

## A Review of MAC Scheduling Algorithms in LTE System

Satheesh Monikandan B<sup>#</sup>, A.Sivasubramanian<sup>\*</sup>, SPK Babu<sup>#</sup>

<sup>#</sup>Department of ECE, Periyar Maniammai University, Thanjavur, Tamil Nadu, 613 403, India  
E-mail: satheeshmb\_rs@pmu.edu, sathy24@gmail.com, spkbabu@rediffmail.com

<sup>\*</sup>Mount Zion College of Engineering, Pathanamthitta, Kerala, 689 649, India  
E-mail: sivapollachi2002@gmail.com

**Abstract**— The recent wireless communication networks rely on the new technology named Long Term Evolution (LTE) to offer high data rate real-time (RT) traffic with better Quality of Service (QoS) for the increasing demand of customer requirement. LTE provide low latency for real-time services with high throughput, with the help of two-level packet retransmission. Hybrid Automatic Repeat Request (HARQ) retransmission at the Medium Access Control (MAC) layer of LTE networks achieves error-free data transmission. The performance of the LTE networks mainly depends on how effectively this HARQ adopted in the latest communication standard, Universal Mobile Telecommunication System (UMTS). The major challenge in LTE is to balance QoS and fairness among the users. Hence, it is very essential to design a down link scheduling scheme to get the expected service quality to the customers and to utilize the system resources efficiently. This paper provides a comprehensive literature review of LTE MAC layer and six types of QoS/Channel-aware downlink scheduling algorithms designed for this purpose. The contributions of this paper are to identify the gap of knowledge in the downlink scheduling procedure and to point out the future research direction. Based on the comparative study of algorithms taken for the review, this paper is concluded that the EXP Rule scheduler is most suited for LTE networks due to its characteristics of less Packet Loss Ratio (PLR), less Packet Delay (PD), high throughput, fairness and spectral efficiency.

**Keywords**— LTE; scheduling; HARQ; OFDM; proportional fair; UMTS

### I. INTRODUCTION

In 4G networks, Orthogonal Frequency Division Multiplexing (OFDM) modulation is adopted as the air interface to combat multi-path delay spread effect in frequency selective fading wireless channel [1] and to design low complexity transceiver with no inter-channel interference [2]. LTE is a full packet switched optimized system with end-to-end QoS support for Internet Protocol (IP) based traffic [3] and it the key standard in 4G cellular mobile communications by its higher data rates, flexible spectrum deployment with higher bandwidths, rising spectral efficiency, high fairness, high throughput rate, less latency and idle time and improved mobility [4], [5]. It is a multicarrier system where radio resources are spread in time and frequency domains and also known as Third Generation Partnership Project (3GPP) Release 8 [6], [7]. It aims to provide users with better wireless network experience while maintaining minimal complexity [4], [8]. LTE aims at ambitious goals of peak data rate of 100 Mbps in the downlink and 50 Mbps in the uplink [9]. But theoretically, the highest peak data rate on the Transport Channel (TC) is 75 Mbps in the uplink, and the downlink peak data rate can

be as high as 300 Mbps using spatial multiplexing, Orthogonal Frequency Division Multiple Access (OFDMA) in combination with 64-Quadrature Amplitude Modulation (QAM) and 20 MHz bandwidth [10]. The simplified structure of LTE network is given in Fig. 1 below.

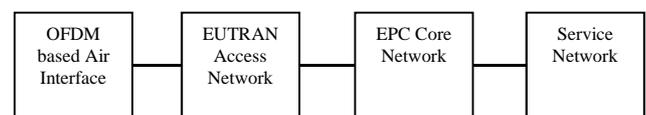


Fig. 1 LTE Network

LTE consists of two entities, the evolved Node-B (eNB) and the Mobility Management Entity/Gateway (MME/GW). The eNB has the responsibility of allocation of available Resource Blocks (RBs) among all users within the cell depending upon the operating bandwidth, based on scheduling algorithms involved [3]. The LTE RBs corresponding to different operating bandwidths is given in Table 1. The resource that is allocated to a user every 1-ms Transmission Time Interval (TTI) in the downlink has both frequency and time domains and is called an RB [3], [7],

[11]. The MME controls the signaling between the User Equipment (UE) and the Core Network (CN) [12]. The Radio Access Network (RAN) in LTE is Evolved Universal Terrestrial Radio Access Network (E-UTRAN) which has a network of eNBs. E-UTRAN is in charge for all radio-related functions, for example, Radio Resource Management (RRM), Header Compression, Security and Connectivity to the Evolved Packet Core (EPC).

TABLE I  
LTE RB'S CORRESPONDING TO DIFFERENT OPERATING BANDWIDTHS [3]

Sl.No.	System Bandwidth (MHz)	PRBs
1	1.4	6
2	3	15
3	5	25
4	10	50
5	15	75
6	20	100

The growing demand for dynamic IP services on the wireless network like Voice over Internet Protocol (VoIP), website browsing, video streaming and online games have given new challenges to the design of 4G networks. Under OFDM, few errors in the bit stream may result in packet failure at MAC layer. The RT multimedia services such as video conferencing and VoIP [13] have to be considered differently in LTE networks to satisfy the required QoS [4]. The RT traffic demands the delivery of packets within the delay upper bound [1]. RT services may be either delay sensitive, loss sensitive or both and they need less PD and PLR in order to receive the information in good quality in time [4], [11]. Priority should be given to the RT services than the Non-real-time (NRT) services such as file downloads and web browsing. For the best operation of the system from the customer's perspective, the system should ensure their QoS of RT flows and/or maximize their throughput (user satisfaction) of best effort (BE) flows [13]. In the event of packet failure at MAC layer, the receiver can get the potential error bits retransmitted prior to decoding. MAC layer performs Cyclic Redundancy Check (CRC) insertion at the transmitter and check at the receiver and delivers successful data units to the Radio Link Control (RLC) layer. The objective of LTE is to build the capacity and speed of wireless systems to a wide range of multimedia applications by achieving maximum throughput gain with the help of HARQ retransmission and proper downlink scheduling schemes [14], [15].

HARQ is utilized for fast retransmissions of erroneous packets to recover the packet errors [1] and to keep the radio interface delay minimum [9]. It imposes stringent time restrictions on MAC layer processing. HARQ functionality is controlled by MAC layer and implemented at the Physical Layer (PHY). A further objective was the upgrade and rearrangements of the system architecture to an IP-based framework with altogether diminished transfer latency compared to the 3G architecture with the help of two-level approach. MAC layer in LTE supports the sub-

tasks such as downlink scheduling, uplink scheduling and scheduling factor calculation with some predefined configurations [16]. The resources required for each sub-task varies with the fluctuation of RT traffic. The traffic fluctuation on the MAC layer exhibits a slowly varying effect than the over-the-air effect. The more resource consuming MAC sub-tasks are scheduling factor calculation and sorting. LTE defines a 1ms TTI and the processing time of MAC sub-tasks should be less than 1ms. Otherwise, the MAC layer cannot send data or instructions to the PHY layer, and 800 $\mu$ s is recommended in LTE to minimize the risk of exceeding 1ms TTI threshold. The MAC sub-tasks are evoked every 1ms TTI which has two-time slots. The Logical Channels (LCs) are scheduled in MAC layer by their packet scheduling algorithms [4]. Intelligent resource allocation can improve the resource utilization and power consumption under high real-time constraints.

#### A. MAC Layer

MAC protocol layer exists in UE and eNB, and it is a part of LTE air interface's user and control planes which controls access to the shared transmission medium. It runs in both UE and eNB with different behaviors in each by giving commands in eNB and responding them in UE. The main functions of the MAC layer are mapping between LC's and Transport Blocks (TBs), Multiplexing/Demultiplexing of MAC Service Data Units (SDUs) belonging to one or different LC's into/from TB conveyed to/from the PHY layer on transport channels, reporting scheduling information using metrics computation [4], error correction using HARQ, priority handling of LC's and UE's based on dynamic scheduling, transport format selection, reporting of traffic volume measurement, service identification and padding. The Service Access Point (SAP) between the PHY layer and MAC sub-layer gives the transport channels, and the SAP between the MAC sub-layer and RLC provides the logical channels. Data is transferred between MAC sub-layers in the UE and eNB using TBs and shared TCs [16]. Users inform their current channel quality in the cell by a method for a quantized feedback called Channel Quality Indicator (CQI) [7]. The channel quality is measured by the UE in the form of Signal to Interference and Noise Ratio (SINR) and reports the channel dependent CQI reports to the eNB in the uplink to give the time and frequency form of channel quality data for various downlink RRM functions such as Packet Scheduling (PS) and Link Adaptation (LA). PS uses CQI to select the scheduling time and frequency for each UE by calculating priority metrics, and LA uses CQI to choose the most efficient Modulation and Coding Scheme (MCS) [9]. The CQI manager in the PHY layer processes each received CQI report for the downlink and Sounding Reference Signal (SRS) for the uplink. These are utilized by the eNB which performs scheduling decisions in each TTI by distributing Physical Resource Blocks (PRBs) to the users and assigning an Adaptive Modulation and Coding Scheme (AMCS) to enable channel diversity for UE [3]. The target of AMCS is to amplify the information rate while keeping up an average Packet Error Rate (PER) [15]. Because the transmission efficiency of the network is characterized by the proportion of effectively transmitted packets to the aggregate number of packets, is exceptionally dependent on

this PER [1]. The allocated PRBs and selected MCS are directed to the scheduled users on the Physical Downlink Control Channel (PDCCH) [5]. Even though the resource allocation is handled in MAC layer, there is no guarantee that the data has been successfully transmitted or received without an efficient scheduling algorithm [4]. Therefore, MAC scheduling schemes are introduced to schedule the transmissions as indicated by the PD constraints and the achievable throughputs of each connection to guarantee auspicious and spectrum-efficient conveyance of the packets [1].

The target of this paper is to review existing literature on QoS/Channel-aware downlink scheduling algorithms. The structure of this paper has been organized as follows. Section II is dedicated to briefly discuss the background, comprehensive, relevant literature survey of possible existing downlink scheduling algorithms and elaboration of the practical usability of scheduling algorithms in LTE system. Section III shows the findings made from the theoretical study and a brief discussion of key points about the different algorithms with the comparison table and Section IV concludes the paper with attention on the theoretical analysis made on MAC schedulers.

## II. MATERIALS AND METHODS

### A. Downlink Scheduling in LTE

LTE specifications represent Frequency Division Duplexing (FDD) using Single-Carrier Frequency Division Multiple Access (SC-FDMA) for uplink traffic and Time Division Duplexing (TDD) using OFDMA for downlink traffic [5], [9], [17]. OFDMA is one of the key access technologies in the 4G wireless systems because of its high degree of flexibility in allocating resources, scalability, simple equalization and strong characteristics against frequency selective fading [9], [13]. Scheduling is an important factor in RRM mechanism, which is implemented at the eNB. It plays an important role in QoS provisioning by providing mechanisms for the resource allocation and multiplexing at the packet level to guarantee that different types of applications meet their service requirements [13]. It is responsible for the distribution and control of time and frequency domain radio resources to the various users with different service needs, achieving QoS requirements and optimizing system performance in the downlink and uplink of LTE networks. The best PS should be designed to support varieties of traffic with different priorities, channel statistics, queue sizes and traffic loads while improving the Quality of Experience (QoE) for end users [3]. RRM is the method for dynamically allocating radio resources to UE based on the scheduling algorithms involved [7], [18]. The role of RRM is to ensure that the radio resources are efficiently utilized, taking advantage of the available adaptation techniques to serve users according to their QoS attributes. These mechanisms include QoS-aware bearer admission control, multi-user time and frequency domain packet scheduling, HARQ management and LA with dynamic switching between different transmission modes. The RRM functions such as QoS profiling, admission control and semi-persistent scheduling at layer-3 of Fig. 2 are portrayed as semi-

dynamic mechanisms since they are mostly executed amid the setup of new data flows.

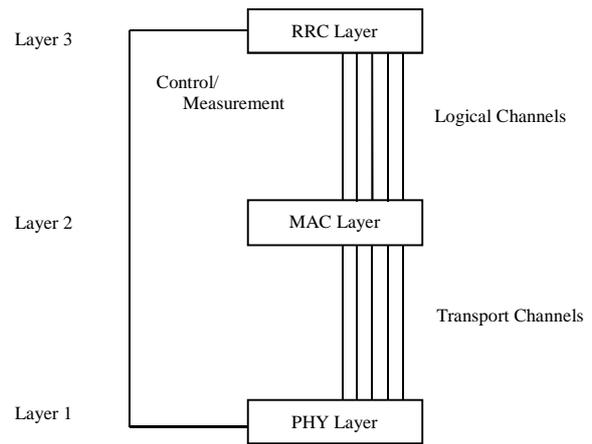


Fig. 2 LTE Layers

QoS is used to determine the performance trade-offs of the LTE network. To achieve the target QoS, many scheduling algorithms have been implemented to allocate the frequency and time resources efficiently to the RT and NRT traffics. Traffic scheduling is associated with channel management to achieve the high-performance networks and to guarantee QoS requirements. RT traffic QoS demands to include the requirement of having a lower PLR (in packets per second) and PD whereas NRT traffic QoS demands more throughput rate [4]. The PLR of a user can be defined as the ratio of the average number of lost packets due to buffer overflow to the average number of arriving packets in one frame period [15]. Note that the packet loss can also occur when errors on the data bit cannot be corrected [19]. The average PD is defined as the ratio of the average number of packets in a queue of a user to the average throughput [15]. The goal is to maximize the cell capacity while fulfilling the minimum QoS for the bearers [5]. Each scheduling algorithm under QoS/Channel-aware scheduling strategy has different methods to determine the users' scheduling priority, such as buffer status, packet delay bound, expected throughput, channel status, past average throughput (in packets per frame) and fairness to provide multiple applications of streaming and live video, VoIP, and data on a single IP-based infrastructure [7], [20]. Scheduling decisions are made on a per user basis even though a user may have several data flows, and MAC protocol decides how much data is sent from each data flow [5]. The required data rate for the QoS provisioning of RT traffic depends on how strictly the scheduling policy is imposed on the scheduling algorithms [13]. An NRT service such as web browsing and File Transfer Protocol (FTP) are BE based, do not have strict requirements and thereby have no minimum throughput requirement. Throughput, PLR, and PD are the key metrics and the main contributors to meet the QoS requirements for different applications in LTE networks. The need to support diverse QoS requirements whilst maximizing the system throughput is the major challenge in the design of downlink scheduling algorithms [11]. Downlink scheduling is whereby data packets are scheduled to transmit from eNB to UE [4].

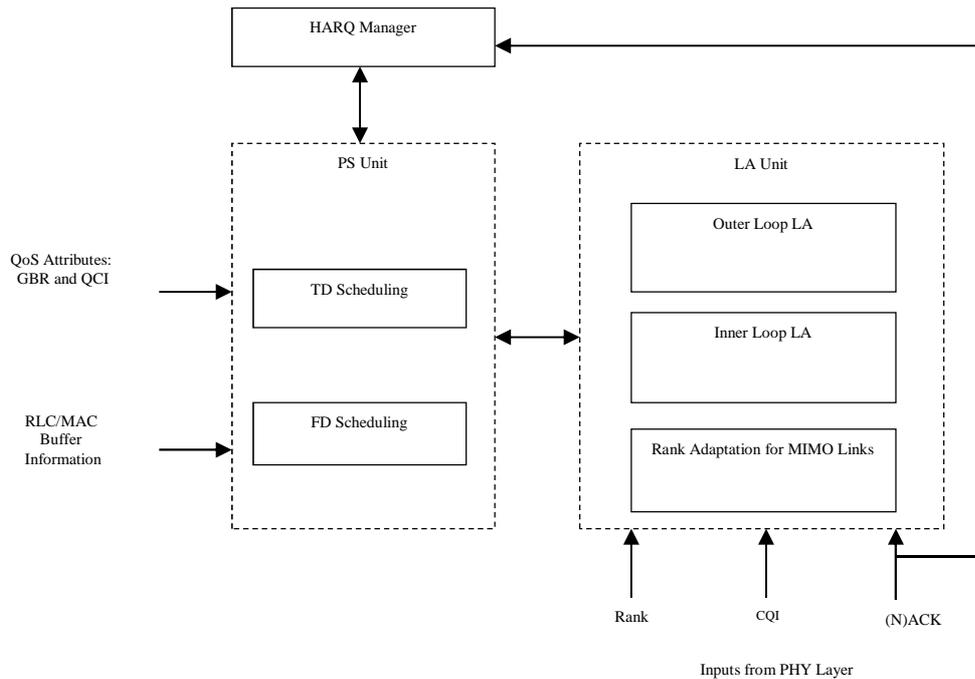


Fig. 3 MAC Layer Functionalities [5]

The shared channel transmission in Downlink Shared Channel (DL-SCH) and Uplink Shared Channel (UL-SCH) is effectively controlled by the scheduler at the MAC layer by managing the resources in the uplink and downlink directions. To deal with traffic load in both directions, the switch point periodicities of LTE frames are used. The downlink information rate and the users can be increased by using this with higher downlink slots [3]. Fig. 3 shows the interaction between the PS and the HARQ manager as it is responsible for scheduling retransmissions. Since the downlink scheduler has the flexibility to dynamically schedule the pending HARQ retransmissions in the time and frequency domains, it must decide to send a new transmission or pending HARQ transmission to each scheduled user during one TTI. The Link Adaptation Unit (LAU) gives information to the PS about the supported AMCS for a user based on the selected PRBs, CQI feedback from the users in the cell and QoS requirements. The outer loop LAU controls the Block Error Rate (BLER) of first transmissions [9] based on HARQ acknowledgements from the past transmissions. Time Domain (TD) scheduler selects the users to be scheduled in the next TTI while the Frequency Domain (FD) scheduler allocates PRBs to the selected users. Here, QoS awareness is achieved by the TD scheduler while radio-channel awareness is done by the FD scheduler. FD scheduler uses the frequency selective fading to schedule the PRBs to the users with high channel quality while avoiding the PRBs to the users having deep fades. It can achieve typically 40% improved cell throughput and coverage gain [9] when the UE having moderate speed up to 20 to 30 km/h since the radio channel is traceable via the periodic CQI reports from the UE. At high speed, the eNB may not be able to accurately track the radio channel due to CQI delays [9], and the best performance can be achieved by designing and incorporating the effective PS algorithm in the

eNB [5]. Thus, FD-PS and LA are essential techniques to improve the performance of LTE networks [9].

The PS considers the traffic volume and QoS profile consisting of the downlink related parameters such as Guaranteed Bit Rate (GBR), Maximum Bit Rate (MBR), Aggregate MBR (AMBR), Allocation and Retention Priority (ARP) and QoS Class Identifier (QCI) of each UE. The GBR parameter is only specified for GBR bearers, and AMBR is specified for non-GBR bearers. The ARP parameter is an integer in the range of 1 to 16 which is defined primarily for prioritization when conducting admission control decisions. QCI is a pointer to a more detailed set of QoS attributes which includes the parameters such as layer-2 Packet Delay Budget (PDB) and PLR. Layer-2 PDB may be used by the eNB packet scheduler to prioritize certain queues in order to fulfill their Head-of-line (HOL) packet delay targets [5]. The more details of QCI types and their applications are given in Table 4 of Section III in this paper.

Schedulers assign radio resources considering the instantaneous channel quality at the UE through measurements made at the eNB [8] and/or reported by the UE. Packet schedulers operate in the time domain with one TTI and in the frequency domain with one RB [18]. As the cells start to contain more active users, the need for scheduling more users in every TTI increases in order to meet the QoS constraints for the different users [5]. In the downlink direction, E-UTRAN dynamically allocates radio resources to UE at each TTI through Cell Radio Network Temporary Identifier (C-RNTI) on PDCCH. PDCCH carries information of dynamic scheduling grants for both downlink and uplink. Each UE scans the contents of PDCCH for Downlink Control Information (DCI) associated to C-RNTI for getting information such as resource allocation type, the bitmap for allocation, AMCS, index to HARQ process and transmit power control. The DCI formats and their purposes

are given in Table 2. The allocated resources combined with AMCS are used to define the used TB size [5]. Downlink channel condition is given by the UE through the channel-quality report in the form of CQI. CQI is a four digit value sent to eNB by UE, and this helps eNB to allocate proper MCS and RB for UE. The downlink data transmission relies on fast LA with AMCS ranging from Quaternary Phase Shift Keying (QPSK) to 64-QAM, assuming constant and equal transmit power of PRBs for a user [5], [8]. The equal power assumption has frequently been used for implementation simplicity as well as analytical traceability in downlink resource allocation problems [11].

Each user has assigned a buffer at eNB. The packets of delay-sensitive priority classes entering the buffer are time stamped by the scheduler [7]. Schedulers consider the buffer status with the help of Buffer Status Report (BSR) and priorities in their scheduling decisions depend upon service types and assign enough resources to schedule the packets before their PDB. BSR is a UE way of informing the network that it has certain data in its buffer and requires a grant to send this data. The performance of downlink scheduling algorithms is studied in terms of system throughput, average RT and NRT throughput, packet loss for RT services, fairness for NRT services and spectral efficiency [11].

TABLE II  
DOWNLINK CONTROL INFORMATION

DCI Formats	Purpose	
Uplink Scheduling	0	Scheduling and TPC for PUSCH
Downlink Scheduling	1	Scheduling for PDSCH and TPC for PUCCH
	1A	Compact Scheduling for PDSCH and TPC for PUCCH
	1B	MIMO Compact Scheduling for PDSCH and TPC for PUCCH
	1C	Very Compact Scheduling for PDSCH
	1D	Compact Scheduling for PDSCH with Power Offset and TPC for PUCCH
	2	Closed Loop MIMO Compact Scheduling for PDSCH and TPC for PUCCH
	2A	Open Loop MIMO Compact Scheduling for PDSCH and TPC for PUCCH
Uplink Power Control	3	TPC for PUSCH and PUCCH – 2bit Power Adjustment
	3A	TPC for PUSCH and PUCCH – 1bit Power Adjustment

To understand the performance of LTE downlink transmission, six scheduling algorithms, based on QoS/Channel-aware scheduling have been studied and focused on in this paper. They are Proportional Fair (PF) [4], [15], [18], [21], Modified Largest Weighted Delay First (MLWDF) [4], [11], [18], [21], Exponential Proportional Fairness (EXP/PF) [4], [11], [18], Exponential Rule (EXP) [4], [18], [21], [22], Logarithmic Rule (LOG) [4], [18], [21], [22] and Frame Level (FL) Scheduler [18], [23]. They vary as far as input parameters, objectives and service targets. The

details are discussed in the following sections.

### B. Downlink Scheduling Procedure

This section gives the exhaustive study of the essential scheduling procedure required in the LTE frameworks. It is a choice of service provider, and there are no specific rules in the LTE standard [3]. In the scheduling framework, it is essential that the scheduler should be flexibly adjusted by customer's changing tendencies and service provider's strategic decisions [13]. A portion of the criteria used to plan and test the schedulers incorporate best multiplexing techniques, fairness among data streams, good channel conditions and priority of data streams [3]. Since LTE employs OFDM in the downlink, users are apportioned a particular number of subcarriers for a foreordained measures of time and are called PRBs. The allocation of PRBs is handled by a scheduling function at the eNB and at one TTI, each PRB can only be allocated to one flow [3], [7], [8].

The scheduling algorithms aim to augment the throughput, fulfill user's QoS and give considerable fairness to NRT users. In the scheduling criteria, they use channel conditions, HOL packet delay, buffer status and service types for making scheduling decisions by the calculation of priority metric. The value of the priority metric determines which packet is to be scheduled first and so on. Priority metric having a higher value will be given a higher priority to schedule the packet first and then followed by the next highest metric value. This is performed by the scheduling algorithms invoked in the MAC layer, until each and every single packet is to be scheduled and served [4].

HOL packet delay is the waiting time of the packet residing in the buffer of a data flow and for efficient scheduling; it must be below the prescribed PDB because a packet violating the preset HOL delay threshold will be dropped from the buffer. Due to diversity in the channel quality and data rate of the different applications, some of the packet flows may be experiencing a higher HOL delay and some less [7]. When a user is selected for transmission, the number of bits to be transmitted is based on the user's instantaneous downlink signal to noise ratio (SNR) value, and it plays a key role in deriving the AMCS [3]. This value of each user varies on each sub-carrier at each TTI due to the frequency selective nature of multi-path propagation and time-selective fading nature of user movement. With higher SNR, more data rate can be achieved in the downlink direction via higher-order modulation schemes [3] because the system service rate is higher than the packet arrival rate [1]. Pathloss in an urban environment, shadow fading, and multi-path fading are used to determine instantaneous downlink SNR value of each user on each RB and channel gain. Usually, users near eNB tend to achieve higher SNR values while users at the end of the cell suffer from lower SNR [3]. The scheduling of PRB with highest SNR allows improved throughput, delay, and fairness. Similarly, scheduling schemes incorporating the service class priorities along with target Bit Error Rate (BER) from LTE specifications allows increased aggregated system throughput when channel quality statistics in addition to the user's earlier data rates and channel occupancy statistics are used [3]. The minimum downlink SNR values and modulation scheme mapping for LTE is given in Table 3.

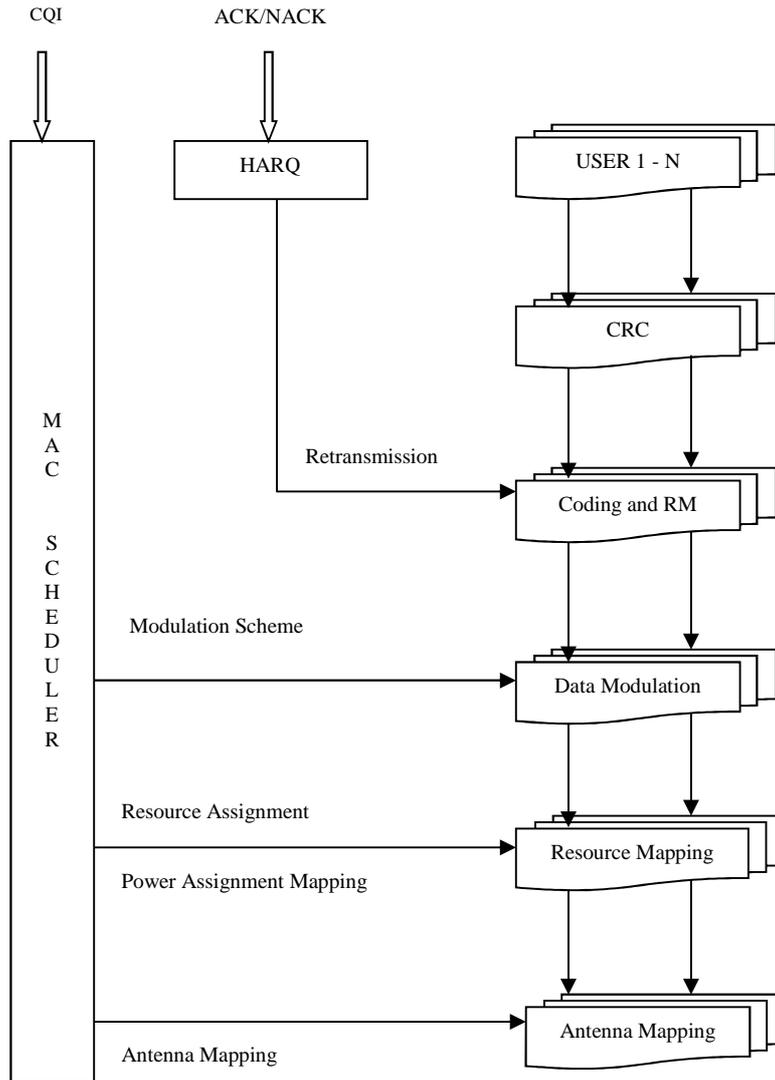


Fig. 4 LTE Downlink MAC Scheduler

TABLE III  
DOWNLINK SNR VALUES AND MODULATION SCHEME MAPPING FOR LTE  
[3]

Sl.No.	Minimum Instantaneous Downlink SNR value (dB)	Modulation Scheme
1	1.7	QPSK
2	3.7	QPSK
3	4.5	QPSK
4	7.2	16QAM
5	9.5	16QAM
6	10.7	16QAM
7	14.8	64QAM
8	16.1	64QAM

UE mobility plays a critical role in improving the throughput and end-to-end delay for the service flows, at the

same time it hinders coverage and incurs limitations on the capacity of the network also. At higher speed and large service flows in the system, the end-to-end delay approaching deadlines for packet transmission earlier and decreasing the average throughput of the users [3]. Therefore, each algorithm should be evaluated based on system throughput, average throughput, PLR, system delay, HOL delay, end-to-end delay for RT traffic and fairness for NRT traffic. The sequence of scheduling procedure of a scheduler [18] is given in the following steps:

1. Creation of the list of data flows to be scheduled in the current TTI by eNB.
2. Recording of CQI report and the queue length of the MAC layer for each flow.
3. Calculation of scheduling metric for each flow based on the scheduling strategy involved.
4. Calculation of the size of TB for every scheduled flow with the amount of data to be transferred at the MAC layer and the selection of best AMCS (QPSK/ 16QAM/ 64QAM).

5. Transmission of DCI in PDCCH for the information about the users, RBs and the AMCS to the UE.
6. Reading of PDCCH payload by the UE.

The RT flow such as VoIP and video has its own QoS parameters like maximum latency or dateline expiry of the individual packet and average traffic rate whereas NRT flow has no explicit QoS parameters. In the system, at each time slot, the scheduler should determine the sub-channel assignment based on each flow's current channel quality, minimum average throughput and individual packet dateline expiry and each sub-channel should be occupied by only one user [13]. To satisfy the QoS requirements of the users, the scheduler should be designed to achieve the guarantee bounded delay for the RT flows to deliver the data packet within a predefined time by taking advantage of best channel conditions while maintaining data rates for all major classes including the BE traffic [3], [13].

### C. Downlink Schedulers

1) *PF Scheduler*: Scheduling is the process in which eNB decide which UEs should be given resources to send or receive data at a sub-frame level. PF scheduler is a content-blind PS which produces a trade-off between throughput performance and fairness in allocating resources to users by considering the user's achieved throughput [4], [7], [15]. Fairness in resource allocation is defined as the ratio between the difference of total packet sizes for the most and least served users to the total accumulated packet size of all the flows that arrive at the eNB scheduler over a time T [3]. This algorithm is providing an efficient amount of bandwidth in its network [4]. Channel-awareness is a fundamental concept for achieving high performance in a wireless environment, and if one can estimate the channel quality perceived by a user on a given RB, it will be easy to allocate radio resources obtaining very high data rate.

This algorithm is able to achieve a good level of fairness among its users by monitoring the channel's quality of each user as well as the previous user's throughput value during its decision making [2]-[4]. The goal is to maximize total network throughput and to guarantee fairness among flows [3]. In the operating principle, the average data rate of each UE is calculated and given as input to the scheduler, and each UE gives the feasible rate to the eNB at the beginning of each TTI. At each time slot, the user experiencing the highest instantaneous rate with respect to its average rate is scheduled [15]. The priority metric  $j$  is defined as the ratio of the expected data rate of the channel state of the  $i^{\text{th}}$  user  $u_i(t)$  at time slot  $t$ , to the mean data rate supported by that channel  $\bar{u}_i$  which acts as a weighting factor and calculated by the eNB as below.

$$j = \frac{u_i(t)}{\bar{u}_i} \quad (1)$$

The user with the highest priority metric will be selected for transmission at the next TTI. This algorithm supports high resource utilization while maintaining good fairness and spectral efficiency (in bits/s/Hz) among users. A user is likely to be scheduled when its instantaneous channel quality

is high relative to its own average channel condition over time so that the users in bad channel conditions also will be served for a certain amount of time surely [21]. This algorithm serves better for BE traffic [4], and it is not ideal for delay-sensitive RT services like video flows since it does not consider the delay metric of the data packet in its scheduling decision [7], [29]. Video traffic covers a wide range of services such as video telephony, entertainment, advertising and video-enabled call center services [8]. The disadvantage of this algorithm is that it is incapable of prioritizing traffic with critical conditions; therefore the traffic may encounter an amount of delay which may lead to packet loss if packets are not served before packet expiry [4].

2) *MLWDF Scheduler*: This algorithm is a QoS/Channel-aware, bounded delay scheduler and is able to provide good performance in terms of fairness measures, spectral efficiency, and less PLR [4], [7]. It supports multiple RT data users in Code Division Multiple Access-High Data Rate (CDMA-HDR) systems to handle multiple services with different QoS requirements [18]. It is used to improve QoS of different flows by defining a probability function that represents the urgency for data transmission corresponding to the time spent in the buffer queue [3]. It considers the packet queuing delay and PLR in addition to the channel condition of each user for its decision making efficiently [3], [4], [7], [13], [20], thereby it is able to provide good performance in terms of fairness measures, spectral efficiency and less PLR [4]. The priority metric  $j$  is defined to give priority to RT flows and to have the radio operator channel with best propagation conditions. A set of QoS parameters is associated with each bearer depending on the application data it carries and thereby enable the differentiation among flows [2].

The metric  $j$  is calculated by using the parameters of QoS factor  $a_i$ , expected data rate of the  $i^{\text{th}}$  user  $u_i(t)$  at time slot  $t$ , mean data rate supported by the channel  $\bar{u}_i$  and HOL packet delay  $W_i(t)$  using the equation(2) to determine the user's priority.

$$j = a_i \frac{u_i(t)}{\bar{u}_i} W_i(t) \quad (2)$$

where,

$$a_i = -\frac{\log \delta_i}{\tau_i} \quad (3)$$

Here,  $\delta_i$  is the acceptable PLR of the  $i^{\text{th}}$  user, and  $\tau_i$  is the delay threshold of the  $i^{\text{th}}$  user. One of the most important QoS requirements is that the packets have to be delivered within a delay bound. So, this algorithm prioritizes the user with higher HOL packet delay and better channel conditions [3], [4], [7], [11]. The HOL delay measures the maximum delay experienced by any packet in the scheduling line [3]. In the MAC queue, packets of RT services who have not been transmitted by the time of its expiry shall be removed from the queue to avoid the consumption of bandwidth. This algorithm is simple, suitable for lower loads, providing good throughput, less average throughput thereby, more PLR and

good fairness among RT and NRT services. When the LTE system is run under high load, this rule only serves higher priority flows such as video conferencing users with good quality using the packet's HOL delay [4]. For this purpose, it requires a proper flow admission control policy which should not increase the arrival rate of packets above the system capacity because the increase in arrival rate incurs delay bound violations. For the best performance of this algorithm, once the delay-sensitive traffic's arrival rate reaches the system capacity, the admission control policy should block further flows from entering the system [7], thereby it has the best performance in video transmissions during UE moving at a minimal speed [4]. Even though the delay-based priority metric makes this rule dynamic, this algorithm fails to dynamically adjust the priorities for RT and NRT services due to the strict delay requirements of RT services. But the earlier delay approaching deadlines for packet transmission at higher speed and during large service flows in the system can be better satisfied by the strict priority mechanism used by this algorithm [3].

3) *EXP/PF Scheduler*: This algorithm is the extension of PF scheduler and offers QoS promises over a shared wireless connection. This algorithm configures the multimedia applications in the Adaptive Coding and Modulation / Time Division Multiplexing (ACM/TDM) system for both RT and NRT services and enhances the priority of RT flows [18]. It guarantees a bounded delay to the packets and also maximizes the system throughput while ensuring proportional fairness between UEs [4]. The metric is calculated for RT and NRT users separately. For the RT users, the metric  $j$  is calculated by using the parameters of the data rate of the channel state of the  $i^{\text{th}}$  user  $u_i(t)$  at time slot  $t$ , mean data rate supported by the channel  $\bar{u}_i$ , HOL packet delay  $W_i(t)$  and the QoS factor  $a_i$ .

$$j = \max_i a_i \frac{u_i(t)}{\bar{u}_i} \exp\left(\frac{a_i W_i(t) - a\bar{W}}{1 + \sqrt{a\bar{W}}}\right) \quad (4)$$

where,

$$a\bar{W} = \frac{1}{N} \sum_i a_i W_i(t) \quad (5)$$

Here,  $a_i > 0$  and  $i = 1, 2, 3, \dots, N$ .

where  $N$  is the number of active RT downlink flows which acts as weights, and the value of  $a_i$  is given as an element of the set  $\left[\frac{5}{0.99\tau_i}, \frac{10}{0.99\tau_i}\right]$ .  $\tau_i$  is a threshold value of the packet's waiting time [4]. In this algorithm, scheduling metric exponentially grows along with delay metric [20].

For NRT users, the metric is calculated by using the parameters of data rate of the channel state of the  $i^{\text{th}}$  user  $u_i(t)$  at time slot  $t$ , mean data rate supported by the channel  $\bar{u}_i$ , average number of RT packets waiting at eNB buffer

$M(t)$ , maximum HOL packet delay of all RT users  $W_{\max}$  and maximum delay constraint of RT users  $\tau_{\max}$ .

$$j = \max_i \frac{w(t)}{M(t)} \cdot \frac{u_i(t)}{\bar{u}_i} \quad (6)$$

where,

$$w(t) = \begin{cases} w(t-1) - \varepsilon \dots W_{\max} > \tau_{\max} \\ w(t-1) + \frac{\varepsilon}{k} \dots W_{\max} > \tau_{\max} \end{cases} \quad (7)$$

Here,  $\varepsilon$  and  $k$  are constants.

By using this algorithm, the RT users have higher priority than NRT users by considering the exponential function of the HOL delay when their HOL packet delay are reaching the delay dateline [7], [11]. At higher load conditions this algorithm provides better performance with more computational complexity. It dynamically adjusts the priority of RT and NRT services, provides less throughput, good average throughput thereby less PLR [4] and good fairness to both RT and NRT services.

4) *EXP Rule Scheduler*: This algorithm is a bounded delay scheme [3] which is the modification of EXP/PF scheduler. It is categorized under the QoS/Channel-aware strategy for packet scheduling and is able to guarantee delay requirements. This rule is known for having a low delay, less PLR and a high measure of fairness as well as throughput [3]. In the principle of operation, the data packets that are not served before their expiration dateline shall be removed from the queue to avoid the bandwidth consumption. While taking scheduling decisions, this rule takes account of the comprehensive network status but does not consider packets that are lost due to the packet expiration dateline [4]. In this non-homogeneous scheduling algorithm [22], the delay of every user is normalized over the sum of observed delay of all the users [4] and provides queue stability without any previous knowledge of the channel information to serve high data rate requirements. Also, this algorithm makes the trade-off between the current transmission rates and unequal queues. In heterogeneous channels, this rule has a bias towards balancing queues thereby compromise the throughput and mean delays for all users. For NRT users, the priority metric of PF algorithm is adopted [24]. For RT users, the priority metric is calculated as below.

$$j = \max_i \frac{u_i(t)}{\bar{u}_i} \exp\left(\frac{a_i W_i(t)}{1 + \sqrt{a\bar{W}}}\right) \quad (8)$$

where,

$$a_i = 6/d_i \quad (9)$$

and,

$$\bar{W} = \frac{1}{N} \sum_i W_i(t) \quad (10)$$

where,  $u_i(t)$  is the expected data rate of the  $i^{\text{th}}$  user at time slot  $t$ ,  $\bar{u}_i$  is the mean data rate supported by the channel,  $a_i$  is

the QoS factor,  $W_i(t)$  is the HOL packet delay,  $d_i$  is the maximal delay target of  $i^{\text{th}}$  user, and  $N$  is the number of active RT downlink flows. This algorithm is well suited for RT applications like live video to satisfy user's demands due to less PD and PLR [4]. This algorithm prioritizes RT traffic with regard to NRT traffic [3] by considering the exponential function of the HOL delay in addition to the good channel condition and supports both application services [4], [7], [13]. It is found to have improvements in delay, fairness measures and throughput [4].

This rule has further enhanced into four types which are well suited for RT applications. They are: 1) simplified EXP rule (sEXP Rule), 2) modified EXP rule (mEXP Rule), 3) EXP rule with maximum throughput (MT) (EXP\_MT Rule), and 4) enhanced EXP rule with MT (E2M) [8]. They will be analysed and explored in future.

5) *Log Rule Scheduler*: The non-homogeneous scheduling algorithm, Log rule has been proposed in [22] to provide a balance between delay and robustness. The goal of this algorithm is to minimize the mean packet delays and thereby to maximize the throughput. It contributes in giving efficient QoS support to the network by maintaining an average amount of delay [4]. This algorithm works with sensitivity to both the scheduler parameters and channel characteristics. It exploits the structures throughput whereby the traffic arrivals, as well as the channel conditions, are known [4]. By using this algorithm, 80% reduction in mean packet delays can be achieved, and it is suitable for the users under reasonably high loads base on the Radial Sum-rate Monotonicity (RSM) property. Also, it satisfies the objective of graceful degradation, meaning that as many users as possible should meet their QoS requirements under overloads. In a system with unpredictable heterogeneous channels, with the help of this algorithm, half of the users achieve their QoS requirements very well. The RSM of this rule ignores queue balancing to maximize the service rate in the algorithm functioning. Under the conditions of changing loads and unpredictable channel conditions and capacities, Log rule is more flexible to take scheduling decisions. This rule satisfies the theorems 1 and 2 as per [22], and the priority metric is calculated based on the logarithmic function of delay experienced by the  $i^{\text{th}}$  user as below.

$$j = \max_i \frac{u_i(t)}{\bar{u}_i} \log(1 + a_i W_i(t)) \quad (11)$$

where,  $a_i = 5/d_i$  (12)

to achieve the optimality [4].

Here,  $u_i(t)$  is the expected data rate of the  $i^{\text{th}}$  user at time slot  $t$ ,  $\bar{u}_i$  is the mean data rate supported by the channel,  $a_i$  is the QoS factor,  $W_i(t)$  is the HOL packet delay, and  $d_i$  is the maximal delay target of  $i^{\text{th}}$  user. This algorithm prioritizes the delay-sensitive traffic by considering the logarithmic function of the HOL delay [7].

6) *Frame Level Scheduler*: A two-level resource allocation and scheduling scheme allow resources in the form of PRB to be distributed among different type of traffics according to the load requirements of flow [3]. This algorithm has two levels in its scheduling operation. A low-complexity scheduling algorithm is implemented on the upper level to calculate the number of data to be transmitted in a single frame to satisfy the delay constraint. The sorting out of flows at an initial level allows a reduction in the complexity of this scheduler where it is required to sort various traffic mixes of users with different priorities [3]. The PF scheduler is implemented in the lower level to allocate the radio resources to the user and to give good fairness among data flows [18]. The amount of data to be transmitted by the  $i^{\text{th}}$  user in the  $k^{\text{th}}$  frame  $u_i(k)$  is calculated below.

$$u_i(k) = h_i(k) * q_i(k) \quad (13)$$

where  $h_i(k)$  is the impulse response of Linear Time Invariant (LTI) filter,  $q_i(k)$  is the signal in queue level for filtering and  $*$  is the discrete time convolution operator. For a video streaming, this scheduler performs well by improving the overall throughput of the network for the good level of user satisfaction even when simplest PF scheduler is used at the lower level [3].

#### D. Practical Usability

The LTE system supports diversified traffic of very high volume when compared to the UMTS with the help of suitable downlink scheduling algorithm [25]. Network services like VoIP, web browsing, chat, FTP, Peer to Peer (P2P) file sharing, video telephony, interactive RT gaming, multimedia, buffered streaming video and live streaming video are the key areas where the above downlink scheduling algorithms are involved in LTE systems with the bandwidth configurations from 1.4 MHz to 20 MHz [2], [5]. To achieve the required QoS to the above services, several classes of QoS parameters such as ARP, MBR, and GBR have been identified through QCIs. QCIs are used as a reference for driving specific packet forwarding behaviors such as scheduling weights, admission thresholds, queue management thresholds and link-layer protocol configuration and guaranteeing interoperability between different service providers and network domains in LTE systems [26]-[28]. LTE defines PDB, PLR and flow priority for a set of nine standard LTE service classes to be followed by the service provider, out of which first four are treated as GBR while the other five are treated as Non-GBR (NGBR) [3] which are given in Table 4. LTE maps QoS constrained flows into parameters like minimum GBR, priority, PDB, and PLR. Based on their QCIs assigned by the network, the RRM procedures will be enabled with the help of above discussed QoS-aware schedulers in the transport network nodes. For downlink packets, the gateway in the Evolved Packet System (EPS) performs the mapping of QCI into the 6-bit Differentiated Services Code Point (DSCP) to make a translation of packets from bearer-level QoS to transport level QoS, and the scheduling algorithms determine the

traffic forwarding treatment of each packet based on the DSCP value.

Table 4 gives the different types of standardized QCI values as per 3GPP Release 8 specifications, their PDB in milliseconds (ms), corresponding access class whether delay

sensitive or delay tolerant and the services used in applications to the users. The goal of standardizing QCI is to ensure that the applications and services that are mapped to that value receive the same minimum level of QoS in multivendor networks and during roaming [27].

TABLE IV  
LTE SERVICE CLASS TABLE [5], [26]

QCI	PDB (ms)	PLR	Access Class	Flow Priority	Services
1	100	10-2	Sensitive	2	Conversational Voice
2	150	10-3	Sensitive	4	Conversational Video
3	300	10-6	Tolerant	5	Buffered Streaming Video
4	50	10-3	Sensitive	3	RT Gaming
5	100	10-4	Sensitive	1	IMS Signaling
6	100	10-3	Sensitive	7	Voice, Live Streaming Video, and Gaming
7	300	10-6	Tolerant	6	Buffered Streaming, TCP-Based, Email, Web browsing, P2P File sharing
8	300	10-6	Tolerant	8	
9	300	10-6	Tolerant	9	

### III. RESULTS AND DISCUSSIONS

The following observations are made from the literature survey of MAC scheduling algorithms used in the LTE system.

(1) *PF Algorithm*: PF algorithm distributes resources to flows on the basis of CQI [3] and gives a better trade-off between fairness and spectral efficiency. It assigns priority to users based on their achievable data rate to the average throughput achieved by the user in the previous sub-frames and thereby the users with poor channel conditions are also allocated with resources during the operation. From the study, it understood that it is not considered a good solution for RT services due to its high PLR, low throughput, and high PD but it is suitable for NRT traffic [28]. This algorithm maximizes the total network throughput and guarantees fairness among flows such as video, BE and VoIP [29].

(2) *MLWDF Algorithm*: MLWDF algorithm is mostly used for RT operating system, streaming video application and wired networks because BE traffic is not provided with any guarantee and resources are only allocated when the guaranteed class QoS is satisfied. It has indirect delay measures to schedule the user's flow and the users having highest delay measure is selected and given access to physical channel [3]. It works based on the system parameter  $\delta_i$ , representing the acceptable probability for the  $i^{\text{th}}$  user that a packet is dropped due to dateline expiration with the aim of avoiding dateline expiration [2]. Hence the user with strongest requirements in terms of acceptable PLR and dateline expiration will be preferred for resource allocation. Thus for the video service, this algorithm takes instantaneous channel variations and delays into account and gives better performance [29]. From the study, it is understood that MLWDF scheduler can be able to adapt an increasing user diversity and channel variation much better

than PF scheduler [3] and it is not suitable for the users with poor average radio propagation conditions suffering from higher delays [28]. Hence it is not suitable for VoIP flows due to its high PD and PLR, but it is the best scheduling rule for delay-sensitive applications in terms of fairness and efficiency [7], [29]. It can also be used with LTE service class constraints for LTE downlink [3].

(3) *EXP/PF Algorithm*: EXP/PF algorithm supports multimedia applications in data communication networks in which the user can belong to either RT or NRT service, and it increases the priority of RT flows [29]. When the HOL packet delays for all the users do not differ a lot, the exponential term in this algorithm is close to 1, and it performs as PF scheduler. At the same time, if one of the users gets the HOL delay very large, the exponential term overrides the channel state-related term, and the user gets a priority [28]. Thus it is mostly used for video flows and VoIP since it provides high fairness index [29]. From the study, it is understood that for the services with a significant user throughput gain, a low PD, a high fairness index and a low PLR, this algorithm is suitable and gives better performance than PF and MLWDF [3]. Also, it is understood that this algorithm is able to support the performance on RT traffics in terms of average throughput and PLR [4]. The main drawback of this algorithm is the positive probability of drop of the services from NRT applications as shifting it to RT one [20].

(4) EXP rule is designed for Single Carrier Code Division Multiple Access (SC-CDMA) networks with a shared downlink channel which is an improvisation of the aforementioned EXP/PF scheduler. It is robust, and it monitors the entire network's status. From the study, it is understood that this algorithm has indirect delay measures to schedule the users flow and users having the highest delay measure is chosen and offered access to physical channel

[3]. The packets that are not served before their expiration date shall be expelled from the queue and perform better than PF [3] for RT traffic which includes video and VoIP traffic transmission. It has a better performance in throughput [3], PLR and spectral efficiency for video traffic and improves the scheduling performance in terms of latency and fairness [4]. This algorithm gives higher priority to the user with more transmission delay besides the channel condition [20]. It does not guarantee the BE traffic because the resources are only allocated when the guaranteed class QoS is satisfied. It can also be used with LTE service class constraints for LTE downlink. For Constant Bit Rate (CBR) traffic, this algorithm performs the best because it considers the buffer queue with an exponentially growing delay priority for user packets [3].

(5) *LOG Rule*: LOG rule contributes much more efficient QoS support to the network by keeping up an average amount of delay [4]. It functions to exploit the structures throughput whereby the traffic arrival in addition to the

condition of the channels is known and gives much better scheduling performance.

(6) *FL Scheduler*: FL scheduler is a two-level scheduler which is suitable for delay sensitive, RT multimedia services like VoIP and video in wireless networks [14]. From the study, it is understood that this algorithm improves the data throughput since it works in both time and frequency domain in a single scheduling process. In the time domain scheduler, the selection of the user is made based on the priority metric by considering channel conditions and PD. Scheduling is done by determining which users are nearest to the deadline and then forwarded to PF algorithm in the frequency domain for the allocation of RBs with priority to transmit the data. The successful transmission of packets is ensured in the eNB. The time domain schedulers are also used to differentiate the users based on their QoS requirements. Based on this, VoIP gets the highest priority followed by video streaming, web browsing, and FTP services.

TABLE V  
SUMMARY OF COMPARISON OF QoS/CHANNEL-AWARE SCHEDULING ALGORITHMS

Scheduling Algorithms	PLR	PD	Through put	Fair ness	Spectral Efficiency
PF	High	High	Less	Good	Good
MLWDF	High	High	Good	Good	Good
EXP/PF	Average	Average	Average	Good	X
EXP Rule	Less	Less	High	High	High
LOG Rule	Average	Less	X	High	X
FLS	X	X	Good	Good	X

Based on the theoretical comparison of six downlink scheduling algorithms taken for the review and shown in the above Table 5, it is found that the EXP Rule scheduler is most suited for LTE networks due to its characteristics of less PLR, less PD, high throughput, fairness and spectral efficiency. Also, it is understood that all the delay-aware scheduling algorithms except PF algorithm must ensure that the arrival rate of the packets should not surpass the system capacity. Otherwise, the QoS performance of the current users in the LTE network would be damaged and result in an increase in the number of unsatisfied users [7].

#### IV. CONCLUSION

This review paper has focused on the study of the MAC layer in the LTE system. Based on the literature review, various QoS/Channel-aware downlink scheduling algorithms have been developed by researchers for solving the scheduling issues in the LTE downlink. The motivation behind this review is to distinguish the current reviews relating to this issue. Most of the studies considered minimizing the PD and PLR as the main objective of their proposed models. In this paper, the basic requirement of scheduling, the radio access network architecture and six types of QoS/Channel-aware scheduling algorithms are discussed with their design principles. As a conclusion, it is understood that for the overall LTE system performance to

improve the throughput, the choice of scheduler should be more pliable to changes in the user's traffic and channel conditions. Also, it is understood that the mean delay optimal schedulers satisfy the RSM and thereby minimize the mean packet delays of the users. This gives more robustness in taking the scheduling decisions to allocate the radio resources to the particular users in the LTE network during RT and NRT flows to achieve their QoS requirements. Finally, it is found that EXP Rule downlink scheduler is having a better performance for RT services than the others due to their characteristics of less PLR, less PD, high throughput, fairness and spectral efficiency. This might be an opportunity for future research, where new algorithms and mathematical models could be developed.

#### ABBREVIATIONS

LTE: Long Term Evolution; RT: Real-time; QoS: Quality of Service; HARQ: Hybrid Automatic Repeat Request; MAC: Medium Access Control; UMTS: Universal Mobile Telecommunication System; PLR: Packet Loss Ratio; PD: Packet Delay; OFDM: Orthogonal Frequency Division Multiplexing; IP: Internet Protocol; 3GPP: Third Generation Partnership Project; TC: Transport Channel; OFDMA: Orthogonal Frequency Division Multiple Access; QAM: Quadrature Amplitude Modulation; eNB: Evolved Node-B; MME/GW: Mobility Management Entity/Gateway; RB:

Resource Block; TTI: Transmission Time Interval; UE: User Equipment; CN: Core Network; RAN: Radio Access Network; E-UTRAN: Evolved Universal Terrestrial Radio Access Network; RRM: Radio Resource Management; EPC: Evolved Packet Core; VoIP: Voice over IP; NRT: Non-real-time; CRC: Cyclic Redundancy Check; RLC: Radio Link Control; PHY: Physical Layer; LC: Logical Channel; TB: Transport Block; SDU: Service Data Unit; SAP: Service Access Point; CQI: Channel Quality Indicator; SINR: Signal to Interference and Noise Ratio; PS: Packet Scheduling; LA: Link Adaptation; MCS: Modulation and Coding Scheme; SRS: Sounding Reference Signal; PRB: Physical Resource Block; AMCS: Adaptive MCS; PDCCH: Physical Downlink Control Channel; FDD: Frequency Division Duplexing; SC-FDMA: Single-Carrier Frequency-Division Multiple Access; TDD: Time Division Duplexing; QoE: Quality of Experience; FTP: File Transfer Protocol; BE: Best-effort; DLSCCH: Downlink Shared Channel; ULSCCH: Uplink Shared Channel; LAU: Link Adaptation Unit; BLER: Block Error Rate; TD: Time Domain; FD: Frequency Domain; GBR: Guaranteed Bit Rate; MBR: Maximum Bit Rate; AMBR: Aggregate Maximum Bit Rate; ARP: Allocation and Retention Priority; QCI: QoS Class Identifier; PDB: Packet Delay Budget; HOL: Head-of-line; C-RNTI: Cell Radio Network Temporary Identifier; DCI: Downlink Control Information; PER: Packet Error Rate; QPSK: Quaternary Phase Shift Keying; BSR: Buffer Status Report; TPC: Transmit Power Control; PUSCH: Physical Uplink Shared Channel; PDSCH: Physical Downlink Shared Channel; PUCCH: Physical Uplink Control Channel; MIMO: Multiple Input Multiple Output; ACK/NACK: Acknowledgment / Negative Acknowledgment; PF: Proportional Fair; MLWDF: Modified Largest Weighted Delay First; EXP/PF: Exponential Proportional Fairness; EXP: Exponential Rule; LOG: Logarithmic Rule; FL: Frame Level; SNR: Signal to Noise Ratio; BER: Bit Error Rate; CDMA-HDR: Code Division Multiple Access-High Data Rate; ACM/TDM: Adaptive Coding and Modulation/Time Division Multiplexing; sEXP: Simplified EXP; mEXP: Modified EXP; MT: Maximum Throughput; E2M: Enhanced EXP rule with MT; RSM: Radial Sum-rate Monotonicity; LTI: Linear Time Invariant; P2P: Peer to Peer; NGBR: Non-GBR; EPS: Evolved Packet System; DSCP: Differentiated Services Code Point; IMS: IP Multimedia Subsystem; TCP: Transmission Control Protocol; SC-CDMA: Single Carrier Code Division Multiple Access; CBR: Constant Bit Rate.

#### REFERENCES

- [1] Ying Jun Zhang, and Soung Chang Liew, "Link-adaptive largest-weighted-throughput packet scheduling for real-time traffics in wireless OFDM networks," in *IEEE Global Telecommunications Conference Proceedings*, 2005, vol.5.
- [2] F.Capozzi, G.Piro, L.A.Grieco, G. Boggia, and P.Camarda, "Downlink packet scheduling in LTE cellular networks: key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol.15. Issue.2, pp.678-700, 2013.
- [3] Salman Ali, Muhammad Zeeshan, and Anjum Naveed, "A capacity and minimum guarantee-based service class-oriented scheduler for LTE networks," *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen Journal, 2013.
- [4] Ee Mae Ang, Kuok Kwee Wee, Ying Han Pang, and Keat Keong Phang, "A performance analysis on packet scheduling schemes based on an exponential rule for real-time traffic in LTE," *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen Journal, 2015.
- [5] Klaus I P, Troels E K, Frank F, Istvan Z K, Daniela L, and Preben E M, "An overview of downlink radio resource management for UTRAN Long-Term Evolution," *IEEE Communications Magazine*, vol.47, Issue.7, pp.86-93, Jul.2009.
- [6] Yasir Zaki, Thushara Weerawardane, Carmelita Gorg, and Andreas Timm-Giel, "Multi-QoS-aware fair scheduling for LTE," in *IEEE 73rd Vehicular Technology Conference*, 2011, pp.1-5.
- [7] Nabeel Khan, and Maria G.Martini, "QoE-driven multi-user scheduling and rate adaptation with reduced cross-layer signaling for scalable video streaming over LTE wireless systems," *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen Journal, 2016.
- [8] Kan Zheng, Xiaoli Zhang, Qiang Zheng, Wei Xiang, and Lajos Hanzo, "Quality-of-Experience assessment and its applications to video services in LTE networks," *IEEE Wireless Communications Magazine*, vol.22, Issue.1, pp.70-78, Feb.2015.
- [9] Niko K, Jani P, Petteri K, Tapani R, Tero H, and Martti M, "Channel quality indication reporting schemes for UTRAN Long Term Evolution downlink," in *IEEE Vehicular Technology Conference*, 2008, pp.2522-2526.
- [10] (2016) The 3GPP website. [Online]. Available: <http://www.3gpp.org/LTE>.
- [11] R.Basukala, H.A.Mohd Ramli, and K.Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP system," in *IEEE First Asian Himalayas International Conference on Internet, AH-ICI 2009*.
- [12] (2016) The Tutorialspoint website. [Online]. Available: [http://www.tutorialspoint.com/lte/lte\\_network\\_architecture.htm](http://www.tutorialspoint.com/lte/lte_network_architecture.htm).
- [13] Youngki Kim, Kyuho Son, and Song Chong, "QoS scheduling for heterogeneous traffic in OFDMA-based wireless systems," in *IEEE Global Telecommunications Conference Proceedings*, 2009.
- [14] Panagopoulous and Athanasios, *Handbook on Research on Next Generation Mobile Communication Systems*, IGI Global, 2015.
- [15] Loren Carrasco, Guillem Femenias, and Jaume Ramis, "Channel-aware MAC performance of AMC-ARQ-based wireless systems," *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen Journal, 2013.
- [16] Bao, Xiaofeng, Geng, Caroline, and Mukherjee, "A pooling prototype for the LTE MAC layer based on a GPP platform," in *IEEE Global Communication Conference (GLOBECOM)*, *IEEE Conference Publications*, 2015, pp.1-7.
- [17] Rami Amad, Elankovan A Sundararajan and Nor Effendy Othman, "A moving direction and historical information assisted fast handover in LTE-A," *International Journal on Advanced Science, Engineering and Information Technology*, vol.6, No.6, pp.882-888, 2016.
- [18] Samia Dardouri and, Ridha Bouallegue, "Comparative study of downlink packet scheduling for LTE networks," *Wireless Pers Commun* (2015) 82:1405-1418, Springer Science+Business Media, 2015.
- [19] Thimma V.J. Ganesh Babu, A.Anpalagan, and Jeremiah F.Hayes, "A study of DiffServ based QoS issues in next generation mobile networks," in *IEEE Canadian Conference on Electrical and Computer Engineering, IEEE Conference Publications*, 2004, vol.4, pp.2359-2362.
- [20] Hindia MN, Reza AW, Noordin KA, and Chayon MHR, "A novel LTE scheduling algorithm for green technology in smart grid," *PLoS ONE* 10 (4):e0121901, Open Access Research Article, 2015.
- [21] S.Fouziya Sulthana, and R.Nakkeeran, "Performance evaluation of downlink packet scheduling algorithms in LTE," in *Proc.of Int.Conf.on Advances in Communication, Network and Computing*, Elsevier, 2014, pp.38-44.
- [22] Bilal Sadiq, Seung Jun Baek, and Gustavo de Veciana, "Delay-optimal opportunistic scheduling and approximations: the log rule," in *Proc.of the IEEE INFOCOM*, 2009, pp.1692-1700.
- [23] Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, and Pietro Camarda, "A two-level scheduling algorithm for QoS support in the downlink of LTE cellular networks," in *IEEE European Wireless Conference (EW)*, *IEEE Conference Publications*, 2010, pp.246-253.
- [24] Taniya Shafiqu, Zia Muhammed, Huy-Dung Han, and Hasan Mahmood, "Cross-layer chase combining with selective retransmission, analysis and throughput optimization of OFDM systems," *IEEE Transactions on Communications*, vol.64, Issue.6, pp.2311-2325, 2016.

- [25] Aswin Kanagasabai, and Amiya Nayak, "Channel aware scheduling algorithm for LTE uplink and downlink," *Network Protocols and Algorithms*, White Paper, vol.7, No.3, 2015.
- [26] Tiago P.C.de Andrade, Carlos A. Astudillo, and Nelson L.S.da Fonseca, "Random access mechanism for RAN overload control in LTE/LTE-A networks," in *IEEE International Conference on Communications (ICC), IEEE Conference Publications*, 2015, pp.5979-5984.
- [27] Hannes Ekstrom, "QoS control in the 3GPP evolved packet system," *IEEE Communication Magazine*, vol.47, Issue.2, pp.76-83, Feb.2009.
- [28] Tara Ali-Yahiya, *Understanding LTE and its Performance*, Springer Science+Business Media, LLC, 2011.
- [29] Stojan Kitanov, and Toni Janevski, "Performance evaluation of scheduling strategies for LTE networks in downlink direction," in *Proceedings of the XI International Conference ETAI*, Republic of Macedonia, 2013.
- [30] Cedric Westphal, "Monitoring proportional fairness in cdma2000 high data rate networks," in *IEEE Global Telecommunications Conference, IEEE Communications Society*, 2004, vol.6, pp.3866-3871.
- [31] Lingyang Song and Jia Shen, *Evolved Cellular Network Planning and Optimization for UMTS and LTE*, CRC Press, 2010.
- [32] Suyash Tripathi, Vinay Kulkarni, and Alok Kuma, "LTE E-UTRAN and its access side protocols," *White Paper*, 2011.
- [33] (2016) The 3GPP website. [Online]. Available: <http://www.3gpp.org/specifications/releases/72-release-8>.
- [34] (2016) The 3GPP website, TS 36.321, E-UTRA: medium access control (MAC) protocol specifications, [Online]. Available: <http://www.3gpp.org>.
- [35] (2016) The Tutorialspoint website. [Online]. Available: [http://www.tutorialspoint.com/lte/lte\\_quickguide.htm](http://www.tutorialspoint.com/lte/lte_quickguide.htm).