been made for installing a dual feedback approach for bandwidth allocation in VBR service in WiMax networks. MAC delay is calculated for rtPS with broadband wireless access to improve the bandwidth utilization and queue length delays [60]. Bandwidth reservation is one of the major problems to be discussed in 802.16e model. To achieve bandwidth optimization, the author proposed a bandwidth allocation algorithm based upon admission control policy [61]. The fairness can be improved and blocking can be decreased during call processing with the help of proposed scheme. A BS assisted scheduling strategy is discussed to provide end-to-end delay reduction between BS and customer premises equipments (CPEs). The traffic from MAC is divided into various queues and sent to BS. The base station assisted scheduling algorithm will check the queues and scheduling is implemented depending upon traffic mapping and integration of traffic at MAC level [62]. To provide QoS provisioning among scheduling algorithms, BWA networks are categorized into different service classes as defined earlier. Two scheduling algorithms have been proposed for connection admission and establishment at MAC layer [63]. The combination of weighted fair queuing (WFQ) and earliest deadline first (EDF) is generated to achieve maximum fairness and minimum delays. The other scheduling algorithms used are priority queuing and random early detection to reduce jitter and delays. A simulation is proposed for large number of nodes in WiMax environment to study and evaluate various QoS parameters as throughput, delay, latency, packet error rate, packet delivery ratio [64]. The Table 2.1 gives the comparative analysis of the QoS based WiMax scheduling schemes.

Table 2.1: Comparison of QoS based WiMax Scheduling Algorithms

Year	**Authors**	**Algorithm**	**Goal/Key Features**	**QoS Framework**
2011	Rizwan et al.	Priority based	Queue management for UGS & BE	BW fully utilized for different service classes
2009	So-In et al.	Scheduling QoS architecture	Scheduling survey & comparison	Enhanced utilization of resources
2002	Hawa et al.	InterServ & DiffServ QoS control mechanism	QoS control architecture for PMP & Mesh mode	Fast and efficient service setup
2006	Roy et al.	Link layer integration and admission control	QoS for VBR services	QoS performance improved
2006	Liu et al.	Priority based on service & channel conditions	Class based QoS requirements for delays & received data rate	BW efficiency & low implementation complexity
2007	Nie et al.	aPS for BW requests by bursty traffic	UL & DL scheduling for large traffic	Reduced BW requirements & signaling overhead
2008	Tung et al.	Handover & BW allocation for all services	Blocking & dropping probability vs traffic loads	Supports all services with different QoS
2010	Mokdad et al.	Markovian model for priority services	Admission control for BE services	Connection enhancement for BE traffic
2014	Othman et al.	QoS markovian model for admission control	Performance analysis of ert, rt & nrt polling services	BW provisioning to reduce delay and connection probability
2009	Park et al.	BW request mechanism with priority asignment	BW provisioning for UL & DL flows of VBR services	Reduced delays & improved BW utilization
2009	Gidlund et al.	Combination of WFQ & EDF	Connection admission and MAC establishment	Maximum fairness and minimum delays

2.3.4 QoS Based Upon Traffic Types

The scheduling algorithms can be designed based upon traffic type and length to achieve maximum fairness and to support QoS. The various traffic classes may be described as UGS, rtPS, nrtPS, ertPS and BE. OPNET simulator is used to design the scheduling algorithms for the various service classes to attain best quality during link design [65]. The DL design is considered to for UGS, rtPS and nrtPS services for RR and WRR algorithms. The comparison of two schemes is done by increasing the number of nodes. The proposed scheme uses the threshold priority algorithm for fixed amount of bandwidth for large number of users with variable type of traffic and reduces the delays and hence throughput is increased [66].

By measuring one-way delay, the interworking between WiMax and DiffServ networks is discussed to calculate jitter and packet drop rate of VoIP traffic for WiMax users in a test-bed scenario with and without DiffServ [67]. It is shown that DiffServ in the core network solves the congestion problem in efficient way by providing service priority to delay-sensitive traffic. The results show valuable gains using DiffServ network. The QoS requirement and configuration may differ for different service classes and traffic types. A mobile WiMax network is deployed [68] to compare the different service classes for delay sensitive applications. VoIP transportation is discussed for BE and ertPS, which fluctuates at larger rate. The performance of data services is also studied for the same traffic types during bandwidth reservation.

The objectives of this paper are to study in detail the architecture of IEEE 802.16 and to simulate video transmission based on MPEG 4 and H.263 coding schemes on the 802.16 network [69]. Specifically, the study examined various performance of QoS parameters included video packet loss (number of packet dropped), end-to-end packet delay, and throughput of several subscriber stations over rtps service flows as defined in WiMax networks. The OPNET modeler with integrated WiMax support has been adopted for this effort. The effect on various QoS parameters for FTP and HTTP traffic in WiMax network is discussed in [70]. The study is conducted on low quality video stream at 128 Kbps over the application of coded modulation quality and

Suppressed silence. Study concludes that there is a tradeoff between delay and quality affected by various parameters.

Scheduling algorithms play an important role in wireless network planning and optimization. The scheduling schemes in WiMax networks are based upon the network capacity, bandwidth allocation and resource allocation methods. The major QoS constraints to be taken for network planning are maximum fairness and goodput. The various UL scheduling schemes are compared for BE and FTP traffic to achieve a reciprocation between the discussed QoS constraints [71]. The algorithms discussed for these service classes are RR, weighted Delay and token bucket algorithms. Streaming video is one of major quality class of service required now a days by large number of users. The author described a QUALNET based emulator working with WiMax networks called WEBS for energy saving algorithms during service mechanism [72]. Maximum amount of time the device works in sleep mode while streaming going on. Hence the idea is generated to save the energy during sleep mode. The results have been improved by almost 50% to save in battery of WiMax enabled devices without affecting the quality of service.

Various wireless models can be implemented with different simulators as discussed in earlier research. The frequently used simulators are OPNET, QUALNET and Network simulator (NS). A new WiMax scheduling architecture model is designed using NS-2 to validate the results for real time traffic classes [73]. The scenario is designed for voice and video calling in real time environment to attain QoS parameters and results are optimized to calculate throughput, delay and packet loss. Study of cross layer design is one of the important parameters in wireless network design. A signaling control mechanism is adopted for IEEE 802.11e to analyze the video quality at the receiver end [74]. The physical and MAC layer are shared for connecting two devices in open system to design a cross layered approach and QoS parameters are evaluated.

Simulations are helpful in designing the networks as they help to investigate the theories regarding implementation of real world problems. The approximate scheduling analysis can be performed with the high computation algorithms to be defined in various mathematical theories [75]. Various test data are generated for testing the algorithms by random procedures in simulations to get the exact analysis and performance parameters in scheduling algorithms.

The MAC layer encloses UL and DL scheduling with either time division or frequency division operation to fully utilize the channel bandwidth to optimally allocate the available bandwidth among the users. One of the critical issues in network planning is bandwidth allocation up to channel rate capacity for manifold QoS requirements. A number of protocol design approaches are suggested to tackle with the bandwidth allocation problem for different channel services [76]. A benchmark has been generated for full resource utilization, power saving and limiting the errors during traffic distribution and communication.

The scheduling algorithms mainly considered for WiMax are RR, WFQ, WRR and SP to evaluate the performance of network for different service classes. The proposed results show that the SP scheduler generates the best results among all for BE and nrt services providing high throughput and less delays [77]. The fairness is also to be calculated for QoS provising in BWA networks. The fairness algorithms discussed are EDF and WFQ and enhancement in EDF is presented with queuing theory to increase the fairness and reducing the delays due to priority starvation problem [78]. The priorities are assigned to traffic as low to high and results are evaluated to improve the throughput in low priority queue and improving the delays in high priority queue. The tremendous development is going on to enhance the services and QoS in WiMax networks by applying different types of scheduling algorithms to get high performance and cost effectiveness for all types of traffic [79]. The various scheduling algorithms discussed and compared in this study are RR, Deficit Round Robin (DRR) and Modified Deficit Round Robin (MDRR) for achieving high efficiency.

The main contribution of the work is to analyze the behavior of different scheduling algorithms for delay and jitter constraints. The real time services are more affected by jitter in fixed as well as mobile environment. The optimization algorithms are designed for IP network for controlling the delays, packet loss and increasing the throughput [80]. Mainly the packets in computer networking are defined in either real time and BE class of service. Sometimes the services may be integrated in the packet switching network to enhance the speed of transmission. Two scheduling algorithms are compared for BE and real time traffic are namely FCFS and EDF to generate best class of service for QoS guaranteed [81]. The parameters studied are error rate, delay and buffer size approximation for packet switching in WiMax networks. An advanced research proposal has been discussed based upon Greedy scheduling algorithm to fulfill the needs of rtPS, represented as Greedy Latency algorithm [82]. Priority is included in the proposed scheme for throughput optimization in BE and rtPS for interclass and intraclass TCP/UDP traffic architecture. It is supposed to be protection algorithm for TCP flows from packet loss and packet drop. As WiMax is supportive for wide range during interoperability operation, the QoS is required to be guaranteed for different data, voice and video applications. A hybrid scheme based upon channel conditions has been proposed for resource allocation in WiMax MAC design [83] for complete bandwidth allocation and channel utilization.

IEEE 802.16 (WiMax) is an innovative and cost effective technology that allows fast and easy access of BWA networks for fixed as well as mobile users. Qualnet simulator is used to study the effects of delay, jitter and losses with CBR traffic to be deployed in the form of UGS and BE service flows in order to use them in a proper way for real life scenarios [84]. High speed mobility is the need of wireless networks to provide telecommunication services at a faster pace. WiMax is able to provide high speed data, voice and video transmission for different QoS classes [85]. Several scheduling algorithms discussed earlier are compared along with proposed algorithm represented as Self-Clocked Fair (SCF) algorithm to measure QoS parameters in WiMax network. The last mile wireless access is provided by WiMax for all service classes and applications. Again, the parametric study of delay, throughput and jitter is represented by allocating the weights dynamically to different traffic types [86]. The weight factors are specified for rtPS and nrtPS during MAC design admission control using OPNET model for simulation. An admission control strategy is designed for multi class traffic by reserving certain amount of bandwidth for UGS and rtPS classes. A bandwidth request algorithm is proposed at BS scheduler and SS are allocated with specific channels [87]. The performance of algorithm is evaluated for UGS, BE, rtPS and nrtPS traffics to calculate delay, throughput and packet loss. An additional bandwidth request grant mechanism is used to discuss the performance enhancement in ertPS, rtPS and nrtPS services in mobile WiMax networks [88]. The performance metrics considered are delay minimization and throughput maximization.

QoS in wireless networks is an important paradigm to attract the attention of network providers in telecommunication world. The present study conducted a comprehensive study about Point-to-multipoint Scheduling Algorithms in WiMax networks [89]. A detailed simulation was performed to investigate the efficiency of the main scheduling algorithms as FIFO, WFQ, PQ and MDRR and the performance of each scheduler is evaluated to support the various classes of quality of service and various applications. The results of simulations showed that appropriate selection of scheduling algorithm can improve the required QoS for different traffic types of users. The best scheduling algorithm in this evaluation is determined based on the minimum jitter, throughput and maximum received traffic for each servicing class and specific application. WiMax networks have been substantiated due to its high data rate, wide area coverage and differentiation of Quality of service. In this paper they adduce weighted round robin scheduling algorithm that not only fulfills the quality of service requirement but also provides a fair scheduling for real-time services [90]. In this work, a comprehensive study was carried out of scheduling algorithms such as Weighted Round robin, round robin, strict priority and weighted fair queuing, analyzing and appraising the performance of schedulers individually in order to support the different service classes. The simulation is carried out and the results implied that Weighted Round Robin out performs other schedulers and provides higher service standard to support the different QOS requirement.

2.3.5 QoS Based upon Call Admission Control and Handover Management

Handover is an important aspect of mobile wireless communications. IEEE 802.16 standard was initially designed for fixed broadband wireless solutions, substituting the cables for last mile connectivity. The importance of network scan in the whole handover procedure has been realized, thus the proposed design strives to reduce the scan latency and optimize the handover message exchange [91]. Mobile WiMax introduces support for handovers which is basic requirement for mobile communication. The promised high speed handoff is designed in this work to use soft handover methods [92]. The handoff latency is measured and simulation parameters are adjusted to achieve best handoff timers. The aim of the work is to find the parameters which affect the handover performance.

A scheme has been proposed for handover management to decrease the number of scanned BS, thus reducing the time for handover [93]. The BS selection scheme performs better than already existing handover scheme in WiMax, thus reducing the delays and wastage of resources. As the BSs in process are reduced, the amounts of messages for connection establishment, processing and connection release among MSs and BSs are decreased. Hence time and resources are saved. Sometimes the handovers are not required and the decision is to taken by the nodes lying within that network. The appropriate decision making and handover procedure for different nodes is described by some algorithms. The research describes the unnecessary handoff and delays during handover initiations for better network performance [94].

The increased demand of high speed IP services and increased number of mobile devices has accelerated the research towards generation of new techniques and services. As the services are increasing, the handoff requirements are increasing to achieve a high speed broadband network. Initially it is represented as WiBro, the wireless broadband IP network to cover broader WLAN regions. Wibro requires managing mobility support in MAC and network layer to minimize the latency and delays during handoff for seamless service provisioning [95]. The cross layer handoff optimization technique is discussed to receive the maximum strength signal and reducing the handoff delays. Fast handoff is achieved by suggested exponential smoothing technique and comparing the results with existing methodology to reduce the handoff delays.

Resource management is one of the important considerations for handoff delays and dropping probability. The handoff is well quantified by using the approach of full resource utilization which is proposed in this work using a fuzzy based bandwidth controller [96]. It is optimally utilized for real time as well as non real time services to deny more bandwidth utilization than required. Two resource management schemes are implement for handoff between WLAN and WiMax networks by taking QoS into consideration. The handoff schemes may be distance based or priority based depending upon the network performance metrics. The metrics as throughput, delays and SINR are considered for distance based handover between BS and access point [97]. It helps in selecting the best connecting point during handoff based upon distance criteria and to achieve high SINR ratio.

Some important parameters for QoS aware handoff decision and CAC in WiMax networks are bandwidth allocation, resource utilization, call blocking probability and call dropping probability. The connections are admitted and get the resources with the priority being assigned by CAC during handoff connectivity. Various service classes are assigned the high and low priorities. The performance of CAC scheme is analyzed for WiMax networks in the proposed model [98] and is compared with connection blocking and dropping schemes for different types of service classes and bandwidth utilization is evaluated.

The seamless integration of different networks is mainly depending upon the handover management aspects. The handovers may be horizontal or vertical decided by the type of the service to be facilitated and the type of network to be used. The integration needs to take into account the procedure for QoS mapping. MIH provides a uniform framework representing the integrating of two networks through common scenarios presented in 4G, including the handover signaling between WLANs and WMANs. The future research includes MIH support for multihop heterogeneous networks, QoS mapping and resource allocation mechanism for integrating networks along with security support and solutions [99].

To integrate heterogeneous networks provisioning individual coverage and capacity along with hotspot connectivity and high data rates, there is need of an optimized vertical handover (VHO). There is a number of decision algorithms available based on certain QoS parameters. The performance measurements of physical and MAC layer are represented in IEEE 802.11x i.e. bandwidth measurements in terms of network allocation vectors (NAV) and error vector measurements (EVM) for data and management frames [100]. EVM can measure modulation errors which should be reduced to strengthen the performance of NGNs and maximum QoS is achieved during VHO of 802.11x .

The handoff scheme and call blocking probability are the major consideration for QoS in Wireless Cellular Networks. As the networks and number of users are growing day by day, the demand for handoff from one network to another is increasing with no channel division. The new scheme has been suggested to overcome the issues during handoff and to reduce to call blocking probability [101]. The channel utilization has been increased for real time services. The work may be further extended for nrtPS as well as other type of services. Next generation 4G wireless networks highly require QoS architecture for service provisioning for end to end real time applications. A coordinated call admission control (CCAC) mechanism is presented to contribute for high QoS architecture for CAC and handoff management. CCAC utilized the total bandwidth and decreased the call blocking probability [102]. A Guard bandwidth is reserved in CCAC for handoff to provide seamless mobility. The call blocking and call dropping probability is decreased during handoff. The new methods presented in CCAC adjust the bandwidth dynamically for handoff initiation and adjusting new calls.

A VHO algorithm can support the provisioning of QoS in mobile WiMax networks by handling over BE low speed WiMax SS to an overlaid wireless LAN network subject to QoS guarantee for the SS. Simulation results show a significant improvement in probability in blocking using proposed strategies over the normal strategy [103]. For future work, more constraints on QoS factors such as data rate, delay and throughput for different type of services need to be considered depending upon the user preferences. Different service mapping techniques may be suggested to compromise between two different service classes.

2.4 Related Work for LTE Networks

Started as a project in 2004 and earlier known as the Third Generation Partnership Project (3GPP) is initiated by the telecommunication research body is defined as LTE is considered a revolution towards 4G networking. 3G packet networks were also revolutionized by SAE (System Architecture Evolution). The term LTE represents both SAE and LTE working both in 3G and 4G environment. LTE is customized to provide flexible bandwidth with high data rates and less delay. Efficiently working with packet switched networks and seamless mobility, LTE architecture is supports great QoS. The aim of global IP connectivity without any delays and errors in packet data networks is achieved with LTE along with mobility applications.

2.4.1 QoS Based LTE Design Considerations

The upcoming 4G mobile technology promises to fulfill the high data rate, high bandwidth capacity, all IP connectivity for fixed and mobile applications [104]. LTE is implemented with MIMO with a downlink bit rate of 100 Mbit/s and the uplink bit rate of 50 Mbit/s for a 20 MHz channel. Its physical interface is represented as Evolved UMTS Terrestrial Radio Access (E-UTRA). A lot of architectural improvements have been done to provide 4G networks with high data rates. The comparative study of different 4G network as Evolved High Speed Packet Access (E-HSPA), WiMax and LTE is described in [104].

A study has been presented on LTE evolution toward LTE-Advanced (LTE-A) represented as OFDMA and MIMO technologies [105]. The work also discusses MIMO enhancements for LTE-A, Coordinated Multi Point transmission (CoMP), repeaters and relays. The service coverage area and signal throughput in LTE-A networks is required to be increased with higher data rates. The high throughput can be achieved with high signal to noise ratio and low symbol error rates. One of the methods to enhance the coverage with throughput is the use of relay stations [106]. The author has implemented the network design of physical layer with more number of relay stations and mathematical analysis proved that OSTBC codes in LTE downlink design achieved the coverage enhancement and decrease in SER.

A single transmitting and single receiving antenna is used to implement OFDMA in LTE and dynamic subcarrier allocation schemes using OFDMA is proposed to assign channels to various users. The energy efficiency is increased with resource utilization as proposed [107] for given data rate requirement. The fading channel is considered for performance evaluation. An interference condition criterion is suggested for scheduling and resource allocation in joint form. The model is optimized for multi hop multi channel environment to improve the throughput in wireless networks [108]. An additional algorithm for power allocation in subcarriers is proposed for OFDM system with multiple frequency bands for operation [109]. A greedy algorithm is designed for subcarrier power allocation at different channels.

A max-min fairness model is proposed to solve the bandwidth allocation problem to achieve the trade-off between fairness and throughput in wireless mesh networks are discussed [110]. To optimally allocate the subcarriers in wireless adhoc networks using OFDMA, the requirement is to reduce the distortions from other channels and maximizing the energy efficiency and throughput. The algorithm is proposed to manage with the available resources using OFDMA and fulfilling the QoS requirements [111]. The capacity of the network is also improved by using Lagrange dual algorithm for subcarrier allocation. Cognitive wireless networking integration scheme is used in LTE networks for controlling the interference, reducing complexity and throughput management functions described to make the system power efficient [112]. The optimization solution is proposed to increase the network efficiency and managing the network load with respect to bandwidth allocation.

A number of performance metrics have been discussed for LTE network to analysis the network in different domains of signal quality. The various QoS metrics are represented on individual cell based upon the performance parameters and user throughput, cell throughput and link throughput has been analyzed [113]. Live call traces have been picked and calculated for result analysis. The throughput measures strongly depend upon last transmission interval and link adaption algorithm used. The objective of LTE is to provide the low latency, high data rate, latency and packet optimized radio access technology. All these features support the deployments of flexible bandwidths. This research presented the overview of features of LTE from release-8 to release-13. The comparison study indicated that the release 10 of 3GPP exceeds and support IMT-Advanced requirements and also accepted as a true 4G system [114]. The result indicated that the current, 3GPP works on release-13 are capable of performing in the more improved system.

The utilization of LTE vehicular environment using mobile femtocell is presented [115]. This research presented the importance of the mobile BS to serve as the vehicular user equipment (UE) in public transportation. Vehicular UEs are exposed to path-loss, high VPL, performance and interference degradation. The author proposed Mobile-Femto technology in this research to improve the performance of cell edge vehicular UEs. To create the comprehensive comparison between Fixed-Femto, eNB and mobile-Femto assisted transmissions. The comparison was performedregarding spectral efficiency, throughput, vehicular UEs Link Ergodic capacity and SINR. The result indicated the improvement in the performance of vehicular UEs after the implementation of this technique in public transportation. Also, 80% of equipment throughput was an improvement over the direct transmission using eNB.

The performance of the LTE based parameters with the help of OFDM in the outdoor and indoor environment is required to be enhanced. The study measured the spectral efficiency throughput and signal to interference ratio (SINR) using various scenarios with a different number of users [116]. The result of the study shows that the QAM and PSK throughput varies significantly in the indoor and outdoor QAM, indoor and outdoor applications. The average consistent performance and throughput are more frequently observed in the outdoor areas than indoor areas. LTE minimized the cost of delivery of data and increased the data output rates. In the LTE network, outside sources of design, configuration, and interface decreases the spectrum efficiency which leads to higher costs for network operators and low data output rates.

The multiple access techniques are used for uplink/ downlink in the wireless communication systems. The study presented the comparison between SC-FDMA and OFDMA. This research demonstrated the issues, basic features of the uplink, and multiple access techniques in the LTE Mobile Communication System. The technology used for uplink the transmission is SC-FDMA. The results indicated that the SC-FDMA has lower PAPR as compared to OFDMA. It is also found that the SC-FDMA has the better performance regarding frame error rate and PAPR because of built-in frequency diversity and coherent single-carrier property. The performance of OFDM is improved in 4G wireless communication such as WiMax and LTE. The performance of the MIMO-OFDM is observed with the help of SNR and BER [117]. With the increase in the SNR, BER decreases, and vice versa. In LTE, the high data rate is obtained by using the multiple antenna at both sides of the wireless link. In the wireless broadband system, OFDM was used to decrease the receiver complexity. The OFDM is used in the underwater communication. Problems in OFDM such as intercarrier interference and intersymbol interference are solved with the help of cyclic prefix or guard period techniques. When the BER is zero, different modulation techniques are used such as QPSK, BPSK, and 16QPSK. The results indicated that combination of MIMO and OFDM gives the effective output and it is cheaper in cost.

The research also presented the performance of LTE systems in the frequency selective fading environment with two receiving and two transmitting antennas [118]. The 4G wireless systems are employed on MIMO (Multiple Input Multiple Output) with OFDM systems. In the spatial multiplexing or transmit diversity, the MIMO system has better performance or throughput as compared to the single antenna based system. The results show that MIMO increases the throughput and OFDM convert the frequency selective fading channels into various flat fading subchannels. The capability and cell range of the LTE network and Mobile WiMax is enhanced to determine the minimum demand and maximum capacity for the network after decreasing the overhead part. In this research, MAC layer and physical layer overhead are analyzed [119]. Minimizing the overhead of the MAC layer and physical layer in 4G LTE and Mobile WiMax networks helps in determining the actual bandwidth necessary to transfer the data. The analysis is done in the DL by modulation, multiplexing, and coding scheme for LTE and WiMax.

2.4.2 QoS Based LTE Scheduling

The scheduling in LTE has posed a challenge for researchers to meet the service flow requirements for different type of traffic to meet QoS constraints. Various classical techniques have been discussed in the earlier research to cater the needs of QoS requirements in different networks. But LTE requirement is constrained on fairness and delay prospective. As different service classes have been assigned different priorities, the services may suffer from resource scarcity. Hence, the intelligent scheme controls the resource allocation in the LTE networks, improves the fairness among different service classes and reduces hand on delays [120]. The queues have been chosen as GBR and non-GBR services to fulfill packet delay budget requirements and improving capacity.

Scheduling in LTE is main function of SSs. An analytical radio scheduler has been designed which is working in both time division and frequency division mode. The time division scheduling mainly works for modeling the channel aware paradigm. The resources have been allocated to different services from SSs to enhance the capacity and delays are reduced [121]. The analytical model is compared with simulation model to achieve QoS differentiation for different service classes. The scheduling algorithms are extended to work upon multidimensional continuous Markov Chain model and AMC techniques are applied on analytical scheduler for enhanced services.

A conceptual framework has been designed to fulfill the requirements of high data rates among machine to machine communication. An enhancement as well as improvement in LTE design is represented as LTE-M which is suitable to efficient machine to machine applications [122]. The existing infrastructure is reused to improve the scheduling strategies, resource allocation, high bandwidth, high data rate transmission as well as smooth coexistence with other communication networks and technologies. Advanced techniques have been described to enhance the performance of wireless networks needed to control interference and manage resources. Smart resource allocation is provided by pico, femto and relays cells acting as low power base stations for range expansion among interfering cells. Fairness and performance gains can be improved along with low latency can be achieved [123].

Wireless mesh networks can operate very well in coexistence with WLAN, WiFi, WiMax and LTE. Defense mesh networks have been proposed in connection with LTE to work on higher data rates as well as to enhance the QoS requirements. Working with OFDM, priority based traffic grooming in defense networks is used to share the increased load traffic as well to reduce the packet loss [124]. The mesh networks can easily manage the congestion in active sectors and optimize the available bandwidth. LTE radio frame working with OFDM/OFDMA technology enhances the bandwidth optimization by allocating the reference bursts among users in a scheduled manner. A transmit receive chain has been discussed to represent the bandwidth division in resource blocks in AMC [125]. The frame format is designed either in TDD or FDD mode. The physical channels may be categorized as communication or control channels. The channel equalization reduces the noise at receiver and less time is taken to receive a reference burst, hence delays are reduced.

Two linear approximation methods have been discussed and analyzed to develop network planning and optimization techniques for LTE [126]. The linearization and the optimization based upon bounding conditions is used potentially for more general convex optimization problems like QoS achievements at high data rates. Different cells with varying load are able to enhance the capacity of the network with the help of discussed algorithm. The poor link quality channels are handed over to hotspots to improve the delay budget requirements. A number of planning and performance models have been designed for channel dependent scheduling but interference limiting is difficult to achieve in cellular networks. OFDMA based scheduled subcarrier optimization algorithm is proposed for optimal SINR with proportional fair scheduling [127]. The model is further refined with uniform modulation and coding, being applied to LTE network for precise SINR distribution. The model can approximately estimate the performance of rate based proportional fair scheduling and can work for simpler prediction models and control algorithms.

The LTE is able to support scalable bandwidth for FDD and TDD modes, operating evenly with already existing networks. Multiantenna techniques and intercell interference coordination are helpful in radio resource scheduling and interference mitigation in LTE. Radio resource scheduling and methods based on interference mitigation to improve the QoS of cell-edge users are discussed in this work [128]. The downlink and uplink scheduling, dynamic and persistent scheduling for VoIP services are compared and channel quality indicator (CQI) values are reported to achieve high

SINR and system throughput. The Table 2.2 gives the comparative analysis of QoS based LTE scheduling algorithms.

Table 2.2: Comparison of QoS based LTE Scheduling Algorithms

Year	Authors	Algorithm	Goal/Key Features	QoS Framework
2013	Zaki et al.	TDM based Channel aware scheduling	Radio scheduler design Markov chain model and AMC	QoS differentiation for different and enhanced services
2012	LSI project	Improvement in scheduling strategies	Design of LTE-M for machine to machine applications	Resource allocation, high bandwidth, high data rate transmission
2011	Qualcomm Report	Smart resource allocation	pico, femto and relays cells acting as low power base stations	Fairness, performance gain and low latency
2017	Shukla et al.	Defense mesh networks with LTE	To work on high data rates & enhanced QoS	Increased load traffic and reduced packet loss
2016	Kamath et al.	Mesh networks with BW division	To manage congestion & optimize BW	Reduced noise at receiver & reduced delays
2012	Siomina et al.	Linear approximation method for network planning	QoS achievement at high data rates	Enhanced capacity of network with varying load
2013	Parruca et al.	Channel dependent scheduling with subcarrier optimization	Poor link quality & interference limiting	Performance estimation of rate based scheduling & controlling SINR distribution
2010	Kwan et al.	Multiantenna techniques with radio resource scheduling	Resource scheduling and interference mitigation to improve QoS	High SINR and system throughput
2012	Kawser et al.	Single antenna & multiantenna transmit diversity techniques	System evaluation for downlink design	Improved data rate, fairness & throughput

A number of scheduling algorithms have been discussed in various research works. The most commonly used are the RR, proportional fair and best CQI. The traffic scheduling plays a vital role in enhancing the spectrum allocation in the most efficient manner. The throughput and BER, SER represent the scheduling performance in the downlink analysis [129]. The CQI performance is found to be most suitable in bad channel conditions during uplink design. As all the algorithms have some advantages and disadvantages, the tradeoff is achieved among all. The investigations made for single antenna and multiantenna transmit diversity techniques in [130] represent the LTE system evaluation for downlink design. RR and proportionate fair schemes are compared in the fading scenario for better data rates, fairness and throughput.

2.5 Load Balancing in Wireless networks

In recent years, with the growth of mobile traffic in wireless environments, the load balancing problem among users and access points (APs) is increasing. Traffic load is randomly distributed throughout the network due to uneven bandwidth allocation methods. The proposed algorithm suggested a fair bandwidth allocation schemes to achieve throughput fairness among APs and users [131]. The bandwidth is divided among all available mobile users in WLAN environment for improving the balancing index. The algorithm can further be implemented for heterogeneous networks.

The three conditions have been considered for load balancing logic implementation as whether the randomly distributed APs are under loaded, overloaded or balanced. The balancing index depends upon the throughput of AP and position of neighboring nodes [132]. The proposed algorithm studied the behavior of Ethernet model for load balancing among APs in a region and initiating the handovers for balancing load among nodes available in the network. The unbalanced load condition in WiMax networks is recovered by implementing the concept of frequency reuse among cells [133]. The algorithm is proposed as handover scheme for congestion control when the cell is overloaded. Frequency reuse concept in the form of frequency assignment is allocated to BS to initiate handovers. The throughput of the APs is increased and handoff latency is reduced. With the increase in available wireless networks, the users are getting the choice to get connected with more large number of APs. The connectivity decision depends upon the type of service as well as the SINR of the available APs. The fairness throughput depends upon the SINR and the delays to get connected. Clients using this technology require up-to-date information about the expected delays at available APs, which can be obtained through active probing [134]. They model this scenario as a load balancing game. Using techniques of empirical game-theoretic analysis, they evaluate range strategies through simulation. They find that variants of the Hedge algorithms are effective at the single unit load balancing game. It is also effective for scheduling multiple jobs per period. However, when delay information can only be obtained from using an access point, all variants of the Hedge algorithm we examined were performed by simple decision-theoretic optimization policies. Another BS controlled handover scheme is proposed for WiMax by implementing the load balancing algorithm in [135]. The efficient load balancing scheme can avoid the ping pong effects and increases the signal quality during handoff.

The load balancing algorithm is one of the well-known topics being discussed these days in cloud environment. It increases the resource utilization as well as reduces the markspan time in the cloud resources. Load balancing can be achieved either by using task scheduling or by machine monitoring approach of task migration. The task scheduling schedules and assigns the task to the available nodes called virtual machines to be performed in specified minimum time duration [136]. The load balancing algorithm helps in managing the upcoming requests and increases the utilization of nodes. The effectiveness of algorithm is to balance the tasks among machines in under loaded as well as overloaded conditions. The spectrum allocation and resource utilization is further improved by cell selection technique in LTE network by considering femto and pico cells [137]. A radio controller and self-organizing network is designed to achieve higher throughput and SINR as compared to conventional Max SINR algorithm. The ping pong effect is decreased and load distribution is improved as the number of mobile users is increasing.

In today's social world, the distributed systems need to work in such a manner that the computing power of the system must be divided among its elements, nodes and network in fair and effective way load balancing scheme [138]. A community based and cluster based load balancing approach has been proposed with a fault tolerance mechanism to deal with event dissemination, replication and load distribution. Filter based functionality approach is designed with each multicast group to achieve fair distribution of load in overall cluster functioning. The fairer load distribution can be achieved by packet schedulers to support different types of traffic with priorities, channel conditions with varying traffic loads. A suitable relationship is required to a generated and maintained for successful load balancing and job scheduling. A number of packet scheduling algorithms have been discussed [139-141] to generate priorities, reduce delays and improve SINR as well as throughput.

2.6 Related Work in the Area of Access Selection in Heterogeneous Networks

The most challenging issue in wireless networks is either to migrate or combine one technology to another as per users prospective or market strategies. The WiMax and LTE are the two most promising and recent 4G technologies to provide support for service competitiveness, ecosystem viability, cost satisfaction and variety of consumer devices to be supportive with. To combine two technologies, planning required is careful and complex process. The requirements, similarities, dissimilarities, features and applications are discussed along with the requirements to combine the two technologies with QoS [142].

QoS mechanism in the heterogeneous NGN is proposed by integrating WLAN and WiMax to ensure that quality is maintained over wireless link. A mapping between the WLAN and WiMax link-layer classes and the NGN Diffserv classes is given, which operate at network layer. The DSCP is marking each packet and causes routers to forward the packets with different priorities, which can be mapped to the wireless link-layer services to ensure that prioritization occurs over the wireless link. The author proposed to map the DSCP to the link-layer scheduling services in WiMax and WLAN according to the priorities [143]. A network architecture showing the entities necessary for interworking each network into the NGN is given. After that, a detailed call flow to set up a QoS enabled call, is described using message sequences.

The demands for wireless connectivity are growing day by day to fulfill quality requirements. The suggested solution was to integrate 3G and IEEE 802.16 networks as per users' requirements. A loosely-coupled architecture is proposed to integrate the 3G and 802.16 networks for attaining QoS and seamless mobility [144]. The architecture consists of session activation, classification procedures and network layer support. A SIP based mobility scheme is also represented that is capable of providing QoS support across different networks as well as across integrated architecture. The sessions in WLAN are initialized and handover occurs between WLAN and WMAN to support service continuity. Handoffs in different networks are based upon bandwidth availability and internetworking simulation scenarios. In future work, the impact of different scheduling techniques of heterogeneous traffic will be studied, making the simulation more applicable to real-life scenarios [145].

Based on CLD (cross layer design) solution, the author proposed a cross layer QoS architecture including the QoS Engine and the cross layer algorithm to guarantee high quality multimedia services in NGNs. The QoS engine consists of daemon, a control module and QoS agent to guarantee high quality wireless multimedia services in 4G environment. The algorithm satisfies all the QoS requirements from applications. Simulation results show that cross layer design strategy has 2% lower average latency, more stable operation and average packet loss ratio about 8.5% lower than the strategy without cross layer design, although it has a slightly lower throughput [146].

A new prioritized resource sharing algorithm for different traffic classes of WiMax-WiFi integrated networks is discussed, which overcomes tradeoff between QoS and channel utilization. The prioritized classes help the other users to not get starved of resource allocation. This algorithm offers reduced probability of blocking and highest channel utilization of WiMax-WiFi users. The number of prioritized channels according to users demands and for various traffic classes with given QoS requirements is considered for future work [147]. Along with that a technology integration framework that defines an optimization of vertical handovers based on context establishment mechanism in heterogeneous environments is discussed. WiMax-WiFi networks are considered which showed handover delays with efficient optimization and managing signaling costs along with QoS management [148].

The current state of the art in heterogeneous networks is 4G deployments focusing 3GPP and LTE interface and mitigation technologies which describe the future of wireless networks. The air interface of NGN consists of spectrum allocation, capacity enhancement, interference management along with the enabling mechanisms for heterogeneous deployments. The load balancing among heterogeneous networks, interference management in networks nodes, cell range expansion and resource portioning is the part of study discussed in this research [149]. An interference cancellation receiver along with control and signal management at the receiver end is discussed. LTE BS and UE transmitter receiver is studied to represent the coexistence of LTE 3GPP with other wireless technologies [150]. When the UE sets the traffic in the heterogeneous network, the congestion occurs because of limited bandwidth availability and delays deteriorate the QoS in the network. To cop up the problem, a self-adaptive packet scheduling algorithm is proposed to reduce the delays and improve the throughput and SINR [151]. Assigning the weight factors to various types of traffic, the packet loss rate is also reduced. The queue overflows may also be controlled by applying probabilistic approach providing minimum cost path to overflow queue with the largest length [152]. The congestion and decay rate is reduced in large length queues which follow a sample path with linear increment. The algorithm is proposed to be optimal in coordinating with large queues and is suitable for heterogeneous traffic networks.

Wireless scenarios are characterized by harsh environments for communication in the indoor as well as outdoor environments. As the WLAN QoS requirements by scarce available resources are increasing, the QoS parameters become difficult to achieve in these environments. The use of OFDM MIMO has extensively been included in several researches to manage the delay sensitive applications. A cross layer scheduling algorithm has been proposed for the downlink multiuser and multiantenna system for heterogeneous traffic scenario [153]. The traffic classes are prioritized with the opportunistic scheduler to maximize the throughput and minimizing delays. The queue length is adjusted as per users' requirements. A LOOP project has been demonstrated for the telecommunication and manufacturing companies providing wireless solution trending towards the convergence in internet and mobile communications [154]. The technology is growing at a rate of 30% faster every year and LOOP is targeting the potential wireless market for process management and manufacturing. The challenges for performance metric and delay sensitive applications have also been addressed and coordinated.

IPv6 mobile networks are considered the future mobile networks for mobility management in all IP networks. The significant latency is involved while mobility has been extended with handover process. The delays and packet loss during handover at MN or access network can be reduces by proposed handover process for IPv6 protocol aiming at providing network-based mobility support [155]. The buffering cost is also evaluated for the given network. As the data and user services are increasing along with traffic variability, the single vector/matrix solutions are ineffective to fulfill the requirements. An efficient classification approach has been proposed to support multiclass, multilevel services in ubiquitous mobile computing environment [156]. The context classification becomes more effective and efficient for multiple subtasks and training processes in the mobile heterogeneous environment. To overcome the challenges in mobile multimedia communication systems, the three smart communities have been discussed namely mobile social learning, mobile event guide and context-aware services [157]. Some new issues related to these systems and some new paradigms for mobile device users are represented which necessitates the development of these systems. Factors such as multimedia information overload and user preferences/ interests as well as mobile device challenges and limitations are also discussed.

To support network mobility in next generation wireless networks, an effective handover scheme is required to be optimized for high speed scenarios. Seamless mobility management architecture is proposed for heterogeneous networks to support vehicular mobility in multiple mobile routers [158]. The architecture is extended to include TCP and UDP service performances. The throughput, service disruption rate, signaling overhead during handover, packet loss rate is improved by simulations providing seamless mobility across heterogeneous networks. The future NGNs communication is converged with Machine to machine communication, where a large number of devices are connected with wired/wireless links. The research finds applications in wide networking areas including e-health care, smart home devices and networks, environment monitoring and automation industries [159]. The various challenges are addressed for 3GPP, 4G and future networks investigating features in characterization and supporting attributes for heterogeneous communication networks.

2.7 Research Gaps

Since IEEE 802.16 standard emerged as a broadband technology, as most applications broadband applications are with heterogeneous QoS requirements. To cope with the issues of QoS requirements, several QoS differentiated schemes have been put forward along with admission control approaches, reservation protocols and a variety of scheduling schemes. To provide better understanding of research challenges of QoS provisioning, the four major aspects being considered are: service classes mapping policy, QoS architecture, combined scheduling and cross layer design [160].

Many queue scheduling schemes have been proposed to achieve tradeoff among fairness, delay and complexity in achieving QoS. Most studies of scheduling algorithms have been focused on optimizing the long term average throughput of users, delay sensitive applications and load balancing conditions. The various issues discussed in NGN are packet loss rate, packet delivery ratio, network scalability, variation in links, routing, energy constraints, security and reliability, internetworking and power consumption.

Multilevel scheduling algorithms are required as heterogeneous computation platforms for parallel computing applications which are quite popular topic of research these days. They offer high computation power and easy to access in different scenarios for heterogeneous networks. The parallel algorithms provided by different studies are still slow in processing and are not able to coordinate with the properties of communication systems [161]. Task scheduling with parallel processing is one of the leading approaches towards modifying traditional parallel algorithms. The speed of job applications can be increased and computational cost and computation time are decreased. The multiprocessor computing consists of more than one processor to execute the parallel tasks for massive amounts of data. The system performance is enhanced depending upon the scheduling algorithm imposed upon the task scheduler. Different task based scheduling algorithms are compared and performance evaluation is done to reduce the time delays [162]. The task scheduling can be classified as static task scheduling or dynamic task scheduling. In static scheduling, the decisions can be made at execution time while dynamic scheduling decision making is done at run time. A static scheduling algorithm is proposed to sequence the tasks during execution and assigning priorities in heterogeneous environment [163]. The main objective of task scheduling algorithms is to minimize the schedule length and hence makespan time and execution time are reduced. Several scheduling algorithms have been proposed for homogeneous and heterogeneous systems may further be categorized as clustering algorithms, random search algorithms, list based and duplication based algorithms. List based scheduling is supposed to better to achieve good QoS [164]. The clustering algorithm for communication networks in heterogeneous environment is discussed in which the number of clusters is greater than number of processors available [165-166]. The clusters are mapped and task scheduling is implemented on each cluster. The congestion is avoided and scheduling time is decreased to attain high QoS. High number of tasks is assigned to unbounded number of processors to obtain duplication based scheduling [167]. Theses algorithms are assigned to heterogeneous networks having high complexity of scheduling.

The dynamic scheduling algorithms for parallel processing are rapidly developing research topic these days. Many dynamic scheduling schemes are discussed to design and achieve the goals such as minimum resource idleness, maximum resource utilization, decreasing make span time, compilation time and reducing task delays. Different aspects of scheduling algorithms re discussed and compared for various issues in homogeneous and heterogeneous systems [168]. More than one number of DAG scheduling onto heterogeneous systems is applied to enhance the fairness in task scheduling and reducing the makespan time, hence increasing utilization [169]. QoS analysis for the same is proposed to be as future work for this research. High degree of parallelism can be achieved with DAG represents the maximum number of tasks to be scheduled concurrently [170]. A new scheduling algorithm has been proposed for heterogeneous computing systems to enhance the scheduling performance [171]. The proposed scheme is compared with number of already existing algorithms and performs better in terms of schedule length and efficiency.

The two main drivers for 5G are mobile internet and internet of things (IoT) which are expected to be the upcoming deployments available by 2020. The networked society in near future provides wireless connectivity to all kind of devices and services. 5G will generate a smart connectivity between users and all things. Be it a smart home, smart automotive industry, intelligent appliances, remote sensing, remote computing and virtual reality, smart hospitals and medical monitoring devices. The scenario for the whole world will change as 5G is expected to entertain almost all the challenges addressed in 4G like higher data rates, capacity enhancement, lesser delays, massive connectivity and reduced cost [172-174]. It also aims to enhance the connectivity, scalability and energy efficiency of the future networks.

2.8 Summary

In this chapter, a number of issues related to scheduling in heterogeneous networks have been studied. Major problems related to QoS in heterogeneous networks and scheduling algorithms have been identified and solutions are proposed by different research works. The proposed solutions are rarely able to match the QoS constraints to be defined and proposed for future generation wireless networks.

Chapter 3

ANALYSIS OF QUEUING MODELS FOR QoS BASED TRAFFIC SCHEDULING

3.1 Introduction

In day to day services, the experience of waiting in queues is annoying always as no one wants to wait. The waiting lines experienced by us are mostly in supermarkets during billing, banks, cinemas and railway station or airport. In this modern world, the waiting lines are also encountered in communication networks, transportation systems, production units and supply chain management systems are some of the examples represented by Fig. 3.1. The requirement of the customers is never to wait in queues or the waiting time should be as small as possible. In telecommunication systems, the packets coming from different channels are required to queue in single interface to forward them towards destination. Hence, queuing plays a big role in packet communications to buffer data packets and aggregating them for common applications. In the era of networking, telecommunications and data processing services, queuing theory is quiet advantageous.

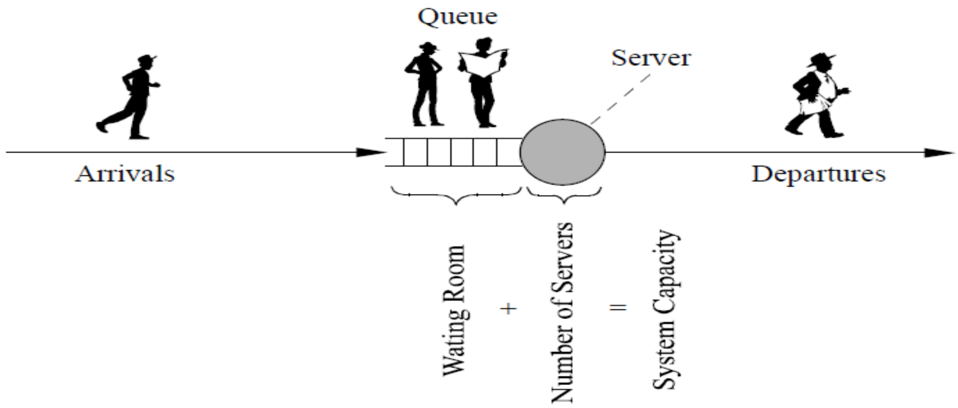

Figure 3.1: Schematic Representation of a Queuing System [175]

Various delays are encountered in queuing theory as delay occur during processing at node level, delays in queuing and during propagation; the aggregation of all these delays represents the total waiting time in processing.

3.2 Queuing Models

The customers waiting in a queuing system generally want to access some type of service. To achieve their aim, they are rarely interested in waiting services as could have utilized the time in better way. So the requirement is to reduce the waiting time and processing delays. The system performance depends upon the server and task manager, may be high or low. If the tasks are increased, the queue length is increased and the quality is degraded along with customer satisfaction is affected. To make the system effective, several measures are taken in the queuing system to satisfy overall performance is characterized as follows [176]:

3.2.1 Number of customers in the system: The random number of customers in steady state is defined by random variable N. If the probability that n customers are present in a system is denoted by p_n at n=N, the average number of customers in a system is,

$$E[N] = \sum_{i=0}^{\infty} n \ . \ p_n \tag{3.2.1}$$

3.2.2 System time and queuing time: The total time spent by the customer in the system, from its start of arrival in the queue to the end of departure from the system is called as system time. The system time is also called service time consisting of waiting time in a queue i.e. queue time and time taken to receive the service. The mean waiting time in a queue in represented as $E[W_q]$.

3.2.3 System utilization: In a queuing system with a single server, the utilization U is defined as the amount of time that the server is busy. If the rate at which customers arrive in the queuing system is λ and if μ is the rate at which these customers are served, then the utilization is equal to $\rho = \frac{\lambda}{\mu}$ (where ρ is identified with the utilization). In the case of queuing system with multiple servers, the utilization is defined as the average fraction of server that is active $U = \frac{\lambda}{c\mu}$.

3.2.4 System throughput: The throughput in a queuing system is denoted by X which is the average number of tasks processed in a unit time. The customers reach the system, getting served and leave, the throughput is denoted by λ, the arrival rate of customers.

3.2.5 Traffic intensity: We define the traffic intensity as the rate of tasks entering the system, so it is given as the product of the average arrival rate of customers and the service time, $\lambda\bar{x} = \frac{\lambda}{\mu}$ where $\bar{x} = \frac{1}{\mu}$ and μ is the mean service rate. In a queuing system with a single server the traffic intensity is equal to the utilization. For multiple servers, the traffic intensity is equal to cU.

Queuing models are generally constructed to represent the steady state of a queuing system that is the long run or average state of the system.

3.3 Moments of Queue Length

The behavior of many communication systems is modeled with the queuing theory as moments of queue length play a major role in describing the communication models. The various tasks may be considered as jobs distributed with the moments of queues may bind the waiting time and may help to reduce it. The different moments affect the mean waiting time in different ways. The performance evaluation of a communication system can be improved or optimized with the knowledge of production and queue distribution. Networks can be well optimized with moment recursion formulas. The recursion models may increase the performance and decrease the cost of a network, hence help in generating an efficient networking architecture.

3.4 Moment Recursive Formula of Various Queues

3.4.1 M/M/1 Queue

M/M/1 is represented as a queue with single server, arrival rate λ following a Poisson process and service time distribution is exponential. The mean service time is $1/\mu$. As the probability of queue with 'n' number of customers or packets is defined as P_n, the state transition equations [175] for single server Markovian model is:

$$\begin{cases} \mu P_1 = \lambda P_0 \\ \mu P_{n+1} = (\lambda + \mu) P_n - \lambda P_{n-1} \qquad (n = 1,2,3 \ldots) \end{cases} \tag{3.4.1}$$

The queue load is denoted by $\rho = \lambda/\mu$ and it is required to be less than 1 to make the system stable. In the equilibrium state, the probability of empty system is considered to be P_0 and according to little's law,

$$1 - P_0 = \frac{\lambda}{\mu} = \rho \tag{3.4.2}$$

Packet number distribution

$$P_n = \rho(1-\rho)^n \qquad (n = 0,1,2,\dots,\infty) \tag{3.4.3}$$

Average number of packet

$$\bar{N} = \sum_{n=1}^{\infty} nP_n = \frac{\rho}{1-\rho} \tag{3.4.4}$$

Moment generating function is obtained by taking the derivative of equation as follows:

$$P(z) = \frac{1-\rho}{1-\rho z} \tag{3.4.5}$$

Considering λ as the arrival rate and load offered in M/M/1 queue is ρ, and $M_0 = \sum_{n=0}^{\infty} P_n = 1$, then in the queue length of kth moment $M_k (k \geq 1)$ represents [177]:

$$M_k = \frac{\sum_{l=2}^{k+2} \binom{k+l}{l} [(-1)^l + \rho] M_{k+1-l} - (-1)^{k+1}[1-\rho]}{(k+1)(1-\rho)} \tag{3.4.6}$$

And the higher moments can also be recursively calculated. The moments of queue length with respect to utilization factor are plotted in Fig. 3.2. With respect to utilization factor, the higher moments of queue length represent the extended utilization. As the moments are increasing, the utilization is also increasing.

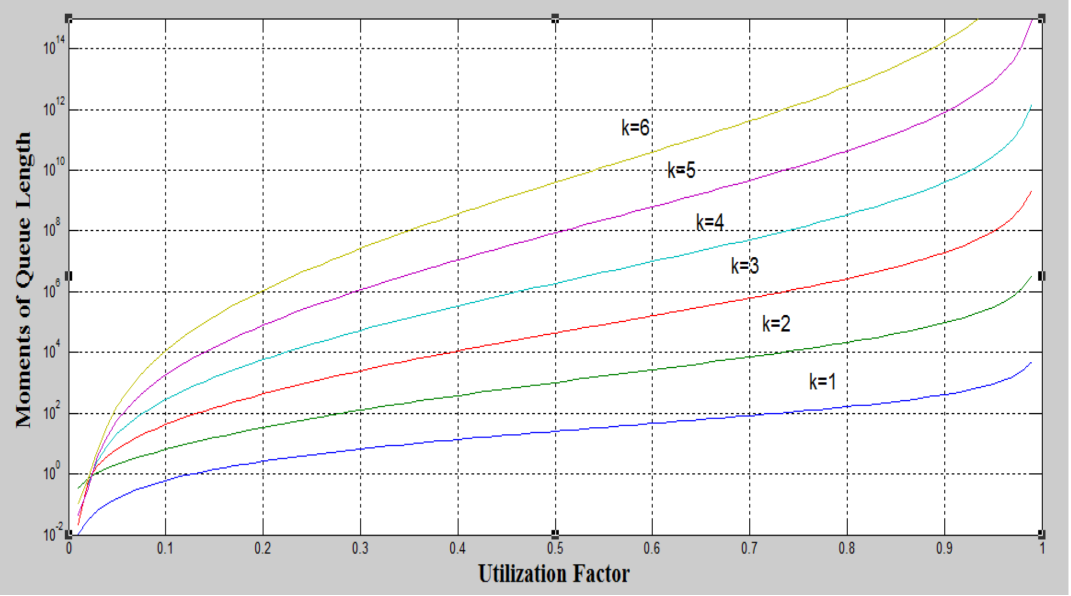

Figure 3.2: M/M/1 Queue Length Moments with Utilization Factor

3.4.2 M/M/1/B queue

If the number of arriving packets is constrained to B, the system model is represented as M/M/1/B queue. Additionally, one more packet is considered to be arrived with B packets already in service, the model will not allow the extended packet to enter the system. The state transition equations for this model are as:

$$\begin{cases} \mu P_1 = \lambda P_0 \\ \mu P_{n+1} = (\lambda + \mu) P_n - \lambda P_{n-1} \quad (n = 1,2,3 \ldots \ldots \ldots, B-1) \\ \mu P_B = \lambda P_{B-1} \end{cases} \tag{3.4.7}$$

P_0 and P_B are the probabilities of empty and blocking model. Applying little's law, the equation is defined as:

$$1 - P_0 = (1 - P_B)\frac{\lambda}{\mu}$$

For $\rho < 1$, the following equations are obtained.

Packet number distribution

$$P_n = \frac{(1-\rho)\rho^n}{1-\rho^{B+1}} \quad (n = 0,1,2, \ldots \ldots \ldots . B)$$

Probability of system being empty,

$$P_0 = \frac{1-\rho}{1-\rho^{B+1}}$$

Packet blocking probability,

$$P_B = \frac{(1-\rho)\rho^B}{1-\rho^{B+1}}$$

By taking the derivative of the above equation, the various moments of queue length are obtained.

With the arrival rate and service time of λ and$1/\mu$, the offered load $\rho = \frac{\lambda}{\mu} < 1$, in an M/M/1/B queue, $M_k = \sum_{n=0}^{B} P_n = 1$, then the kth moment $M_k (k \geq 1)$ of queue length is [177]:

$$M_k = \frac{\sum_l^{k+1} \binom{k+1}{l} [(-1)^l + \rho] M_{k+1-l} - (-1)^{k+1}[1-\rho]}{(k+1)(1-\rho)} - \frac{\rho P_B[(B+1)^{k+1} - B^{k+1} + (-1)^{k+1}]}{(k+1)(1-\rho)} \qquad (3.4.8)$$

By considering $B \rightarrow \infty$ and $P_B \rightarrow 0$, the equation for moments of queue length may be verified as:

$$M_1 = \frac{\rho}{1-\rho} - \frac{(B+1)\rho P^B}{1-\rho}$$

$$M_2 = \frac{(1+\rho)M_1}{1-\rho} - \frac{[(B+1)^3 - B^3 - 1]\rho P^B}{3(1-\rho)}$$

$$M_3 = \frac{6(1+\rho)M_2 - 4(1-\rho)M_1 + 2\rho}{4(1-\rho)} - \frac{[(B+1)^4 - B^4 + 1]\rho P_B}{4(1-\rho)}$$

The moments of queue length with respect to utilization factor are plotted in Fig. 3.3. With limited number of packets, the utilization factor is reduced for higher moments of queue length as compared to M/M/1 model. As the load is decreased, the utilization is decreased. But for higher moments, the response of the system is better as compared for the lesser value of k.

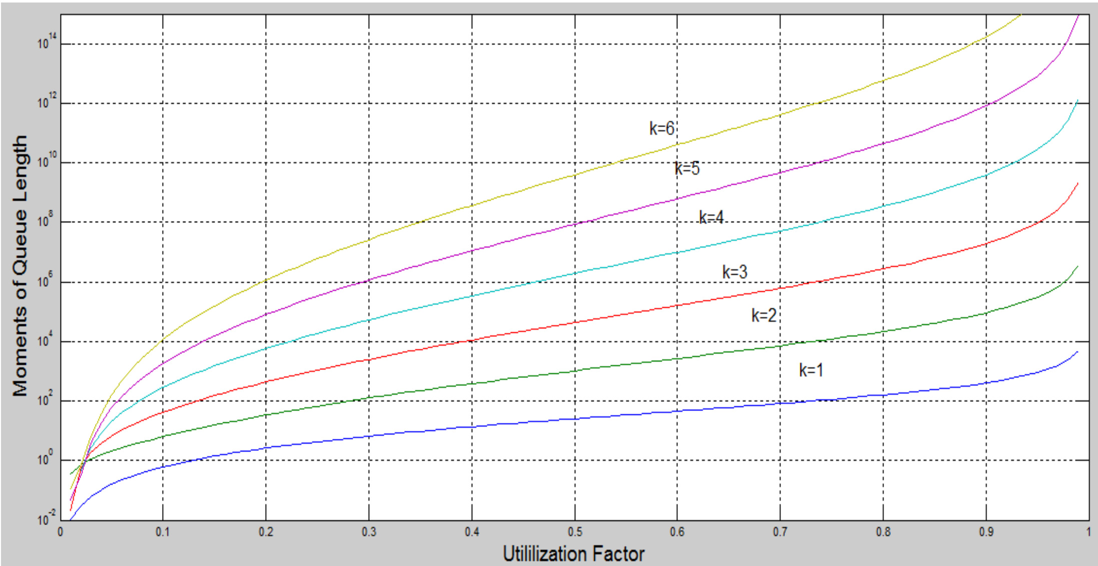

Figure 3.3: M/M/1/B Queue Length Moments with B = 8

3.4.3 M/M/c queue

The queue model with c server having identical characteristics is presented as M/M/c with packet arrival rate and a mean service distribution time of λ and 1/μ. If the number of packets n is less than number of servers c, all the packet will get direct access to servers else the packets are send to queue. If the load offered to the system is defined as $\rho = \frac{\lambda}{\mu}$, the corresponding transition equations for the system model are represented as[178]:

$$\begin{cases} \mu P_1 = \lambda P_0 \\ (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n - \lambda P_{n-1} \quad (n = 1,2,3 \ldots \ldots \ldots, c-1) \\ c\mu P_{n+1} = (\lambda + c\mu)P_n - \lambda P_{n-1}, \qquad n \geq c \end{cases} \tag{3.4.9}$$

The packet number distribution as follows:

$$P_n = \begin{cases} \frac{(c\rho)^n}{n!} P_0, & n = 0,1, \dots, c-1 \\ \frac{c^c \rho^n}{c!} P_{0,} & n = c, c+1, \dots \end{cases} \tag{3.4.10}$$

The average number of packets in the queue could be written as:

$$N = M_1 = c\rho + \rho \frac{(c\rho)^c}{c!} \frac{P_{0,}}{(1-\rho)^2}$$

In an M/M/c queue, with the load offered is $\rho = \lambda / n\mu < 1$, and $M_0 = \sum_{n=0}^{c} P_n = 1$, the kth moment $M_k (k \geq 1)$ of queue length satisfy:

$$M_k = \frac{\sum_{l=2}^{k+1} \binom{k+1}{l} M_{k+1-l} (\rho + (-1)^l) - (-1)^{k+1} P_0}{(k+1)(1-\rho)} - \frac{\sum_{n=1}^{c-1} P_0 \frac{n^{k+1} (\rho c)^n}{n!} \left[\rho \left(\frac{c}{n+1} - 1 \right) - \left(1 - \frac{n}{c} \right) \right]}{(k+1)(1-\rho)} \tag{3.4.11}$$

The moments of queue length with respect to utilization factor are plotted in Fig. 3.4. The number of servers has increased the utilization factor along with the higher moments of queue length as represented in Fig. 3.4.

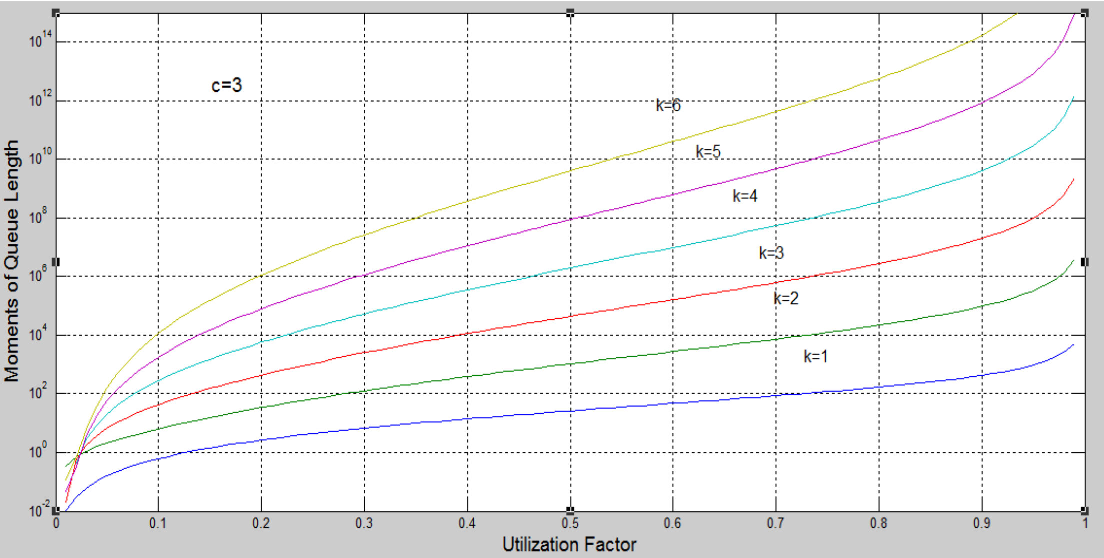

Figure 3.4: M/M/c Moments of Queue Length

3.4.4 M/M/c/0 queue

When the waiting space is removed in the M/M/c/0 queue then either the packets will be occupied by the servers, denied permission or dropped by the system. The system with transition equations is represented as:

$$\mu P_1 = \lambda P_0$$
$$(n+1)\mu P_{n+1} = (\lambda + n\mu)P_n - \lambda P_{n-1} \quad (n = 1,2,3 \dots\dots\dots, c-1)$$
$$c\mu P_{n+1} = (\lambda + c\mu)P_n - \lambda P_{n-1}, n \geq c \tag{3.4.12}$$

With a mean arrival rate and service rate of λ and 1/μ in an M/M/c/0 queue, when the offered load $\rho = \lambda/\mu < 1$, and $M_0 = \sum_{n=0}^{c} P_n = 1$, the kth moment $M_k (k \geq 1)$ of queue length satisfy [178]:

$$M_k = \frac{\sum_{l=2}^{k+1} \binom{k+1}{l} [(-1)^l \mu M_{k+2-l} + \lambda M_{k+1-l}]}{(k+1)\mu} + \frac{\sum_{n=1}^{c-1} P_0 \frac{n^{k+1}(\rho c)^n}{n!} \left[\rho\left(\frac{c}{n+1} - 1\right) - \left(1 - \frac{n}{c}\right)\right]}{(k+1)\mu} \tag{3.4.13}$$

3.4.5 M/M/∞ Queue

If the number of servers have been increased up to infinity, none of the packets are required to wait or no packet is discarded and no service denials. Here, we consider that M/M/∞ queue have infinite number of servers. If mean arrival rate of the queue is λ and mean service time $1/\mu$, when the offered load, $\rho = \frac{\lambda}{\mu} < 1$ define $M_0 = \sum_{n=0}^{\infty} P_n = 1$, then the kth moment $M_k (k \geq 1)$ of queue length satisfies [178]:

$$M_k = \frac{\sum_{l=2}^{k+1} \binom{k+1}{l} [(-1)^l \mu M_{k+2-l} + \lambda M_{k+1-l}] + \lambda(k+1)M_k}{(k+1)\mu} \tag{3.4.14}$$

3.4.6 $M/E_r/1$ Queue

Besides specifying the number of customers in the system, the customers have to satisfy the number of stages in a particular service model. If for k number of customers, j is the number of stages left out of total r stages in a system and the server is at ith stage, the stages may be represented with the relation [179]:

$$j = number\ of\ stages\ left\ in\ the\ system = (k-1)r + (r-i+1)$$
$$= rk - i + 1.$$

For this case, the probability for k number of customers is defined as P_k and at equilibrium condition, it is analysed as:

$$p_k = \sum_{j=(k-1)r+1}^{kr} P_j, k = 1,2,3, \ldots$$

Focusing on stage n, it is seen that transition to stage n from a state which is r positions below (stage n-r) is because of new r stages for an additional customer entering the system. Transition to next stage is occurs when one stage is completed from r processes. In this system, the mean arrival rate is and the average service time is independent of r and held fixed at 1/μ. Thus the utilization factor is $\rho = \lambda/\mu$. The equilibrium state transition equations can be written by using flow conservation method:

$$\begin{cases} \lambda P_0 = r\mu P_1 \\ r\mu P_{n+1} = (\lambda + r\mu)P_n - \lambda P_{n-r}, \qquad 1 \le n \le \infty \end{cases} \tag{3.4.15}$$

With mean arrival rate and mean service time of λ and $1/\mu$ in an $M/E_r/1$ queue, the load offered is $\rho = \lambda/\mu < 1$, and $M_0 = \sum_{n=0}^{\infty} P_n = 1$, the kth moment $M_k(k \ge 1)$ of queue length is defined as [178]:

$$M_k = \frac{\sum_{l=2}^{k+1}\binom{k+1}{l}[(-1)^l r + r^l \rho]M_{k+1-l} - r(-1)^{k+1}[1-\rho]}{r(k+1)(1-\rho)} \tag{3.4.16}$$

Based on the previous equation we could obtain the formula. When r=1 in the previous equation (3.4.16), we get the recursive formula for M/M/1 queue.

The moments of queue length with respect to utilization factor are plotted in Fig. 3.5 and Fig. 3.6. Fig. 3.7 compares 1st moment and 2nd moment at various values of r.

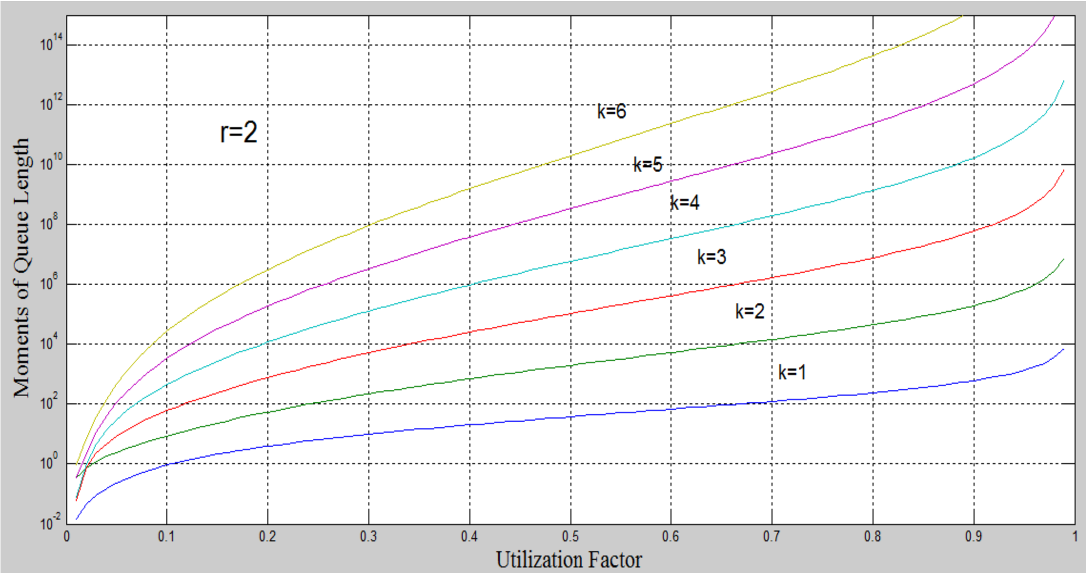

Figure 3.5: M/ E_r /1 (r=2) Queue Length Moments

Comparing the number of stages, the higher value of moment is achieved with respect to increase in utilization factor. The number of stages may able to handle the larger load represented by the utilization factor.

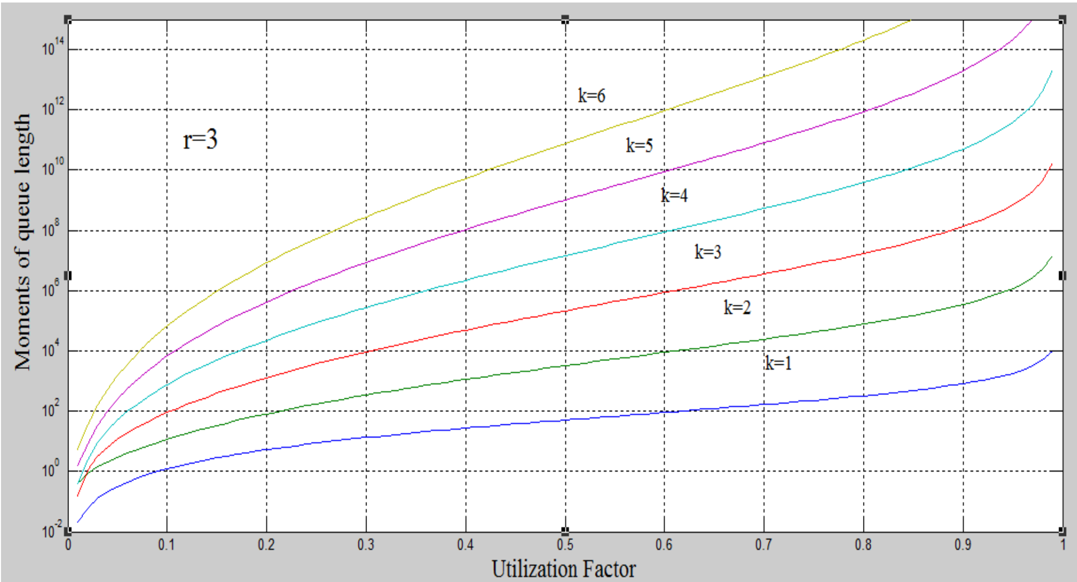

Figure 3.6: M/ Er /1 (r=3) Queue Length Moments

As the number of stages 'r' in a system is increasing, the utilization factor is increasing along the number of moments of queue length are also increasing.

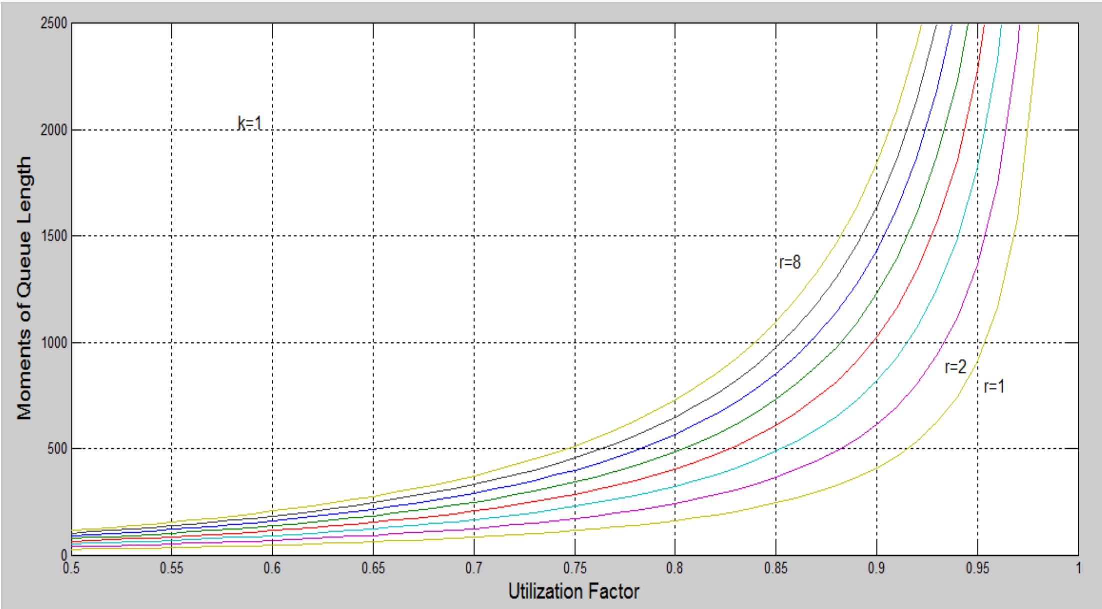

Figure 3.7: 1st Moment of Queue Length for M/ E_r /1 (r=1 to 8)

3.4.7 $E_r/M/1$ Queue

This system is represented as dual of M/E_r/1 as function of service time and arrival time have been interchanged. Using the same method as above for $E_r/M/1$ queue with mean arrival rate $r\lambda$ and mean service time $1/\mu$, when the load offered is $\rho = \frac{\lambda}{\mu} < 1$, and , we get the recursive formula for the kth moment M_k $(k \geq 1)$ of queue length as follows [178]:

$$M_k = \frac{\sum_{l=2}^{k+1}\binom{k+1}{l}[(-1)^l r + r^l \rho]M_{k+1-l} - r(-1)^{k+1}[1-\rho]}{r(k+1)(1-\rho)} \qquad (3.4.17)$$

The moments of queue length with respect to utilization factor are plotted in Fig. 3.8. The Fig. 3.9 represents variations in 2nd moment of E_r/M/1 at various values of r.

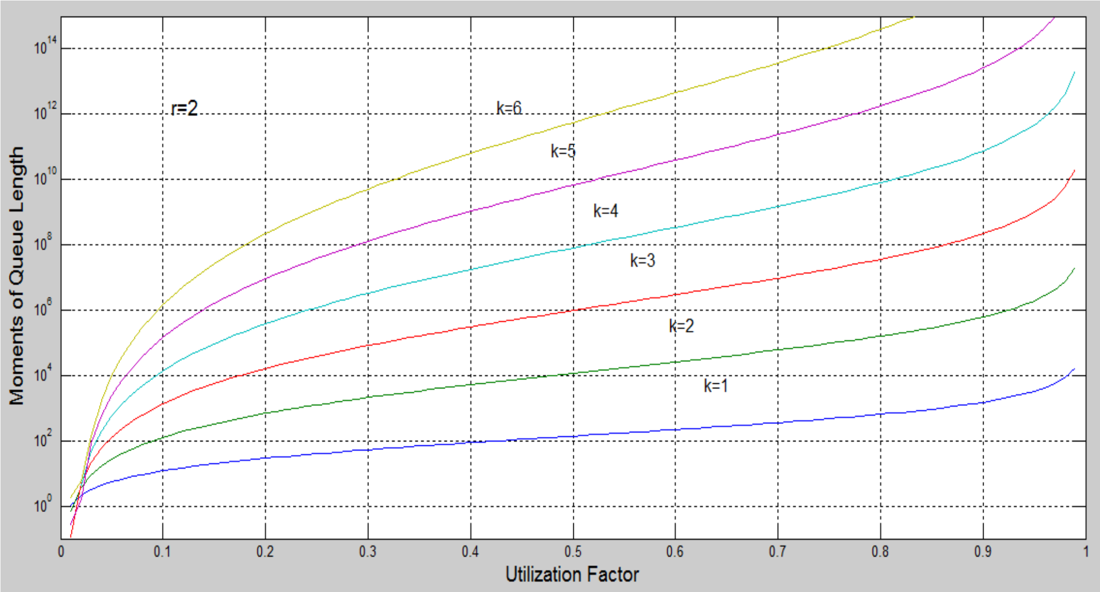

Figure 3.8: E_r /M/1 (r=2) Queue Length Moments

The result represent the moment of Erlang queue lengths with respect to different stages in the queue. The number of stages may enhance the utilization as shown in Fig. 3.9. The higher utilization factor is achieved with the higher moments of queue length.

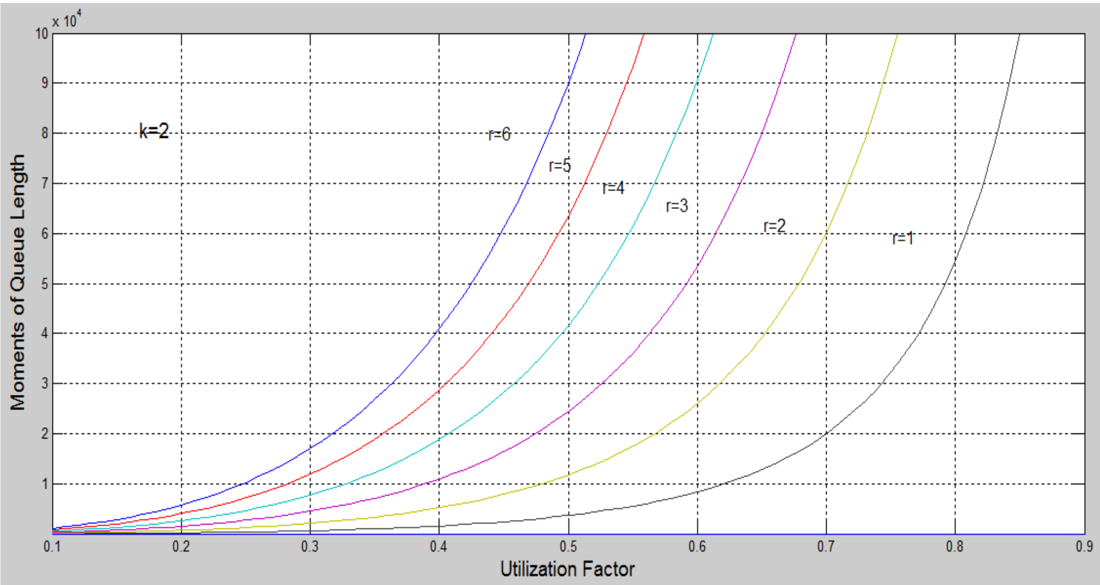

Figure 3.9: 2nd Moments of Queue Length for E_r /M/1 (r=1 to 6)

3.4.8 M/M$^{[K]}$/1 Queue

In this model, the single server is able to serve k number of customers at a time with an arrival rate λ, if the number of customers in service are less than k. The new customers will enter the server and will finish up with already existing customers with no waiting time. This is represented as batch service with bulk arrivals, when the capacity of batch is full with k customers [175]. The steady state equation for the model is as follows:

$$0 = -(\lambda + \mu)p_n + \mu p_{n+K} + \lambda p_{n-1}, \quad n \geq 1 \qquad (3.4.18)$$

$$0 = -\lambda p_0 + \mu p_1 + \mu p_2 + \ldots \ldots + \mu p_{K-1} + \mu p_K$$

If $r_1, r_2, \ldots \ldots \ldots \ldots . . r_{k+1}$ are the roots of the characteristics equation, then

$$p_n = \sum_{i=1}^{K+1} a_i r_i^n , \quad n = 0,1, \ldots \ldots \ldots .,$$

Where a_i are the constant

$$p_n = (1 - r_0) {r_0}^n, \qquad n \geq 0; \;\; 0 < r_0 < 1$$

The average number of customer in the model is

$$L_s = \frac{r_0}{1 - r_0}$$

$$p_K = \frac{\lambda}{\mu} p_0 = C r_0^K$$

$$\Rightarrow C = \frac{\lambda}{\mu r_0^K} p_0$$

$$\therefore \; p_i = \frac{p_0 \lambda}{\mu} r_o^{i-K} , \qquad i = K, K+1, \ldots \ldots \ldots \ldots ....$$

Hence the probabilities p_n for n = 1, 2........are given by

$$p_n = \begin{cases} \frac{p_0\left(1-r_0^{n+1}\right)}{1-r_0} & , n = 1,2, \dots\dots\dots\dots\dots K-1 \\ \frac{p_0\lambda}{\mu} r_0^{n-K} & , n = K, K+1, \dots\dots\dots\dots. \end{cases} \quad (3.4.19)$$

The only probability yet to be found is p_0 which is indeed obtained using the normalizing equation

$$\sum_{n=0}^{\infty} p_n = 1$$

$$p_0 = \frac{\mu(1-r_0)^2}{\mu K(1-r_0)} = \frac{1-r_0}{K}$$

With arrival rate and mean service time of λ and $1/\mu$ in a $M/M^{[K]}/1$ queue, when the load offered is $\rho = \frac{\lambda}{\mu} < 1$, and define $M_0 = \sum_{n=0}^{\infty} P_n = 1$, then the kth moment $M_k (k \geq 1)$ of queue length is given as:

$$M_k = \frac{\sum_{l=2}^{k+1}\binom{k+1}{l} M_{k+1-l}(\rho + (-K)^l)}{(k+1)(K-\rho)} - \frac{\sum_{n=1}^{K-1} P_0 \frac{n^{k+1}}{1-r_0}\left[l - r_0^{n+K+1} + (\rho+1)^{1-r_0^{n+1}} - \rho(1-r_0^n)\right]}{(k+1)(K-\rho)} \quad (3.4.20)$$

The equation can be verified by putting $r_0 = \rho$ then K=1, equation (3.4.20) will be same as (3.4.6). The proof of the bulk arrival equation is as follows:

When $k \geq 1$, define $M_0 = \sum_{n=0}^{\infty} P_n = 1$. Multiplying n^{k+1} on both sides of the second equation (3.4.7) yields

$$\mu P_{n+K} = (\lambda + \mu)P_n - \lambda P_{n-1}$$

Multiply by n^{k+1} on both sides

$$n^{k+1}P_{n+K} = [(\lambda+\mu)P_n\, n^{k+1}] - (\lambda P_{n-1}\, n^{k+1})$$

$$\mu\, [(n+K) - K]^{k+1}\, P_{n+K} = [(\lambda+\mu)P_n\, n^{k+1}] - \left[\lambda\left((n+1)+1\right)^{k+1} P_{n-1}\right]$$

$$\mu\left[\sum_{l=0}^{k+1}\binom{k+1}{l}(n+K)^{k+1-l}(-K)^{l}\right]P_{n+K}$$

$$= [(\lambda+\mu)P_n n^{k+1}] - \left[\lambda\sum_{l=0}^{k+1}\binom{k+1}{l}(n-1)^{k+1-l}P_{n-1}\right]$$

$$\mu\sum_{l=0}^{k+1}\binom{k+1}{l}(n+K)^{k+1-l}(-K)^{l}P_{n+K}$$

$$= \left[(\lambda+\mu)\sum_{n=K}^{\infty}n^{k+1}P_n\right] - \left[\lambda\sum_{l=0}^{k+1}\binom{k+1}{l}\sum_{n=K}^{\infty}(n-1)^{k+1-l}P_{n-1}\right]$$

$$\left[\mu\left(\sum_{l=0}^{k+1}\binom{k+1}{l}(-K)^{l}\right)\left(\sum_{n=K}^{\infty}(n+K)^{k+1}P_{n+K}^{l}\right)\right]$$

$$= \left[(\lambda+\mu)\left(\sum_{n=K}^{\infty}n^{k+1}P_n\right)\right]$$

$$-\left[\lambda\sum_{l=0}^{k+1}\binom{k+1}{l}\left(\sum_{n=K}^{\infty}(n-1)^{k+1-l}P_{n-1}\right)\right]$$

$$\sum\nolimits_{l=2}^{k+1}\binom{k+1}{l}M_{k+1-l}\left(\lambda+\mu(-K)^{l}\right)-(-K)^{k+1}P_0\big)-\mu\sum_{n=1}^{K-1}P_0\frac{(1-r_0{}^{n+K+1})}{1-r_0}$$

$$= (-\lambda-\mu)\sum_{n=1}^{K-1}P_0\frac{(1-r_0^{n+1})}{1-r_0}n^{k+1}+\lambda\sum_{n=1}^{K-1}n^{k+1}P_0\frac{(1-r_0^{n})}{1-r_0}$$

$$+ M_K(K+1)[\mu K-\lambda]$$

$$M_k = \frac{\sum_{l=2}^{k+1}\binom{k+1}{l}M_{k+1-l}\left(\rho+(-K)^{l}\right)-(-K)^{k+1}P_0\big)}{(k+1)(K-\rho)}$$

$$-\ \frac{\sum_{n=1}^{K-1}P_0\dfrac{n^{k+1}}{1-r_0}\left[1-r_0^{n+K+1}+(\rho+1)^{1-r_0^{n+1}}-\rho(1-r_0^{n})\right]}{(k+1)(K-\rho)}$$

The moments of queue length with respect to utilization factor are plotted in Fig. 3.10 and Fig. 3.11. Bulk queuing undoubtly may able to handle larger loads. As shown in Fig. 3.10, the utilization factor increases the moments handled by the bulk queuing model. The increase in value of k increases the moments handled by the bulk queuing model as shown in Fig. 3.11.

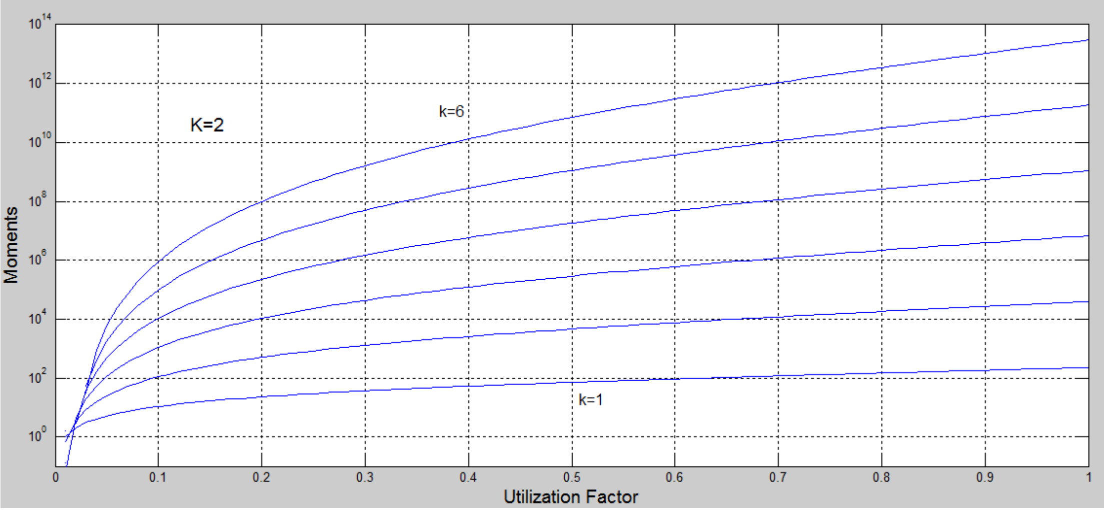

Figure 3.10: Moments of Queue Length for $M/M^{[K]}/1$ (K=2)

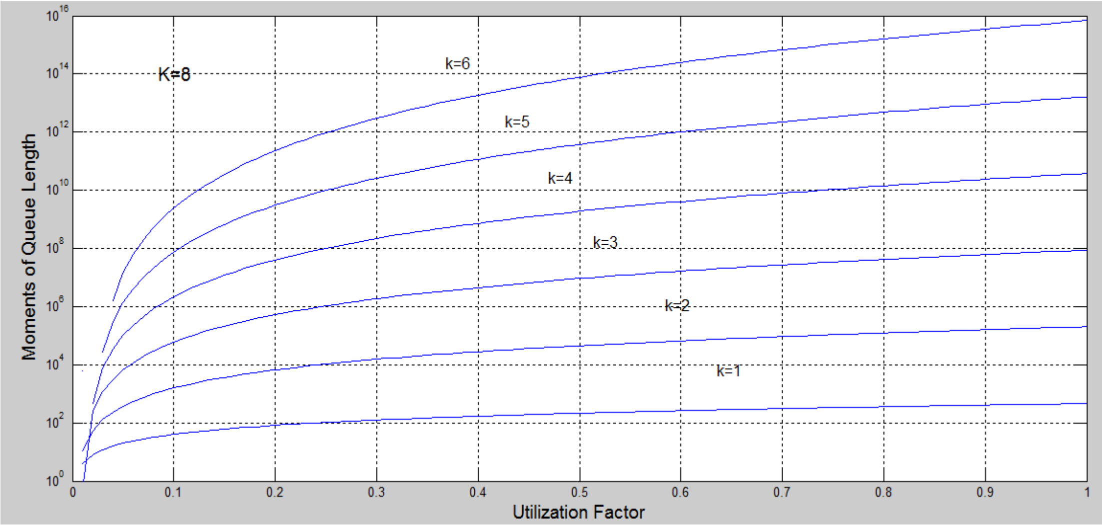

Figure 3.11: $M/M^{[K]}/1$ (K=8) Queue Length Moments

The moments of various queues such as M/M/1, $M/E_r/1$, $E_r/M/1$ and the derived $M/M^{[K]}/1$ is compared in Table 3.1 given below. The Fig. 3.12 and Fig. 3.13 illustrate the difference between above mentioned moments of queues for 2^{nd} and 3^{rd} moment.

Both the results reflect the bulk queuing model as most efficient as the utilization is highly achieved. By using these values various parameters are analyzed which depends on the utilization factor.

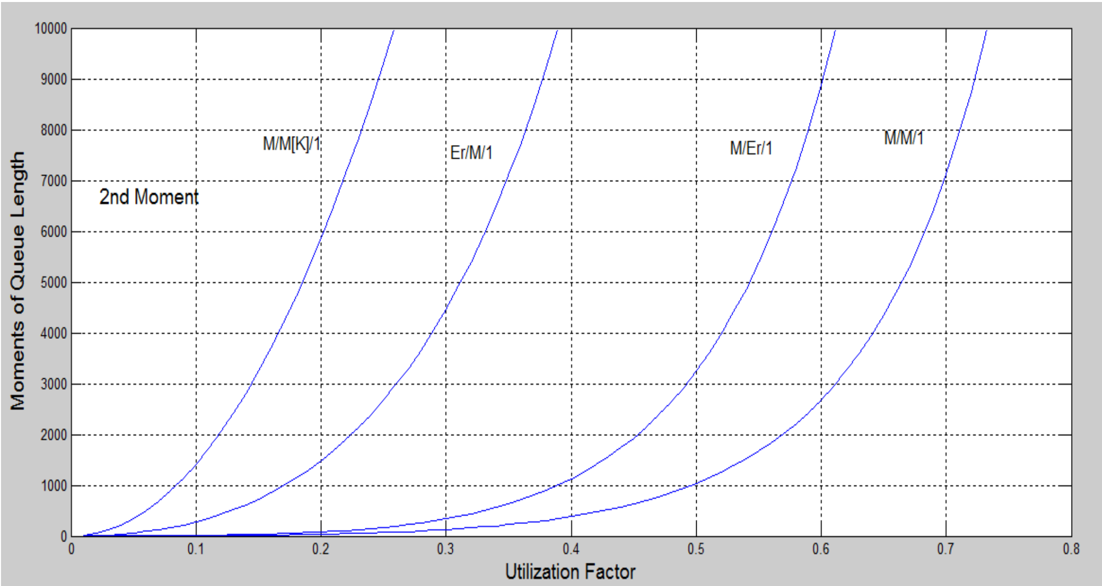

Figure 3.12: Comparing 2nd Moment of Queue Length for Different Queues

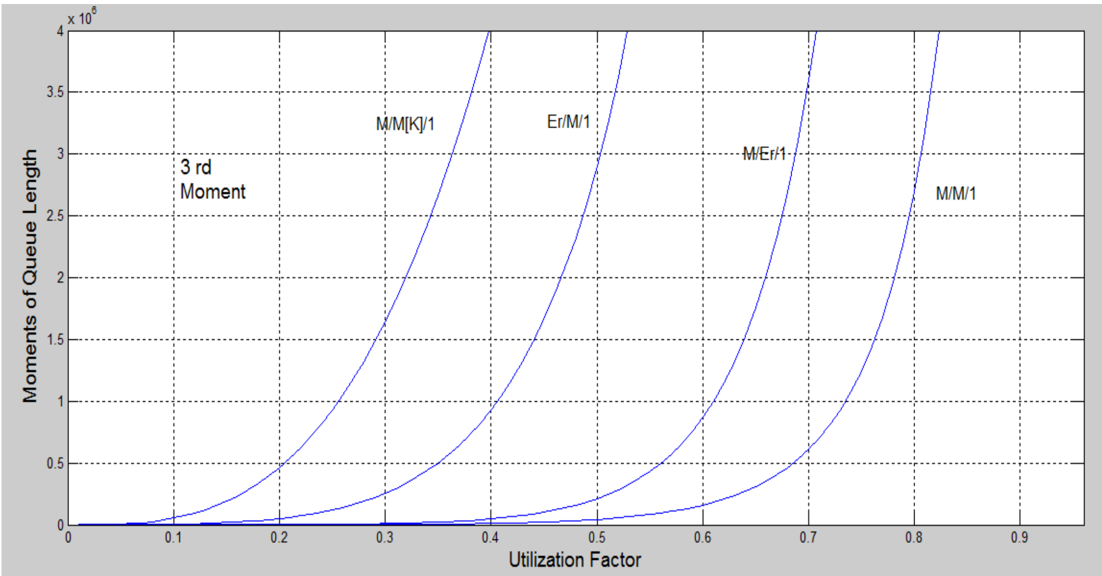

Figure 3.13: Comparing 3rd Moment of Queue Length for Different Queues

With the lesser value of utilization factor, the moments of queue length are increasing in the bulk queuing model as compared to Erlang and markovian queuing model with single server and enhanced number of servers as shown in Fig. 3.13 and Table 3.1.

Table 3.1 Comparison Table for Moments of Queue Length for Different Queues

Utilization Factor			*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*
M/M/1	*k=1*		0.611	2.625	6.643	13.67	25.5	45.75	82.83	162	409.5
	k=2		6.358	33.74	125.7	381.6	1036	2691	7142	2.135e4	9.294 e4
M/M/c	*k=1*	*c=3*	0.614	2.655	6.759	13.99	26.25	47.37	86.26	169.7	431.4
		c=8	0.611	2.269	6.762	15.06	35.46	99.26	327.7	1231	5859
	k=2	*c=3*	6.374	34.01	127.4	388.7	1060	2769	7385	2.2 e4	9.722 e4
		c=8	6.358	33.77	127.1	403	12.45	4264	1.764 e4	9.34 e4	7.46 e5
$M/E_r/1$	*k=1*	*r=2*	0.916	3.937	9.964	20.5	38.25	68.63	124.3	243	614.3
		r=3	1.222	5.25	13.29	27.33	51	91.5	165.7	324	819
	k=2	*r=2*	8.708	54.77	221.3	706.7	1994	5348	1.46 e4	4.477 e4	1.99 e5
		r=3	11.67	81.07	342.6	1127	3253	8891	2.266 e4	7.676 e4	3.473 e5
$E_r/M/1$	*k=1*	*r=2*	12.03	28.94	52.82	87.17	138.3	218.6	357.6	643	1514
		r=3	17.89	42.75	77.57	127.3	201	316.5	515.7	927	2169
	k=2	*r=2*	128.2	686.5	2093	5154	1.17 e4	2.609 e4	6.023 e4	1.585 e5	6.09 e5
		r=3	282.4	1490	4476	1.093	2.44 e4	5.35 e4	1.219 e5	3.164 e5	1.199 e6
$M/M^{[K]}/1$	*k=1*	*K=8*	35.74	72.88	111.4	151.3	192.6	235.4	279.6	325.3	372.5
		K=16	75.75	152.9	231.4	3.11.3	392.5	475	558.8	643.9	730.4
	k=2	*K=8*	1420	5877	1.362	2.49 e4	3.999 e4	5.918 e4	8.279 e4	1.111 e5	1.446 e5
		K=16	5934	2.434	5.57	1 e5	1.592 e5	2.324 e5	3.204 e5	4.24 e5	5.436 e5

3.5 Summary

The performance of a queuing system can be analyzed by moment generating function and hence calculating the queue lengths for various queuing models. The Markov and Erlang queuing models have been studied and discussed in this chapter. Various moment of queue length distribution are generated and analyzed with respect to utilization factor. The bulk queuing model is derived in comparison with single queuing model and utilization factor is compared with queue length moments for various queuing models. This analysis can further be expanded to different communication models for calculating waiting time, processing time, processing length, queue length, throughput, capacity and many more parameters. The birth-death processing models can also be designed for different queuing models. The greater will be the utilization factor, the better will be the system performance and the system QoS can be described as optimal.

Chapter 4

QoS BASED HYBRID SCHEDULING ALGORITHM IN WiMax NETWORK

4.1 Introduction

The WiMax network is based upon all IP platforms and able to support both fixed and mobile applications. The core networks are easily integrated with WiMax without any financial constraints. The Network Working Group (NWG) has defined the Network Architecture of WiMax as shown in Fig. 4.1. In wireless communication systems, transmitting multiple copies of a data flow over more than one number of antennas and to receive numerous forms of data is the main requirement to advance the accuracy of data transfer. WiMax is the technology working with OFDM and utilizes AMC over different types of environment and able to monitor the radio channel quality. The chapter presents the proposed hybrid scheduling algorithm for WiMax to enhance the quality of service parameters.

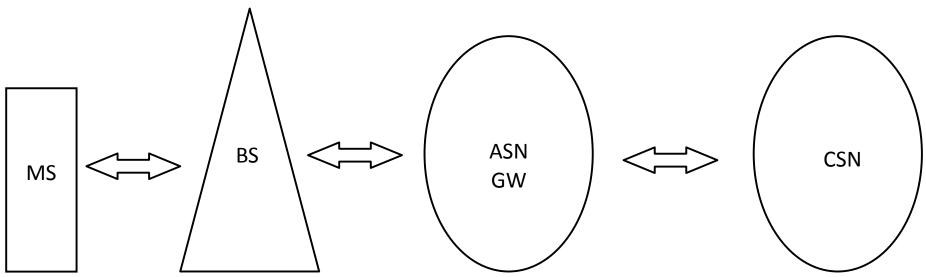

Figure 4.1: Network Architecture of WiMax

4.2 QoS in WiMax

As the internet services are evolving over time, the need to provide high QoS for internet traffic and its associated rich patterns is becoming the fundamental requirement. To manage network resources and satisfactory service delivery, the traditional QoS framework is going to be inefficient to support these mobile Internet applications required user experience. QoS is provided via scheduling the process of

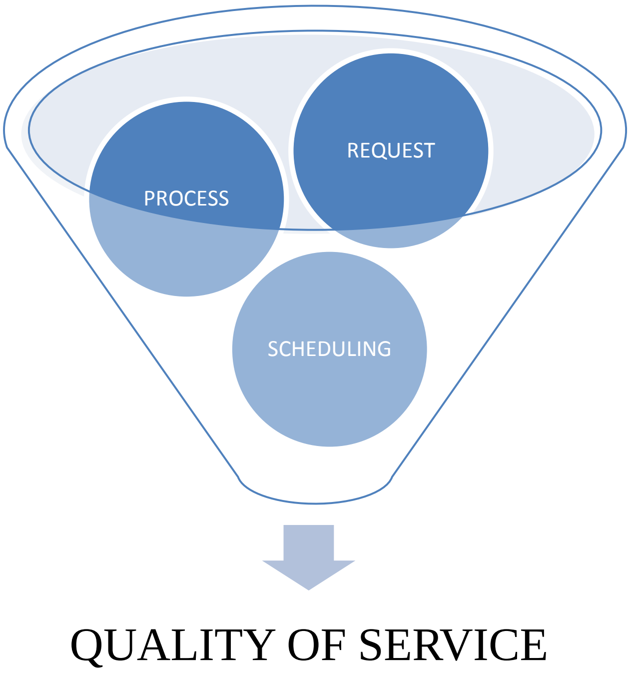

Figure 4.2: QoS in WiMax

service flows generated by request from users in the Mobile WiMax MAC layer as unidirectional flow of packets as shown in Fig. 4.2. Traffic mapping is done in the classifier depending upon the service flows based upon QoS attributes. The MAC interface then associates the packets for transmission over the connection established. The MAC also helps in managing the dynamic traffic over the interface to accommodate the dynamic service demands for uplink as well as downlink scheduling [13]. The service flows may be created, changed or deleted as per user prospective to fulfill QoS requirements.

The MAC layer interface along with scheduler is responsible to provide the required bandwidth to service flows. A number of methods have been proposed for UL and DL bandwidth request/grant mechanism. An UL bandwidth request method is represented to reduce the bandwidth signaling and channel access delays. An improved novel adaptive polling service is described for real time applications [55] to reduce signaling overhead by 50% and traffic latency. An integrated QoS control architecture is represented which implements a cross layer traffic based mechanism for DL and UL separately to perform traffic classification & mapping functions [160]. Also an integrated QoS mechanism is described with CAC signaling in NGNs [52,143]. To support emerging mobile internet applications, the next generation AAI has been introduced as IEEE 802.16m to support priority controlled access and to improve efficiency [5]. In this work, efficient traffic scheduling scheme is proposed for different type of service flows with required QoS attributes.

4.2.1 System Model

The WiMax model consists of a network with multiple subscriber stations. Multiple connections may be served by a single subscriber station. The number of ongoing connections through a subscriber station may be limited by admission control policy. All the types of traffic coming from different queues are aggregated into a single queue to be represented as packets entering the queue or may be considered as load on the queue. The size of queue is the number of packets entered the queue in which some packets are dropped if the queue is full. The OFDM transmitter accesses all the packets in queue depending upon the type of scheduling scheme imposed upon it. The transmitter transmits the queue of packets to base station and then the packets are channelized. The base stations can allocate different channels to subscriber stations. The scheduling schemes used in the work are FCFS, RR, Prioriry and hybrid schemes. The SS with high priority will get more number of sub channels.

The WiMax scheduling algorithms are frame based. The BS collects all the service requests in the beginning of each frame represented by Fig. 4.3. These requests are passed to the scheduler as inputs and the scheduling results are broadcasted to all SS in the current frame or in the successive frames as shown in Fig. 4.4. The scheduling process is repeated for all type of services. Different SS may choose different modulation and coding schemes which are employed in WiMax network. Hence the bandwidth efficiency of all SS varies. The bandwidth request is expressed in terms of bytes and is converted in the form of time slots according to AMC scheme used by the SS. Some services require more bandwidth or in other services delays may be increased. Hence, QoS mechanism is needed to balance the bandwidth and delay constraints.

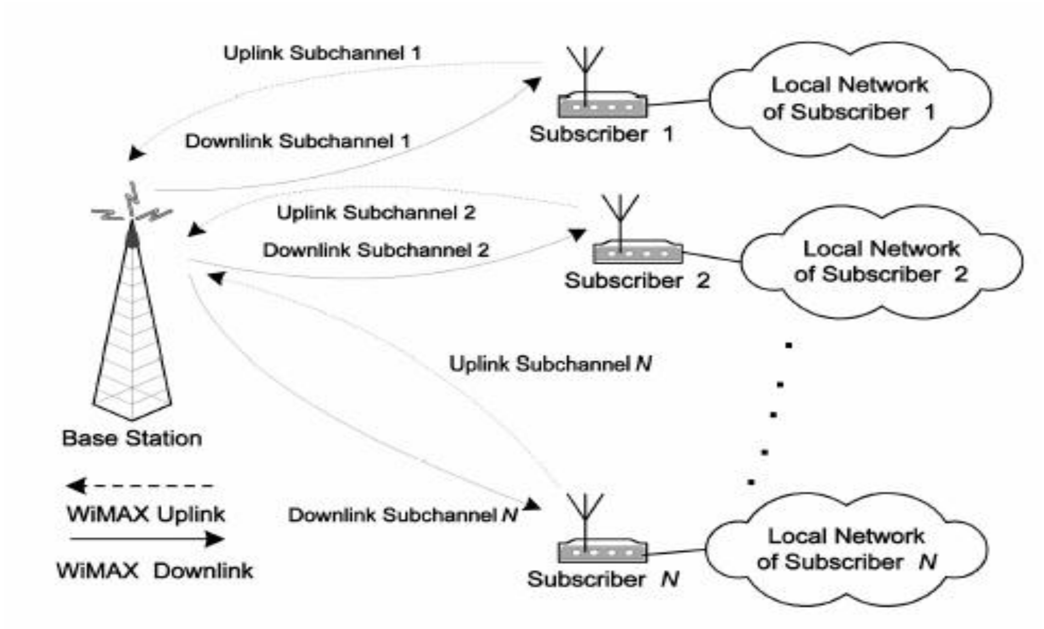

Figure 4.3: WiMax Point to Multipoint Communication [83]

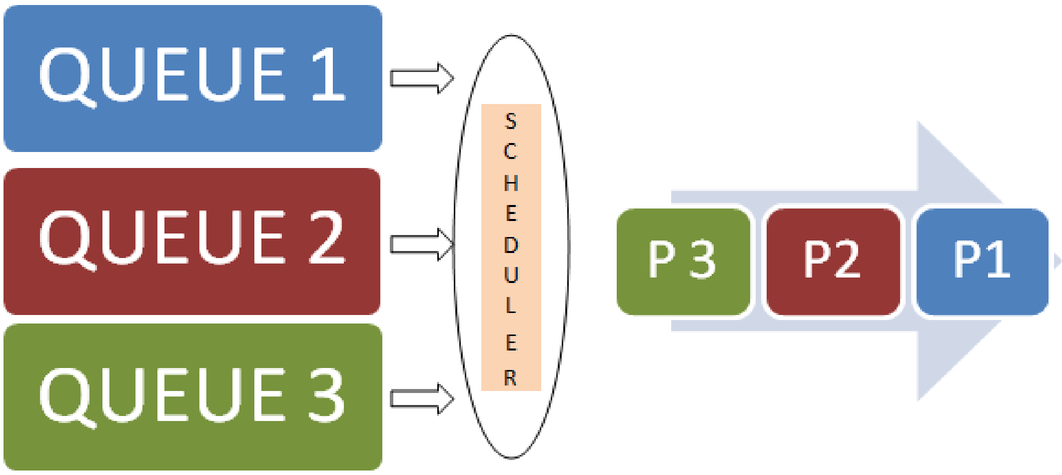

Figure 4.4: Scheduling Queues Generated by SS

In this work, a scheduling model for WiMax is designed in which each class is designated as M/M/1/Li. Li shows the number of requests to be served over Li time-slots in the given frame duration. The blocking probability of M/M/1/Li is represented as follows:

$$q\,(Li) = (1- qi)\, qi^{Li} / (\,1- qi)\, qi^{Li\ +1} \qquad (4.1)$$

Where qi= $\lambda i/\mu i$

The blocking probability can be minimized by allocating the timeslots to all the service requests and may be optimally calculated as:

$$\text{Min } L_i \sum \beta_i \, q(L_i) \quad (4.2)$$

Subject to

$L \leq \sum L_i$

$L_i^{min} \leq L_i \leq L_i^{max}$

$L_i^{min} \leq L_i^{max}$

$L_2^{min} \geq 0$

$L_3^{min} \geq 0$

$L_4^{min} \geq 0$

Where 0 is the number of admitted service conditions in WiMax. The strict convex function shows the probabilities of all classes.

The minimum β^{min} and β^{max} bandwidth required for each connection is given as:

$L^{min} = \beta^{min} / R_m$

$L^{max} = \beta^{max} / R_m$

where R_m is the slot required for each connection or task.

4.2.2 OFDM in WiMax

OFDM transmitter includes a baseband modulator, subcarrier mapper, IFFT, CP, parallel to serial converter and digital to analog converter which is further followed by an RF I-Q modulator. In OFDM, a block of data symbols transmitted simulataneously over single OFDM symbol. The time used to transmit total sub-carriers is defined as an OFDM symbol. The type of modulation such as BPSK, QPSK, 16 or 64-QAM used repeatedly relies upon the signal-to-noise ratio of the received signal and the receiver capability to decipher them accurately. The symbols are modulated and further mapped to subcarriers. For transforming the adjusted subcarriers in frequency field to time domain regions an IFFT is employed. Keeping the small control information overhead, the same modulation techniques are used in all subcarriers. The channel is frequency selective across its huge system bandwidth in a broadband scheme; it means the signal fading on every subcarrier is absolute. There can also be different amount of interference on each subcarrier and alter differently with time. Due to this, signal to impairment level on every subcarrier is different. By choosing an appropriate modulation format on these subcarriers the overall system throughput can be maximized.

MIMO in combination with OFDM is a technique used in frequency selective fading environments for achieving system capacity and high data rates. The quality of wireless communication system is basically dependent on three factors i.e. transmission range, transmission reliability and rate of transmission. By using MIMO with OFDM system, these factors can be improved. The main advantages of MIMO-OFDM are less interference, diversity gain, increase capacity of data, bandwidth gain also increases and power efficiency. MIMO multiplexing enhances the capacity of a system by splitting a high rate streams into multiple lower data rates. It also improves the link reliability by antenna diversity gain. There are various parameters on which performance of OFDM depends. Few are discussed below along with enhancement techniques.

- BER: It is defined as the rate of occurrence of error in a transmission system. It is represented as the number of bits corrupted out of total bits transmitted. BER =number of bit errors / total number of bits transmitted

  The effect of channel noise, interference and bit synchronization are some of the problems responsible for BER.

- Symbol error rate: The symbol error rate (SER) is the rate of number of incorrectly received symbols with respect to total symbols received. The expectation value of the SER depends upon its error probability (S_e), with a symbol length of N bits can be expressed as

$$SER = 1 - (1 - S_e)\ N$$

- Data rate: The number of bits that are processed by the system per unit of time is called data rate. It is an important factor, as more number of bits conveyed in particular time, more is the speed.
- Signal-to-noise ratio: Signal-to-noise ratio is defined as the ratio of average signal power to average noise power.

$$SNR = \frac{P_{Signal}}{P_{Noise}}$$

4.3 Scheduling Algorithms in WiMax

4.3.1 Non- preemptive Scheduling Algorithms

The algorithms that may not accept interruptions are non-preemptive scheduling algorithms. The types of non-preemptive scheduling are discussed along with calculating the scheduling parameters.

4.3.1.1 FCFS (First come First Serve) Scheduling

In this type of scheduling algorithms FIFO process is going on. All the jobs entering into the ready queue are executed on the first come first serve basis. As shown in Fig. 4.5, there is more number of jobs present in the ready queue, the algorithm starts executing the process from the ready queue, no matter which process with high priority and which process with less priority. As the jobs can be served on the first come first serve basis the starvation problem can also be reduced in FCFS algorithms, but the turnaround time and response time is increased.

For any scheduling algorithm it is necessary to analyze the behavior and outcome of the algorithm. There are certain parameters like waiting time and turnaround time from which the algorithm results can be analyzed using Gantt chart. A Gantt chart is a chart which shows the starting and end time for a process. The burst time for each process is considered and placed into the Gantt chart. This chart helps to find the turnaround time and waiting for all the processes.

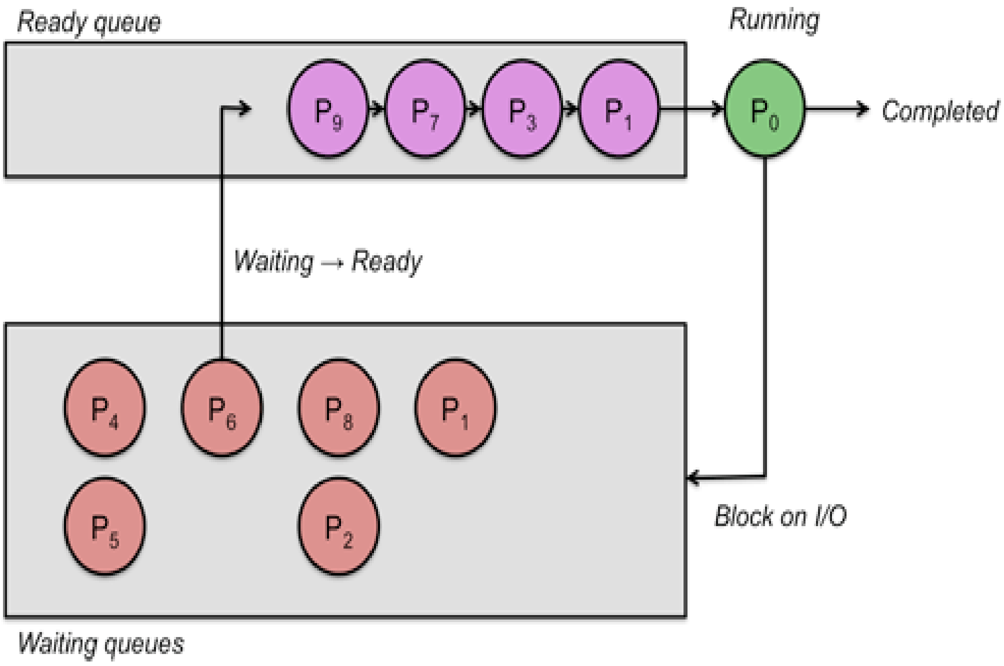

Figure 4.5: FCFS Process

Table 4.1: Process Table for FCFS

PROCESS	PRIORITY	BURST TIME(ms)
P0	5	20
P1	7	27
P2	4	34
P3	2	52
P4	6	17
P5	1	12
P6	3	10

As shown in the Table 4.1, jobs are entered from P0 to P6. Fig. 4.6 shows the Gantt chart for the FCFS Algorithm.

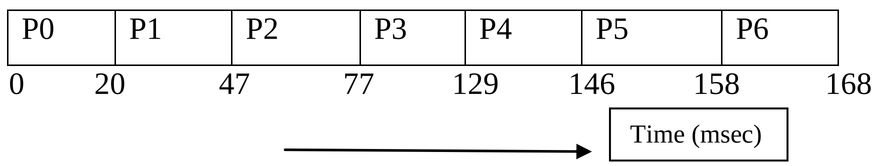

Figure 4.6: Gantt Chart for FCFS Algorithm

The average turnaround time and average waiting time for single process can be calculated and the behavior of the FCFS Algorithm is examined.

Average waiting time = 0 + 20 + 47 + 77 + 129 + 146 + 158 / 7 = 577/7 = 82.42ms

Now, 82.42ms will be the waiting time for single process.

Turnaround time = burst time + waiting time

The total of turnaround time for all process =577+168 = 745ms

So the Average turnaround time =745/7 =106.42ms

As no method of prioritization is followed, hence there is no problem of starvation; the jobs takes response on first come first serve basis and the response time is high. As shown in the above calculations, the average waiting time is more in case of FCFS algorithm, shows poor performance.

4.3.1.2 Strict Priority Scheduling or Priority Scheduling

As there are more number of jobs present in the ready queue, by taking the average waiting time kept in mind there is need of such algorithms which can execute the higher priority jobs first. The high priority jobs are more important to execute. The objective of this algorithm is to execute the higher priority jobs first and after that lower priority jobs are executed. The advantage of using this technique is that if higher priority jobs are present in the queue then these jobs will not wait for long time as in the case of FCFS.

The flow diagram in Fig. 4.7 presents the priority scheduling which helps to execute important jobs which are having higher priority. There is also a disadvantage of priority scheduling as the higher priority jobs are executed first if any important job is present in the lower priority queue then this job will wait for long time. This is called the problem of starvation. The lower priority jobs are waiting for long time as the higher priority jobs will be executed first. The strict priority scheduling is helpful for the higher priority jobs but not for the lower priority jobs because waiting time for lower priority jobs is more. The analysis of SP is represented with the help of process table shown in Table 4.2.

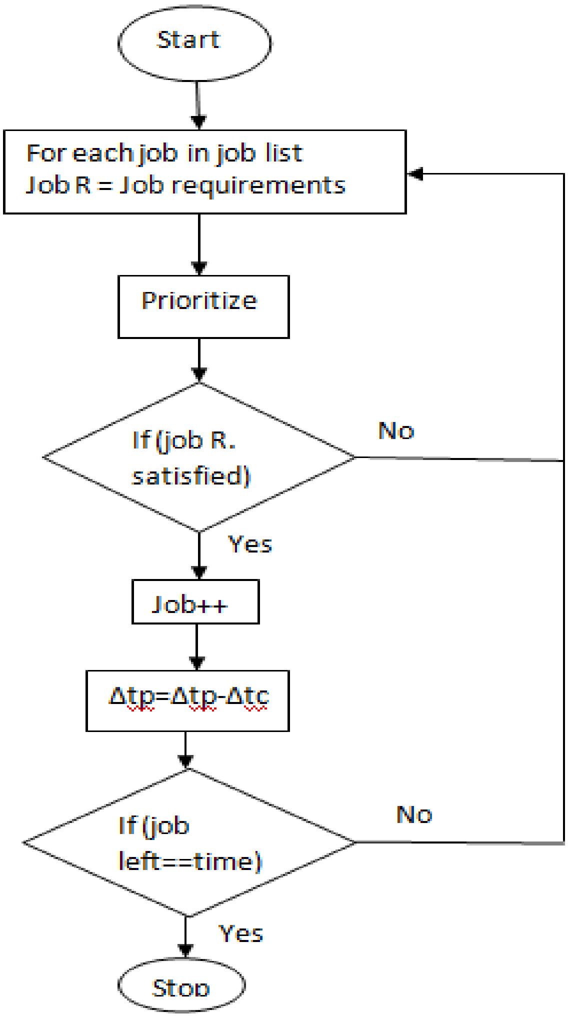

Figure 4.7: Flow Diagram of Priority Scheduling Algorithm

Table 4.2 Process Table for SP

PROCESS	**PRIORITY**	**BURST TIME(ms)**
P0	5	20
P1	7	27
P2	4	34
P3	2	52
P4	6	17
P5	1	12
P6	3	10

The Gantt chart is designed by taking SP algorithm accordingly as the higher priority jobs will be placed first than lower priority jobs as shown in Fig 4.8.

P5	P3	P6	P2	P0	P4	P1	
0	12	64	74	108	128	145	172

→ Time (msec)

Figure 4.8: Gantt Chart for SP

From above Gantt chart, the average turnaround time and average waiting for single process is calculated and the behavior of the SP algorithm is examined.

Average waiting time = 0 + 12 + 64 + 74 + 108 + 128 + 145 / 7 = 531/7 = 75.85ms

Now, 78.85ms will be the waiting time for single process.

Turnaround time = burst time + waiting time

The total of turnaround time for all process = 703ms

So the Average turnaround time = 531+172 = 703/7 = 100.42ms

For higher priority jobs, the waiting time is reduced and the important jobs will finish first. The waiting time for lower priority jobs is increased.

4.3.2 Preemptive Scheduling Algorithms

The algorithms accepting interruptions are preemptive scheduling algorithms. The type of preemptive scheduling algorithm along with the scheduling parameters is discussed as follows:

4.3.2.1 Round Robin (RR) Scheduling

In round robin scheduling, the process is going by rounds. For execution of process the time slot is given to every process. The scheduler picks the process for a time slot from each queue. This scheduling is called preemptive because the running process is interrupted by another process. The Fig. 4.9 shows the process of RR scheduling. It is not a priority scheduling as the scheduler cannot give priority to any process. The main advantage of this scheduling is that it is starvation free scheduling that is every process is picked by the scheduler on time slot basis. The drawback of this scheduling is that execution of small process can wait for a long time because time is divided for each process in the queue.

As there are more number of jobs present in the ready queue. By taking starvation problem kept in mind there is need of such algorithms which can give time to each process, as all the jobs in a queue are important. The objective of this algorithm is to assign time slot to every process so that no process in the queue waits for a long time. The analysis of RR is represented with the help of process table shown in Table 4.3.

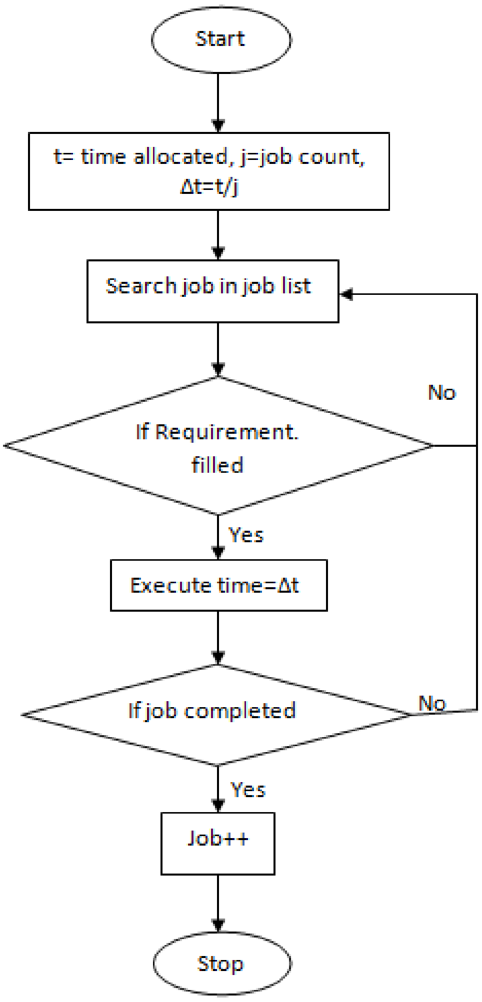

Figure 4.9: Flow Diagram of RR Scheduling Algorithm

Average waiting time = 828/7 = 118.28ms

Turnaround time = burst time + waiting time

The total of turnaround time for all process = 906ms

So the Average turnaround time = 906/7 =129.42ms

The main advantages of RR is that time is given to all the processes, starvation free and easy to implement. Even a short process may take long time to execute is the main shortcoming.

Table 4.3: Process Table for RR

PROCESS	**PRIORITY**	**BURST TIME(ms)**
P0	5	20
P1	7	27
P2	4	34
P3	2	52
P4	6	17
P5	1	12
P6	3	10

The Fig. 4.10 shows the Gantt chart for the RR Algorithm. The time slots have been allocated to every process. All the new processes getting entered into the system will be queues for the next phase till the older one requests are not being compiled.

P0	P1	P2	P3	P4	P5	P6	
0	6	12	18	24	30	36	42

P0	P1	P2	P3	P5	P6	P1	
48	54	60	66	72	78	84	90

P1	P2	P3	P4	P5	P1	P2	
96	102	108	114	120	126	132	138

P1	P2	P3	P4	P5	
144	150	156	162	168	174

→ Time (msec)

Figure 4.10: Gantt Chart for RR Algorithm

4.4 WiMax Network Design for Requests Generation and Scheduling

The WiMax network design flow is represented in Fig. 4.11. The WiMax network is designed by initializing number of nodes in a network. Then the tasks are generated by the various nodes by initializing the source and destination nodes. The source node is from where the data is to be sent and destination node is where the data is to be received. The available nodes in a network will decide the number of jobs generated

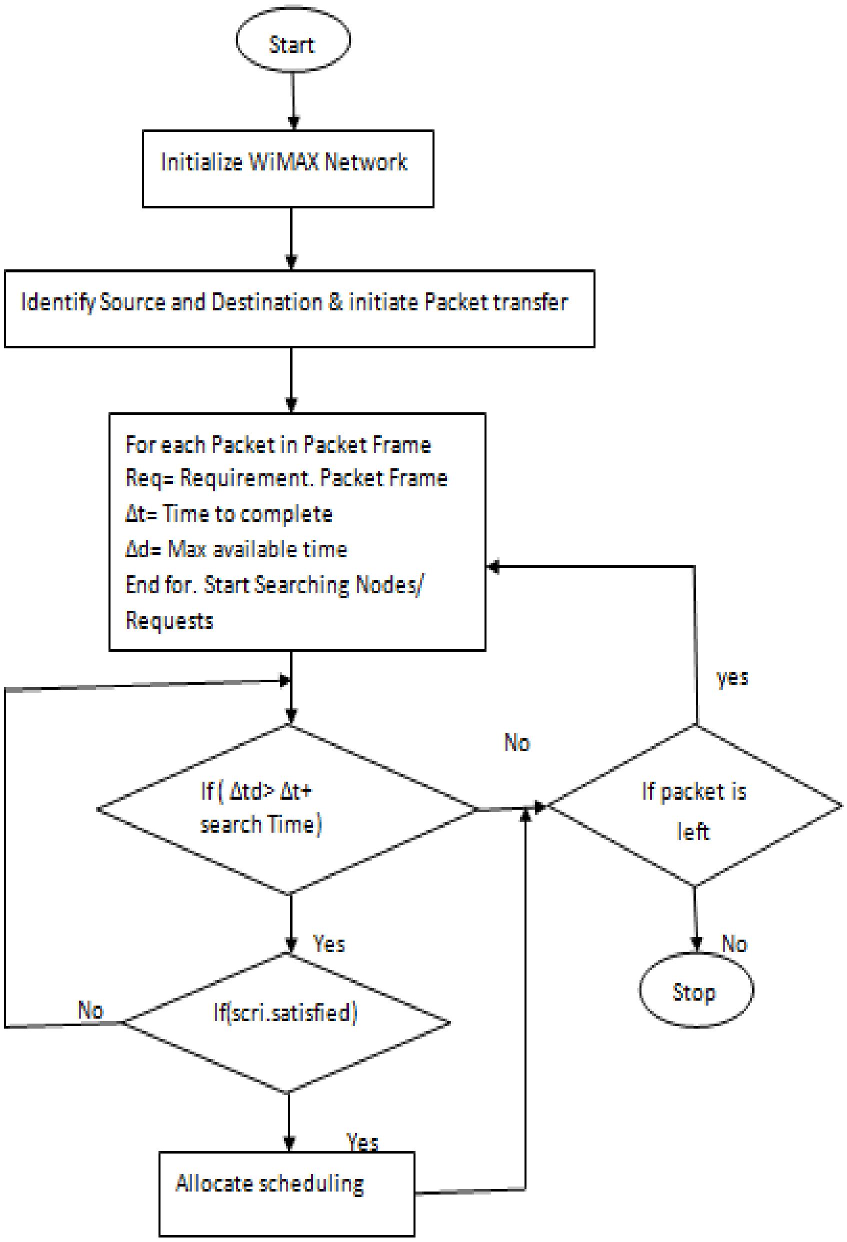

Figure 4.11: Flow Diagram of WiMax Network Design

and completed. The number of jobs generated is one of the major parameter considered for efficient scheduling algorithm. As there is no fixed scenario for network design considerations, the source and the destination nodes are taken at random. For each packet in the frame, the requirements are to be specified as time required for completing the jobs and total time available. The packet identification and energy required to transfer the packets from source to destination is calculated. Total energy consumed in a network will be the sum of the energy consumed by all nodes as well as packet life time. If all the requirements are satisfied, the scheduling algorithm has been allocated to the jobs.

4.5 Proposed method: An Efficient Scheduling Algorithm for WiMax

The job model for the proposed efficient algorithm consists of the hybrid combination of FCFS, RR and priority scheduling as shown in Fig. 4.12.

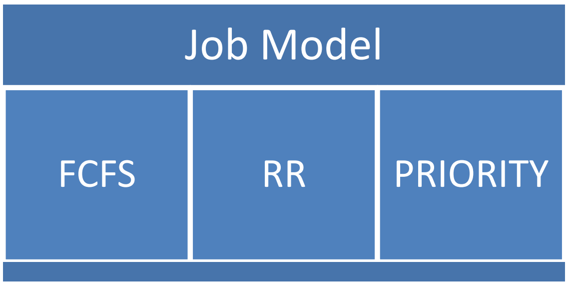

Figure 4.12: Job Model as Combination of FCFS, RR and Priority Scheduling

As described above, WiMax is a high access data point medium which provides high rate of data transfer with accuracy in the transfer, it becomes a necessary part for the WiMax system to handle huge data request at the same time. Algorithms like FCFS, RR and Shortest Path First is already in existence in the WIMAX technology. The purpose of this research work also involves the working methodology of FCFS algorithm and RR algorithm. When a huge amount of data packets fall into the category at one node, it becomes quite sophisticated for a node to judge which request has to be handles first and which has to be handled at later stage. Hence, the priorities are required to be involved. The flow diagram for Hybrid algorithm for the implementation is shown in Fig. 4.13

The prioritized jobs are firstly scheduled in the FCFS algorithm and then RR is applied to rest of the jobs as RR algorithm is an alternate solution to the FCFS and time factor is considered as one of the important input component in RR component. Similar execution time is assigned to each job in the queue. The analytical model of RR consists of a traffic model with traffic intensity λ, the data flows of M queues with intensities $\lambda 1$, $\lambda 2$ to λM. The executed data flow is represented by $\lambda 1S$, $\lambda 2S$ to λMS in the node and the discarded part by the node is $\lambda 1L$, $\lambda 2L$ to λML. The maximum

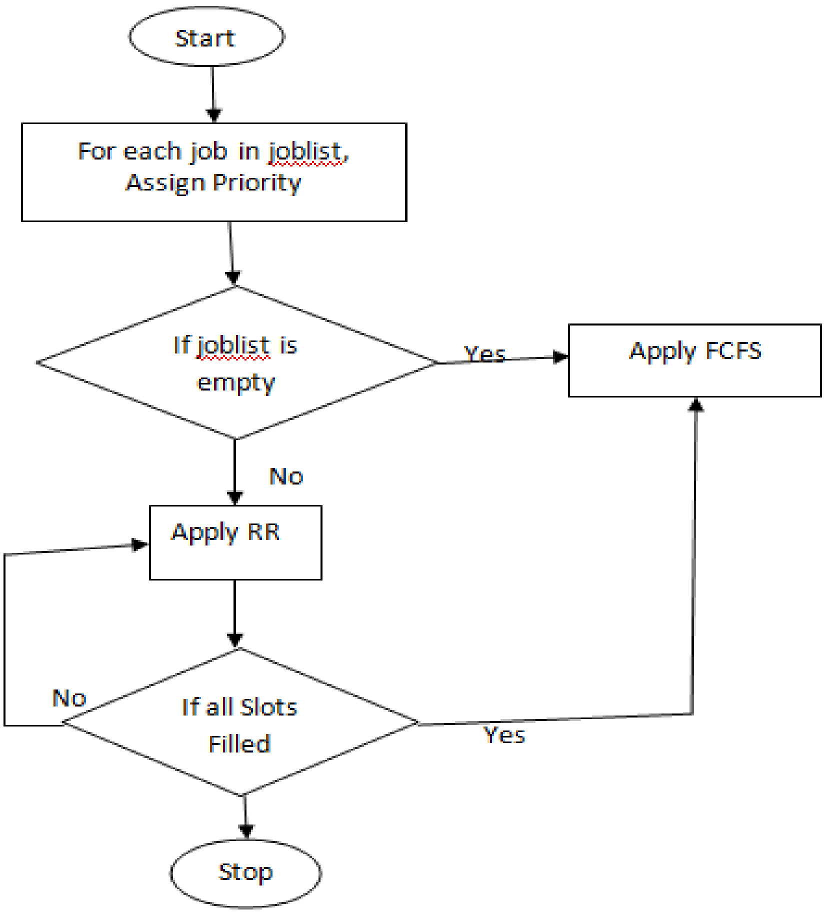

Figure 4.13: Flow Chart for Hybrid Scheduling in WiMax

number of packets stored in the i^{th} queue is Nimax. The intensity of service utilization of i^{th} queue is μi and average time of execution is Ti. Services of different queues in

RR are processed during certain time intervals Δi. The probability of k^{th} queue in service is represented by the relation:

$\Delta(i=k)=\Delta_k/1+\Delta_2+\cdots+\Delta_M$

(4.3)

$\Delta=BR_k/BR$

When the n^{th} queue is served and k^{th} queue has entered the network, the average waiting time of packets of k^{th} queue is $W^n k$, then the average waiting time of packet in the k^{th} queue is generally given as:

$Wk=W^1k\cdot(i=1)+\cdots+W^Mk\cdot(i=M)$ (4.4)

In the case of three queues, the packets in the first queue have average waiting time given by the relation:

$W^1 1\cdot BR_1/BR+W^2 1\cdot BR_2/BR+W^3 1\cdot BR_3/BR$ (4.5)

To reduce the waiting time for longer queues, the priorities have been involved in the traffic model. The queues will be generated with priority and given to the hybrid scheduler. In the hybrid scheduling algorithm, the requests entered the network with delay are required to wait till the requests already in queue are not fulfilled. To reduce the waiting time for such requests, the priority is assigned on length basis. The requests with the shortest queue length will be served earlier as compared to queues with longer duration. The average waiting time in FCFS-RR scheduling is 20-30msec and in hybrid scheduling it has been reduce to 3-4msec.

4.5.1 Algorithm for Hybrid Scheduling in WiMax Network

The algorithm for network initialization is configured in the following manner

1) *START*
2) *Initialize data_packets*
3) *Initialize source node*
4) *Initialize destination node*
5) *Initial Energy, E=0;*
6) *T=network life time for the completion of the tasks*

7) *For i=1:length(jobs)*
8) *E(i)=E+Ej; Tc=T+Tj; Ej=energy per job; Tj=time of execution per job;*
9) *While Tc<T*
10) *Repeat procedure*
11) *Stop*

Initialization parameters would be same for round robin algorithm also. The algorithm is defined as follows:

1) *Start*
2) *Initialize data_packets*
3) *Initialize source node*
4) *Initialize destination node*
5) *E=0;*
6) *T=network life time for the completion of the tasks*
7) *For i=1:length(jobs)*
8) *tpc=Total_jobs/T;*
9) *For k=1:tpc:T*
10) *If k==tpc*
11) *Switch job.*
12) *E(i)=E+Ej; Tc=T+Et;*
13) *While Tc<T*
14) *Repeat procedure*
15) *Stop*

4.5.2 Parameters for Simulation

The simulation has been carried out on MATLAB environment. The following parameters have been configured.

1) The size of the OFDM symbols to be configured, N: 512,1024,2048,4096
2) Number of OFDM symbols to be simulated, m: 100, 500,1000,2000,3000
3) Size of Alphabet M : 4, 8, 16 generally
4) Up-sampling factor L: 1 generally
5) Type of mapping : PSK or QAM

6) Constellation phase offset : 1 generally

7) Symbol order : Binary or gray

8) Size of cyclic prefix samples, N_{cp} : 1/4, 1/8, 1/16, 1/32

4.5.3 Results and Discussions

4.5.3.1 Performance Evaluation of proposed algorithm based on QoS parameters

The hybrid scheduling algorithm is applied on WiMax networks consisting of N packets. The size of the packet, m may change depending upon the type of service involved and the requests generated by the user. Fig. 4.14 represents the requests

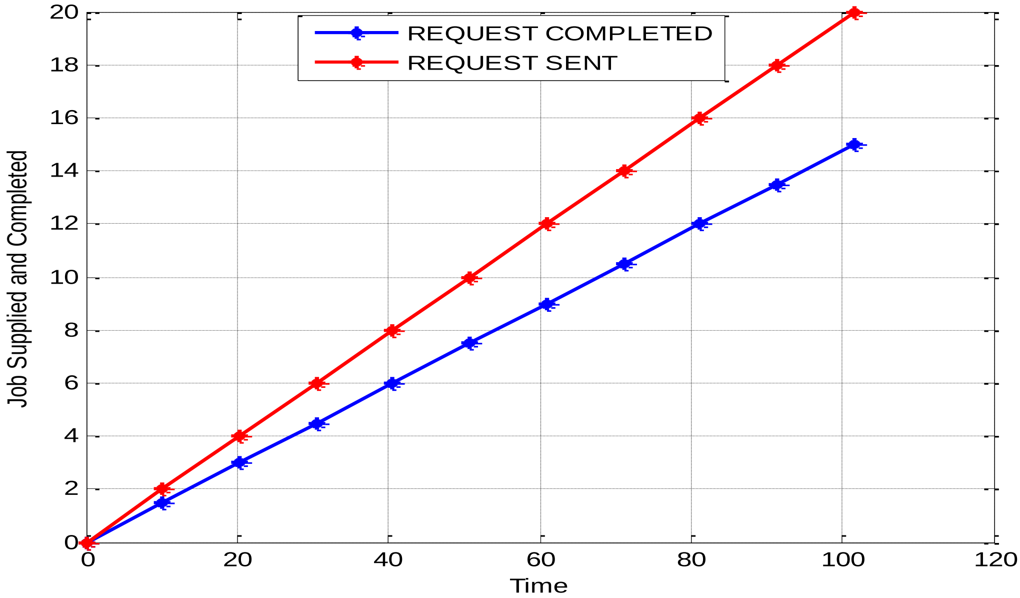

Figure 4.14: Number of Requests Sent and Completed in Hybrid Scheduling

generated by the user as requests sent and the requests successfully completed. Almost 75% of requests are able to reach the destination as 15 requests out of 20 are successfully completed. The Table 4.4 represents the jobs provided to the different algorithm and their completion accuracy. Vital growth in the completed job and its accuracy is represented when we are using hybrid algorithm in comparison to the FCFS and RR algorithms.

Table 4.4: Performance Accuracy of FCFS, RR and Hybrid Algorithm

ALGORITHM	JOBS PROVIDED	JOBS COMPLETED	PERCENTAGE ACCURACY
FCFS	100	36	36%
RR	100	51	51%
HYBRID	100	75	75%

4.5.3.2 Energy Consumed in Hybrid Scheduling Algorithm

The energy consumed in WiMax network environment depends upon the number of nodes available in the network and the number of requests generated by different nodes. The energy consumed during transmission increases with the time required for job completion as shown in Fig. 4.15. The amount of energy consumed for 20 requests is 68mJ for a job completion time of 112sec.

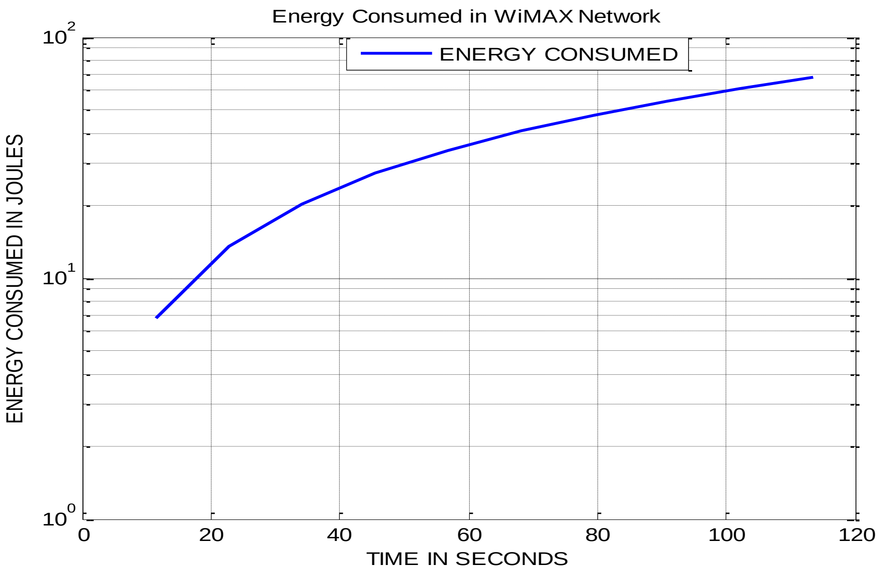

Figure 4.15: Energy Consumed During Transmission in Hybrid Algorithm

Fig. 4.16 represents the amount of energy consumed for different number of requests. As the number of requests is increasing, the energy required to fulfill the job completion also increases. As the network will work for longer duration, the energy consumption is increasing to entertain large requests. The energy efficiency is also increased for larger number of requests. The change in number of requests varies the amount of energy consumed. It is decreasing as the packets are going to be increased.

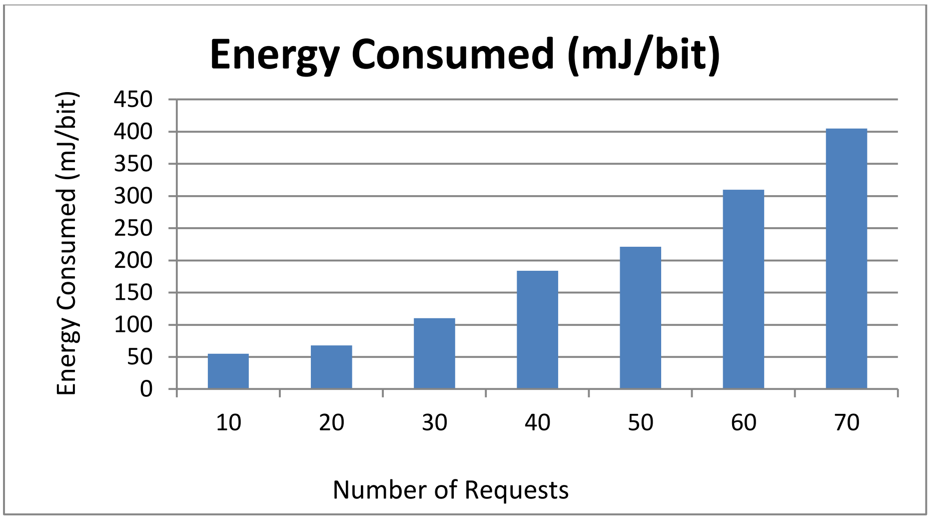

Figure 4.16: Energy Consumed for Different Number of Requests

4.5.3.3 Performance Evaluation Based on BER Comparison

The BER for the three scheduling algorithms is calculated and compared as shown in Fig. 4.17. Hybrid scheduling provides less amount of BER as compared to FCFS-RR and priority scheduling algorithms. As the error rate is reduced, the accuracy during packet transmission increased. The error rate is calculated for large number of iterations to get the desired results by means of repeated actions.

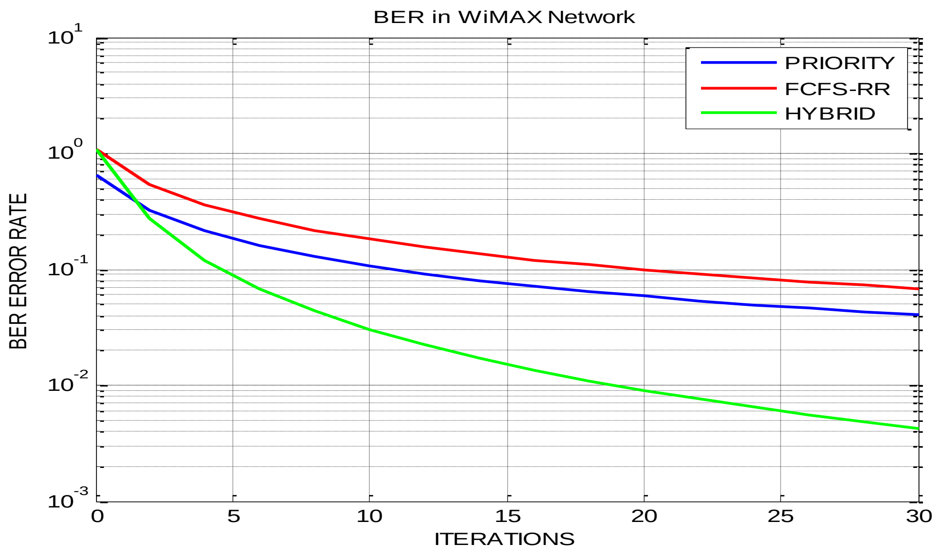

Figure 4.17: BER Comparison of Three Scheduling Algorithms

Fig. 4.18 represents the comparison of the hybrid algorithm in terms of the bit error rate calculation by varying the block size.

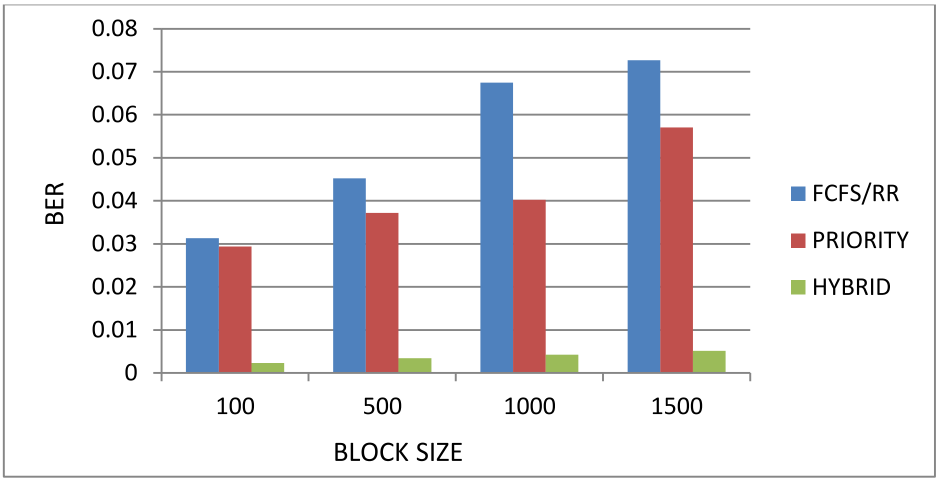

Figure 4.18: BER in WiMax Network for Different Scheduling Algorithms by Varying Block Size

The hybrid scheduling removes all the conflicts occurring during FCFS/RR and priority scheduling, it represents BER results with less errors and higher efficiency. As the block size is increased, the BER is increasing for the three scheduling algorithms. Table 4.5 shows a decrease in the BER when it comes to the hybrid algorithms for different block sizes.

Table 4.5: BER in WiMax Network for Different Scheduling Algorithms

BER in WiMax Network				
Scheduling Type	**BLOCK SIZE**			
	100	**500**	**1000**	**1500**
FCFS/RR	0.03132	0.0452	0.06749	0.07265
PRIORITY	0.0294	0.03718	0.04022	0.05703
HYBRID	0.002321	0.00341	0.004243	0.005161

4.5.3.4 Performance Evaluation Based on Throughput Comparison

The throughput for the three scheduling algorithms is calculated and compared as shown in Fig. 4.19. Hybrid scheduling provides greater amount of throughput as compared to FCFS-RR and priority scheduling. As the throughput is increased, the accuracy during packet transmission increased and hence quality is enhanced.

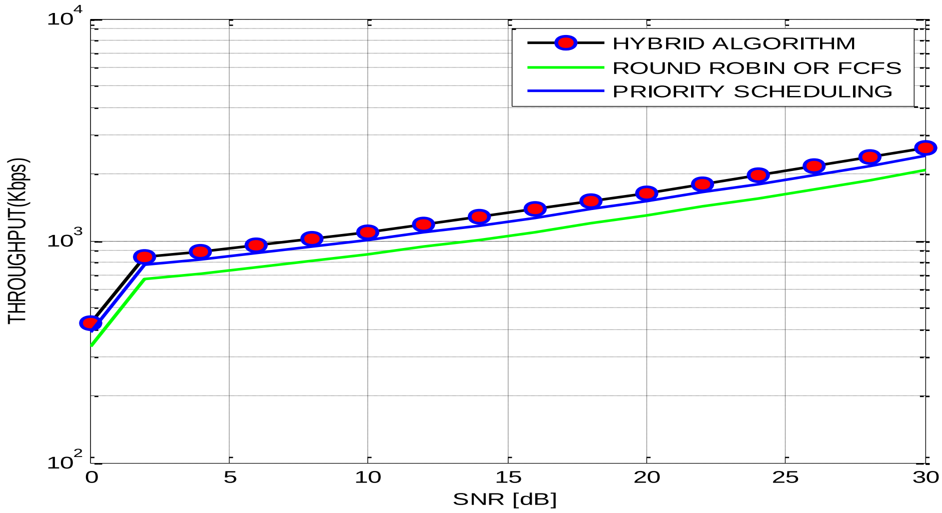

Figure 4.19: Throughput Comparison of Three Scheduling Algorithms

The throughput with respect to SNR is also evaluated for cyclic prefix length of 1. The results are calculated for 100, 500, 1000 and 1500 number of blocks. Fig. 4.20 and Table 4.6 represent the throughput performance comparison for the three scheduling algorithms. The best results for throughput are provided by hybrid algorithm as the throughput remains same by varying the number of packets.

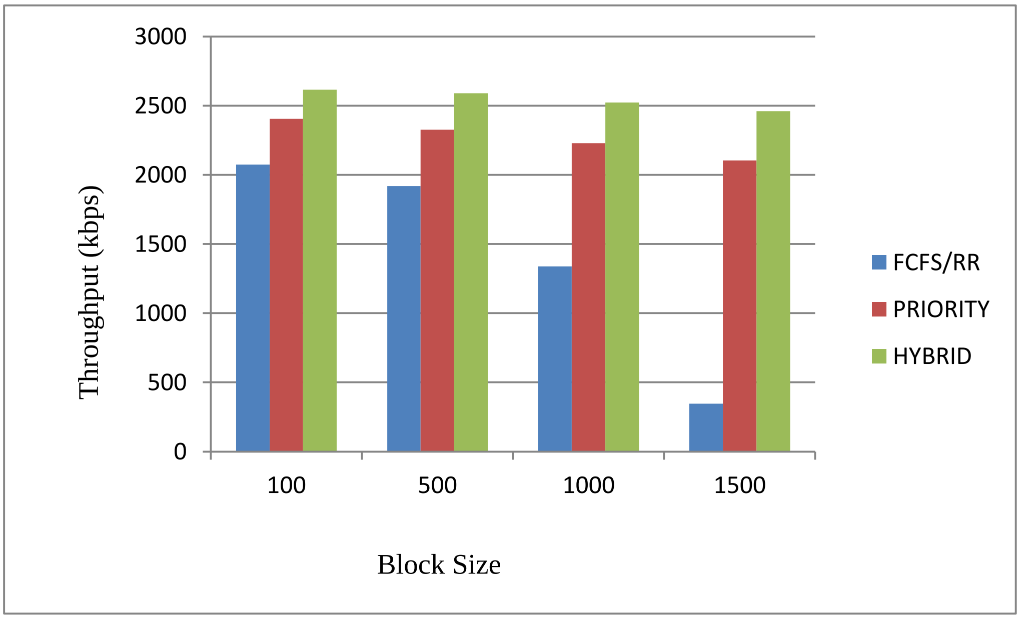

Figure 4.20: Throughput in WiMax Network for Different Scheduling Algorithms by Varying Block Size

Table 4.6: Throughput in WiMax Network for Different Scheduling Algorithms

Throughput(Kbps) in WiMax Network				
Type of Scheduling	**BLOCK SIZE**			
	100	**500**	**1000**	**1500**
FCFS/RR	2073	1920	1339	347
PRIORITY	2405	2327	2228	2105
HYBRID	2616	2589	2523	2460

4.5.3.5 Performance Evaluation based on Symbol Error Rate

The symbol Error rate performance is evaluated with SNR by changing the size of the packets as 512, 1024, 2048, and 4096 when the number of packets considered is 1000 and CP is set to 1. The SER is decreasing as the SNR is going to be increased. The increase in number of packets also decreases the value of SER with SNR. The various performance metrics in WiMax networks scheduling algorithms may vary by changing the evaluation parameters. Fig. 4.21 and Table 4.7 represent the comparison of SER with respect to SNR by varying the packet size.

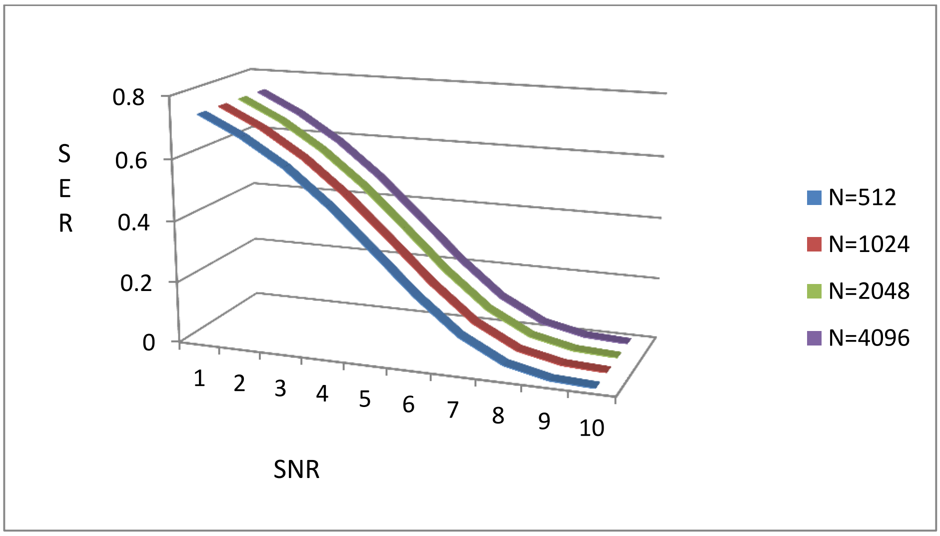

Figure 4.21: SER for Hybrid Scheduling Algorithm by Varying Packet Size

Table 4.7: SER for Hybrid Scheduling Algorithm by Varying Packet Size

SER										
SNR	1	2	3	4	5	6	7	8	9	10
N=512	0.740244	0.674758	0.59051	0.480158	0.353262	0.222061	0.110178	0.037377	0.00693	0.000611
N=1024	0.740932	0.675519	0.589303	0.480593	0.353688	0.222245	0.109213	0.036811	0.007168	0.000604
N=2048	0.740841	0.675308	0.588517	0.480484	0.353174	0.221893	0.10934	0.037151	0.007094	0.000577
N=4096	0.740984	0.674496	0.588575	0.480329	0.353517	0.222342	0.109277	0.037089	0.007213	0.000565

4.6 Summary

An Efficient Scheduling Algorithm for WiMax is presented. From the results obtained, it is concluded that the WiMax system is one of the efficient platform for high speed data transfer but when the load increases over the nodes there must be a suitable mechanism which can handle the number of nodes and its architecture system. This chapter concludes that there are several algorithms which are based on scheduling of packets. We have presented three algorithms namely FCFS-RR, priority and hybrid algorithm which is a combination of FCFS-RR and Priority algorithm. The hybrid algorithm shows an increase in the performance of the jobs completed and error rate reduction as well as the throughput is enhanced comparatively. As the SNR is increased, the SER is reduced irrespective of packet size and number of packets transmitted.

Chapter 5

QoS BASED SCHEDULING IN LTE AND LOAD BALANCING

5.1 Introduction

The third generation of wireless communication is defined by LTE as the evolution of 3GPP represented as standard developing body to provide higher data rates, less errors, low latency and high spectral efficiency. The concept of RAN was developed in 2004 proceeded with a lot of research proposals for the development of 3GPP and presented UTRAN, evolved UMTS and E-UTRA in general referred as LTE. The BS and MS in LTE are represented as eNodeB and UE. The physical resource consists of time frequency grid with multiple resource blocks which is further divided into resource elements. The function of the scheduler is to assign the time-frequency resource elements to different users available in a cell. This chapter discusses the structure of LTE, resource allocation along with frame structure. The research is focused on the FCFS/RR, Priority and Hybrid Scheduling algorithms. The three algorithms are compared for different QoS parameters. The problems prevailed due to fixed resource allocation is improper load balancing and job scheduling. A job scheduling algorithm is proposed for load balancing in LTE networks.

5.2 LTE Network Architecture

The network architecture of LTE consists of EU, E-UTRAN, EPC and server PDNs as shown in Fig. 5.1. The radio access in LTE is generated by E-UTRAN also accompanied by the evolution of non-radio aspects which include EPC network. The flow of IP packet with a defined QoS is between the gateway and EU. The packet is then set up and released by E-UTRAN and EPC depending upon the requirement of application. The EPC communicates with the different PDNs attached with the server in the outside networks such as IP networks and IP systems or private corporate networks. The interfacing among different components is defined as Uu, S1 and SGi.

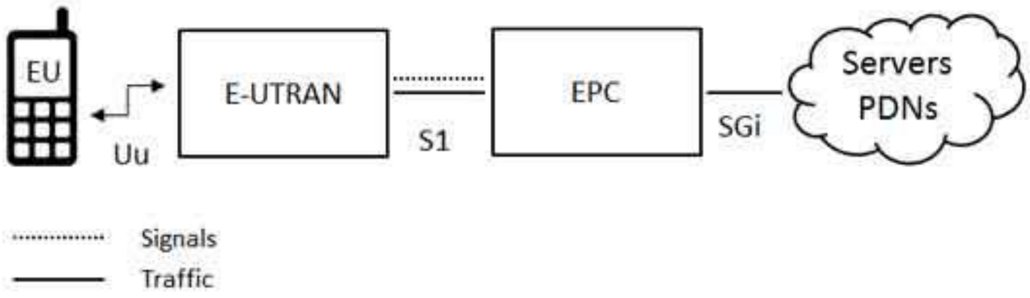

Figure 5.1: Components of LTE Network

The actual architecture design is much wider than the components represented earlier. The various recompenses required to enhance the architecture are the network topology, relay nodes to increase seamless integrity, advancement in low power nodes, improving coverage and capacity and multicarrier introduction up to high ranges.

5.2.1 LTE Control Structure Design

The LTE protocol layered structure consists of a mainly three layers as shown in Fig. 5.2. The layer 1 is the physical layer carries the information from or towards the air

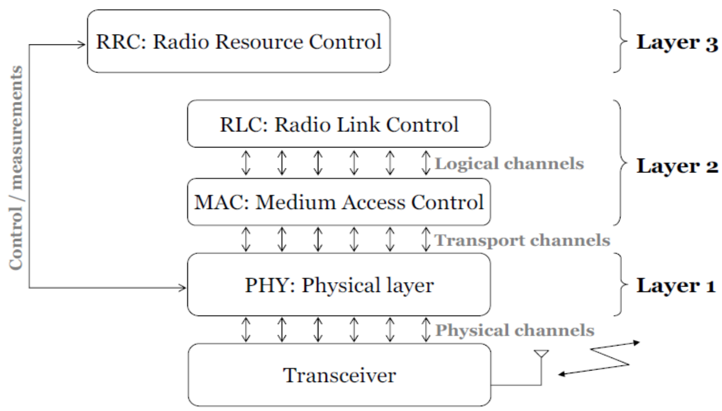

Figure 5.2: Physical and MAC Architecture [120]

interface by taking care of AMC, link adaptation, energy consumption. The layer 2 is defined as MAC layer consisting of logical and transport channel control and management, MAC multiplexing from SDU for different channels, dynamic scheduling along with priority management for various channels, error detection and correction information reporting.

The RLC layer transfers the upper layer PDUs, concatenation and segmentation of PDUs, RLC establishment and protocol error detection. The information broadcasting related to paging, establishment and release of connection is handled by the third layer defined as RRC. Security is one of the concerns to be established, handled and managed by RRC sublayer for point to point radio links. The IP packets are handled by MAC SDU which is further split into PDUs if SDU size is large. The MAC layer attaches the Header and padding to the MAC SDU and submits it to physical layer for transmission. The MAC PDU frame with RLC segmentation is shown in Fig. 5.3. The physical layer allocates resource blocks to MAC PDUs for transmission.

Figure 5.3: MAC PDU Frame

5.2.2 LTE Resource Block Structure

LTE resource block consists of time and frequency units correspond to one OFDM symbol and its subcarrier. LTE time-frequency grid is represented in the form of resource elements. The TDD frame structure consists of 10 sub frames with two slots having duration of 0.5msec each and a total duration of 10msec. The frame and sub frame length is same for uplink and downlink resource allocation. The FDD frame composed of two sub frames of length 5msec separately for uplink and downlink. Each subframe is further subdivided into 5 slots with duration of 1msec.

The LTE resource blocks are divided among time and frequency grids and a number of resource blocks are allocated to each user. The number of allocated resource blocks will describe the resource elements a user gets along with AMC technique applied on. The higher the resource elements, higher the rate of modulation and higher the bit rate. The number of resource blocks will be decided by the scheduling mechanism being deployed in time and frequency domain. The scheduling schemes will be helpful in optimal performance of LTE networks in different QoS scenarios.

5.3 Concept of Load Balancing

Load balancing in telecommunication is described as the procedure of distributing the information or packets among available base stations to manage more requests or calls so that the entire system can accomplish the jobs more efficiently. The load in the different networks can be balanced in two ways. The static load balancing is not dependent on the available resources in the system whereas the dynamic load balancing takes decisions based upon the current load availability of the system and availability of resources.

5.3.1 Load Metric

The role of load metric is to describe the load balancing status of the network accurately. Load metric can be described in number of ways based upon the characteristics of the network. The metrics such as total number of calls, probability of call blocking, throughput, delay, packet error are used in wireless network.

The proposed algorithm represents node throughput and symbol error rate as load metric because it depends upon the amount of traffic generated by each node and number of jobs scheduled. The nodes are randomly distributed within the network. The system is considered as under loaded, overloaded or balanced depending upon the traffic level. The random distribution of nodes in the LTE networks is shown in Fig. 5.4 where red node is considered as source and green node is destination. The black nodes help the scheduling algorithm to send the traffic from source to destination along with load balancing.

The load on a node depends upon the throughput of that node and the number of neighboring nodes sharing the load. Due to varying nature of traffic and traffic level at every time instant, it is difficult to calculate the load balancing metric exactly for every time span. Hence, load balancing index parameter is considered to presume the average traffic load on each node.

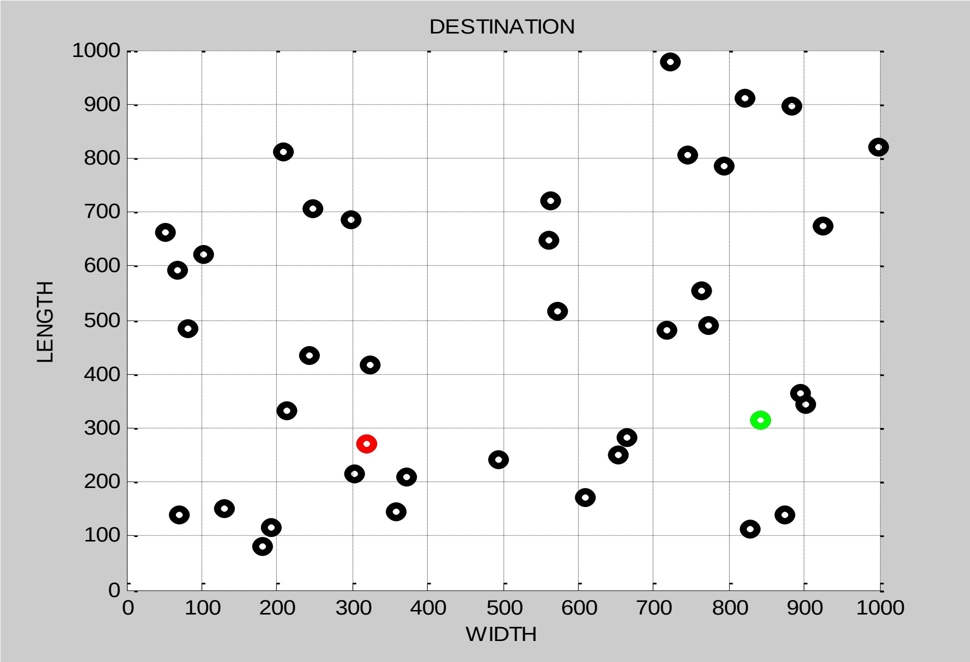

Figure 5.4: Traffic Distribution from Source to Destination

The load on node *i* is represented as follows:

$$L = \sum B_i^2 / n$$

and the load balancing index is:

$$£(t) = (\sum B_i)^2 / (n \sum B_i^2)$$

where B_i is the throughput of node i and n is the number of neighboring nodes.

5.3.2 Load Balancing Mechanism

Load balancing schemes are utilized to resolve the problem of overloading cells. These schemes are classified into two categories as Resource Allocation Schemes and Load Distribution Schemes as shown in Fig. 5.5.

5.3.3 Resource Allocation Schemes

The idea for balancing the load in the system is to allocate the resources in such a manner that almost all the users are satisfied. This scheme includes two types of channel allocation methods. In Fixed Channel Allocation (FCA) Scheme, a fixed

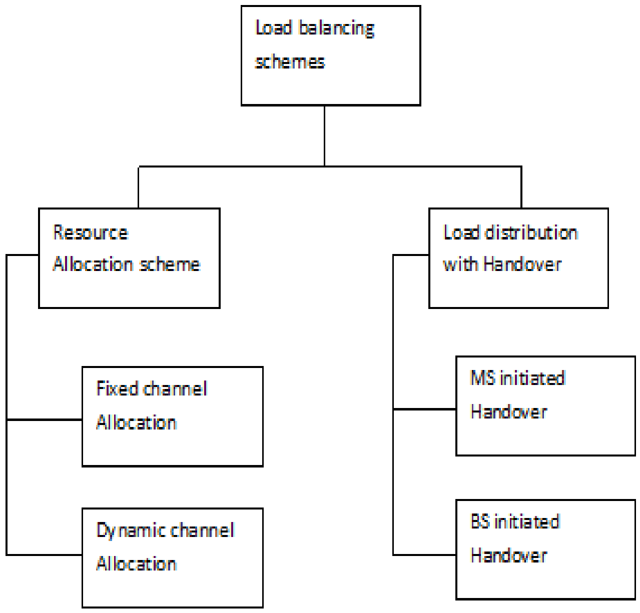

Figure 5.5: Load Balancing Schemes in LTE Networks

numbers of channels are allocated to each base station. Though, this scheme does not use the channel efficiently because of the varying nature of traffic. Dynamic Channel Allocation (DCA) scheme is an enhancement to the FCA which can adapt itself with changing traffic and adjusts frequency assignments relevant to the traffic load.

5.3.4 Load Distribution Schemes

In load distribution scheme, the traffic is directed towards the resources. This load balancing scheme includes two ways. In the first process, the mechanism checks the overloaded cells and load on MSs. The MSs in that region are connected to less congested access points [139]. In the second process, the load balancing algorithm is applied on BSs to force the MSs to handover the load to less congested BSs.

5.3.5 Load Balancing with Mobile IP

The operating mode of Mobile IP is similar to the post-office service. Every time a mobile node is required to be attached to new point at any different place. The requirement for service to register its location i.e. care-of address with its home agent to maintain the mobile node's permanent and temporary care-of addresses is represented as a mobility binding [180]. After successful registration, the home agent works for traffic management in mobile nodes and tunnels all the incoming traffic to the recent registered location of mobile node.

Load balancing problem is designed using optimization algorithm for LTE network in two steps. In the first step, minimum number of resources is considered to be utilized to send the traffic in LTE network to make the recourse utilization ratio α (t) minimum. In the next step, load is required to be evenly distributed among resources available to maximize the load balancing index £ (t). The load is unbalanced if α (t) is minimized and increasing £(t) may increase resource utilization. Hence, to balance the trade-off generated between the two parameters, a new aggregation parameter is defined as:

$$Z = £\alpha (t) - (1 - £) (\alpha)$$

The value of £ (t) lies between 0 and 1. £ (t) tends to 0 implies that load balancing objective is fulfilled with minimum number of resources and £(t) tends to 1 implies resources are fully utilized for load balancing. The suitability of the algorithm is to set £(t) between 0 and 1. Therefore, the mathematical formulation of objective is as follows:

$$Min\ Z = £\alpha (t) - (1 - £) (\alpha)$$

5.4 LTE Uplink and Downlink Design

LTE uses OFDMA and SC-FDMA as multiple carrier access schemes in downlink and uplink design. SC-FDMA is a single carrier multiple access with Orthogonal Frequency Multiplexing technique and Frequency Domain Equalization having almost similar structure and performance as OFDM. SC-FDMA is adopted as the uplink multiple access scheme for LTE. The information bits can be adjusted using any modulation schemes like QPSK, QAM etc. In this thesis work, the QAM modulation is used. The purpose of using this modulation technique is not only to transmit message signals through a radio channel but it helps in achieving system capacity with quality of service, power efficiency and minimum bandwidth.

In SC-FDMA, the data is mapped into modulation symbols or into a block of N symbols. The DFT transforms N symbols in time domain into frequency domain. The output of DFT block i.e. samples of frequency domain is then mapped to a subset of M carriers. In this M is typically greater than N. Similarly as in OFDM, an M-point IFFT is used to originate the samples of time domain of subcarriers which is further followed by Cyclic Prefix (CP), P/S and DAC block and then inserted into RF subsystems. The data is DFT converted before aligning into subcarriers, therefore SC-FDMA is known as DFT pre-coded OFDM. In case of OFDM system, every data symbol is transported on separate sub-carriers whereas in SC-FDMA multiple subcarriers transport each data symbol due to frequency domain. While every data symbol is spread over multiple subcarriers, SC-FDMA proposes frequency diversity gain in a selective channel. In this manner, SC-FDMA can be considered as frequency spread or DFT OFDM.

5.5 LTE System Model

The system model may be described as the combination of network design model, link design model, load balancing conditions and scheduling algorithms.

5.5.1 Network Model

The LTE network model is consists of combination of number of cells represented as $N= (c_1+ c_2+c_3+ - - - - - +c_n)$ where c is the area of one cell. The total number of jobs scheduled is represented by K. As the network is working for heterogeneous QoS services, the K is defined as the sum of all types of CBR and VBR traffic. In the scenario discussed, mixed traffic type is considered and the packets are generated randomly. Time t is considered to provide the service to K jobs generated by M number of users is used as the time required for load balancing. R is the number of requests sent by users to complete k jobs. The probability of completion of requests is more than 70% for the designed network.

5.5.2 Link Model

The link model design for LTE simulation network is shown in Fig. 5.6. The traffic generated from the transmitter, passes through the channel model along with signalling, reaches the receiver node and fed back Channel State Information (CSI) to the transmitting node. The output is produced at the receiver end in the form of BER, SER and throughput with respect to SNR.

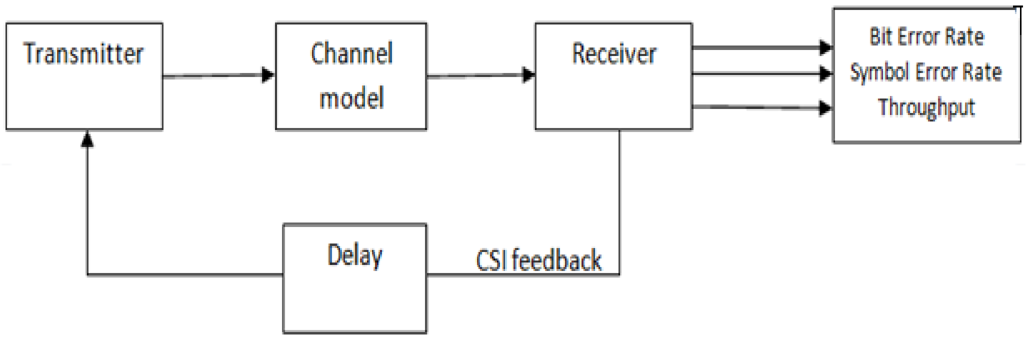

Figure 5.6: Link Model for LTE Network Simulation

The mapping between the load and throughput in the system model is represented as:

$R\ (SINR_u) = log2\ (1 + SINR_u)$

where $R\ (SINR_u)$ is the throughput with respect to $SINR_u$ which is the signal to interference ratio at end user. The bandwidth required for transmission depends upon the relation:

$N_u = \frac{\text{Du}}{\text{R (S I N Ru).BW}}$

where N_u is the addictive white Gaussian noise and D_u is the distance between transmitting and receiving node. Since the load balancing is to done periodically for all the nodes available in a network, the average SINR is *E(SINR(t))* to represent expected QoS SINR between time (t-1, t). $I_{t,k}(t)$ is considered an assignment indicator with value varying from 0 to 1 depending upon the time required to serve K users. The maximum time span of service for one user is taken as 1msec long. The average bandwidth required is defined as:

$E_{I,k}\ (t) = log2\ [1 + E\ (SINR_{I,k}\ (t))]\ bps/Hz$

The resource allocation depends upon the type of service for K users for QoS considerations.

5.5.3 Load Balancing in the Network Model

The network proficiency is required to be increased to get the required QoS in traffic transmission and to get the decent performance of a network. The load balancing is done in self-organizing networks to balance the load among cells and nodes and keeping the network throughput at high rate. The approach works well by introducing scheduling algorithms with load balancing. The power allocation in a cell P_c with respect to cell load is represented as:

$$Pc = \frac{1}{\frac{M}{PRB}} . \sum N u$$

If the power consumed in a cell increases, the load metric is exceeded to 100% cell load and the cell may not be able to serve all the users. The unsatisfied users in that network are:

$$Z = \sum max\ (o, Mc\ .\ (1- \frac{1}{Pc}))$$

The network in this condition is considered as overloaded network and can be balanced by transferring the load to the neighboring idle nodes. The load balancing problem is then resolved if the load will not be more than the capacity of neighboring nodes. If S_c is the signal power from neighboring nodes, the relation may be expressed as:

$$Sc = Pc\ .\ Lc\ (qu)$$

$$SINR_u = \frac{S2}{\frac{S1\ +S1-S2}{S\ I\ N\ Ru}}$$

5.5.4 Job Scheduling with Load Balancing

The job scheduling algorithm as shown in Fig. 5.7 helps the nodes to balance the load among different nodes available in the network to retain network stability and increasing throughput and reducing delays.

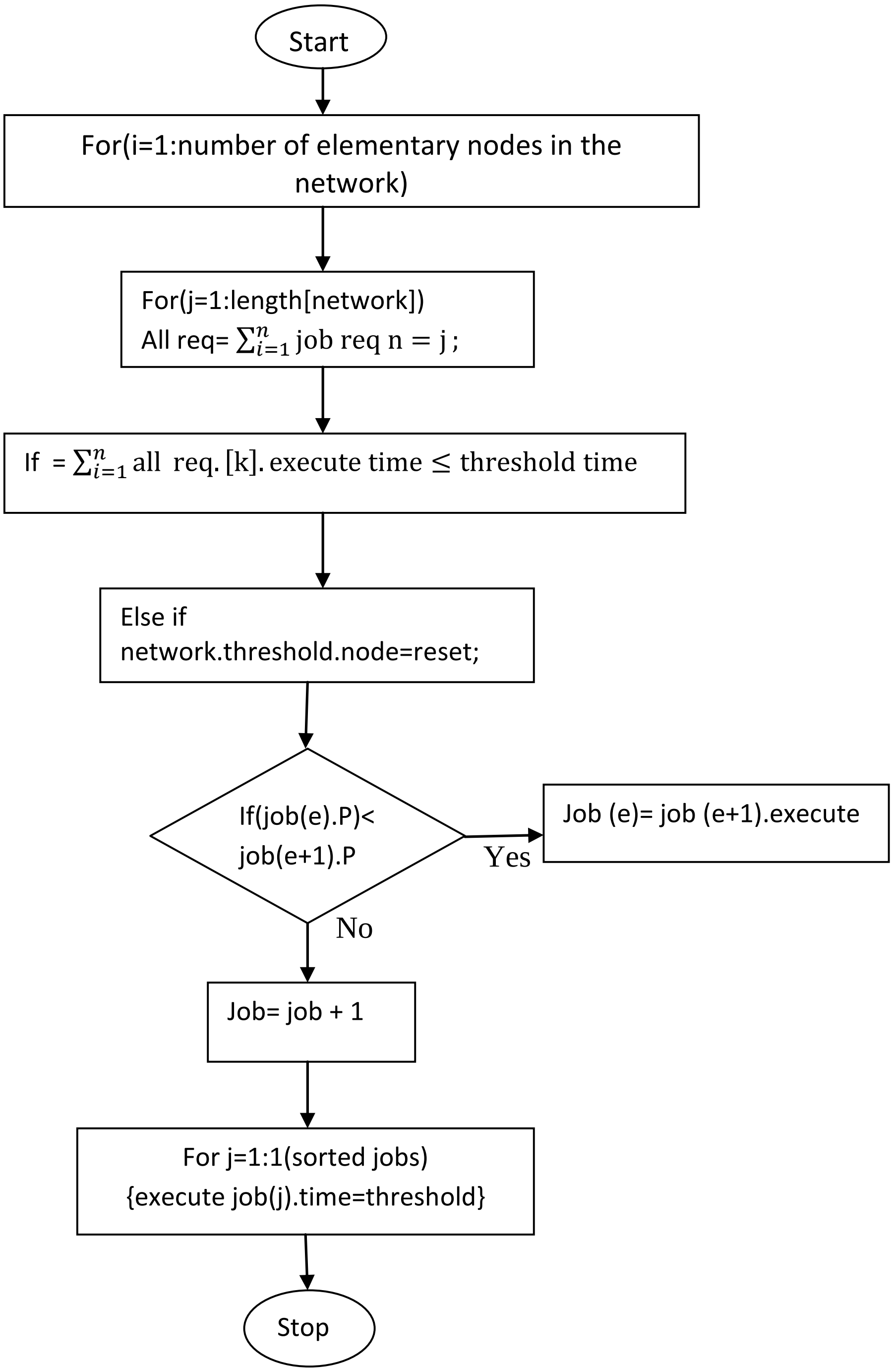

Figure 5.7: Flow Diagram for Job scheduling in LTE

The operating system works for scheduling the jobs by applying different scheduling algorithms. Before that the requirement is to prioritise the jobs in a particular time slot allocated for processing. The jobs are arranged in a queue in the classifier which assigns the priority to queue of jobs and the scheduling is done to process all jobs without delays and required QoS. There are a number of scheduling algorithms discussed in previous research but the proposed work is mainly concentrated on FCFS, RR, the hybrid algorithm which is the combination of FCFS/RR with priorities assigned to jobs.

As the priority is assigned to FCFS /RR queue, the queue will decide the processing as:

If p2 (A) > p1 (A)
Then p2 = TRUE.
Else p1.
If p2= TRUE
Till p1…pn =wait
Until
Arrival time (A) > other.

5.5.5 The Algorithm Description

For any scheduling algorithm, number of tasks generated has to be a major parameter. In our simulation environment the following parameters have been configured.

Sn = Source node from where the data has to be transmitted.
Dn = destination node where the data has to be sent
1) As there is not any fixed scenario hence the source and the destination nodes are taken as random
2) En = Energy getting consumed by node per second in the simulation
Enall = $\sum$En(i) where i is the current node
Et = Current Job Energy
Es = Total Energy
e = Error Rate
Tpc = Total Jobs
3) Once the energy, source node and the destination node is decided number of data packets which has to be send are generated.
The specifications of the data packets are Packet_Id, Packet_transfer_energy, Packet_life_time

The algorithm is configured in the following manner:

1) START

2) Initialize data_packets

3) Initialize source node

4) Initialize destination node

5) E=0;

6) T=network life time for the completion of the tasks

7) For i=1:length (jobs)

8) E(i)=E+Ej; Tc=T+Et;

9) While Tc<T

10) Repeat procedure

11) Stop

Initialization parameters would be same for round robin algorithm represented by Steps 1 to 6. The proceeding algorithm is defined as follows:

7) For i=1:length (jobs)

8) Tpc=Total_jobs/T;

9) For k=1:tpc:T

10) If k==tpc

11) Switch job.

12) E(i)=E+Ej; Tc=T+Et;

13) While Tc<T

14) Repeat procedure

15) Stop

The Proposed Hybrid Algorithm:

1) START

2) Initialize data_packets

3) Initialize source node, Sn

4) Initialize destination node, Dn

5) E=0;

6) T=network life time for the completion of the tasks

7) For i=1:length(jobs)

8) *If job.queue.count==1*
9) *Execute.queue job*
10) *Elseifjob.count>1*
11) *Total_time_left=total_time-sum(alljobs.time)-firstjob.time;*
12) *Allocated_slot_time=Total_time_left/Total_jobs-1*
13) *Sn.initialize data packets*
14) *Dn=Receive job_request (Tcp)*
15) *Tpc=Total_jobs/T;*
16) *For k=1:tpc:T*
17) *If k==tpc*
18) *Switch job.*
19) *E(i)=E+Ej; Tc=T+Et;*
20) *While Tc<T*
21) *Repeat procedure*
22) *Stop*

5.6 Results and Discussions

5.6.1 Simulation Parameter Environment

The simulation parameters for LTE system consists of system bandwidth of 10 and 20MHz working with a 1000 packets per symbol with number of OFDM symbols simulated are 100, 500, 1000 and 1500 for uplink and downlink and may be extended. The channel considered for simulation with transmitter is AWGN channel. The mapping used for simulation is 16PSK, 32QAM or 64QAM. Generally 64 QAM is preferred for LTE symbol transmission. Random distribution of nodes is considered to move around the selected region in LTE environment. Let the 60 nodes are working in the selected area of 1000mX1000m generating 25 jobs, then almost 15 jobs have been successfully completed as shown in Fig 5.8. The number of nodes working in the region may vary up to 300 and able to support large number of jobs.

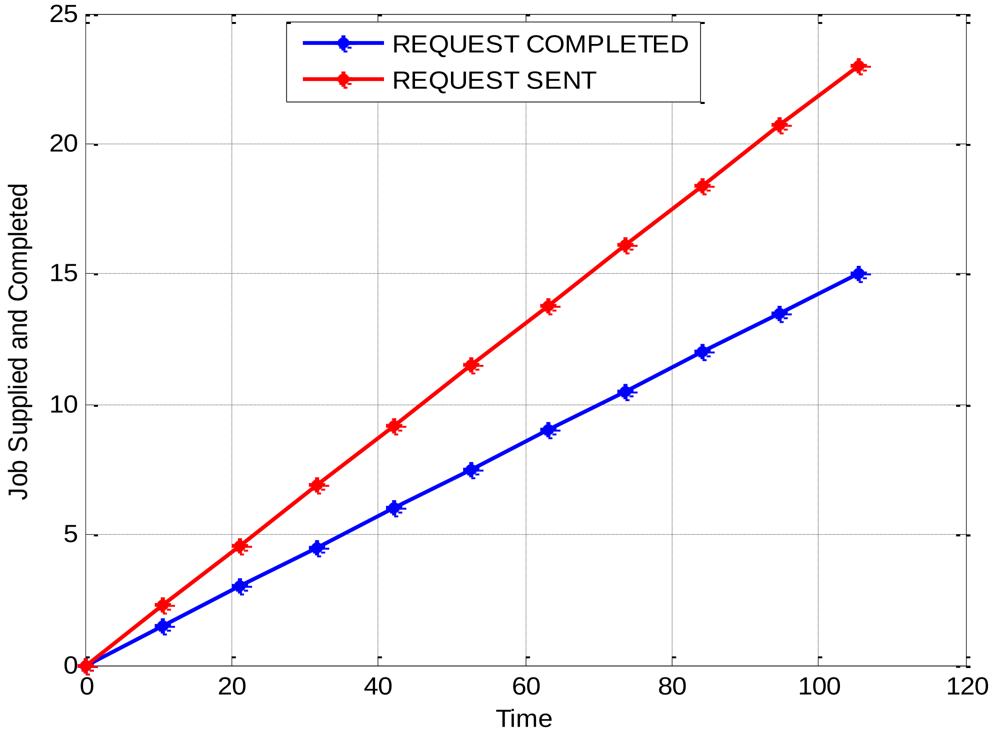

Figure 5.8: Requests Sent and Completed in LTE Environment

5.6.2 LTE Simulation for Three Different Scenarios

The three scenarios have been considered for the simulation of LTE model designed for different scheduling algorithms. The scenario I consists of a frequency range varying from 0-6GHz, working for LOS and NLOS in the building, indoor small offices or residential environment. The scenario II consists of indoor to outdoor communication, outdoor urban environment covering a range of 0-6GHz, working for LOS and NLOS conditions. The scenario III covers a typical urban microcell in LOS and NLOS, working as hotspot in a definite range with a frequency of 0-20MHz.The actual range depends upon the frequency of transmission and antenna height. The BER and throughput results have been analyzed for the three scenarios.

5.6.2.1 Simulation for Indoor Environment: Scenario I

As the simulation environment considered for this case is indoor office building, the strength of the signal getting affected. Generally the antennas are placed outside the building or above on the roof top, the signal gets weaker while reaching the indoor environment to access the mobile devices. The parameters for the scenario I are considered the same as described in previous section. The BER is measured with respect to number of iterations to get the desired results by means of repeated cycle of operations. The simulation is run for the BER comparison of three scheduling algorithms as FCFS/RR, Priority and hybrid scheduling as shown in Fig. 5.9.

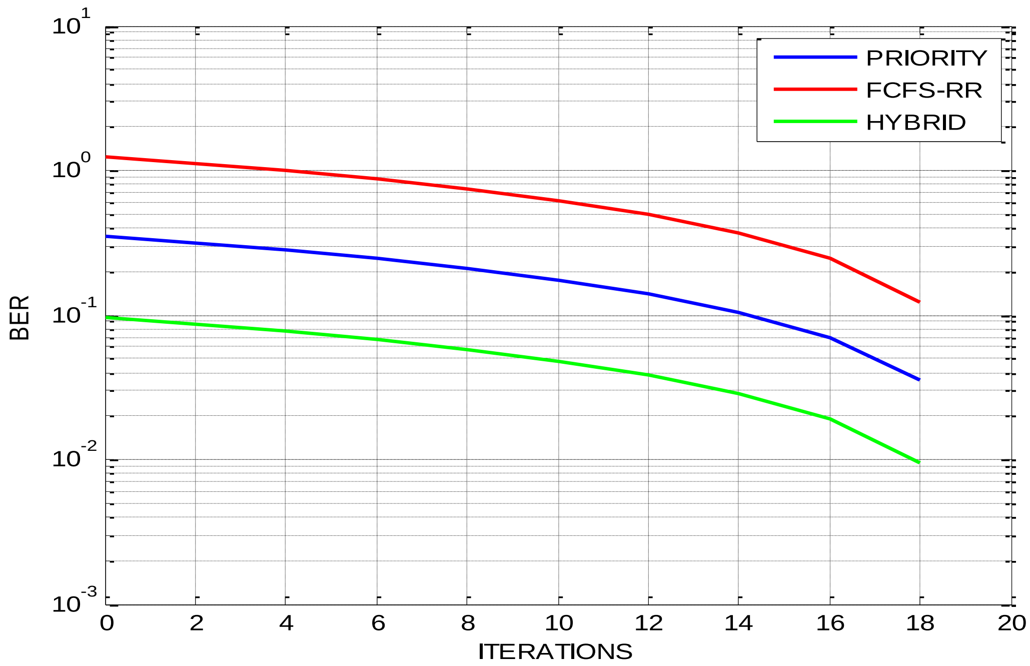

Figure 5.9: BER Comparison of Three Scheduling Algorithms in Scenario I

Hybrid scheduling provides less amount of BER as compared to FCFS-RR and priority scheduling algorithms. As the error rate is reduced, the accuracy during packet transmission increased. The results are obtained for BER by varying number of packets in a block. Fig. 5.10 and Table 5.1 compare the results for the above mentioned parameters by varying the block size.

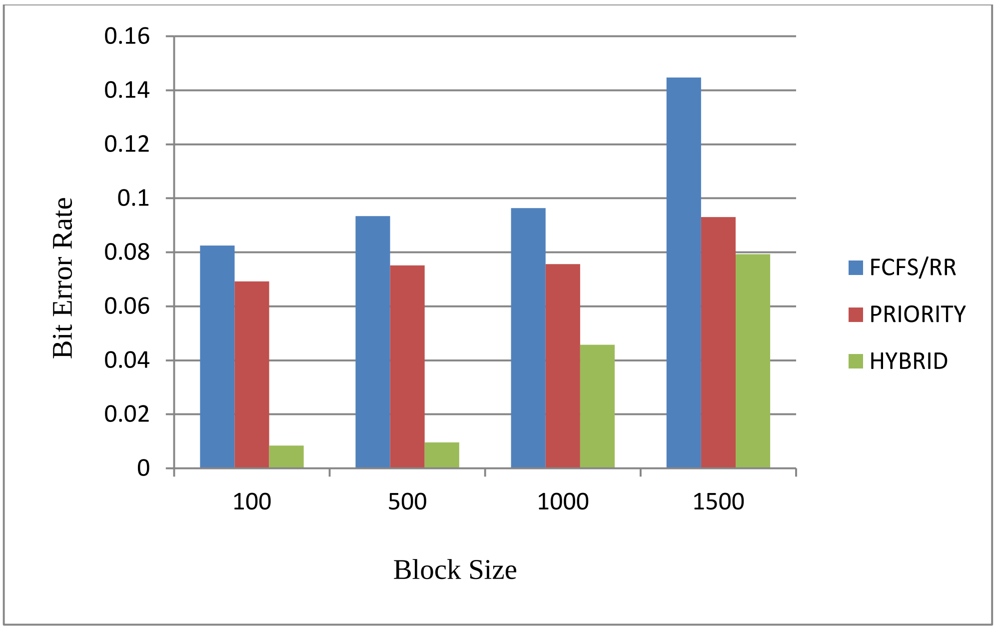

Figure 5.10**:** BER for Three Scheduling Algorithms by Varying Number of Symbols in Scenario I

Table 5.1: BER for Three Scheduling Algorithms by Varying Number of Symbols in Scenario I

Type of Scheduling	**BLOCK SIZE**			
	100	500	1000	1500
FCFS/RR	0.08253	0.09336	0.09638	0.1448
PRIORITY	0.06923	0.07516	0.07561	0.09302
HYBRID	0.00845	0.009571	0.04577	0.07935

The simulation is run for the throughput comparison of three scheduling algorithms as FCFS/RR, Priority and hybrid scheduling as shown in Fig. 5.11. Hybrid scheduling provides greater amount of throughput as compared to FCFS/RR and priority scheduling algorithms. As the throughput is increased, the accuracy during packet transmission increased and hence quality is enhanced.

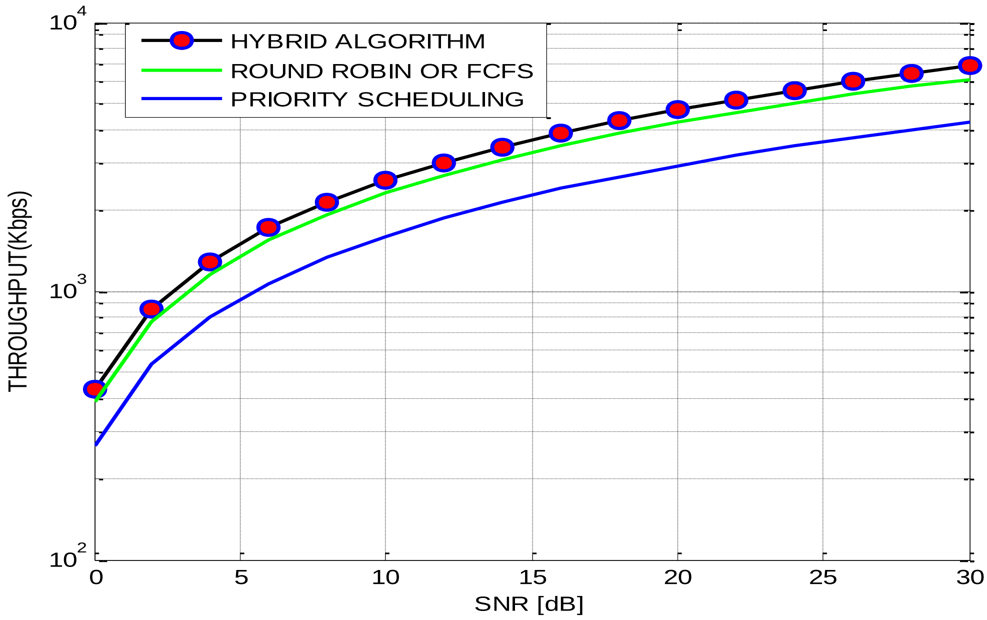

Figure 5.11: Throughput Comparison of Three Scheduling Algorithms in Scenario I

The results are obtained for throughput by varying number of packets in a block. Fig. 5.12 and Table 5.2 compare the results for the above mentioned parameters. The results represent the high throughput for hybrid algorithm as compared to FCFS/RR and priority scheduling.

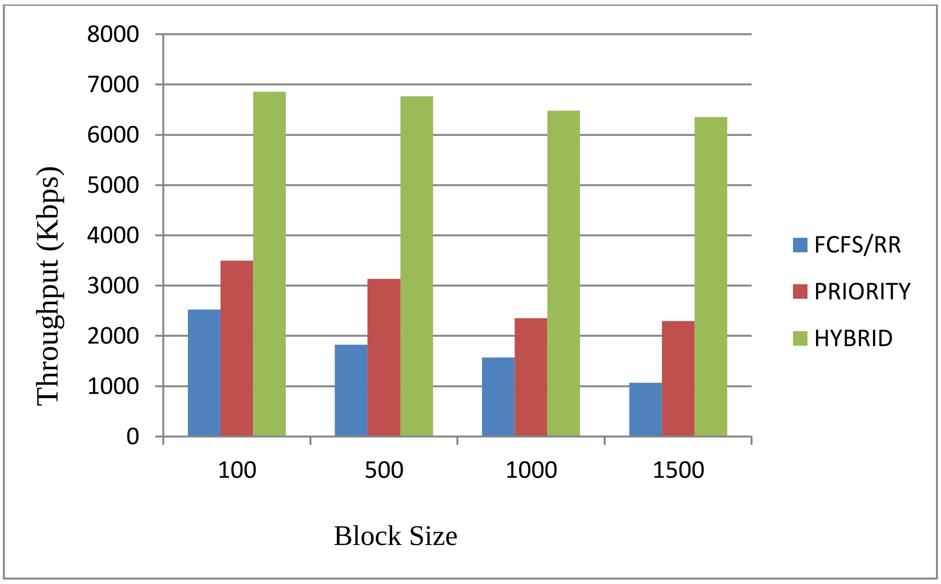

Figure 5.12: Throughput for Three Scheduling Algorithms by Varying Block Size in Scenario I

Table 5.2: Throughput for Scheduling Algorithms by Varying Number of Symbols in Scenario I

Type of Scheduling	BLOCK SIZE			
	100	500	1000	1500
FCFS/RR	2523	1828	1568	1069
PRIORITY	3497	3137	2354	2297
HYBRID	6856	6764	6481	6354

5.6.2.2 Simulation for Indoor to Outdoor Communication: Scenario II

The simulation environment is considered for indoor to outdoor communication, or an urban outdoor environment. The results for the BER are compared for the three scheduling algorithms by varying the block size as represented in Fig. 5.13 and Table 5.3. Hybrid scheduling provides less amount of BER as compared to FCFS/RR and priority scheduling algorithms. As the error rate is reduced, the accuracy during packet transmission increased.

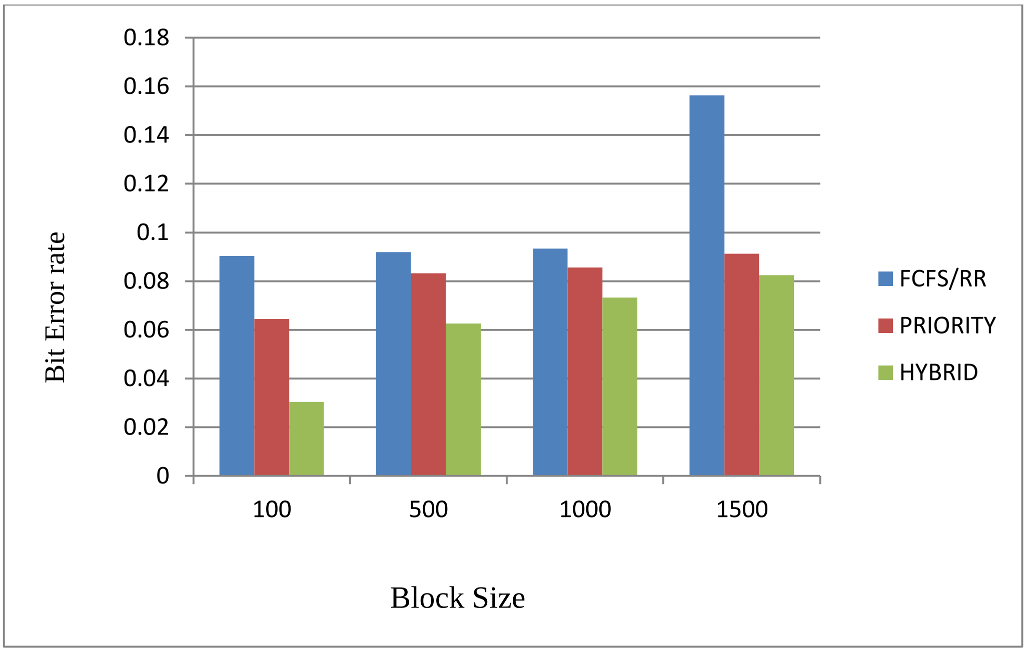

Figure 5.13: BER for Three Scheduling Algorithms by Varying Block Size in Scenario II

Table 5.3: BER for Scheduling Algorithms by Varying Block Size in Scenario II

Type of Scheduling	**BLOCK SIZE**			
	100	500	1000	1500
FCFS/RR	0.09038	0.09199	0.09341	0.1564
PRIORITY	0.0645	0.08326	0.08561	0.09126
HYBRID	0.03037	0.06256	0.07326	0.08247

Hybrid scheduling provides greater amount of throughput as compared to FCFS/RR and priority scheduling algorithms. As the throughput is increased, the accuracy during packet transmission increased and hence quality is enhanced. The results are obtained for throughput by varying number of packets in a block. Fig. 5.14 and Table 5.4 compare the results for the above mentioned parameters.

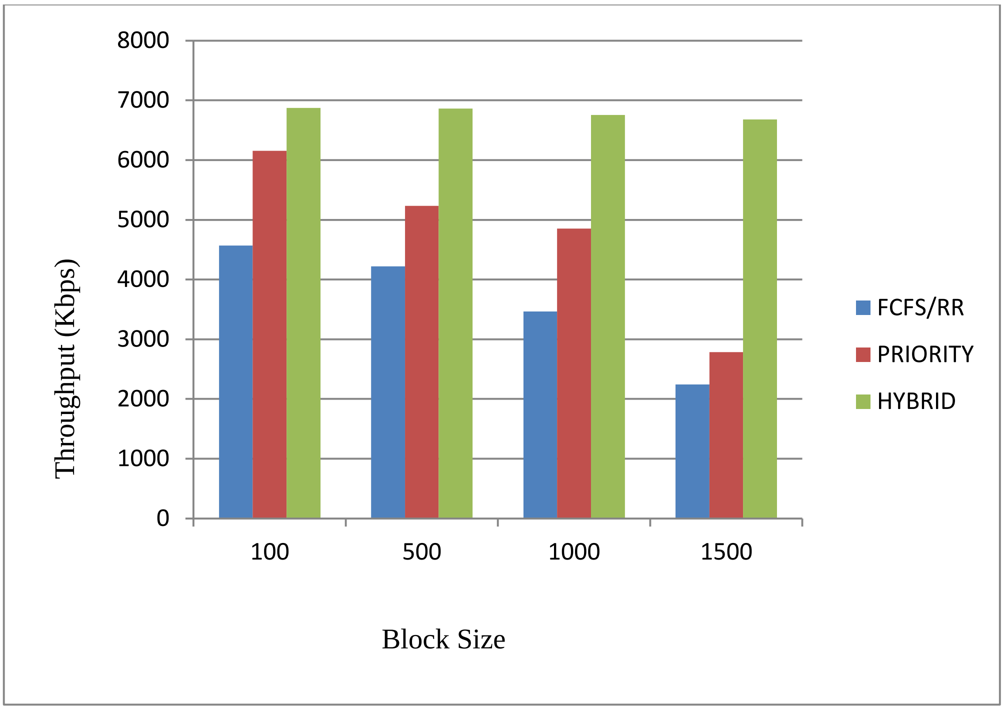

Figure 5.14: Throughput for Three Scheduling Algorithms by Varying Block Size in Scenario II

Table 5.4: Throughput for Scheduling Algorithms by Varying Block Size in Scenario II

Type of Scheduling	**BLOCK SIZE**			
	100	500	1000	1500
FCFS/RR	4568	4220	3467	2241
PRIORITY	6155	5231	4853	2785
HYBRID	6875	6863	6754	6682

5.6.2.3 Performance Evaluation in Urban Microcell: Scenario III

The scenario III consists of an urban microcell covering large area with the hotspot or the access point may also work as hot spot. The three scheduling algorithms are compared as shown in Fig. 5.15. Hybrid scheduling provides less amount of BER as compared to FCFS/RR and priority scheduling algorithms. As the error rate is reduced, the accuracy during packet transmission increased.

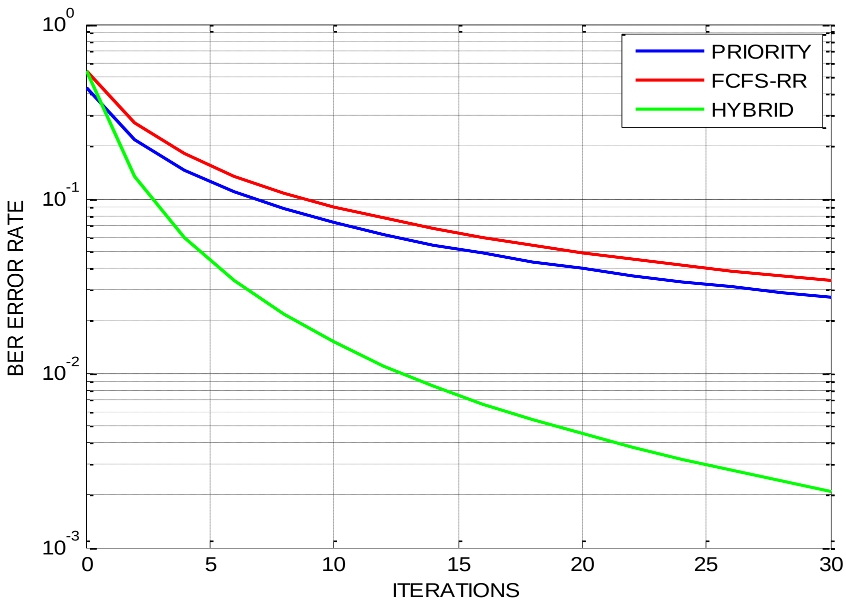

Figure 5.15: BER Comparison of Three Scheduling Algorithms for Scenario III

The results for the three scheduling algorithms are compared for BER and throughput by varying the block size as shown in Fig. 5.16 and Table 5.5. The throughput variation is not so high by varying the no of packets but BER rate is reducing for hybrid scheduling algorithm as compared to FCFS/RR and Priority scheduling.

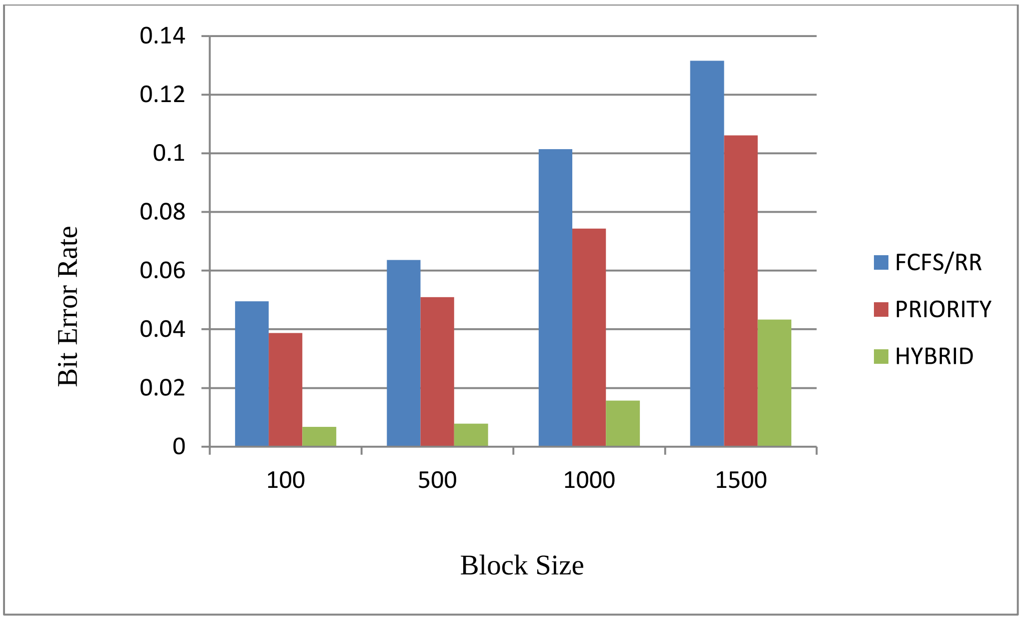

Figure 5.16: BER for Three Scheduling Algorithms by Varying Number of Symbols in Scenario III

Table 5.5: BER for Scheduling Algorithms by Varying Number of Symbols in Scenario III

Type of Scheduling	**BLOCK SIZE**			
	100	500	1000	1500
FCFS/RR	0.04958	0.0636	0.1014	0.1315
PRIORITY	0.03872	0.05092	0.07433	0.1061
HYBRID	0.006691	0.007868	0.01567	0.04327

Hybrid scheduling provides greater amount of throughput as compared to FCFS-RR and priority scheduling algorithms. As the throughput is increased, the accuracy during packet transmission increased and hence quality is enhanced. The results are obtained for throughput by varying number of packets in a block. The Fig. 5.17 and Table 5.6 compare the results for the above mentioned parameters.

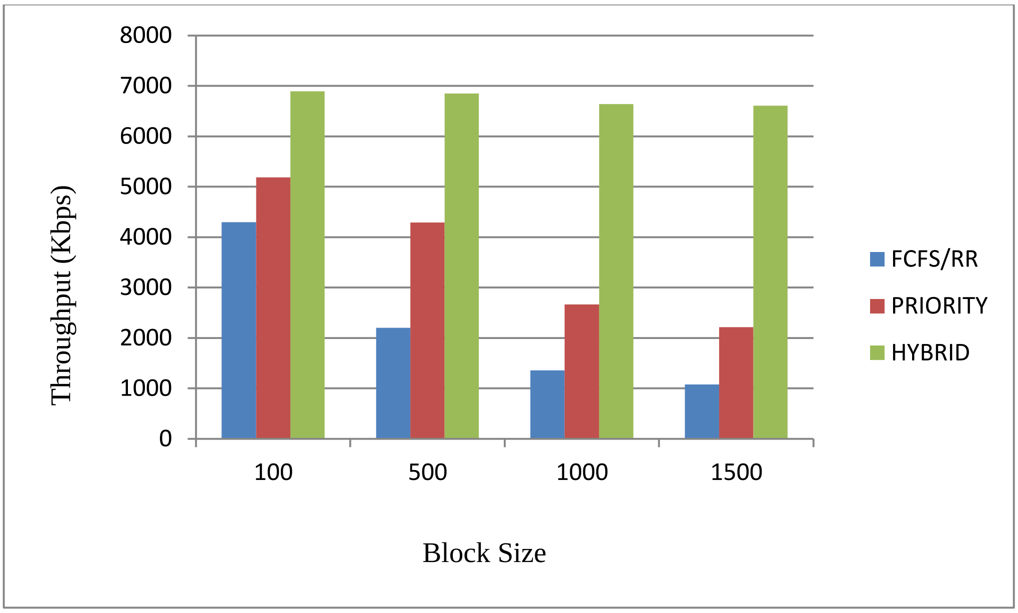

Figure 5.17: Throughput for Three Scheduling Algorithms by Varying Number of Symbols in Scenario III

Table 5.6: Throughput for scheduling algorithms by varying Number of symbols in Scenario III

Type of Scheduling	**BLOCK SIZE**			
	100	**500**	**1000**	**1500**
FCFS/RR	4295	2200	1356	1082
PRIORITY	5188	4289	2668	2217
HYBRID	6894	6851	6642	6605

5.6.3 Energy Consumed in Hybrid Scheduling Algorithm

The energy consumed in LTE network environment depends upon the number of nodes available in the network and the number of requests generated by different nodes. The energy consumed during transmission increases with the time required for job completion as shown in Fig. 5.18. The amount of energy consumed for 25 requests is 50mJ for a job completion time of 108sec which is lesser as consumed in WiMax network.

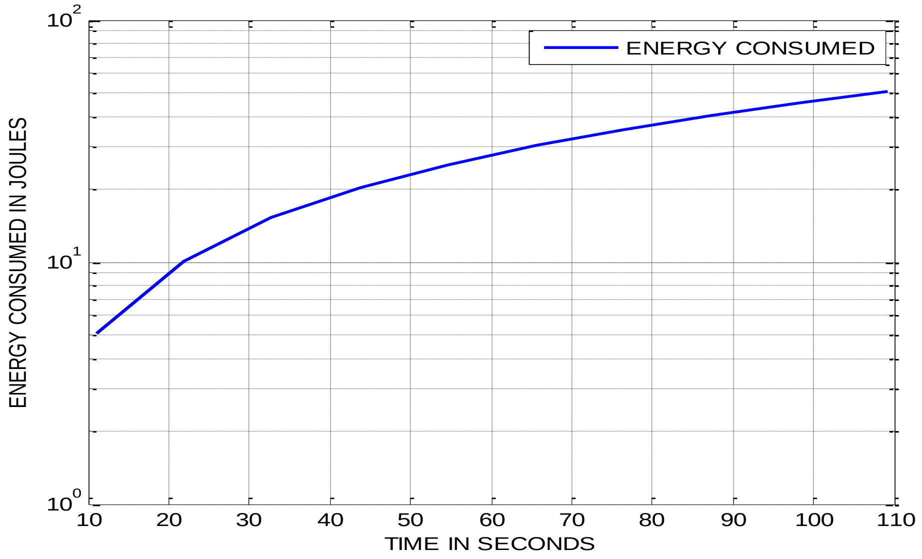

Figure 5.18: Energy Consumed in LTE Network for Hybrid Scheduling Algorithm

5.6.4 Performance Evaluation based on SER

The comparison of symbol Error rate for the above mentioned scheduling algorithms is represented in Fig 5.18. The symbol error rate is decreased as the SNR is increasing in case of Hybrid scheduling as compared to FCFS/RR as shown in Fig. 5.19.

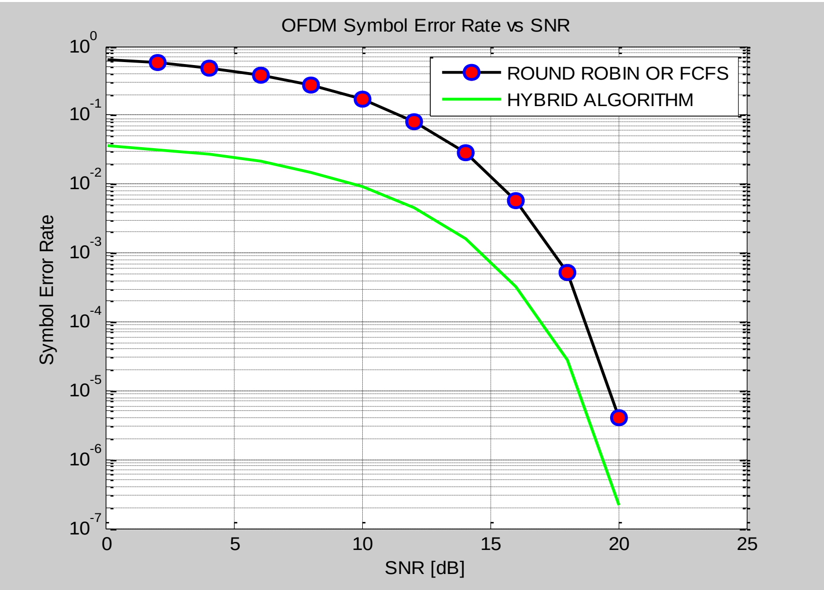

Figure 5.19: Symbol Error Rate vs SNR for Hybrid Scheduling in LTE

For hybrid scheduling, the results are also compared by changing the packet size as 1024, 2048 and 4096 along with changing the number of symbols 200 and 500. As the SNR is increasing, SER is decreasing as shown in Fig. 5.20.

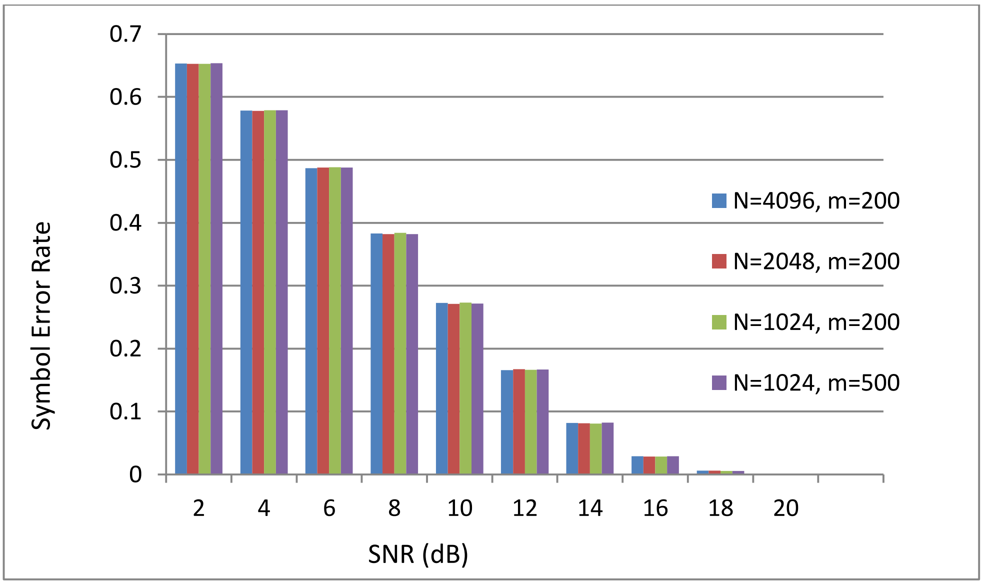

Figure 5.20: Symbol Error Rate Vs SNR for Hybrid Scheduling Algorithm

The comparative analysis of SER with SNR is by changing the value of CP=1, 1/4, 1/8, 1/16 is compared in Table 5.7.

Table 5.7: SER for Scheduling Algorithms by Varying CP

	CP=1		CP=1/4		CP=1/8		CP=1/16	
SNR	RR/FCFS	Hybrid	RR/FCFS	Hybrid	RR/FCFS	Hybrid	RR/FCFS	Hybrid
0	0.7467	0.1306	0.7356	0.5961	0.7354	0.1515	0.7466	0.4791
2	0.6742	0.1179	0.6758	0.4791	0.6741	0.1388	0.6714	0.4309
4	0.5931	0.1038	0.5912	0.3891	0.5913	0.1218	0.591	0.3793
6	0.4755	0.08318	0.4801	0.288	0.4868	0.1003	0.4762	0.3056
8	0.352	0.06158	0.3554	0.1785	0.3547	0.7306	0.3569	0.229
10	0.2244	0.3925	0.2203	0.0889	0.2226	0.4585	0.2267	0.1455
12	0.1091	0.1909	0.1097	0.2885	0.1084	0.2233	0.1135	0.07284
14	0.0379	0.00663	0.0356	0.00591	0.038	0.007827	0.0363	0.0233
16	0.0066	0.00115	0.0073	0.00032	0.0062	0.001277	0.0063	0.00404
18	0.0004	0.0007	0.0004	0.0005	0.0005	0.000010	0.0006	0.00038

5.7 Summary

A number of scheduling algorithms are proposed by various researches discussed already having advantages and disadvantages depending upon the amount of traffic, number of jobs scheduled and environment conditions. The low complexity scheduling algorithm is implemented in this chapter by considering the concept of load balancing. The load balancing is either dependent upon resource sharing or handover strategy. The network model and link model is designed based upon load balancing to optimize the problem. The chapter also describes the three scheduling algorithms for LTE network working in three different environments. The results are analyzed and compared for BER, SER and throughput for different values of packets and by varying the size of packets. Hybrid algorithm performs better as compared to FCFS/RR algorithm as represented in results.

Chapter 6

QoS BASED SCHEDULING ALGORITHMS IN HETEROGENEOUS NETWORKS

6.1 Introduction

The motivation behind the integration of different access technologies in a converged 4G network is to schedule the traffic between different access systems or networks. The greatest challenge is the effective load balancing among heterogeneous networks. Traffic scheduling along with load balancing requires a strong control from the network side as well as service provider. Firstly, the access network selection decides the access node in a particular network. If more than one number of users will select the same access node, it will get overloaded. Hence, for the purpose of load balancing, the users are required to transfer the load to other nearby access node or the network may be changed in some cases. In this work, a hybrid approach has been designed to schedule the traffic which can hide the heterogeneity of different networks. The objective to design the scheduling algorithm is to improve the overall network performance and to minimize unnecessary load balancing operations. The work is further improved by considering parallel processing scheduling schemes.

6.2 Access Selection in Heterogeneous Networks

In an environment of multiple access technologies consisting of multitude of devices, the concept of being always connected has become the hot topic of research. This refers to being not only always connected to best available network, but also utilizing the available networks and systems to provide high QoS. Heterogeneous networks represent the combination of more than one number of networks in which users are connected with the help of same or different network operator and service provider. The relationships between the user and the access network operators & service provider is shown in Fig. 6.1 and Fig. 6.2, while Fig. 6.3 introduces the application service provider as well. Scenarios may also be represented as depending upon

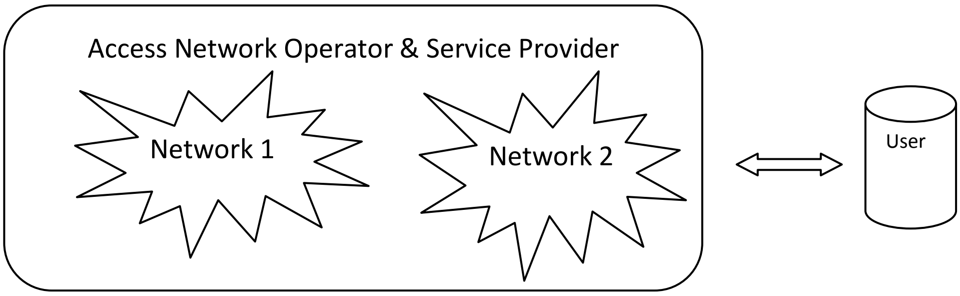

Figure 6.1: Network Operator and Service Provider with Multiple Access Networks

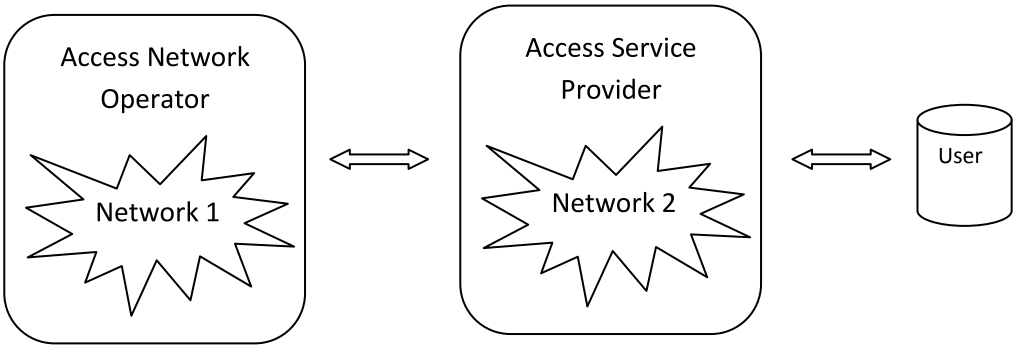

Figure 6.2: Network Operator and Service Provider with Different Networks

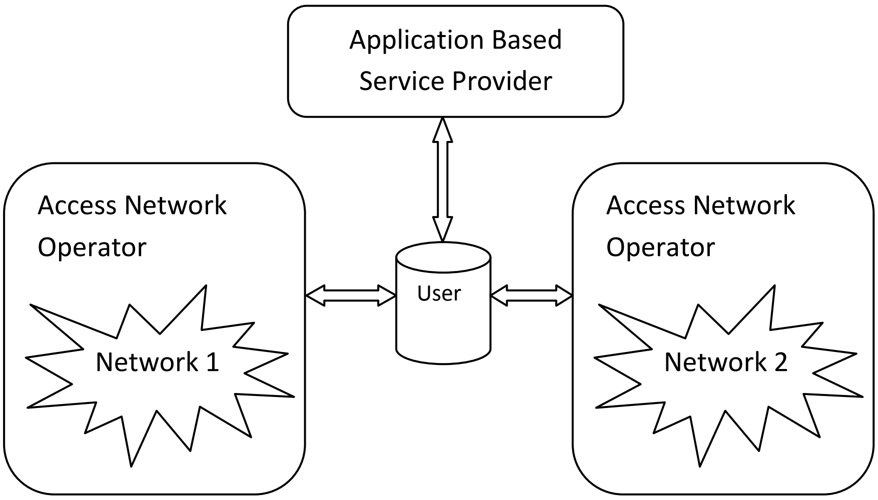

Figure 6.3: Network Operator with Application Based Service Provider

combinations of different conditions. The service provider has agreements with operators to provide the user with access to these networks. The resources for the users may be requested or purchased from any other service provider also. The requirement is to allow the users to transfer an application session from one network to another, without losing data or needing to restart the transmission.

The decision based process of a network to select the access networks for maintaining the service contiuity at any point is defined as access selection. The criteria for access selection is based upon terminals available, networks available or user involvements. Based upon the criteria, different aspects are required to be taken into consideration. The choice of access also depends upon the user preferences and service provider's ABC service cooperation. The network criteria depends upon the bandwidth available, cost and error rate, device capability and applications to be accessed. Initially, the terminal has no information for ABC network function. Hence, a stored data base is required to get connectivity while choosing the access. Network based selection is quite beneficial for load balancing, to reduce interface signalling and best connectivity among number of access terminals. The network based access selction provides full resource utilization with efficient selection performance along with maximizing the system throughput and minimizing the error rate. The issue related to access selection is providing easy methodology to allow terminals to choose ABC network with successful utilization.

The convergence towards heterogeneous networks is one of the evolutions of 4G platform to get the advantages of full bandwidth coverage. The cellular networks, WiFi, WiMax and satellite networks will ensure seamless roaming and global mobility with end to end QoS with affordable benefits. Heterogeneous network is all about convergence of different networks and technologies representing the umbrella of multiple technologies to represent seamless mobility approach. Interconnectivity and mobility between WiMax and LTE is the aim of this chapter.

6.3 Heterogeneous Network Design Model

A heterogeneous network comprising of LTE and WiMax is considered as a system model shown in Fig. 6.4. The requests have been generated by different users within a network and initialized with the nodes available in that network. The requests are supposed to be initialized in the LTE environment. If the LTE is not able to satisfy the requests, the requests are transferred to the WiMax network. The job model is generated for the initialized requests. The job Model consists of FCFS, RR and Priority algorithms. The three algorithms are combined to generate a hybrid scheduling algorithm for the requests available in heterogeneous environment.

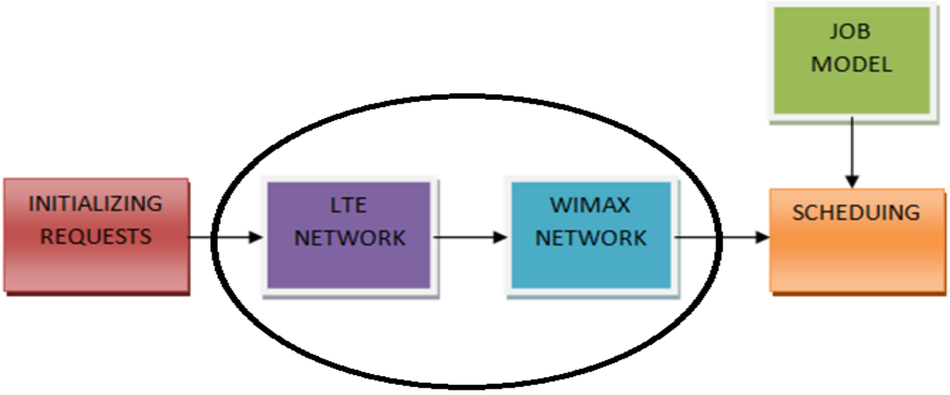

Figure 6.4: Design of Proposed Heterogeneous Network

Fig.6.5 represents a flow diagram to describe the heterogeneous network consisting of LTE and WiMax. The source node, destination node and number of packets are identified. The number of packets may vary depending upon the amount of traffic generated by the nodes in the network. The requests are generated by the packets and number of completed jobs depends upon the load distribution of nodes.

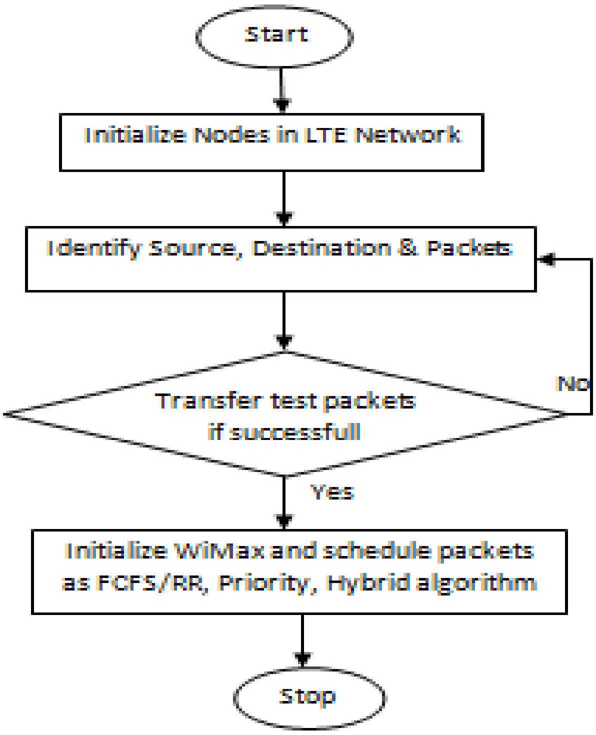

Figure 6.5: Flow Diagram of Proposed Heterogeneous Network

6.3.1 Network Model

The network model consists of 2X2 MIMO Channel considered to be based on path loss models working as line of sight(LOS) or Non-line of sight(NLOS). These models may be defined in a frequency range varying from 2GHz to 6GHz for different antenna heights. The different path loss models may be represented with different scenarios as indoor to indoor, indoor to outdoor, urban microcell, urban macro cell, rural microcell etc. The free space path loss model [181] may be defined as:

$$PL_{free} = 20\log_{10}(d) + 46.4 + 20\log_{10}(f_c/5.0)$$

where PL_{free} is the path loss, *d* is the distance between the transmitter and the receiver in meters, f_c is the system frequency in GHz. The transition from LOS to NLOS models can be done depending upon the requirement of the users and the type of traffic, the location of transmitting and receiving node as well as the type of network through which the packets generated and passed on to destination. The signal to noise ratio for uplink for the defined system model is represented as [182]:

$$R(SINR_u) = log2(1 + SINR_u)$$

A number of issues have been generated while discussing about the decent performance and gain of a network along with network proficiency. Lot of research is underway on load balancing in different networks individually or in hybrid manner. As LTE belongs to the category of self organizing network, the load balancing is not much required. The recent research represented a load balancing algorithm to enhance the network proficiency [141,182]. WiMax on the other hand, may also able to share the load among the cells in LTE networks by introducing scheduling algorithms. The FCFS/RR, Priority based and a newly developed hybrid algorithm have been implemented on the LTE-WiMax Network to add on the solution of issues in load balancing.

6.3.2 Network Model Configuration Parameters

To simulate the heterogeneous network, the length (L) and width (W) of simulation environment is considered in meters. The area defined for simulation is A (L*W). The number of nodes available and requests are distributed randomly in the specified area. The Up sampling factor (*Ups*) is generally considered for heterogeneous networks is chosen as that of the LTE network. The number of packets is also to be defined as N and number of simulated bits is taken as m. The network architecture for heterogeneous network has been designed by considering above mentioned parameters [182]. Parallel to serial conversion of the data takes place on it and has to be transferred in the available network in the heterogeneous environment.

6.3.3 The Algorithm Description

For any scheduling algorithm, number of jobs generated is one of the major parameter. In our simulation environment, the network consisting of LTE and WiMax is generated. Traffic in terms of jobs is passed on randomly in the heterogeneous environment. The choice of the network is done by nodes randomly depending upon the network availability and user preferences. The three scheduling algorithms have been designed in the heterogeneous environment namely FCFS or RR, Priority based algorithm and Hybrid algorithm. The hybrid algorithm is the proposed scheme which is the combination of FCFS and RR including job priority. The proposed scheduling algorithm's performance is compared with other mentioned algorithms and measured in terms of bit error rate and Signal to noise ratio and throughput analysis.

Proposed Algorithm for Scheduling

1) START

2) Initialize data_packets

3) Initialize source node,Ss

4) Initialize destination node, ds

5) T=network simulation time for the completion of the tasks

6) J=jobs in a network

7) Jf=feature set of the job

8) Ex=total execution time

9) Jcom= jobs completed

10) While T >Ex ll J > Jcom

11) If (queue.length==1)

12) Jcom= Job count1;

13) Ex= Ex+Jfe; where Jfe=execution time of current job

14) end

15) else if if(job.queue.list count >1)

16) Assign priority (Jlist)

17) Task count==1;

18) Execute job as per priority( );

19) Time slot for each job= T/J;

20) If (Job.Jfe <time slot)

21) Ex= Ex+Jfe else

22) Pending task [task count]=J;

23) end

24) while end if task count =task count;

To assign priorities to the task:

1) Function Assign priority (Jlist); to reach J in Jlist

2) temp J;

3) next= Jlist(J+1);

4) if(J.Jf.weight < Jlist(J+1).fe)

5) J= Jlist(J+1);

6) end

7) else if J.fe.weight==Jlist

8) d= tie breaker(J,Jlist(J+1))

9) end

6.3.4 Parameters Used for Simulation

The heterogeneous network topology is considered the same as in system model represented for LTE networks [182], nodes are considered to be randomly moving in the environment. The area defined for simulation is 1000 metersX1000 meters. The network bandwidth is considered as 10MHz to 20MHz which includes OFDM symbols and block size varying from 100, 500, 1000 and 2000 packets along with packet size of 512bits or 1024bits. There are number of channel models discussed in [181] working with LOS and NLOS path loss models. The heterogeneous network considered is combination of LTE and WiMax. The random mobility model is designed for distribution of user nodes throughout the area and move around the selected area and numbers of nodes are working for 70 job requests. Depending upon the parameters set for the configuration as shown in Table 6.1, more than 70% of jobs have been completed successfully for all the simulations being run by setting different set of parameters as shown in Table 6.1.

Table 6.1**:** Parameters for Simulation

Parameters for Simulation	**Parameter Values**
Size of OFDM Symbol N	4096, 2048, 1024
Number of OFDM symbols to be simulated, m	100 to 2000
Size of Alphabet M	2, 4, 8,16,32,64
Up-sampling factor L	1-4
Type of Mapping	PSK or QAM
Constellation Symbol Order	Binary or Gray
Size of cyclic prefix samples Ncp	1/4,1/8,1/16,1/32
Maximum number of RBs for DL	110
Number of subcarriers per RB	12
Number of OFDM symbols per RB for DL	7
Modulation level	2,4 for QPSK, 16 or 64QAM
Number of Cyclic Prefix	0,1 for normal or extended
CP=0	
Number of reference signals per RB	8
Cyclic prefix length of symbol	160
Cyclic prefix length of symbol	144
Number of RB in Downlink	80
Number of bit streams to simulate	18976

6.3.5 Performance Evaluation of Proposed Scheduling Algorithm

The hybrid scheduling algorithm is designed as the combination of FCFS/RR and Priority algorithm for heterogeneous network. The performance evaluation of hybrid scheduling algorithm is done on the basis of various QoS parameters as BER, throughput and energy consumed. The simulation is carried out in heterogeneous environment considered as combination of LTE and WiMax networks. Fig. 6.6 represents the number of jobs supplied and completed in heterogeneous environment with time duration of 106sec.

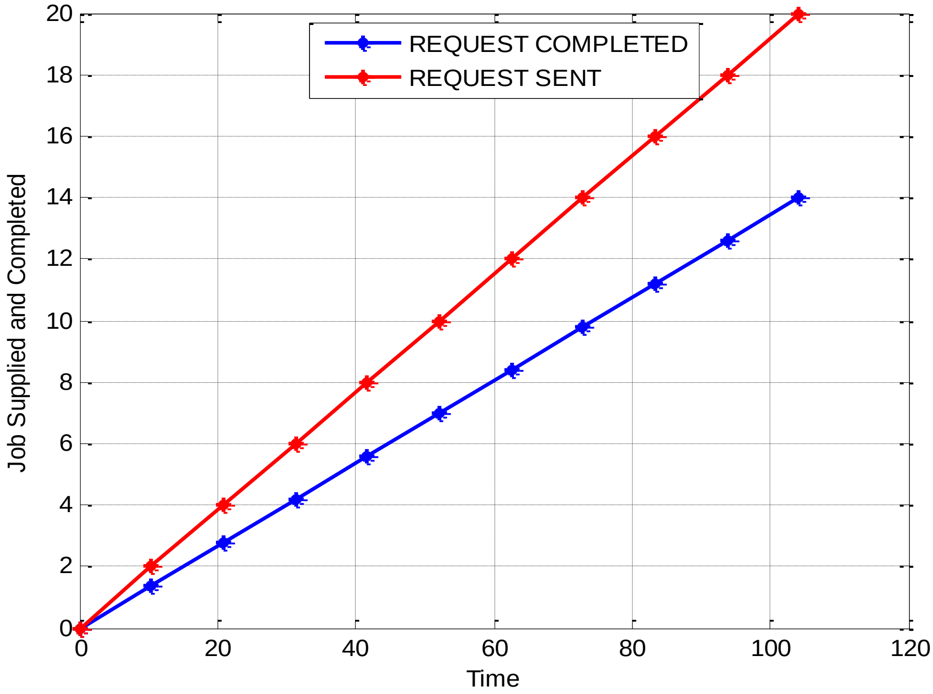

Figure 6.6: Jobs Supplied and Completed in Heterogeneous Environment

6.3.5.1 Energy Consumed in Heterogeneous Environment

The energy consumed in a wired or wireless network depends upon the power supplied for any fixed time duration. The energy consumed is calculated with respect to simulation time. The simulation time considered for requests generated by different nodes is 100sec. The total energy consumed by a network for any number of requests is the sum of initial energy and energy consumed per request. As the number of requests is increasing, the energy consumption in the network will increase. Fig 6.7 represents the energy consumption of 106 mJoules when the time allocated for 50 requests completion is 100sec. The energy consumption is more in the heterogeneous network as compared to WiMax and LTE network.

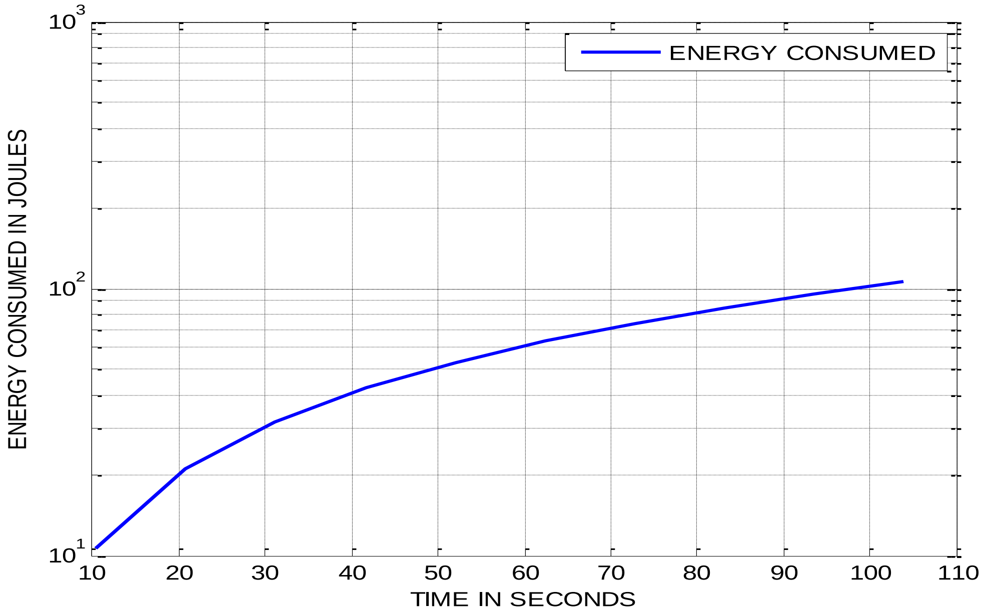

Figure 6.7: Energy Consumed for 50 Requests

By varying the number of requests generated, the energy consumed in the heterogeneous network is calculated as shown in Fig. 6.8. The number of requests is varied for the fixed simulation time and the energy consumption is increasing up to 254mJ for 70 requests.

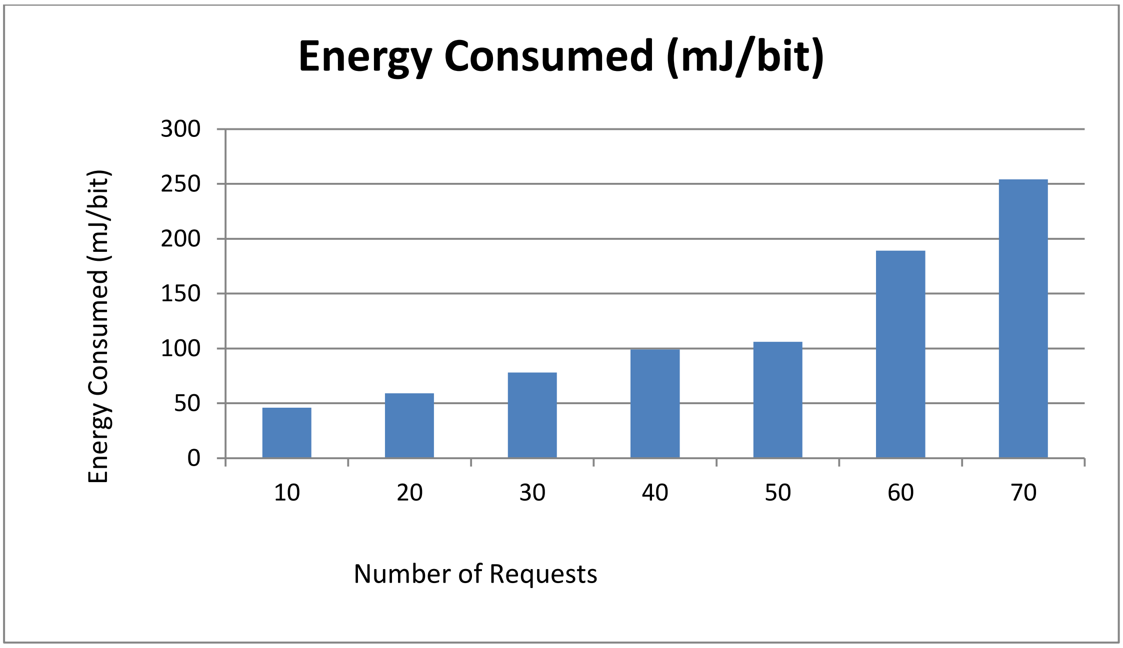

Figure 6.8: Energy Consumed by Varying Number of Requests

6.3.5.2 BER Analysis by Varying Block Size

The three scheduling algorithms FCFS-RR, Priority and hybrid are compared for BER analysis in the heterogeneous environment. The BER is calculted for 50 requests generated over fixed simulation time. Fig. 6.9 shows that hybrid scheduling algorithm has error rate of $4.9X10^{-3}$ is less as compared to FCFS-RR and priority scheduling.

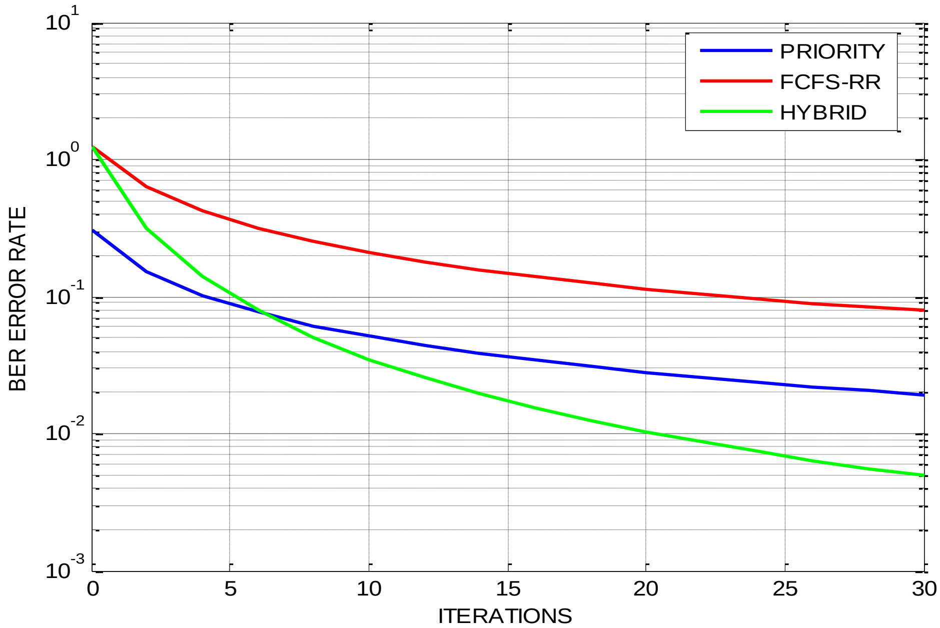

Figure 6.9: BER Comparison of FCFS-RR, Priority and Hybrid Scheduling

Table 6.2 and Fig. 6.10 represent the variation of BER for three scheduling algorithms by varying the block size. The block size is varied as 100, 500,1000 and 1500. As the amount of data is increasing, the number of errors have been increased. From the results shown, it is concluded that hybrid scheduling algorithm generates less errors as compared to other two scheduling schemes.

Table 6.2: BER of Heterogeneous Networks for Different Scheduling Schemes

Scheduling Type	**BLOCK SIZE**			
	100	**500**	**1000**	**1500**
FCFS/RR	0.02841	0.03531	0.06174	0.07853
PRIORITY	0.01221	0.02519	0.03811	0.05612
HYBRID	0.001776	0.002207	0.003859	0.004908

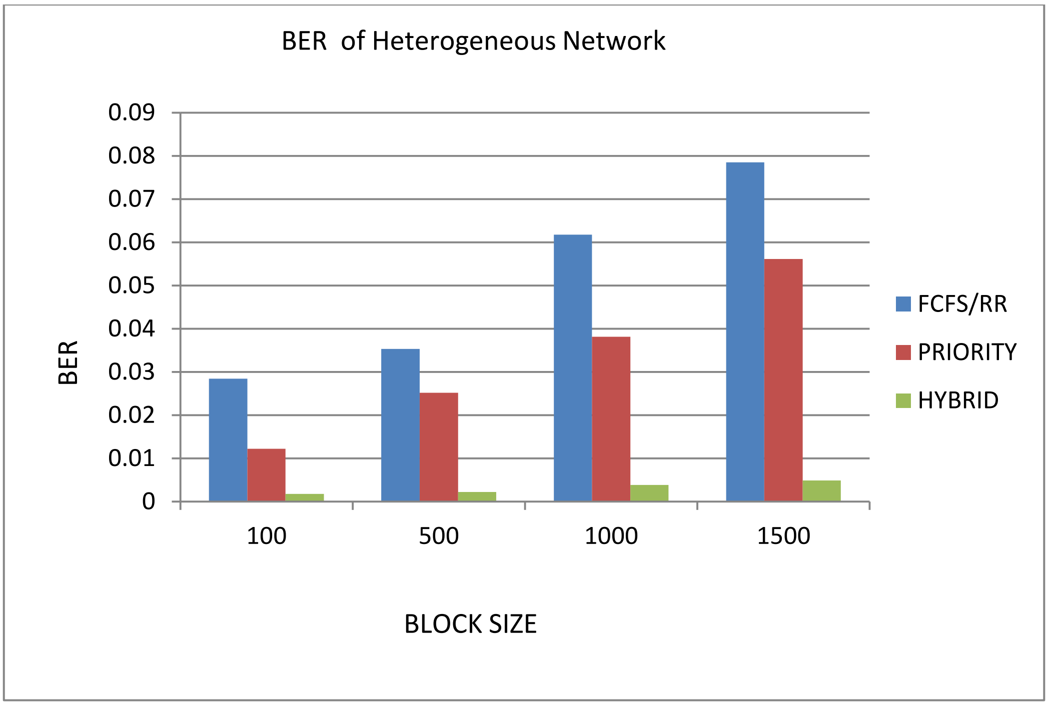

Figure 6.10: BER of Heterogeneous Network for FCFS-RR, Priority and Hybrid Algorithm

6.3.5.3 Throughput Analysis by Varying Block Size

The three scheduling algorithms namely FCFS-RR, Priority and hybrid are compared for throughput analysis in the heterogeneous environment. The throughput is calculated with respect to signal to noise ratio. As shown in simulation results, the proposed hybrid scheduling algorithm gives high throughput of 422Kbps as compared to FCFC-RR and Priority schemes which have throughput values of 145Kbps and 360Kbps.The average throughput is calculated by varying the block size ‘M’ as well as number of symbols per block ‘n’. The M is considered as 1024, 2048, 4096 and n is considered as 200 and 500. The variation of data rate represents the change in throughput for the three scheduling algorithms as shown in Fig. 6.11. A high throughput value is achieved by the hybrid scheduling for different values of M and n. The average throughput almost remains same for hybrid scheduling algorithm as compared to FCFS-RR and Priority scheduling.

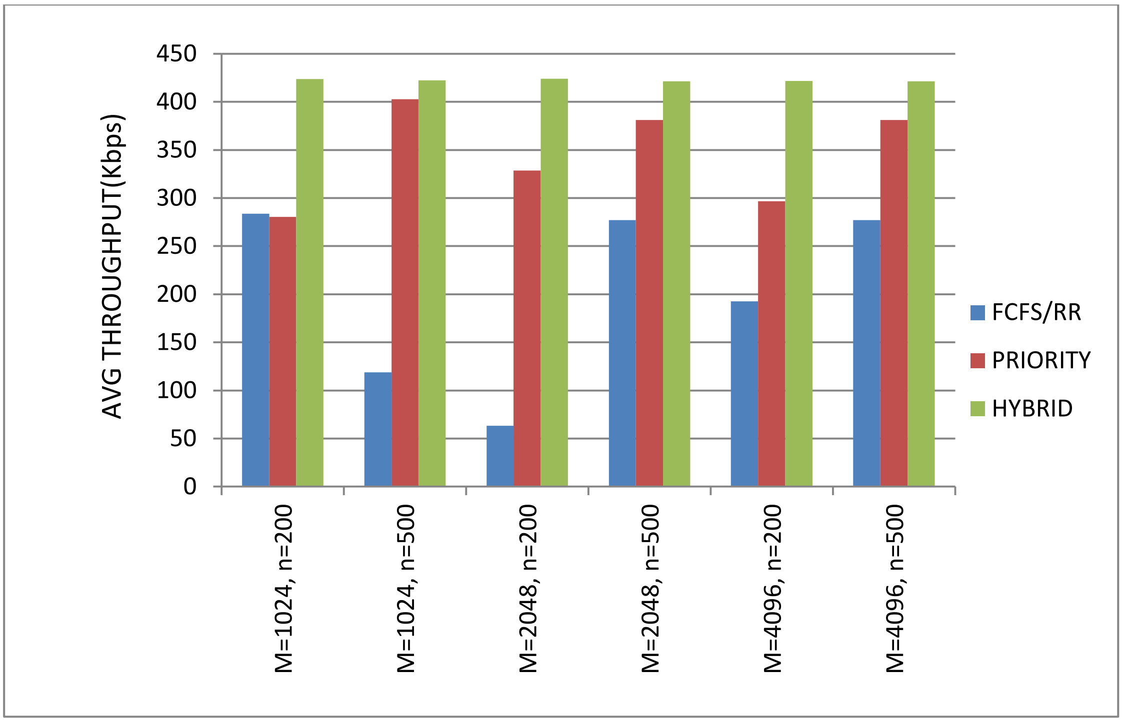

Figure 6.11: Average Throughput Comparison of FCFS-RR, Priority and Hybrid Scheduling

The scheduling algorithms implemented in heterogeneous environment provide better results in terms of BER but throughput is reduced as compared to LTE and WiMax networks. the results can further be enhanced to improve these QoS parameters as presented in next sections of this chapter.

6.4 Task Scheduling

In scheduling problems, some tasks are dependent on certain other tasks. So they are to be completed before starting the next task. In such a condition one has to decide the order in which these tasks are to be performed keeping in view their dependency. This way of ordering tasks is known as Topological Sorting. The increasing demand of tasks has resulted in the increase of Task Scheduling in Cloud. Cloud computing makes use of low power providers to provide high usage. So selection of nodes for performing a task in cloud is to be considered. In cloud, user applications are generally run on virtual systems/machines. Dynamic load balancing has to assign tasks dynamically to the processor. During the overloading of processors, the tasks are redistributed. All the applications are completely different from each other in nature and are also independent of each other. Some need more CPU time for computation of task whereas others may differ in the need of memory. The scheduling algorithm can increase the efficiency and can maintain the load balancing of a particular system. The efficiency of cloud is dependent on the algorithms which are used for the purpose of task scheduling. From the discussion carried out above we have to make a decision regarding each task as to whether it would be prudent to perform it on the device or to process it on the cloud and transmit the results rendered to the smart device to use energy in an optimum manner. Since each task requires execution/processing of certain instructions, generation and transmission of data hence we have to keep in mind the following factors for task scheduling:

- The tasks may be executed on the mobile device.
- Every task might have a rigid compression ratio of output and input data and each task might have different processing time.
- The network bandwidth available among the mobile device with the remote cloud.

The energy consumption in a smart device thus depend on the following two types of load:

- Computational and Processing task
- Communication and Data transmission cost

Thus we need to develop an algorithm to make wise choices in offloading so as to save energy and time at the same time. Clearly Energy optimization requires optimum choices between communication and computational tasks. A number of parallel processing algorithms are already being implemented for task scheduling. A hybrid task scheduling algorithm is implemented on heterogeneous network to enhance the QoS parameters environment considered as combination of WiMax and LTE.

6.5 WiMax–LTE Packet Scheduling using Parallel Structure

In the last couple of years, speed has become an essential part of the modern-day computation and communication framework. 3G and 4G networks have marked them valuable especially for worldwide communication at higher data rates. LTE and

WiMax both suffer from load management and congestion. Algorithms like FCFS, RR and priority scheduling are not sufficient enough to cope up with high demand network traffic [183-184]. The memory buffer of serving architecture is not so big that it can compensate for all the node requests at the same time. As the requests (R_1, ----------, R_n) are increasing, the scheduler is unable to manage the resources and providing services to all the users as shown in Fig. 6.12.

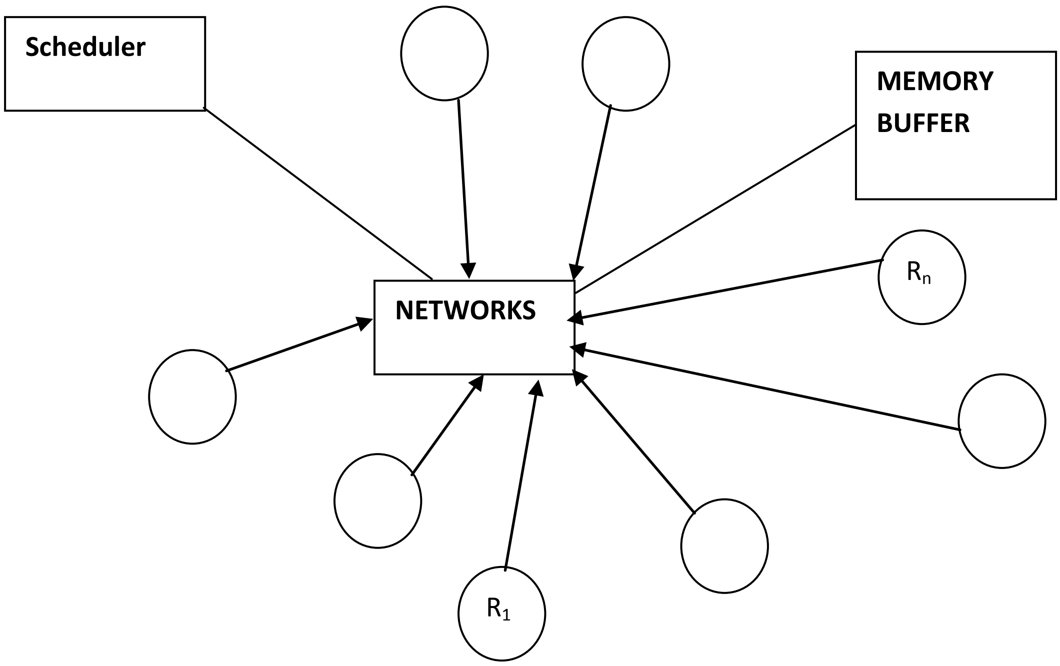

Figure 6.12: Requests from Nodes in a Network

6.5.1 Task Scheduling in LTE and WiMax

Scheduling is the procedure of ordering, directing and enhancing the work and workloads in a production process or manufacturing process. It is a significant means for manufacturing and engineering, where it can have a foremost effect on the output of a procedure. In scheduling, several diverse ways are employed to decide which particular job to run. Following are some of the parameters that can be considered in the process of job scheduling:

- Job priority
- Accessibility of computing resource
- License key if the job is using a licensed software
- Execution time allocated to the user
- Number of corresponding jobs allowed for a user
- Expected execution time
- Passed execution time
- The existence of peripheral devices
- Number of cases of specified actions

Scheduling has been implemented to organizing the jobs, procedures, and tasks into the utmost effective order, increasing output and utilization to meet the business necessities. Job scheduling is a major assignment of the operating system which distributes the system resources between the several tasks. The system upholds prioritized queues of jobs waiting for CPU time and must decide which job to take from which queue and how much time to allocate to it so that all jobs are completed in a fair and timely manner. Job scheduling tools manage a logical procedure (i.e. numerous jobs or programs) while they execute within a mainframe or distributed scenario, offering scheduling and dependency management of the procedure as it runs, primarily in series, over different systems, topographies, and applications. The tools in this category are often utilized for "batch integration" of heterogeneous applications and data provisions.

The architecture of LTE along with performance issues of packet scheduling using prioritization has been discussed in [183]. The evaluation has been made on the basis of throughput and packet delivery ratio. The author has discussed the problems of scheduling architecture in terms of high load. The same issues in WiMax framework have been discussed in [184]. The architecture of parallel processing in WiMax and LTE has been presented [185] as the solution. The algorithm of task scheduling utilizing a model of nonlinear programming for different task scheduling is proposed for different load and network bandwidth. The proposed scheduling function in the job scheduling mechanisms first creates a set of jobs and resources to generate the population by randomly assigning the jobs to the resources and evaluating the population by using fitness values that represent jobs execution time [186]. Second, function utilizes iterations to regenerate the population based on the behavior of the fireflies for producing the best job plan for providing the minimum execution time for the job.

Job scheduling primarily satisfies the business conditions of upholding data integrity and guaranteeing that infrastructure can counterattack and recuperate from mistakes and improved by concentrating on satisfying operational service levels for scheduled data processing, defending subtle output, and controlling and upholding the infrastructure. Job Scheduling can be carried out through a proper scheduling technique. The existing solutions do not provide efficient output for complex data architectures like WiMax and LTE.

6.5.2 The Proposed Solution

The proposed solution aims to reduce the load of the serving network is divided into two sections

a) Priority Decision
b) Scheduling

Consider a network with adjoining jobs. An adjoining job framework cannot execute child's job before a parent job. Figure 6.13 demonstrates a connected adjoined job model.

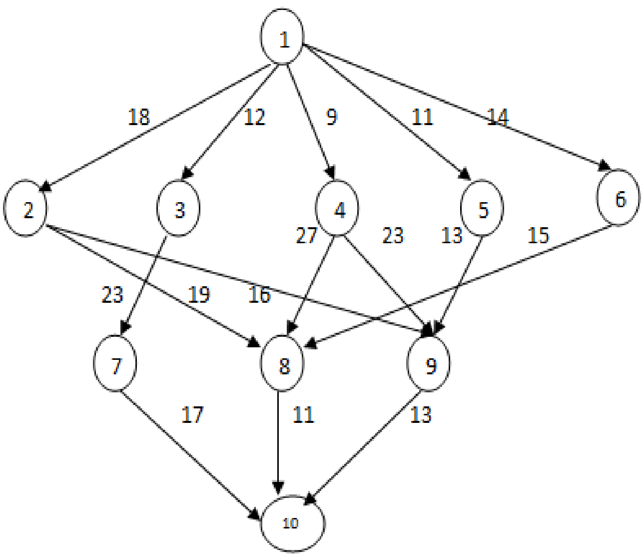

Figure 6.13: Adjoined job model

As shown in Fig. 6.13, 1 is the parent job which has children organized at level two. The proposed solution takes the following assumption for the processing:

- The proposed framework has 5 processing units.
- Each processing unit is powered by the same type of configuration.
- Each processing unit is capable of performing or scheduling any request.

The proposed solution applies the priority to the children. The nodes are hypothetically assumed in the following level frame.

1	Level 1
2,3,4,5,6	Level 2
7,8,9	Level 3
10	Level 4

As at level 1 and level 4, there is only 1 node, hence there would be no prioritization. The prioritization will be required at level 2 and level 3. The prioritization of nodes is done using the average computation cost (ACC) of each node at each processing unit [186].

ACC= Sum (All computations)/ Total no of computations

There are two types of costs for the processing of any request at any processing unit:

- Transfer Cost: Applicable when information is transferred from one processing unit to another processing unit. This cost is applicable if the parent request is executed on different processing unit but the child is to be executed on the different processing unit.
- Execution Cost: This cost is applicable when it is decided that the request will be executed on which processing unit. Each processing unit has its own execution cost for each request in the queue. The aim is to minimize the total cost.

Total Cost= Execution Cost + Transfer Cost

If the parents processing unit in same as that of the child processing unit then there would be no transfer cost for the executing request. High ACC results in high priority.

Once the priority of the nodes is decided, the scheduling takes place. Algorithm briefs about setting up the priority order of the nodes.

Algorithm: Priority (Nodes, Cost)

1. *ALL_ACC = [ ]; Level = [ ] // Initializes Average computation Cost and Level array as empty*
2. *Foreach node in Nodes // Considering each node in the node list*
3. *ALL_ACC = CalculateACC(); // comuting ACC*
4. *[node]*
5. *Level = find Level(Node); // Finding node levels*
6. *end*
7. *for i=1:Nodes*
8. *Level find Nodes at some Level;*
9. *Sort ALL_ACC[Level_X], descending] // Sorting the acc of each level in descending order*
10. *Set Priority (descending) // setting*
11. *End Of*

End Algorithm

The scheduling architecture is simple but classy at the same time. Consider job no 2 as it has one parent request 1. Request 1 will be proposed on that processing unit, which would result in at least time. Request 2 will not have to pay the transfer cost at request 1's parent.

Algorithm presented in next section demonstrate the structure of scheduling

Algorithm: Schedule (Priority, Transfer_Cost, Execution_Cost, Job_Graph)

1. *Foreach job in Priority // Taking each job in the priority list*
2. *Execution_order []; // Initializing the the execution order as empty*
3. *Job_id = Priority (Job, 1);*
4. *Read(Job_Graph) // Reading Job Graph*
5. *Job_Parents= find (Parent (Job_id));*
6. *If Job_parents.count==0*
7. *Execution_order [Job_id, 1]= Job_id;*
8. *Execution_order [Job_id, 2] = Execution_Cost, min_processing_units;*
9. *Execution_order [Job_id, 3] = Processing_Unit[MinSelected];*
10. *else*

11. *All_Cost [ ]; // Initializing all cost to empty*
12. *Foreach parent in Job_parent*
13. *Current_parent = Parent;*
14. *S= find (Execution List, Current parent);*
15. *P= find S.Execution_Unit;*
16. *Apply transfer_cost at each processing unit other than P.*
17. *All_Cost[Current, Parent] =* $Cost_{(transfer)} + Cost_{(execution)}$ *;*
18. *end for*
19. S_p *= find_min(All_Cost);*
20. *Allocate to min value execution unit;*
21. *end for*

End Algorithm;

The flow diagram of proposed task scheduling in Heterogeneous environment is represented in Fig. 6.14. The heterogeneous network is considered as the combination of WiMax and LTE. The hybrid scheduling is considered as the combination of directed acyclic graph (DAG) and critical path on a Processor (CPOP).

The scheduling algorithm takes priority of each job as input along with the transfer and the execution cost of each job. If the job has no parent, then it is executed on the processing unit which has least execution cost. If the job has multiple parents, then the proposed algorithm first computes all its related parents followed by the execution cost of parents on its corresponding execution unit. It finds out the minimum out of all execution units and then it finds the minimum value for positive execution.

The proposed research work collaborates the possibilities of the best outcome of the LTE scheduling and WiMax environment. The hybrid model utilizes both the networks at the same time in which the constraints of LTE and WiMax are satisfied altogether.

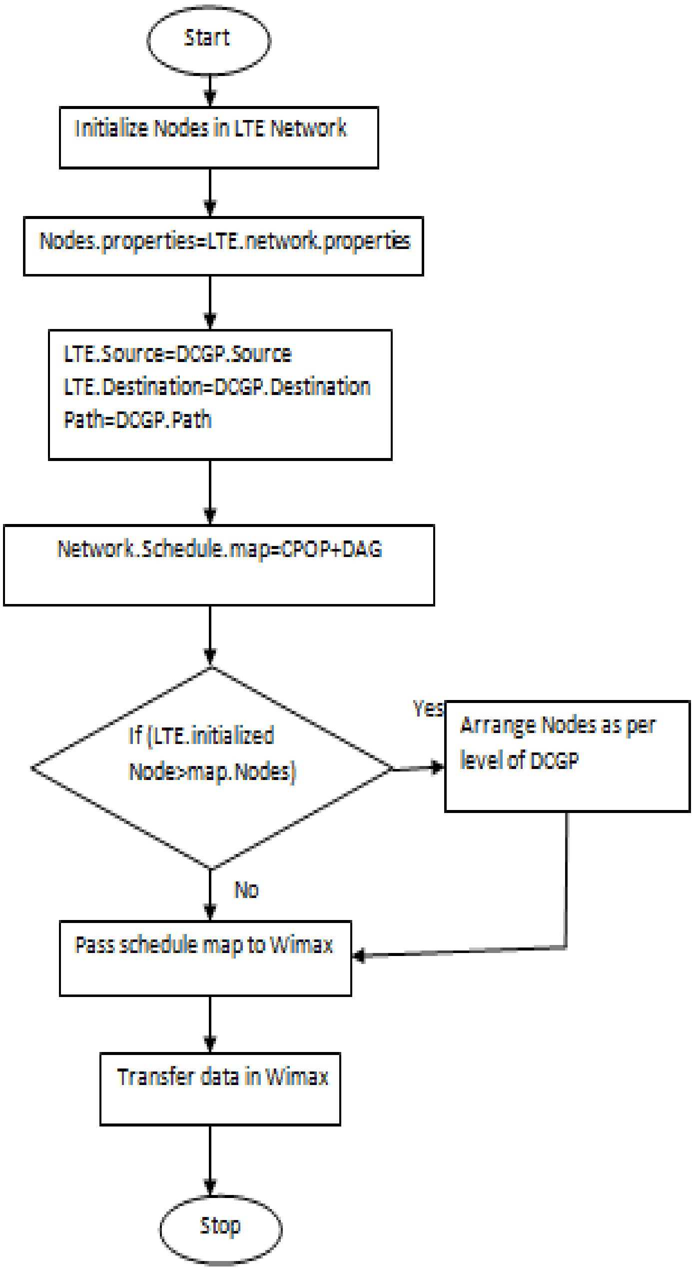

Figure 6.14: Flow Diagram of Proposed Task Scheduling in Heterogeneous Environment

6.5.3 Performance Evaluation of Proposed Task Scheduling Algorithm

The task scheduling algorithm is designed for heterogeneous processor using critical path algorithm for scheduling task graphs. The performance evaluation of task scheduling algorithm is done on the basis of various QoS parameters as throughput, packet delivery ratio (PDR) and energy consumed. The simulation is carried out in heterogeneous environment considered as combination of LTE and WiMax networks.

6.5.3.1 Energy Consumed using task scheduling in Heterogeneous Environment

The energy consumed during task scheduling depends upon the total energy supplied and consumed by the nodes in fixed time duration. The simulation time considered for requests generated by different nodes is 100sec. As the number of requests is increasing, the energy consumption in the network will increase. Fig 6.15 represents the energy consumption of 60mJ when the time allocated for 60 requests completion is 100sec.

As the simulation is run for 150sec, the 100 requests are completed in prescribed time consuming energy of not more than 100mJ as shown in Fig. 6.15. Compared to the results for hybrid scheduling algorithm, a large amount of energy is saved while using the proposed solution of task scheduling for heterogeneous network.

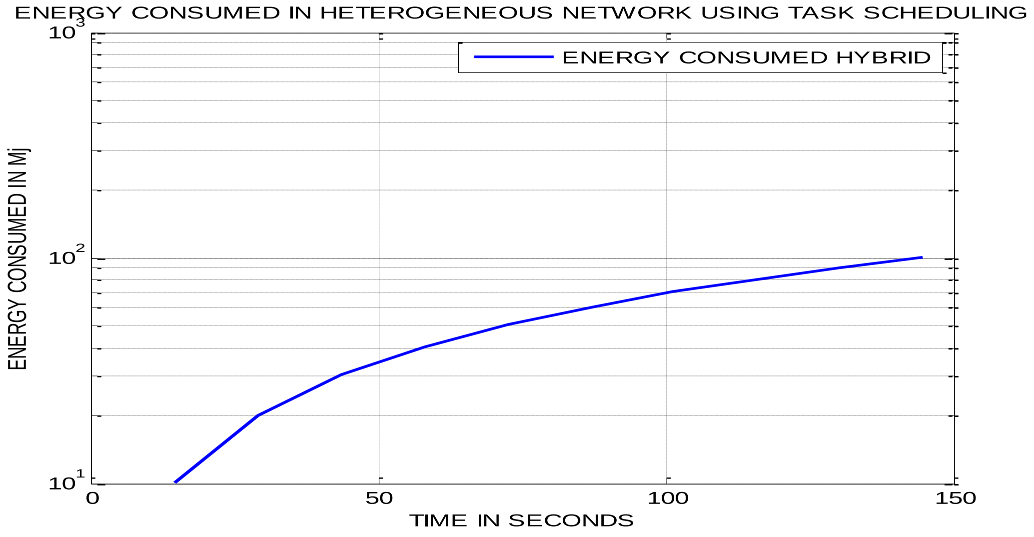

Figure 6.15: Energy Consumed in Heterogeneous Network using Task Scheduling

Fig. 6.16 presents the number of requests sent and completed in heterogeneous environment using task scheduling. As compared to hybrid algorithm implemented without task scheduling, the number of requests completed is increased to almost 100%.

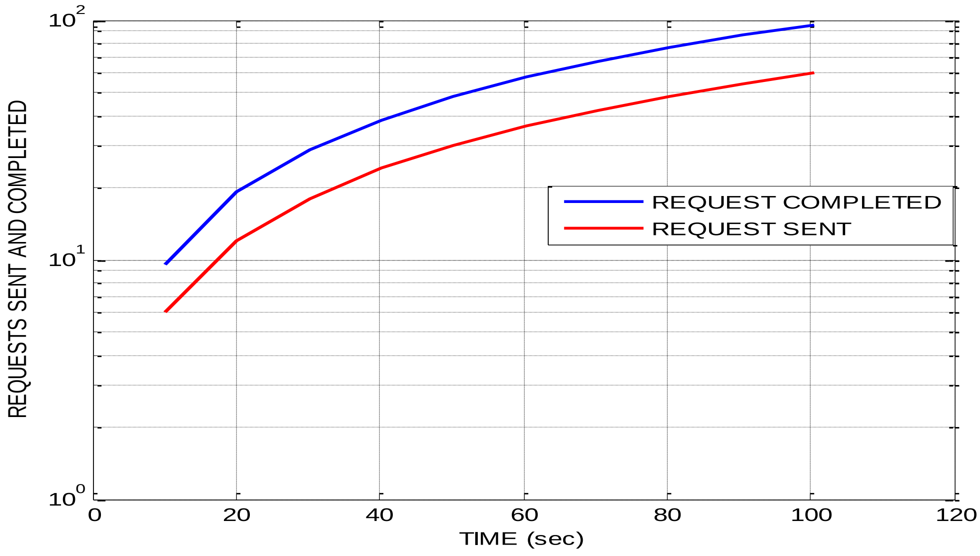

Figure 6.16: Requests Sent and Completed using Task Scheduling

6.5.3.2 Throughput for Heterogeneous Network Using Task Scheduling

A total of 50 different sets are executed for a different set of the acyclic structure. The evaluation is made on the base of the following parameters.

- Throughput: Total number of delivered packets per time frame
- Packet Delivery Ratio (PDR) : Total number of transferred packets per packet frame.

Fig. 6.17 shows the throughput in a heterogeneous network using task scheduling algorithm. The throughput is increasing with respect to number of iterations. The average throughput is calculated by generating different number of jobs and results are compared with WiMax and LTE separately as shown in Table 6.3 and Fig. 6.18. The throughput is going to be increased as the number of jobs is increasing. The hybrid scheduling algorithm in heterogeneous environment performs better as compared to WiMax and LTE networks.

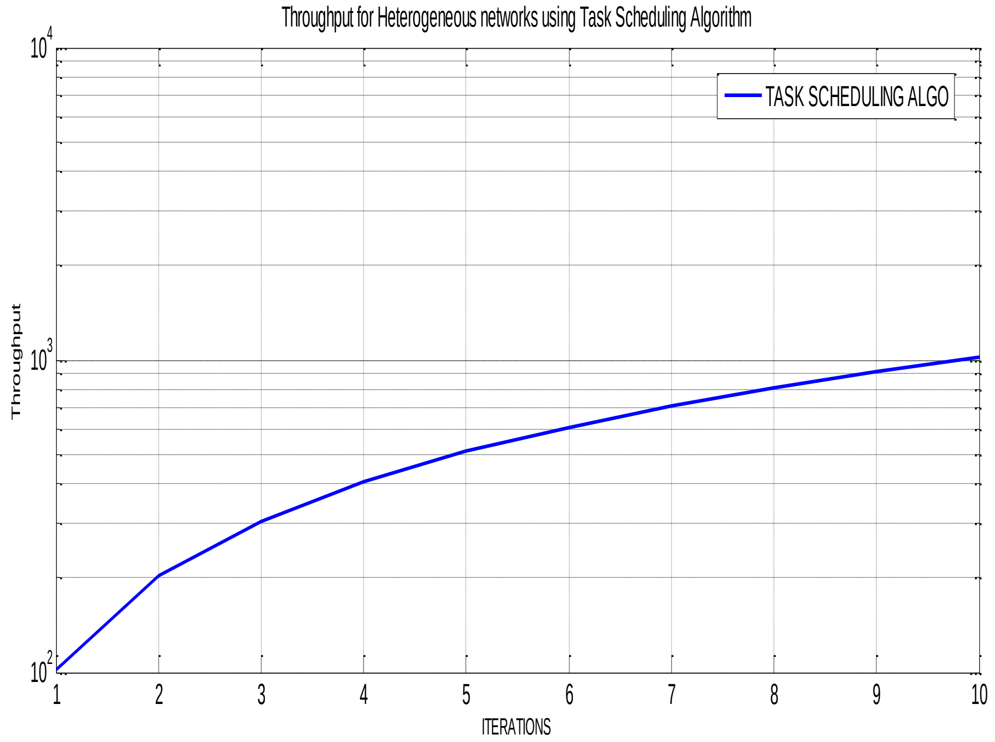

Figure 6.17: Throughput for Heterogeneous Network Using Task Scheduling

Table 6.3: Average Throughput

TOTAL JOBS	AVG THROUGHPUT WIMAX	AVG THROUGHPUT LTE	AVG THROUGHPUT HYBRID
10	125000	119585	131020
20	126879	138245	144562
30	98125	97845	102541
40	135648	132456	141231
50	135654	135458	140024

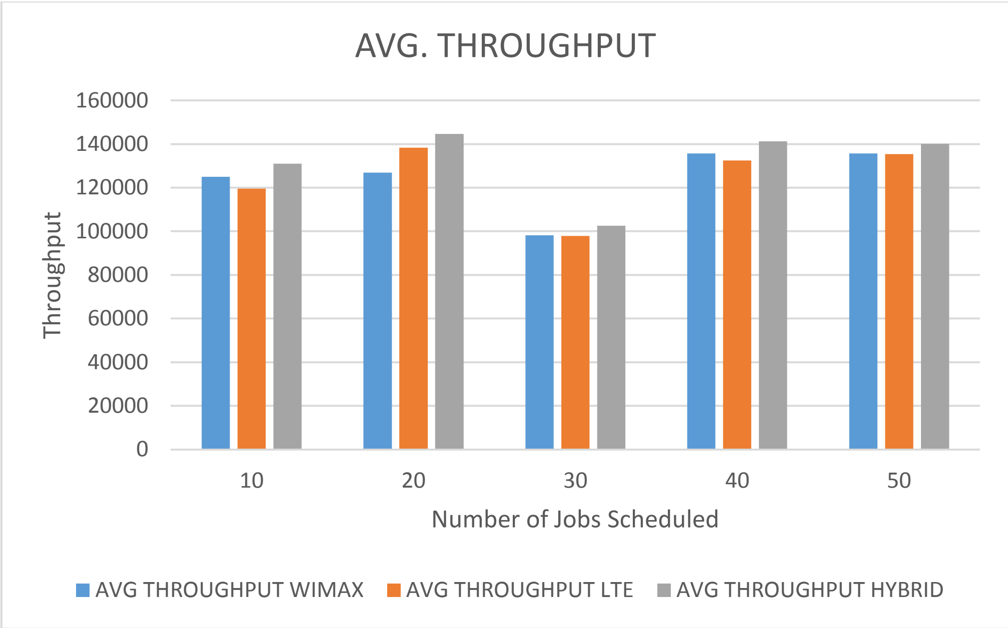

Figure 6.18: Throughput in Heterogeneous Network for different job models

6.5.3.3 PDR for Heterogeneous Network using Task Scheduling

The PDR is calculated for heterogeneous network using task scheduling algorithm as shown in Fig. 6.19. The PDR is increasing as the number of iterations is increased.

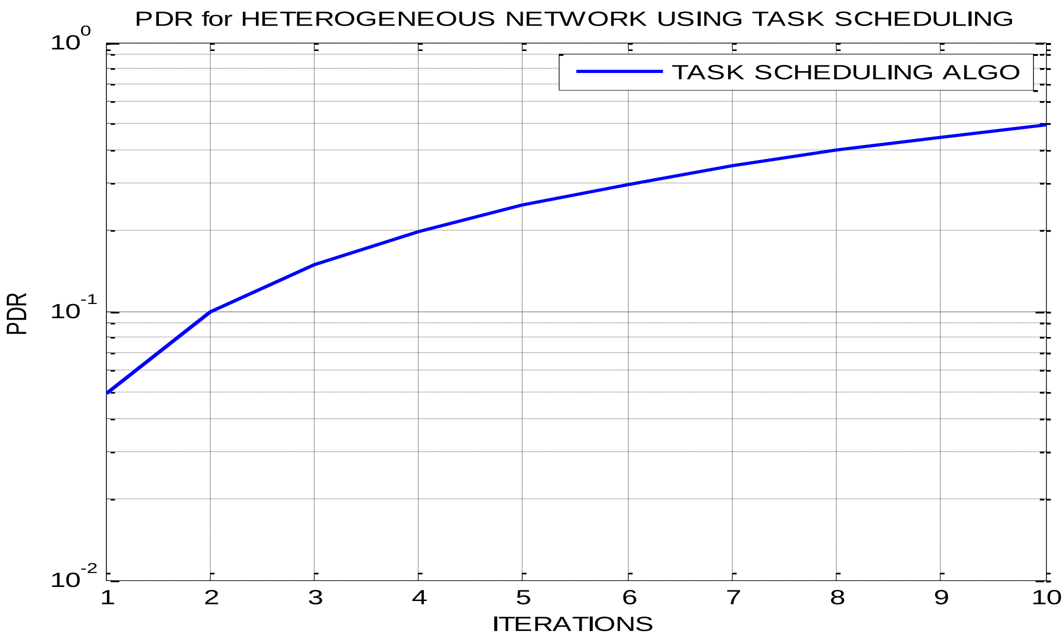

Figure 6.19: PDR for Heterogeneous Network using Task Scheduling

Table 6.4 and Fig. 6.20 represent the average PDR value for different job models. It is monitored that the average PDR value lies between 0.74 to 0.78. It is a high delivery rate as the architecture of scheduling is well designed in layers. The average value of PDR for WiMax is more as compared to the LTE. This is due to the complex architecture of LTE as compared to WiMax. The results show that PDR in proposed hybrid scheduling algorithm is better as compared to both LTE and WiMax. The advantage of the proposed work is that it does not leave the jobs in the queue due to its parent-child solution architecture.

Table 6.4: Average PDR

TOTAL JOB MODELS	AVG PDR WIMAX	AVG PDR LTE	AVG PDR HYBRID
10	0.76	0.71	0.78
20	0.72	0.74	0.76
30	0.76	0.77	0.79
40	0.74	0.76	0.78
50	0.78	0.73	0.79

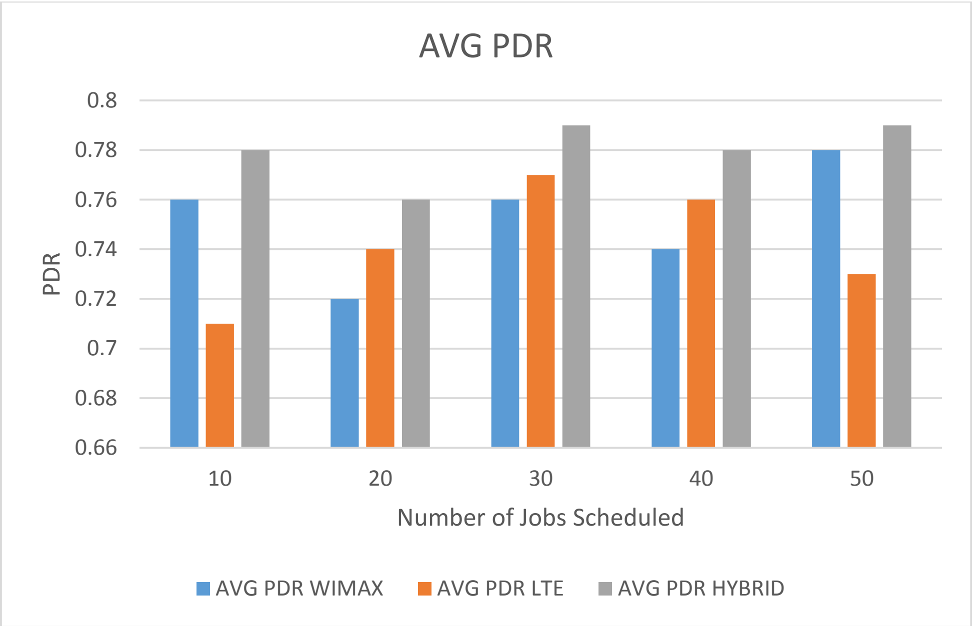

Figure 6.20: Average PDR for Different Jobs Scheduled

Throughput is much similar to PDR and also they are directly proportional to each other. If PDR increases then throughput also increases.

6.6 Summary

The network performance is affected by scheduling algorithm and has a direct impact on jobs requested and completed successfully. In this chapter, a hybrid scheduling algorithm is proposed for heterogeneous networks. The heterogeneous network is comprised of overlapped LTE and WiMax network implemented in same environment by considering common parameters for both networks. The three scheduling algorithms are implemented separately and in hybrid manner over the simulation environment and the performance of these algorithms is compared. The obtained simulation results demonstrated that the proposed algorithm can select the best available network in heterogeneous environment based on jobs scheduled, load balancing and throughput and error rate is improved.

The proposed methodology is further enhanced by implementing parallel scheduling architecture to overcome the problem of overloading in WiMax and LTE network. It presented a parent-child concept which reduces the waiting queue length. The evaluation is made on the base of throughput and PDR. Due to its simple yet classic nature, it reduces the packet drop. As the child job can't be executed without the parent job, hence the child job is not initiated till the parent job is not finished. This reduces the packet drop as well and enhances the PDR. The proposed work has used the hybrid environment which is a combination of LTE and WIMAX. As the hybrid network has capabilities of both the networks hence the outcome of hybrid environment is better as compared to WiMax and LTE individually.

Chapter 7

CONCLUSION AND FUTURE SCOPE

7.1 Summary

As NGNs are in a stage of constant development, the scenario is moving towards providing the heterogeneity among various wired and wireless networks. It has been observed that combination of different technologies will produce a network with competent features to satisfy upcoming market trends. To satisfy growing demands of users, the suitable choice of technology supported by MIMO/OFDM is either WiMax or LTE or the combined approach for heterogeneous networks for faster, reliable and cost effective connections. QoS in wireless networks is also an important paradigm to attract the attention of network providers in telecommunication world. An effective combination of the two technologies is able to complement each other and provide high QoS. Scheduling algorithms play an important role in wireless network planning and optimization. The scheduling schemes in heterogeneous networks are based upon the network capacity, bandwidth allocation and resource allocation methods. The underlying idea of thesis was to study, design and simulate various scheduling algorithms for heterogeneous wireless communication networks and find out solutions to maintain heterogeneity by designing a proposed hybrid scheduling algorithm as a combination of already existing algorithms.

The thesis work is started with the study of queuing theory which can solve a number of problems regarding service flows in queues and how to schedule the flows with QoS constrains. The queuing models can be designed depending upon job size distribution. The models are able to study and enhance the performance of a network in terms of error rate reduction and throughput maximization. The recursive formulas for queue length distribution are analyzed to generate higher moments of queue length distribution. In this thesis, queue length moment distribution of various queues is discussed. Analysis of moments of queue length in M/M/1, M/M/1/B, M/M/c, M/Er/1 and Er/M/1 queue models is represented. Also the recursive formula of M/M[K]/1 queue is derived and compared with other queue models. The bulk queuing model is derived in comparison with single queuing model and utilization factor is compared with queue length moments for various queuing models. The greater is the utilization factor, the better is the system performance and the system QoS can be described as optimal.

We have proposed a WiMax network model for BER evaluation of the real time data by comparing different modulation and channel encoding schemes. The parameters are chosen based upon the standard for transmitter and receiver section physical layer and performance evaluation is done. WiMax systems is one of the major analysis platform for high speed data transfer but when the load increases over the nodes there must be a suitable mechanism which can handle the number of nodes and its architecture system. The WiMax network is designed by initializing number of nodes in a network. Then the tasks are generated by the various nodes by initializing the source and destination nodes. The job model for the proposed efficient algorithm consists of the hybrid combination of FCFS, RR and priority scheduling. The hybrid algorithm shows a vital increase in the performance of the jobs completed and error rate reduction as well as the throughput is enhanced comparatively. As the SNR is increased, the SER is reduced irrespective of packet size and number of packets transmitted.

In addition to above, the load balancing structure in LTE networks is discussed as the procedure of distributing the information or packets among available base stations to manage more requests or calls so that the entire system can accomplish the jobs more efficiently. The number of jobs completed has been reached almost 70%. The energy consumed in LTE network is lesser as compared to WiMax network. A job scheduling algorithm is designed based upon load balancing. The hybrid structure including job scheduling is designed by combining more than one number of scheduling algorithms to achieve best of QoS as well as energy efficiency during load balancing. The three scenarios have been considered for LTE network and the scheduling algorithms namely FCFS/RR, Priority and proposed hybrid algorithm are compared. The results are analysed and compared for BER, SER and throughput for different values of packets and by varying the size of packets. Hybrid algorithm performs better as compared to FCFS/RR and priority algorithm as represented in results.

We further proposed heterogeneous network overlapping LTE and WiMax implemented in same environment by considering common parameters for both networks and a hybrid scheduling algorithm is proposed as a combination of FCFS, RR and priority scheduling. The three scheduling algorithms namely FCFS/RR, priority and hybrid have been implemented on the heterogeneous network and the performance of these algorithms is compared. The obtained simulation results demonstrated that the proposed algorithm can select the best available network in heterogeneous environment based on jobs scheduled, load balancing and throughput and error rate is improved. As compared to WiMax and LTE networks, the number of jobs completed is reduced and energy consumption is increased in the heterogeneous environment. Hence a more advanced multilevel scheduling architecture is proposed as a solution to the prevailing problem in heterogeneous environment. Multilevel scheduling algorithms are required as heterogeneous computation platforms for parallel computing applications are quite popular topic of research these days. They offer high computation power and easy to access in different scenarios for heterogeneous networks. Task scheduling with parallel processing is one of the leading approaches towards modifying traditional parallel algorithms. It presented a parent-child concept which reduces the waiting queue length. The evaluation is made on the basis of throughput and PDR. Due to its simple yet classic nature, it reduces the packet drop. This reduces the packet drop as well and enhances the PDR. The proposed work has been implemented in the hybrid environment which is a combination of LTE and WiMax. As the hybrid network has the capabilities of both the networks hence the outcome of hybrid environment is better as compared to WiMax and LTE individually.

7.2 Future Work

The proposed solution has a lot of future possibilities. Some of the future directions related to QoS based traffic scheduling in heterogeneous environment are directed as follows:

- Hybridization of machine learning and swarm intelligence optimization methods may be implemented as future work for the heterogeneous networks. Energy optimization along with quality scheduling could have been a critical parameter to find. There are various aspects of energy optimization which requires a good hand in this area. It also aims to enhance the connectivity, scalability and energy efficiency of the future networks.

- The dynamic scheduling algorithm for parallel processing is rapidly developing research area these days. Many dynamic scheduling schemes are discussed to design and achieve the goals such as minimum resource idleness, maximum resource utilization, decreasing make span time, compilation time and reducing task delays. Considering other QoS parameters like jitter, latency analysis for the same is proposed to be as future work for this research.
- The two main drivers for 5G are mobile internet and internet of things (IoT) which are expected to be the upcoming deployments available by 2020. The networked society in near future provides wireless connectivity to all kind of devices and services. The work can be enhanced to prevent the security threats in LTE and WiMax networks and to gain the attention of artificial intelligence in this area.

REFERENCES

[1] Ahuja, K., Singh, B., Khanna, R., “Network Selection Based on Weight Estimation of QoS Parameters in Heterogeneous Wireless Multimedia Networks”, *Wireless Personal Communications*, Vol. 77, Issue 4, pp. 3027-3040, April 2014.

[2] Xiang, W., Zheng, K., Shen, X., “5G Mobile Communications”, *Editors Springer International Publishing*, 2017.

[3] Chen, H., “Scheduling and Resource Optimization in Next Generation Heterogeneous Wireless Networks”, *A thesis submitted to the University of Louisiana at Lafayette*, 2008.

[4] Gustafsson, E., Jonsson, A., “Always Best Connected”, *IEEE Wireless Communication*, Vol. 10, Issue 1, pp. 49–55, Feb 2003.

[5] Alasti, M., Neekzad, B., Hui, J., Vannithamby, R., “Quality of Service in WiMax and LTE Networks”, *IEEE communications Magazine*, Vol. 48, Issue 5, pp. 104-111, May 2010.

[6] Bhagat, K., Malhotra, J., "A Survey of Uplink Multiple Access Techniques in LTE Mobile Communication System”, in proceedings of *IEEE International Conference on Advances in Engineering & Technology Research*, pp. 1-4, August 2014.

[7] Jain, V. K., Franz, J.H., “Some Design Issues in Infrared Wireless Networks”, *IETE Technical Review*, Vol. 21, Issue 5, pp. 335-343, Mar 2015.

[8] Reka, R., Wahidabanu, R., “A Survey on Uplink Bandwidth Request in WiMax Networks” *International Journal of Emerging Trends & Technology in Computer Science,* Vol. 1, Issue 2, pp. 280-284, August 2012.

[9] Al-Howaide, A., Doulat, A., Yaser, M., Khamayseh, “Performance Evaluation of Different Scheduling Algorithms in WiMax”, *International Journal of Computer Science, Engineering and Applications,* Vol.1, Issue 5, pp. 81-94, October 2011.

[10] Kishor, L., Goyal, D., "Time Quantum Based Improved Scheduling Algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering,* Vol. 3, Issue 4, pp. 955-962, April 2013.

[11] WiMax Forum, WiMax and the IEEE 802.16m Air Interface Standard- April 2010, published by Intel.

[12] IEEE Std 802.16-2004, Part 16: Air Interface for Fixed Broadband Wireless Access Systems. *Technical report on* IEEE Standard for Local and Metropolitan Area Networks, published by IEEE, October 2004.

[13] IEEE Std 802.16e/D9. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands. *Technical report on* IEEE Standard for Local and metropolitan area networks, June 2005.

[14] Zorba, N., and Perez-Neira, A.I., "CAC for Multibeam Opportunistic Schemes in Heterogeneous WiMax Systems under QoS Constraints," in Proceedings of *IEEE Global Telecommunication Conference,* pp.1-5, November 2007.

[15] Sonia, Khanna, R., Kumar, N., "Analysis of Bulk Arrival Queuing Model for Quality of Service based Traffic Scheduling" in Proceedings of *IEEE Nirma University International Conference on Engineering*, pp. 1-5, December 2012.

[16] Chang, S. H., Choi, D. W., "Performance analysis of a finite-buffer discrete-time queue with bulk arrival, bulk service and vacations", *Computers & Operations Research*, Vol. 32, Issue 9, pp. 2213–2234, September 2005.

[17] Jin, X., Min, G., Velentzas, S., Jiang, J., "Quality-of-Service Analysis of Queuing Systems with Long-Range-Dependent Network Traffic and Variable Service Capacity", *IEEE Transactions on Wireless Communications*, Vol. 11, Issue 2, pp. 562-570, FEBRUARY 2012.

[18] Fedorcakova, M., Sebo, J., "Application of the queuing theory model with stochastic creation of servicing channels", *International Journal of Engineering*, Vol. 1, pp. 115-120, 2011.

[19] Murthy, T.S.R., Krishna, D.S.R., Raju, G.V.S., "M / M (k) /1 Queuing model with varying bulk service", *International Journal of Mathematics and Soft Computing*, Vol. 2, Issue 1, pp. 109 – 117, 2012.

[20] Myers, D.S., Vernon, M.K., "Estimating Queue Length Distributions for Queues with Random Arrivals", *ACM Sigmetrics Performance Evaluation Review*, Vol. 40, Issue 3, pp. 77-79, December 2012.

[21] Takine, T., "Queue length distribution in a FIFO single-server queue with multiple arrival streams having different service time distributions", *Queuing Systems*, Vol. 39, Issue 4, pp. 349-375, December 2001.

[22] Horvath, A., Horvath, G., Telek, M., "A Joint Moment Based Analysis of Networks of MAP/MAP/1 Queues", *Performance Evaluation*, Vol. 67, Issue 9, pp. 759-778, September 2010.

[23] Manjunath, D., Sikdar, B., "Input Queued Switches for Variable Length Packets: Analysis for Poisson and Self-similar Traffic", *Computer Communications*, Vol.25, Issue 6, pp. 590-610, 2002.

[24] Casale, G., "Exact analysis of performance models by the Method of Moments", *Performance Evaluation*, Vol. 68, Issue 6, pp. 487-506, June 2011.

[25] Casale, G., "The Multi-Branched Method of Moments for Queueing Networks", in Proceedings of IEEE *International Conference on Quantitative Evaluation of Systems*, pp. 227–236, February 2009.

[26] Kaur, *P.*, Khosla, A., Uddin, M., "Markovian Queuing Model for Dynamic Spectrum Allocation in Centralized Architecture for Cognitive Radios", *International Journal of Engineering and Technology*, Vol. 3, Issue 1, February 2011.

[27] McKay, M.R., Smith, P.J., Suraweera, H.A., Collings, I.B., "On the Mutual Information Distribution of OFDM-based Spatial Multiplexing: Exact variance and outage approximation," *IEEE Transactions on Information Theory*, Vol. 54, Issue 7, pp. 3260–3278, 2008.

[28] Musavian, L., Aissa, S., "Capacity and Power Allocation for Spectrum Sharing Communications in Fading Channels," *IEEE Transactions on Wireless Communications*, Vol. 8, Issue 1, pp. 148–156, 2009.

[29] Abbas, M. J., Kumar, M., Jaiswal, A. K., "Enhancement of the Channel Capacity using MIMO-OFDM Technology for Wireless Transmission System", *International Journal of Electronics, Communication &*

Instrumentation Engineering Research and Development, Vol.4, Issue 3, pp. 1-8, June 2014.

[30] Li, Y., "Simplified Channel Estimation for OFDM systems With Multiple Transmit Antennas*", IEEE Transactions on Wireless Communications*, Vol.1, No.1, Page 67-75, January 2002.

[31] Li, Y.G., Winters, J.H., Sollenberger, N.R., "MIMO-OFDM for Wireless Communications: Signal Detection with Enhanced Channel Estimation", *IEEE Transactions on Communications*, Vol. 50, Issue 9, pp. 1471-1477, September 2002.

[32] Sinha, N.B., Nandy, M., Kapri, K.K., Bera, R., Mitra, M., "Channel Estimation and Performance Enhancement of Hybrid Technology for Next Generation Communication System", *International Journal of Research and Reviews in Applied Sciences,* Vol. 1, Issue 2, pp. 127-140, November 2009.

[33] Thomas, G., "OFDM Capacity Enhancement by Selective Channel Use", in Proceedings of *IEEE Military Communications Conference, October 2009.*

[34] Alen, T.C.H., Madhukumar, A.S., Chin, F., 2003. "Capacity enhancement of a multi-user OFDM system using Dynamic Frequency Allocation", *IEEE Transactions on Broadcasting*, Vol. 49, Issue 4, pp. 344-353, December 2003.

[35] Shamganth, K., Reena, M.P., "Capacity Enhancement in WLAN using MIMO", in Proceedings of *International Conference ACCTA*, Vol. 1, Issue 2, 3, 4, pp. 319-324, August 2010.

[36] Gautam, A., and Sharma, M., "Capacity Enhancement of MIMO-OFDM system in Rayleigh Fading Channel", *IOSR Journal of Electronics and Communication Engineering,* Vol. 6, Issue 1, Page 49-58, June 2013.

[37] Ameta, L., and Porwal, M.K., "Channel Capacity Estimation in MIMO-OFDM system for Different Fading Channels using Water filling Algorithm", *International Journal of Digital Application & Contemporary research,* Vol. 3, Issue 3, pp. 1-5, October 2014.

[38] Shoba, B., and Jayanthi, K., "Performance Improvement of MIMO OFDM Systems through Channel Estimation*", International Journal of Wireless & Mobile Networks,* Vol.4, Issue 5, pp. 49-59, October 2012.

[39] Elavarasan, P., Nagarajan, G., Narayanan, A., "PAPR reduction in MIMO-OFDM systems using Joint channel estimation and Pre-coding", in Proceedings of *International Conference on Advanced Communication Control and Computing Technologies*, pp. 327-331, August 2012.

[40] Yaacoub, E., and Dawy, Z., "A Survey on Uplink Resource Allocation in OFDMA Wireless Networks", *IEEE Communications Surveys and Tutorials*, Vol. 14, Issue 2, Page 322-337, 2012.

[41] Jain V.K., Franz, J.H., "Some Design Issues in Infrared Wireless Networks", *IETE Technical Review on Broadband Wireless Technologies*, Vol. 21, Issue 5, pp 335-343, March 2015.

[42] Gopal, P., Jain, V.K., Kar, S., "Performance Enhancement of Free Space Optical Satellite Uplink with Transmitter Spatial Diversity Technique", *Optical and Quantum Electronics*, pp. 1-14, August 2016.

[43] Heath, R.W., Paulraj, A.J.,"Switching Between Diversity and Multiplexing in MIMO Systems", *IEEE Transactions on Communications*, Vol. 53, Issue 6, pp. 962-968, June 2005.

[44] Payidar, M., Dubey, R., Jain, N.K., Kulpariya, S., "Performance analysis of WiMax 802.16e physical layer model", in Proceedings of IEEE *Ninth International Conference on Wireless and Optical Communications Networks, pp 1-4, October 2012.*

[45] Omerovic, S., "WiMax Overview", Technical Report, *University of Ljubljana, Slovenia, pp. 1-35, 2014.*

[46] Verikoukis, C.V., "An adaptive hidden Markov model for indoor OFDM based wireless systems", *Journal of Telecommunications and Information Technology*, Vol. 2, pp. 61-65, June 2004.

[47] Sacchi, C., Scorza, G.B., Granelli, F., De Natale, F.G.B., "A Joint MAC-PHY Approach for Medium Access Control in VBR MC-CDMA Broadband Indoor Connections", *Wireless Personal Communications*, Vol. 36, Issue 4, pp. 403–424, March 2006.

[48] Sacchi, C., Panizza, M., "Multi-rate group-orthogonal OFDMA-CDMA for broadband mobile transmission", *Telecommunication Systems*, Vol.52, Issue 51, pp. 15–29, January 2013.

[49] Riizwan, M., Ibrahiim, M., "Efficient Architecture for QOS Proviidence across WiiMAX" *Computer Science Series,* Vol. 9, Issue 1, pp. 203-214, 2011.

[50] So-In, C., Jain, R., Tamimi, A.K., "Scheduling in IEEE 802.16e Mobile WiMax Networks: Key Issues and a Survey", *IEEE Journal on Selected Areas in Communications,* Vol. 27, Issue 2, pp. 156-171, February 2009.

[51] Hawa, M., Petr, D.W., "*Quality of Service Scheduling in Cable and Broadband Wireless Access* Systems", in Proceedings of *IEEE International Workshop on Quality of Service***,** pp. 247-255, May 2002.

[52] Chen, J., Jiao, W., Guo, Q., "An Integrated QoS Control Architecture for IEEE 802.16 Broadband Wireless Access Systems", in Proceedings of *IEEE Global Telecommunications Conference*, pp. 3330-3335, December 2005.

[53] Roy, J.J.J., Vaidehi, V., Srikanth, S., "*Always best-connected QoS integration model for the WLAN, WiMax Heterogeneous Network*," in Proceedings of *IEEE First International Conference on Industrial and Information systems,* pp. 361-366, August 2006.

[54] Liu, Q., Zhou, S., Giannakis, G.B., "A Cross-layer scheduling Algorithm with QoS support in wireless networks," *IEEE Transactions on vehicular Technology*, Vol. 55, Issue 3, pp. 839-847, May 2006.

[55] Nie, C., Venkatachalam, M., Yang, X., "Adaptive Polling Service for Next Generation IEEE 802.16 Wireless netwoks", in Proceedings of *IEEE Global Telecommunications Conference*, pp. 4754-4758, November 2007.

[56] Grewal, V., Sharma, A.K., "On Performance Evaluation of Different QoS Mechanisms and AMC scheme for an IEEE 802.16 based WiMax Network", *International Journal of Computer Applications, Vol. 6, Issue 7, pp. 12-17, September 2010.*

[57] Tung, H.Y., Tsang, K.F., Lee, L.T., Ko, K.T., "QoS for Mobile WiMax Networks: Call admission Control & Bandwidth allocation", *5th IEEE Consumer Communications and Networking Conference,* pp. 576-580, January 2008.

[58] Mokdad, L., Othman, J.B., Cheikh, M.O., "Improving QoS for BE Traffics in WiMax Networks", in Proceedings of *IEEE International Conference on Computer Systems and Applications*, pp. 1-5, May 2010.

[59] Othman, J.B., Mokdad, L., Ghazal, S., "Performance Analysis of WiMax Networks AC", *Wireless Personal Communications*, Vol. 74, Issue 1, pp. 133-146, January 2014.

[60] Park, E.C., Kim, H., "Dual Feedback Approach of Request Based Bandwidth Allocation For Real-Time Service in Broadband Wireless Access Networks", *IEEE Transactions on Communications,* Vol. 57, issue 3, pp. 610-613, March 2009.

[61] Xie, L., Xiang, J., Zhang, Y., Zhang, J., "Joint bandwidth Reservation and Admission Control in IEEE 802.16e based Networks", in Proceedings of *IEEE International Conference on Communications*, pp. 1-6, 2009.

[62] Prasath, G.A., Raghu, K.R., Ma, M., "Integration of WLAN and WiMax with Base Station Assisted QoS", in Proceedings of *IEEE 5th IFIP International Conference on Wireless and Optical Communications Networks*, pp. 1-5, May 2008.

[63] Gidlund, M., Wang, G., "Uplink Scheduling Algorithms for QoS Support in Broadband Wireless Access Networks" *Journal of Communications*, Vol. 4, Issue 2, pp. 133-142, March 2009.

[64] Mehta, V., Gupta, N., "Performance Analysis of QoS Parameters for WiMax Networks". *International Journal of Engineering and Innovative Technology, Vol. 1, Issue 5, pp. 105-110, May 2012*.

[65] Nie, W., Wang, H., Park, J.H., "Packet Scheduling with QoS and Fairness for Downlink Traffic in WiMax Networks", *Journal of Information Processing Systems, Vol.7, Issue 2, pp. 261-270, June 2011*.

[66] Ali, D.M., Dimyati, K., "Performance Analysis of Delay Jitter in Mobile WiMax Systems", in Proceedings of *International Conference on Information and Electronics Engineering, Vol. 6, pp. 1-5, 2011.*

[67] Haseeb, A., Tralli, V., 2011."QoS performance analysis for VoIP traffic in heterogeneous networks with WiMax access", in Proceedings of IEEE *13th*

International Conference on Advanced Communication Technology, pp. 960-965, February 2011.

[68] Chandra, Prabha, H., "Enhancement of Quality of Service Parameters in WiMax Mobile Networks", *American Journal of Applied Sciences, Vol 9, Issue 12, pp. 1906-1915, 2012.*

[69] Mohamad, N.H.A.B., "Performance Analysis of Video Transmission Over IEEE 802.16 Architecture", *Project Report, Universiti Teknikal Malaysia Melaka*, April 2010.

[70] Chowdhury, A., Boyer, E., "Analysis of Quality of Service (QoS) for Video Conferencing Over WiMax Networks", Final Project Presentation on *ENSC 427: Communication Networks, 2010.*

[71] Bestetti, A., Giambene, G., Hadzic, S., "Fair traffic scheduling for WiMax systems", in Proceedings of IEEE *6th International Symposium on Wireless Communication Systems, pp. 254-258, September 2009.*

[72] Kumar, K.R., Raj, M., Balakrishnan, K., Das, D., "WiMax emulation test bed to benchmark streaming multimedia QoS", in Proceedings of 3rd International Conference on *Internet Multimedia Services, Architecture and Applications, pp. 193-198, December 2009.*

[73] Talwalkar, R.A., Llyas, M., "Analysis of Quality of Service (QoS) in WiMax Networks", in Proceedings of *16th IEEE International Conference on Networks, pp. 1-8, December 2008.*

[74] Sai, S.N., Schaar, M.V., "Performance Analysis of Video Transmission over IEEE 802.11a/e WLANs", *IEEE Transactions on Vehicular Technology, Vol. 56, Issue 4, pp. 2346-2362, July 2007.*

[75] Bini, E., Buttazzo, G.C., "Measuring the Performance of Schedulability Tests" *Real-Time Systems*, Vol .30, Issue 1-2, pp. 129-154, May 2005.

[76] Wang, X., Giannakis, G.B., Marques, A.G., "A Unified Approach to QoS-Guaranteed Scheduling For Channel-Adaptive Wireless Networks" in proceedings of *IEEE*, Vol. 95, Issue 12, pp. 2410-2431, December 2007.

[77] Rashwan, A.H., ElBadawy, H.M., Ali, H.H., "Comparative Assessments for Different WiMax Scheduling Algorithms" in Proceedings of *World Congress on Engineering and Computer Science*, Vol. 1, pp. 1-5, October 2009.

[78] Nagaraju C. and Sarkar M., "A Packet Scheduling To Enhance Quality of Service in IEEE 802.16" in Proceedings of *World Congress on Engineering and Computer Science*, Vol. 1, pp. 1-6, October 2009.

[79] Ravichandiran, C., Raj, C.P., Vaidhyanathan, V., "Analysis, Modification, and Implementation (AMI) of Scheduling Algorithm for the IEEE 802.116e (Mobile WiMax)", *International Journal of Computer Science and Information Security, Vol. 7, Issue 2,* pp 94-103, February 2010.

[80] Dahmouni, H., Ghazi H.E., Bonacci, D., Sansò, B., Girard, A., "Improving QoS of all-IP Generation of Pre-WiMax Networks Using Delay-Jitter Model" *Journal of Telecommunications*, Vol. 2, Issue 2, pp. 99-103, May 2010.

[81] Saleh, M., Dong, L., *2010 "*Comparing FCFS & EDF Scheduling Algorithms for Real-Time Packet Switching Networks" in Proceedings of *IEEE International Conference on Networking, Sensing and Control*, pp. 698 -703, April 2010.

[82] Alizadeh, M., Dziyauddin R.A., Kaleshi, D., Doufexi, A., "A Comparative Study of Mixed Traffic Scenarios for Different Scheduling Algorithms In WiMax" in Proceedings of *IEEE 75th Vehicular Technology Conference*, pp 1-6, May 2012.

[83] Awan K.M., Abdullah, A.H., Qureshi, K.N., "Resource Allocation in IEEE 802.16e Mobile WiMax Networks: Survey", *World Applied Sciences Journal,* Vol.28, Issue 1, pp. 103-113, 2013.

[84] Khatkar, A., "Performance Analysis of UGS and BE QOS Classes In WiMax", *Advance in Electronic and Electric Engineering,* Vol. 3, Issue 7, pp. 805-810, 2013.

[85] Gupta, S., "Comparison of Various Scheduling Algorithms In WiMax: A Brief Review", *International Journal of Computer Applications,* pp. 34-36, February 2013.

[86] El-Shennawy, N.M., Youssef, M.A., El-Derini, M.N., Fahmy, M.M., "A Proposed Real–Time Scheduling Algorithm for WiMax Networks", *The Ninth International Conference on Wireless and Mobile Communications,* pp. 167-172, 2013.

[87] Kibria, M.G., Ali, M.A., Bhuiyan, M.N., Kundu, R., "An Improved Algorithm For Ensuring Quality Of Services (QOS) In WiMax Network" *IPASJ*

International Journal of Computer Science, Vol. 1, Issue 2, pp 5-10, July 2013.

[88] Shwetha, D., Thontadharya, H.J., Devaraju, J.T., "A Bandwidth Request Mechanism For QoS Enhancement in Mobile WiMax Networks" *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* Vol. 3, Issue 1, pp. 6881-6888, January 2014.

[89] Khoeia, A.H., Mirjalily, G., Sarram, M.A., 2014 "Improving the Quality Of Services by Scheduling Algorithms in WiMax Networks" *Indian Journal of Science and Research,* Vol. 7, Issue 1, pp. 1174-1184, 2014.

[90] Luther, N.A., Paulus, R., Jaiswal, A.K., Agarwal, A., " Improving The Quality of Service Based On Fair Distribution Of Load and Request in IEEE 802.16 Using Scheduling Algorithms", *International Journal of Current Engineering and Technology,* Vol. 4, Issue 1, pp. 846-849, February 2014.

[91] Kumar, M., Chetan, A., Gupta V., Sharma, S., "Mobility Improvement In IEEE 802.16", *Research Library*, IIT, Delhi.

[92] Makelainen, A., "Analysis of Handoff Performance In Mobile WiMax Networks". *MS thesis* submitted to Helsinki University of Technology, Finland, 2007.

[93] Ray, S.K., Mandal, A., Pawlikowski, K., Sirisena, H., "Hybrid Predictive Base Station Selection Procedure in IEEE 802.16e based WMAN", in Proceedings of Australasian *Telecommunication Networks and Applications Conference,* pp. 93-98, December 2007.

[94] Becvar, Z., Zelenka, J., "Implementation of Handover Delay Timer into WiMax", in Proceedings of the *6th Conference on Telecommunications*, pp. 107–114, January 2007.

[95] Pyo, S., Choi, Y., "A Fast Handover Scheme Using Exponential Smoothing Method", *International Journal of Computer Science and Network Security*, Vol. 9, Issue 2, pp. 61-64, February 2009.

[96] Hwang, I.S., Hwang, B.J., Lai, K.R., Ku, L.F., Hwang, C.C., "Adaptive QoS Aware Resourse Management In Hetrogeneous Wireless Networks", in proceedings of *22nd International Conference on Advanced Information Networking and Applications - Workshops*, pp. 880-885, March 2008.

[97] Al-Ghadi, M.Q., Ababnah, I.M., Mardini, W.E., "Performance Study of SINR Scheme for Vertical Handoff in Wireless Networks*"*, in Proceedings of *International Conference on Information and Communication System, pp. 137-139, January* 2011.

[98] Kalikivayi, S., Misra, I.S., Saha, K., "Bandwidth and Delay Guaranteed Call admission Control Scheme for QoS Provisioning in IEEE 802.16e Mobile WiMax", in Proceedings of *IEEE Global Telecommunications Conference*, pp. 1-6, December 2008.

[99] Pontes, A.B., Silva, D.D.P., Jailton J., Rodrigues O., Dias, K.L., "Handover Management in Integrated WLAN & Mobile WiMax networks", *IEEE Wireless Communications,* Vol.15, Issue 5, pp. 86 – 95, October 2008.

[100] Kumar, R., Singh, B., "Measurement of management and Data Frames in IEEE 802.11x to Model New (QoS) Trigger for vertical Handover Initialization", in Proceedings of *IET International Conference on Wireless, Mobile and Multimedia Networks,* pp. 210-214, January 2008.

[101] Chaudhary, V., Tripathi, R., Shukla, N.K., Nasser, N., "A New Channel Allocation Scheme for Real-Time Traffic in Wireless Cellular Networks", in Proceedings of *IEEE International Performance, Computing and Communication Conference*, pp. 551-555, April 2007.

[102] Cheng, S.T., Lin, J.L., "IPv6-Based Dynamic Coordinated Call Admission Control Mechanism over Integrated Wireless Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 23, Issue 11, pp. 2093-2103, November 2005.

[103] Eshanta, O.M., Ismail, M., Jumari, K., "VHO Strategy for QoS Provisioning in WiMax/WLAN Interworking System", *Asian Journal of applied Sciences*, Vol. 2, issue 6, pp. 511-520, 2009.

[104] Singh, R.k., Singh, R., "4G LTE Cellular Technology: Network Architecture and Mobile Standards", *International Journal of Emerging Research in Management & Technology, Vol. 5, Issue 12, pp. 1-6, December 2016.*

[105] Abdullah, M.F.L., Yonis, A.Z., "Performance of LTE Release 8 and Release 10 in Wireless Communications", in Proceedings of *International Conference on Cyber Security, Cyber Warfare and Digital Forensics*, pp. 236-241, June 2012.

[106] El-Mashed, M.G., El-Rabaie, S., “MIMO AF RSs Based OSTBCs for LTE-A Downlink Physical Layer Network”, *Wireless Personal Communication*, Vol. 95, Issue 3, pp. 2931-2946, August 2018.

[107] Salman, M.I., Ng, C.K., Noordin, N.K., Ali, B.M., Sali, A., “Energy Efficient Transmission for LTE Cellular System”, in Proceedings of *International Conference on Computer and Communication Engineering*, pp. 222-227, July 2012.

[108] Tang, J., Misra, S., Xue, G., “Joint spectrum allocation and scheduling for fair spectrum sharing in Wireless wireless networks,” *Computer Networks*, Vol. 52, Issue 11, pp. 2148-2158, April 2008.

[109] Zhang, Y., Leung, C., “Resource Allocation in an OFDM–based Cognitive Radio System,” *IEEE Transactions on Communications*, Vol. 57, Issue 7, pp. 1928-1931, July 2009.

[110] Tang, J., Hincapie, R., Xue, G., Zhang, W., Bustamante, R., “Fair Bandwidth Allocation in Wireless Mesh Networks with Cognitive Radios,” *IEEE Transactions on Vehicular Technology*, Vol. 59, Issue 3, pp. 1487-1496, March, 2010.

[111] Ngo, D.T., Le-Ngoc., T., “Distributed Resource Allocation for Cognitive radio Networks with Spectrum Sharing Constraints,” *IEEE Transactions on Vehicular Technology*, Vol. 60, Issue 7, pp. 3436-3449, September 2011.

[112] Alqerm, I., Shihada, B., Shin, K.G., “CogWnet: A Resource Management Architecture for Cognitive Wireless Networks”, in Proceedings of *International Conference on Computer Communications and Networks*, pp. 1-7, August 2013.

[113] Buenestado, V., Ruiz-Aviles, J.M, Toril, M., Luna-Ramirez, S., Mendo, A., “Analysis of Throughput Performance Statistics for Benchmarking LTE Networks”, *IEEE communication Letters*, Vol. 18, Issue 9, September 2014.

[114] Ali, A., Barakabitze, A.A., "Evolution of LTE and Related Technologies Towards IMT- Advanced”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, Issue 1, pp. 16-22, January 2015.

[115] Raheem, R., Lasebae, A., Aiash, M., Loo, J., Colson, R.H., “Mobile Femtocell Utilization in LTE Vehicular Environment: Vehicular Penetration Loss

Elimination and Performance Enhancement", *Vehicular Communications*, Vol. 9, pp. 31–42, July 2017.

[116] Riyazuddin, K., Sharma, A.K., "Performance Evaluation of LTE-based System Parameters Using OFDM in Indoor and Outdoor Environment", *Indian Journal of Science and Technology*, Vol. 9, Issue 48, pp. 1-6, December 2016.

[117] Sharma, S., Sharma, B., Kanwar, V., "Review Paper on Performance of OFDM in 4g Wireless Communication", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 6, Issue 9, pp. 46-49, September 2016.

[118] Ajey, S., Srivalli, B., Rangaraj, G., "On Performance of MIMO-OFDM Based LTE Systems", In Proceedings of *International Conference on Wireless Communication and Sensor Computing*, pp. 1-5, January 2010.

[119] Tiwari, H., Nath, R.K., "Performance Evaluation of LTE and WiMax Network as per Use Cases", *International Journal of Advanced Research in Computer Engineering & Technology*, Vol. 4, Issue 7, pp. 3081-3088, July 2015.

[120] Ali, S., Zeeshan, M., Naveed, A., "A Capacity and Minimum Guarantee-based Service Class-oriented Scheduler for LTE Networks", *EURASIP Journal on Wireless Communications and Networking, pp. 1-15, March* 2013.

[121] Zaki, Y., Weerawardane, T., Li, Xi., Gorg, C., "LTE Radio Schedulers Analytical Modeling using Continuous Time Markov Chains", in Proceedings of *6th Joint IFIP Wireless and Mobile Networking Conference*, pp. 1-10, April 2013.

[122] "EXALTED, Expanding LTE for Devices D3.3," Final Report on *LTE-M Algorithms and Procedures, large Scale Integration Project*, July 2012.

[123] "LTE Advanced- Heterogeneous Networks", A Report, *QUALCOMM Incorporated* January 2011.

[124] Shukla, S., Bhatia, V., " Priority Scheduling Algorithm for Traffic in a LTE-based Defence Mesh Network Incorporating Centralized Scheduling Architecture", *Defence Science Journal*, Vol. 67, No. 5, pp. 581-587, September 2017.

[125] Kamath, H.S., Singh, H., Khanna, A.,“Transmit Receive Chain in LTE”, in Proceedings of *IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, pp. 1-7, October 2016.

[126] Siomina, I., Yuan, D., “Analysis of Cell Load Coupling for LTE Network Planning and Optimization”, *IEEE Transactions on Wireless communications*, Vol. 11, Issue 6, pp. 2287-2297, June 2012.

[127] Parruca, D., Grysla, M., Gortzen, S., Gross, J., “Analytical Model of Proportional Fair Scheduling in Interference-limited OFDMA/LTE Networks”, in Proceedings of *IEEE 78th Vehicular Technology Conference*, pp. 1-7, March 2013.

[128] Kwan, R., Leung, C., “A Survey of Scheduling and Interference Mitigation in LTE”, *Journal of Electrical and Computer Engineering*, Vol. 2010, pp. 1-10, 2010.

[129] Habaebi, M.H., Chebil, J., Al-Sakkaf, A.G., Dahawi, T.H., “Comparison Between Scheduling Techniques in Long Term Evolution”, *IIUM Engineering Journal*, Vol. 14, Issue 1, pp. 66-75, 2013.

[130] Kawser, M.T., Farid, H.M.A.B., Hasin, A.R., Sadik, A.M.J., Razu, I.K., “Performance Comparison between Round Robin and Proportional Fair Scheduling Methods for LTE”, *International Journal of Information and Electronics Engineering, Vol. 2, Issue 5, pp. 678-681, September 2012.*

[131] Venkatesan, S. and Manoharan, C., “Access Point Selection for Fair Load Balancing in Wireless LAN”, *Information Technology Journal,* Vol. 11, Issue 2, pp. 283-288, 2012.

[132] Velayos, H., Aleo, V., and G. Karlsson, “Load balancing in overlapping wireless LAN cells**”**, *in Proceedings of IEEE International Conference on Communications*, pp. 3833–3836, June 2004.

[133] Lee, S.H., Han, Y., “A Novel Inter-FA Handover Scheme for Load Balancing in IEEE 802.16e System,” in Proceedings of IEEE 65th *Vehicular Technology Conference*, pp. 763–767, April 2007.

[134] Cassell, B.A., Alperovich, T., Noble, B., Wellman, M.P., “Access Point Selection Under Emerging Wireless Technologies”, in Proceedings of *6th*

Workshop on Economics of networks, Systems and Computations, pp.1-7, June 2011.

[135] Casey, T., Veselinovic, N., jantti, R., "Base Station Controlled Load Balancing with Handovers in Mobile WiMax" in Proceedings of *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-5, September 2008.

[136] Kumar, M., Sharma, S.C., "Dynamic Load Balancing Algorithm to Minimize the Makespan Time and Utilize the Resources Effectively in Cloud Environment", *International Journal of Computers and Applications*, pp. 1-10, November 2017.

[137] Aghazadeh, Y., Kalbkhani, H., Shayesteh, M.G., Solouk, V., "Cell Selection for Load Balancing in Heterogeneous Networks", *Wireless Personal Communications*, Vol. 101, Issue 5, pp. 305-323, April 2018.

[138] Xia, F., Ahmed, A.M., Yang, L.T., Luo, Z., "Community-Based Event Dissemination with Optimal Load Balancing", *IEEE Transactions on Computers*, Vol. 64, Issue 7, pp. 1857-1869, July 2015.

[139] Wang, H., Ding, L., Wu, P., Pan, Z., Liu, N., & You, X. (2010). Dynamic load balancing and throughput optimization in 3GPP LTE networks", in Proceedings of the *6th International Wireless Communications and Mobile Computing Conference*, pp. 939-943, June 2010.

[140] Son, K., Chong, S., Veciana, G., "Dynamic association for load balancing and interference avoidance in multi-cell networks", *IEEE Transactions on Wireless Communications*, Vol. 8, Issue 7, pp. 3566–3576, July 2009.

[141] Tonguz, O.K., Yanmaz, E., "The mathematical theory of dynamic load balancing in cellular networks", *IEEE Transactions on Mobile Computing*, Vol. 7, Issue 12, pp. 1504–1518, December 2008.

[142] "WiMax-LTE Technology Migration", *Telesystem Innovations* white paper, pp. 1-16, 2010

[143] Haffajee, H.A., "Low-cost QoS-enabled Wireless Network with Interworked WLAN and WiMax," in Proceedings of the *First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, pp. 1-6, March 2006.

[144] Kim, D., Ganz, A., "Architecture for 3G and 802.16 Wireless Networks Integration with QoS Support," in Proceedings of *Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, pp. 1-8, August 2005.

[145] Ali-Yahiya, T., Sethom, K., Pujolle, G., *"Seamless Continuity of Service across WLAN and WMAN Networks: Challenges* and *Performance Evaluation", in Proceedings of 2nd IEEE/IFIP International Workshop on Broadband Convergence Networks, pp. 1-12, May 2007.*

[146] Chen, J.L., Chen, M.C., Liu, C.W., Jhuo, J.Y., "Cross-layer QoS architecture for 4G heterogeneous network services", in Proceedings of ***11**th International Conference on Advanced Communication Technology*, Vol. 1, pp. 783-785, February 2009.

[147] Nirmal, A., Kondareddy, Y., Agrawal, P., "Prioritized Resource sharing In WiMax & WiFi Integrated Networks", in Proceedings of *IEEE Wireless Communications and Networking Conference (WCNC)- 2010*, pp. 1-6, April 2010.

[148] Kassab, M., Bonnin, J.M., Belghith, A., "Technology Integration Framework for Fast and Low Cost Handovers-Case Study: Wi-Fi and WiMax Networks", *Journal of Computer Systems, Networks and Communications*, Vol. 2010, pp. 1-22, April 2010.

[149] Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., YOO, T., Song, O., Malladi, D., "A Survey on 3GPP Heterogeneous Networks", *IEEE Wireless Communications*, Vol. 18, Issue 3, pp. 10-21, June 2011.

[150] Ng, M.H., Lin, S.D., Li, J., Tatesh, S., "Coexistence Studies for 3GPP LTE with Other Mobile Systems", *IEEE Communications Magazine*, Vol. 47, Issue 4, pp. 60-65, April 2009.

[151] Chen, W., Yu, H., Feng, W., "A self adaptive Packet scheduling algorithm for hybrid Traffic in Heterogeneous Wireless Networks", *International Journal of Future Generation Communication and Networking*, Vol. 7, Issue 6, pp. 69-80, 2014.

[152] Ruby, R., Michelson, D.G., Leung, V.C.M., "Performance Analysis of a Heterogeneous Traffic Scheduler using Large Deviation Principle", *Computer Networks*, pp. 1-35, April 2016.

[153] Zorba, N., Verikoukis, C., "A QoS-based Dynamic Queue Length Scheduling Algorithm in Multiantenna Heterogeneous Systems", *EURASIP journal on Wireless Communications and Networking*, pp. 1-25, December 2009.

[154] Rodriguez, J., Monteiro, V., Gomes, A., Renzo, M.D., Alonso-Z´arate, J., Verikoukis, C., Gonzalez, A., Lazaro, O., Akan, A., Vila, A.P., Kormentzas, G., Boixade, D., laMaza, S.D., "Converged Wireless Networking and Optimization for Next Generation Services", *EURASIP Journal on Wireless Communications and Networking*, pp. 1-11, December 2010.

[155] Lee, J.H., Yan, Z., You, I., "Enhancing QoS of Mobile Devices by a New Handover Process in PMIPv6 Networks", *Wireless Personal Communications*, Vol. 61, Issue 4, pp. 591–602, December 2011.

[156] Tang, F., You, I., Tang, C., Guo, M., "An Efficient Classification Approach for Large-scale Mobile Ubiquitous Computing", *Information Sciences*, Vol. 232, pp. 419–436, May 2013.

[157] Xia, F., Asabere, N.Y., Ahmed, A.M., Li, J., Kong, X., "Mobile Multimedia Recommendation in Smart Communities: A Survey", *IEEE Access*, Vol. 1, pp. 606-624, September 2013.

[158] Prakash, A., Verma, R., Tripathi, R., Naik, K., "A Seamless Handover Scheme for Vehicles Across Heterogeneous Networks", *International journal of Communication Networks and Distributed Systems, Vol. 8, Issue 1-2, pp. 4-23, December 2011.*

[159] Verma, P.K., Verma, R., Prakash, A., Agrawal, A., Naik, K., Tripathi, R., Khalifa, T., Alsabaan, M., Abdelkader, T., Abogharaf, A., "Machine-to-Machine (M2M) Communications: A Survey", *Journal of Network and Computer Applications*, Vol. 66, pp. 83-105, May 2016.

[160] Yajun Li, Y. Yang, C. Cao, L. Zhou, "QoS Issues in IEEE 802.16 Mesh Networks", in Proceedings of *1st International Conference on Information Science and Engineering*, pages: 2687-2690, December 2009.

[161] Votava, O., "A Network Simulation Tool for Task Scheduling", *Acta Polytechnica*, Vol. 52, Issue 5, pp 112-119, 2012.

[162] Kaur, G., "A DAG based Task Scheduling Algorithms for Multiprocessor System - A Survey", *International Journal of Grid and Distributed Computing*, Vol. 9, Issue 9, pp.103-114, 2016.

[163] Eswari, R., Nickolas, S., "A Level-wise Priority Based Task Scheduling for Heterogeneous Systems", *International Journal of Information and Education Technology*, Vol. 1, Issue 5, pp. 371-376, December 2011.

[164] Yang, C., Lee, P., Chung, Y., "Improving Static Task Scheduling in Heterogeneous and Homogeneous Computing Systems," in Proceedings of *International Conference on Parallel Processing*, pp. 1-8, September 2007.

[165] Wu, M., Gajski, D., "Hypertool: A Programming aid for message passing systems," *IEEE Transaction on Parallel and Distributed Systems, V*ol.1, Issue 3, pp.330-343, July 1990.

[166] Boeres, C., Filho, J.V., Rebello, V.E.F., "A Cluster Based Strategy for Scheduling Task on Heterogeneous Processors," in Proceedings of *16th Symposium on Computer Architecture and High Performance Computing ,* pp.214-221, October 2004.

[167] Baskiyar, S., Dickinson, C., "Scheduling Directed A-cyclic Graph on a Bounded Set of Heterogeneous Processors Using Task Duplication", *Journal of Parallel Distributed Computing*, Vol. 65, pp. 911-921, May 2005.

[168] Alam, M., Khan, A., Varshney, A.K., "A Review of Dynamic Scheduling Algorithms for Homogeneous and Heterogeneous Systems", *Advances in Intelligent Systems and Computing*, Vol. 732, pp. 73-83, May 2018.

[169] Zhao, H., Sakellarious, R., "Scheduling Multiple DAGs onto Heterogeneous Systems", in Proceedings of the *20th IEEE International Parallel and Distributed Processing Symposium*, pp.1-14, April 2006.

[170] Olteanu, A., Marin, A., "Generation and Evaluation of Scheduling DAGs: How to Provide Similar Evaluation Conditions", *Computer Science Master Research*, Vol. 1, Issue 1, pp 57-66, January 2011.

[171] Nasr, A.A., El-Bahnasawy, N.A., El-Sayed, A., "Performance Enhancement of Scheduling Algorithm in Heterogeneous Distributed Computing Systems", *International Journal of Advanced Computer Science and Applications*, Vol. 6, Issue 5, pp 88-96, 2015.

[172] Chen, S., Zhao, J., "The Requirements, Challenges and Technologies for 5G of Terrestrial Mobile Telecommunication", *IEEE Communications Magazine*, Vol. 52, Issue 5, pp. 36–43, May 2014.

[173] Demestichas, P., Georgakopoulos, A., Karvounas, D., Tsagkaris, K., Stavroulaki, V., Lu, J., Xiong, C., Yao, J., "5G on the Horizon: Key Challenges for the Radio-Access Network", *IEEE Vehicular Technology Magazine*, Vol. 8, Issue 3, pp. 47–53, September 2013.

[174] Gupta, A., Jha, R.K., "A Survey of 5G Network: Architecture and Emerging Technologies", *IEEE Access*, Vol. 3, pp. 1206–1232, 2015.

[175] http://en.wikipedia.org/wiki/Moment-generating _function

[176] Dudin, A., Klimenok, V., Lee, M.H., "Recursive Formulas for the Moments of Queue Length in the BMAP/G/1Queue", *IEEE Communication Letters*, Vol. 13, Issue 5, pp. 351 – 353, 2009.

[177] Makaronidis, M., Casale, G., "Efficient Parallelization of the Method of Moments for Queueing Networks Using Multi-Modular Algebra", in Proceedings of *5th International ICST Conference on Performance Evaluation Methodologies and Tools*, pp. 176-185, May 2011.

[178] Tomic, D., Radusinovic, I., "Recursive Formula for the Moments of Queue Length in the M/M/c and Erlangian Queues", *17th Telecommunications forum TELFOR*, pp. 202-205, November, 2009.

[179] Whitt, W., "Fitting birth-and-death queuing models to data", *Statistics and Probability Letters*, Vol. 82, pp. 998–1004, 2012.

[180] Furht, B., & Ahson, S. A., "Long term evolution: 3GPP LTE radio and cellular technology" published by *Boca Raton: CRC Press*, April 2009.

[181] IST-WINNER II, D1.1.2, "WINNER II Channel Models", Ver 1.0, 2007.

[182] Sonia, Khanna, R., Kumar, N., "Load Balancing Efficiency Improvement using Hybrid Scheduling Algorithm in LTE Systems', *Wireless Personal Communications*, Vol. 96, Issue 343, pp 1-13, July 2017.

[183] Al-Jaradat, H., Sandrasegaran, K., "On the Performance of PF, MLWDF and EXP/PF algorithms in LTE." *International Journal of Computers & Technology*, Vol. 8, Issue 1, 2013.

[184] Dardouri, S., Bouallegue, R., "Comparative study of downlink packet scheduling for LTE networks." *Wireless Personal Communications,* Vol. 82, Issue 3, pp. 1405-1418, June 2015.

[185] Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., "Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research." *Bioinformatics,* Vol. 21, Issue 18, pp. 3674-3676, Sept 2005.

[186] Wang, L., "Directed acyclic graph". In Encyclopedia of *Systems Biology*, pp. 574-574, 2013.