ISLAMIC UNIVERSITY OF TECHNOLOGY

(IUT)

IMPROVED RESOURCE SCHEDULING FOR LTE

BY

SK. TAIFUR RAHMAN

MD. SADEK HOSSAIN ALVE

INTISAR TAHMID

MEHEDI HASAN

A Dissertation
Submitted in Partial Fulfillment of the Requirement for the
Bachelor of Science in Electrical and Electronic Engineering
Academic Year: 2013-2014

Department of Electrical and Electronic Engineering.
Islamic University of Technology (IUT)
A Subsidiary Organ of OIC
Dhaka, Bangladesh.

A Dissertation on

IMPROVED RESOURCE SCHEDULING FOR LTE

Submitted by

SK. TAIFUR RAHMAN **MD SADEK HOSSAIN ALVE** **INTISAR TAHMID**

MEHEDI HASAN

Approved By

Prof. Dr. Md. Shahid Ullah

Head of the Department
Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur-1704, Bangladesh.

Supervised by

MD. TAWHID KAWSER

Assistant Professor
Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT).

ABSTRACT

LTE, an abbreviation for Long-Term Evolution, commonly marketed as 4G LTE, is a standard for wireless communication of high-speed data for mobile phones and data terminals. It is based on the GSM/EDGE and UMTS/HSPA network technologies, increasing the capacity and speed using a different radio interface together with core network improvements. The standard is developed by the 3GPP (3rd Generation Partnership Project) and is specified in its Release 8 document series, with minor enhancements described in Release 9.

LTE is the natural upgrade path for carriers with both GSM/UMTS networks and CDMA 2000 networks. The different LTE frequencies and bands used in different countries will mean that only multi-band phones will be able to use LTE in all countries where it is supported.

MOTIVATION

Scheduling of resources in LTE is a very important factor as it determines the improvements in data rate, SINR QOS and other key factors.

Long Term Evolution (LTE), proposed by 3rd generation Partnership Project (3GPP) as a 3.9G technology, represents a very promising answer to the ever rising bandwidth Demand of mobile applications. To support vast range of multimedia and internet services at high data rates that too with increased spectral efficiency; LTE incorporates various Radio Resource Management (RRM) procedures. The key to achieve optimal performance of base station is dynamically scheduling limited resources like power and bandwidth to offer the best service for terminals with the lowest cost. In this context, radio Resource allocation strategies play a key role in distributing radio resources among different stations by taking into consideration the channel conditions as well as QoS requirements. The present paper provides review of radio resource allocation strategies present in the literature

CONTENTS

Introduction:

The evolution of wireless telephone technologies can be discretely grouped into various generations based on the level of maturity of the underlying technology. The classification into generations is not standardized on any given metrics or parameters and as such does not represent a strict demarcation. However, it represents a perspective which is commonly agreed upon, both by industry and academia, and hence conceived to be an unwritten standard. At this time, there are two major efforts towards the development of the next generation - "4G" wireless access technology. The 3GPP or 3rd Generation Partnership project (brand named as Long Term Evolution) is the name of the 4G efforts being undertaken in Europe and the 3GPP2 or 3rd Generation Partnership project 2 (brand named as Ultra Mobile Broadband) is the 4G effort of North America and parts of Asia. This book tries to present an evolutionary and objective sketch to the development efforts of these technologies that mark the future of wide area broadband wireless access technologies.

Evolution:

The evolution of wireless telephone technologies can be discretely grouped into various generations based on the level of maturity of the underlying technology. The classification into generations is not standardized on any given metrics or parameters and as such does not represent a strict demarcation. However, it represents a perspective which is commonly agreed upon, both by industry and academia, and hence conceived to be an unwritten standard.

The first generation or 1G was represented by an analog wireless access system primarily for voice traffic. The AMPS (Advance Mobile Phone System) in United States and the TACS (Total Access Communication System) in most parts of Europe represented this generation. The analog channel was susceptible to static noise and did not provide any protection from eavesdropping on the shared media. However, AMPS laid the foundation to the "cellular" technology which pioneered the use of small hexagonal service areas and hence supported frequency re-use across the "cells" without interference.

The 1G technology was soon replaced by the second generation or 2G technologies which represented the replacement of the analog radio network with digital radio network. The digital technology was much superior than its analog counterpart in the sense that digitized data could be subjected to superior processing techniques making it less susceptible to noise. Also, digital technology is based on discreet bi-level signals as against continuous analog signals making it easier to calibrate and maintain and hence cheaper than analog devices. 2G technologies could be further classified into Time Division Multiple Access (TDMA) based and Code Division Multiple Access (CDMA) based. The TDMA based technology was adopted mostly in Europe and was called Global system for Mobile communications or GSM (originally Groupe Special Mobile) while USA adopted the CDMA based technology and it was called CDMA one or standardized as IS-95a. CDMA had the advantage of supporting more users than GSM due to the better usage of the spectrum. CDMA is a spread spectrum technology in which each user is allowed to transmit over the whole spectrum using a different orthogonal code. Plainly speaking, each user uses a distinct code of one's and zero's to represent a one and zero at the other end. All the codes are orthogonal to each other and hence don't interfere. Neighboring cells may reuse the same frequency band and not interfere as long as they use different code, thus allowing better use of the available spectrum. CDMA one supported digital data transfer rates varying between 4.8-14.4 kbps while CDMA two or IS-95b supported data rates of around 115.2 kbps.

The 2G technology led to an interim generation of 2.5G which represented 2G systems which implemented a packet switched domain in addition to the circuit switched domain. General Packet Radio Service (GPRS) was the 2.5G technology adopted by GSM.GPRS provides a packet switched service over GSM offering data speeds between 56-114 kbps. Enhanced Data Rates for GSM Evolution (EDGE) over GSM and CDMA2000 1xRTT over CDMA were touted as 2.75G technologies though they may well be called 3G technologies as the surpass data rates of 144kbps required to qualify as 3G technology because their data rates were far below the data rates of actual 3G technologies. EDGE provides data rates of 236.8 kbps while CDMA2000 deployments limit the data rates at 144kbps.

This interim period led to the evolution of the third Generation of Mobile technology, better known as 3G. The International Telecommunication Union (ITU) under the International Mobile Telecommunications Program fixed the minimum data rate of 144kbps for any technology to qualify to be a 3G technology. However, most technologies which fall under this category by far surpass this minimum limit and provide data rates typically between 5-10 Mbps 3G Technologies attain better spectral efficiency (more bits/ Hertz) over wide area cellular telephone networks allowing for higher data rates and enhanced services. The first pre-commercial and commercial 3G technology was installed in Japan followed by South Korea. In Europe the Universal Mobile Telecommunication System (UMTS) is the adopted 3G technology using W-CDMA (Wideband Code Division Multiple Access) as the air interface.

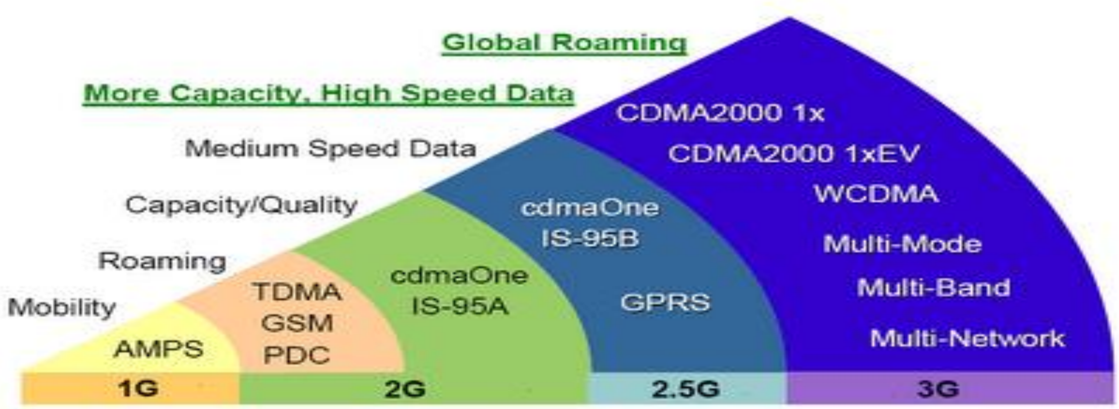
UMTS is sometimes called 3GSM to emphasize the fact that it is the 3rd generation technology succeeding GSM. The

evolution of CDMA based technologies to 3G was through the CDMA2000 family of protocols, especially EV-DO (Evolution-Data Optimized) which uses multiplexing techniques including CDMA and TDMA to increase per user as well as system throughput.

UMTS based 3G technologies have raised themselves to 3.5G with HSDPA (High Speed Downlink Packet Access) allowing data rates up to 7.2Mbps. Also these networks are planning their progress into 4G through the 3GPP (3rd Generation Partnership Projects) where they aim to attain data rates in the order of 100Mbps downlink and 50 Mbps Uplink. Similarly, the North American counterpart is planning their 4th Generation through the 3GPP2 (3rd Generation Partnership Project 2) and aim at comparable data rates.

LTE or Long Term Evolution is the brand name given to the efforts of 3GPP 4th Generation technology development efforts mostly in Europe and UMB (Ultra-Mobile Broadband) is the brand name for similar efforts by 3GPP2 in North America.

Having presented some evolutionary details and a historical roadmap, our main aim in this survey is to study the two competing 4G technologies of LTE and UMB in more detail. In the rest of the survey, we shall provide a brief introduction into the LTE and UMB projects and look into the various technical challenges and the technologies adopted which shall go on to make these 4G efforts successful. We shall then look into the various services that shall be made possible as a result of these 4G technologies and finally try to present an objective comparison between the two 4G efforts and also between these 4G efforts and WIMAX(Wireless Interoperability for Microwave Access).



4G wireless standards

The two competing bodies involved in churning out 4G wireless technologies are the 3GPP in Europe and the 3GPP2 in North America. The 3GPP is marketed under the brand name of Long Time Evolution or LTE and is working on the 4G technology which is to succeed the 3G technology of UMTS. The 3GPP2 project is marketed under the brand name Ultra Mobile Broadband or UMB and their effort is to make transition to 4G from the existing CDMA2000 family of standards in North America.

The High Level requirements for a 4G technology were identified as:

1. Higher spectral Efficiency

2. Reduced cost per bit

3. Increased Service Provisioning by lowering the cost and increasing efficiency and experience

4. Open Interfaces as against closed technologies of the past

5. Power consumption efficiency

6. Scalable and flexible usage of frequency bands

The 3rd Generation Partnership project was established in December 1998, and is a collaborative agreement to bring together a number of Telecommunications standard bodies known as "Organizational Partners" [3GPP]. The stated aim of the collaboration is to "co-operate for the production of a complete set of globally applicable Technical Specifications for a 3rd Generation Mobile System based on the evolved GSM core networks and the radio access technologies supported by 3GPP partners (i.e., UTRA both FDD and TDD modes)" . The 3GPP organization is based on a layered hierarchy with a "Technical specifications" Group working under the directions of a "Project Coordination" group to role out technical specifications as shown in Figure 2 (adapted from reference). The "Market Representation Partners" is an organization invited by the Organization Partners to advice them about market requirements and strategies. Individual members make technical contributions to the "Technical specifications Group". The "Organizational Partners" shall have joint ownership and copyright to the technical specifications churned out of the project.

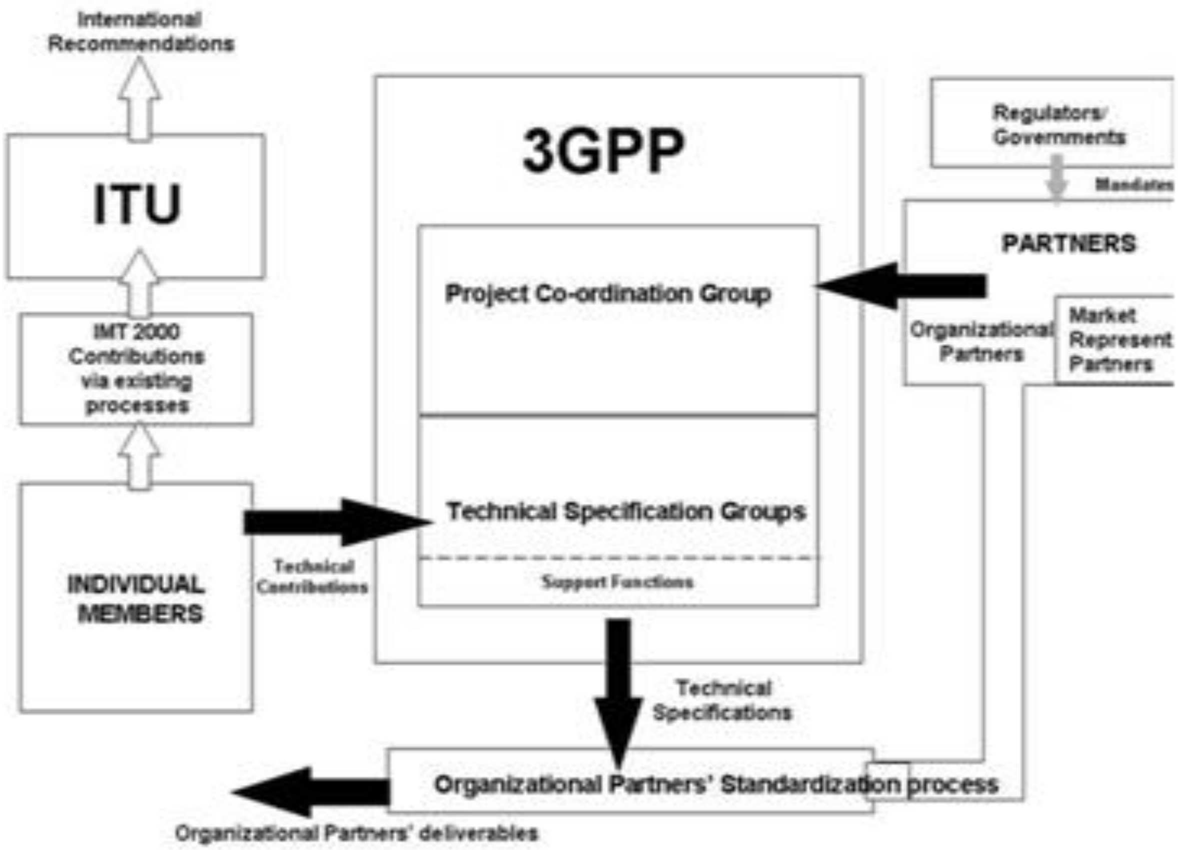


Figure: 3GPP Overview

The technical specifications approved by 3GPP for the LTE project include the use of Orthogonal Frequency Division Multiplexing (OFDM) and advanced antenna technologies such as MIMO (Multiple Input Multiple Output).It specifies downlink peak speeds of 326Mbps and uplink peak speeds of 86Mbps, both in a 20 Mhz bandwidth. It also mandates the roundtrip latency between the base station and handsets to 10-milliseconds. The specification documents extend numerous documents and thousands of pages and the information provided above is just a summary of the final results.

The 3GPP2

It is the Asian-North American effort for achieving similar capabilities for their CDMA2000 group of specifications and called by the brand-name of UMB. 3GPP2 organization is very similar to the 3GPP organization being a collaborative effort between 5 standards development organization from Asia(China, Japan, North Korea) and North America and multiple Market Representation Partners, providing market advice to the SDO's.

The technical specifications approved by 3GPP2 for UMB include and OFDMA based air interface with Frequency Division Duplexing. The specifications specify downlink peak data rates of 275 Mhz and uplink peak data rates of 75Mbps on a scalable bandwidth of 1.25-20 Mhz. It also supports the use of advanced antenna systems such as MIMO and Beam forming antennas.

It can be noticed that the technical specification of LTE and UMB are very similar to each other and are based on the same underlying technologies of OFDM and AAS(Advanced Antenna Systems). Both these standards represent a shift towards an All IP network and as such an All IP network has been specified as a part of the System Architecture Evolution (SAE), the core network architecture for LTE.

Differences among the Generations :

Generation (1G,2G,3G,4G, 5G)	**Definition**	**Throughput/ Speed**	**Technology**	**Time period**	**Features**
1G	Analog	14.4 Kbps (peak)	AMPS,NMT,TACS	1970 – 1980	During 1G Wireless phones are used for***voice only***.
2G	Digital Narrow band circuit data	9.6/14.4 Kbps	TDMA,CDMA	1990 to 2000	2G capabilities are achieved by allowing***multiple users on a single channel via multiplexing***.During 2G Cellular phones are

					used for ***data also along with voice***.
2.5G	Packet Data	171.2 Kbps(peak) 20-40 Kbps	GPRS	2001-2004	In 2.5G the***internet*** becomes popular and data becomes more relevant.2.5G***Multimedia services*** and streaming starts to show growth.***Phones start supportingweb browsing***though limited and very few phones have that.
3G	Digital Broadband Packet Data	3.1 Mbps (peak) 500-700 Kbps	CDMA 2000 (1xRTT, EVDO) UMTS, EDGE	2004-2005	3G has ***Multimedia services support***along with streaming are more popular.In 3G,***Universal access***and ***portability*** across different device types are made possible. (Telephones, PDA's, etc.)
3.5G	Packet Data	14.4 Mbps (peak) 1-3 Mbps	HSPA	2006 – 2010	3.5G supports ***higher throughput and speeds*** to support higher data needs of the consumers.
4G	Digital Broadband Packet All IP	100-300 Mbps (peak) 3-5 Mbps	WiMax LTE Wi-Fi	Now (Read more onTransitioning to 4G)	***Speeds*** for 4G are further increased to keep up with data access demand used by various

	Very high throughput	100 Mbps (Wi-Fi)			services.***High definition streaming*** is now supported in 4G. New phones with HD capabilities surface. It gets pretty cool.In 4G,***Portability*** is increased further.***World-wide roaming*** is not a distant dream.
5G	Not Yet	Probably gigabits	Not Yet	Soon (probably 2020) **Update**:Samsung conducts tests on 5G	***Currently there is no 5G technology deployed***. When this becomes available it will provide very high speeds to the consumers. It would also provide efficient use of available bandwidth as has been seen through development of each new technology.

In the next section we shall discuss about the various technical challenges and discuss in detail the key technologies which make the transition from 3g to 4g possible.

Technical Challenges and Technologies Adopted

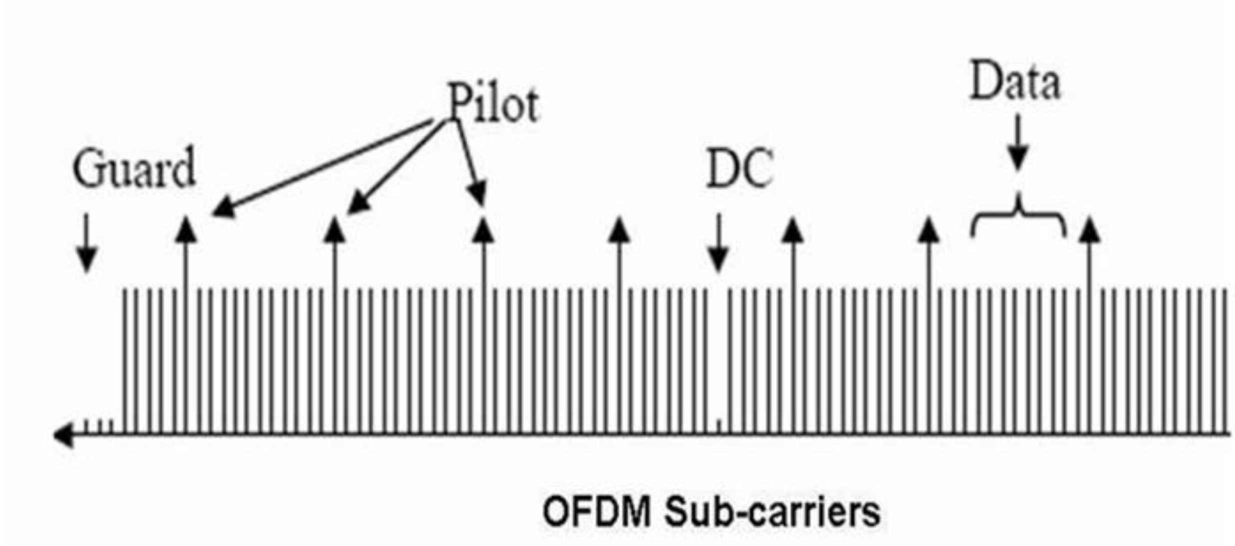
In this section, we first discuss the two generic technologies of OFDM and MIMO that are adopted by both standards(LTE and UMB) and then look into the details System Architecture Evolution- the proposed architectural framework proposed specifically for LTE.

OFDM:

Orthogonal Frequency Division Multiplexing is a superior air access method compared to its predecessor CDMA. Also OFDM is one of the key technologies which enable non-line of sight wireless services making it possible to extend wireless access system over wide-areas. It is a variant of the Frequency Division Multiplexing scheme in which the frequency channel is divided into multiple smaller sub-channels. In FDM, sub-channelization requires provisioning of guard bands between two sub-channels to avoid interference between them. OFDM (as shown in Figure 3) divides the frequency bandwidth in narrow orthogonal sub-parts called sub-carriers. A sub-channel is an aggregation of a number of these sub-carriers. The sub-carriers include data carriers, pilot carriers and a DC. The data carriers are used to carry data, the pilot carriers are used for channel sensing purposes and the DC mark the centre of the channel. Each subcarrier is modulated with conventional modulation scheme such as Quadrature Amplitude Modulation or Phase Shift Keying at a low symbol rate. Each user is provided with a integer number of sub-channels which is composed of a number of sub-carriers. User data is carried parallely on each sub-carrier at a low rate. The combination of the parallel sub-carriers at the destination provide for the high data rates.

Since the sub-carriers carry data at a low rate and thus higher symbol time it is more resilient to multi-path effects, thus making it more suitable for wide-area non-line of Sight wireless access technology. Also, the use of overlapping orthogonal sub-carriers without guard bands make it more efficient than FDM scheme. OFDM resembles CDMA in that it is also a spread-spectrum technology in which energy generated at a particular bandwidth is spread across a wider bandwidth making it more resilient to interference and "jamming". However, unlike CDMA, OFDM allows adaptive assignment of sub-carriers to sub-channels based on channel conditions making it more robust and achieving higher spectral efficiency than CDMA.

The Multi-User version of OFDM is called OFDMA(Orthogonal Frequency Division Multiple Access).


Pilot
Data
Guard
DC
OFDM Sub-carriers

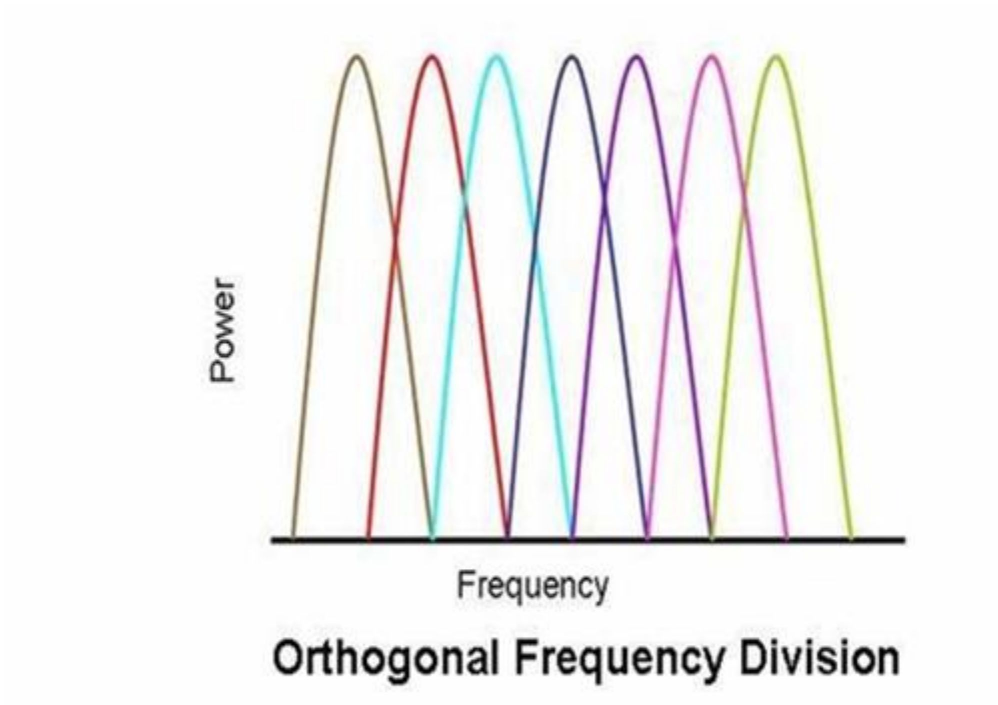

Power
Frequency
Orthogonal Frequency Division

Figure :OFDM

MIMO:

Multiple Input Multiple Output (MIMO) is one of the most popular Advanced Antenna Technologies which is supported both by LTE and UMB. The salient features of MIMO is that it offers higher throughput for a given bandwidth and higher link range for a given power value. A detailed discussion of the MIMO technology is beyond the scope of this survey and we provide a cursory glance at the key features of the technology. In MIMO the transceiver and receiver have multiple antennas giving MIMO multiple flavors based on the number of antennas present on each side. However, the key idea is that a transmitter sends multiple streams on multiple transmit antennas and each transmitted stream goes through different paths to reach each receiver antenna as shown in Figure 3.The different paths taken by the same stream to reach multiple receivers allow canceling errors using superior signal processing techniques. MIMO also achieves spatial multiplexing to distinguish among different symbols on the same frequency. MIMO thus helps in achieving higher spectral efficiency and Link reliability.

Architectural Review of UMTS and GSM

High-Level Architecture

LTE was designed by a collaboration of national and regional telecommunications standards bodies known as the Third Generation Partnership Project (3GPP) [1] and is known in full as 3GPP Long-Term Evolution. LTE evolved from an earlier 3GPP system known as the Universal Mobile Telecommunication System (UMTS), which in turn evolved from the Global System for Mobile Communications (GSM). To put LTE into context, we will begin by reviewing the architectures of UMTS and GSM, and by introducing some of the important terminology. A mobile phone network is officially known as a public land mobile network (PLMN), and is run by a network operator such as Vodafone or Verizon. UMTS and GSM share a common network architecture, which is shown in Figure 1.1. There are three main components, namely the core network, the radio access network and the mobile phone.

The core network contains two domains. The circuit switched (CS) domain transports phone calls across the geographical region that the network operator is covering, in the same way as a traditional fixed-line telecommunication system. It communicates with the public switched telephone network (PSTN) so that users can make calls to land lines and with the circuit switched domains of other network operators. The packet switched (PS) domain transports data streams, such as web pages and emails, between the user and external packet data networks (PDNs) such as the internet. The two domains transport their information in very different ways. The CS domain uses a technique known as circuit switching, in which it sets aside a dedicated two-way connection for each individual phone call so that it can transport the information with a constant data rate and minimal delay.

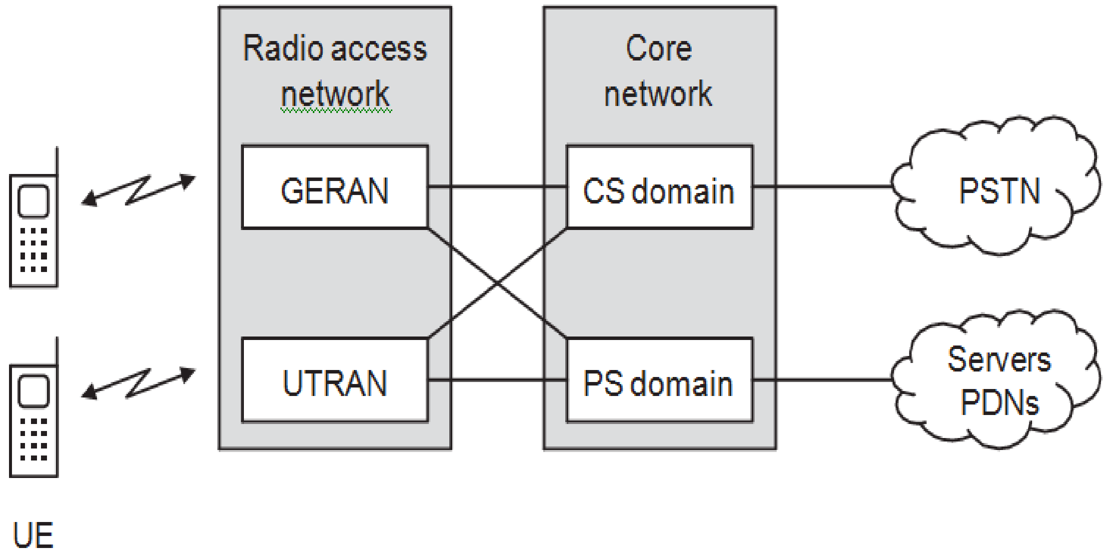


Figure: High-level architecture of UMTS and GSM

This technique is effective, but is rather inefficient: the connection has enough capacity to handle the worst-case scenario in which both users are speaking at the same time, but is usually over-dimensioned. Furthermore, it is inappropriate for data transfers, in which the data rate can vary widely.

To deal with the problem, the PS domain uses a different technique, known as packet switching. In this technique, a data stream is divided into packets, each of which is labelled with the address of the required destination device. Within the network, routers read the address labels of the incoming data packets and forward them towards the corresponding destinations. The network's resources are shared amongst all the users, so the technique is more efficient than circuit switching. However, delays can result if too many devices try to transmit at the same time, a situation that is familiar from the operation of the internet.

The radio access network handles the core network's radio communications with the user. In Figure 1.1, there are actually two separate radio access networks, namely the GSM EDGE radio access network (GERAN) and the UMTS terrestrial radio access network (UTRAN). These use the different radio communication techniques of GSM and UMTS, but share a common core network between them.

The user's device is known officially as the user equipment (UE) and colloquially as the mobile. It communicates with the radio access network over the air interface, also known as the radio interface. The direction from network to mobile is known as the downlink (DL) or forward link and the direction from mobile to network is known as the uplink (UL) or reverse link.
A mobile can work outside the coverage area of its network operator by using the resources from two public land mobile networks: the visited network, where the mobile is located and the operator's home network. This situation is known as roaming.

Architecture of the Radio Access Network

In the figure the radio access network of UMTS is shown. The most important component is the base station, which in UMTS is officially known as the Node B. Each base station has one or more sets of antennas, through which it communicates with the mobiles in one or more sectors. As shown in the diagram, a typical base station uses three sets of antennas to control three sectors, each of which spans an arc of 120°. In a medium-sized country like the United Kingdom, a typical mobile phone network might contain several thousand base stations altogether.

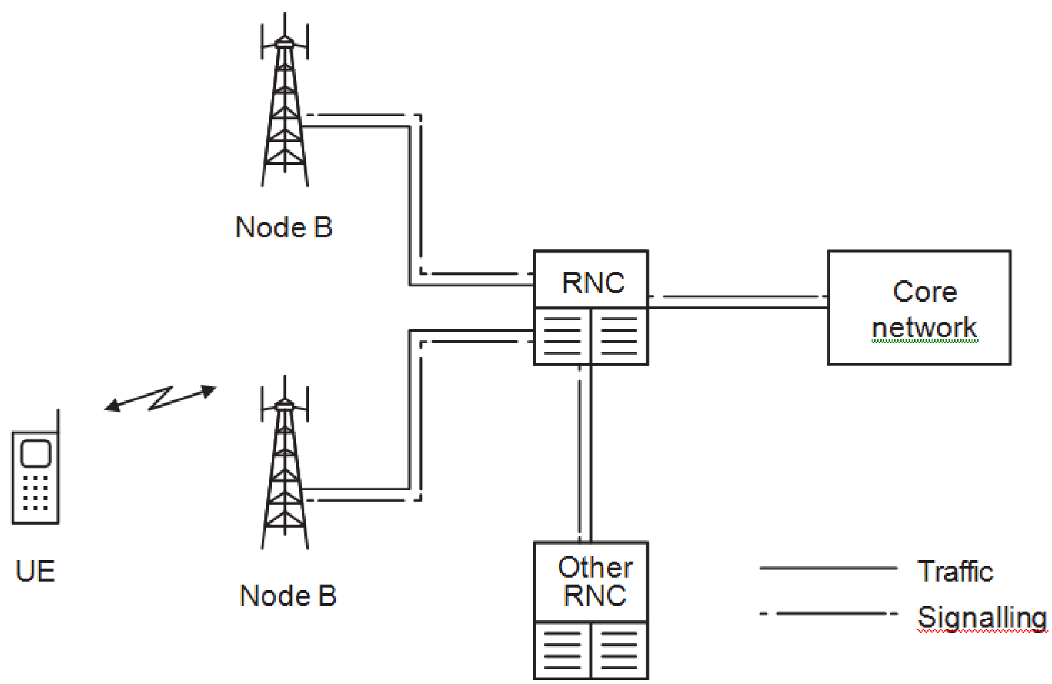


Figure: Architecture of the UMTS terrestrial radio access network

The word cell can be used in two different ways [2]. In Europe, a cell is usually the same thing as a sector, but in the United States, it usually means the group of sectors that a single base station controls. We will stick with the European convention throughout this book, so that the words cell and sector mean the same thing.

Each cell has a limited size, which is determined by the maximum range at which the receiver can successfully hear the transmitter. It also has a limited capacity, which is the maximum combined data rate of all the mobiles in the cell. These limits lead to the existence of several types of cell. Macrocells provide wide-area coverage in rural areas or suburbs and have a size of a few kilometres. Microcells have a size of a few hundred metres and provide a greater collective capacity that is suitable for densely populated urban areas. Picocells are used in large indoor environments such as offices or shopping centres and are a few tens of metres across. Finally, subscribers can buy home base stations to install in their own homes. These control femtocells, which are a few metres across.

Looking more closely at the air interface, each mobile and base station transmits on a certain radio frequency, which is known as the carrier frequency. Around that carrier frequency, it occupies a certain amount of frequency spectrum, known as the bandwidth. For example, a mobile might transmit with a carrier frequency of 1960 MHz and a bandwidth of 10 MHz, in which case its transmissions would occupy a frequency range from 1955 to 1965 MHz. The air interface has to segregate the base stations' transmissions from those of the mobiles, to ensure that they do not interfere. UMTS can do this in two ways. When using frequency division duplex (FDD), the base stations transmit on one carrier frequency and the mobiles on another. When using time division duplex (TDD), the base stations and mobiles transmit on the same carrier frequency, but at different times. The air interface also has to segregate the different base stations and mobiles from each other. We will see the techniques that it uses later on.

When a mobile moves from one part of the network to another, it has to stop communicating with one cell and start communicating with the next cell along. This process can be carried out using two different techniques, namely handover for mobiles that are actively communicating with the network and cell reselection for mobiles that are on standby. In UMTS, an active mobile can actuallycommunicate with more than one cell at a time, in a state known as soft handover.

The base stations are grouped together by devices known as radio network controllers (RNCs). These have two main tasks. Firstly, they pass the user's voice information and data packets between the base stations and the core network. Secondly, they control a mobile's radio communications by means of signalling messages that are invisible to the user, for example by telling a mobile to hand over from one cell to another. A typical network might contain a few tens of radio network controllers, each of which controls a few hundred base stations.

The GSM radio access network has a similar design, although the base station is known as a base transceiver station (BTS) and the controller is known as a base station controller (BSC). If a mobile supports both GSM and UMTS, then the network can hand it over between the two radio access networks, in a process known as an inter-system handover. This can be invaluable if a mobile moves outside the coverage area of UMTS, and into a region that is covered by GSM alone. In Figure 1.2, we have shown the user's traffic in solid lines and the network's signalling messages in dashed lines. We will stick with this convention throughout the book.

Architecture of the Core Network

In the figure the internal architecture of the core network is shown. In the circuit switched domain, media gateways (MGWs) route phone calls from one part of the network to another, while mobile switching centre (MSC) servers handle the signalling messages that set up, manage

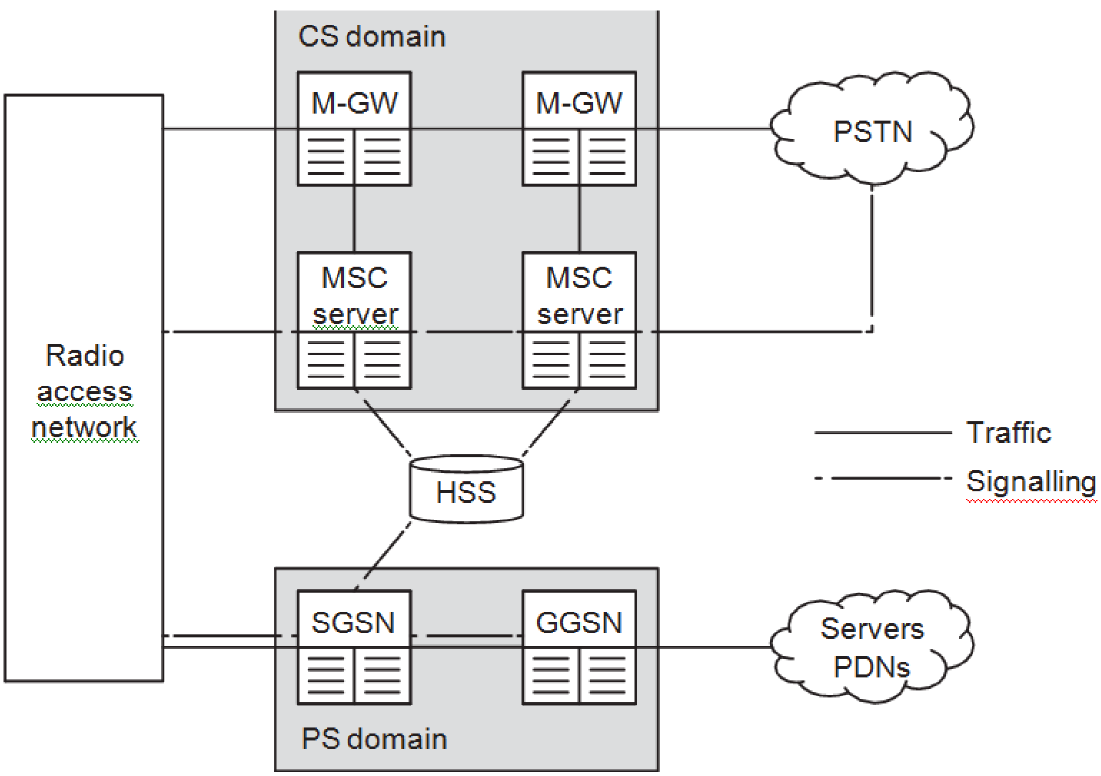


Figure: Architecture of the core networks of UMTS and GSM and tear down the phone calls. They respectively handle the traffic and signalling functions of two earlier devices, known as the mobile switching centre and the visitor location register (VLR). A typical network might just contain a few of each device.

In the packet switched domain, gateway GPRS support nodes (GGSNs) act as interfaces to servers and packet data networks in the outside world. Serving GPRS support nodes (SGSNs) route data between the base stations and the GGSNs, and handle the signalling messages that set up, manage and tear down the data streams. Once again, a typical network might just contain a few of each device.

The home subscriber server (HSS) is a central database that contains information about all the network operator's subscribers and is shared between the two network domains. It amalgamates the functions of two earlier components, which were known as the home location register (HLR) and the authentication centre (AuC).

Communication Protocols

In common with other communication systems, UMTS and GSM transfer information using hardware and software protocols. The best way to illustrate these is actually through the protocols used by the internet. These protocols are designed by the Internet Engineering Task Force (IETF) and are grouped into various numbered layers, each of which handles one aspect of the transmission and reception process. The usual grouping follows a seven layer model known as the Open Systems Interconnection (OSI) model.

As an example, let us suppose that a web server is sending information to a user's browser. In the first step, an application layer protocol, in this case the hypertext transfer protocol (HTTP), receives information from the server's application software, and passes it to the next layer down by representing it in a way that the user's application layer will eventually be able to understand. Other application layer protocols include the simple mail transfer protocol (SMTP) and the file transfer protocol (FTP).

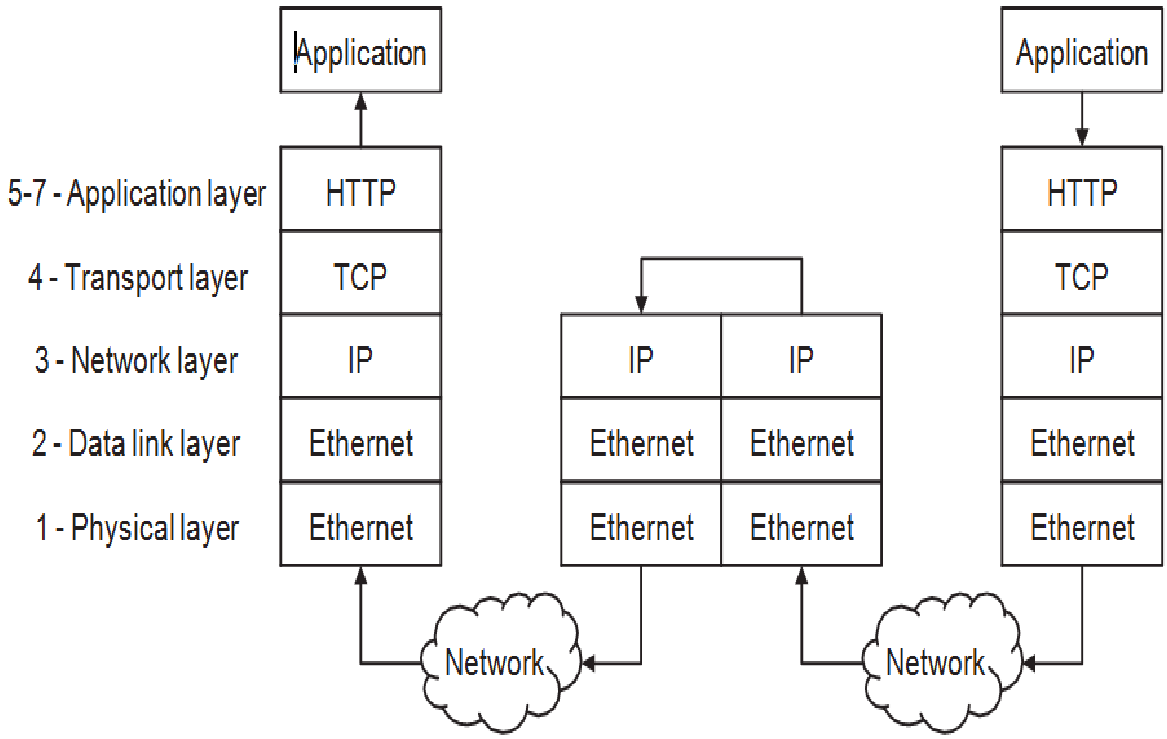


Figure : Examples of the communication protocols used by the internet, showing their mapping onto the layers of the OSI model

The transport layer manages the end-to-end data transmission. There are two main protocols. The transmission control protocol (TCP) re-transmits a packet from end to end if it does not arrive correctly, and is suitable for data such as web pages and emails that have to be received reliably. The user datagram protocol (UDP) sends the packet without any re-transmission and is suitable for data such as real time voice or video for which timely arrival is more important.

In the network layer, the internet protocol (IP) sends packets on the correct route from source to destination, using the IP address of the destination device. The process is handled by the intervening routers, which inspect the destination IP addresses by implementing just the lowest three layers of the protocol stack. The data link layer manages the transmission of packets from one device to the next, for example by re-transmitting a packet across a single interface if it does not arrive correctly. Finally, the physical layer deals with the actual transmission details; for example, by setting the voltage of the transmitted signal. The internet can use any suitable protocols for the data link and physical layers, such as Ethernet.

At each level of the transmitter's stack, a protocol receives a data packet from the protocol above in the form of a service data unit (SDU). It processes the packet, adds a header to describe the processing it has carried out, and outputs the result as a protocol data unit (PDU). This immediately becomes the incoming service data unit of the next protocol down. The process continues until the packet reaches the bottom of the protocol stack, at which point it is transmitted. The receiver reverses the process, using the headers to help it undo the effect of the transmitter's processing.

This technique is used throughout the radio access and core networks of UMTS and GSM. We will not consider their protocols in any detail at this stage; instead, we will go straight to the protocols used by LTE .

Downlink Resource Scheduling in an LTE System

Orthogonal Frequency Division Multiplexing (OFDM) is an attractive modulation technique that is used in a variety of communication systems such as Digital Subscriber Lines (DSLs), Wireless Local Area Networks (WLANs), Worldwide Interoperability for Microwave Access (WiMAX) Andrews et al. (2007), and Long Term Evolution (LTE) cellular networks. In order to exploit multiuser diversity and to provide greater flexibility in resource allocation (scheduling), Orthogonal Frequency Division Multiple Access (OFDMA), which allows multiple users to simultaneously share the OFDM sub-carriers, can be employed. The problem of power and sub-carrier allocation in OFDMA systems has been extensively studied, e.g. Liu & Li (2005); Wunder et al. (2008), and the references therein.

What distinguishes packet scheduling in LTE from that in earlier radio access technologies, such as High Speed Downlink Packet Access (HSDPA), is that LTE schedules resources for users in both the time domain (TD) and the frequency domain (FD) whereas HSDPA only involves the time domain. This additional flexibility has been shown to provide throughput and coverage gains of around 40% Pokhariyal et al. (2006).Because packet scheduling for LTE involves scheduling users in both TD and FD, various TD and FD schemes have been proposed in Pokhariyal et al. (2006)-Monghal et al. (2008). Assume that we have packets for Nusers users waiting in the queue and that resources can only be allocated at the beginning of a pre-defined time period known as the Transmission Time Interval (TTI) or scheduling period. In TD scheduling, U users from the total of N_{users} users are selected based on some priority metric. After the U users have been selected, appropriate subcarrier frequencies and modulation and coding schemes (MCSs) are then assigned by the FD scheduler. Note that the metrics used for TD and FD scheduling can be different in order to provide a greater degree of design flexibility. Examples of TD/FD scheduling metrics have been proposed in Kela et al. (2008); Monghal et al. (2008). In order to make good scheduling decisions, a scheduler should be aware of channel quality in the time domain as well as the frequency domain. Ideally, the scheduler should have

knowledge of the channel quality for each sub-carrier and each user. In practice, due to limited signalling resources, sub-carriers in an OFDMA system are often allocated in groups. On the downlink in LTE systems, sub-carriers are grouped into resource blocks (RBs) of 12 adjacent sub-carriers with an inter sub-carrier spacing of 15 kHz Dahlman et al. (2008); Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8) (2007).
Each RB has a time slot duration of 0.5 ms, which corresponds to 6 or 7 OFDM symbols depending on whether an extended or normal cyclic prefix is used. The smallest resource unit that a scheduler can assign to a user is a Scheduling Block (SB), which consists of two consecu-tive RBs, spanning a sub-frame time duration of 1 ms Dahlman et al. (2008); Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8) (2007) (see Fig. 1).
From the perspective of downlink scheduling, the channel quality is reported by the user via a Channel Quality Indicator (CQI) over the uplink. If a single CQI value is used to convey the channel quality over a large number of SBs, the scheduler may not be able to distinguish the quality variations within the reported range of subcarriers. This is a severe problem for highly frequency-selective channels. On the other hand, if a CQI value is used to represent each SB,
many CQI values may need to be reported back, resulting in a high signalling overhead. A number of CQI reporting schemes and associated trade-offs are discussed in Kolehmainen (2008).

Given a set of CQI values from different users, the multiuser scheduling problem in LTE involves the allocation of SBs to a subset of users in such a way as to maximize some objective function, e.g. overall system throughput or other fairness-sensitive metrics. The identities of the assigned SBs and the MCSs are then conveyed to the users via a downlink control channel.
Studies on LTE-related scheduling have been reported in Kwan et al. (2008); Liu et al. (2007); Ning et al. (2006); Pedersen et al. (2007); Pokhariyal et al. (2007) and the references therein. As pointed out in Jiang et al. (2007), the type of traffic plays an important role in how scheduling should be done. For example, Voice-over IP (VoIP) users are active only half of the time. Also, the size of VoIP packets is small, and the corresponding inter-arrival time is fairly constant. While dynamic scheduling based on frequent downlink transmit format signalling and uplink CQI feedback can exploit user channel diversity in both frequency and time domains, it requires a large signalling overhead. This overhead consumes time-frequency resources, thereby reducing the system capacity. In order to lower signalling overhead for VoIP-type traffic, persistent scheduling has been proposed Discussion on Control Signalling for Persistent

Scheduling of VoIP (2006); Persistent Scheduling in E-UTRA (2007). The idea behind persistent scheduling is to pre-allocate a sequence of frequency-time resources with a fixed MCS to a VoIP user at the beginning of a specified period. This allocation remains valid until the user receives another allocation due to a change in channel quality or an expiration of a timer. The main disadvantage of such a scheme is the lack of flexibility in the time domain. This shortcoming has led to semi-persistent scheduling which represents a compromise between rigid persistent scheduling on the one hand, and fully flexible dynamic scheduling on the other. In semi-persistent scheduling, initial transmissions are persistently scheduled so as to

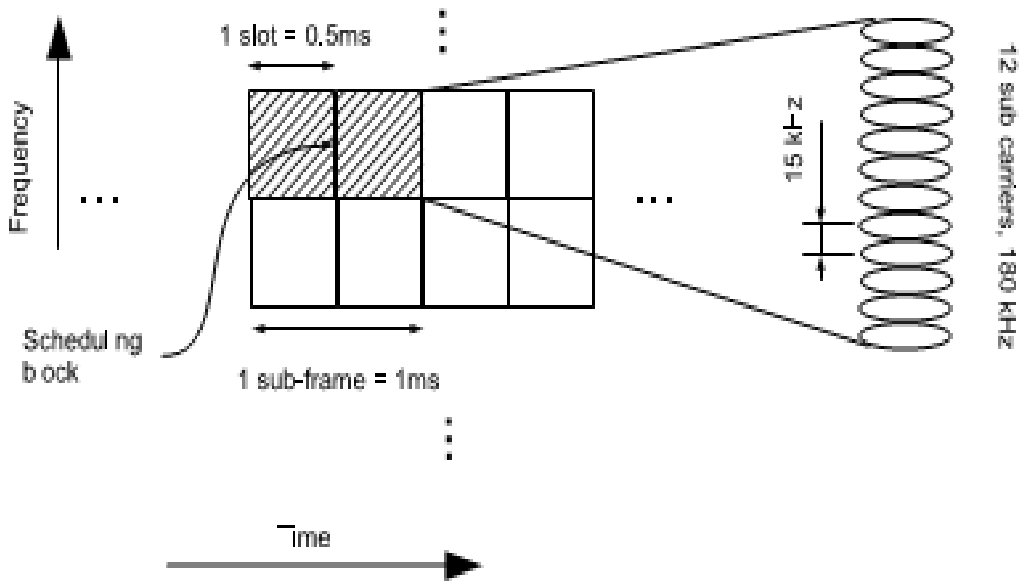


Figure: LTE downlink time-frequency domain structure.

reduce signalling overhead and retransmissions are dynamically scheduled so as to provide adaptability. The benefits of semi-persistent scheduling are described in Jiang et al. (2007). An important constraint in LTE downlink scheduling is that all SBs for a given user need to use the same MCS in any given TTI[1] Dahlman et al. (2008). In the rest of this chapter, we focus on the FD aspect of dynamic scheduling. Specifically, the challenging problem of multiuser FD scheduling is formulated as an optimization problem, taking into account this MCS restriction. Simpler sub-optimal solutions are also discussed.

PROCEDURE OF DOWNLINK SCHEDULING

The per-RB metrics' comparison that serves as the transmission priority of each user on a specific RB is taken into account for resource allocation for each UE. For example the k-th RB is allocated to the j-th user if its metric $m_{j;k}$ is the largest one among all i-UEs, i.e., if it satisfies the equation:

$m_{j;k} = \max_i \{m_{i;k}\}$…. (1)

The whole process of downlink scheduling can be divided in a sequence of operations that are repeated, in general, every TTI (see fig.1):

1) The Evolved Node B prepares the list of flows which can be scheduled in the current TTI .Flows could be formulated only if there are packets to send at MAC layer and UE at receiving end is not in the idle state.

2) Each UE decodes the reference signals, reports CQI (Channel Quality Indicator) to eNB which helps to estimate the downlink channel quality. The eNB can configure if the CQI report would correspond to the whole downlink bandwidth or a part of it which is called sub-band.

3) Then the chosen metric is computed for each flow according to the scheduling strategy using the CQI information. The sub-channel is assigned to that UE that presents the highest metric.

4) For each scheduled flow, the eNB computes the amount of data that will be transmitted at the MAC layer i.e. the size of transport block during the current TTI. The AMC (Adaptive Modulation and Coding module) at MAC layer selects the best MCS (Modulation and Coding Scheme) that should be used for the data transmission by scheduled users. Link adaptation involves tailoring the modulation order (QPSK, 16-QAM, 64-QAM) and coding rate for each UE in the cell, depending on the downlink channel quality.

5) Physical Downlink Control Channel (PDCCH) is used to send the information about the users, the assigned Resource Blocks, and the selected MCS to terminals in the form of DCI (Downlink Control Information).

6) Each UE reads the PDCCH payload .If a particular UE has been scheduled; it will try to access the proper PDSCH

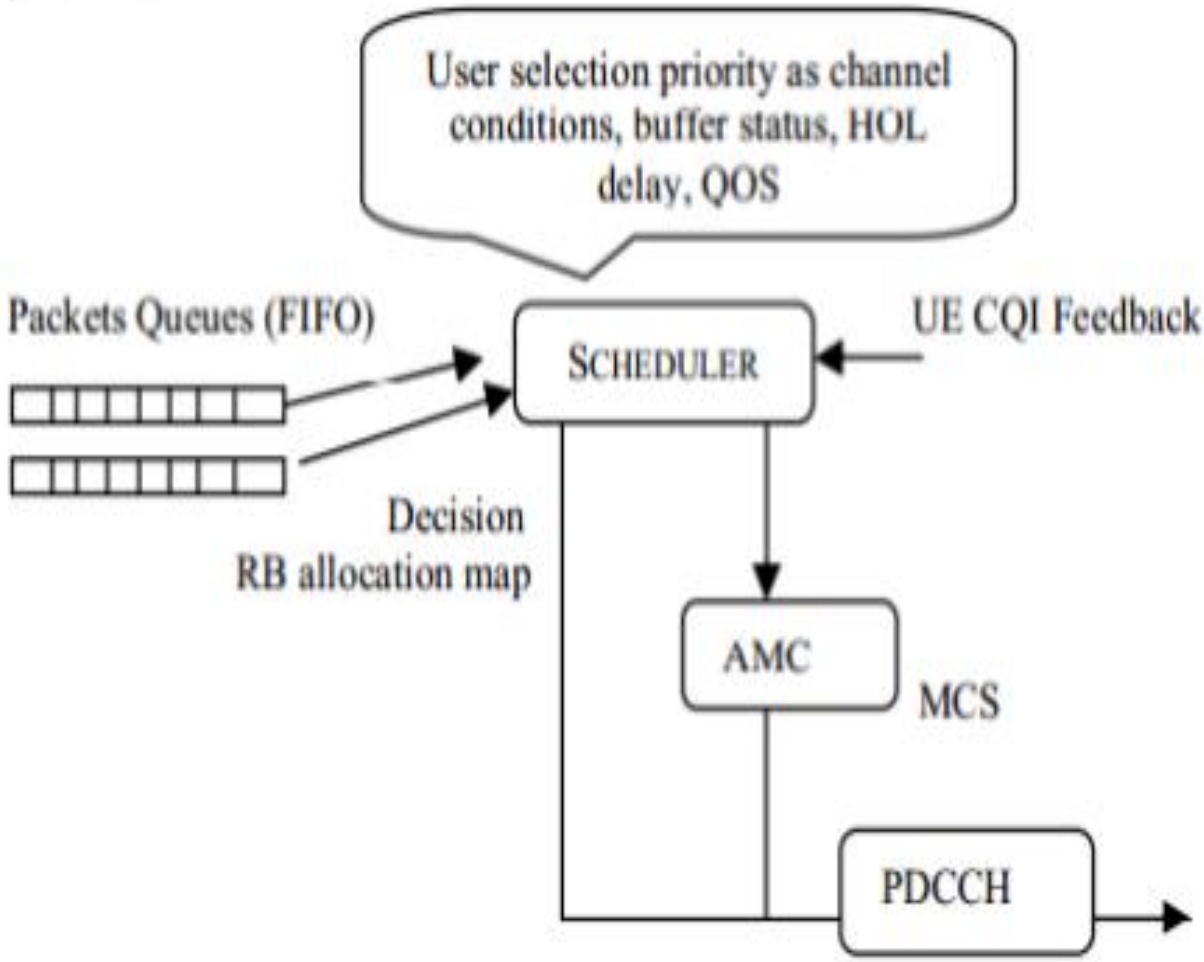


Fig.1. General Model of packet scheduler

The users are prioritized by packet scheduler on the basis of a scheduling algorithm being used. These algorithms while making scheduling decisions, takes into account the instantaneous or average channel conditions, Head of Line (HOL) packet delays, status of receiving buffer or type of service being used .

DOWNLOINK RESOURCE SCHEDULING ALGORITHMS:

Generally, scheduling can be divided into two classes: channel-independent scheduling and channel-dependent scheduling. The performance of channel independent scheduling can never be optimal due to varying nature of the instantaneous channel condition. On the contrary, channel-dependent scheduling can achieve better performance by allocating resources based on channel conditions with optimal algorithms. Channel dependent schedulers can further be classified on the basis of QOS support as QOS unaware or QOS aware channel dependent scheduling algorithms. Apart from these semi-persistent and energy aware solutions also exist in the literature.

A. Channel Independent Scheduling Strategies

Channel independent strategies were firstly introduced in wired networks and are based on the assumption of time-invariant and error-free transmission media. Being unrealistic for LTE networks, they are typically used in conjugation with channel-dependent strategies to improve system performance.

1) First in First out (FIFO):

Though FIFO is the simplest of all possible scheduling disciplines but it is inefficient and unfair. This scheduler serves the packets in the queue in order of arrival and when the queue is full, it drops the packets that are just arriving. The major setback is that it cannot differentiate among connections; therefore all packets experience the same delay, jitter and packet loss irrespective of which packet it is.

The metric of i-th user on the k-th RB can be translated from its behaviour as:

$$m_{i,k}^{FIFO} = t - T_i \quad \text{.......} (2)$$

Where t is the current time and T_i is the time instant when request was issued by i-th user.

2) Round Robin:

Round Robin allocates resources to each UE, completely neglecting channel quality or data rate. Initially, the terminals are ordered randomly in a queue. The new terminals are inserted at the end of the queue. The first terminal of this queue is assigned all the available resources by scheduler, and then put it at the rear of the queue. The rest of steps follow the same way, until no terminal applies for resources. Round Robin (RR) metric is similar to the one defined for FIFO. The only difference is that, in this case, Ti refers to the last serving time instant of the user.

On one hand, it seems to be a fair scheduling, since every terminal is given the same amount of resources. On the other hand, it neglects the fact that certain terminals in bad channel conditions need more resources to carry out the same rate, so it is absolutely unfair. This scheme is impractical in LTE because different terminals have different service with different QoS requirements.

3) Weighted Fair Queuing:

In Weighted Fair scheduling introduced in,the packets are grouped into various queues. A weight is assigned to each queue which determines the fraction of the total bandwidth available to the queue. In this case, a specific weight (wi) is associated to the i-th user (or class of users) and then it is used to modify Round-Robin metric as:

$$m_{i,k}^{WFQ} = w_i \cdot m_{i,k}^{RR} \quad \ldots\ldots (3)$$

To assure that flows with larger packets are not allocated more bandwidth than flows with smaller packets, it also supports variable-length packets. The Weighted Fair scheduling assigns the bandwidth for each service based on the weight assigned to each queue and not based on the number of packet.

4) Blind Equal Throughput:

The Blind equal throughput (BET) is a channel unaware strategy that aims at providing throughput fairness among all the users. To counteract the unfair sharing of the channel capacity, the BET scheduler user priority metric which considers past average user throughput as follows:

$$m_{i,k}^{BET} = \frac{1}{\overline{R^i}(t-1)} \quad \text{................(4)}$$

Where $\overline{R^i}(t-1)$ is the average throughput of terminal i over windows in the past.

The smoothed value of $R^i(t)$ is computed using any weight moving average formula, e.g.,

$$\overline{R^i(t)} = \left(1 - \frac{1}{T}\right)\overline{R^i}(t-1) + \frac{1}{T}R^i\left(t\right) \quad \text{.....................(5)}$$

Where $R^i(t)$ is the instantaneous value of data rate at time instant t .

It is clear from equation (4) that the BET scheduler prioritizes users with lower average throughput in the past.This implies that users with bad channel conditions are allocated more resources compared to the users with good channel conditions. Thus throughput fairness is achieved but at the cost of spectral efficiency.

5) Largest Weighted Delay First:

To avoid packet drops, it is required that each packet has to be received within a certain delay deadline in Guaranteed delay services. It incorporates the information

about the specific packet timing, when the packet was created and its deadline while calculating the priority metric. For Real-Time flow, its metric is calculated as

$$m_{i,k}^{MLWDF} = \alpha_i \cdot D_{HOL,i} \quad \text{where} \quad \alpha_i = -\frac{\log \delta_i}{\tau_i} \quad \text{..........(6)}$$

Where $D_{HOL,i}$ is waiting time of the packet at the head of the line and δ_i represents drop probability and τ_i defines target delay for the i-th user.

Similar to Round Robin, neglecting channel conditions leads to poor throughput in LWDF

B. Channel Dependent/QOS unaware Scheduling Strategies

Channel-dependent scheduling strategies allocate resources with optimal algorithms by taking into consideration the channel conditions. The user channel quality can be estimated from CQI reports which help the scheduler to estimate the

channel quality perceived by each UE and serves as an indication of the data rate which can be supported by the downlink channel.

The following text discusses the the channel dependent but unaware strategies that exist in the literature:

1) Maximum Throughput:

Being a channel dependent scheduling, Max Throughput takes advantage of multiuser diversity to carry out maximum system throughput. First, scheduler analyzes CQI reports from UEs to obtain data rate of an identical sub channel for different terminals. This information can be used in the priority metric to prioritize

users with good channel conditions over users with bad channel conditions. Thus scheduler assigns the resource to the user which can achieve the highest throughput in this sub-channel based on SNR. The priority metric for the MaxT scheduler is given as follows [8]:

$$m_{i,k}^{Max-T} = d_k^i(t) \quad \text{……………..(7)}$$

Where $d_k^i(t)$ is the expected data-rate for i-th user at time t on the k-th Resource-block. It can be calculated by considering the Shannon expression for the channel capacity as:

$$d_k^i(t) = \log\left[1 + SINR_k^i(t)\right] \quad \text{………………(8)}$$

MaxT performs unfair resource sharing of the resources since it aims at blind maximization of throughput only.

2) Proportional Fair:

The Proportional Fair (PF) algorithmcan improve the fairness among users without losing the efficiency in terms of average (or aggregate) throughput. The terminals are ranked according to the priority function which is defined as the ratio of the instantaneous to average throughput.. Then scheduler assigns resources to terminal with highest priority. Repeat the last two steps until all the resources are used up or all the resources requirements of terminals are satisfied [10]-[13].

The PF was designed specifically for the Non-Real Time class and hence does not assure any QoS requirement such as delay, jitter and latency. The preference metric or priority function is obtained by merging the metrics of MaxT and BET and is given as:

$$m_{i,k}^{PF} = m_{i,k}^{Max-T} \cdot m_{i,k}^{BET} = \frac{d_k^i(t)}{\overline{R^i}\left(t-1\right)} \quad \text{...............(9)}$$

$d_k^i\left(t\right)$ is the estimation of supported data rate of terminal i for the resource block k. $\overline{R^i}\left(t-1\right)$ is the average data rate of terminal i over a windows in the past.

T_{PF} is the windows size of average throughput and can be adjusted to maintain fairness. Normally T_{PF} should be limited in a reasonable range so that terminals cannot notice the quality variation of the channels.

If $R^i(t)$ is the instantaneous value of data rate a time instant t ,then if i-th terminal is selected

$$\overline{R^i(t)} = \left(1 - \frac{1}{T_{PF}}\right) \cdot \overline{R_i}\left(t-1\right) + \frac{1}{T_{PF}} \cdot R^i\left(t\right) \quad \text{...........(10)}$$

If i-th terminal is not selected, then

$$\overline{R^i\left(t\right)} = \left(1 - \frac{1}{T_{PF}}\right) \cdot \overline{R_i}\left(t-1\right) \quad \text{.................(11)}$$

Rabie K. Almatarneh et. al. evaluated the performance of two dimensional (time slot and frequency subcarrier) PF scheduling in OFDMA wireless systems ; both analytically and by simulation. Closed-form expressions for the average throughput and Jain's fairness index as the performance metrics; have been derived. The algorithm performance is investigated for a broad range of the traffic load and the number of subbands.

In [13], the approach of PF was formulated as an optimization problem in order to maximize the achieved throughput of a LTE system. Here, a multiuser scheduler with PF is proposed. A suboptimal PF scheduler, which has a much lower complexity at the cost of some throughput degradation, is also proposed. Numerical results show that the proposed PF scheduler provides a superior fairness performance with a modest loss in throughput, as long as the user average SINRs are fairly uniform.

3) Throughput to Average:

Throughput to Average (TTA) scheduling algorithm [10] tries to divide the available resources between all users with the priority metric:

$$m_{i,k}^{TTA} = \frac{d_k^i(t)}{d^i(t)} \quad \text{............(12)}$$

The above metric performs averaging of resources evenly between the users. Here, the achievable throughput in the current TTI is used as normalization factor of the achievable throughput on the considered k-th RB. It is evident from its metric that the higher the overall expected throughput of a user is, the lower will be its metric on a single Resource Block.

\

C. Channel Dependent/QOS aware Scheduling Strategies

In LTE, QoS differentiation is managed by associating a set of QoS parameters to each flow. Minimum required performance can be guaranteed by the scheduler if it knows the values of QOS parameters, either in terms of guaranteed data rates or of delivery delays.

In this subsection, a comprehensive overview on QoS-aware solutions presented in literature for LTE systems is presented.

1) Schedulers for Guaranteed Data Rate:

G. Monghal et.al. in [14] proposed QOS oriented Time and frequency domain schedulers that focus on GBR considerations. The proposed Time Domain Priority Set Scheduler (TDPSS) has been devised to select users with the highest priority. Users are separated into two sets. Set 1's users with bit rate below target bit rate are managed by using BET and prioritized over all the other users which form Set 2. Furthermore, within each set; prioritization is according to priority metrics.While TD- PSS will tend to maintain the throughput of low signal quality users to Target Bit Rate, Frequency Domain- PF i.e PF scheduled (PFsch) will tend to reduce their allocation share in the frequency domain with priority metric given as:

$$m_{i,k}^{PFsch} = \frac{d_k^i(t)}{\overline{R_{sch}^i}\left(t-1\right)} \quad \ldots\ldots\ldots\ldots\ldots(13)$$

Where $\overline{R_{sch}^i}(t-1)$ is similar to the past avg throughput defined in eq. (4), with the difference that it is updated only when the i-th user is actually served.

In [15] a Dynamic Hybrid Scheduler (DHS) composed by two basic components, corresponding to a guaranteed and a dynamic delay based rate allocation policy respectively is presented. Used priorities are calculated, for the i-th user, as:

$$P_i = \frac{D_{HOL,i}}{\tau_i} \quad \text{................(14)}$$

It is important to note that the transmission of the head of line packet becomes more urgent, when the value of Pi is increased. To attain the guaranteed bit-rate, the resources are allocated to the user with the highest priority. The user with second highest priority is considered thereafter for allocation in case the RBs are left free and so on.

A similar approach is followed in [16] by Y. Zaki et.al.. In order to simplify the LTE MAC scheduling, two stages have been defined: Time Domain (TD) and Frequency Domain (FD) schedulers. The TDPS differentiates the users according to their QoS characteristic whereas FDPS assigns the RBs between the priority users. Based on the QoS Class Identifier (QCI), the incoming packets are categorized upon their priority sets are classified as GBR and non

GBR set. After this step, the FDPS orderly assigns the best RB to each user in the GBR set, updating the achieved bitrate. When all users in the list have reached their target bit-rate, if RBs are still available, the scheduler assigns them to users in the non-GBR list using PF metric.Thus, all these approaches use ordered lists to prioritize the most delayed flows and to achieve their target bit-rate.

2) Schedulers for Guaranteed Delay Requirements:

Real-Time flows have more strict delay restraint than Non-Real-Time flows resulting in the reduction of influence of error correction. Scheduling strategies that aim to guarantee bounded delay fall in the category of the QoS-aware schemes.

Herein, QOS aware algorithms present in the literature that makes use of per-RB metrics are described.

The Modified Largest Weighted Delay First (M-LWDF) [17] combines both channel conditions and the state of the queue with respect to delay in making scheduling decisions. It ensures that the probability of delay packets does not exceed the discarded bound below the maximum allowable packet loss ratio i.e.

$$\Pr\left(D_{HOL,i} \succ \tau_i\right) \leq \delta_i$$

The scheduler allocates resources to the user with the maximum priority index which is made up of the product of the HOL packet delay of the user, the channel capacity with respect to flow and the QoS differentiating factor:

$$m_{i,k}^{MLWDF} = \alpha_i \cdot D_{HOL,i} \cdot \frac{d_k^i(t)}{R^i(t-1)} \quad \text{............(15)}$$

Where DHOL,i is waiting time of the packet at the head of the line and αi = - log δi /τi ;δi represents acceptable packet loss rate (i.e. the maximum probability for HOL packet delay of user i to exceed the delay threshold of user i.) and τi defines

Delay threshold for the i-th user.

EXP/PF is a QOS aware extension of PF that can support both Non-Real Time and Real Time flows at the same time [18].For real-time flows the metric is calculated as:

$$m_{i,k}^{\frac{EXP}{PF}} = \exp\left(\frac{\alpha_i \cdot D_{HOL,i} - \chi}{1+\sqrt{\chi}}\right) \cdot m_{i,k}^{PF} \quad \text{............(16)}$$

Where $\chi = \frac{1}{N_{RT}} \cdot \sum_{i=1}^{N_{RT}} D_{HOL,i}$ and N_{RT} is the number of active downlink real time flows The metric when calculated for Non-real time flows is given as:

$$m_{i,\mathrm{k}}^{EXP/PF} = \frac{w(t)}{M(t)} \square \frac{d_k^i(t)}{\overline{R^i}(t-1)} \text{...........(17)}$$

$$\text{where } w(t)= \begin{cases} w(t-1)-\varepsilon \text{.........} (D_{HOL})_{\max} > \tau_{\max} \\ w(t-1)+ \varepsilon / p \text{.......} (D_{HOL})_{\max} < \tau_{\max} \end{cases}$$

M(t) is the average number of RT packets waiting at e-Node B buffer at time t, ε andp are constants, (DHOL)max is themaximum HOL packet delay of all RT service users and τ isthe maximum delay constraint of RT service users. Here, RT users are prioritized over NRT users when their HOL packet delays are approaching the delay deadline. The exponential term is closer to 1 if HOL delays of all users are about the

same. Thus above formula becomes Proportional Fair. If one of the user's delays becomes large, the exponential term in will override the left term in (16) and dominate the selection of a user.

EXP rule [19] can be considered as modified form of the above mentioned EXP/PF and its priority metric is calculated as:

$$m_{i,k}^{EXP-Rule} = b_i \square \exp\left(\frac{\alpha_i \cdot D_{HOL,i}}{c + \sqrt{(\frac{1}{N_{RT}}) \sum_j D_{HOL,j}}} \right) . \Gamma_k^i \text{ ....(18)}$$

Where Γ_k^i represents Spectral efficiency for the user i over the k-th RB and the optimal parameter set according to [19] is:

$$\{a_i \in [(5/0.99\tau_i, 10/0.99\tau_i)],$$
$$b_i = \frac{1}{E[\Gamma^i]} \, and$$
$$c = 1$$

In [20] performance of Exponential Rule is evaluated in comparison to PF Scheduler and MLWDF .A variant of the Exponential rule i.e. EXPQW is also proposed which assigns weights to the subscriber stations based on their queue length and waiting time. Three hierarchical schedulers which use a combination of the exponential rule for waiting time and queue-length and other scheduling rules have also been presented. The results indicate that EXPQW and the hierarchical schedulers have comparable throughput and fairness values with algorithms like PF and MLWDF in moderately loaded and heavily loaded scenarios.

In [21] M. Iturralde et. al. proposed a two level resource allocation scheme is to enhance the QoS for multimedia services. It corresponds to a procedure that combines cooperative game theory, a virtual token mechanism, and the EXP-RULE algorithm. It works in two phases: in the first one the game is run to partition available RBs among different groups of flows, populated depending on the type of application they carry. The second phase uses of EXP rule modified by using a virtual token mechanism in order to meet bounded delay requirements and to guarantee, at the same time, a minimum throughput to all flows. In this way, a significant performance gain over the EXP rule is achieved in terms of both packet loss rate and fairness.

LOG Rule algorithm has been described in [22]. For the LOG rule, the preference function is calculated as:

$$m_{i,k}^{LOG-Rule} = b_i . \log(\mathrm{c} + \alpha_i D_{HOl,i}) . \Gamma_k^i \quad \text{.............(19)}$$

Where bi, c, and ai are tuneable parameters; represents the spectral efficiency for the i-th user on the k-th RB. Optimal parameters are given as :

$$b_i = 1 \Big/ E[\Gamma^i], c = 1.1, and$$

$$a_i = 5 \Big/ 0.99\tau_i$$

Prio et al. proposed a two-level downlink scheduling for real-time flows in LTE networks [23]. At the highest level, a discrete time linear control law is applied every LTE frame. The total amount of data that real-time flows should transmit in 10 ms is thus pre-calculated while considering their delay constraints. When FLS completes its task, the lowest layer scheduler works every TTI. The lower PF algorithm allocates radio resources by considering bandwidth requirements of FLS to flows hosted by UEs experiencing the best channel quality. In particular, the lowest layer scheduler decides the number of TTIs/RBs (and their position in the time/frequency domains) in which each real-time source will actually transmit

Its packets. The resources left free by real time flows are assigned to NRT flows.

D. Dynamic and Semi-resistent Scheduling for VoIP support

Dynamic packet scheduling for VoIP traffic in the LTE Downlink is presented in [24][25]. In [24] the aim is to optimize the performance of dynamic scheduling when mix of VoIP traffic and best effort flows are available .The proposed algorithm is divided into time domain and frequency domain packet schedulers. At every TTI ,scheduler called as Required Activity Detection with Delay Sensitivity(RAD-DS)prioritizes each schedulable user according to the time domain metric M^{TD} [n, t],which is combination of 3 metrics given as:

$$M^{TD}[n,t] = m[n,t] \cdot \mathrm{RA}^{traf}[n,t].DS^{traf}[n,t] \quad \text{..........(20)}$$

$RA^{traf}[n,t]$ (i.e. the required activity) implies the time share required by user n where a user should be scheduled. m [n, t] is a counter incremented every TTI that guarantees some fairness in resource scheduling. Finally, DS^{traf} [n, t] (i.e. the delay sensitivity) function imposes time constraints to users with a delay bound that increases with HOL packet delay.

The frequency domain scheduler allocates Resource Blocks to different users using the Proportional Fair scheduled (PFsch) metric.

To support high number of VoIP flows, semi-persistent allocation solutions (generally considered as channel-independent schemes) aim at increasing the VoIP capacity of the network by maximizing the number of supported VoIP calls. One such scheme presented in [26], improves the VoIP capacity of the network with the use of semi-persistent scheme. Here, the radio resources are divided in several groups of RBs. Each pre-configured block is associated only to certain users. Furthermore, RB groups are associated to each user in contiguous TTIs. Resource allocation of each RB group to the associated UEs is performed dynamically. The proposed scheme reduces the control overhead with respect to the dynamic scheduling. Semi-persistent schemes for VOIP have also been proposed in [27] [28].

E. Energy Aware Solutions :

Energy consumption is heavy in LTE due to tremendous processing load on UE. Energy conserving solutions curb energy waste and hence extend the battery life ofUE among which Discontinuous reception (DRX) is useful. In DRX,When there are no data transmissions; UE turn off its radio equipment UE to save energy. In [29], the light sleeping mode is introduced to further improve the performance of DRX for QOS -aware traffic. The key idea is to turn off the power amplifier .Other components in transceiver cut down their power consumption while allowing fast wakeup. Proposed scheme reduces energy consumption while satisfying the delay constraints.

In [30], the impact of different scheduling schemes from an energy efficiency point of view is analyzed. It is demonstrated that the MaxT scheme is more energy efficient than both PF and RR. In scenarios with low traffic load, Bandwidth Expansion Mode algorithm is used for achieving energy savings for the eNB

[31]. The eNB transmission power is reduced by assigning a coding scheme with lower rate to each user. Consequently their spectrum occupation is expanded.

All aspects and targets of scheduling strategies discussed in this subsection, as well as parameters they use for computing scheduling metrics, have been summarized in Table II.

TABLE II
PARAMETERS USED BY EACH SCHEDULER

Author/Year	Scheduling Strategy	Requested Bit rate	SINR	Average Data Rate or Throughput	HOL Packet Delay	Target Delay	Max PLR	Queue Size
P. Kela *et. al.* (2008)	MaxT [8]		X					
Rabie K. Almatarneh(2010)	PF [12]		X	X				
R. Kwan *et. al.* (2009)	PFMultiuser [13]		X	X				
G. Monghal *et. al.* (2008)	PSS/PFsch [14]	X	X	X				
D. Skoutas *et. al.* (2010)	[15]	X	X					
Y. Zaki *et. al.* (2011)	[16]	X	X					
H. Ramli *et. al.* (2009)	MLWDF [17]		X	X	X	X	X	
R. Basukala *et. al.* (2009)	EXP/PF [18]		X	X	X	X	X	
M. Iturralde *et. al.* (2012)	[21]		X	X	X	X		
B. Sadiq *et. al.* (2009)	EXP Rule [19]		X	X	X	X		
Seung Jun Baek *et. al.* (2011)	LOG Rule [19][22]		X	X	X	X		
G. Piro *et. al.* (2011)	FLS [23]					X		X

Our proposal:

Aim:

The aim our scheduling algorithm is to achieve the following critera

- Ensure fairness
- Better throughput
- Comparatively good SINR
- Buffer status fairness
- Better QOS

Algorithm:

Our proposed algorithm consists of two terms-

- Weight factor
- PF metric

Weight Factor:

Weight factor is a number assigned to each user depending on their buffer status and priority.

The buffer status of each user is calculated at the beginning of each time slot using the following equation.

$$BF_{m,i} = (1 - \frac{1}{T})\overline{BF_{m,i-1}} + \frac{1}{T} BF_{m,i}$$

At the time of assigning weight factors, two key facts were ensured

- Fairness
- Priority

as users having higher priorities are given higher weight factors, still users having lower priorities might get resources earlier because of their buffer status. The following buffer levels are distributed among each user and different weight factors depending on the priority levels were added to each level.

Index	Buffer Size in Bytes
1	0 < Buffer Size <= 10
2	10 < Buffer Size <= 12
3	12 < Buffer Size <= 14
4	14 < Buffer Size <= 17
5	17 < Buffer Size <= 19
	....
	....
31	967 < Buffer Size <= 1132
32	1132 < Buffer Size <= 1326
33	1326 < Buffer Size <= 1552
34	1552 < Buffer Size <= 1817
	....
	....
58	68201 < Buffer Size <= 79846
59	79846 < Buffer Size <= 93479
60	93479 < Buffer Size <= 109439
61	109439 < Buffer Size <= 128125
62	128125 < Buffer Size <= 150000
63	Buffer Size > 150000

- All these stages are present in each user's cases. But the weight factors for the same buffer ranges are different for different users having different priorities.
- Same buffer status will assign larger weight factor to the higher priority users than the one having lower priority.

PRIORITY:

The priority levels of different users are assigned according to the following chart.

Priority	Application
1	IMS Signaling
2	Voice
3	Real Time Gaming
4	Live Streaming Video
5	Buffered Streaming Video
6	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.
7	Voice, Live Streaming Video, Interactive Gaming
8	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.
9	Buffered Streaming Video, Progressive Video, HTTP, E-mail, Chat, FTP, P2P File Sharing, etc.

Our proposed algorithm overcomes the following criteria

- Clearing packets for small BS(Buffer Size)
- Computation of WF(Weight Factor)
- Overlapping among priorities
- Setting Resource allocation target

CLEARING SMALL PACKETS:

Threshold values for buffer status can be defined for each priority. Users will be taken one after another according to their priority. If the buffer amount is less than the threshold, the resource will be allocated in full on their respective best frequency.

WEIGHT FACTOR ASSIGNMENT

- PF={2,3,4,……….,9}
- PRm PR

$$WF\_Range_{PRm,i} = \{\ \}$$

$$WF\_Range_{PRm,i} = 16$$

\

- 25% overlapping between adjacent weight factor ranges are used

$$WF\_Range_{PRm,i} \bigcap WF\_Range_{PRm-1,i}$$

- Number of overlapping = 4

$$WF\_Range_{9,i} = \{1, 2, 3 \ldots\ldots, 16\}$$

$$WF\_Range_{8,i} = \{13, 14, \ldots\ldots, 28\}$$

So number of overlapping = 4

ASSUMPTIONS

Total allocation period= 5000ms
Total sub frames=5000
Target period=100ms
So total allocation under our allocation period=50 times
Here to find out the request of buffer allocation from different users we have used Poisson distribution, where lambda is taken as packet arrival rate. As an example
1 time slot=10ms or radio frame
Monitor 10 time slots=100 ms = target period for resource distribution

Assumed packet size=0.125kB
Assumed data rate for video streaming = 800kbps =100kBps=800k packets/sec

Average packet per 10ms or radio frame = 8 packets = lambda
To find out different users request we have taken
Lambda (m) = randi ([3, 30], 1)

Though we monitor 10 time slots and then allocate the bits, but for make it more realistic we have taken the arrival in every time slot.

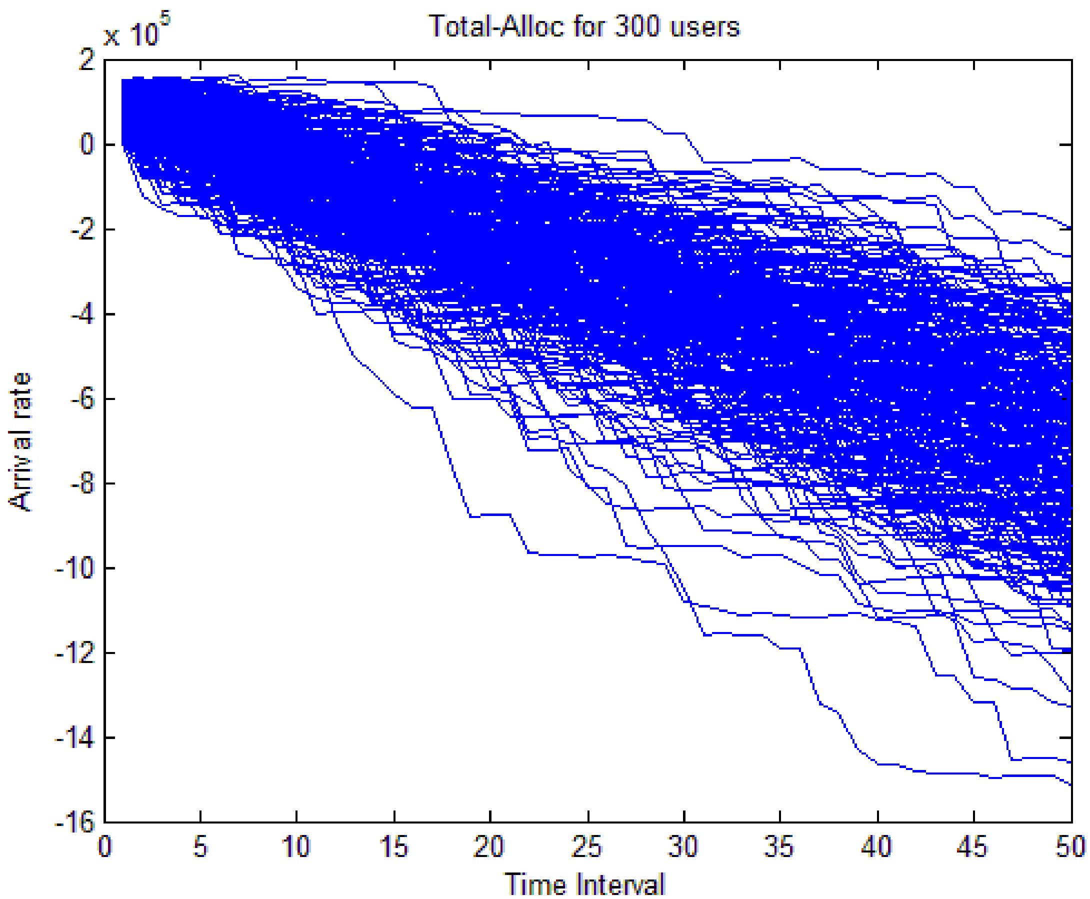
For priority we have taken randomly values from 1 to 9. But because of carrying small packets of IMS signaling and voice they will be cleared in the first step.

Outcomes

- A single algorithm which can clearly justify the major three items of making decisions in an attempt to make appropriate apportionment of the resources.
- Real time analysis for the weighted proportional fair metric.
- Real time analysis for the Proportional Fair Metric.
- Comparison of PF and WPF metric in the case of Queue Aware Scheduling.
- Comparison of PF and WPF metric in the case of priority Aware Scheduling.

Graphical Outcomes:

1.



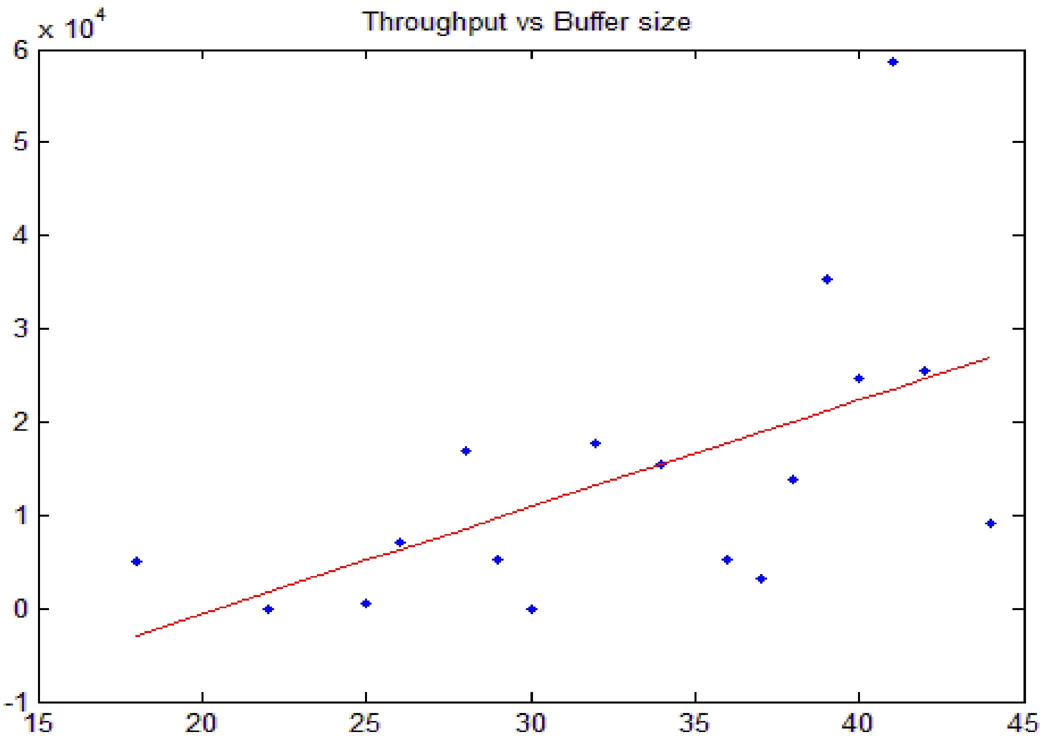
The graph showes total allocation for 300 users.For each user there will be different allocation according to our weighted fair pf metrices. And by implementing allocation in the programme we got the output that is shown in the above figure.

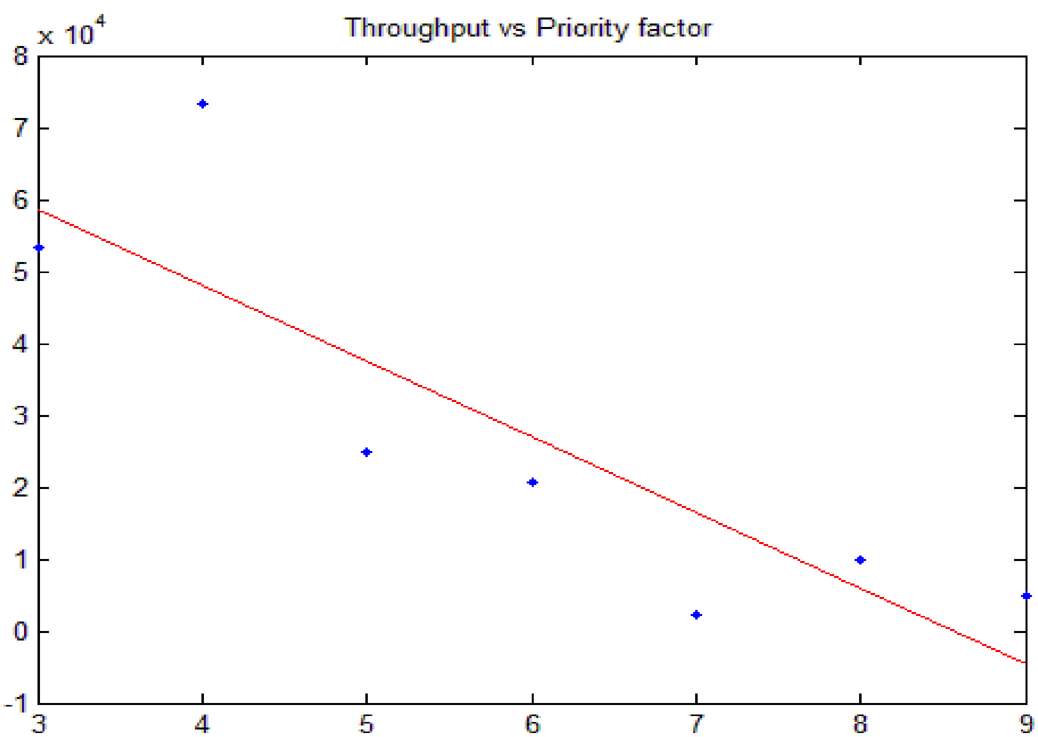
A portion of data that we illustrated in this graph in shown in the next page

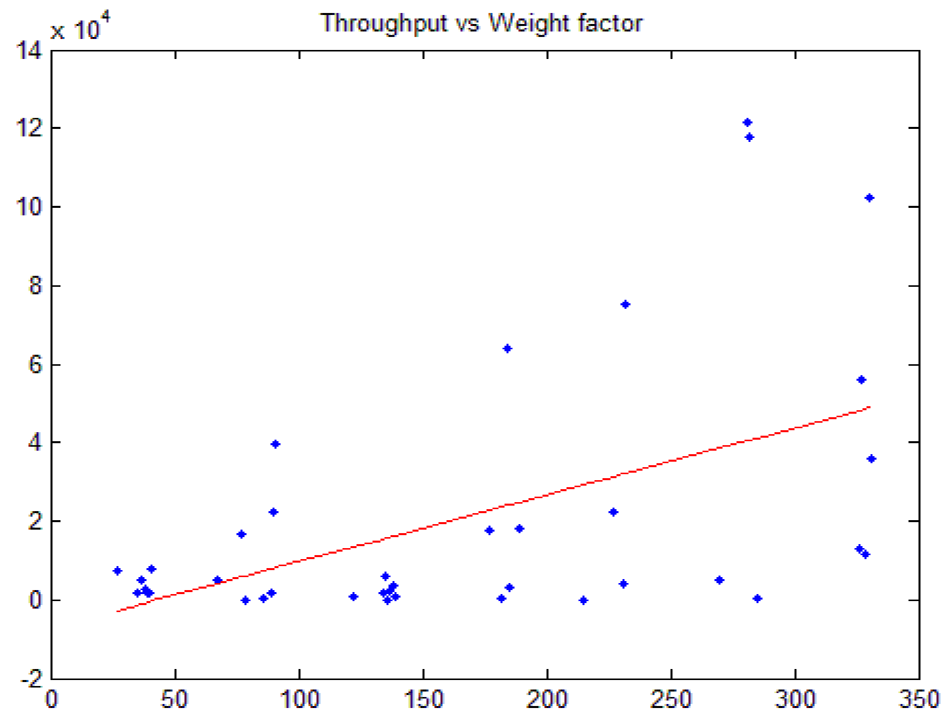
Data:

35150	35500.48	39884.19	43035.68	44259.98	47962.18	50633.68	50482.95
135816	127106.7	130391.8	130668.3	135456.1	136203.2	114322	114322
94390	92198.21	94793.42	100295.4	102727.3	105594.7	106463.1	107237.3
63537	57577.4	52601.84	52422.46	45388.6	49964.07	50665.83	-9405.07
99987	103699	107521.8	51547.47	46668.61	43660.77	44624.58	36428.43
65980	30102.36	29468.34	15871.7	-12395.5	-15417.6	-13640.2	-64589.5
58626	-64793.3	-68968.9	-67724.7	-75004.7	-69334	-103223	-104919
65364	70109.27	-45997.3	-55575.9	-122714	-119230	-116539	-115283
81688	84019.13	85754.16	86499.4	89756.89	84746.69	72973.14	74918.21
126988	129786.2	103738.8	105497.9	35170.85	-22969.1	-35540	-71978.9
826	-7723.21	-81983.5	-76774.1	-71939.2	-69779.1	-71377.9	-224931
93542	78719.6	81262.42	68659.82	70544.5	74017.05	64174.24	274.8585
12308	16459.53	11898.57	7730.613	10620.98	15027.88	16116.43	18626.38
140047	142941.7	88226.92	80095.2	81895.55	80099.04	6335.425	-9149.94
93159	84887.43	86167.43	71658.18	60682.46	60612.62	62984.38	62623.72
147674	141551	144313.2	145678.6	37464.39	42633.65	46764.95	46296.99
82660	87129.11	29442.87	-824.05	-99564.7	-125228	-119797	-143641
131254	132511.4	130518.2	28732.73	23419.39	21998.76	28396.6	16065.4
59162	61004	64656.24	66594.18	65483.72	-82362	-80021.3	-150182
34895	36175.82	36684.94	39889.98	29702.05	-26184.3	-20935.3	-42052.4
15869	17499.54	21396.94	16110.79	19779.15	16492.36	23459.21	26517.07
108587	110457.7	115539.7	104440.7	98729.76	94761.79	92950.65	94282.98
143257	147824.6	149069.1	150282.2	154981.4	138043.7	86369.72	85076.67
132415	62864.72	45853.16	45833.66	8197.882	10685.42	-2138.12	-61481.8
38044	-5850.56	-64885.6	-106316	-105103	-105540	-120180	-115888
44936	35546.24	39878.55	43290.02	-76908.1	-107620	-119754	-123094
74696	74145.23	21504.53	5047.462	-32261.8	-29400.4	-42832.5	-51660.3
61533	-18273.8	-34802.5	-80448.5	-83672	-85480.5	-86377.3	-135924
142961	143904.6	150136.7	149163	118904.1	57879.5	27724.65	24418.61
130918	113734.8	67923.09	50284.63	59753.98	62158.98	-6936.99	-2365.55

Graphical Outcomes:







Here we have got 3 out graphs.:

I. Throughput vs BS
II. Throughput vs priority and
III. Throughput vs Weight Factor

We were supposed to get linear outputs. But unfortunately we got a slightly deviated output. The reason behind the deviated outputs are

a) PF Metric:

Here we have used 10 resource blocks. And there are also 300

Different users. So each user will have a different pf metric. Now due to these different pf metrices we get our deviated outputs.

b) 25% Overlapping:

Overlapping has been discussed earlier in the proposed models. Now due to this 25% overlapping we have got this deviated output.

c) Efiiciency :

Different users have different efficiency due to their Channel Quality Index (CQI). As the name implies, it is an indicator carrying the information on how good/bad the communication channel quality is.
CQI is the information that UE sends to the network and practically it implies the following two
i) Current Communication Channel Quality is this-and-that.
ii) I (UE) wants to get the data with this-and-that transport block size, which in turn can be directly converted into throughput.
the CQI value ranges from 0 ~ 30. 30 indicates the best channel quality and 0,1 indicates the poorest channel quality. Depending which value UE reports, network transmit data with different transport block size. If network gets high CQI value from UE, it transmit the data with larger transport block size and vice versa.

Now due to this different CQI we didn’t get our desired efficiency and as a result output was deviated.

REFERENCES

[1] 3GPP, Technical Specifications Group Radio Access Network - Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), 3GPP TS 25.913. [Online]. Available: http:// www.3gpp.org

[2] D. McQueen, “The momentum behind LTE adoption,” IEEE Communication Magazine, vol. 47, no. 2, pp. 44-45, Feb. 2009.

[3] 3GPP, Technical Specifications Group Radio Access Network - Physical channel and Modulation (release 8), 3GPP TS 36.211.

[4] F Capozzi, G Piro, La Grieco, G boggia, P Camarda, “Downlink packet scheduling in lte cellular networks: Key Design Issues and A survey,” IEEE Commun. Surv. Tutorials, 2012.

[5] A. S. Tanenbaum, Modern Operating Systems, 3rd Ed. Upper Saddle River, Nj, Usa: Prentice Hall Press, 2007.

[6] S. Hussain, “Dynamic Radio Resource Management in 3GPP LTE”, Blekinge Institute Of Technology 2009.

[7] Demers, A., Keshav, S., And Shenker,”Analysis and Simulation Of A Fair Queuing Algorithm.” Symposium Proceedings on Comm. Architectures and Protocols" ACM, New York, USA, 1989.

[8] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, “Dynamic Packet Scheduling Performance in UTRA Long Term Evolution Downlink,” In Proc. of International Symposium On Wireless Pervasive Comput.,, Santorini, Greece, May 2008, pp. 308 - 313.

[9] A. Stolyar and K. Ramanan, “Largest Weighted Delay First Scheduling: Large Deviations and Optimality,” Annals of Applied Probability, Vol.

11, pp. 1-48, 2001.

[10] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Resource Allocation in OFDMA Wireless Communications Systems Supporting Multimedia Services," IEEE/ACM Trans. On Networking, Vol. 17, pp. 926-935, June 2009.

[11] R. Almatarneh, M. H. Ahmed and O. A. Dobre, "'Frequency-Time Scheduling Algorithm For OFDMA Systems," in Proc. IEEE Canadian Conference on Electrical And Computer Engineering, 2009, Pp. 766 - 771.

[12] Rabie K. Almatarneh, Mohamed H. Ahmed, Octavia A. Dobre," Performance Analysis of Proportional Fair Sceduling in OFDMA Wireless Systems,"in Proc. of 72nd IEEE Vehicular Tech. Conference, 6-9 Sept., 2010.

[13] R. Kwan, C. Leung, And J. Zhang, "Proportional Fair Multiuser Scheduling In Lte," Proc. Of Ieee Signal Processing Letters, Vol. 16, No. 6, Pp. 461-464, Jun. 2009

[14] G. Monghal, K. I. Pedersen, I. Z. Kovacs, and P. E. Mogensen, "QOS oriented Time And Frequency Domain Packet Schedulers for the Utran Long Term Evolution," In Proc. of IEEE Veh. Tech. Conf., VTC-Spring, Marina Bay, Singapore, May 2008.

[15] D. Skoutas And A. Rouskas, "Scheduling with Qos Provisioning in Mobile Broadband Wireless Systems," In Wireless Conference (EW), 2010 European, April 2010, Pp. 422 -428.

[16] Y. Zaki, T. Weerawardane, C. Gorg, And A. Timm-Giel, "Multi-Qos-Aware Fair Scheduling for LTE," in Veh. Technology Conference (Vtc

Spring), 2011 IEEE 73rd, May 2011, Pp. 1 -5.

[17] H. Ramli, R. Basukala, K. Sandrasegaran, And R. Patachaianand, “Performance Of Well Known Packet Scheduling Algorithms In The Downlink 3gpp Lte System,” In Proc. Of Ieee Malaysia International Conf. On Comm., Micc, Kuala Lumpur, Malaysia, 2009, Pp. 815 -820.

[18] R. Basukala, H. Mohd Ramli, and K. Sandrasegaran,“Performance Analysis of EXP/PF And M-LWDF in Downlink 3GPP LTE System,” In Proc. of First Asian Himalayas International Conf. on Internet, Ah-ICI, Kathmundu, Nepal, Nov. 2009, Pp. 1 -5.

[19] B. Sadiq, R. Madan, And A. Sampath, “Downlink Scheduling for Multiclass Traffic In LTE,” Eurasip J. Wirel. Commun. Netw., Vol. 2009, pp. 9-9, 2009.

[20] Preethi Chandur, Karthik R and Krishna M. Sivalingam,”Performance Evaluation Of Scheduling Algorithms for Mobile Wimax Networks,” Eighth IEEE Percom Workshop On Pervasive Wireless Networking 2012.

[21] M. Iturralde, A. Wei, and A. Beylot, “Resource Allocation for Real Time Services using Cooperative Game Theory and A Virtual Token Mechanism in LTE Networks,” in IEEE Personal Indoor Mobile Radio Communications, Jan. 2012.

[22] Bilal Sadiq, Seung Jun Baek,”Delay-Optimal Opportunistic Scheduling and Approximations: The Log Rule”, IEEE/Acm Transactions on Networking, Vol. 19, No. 2, April 2011

[23] G. Piro, L. Grieco, G. Boggia, R. Fortuna, And P. Camarda, “Two-Level Downlink Scheduling For Real-Time Multimedia Services in

LTE Networks," in IEEE Trans. Multimedia, Vol. 13, No. 5, Oct. 2011.

[24] G. Monghal, D. Laselva, P. Michaelsen, and J. Wigard, "Dynamic Packet Scheduling for Traffic Mixes of Best Effort and VoIP Users in E-UTRAN Downlink," in Proc. of IEEE Veh. Tech. Conf., VTC-Spring, Marina Bay, Singapore, May 2010, pp. 1 -5.

[25] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, "MAC Scheduling Scheme for VoIP Traffic Service in 3G LTE," in Proc. of IEEE Veh. Tech. Conf., VTC-Fall, Baltimore, MD, USA, Oct. 2007.

[26] Y. Fan, P. Lunden, M. Kuusela, and M. Valkama, "Efficient Semi-Persistent Scheduling for VoIP on EUTRA Downlink," in Proc. of IEEE Veh. Tech. Conf., VTC-Fall, Calgary, Alberta, Sep. 2008.

[27] Y.-S. Kim, "An Efficient Scheduling Scheme to Enhance the Capacity of VoIP Services in Evolved UTRA Uplink,"EURASIP J. Wirel. Commun. Netw., vol. 2008, 2008.

[28] S. Saha and R. Quazi, "Priority-coupling-a semi-persistent MAC scheduling scheme for VoIP traffic on 3G LTE," in Proc.of Conf. on Telecommunications, ConTEL, Zagreb, Croatia, 2009, pp. 325 -329.

[29] Kuo chang,Hwang Cheng, Chih-Cheng Tseng," Energy-Efficient DRX scheduling for QOS traffic in LTE networks ,"IEEE 9th International Symposium on parallel and distributed processing with applications ,pp 213-218 ,2011.

[30] D. Sabella, M. Caretti, and R. Fantini, "Energy efficiency evaluation of state of the art packet scheduling algorithms for lte," in Proc. of IEEE European Wireless Conf., EW, pp. 1-4, Apr. 2011.