Study of Downlink Scheduling Algorithms in LTE Networks

S. Fouziya Sulthana and R. Nakkeeran
Department of Electronics Engineering, School of Engineering and Technology, Pondicherry University, Puducherry-605014, India
Email: fouziya@pec.edu, nakkeeranpu@gmail.com

Abstract—Long Term Evolution (LTE) is one of the fastest growing technologies which supports variety of applications like video conferencing, video streaming, VoIP, file transfer, web browsing etc. In order to support multiple applications, Radio Resource Management (RRM) procedure is one of the key design roles for improving the system performance. LTE system effectively utilizes the resources by dynamically scheduling the users in both frequency and time domain. However, scheduling algorithms are not defined in the Third Generation Partnership Project (3GPP) specifications. Therefore, it becomes one of the special interests for service providers. In this paper a study of downlink scheduling algorithms present in the literature is put forth and performance evaluation of four algorithms proposed for LTE downlink is carried out. This paper also discusses the key issues of scheduling algorithms to be considered for future traffic requirements.

Index Terms—LTE; Resource Allocation; Scheduling

I. INTRODUCTION

The emerging applications with different throughput, delay, Packet Loss Rate (PLR) and bandwidth requirements emphasize the need of a network capable of supporting range of services. To fulfil this need Long Term Evolution (LTE) was introduced by Third Generation Partnership Project (3GPP) [1]. The main objective of the LTE network is to enhance the data rate so as to provide the radio resources for variety of highly demanded services, while taking into consideration a satisfied level of Quality-of-Service (QoS) to all active users. For this requirement, LTE system uses Orthogonal Frequency Division Multiple Access (OFDMA) technology in the Downlink (DL) and Single Carrier-Frequency Division Multiple Access (SC-FDMA) in the Uplink (UL). The OFDMA technology divides the available bandwidth into multiple sub-carriers and allocates a group of sub-carriers to a user based on its QoS requirements. Hence, the design of efficient resource allocation algorithm is important for effective use of radio resources to meet the system performance targets.

Packet scheduler at radio base station (evolved Node B (eNB) in LTE specification) is in charge of assigning portions of spectrum shared among users. The performance of the network can differ according to the algorithms used for scheduler. Designing an effective scheduler is therefore an important task in order to differentiate the performance of one wireless system from another. The packet scheduler in LTE aims to maximize the spectral efficiency and makes the negative impact of channel quality drops into negligible [2].

Several packet scheduling algorithms are proposed to support the increasing traffics. In this paper, the key design aspect of LTE scheduling and the performance analysis of four existing algorithms are given. The rest of the paper is organised as follows: In section II, the overview of LTE network is given. Radio Resource Management (RRM) concepts and different scheduling algorithms for LTE downlink is discussed in section III. In section IV, performance analysis of different algorithms is carried out, the future challenges of LTE are given in section V and conclusion is given in section VI.

II. OVERVIEW OF LTE NETWORK

In order to support wide variety of applications, LTE network is designed with challenging requirements that overtake the features of 3G networks mainly designed for voice services [1]. LTE network provides spectrum flexibility where the transmission bandwidth can be selected between 1.4 MHz and 20 MHz depending on the available spectrum. The peak data rate, which is the important parameter by which different technologies are usually compared, generally depends on the amount of spectrum used. The allowed peak data rate for the DL and UL is equal to 100 Mbps and 50 Mbps respectively. LTE targets to provide spectral efficiency two to four times better than 3G systems (15 bps/Hz in DL and 3.75 bps/Hz in UL).

LTE is flat, Internet Protocol (IP) based architecture with respect to the previous 3G systems [3]. In previous system, separate Radio Access Network (RAN) that consists of Radio Resource Control (RRC), Radio Link Control (RLC) and Medium Access Control (MAC) protocols is used to interface with User Equipment (UE). But in LTE, eNB takes care of the above mentioned protocol functions. So it requires lesser number of nodes that reduces the system latency and improves overall performance [4]. The network architecture of LTE consists of core network called Evolved Packet Core (EPC) and access network called Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) as shown in Fig. 1.
The responsibility of eNB in the access network is to ensure that the necessary QoS for a bearer over the interface is met. Each bearer has an associated QoS Class Identifier (QCI) and each QoS class is characterized by priority, tolerable packet loss, and tolerable delay as shown in Table 1.

Generally bearers can be classified into two categories based on the nature of the QoS they provide: Guaranteed Bit-Rate (GBR) bearers which are real time bearers and non-GBR bearers which are non real time bearers.

At the physical layer, LTE supports both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) modes. OFDMA is chosen as the DL access technology. The available bandwidth is divided into multiple Resource Blocks (RBs) based on time and frequency domains. A RB is the smallest allocation unit in LTE which can be modulated independently. In the frequency domain the RB consists of 12 consecutive subcarriers and in the time domain it is made up of one time slot of 0.5 ms duration and adopts two slots as allocation period. The scheduling period is called as one Transmission Time Interval (TTI) and it lasts for 1 ms duration as shown in Fig. 2.

III. RADIO RESOURCE MANAGEMENT IN LTE

RRM is the general word used in wireless systems to cover all radio related functions like assignment, management and sharing of radio resources among users. The scheduler which is found in the eNB, controls the assignment of RBs to UEs to avoid intra-cell interference. In general the function of scheduler is to find the optimal allocation of the resource unit (time, frequency, power etc) to UEs such that QoS requirements of users are satisfied.

The radio interface in LTE uses one common shared channel which is shared by all users in the cell. The eNB controls the allocation of RBs both on UL and DL shared channel; Physical Uplink Shared Channel (PUSCH) and Physical Downlink Shared Channel (PDSCH), respectively [6]. Both the UL and DL scheduling are carried out at eNB. To indicate the scheduled RBs for a particular UE the Physical Downlink Control Channel (PDCCH) is used. The PDCCH is carried in the first 1-3 OFDM symbols in each TTI.

The scheduler selects the UE to be scheduled and number of RB to be assigned based on two factors: the channel quality and the QoS requirements. In DL, the scheduler can assign any random set of RBs for a particular UE whereas in the UL the RBs allocated have to be adjacent to each other because of single carrier property. To facilitate the channel dependent scheduling on DL, the eNB has to get the channel quality reports from the UE. Each UE calculates the signal-to-noise (SNR) ratio based on its channel condition. It sends the Channel Quality Indicator (CQI) value to eNB based on its calculated SNR to choose the appropriate modulation and coding scheme (MCS).

CQI reporting method is to find balance between channel quality estimation and minimum signalling overhead. Many CQI reporting methods are proposed in [7]. In case of erroneous transmission, eNB performs retransmission by Hybrid Automatic Repeat Request (HARQ) procedure. HARQ is based on well known stop and wait algorithm [8].

A. Dynamic Resource Allocation in LTE

Dynamic resource allocation or Packet Scheduling (PS) takes care of QoS aspects on the access side by employing suitable algorithm for scheduling the data in both UL and DL. The main task of any scheduling algorithm is to maximize the network utilization and provide fairness among users. The PS is an entity of RRM in LTE which is present in the MAC layer of eNB. The MAC layer also provides most important procedures for the LTE radio interface like multiplexing/demultiplexing, random access procedures, scheduling requests etc [9].

In multiuser environment, a good PS scheme makes use of multiuser diversity and channel fading. When many users fade independently, at any time there is a high probability that one of the users will have a good channel. By allowing only that user to transmit, the shared channel resource is used in the most efficient way and the total system throughput is maximized. Thus with increasing number of users the multiuser diversity gain increases [10]. The difficulty lies in the fact that radio resource allocation should also satisfy fairness among UEs. Moreover, in slow fading, multiuser diversity hardly satisfies all QoS parameters at the same time, especially fairness. Ultimately, RRM should follow a combined form of multiuser diversity and fairness scheduling.

The two main entities of PS are Time Domain Packet Scheduling (TDPS) and Frequency Domain Packet
TABLE I. STANDARDIZED QCI FOR LTE [5]

<table>
<thead>
<tr>
<th>QCI</th>
<th>Resource Type</th>
<th>Priority</th>
<th>Packet Delay Budget [ms]</th>
<th>Packet Loss Rate</th>
<th>Example services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBR</td>
<td>2</td>
<td>100</td>
<td>$10^3$</td>
<td>Conversational voice</td>
</tr>
<tr>
<td>2</td>
<td>GBR</td>
<td>4</td>
<td>150</td>
<td>$10^3$</td>
<td>Conversational video (live streaming)</td>
</tr>
<tr>
<td>3</td>
<td>GBR</td>
<td>5</td>
<td>300</td>
<td>$10^3$</td>
<td>Non-Conversational video (buffered streaming)</td>
</tr>
<tr>
<td>4</td>
<td>GBR</td>
<td>3</td>
<td>50</td>
<td>$10^3$</td>
<td>Real time gaming</td>
</tr>
<tr>
<td>5</td>
<td>non-GBR</td>
<td>1</td>
<td>100</td>
<td>$10^3$</td>
<td>IMS signalling</td>
</tr>
<tr>
<td>6</td>
<td>non-GBR</td>
<td>7</td>
<td>100</td>
<td>$10^3$</td>
<td>Voice, video (live streaming), interactive gaming</td>
</tr>
<tr>
<td>7</td>
<td>non-GBR</td>
<td>6</td>
<td>300</td>
<td>$10^3$</td>
<td>Video (buffered streaming)</td>
</tr>
<tr>
<td>8</td>
<td>non-GBR</td>
<td>8</td>
<td>300</td>
<td>$10^3$</td>
<td>TCP based (e.g., WWW, e-mail), chat, FTP, P2P file sharing</td>
</tr>
<tr>
<td>9</td>
<td>non-GBR</td>
<td>9</td>
<td>300</td>
<td>$10^3$</td>
<td>Voice, video (live streaming), interactive gaming</td>
</tr>
</tbody>
</table>

Scheduling (FDPS). The TDPS selects a subset of schedulable UEs and FDPS determines the transport block size, MCS, Physical Resource Block (PRB) to UE mapping. Resource allocation for any UE is based on the scheduling decision of the algorithm. The factors that need to be considered before designing an algorithm are QoS provisioning, throughput maximization, fairness, complexity and scalability.

In LTE downlink, the QoS aspects depend on number of factors like channel conditions, resource allocation policies, available resources, delay sensitive/insensitive traffic etc. The resource allocation is realized in every TTI, that is exactly every two consecutive RBs. That is, resource allocation is done on a resource block pair basis. Fig. 3 shows the generalised model of a packet scheduler.

Resource allocation for each UE is usually based on the comparison of per-RB metric. This metric can be interpreted as the transmission priority of each UE on a specific RB. The detailed key issues in designing a scheduler are given in [11]. The scheduling strategies of any wireless network can be broadly classified as shown in Fig. 4.

Channel independent scheduling is based on the assumption that channel is time invariant and error-free. The channel independent scheduling is first introduced in wired networks [12]. Examples of channel independent scheduling are First-in-First-out (FIFO), Round Robin (RR), Weighted Fair Queuing (WFQ), Earliest Deadline First (EDF), Largest Weighted Delay First (LWDF) etc. Here some algorithms satisfy the QoS requirements and some simply schedules.

Channel sensitive scheduling is based on the assumption that channel is time variant and error-free. The channel sensitive scheduling is first introduced in wired networks [12]. Examples of channel sensitive scheduling are Maximum Throughput (MT), Proportional Fairness (PF), Throughput To Average (TTA), Modified- Largest Weighted Delay First (M-LWDF), Exponential Proportional Fairness (EXP/PF), Exponential rule (EXP rule), Logarithmic rule (LOG rule) etc. In LTE only channel sensitive scheduling is done based on the CQI reports from the UE.

B. Channel Sensitive Scheduling

Many scheduling algorithms are proposed based on the channel estimations in LTE network. Some algorithms aim to maximize the throughput (like MT, M-LWDF,
EXP/PF, EXP rule, LOG rule) while some aim to provide fairness among UEs (like PF, TTA). Some of them are discussed here.

**Maximum Throughput:** This algorithm provides the maximum overall throughput by allocating each RB to the UE which experience the good channel conditions. That is, the UE which experience the good channel quality will always be scheduled. The priority metric used in MT for $i^{th}$ user on $k^{th}$ RB can be expressed as

$$p^k_i = \arg \max_i \left( d_i(t) / \bar{R}_i(t-1) \right); \ 1 \leq i \leq N$$  \hspace{1cm} (1)$$

where $d_i(t)$ is the expected data rate for $i^{th}$ user at time $t$ on $k^{th}$ RB and $N$ is the number of users in a network. This algorithm maximizes the cell throughput but at the same time it provides unfairness resource sharing among UEs especially, cell edge users.

**Proportional Fairness:** This algorithm provides balance between spectral efficiency and fairness among UEs. The priority metric used for PF for $i^{th}$ user on $k^{th}$ RB can be expressed as

$$p^k_i = \arg \max_i \left( d_i(t) / \bar{R}_i(t-1) \right); \ 1 \leq i \leq N$$  \hspace{1cm} (2)$$

where $d_i(t)$ is the expected data rate for $i^{th}$ user at time $t$ on $k^{th}$ RB, $\bar{R}_i(t-1)$ is the past average throughput up to time slots $t-1$ and $N$ is the number of users. The denominator value will be decreased for the users experiencing bad channel conditions, which maximizes the priority metric, so that the poor channel users will also be allocated resources. The past average throughput can be calculated as follows:

$$\bar{R}_i(t) = (1 - 1/T)\bar{R}_i(t-1) + 1/T r_i(t)$$  \hspace{1cm} (3)$$

here $T$ is the fairness window would be equal to one TTI. $r_i(t)$ is the data rate achieved by the $i^{th}$ user at time $t$.

Many algorithms have been proposed in literature with modifications to PF algorithm. A new scheduling approach based on PF algorithm is introduced in [13] which try to balance coverage and cell throughput. In [14], the problem is formulated which aims to maximize the throughput using PF algorithm. The results showed that the performance obtained by using different PF implementation increase with the complexity of the optimization problem.

The Generalized Proportional Fair (GPS) scheduling is developed in [15]. The PF metric is modified by means of two parameters $\xi$ and $\psi$. The priority metric can be expressed as

$$p^k_i = \arg \max_i \left( \frac{[d_i(t)]^\xi}{[\bar{R}_i(t-1)]^\psi} \right); \ 1 \leq i \leq N$$  \hspace{1cm} (4)$$

The parameters $\xi$ and $\psi$ act as weighting factor for resource allocation produce on the instantaneous data rate and the past achieved throughput, respectively. Similarly to this approach, in [16] and [17] use adaptive approach which can be able to adjust the fairness level depending on the system requirements. In [18], PF scheme is applied to both time and frequency domain. In time domain the scheduler selects a subset of active users in the current TTI and in frequency domain RBs are allocated to each UE.

**Throughput to Average:** This algorithm can be considered as an intermediate approach between MT and PF. Its priority metric can be expressed as

$$m^k_i = \arg \max_i \left( d_i(t) / \bar{d}_i(t) \right); \ 1 \leq i \leq N$$  \hspace{1cm} (5)$$

where $d_i(t)$ is the expected data rate for $i^{th}$ user at time $t$ on $k^{th}$ RB, $\bar{d}_i(t)$ is the long term average expected data rate for $i^{th}$ user at time $t$ and $N$ is the number of users. From the metric it is easy to understand that the higher the overall expected throughput of UE lower will be its metric on a single RB.

**Modified-Largest Weighted Delay First:** It is channel aware extension of LWDF [19] used in wireless networks. In this algorithm non-real time and real time flows are treated differently [20]. The priority metric can be expressed as follows:

$$p^k_i = \arg \max_i \left( \alpha_i D_i d_i(t) / \bar{R}_i(t-1) \right); \ 1 \leq i \leq N$$  \hspace{1cm} (6)$$

where $\alpha_i$ is weight parameter, $D_i$ is the head-of-line packet delay, i.e., delay of the first packet to be transmitted by $i^{th}$ user, $d_i(t)$ is the expected data rate for $i^{th}$ user at time $t$ on $k^{th}$ RB, $\bar{R}_i(t-1)$ is the average throughput up to time slots $t-1$ and $N$ is the number of users.

This algorithm tries to guarantee good throughput and acceptable level of fairness. A theoretical analysis of M-LWDF fairness is given in [21]. It is shown that M-LWDF fairness depends on the channel condition, packet’s arrival process and the ratio of QoS requirements of different service queues. Based on the theoretical analysis an enhanced M-LWDF algorithm was proposed which improves the fairness compared to M-LWDF algorithm. In [22], a modified version of M-LWDF algorithm based on token mechanism is proposed which gives better performance to real time flows in the DL systems.

**Exponential PF: **EXP/PF algorithm considers both the characteristics of PF for handling non real time flows and exponential function of the end-to-end delay for real time flows. It is first developed to support multimedia applications in time multiplexed systems [23]. It tries to guarantee the delay bound of real time services and maximizes the throughput with acceptable level of fairness. For real time flows the priority metric is calculated as:

$$p^k_i = \arg \max_i \left( \exp \left( \alpha_i D_i - \bar{D}_i \right) d_i(t) / \bar{D}_i \right); \ 1 \leq i \leq N$$  \hspace{1cm} (7)$$
TABLE II. COMPARISON OF DIFFERENT SCHEDULING APPROACHES

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Scheduling parameter</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIFO</td>
<td>Request time</td>
<td>Simple</td>
<td>Inefficient, Channel conditions not known</td>
</tr>
<tr>
<td>RR</td>
<td>Serving time instant</td>
<td>Simple</td>
<td>Inefficient, Channel conditions not known</td>
</tr>
<tr>
<td>WFQ</td>
<td>Priority weight</td>
<td>Introduces priority</td>
<td>Channel conditions not known</td>
</tr>
<tr>
<td>EDF</td>
<td>Delay threshold</td>
<td>Avoids deadline expiration</td>
<td>Channel conditions not known</td>
</tr>
<tr>
<td>LWDF</td>
<td>Acceptable packet loss rate</td>
<td>Provides QoS in terms of delay</td>
<td>Channel conditions not known</td>
</tr>
<tr>
<td>MT</td>
<td>Expected data rate</td>
<td>Maximize overall throughput</td>
<td>unfair</td>
</tr>
<tr>
<td>PF</td>
<td>Expected data rate, Past average throughput</td>
<td>Provides fairness</td>
<td>Low spectral efficiency</td>
</tr>
<tr>
<td>TTA</td>
<td>Expected data rate, Wideband expected data rate</td>
<td>Strong level of fairness</td>
<td>Low spectral efficiency, does not exploit multiuser diversity</td>
</tr>
<tr>
<td>M-LWDF</td>
<td>Head-of-line packet delay</td>
<td>Real time and non real time flows are treated differently</td>
<td>Inefficient in overloaded conditions.</td>
</tr>
<tr>
<td>EXP/PF</td>
<td>Head-of-line packet delay</td>
<td>Real time and non real time flows are treated differently</td>
<td>complex</td>
</tr>
<tr>
<td>EXP rule</td>
<td>Head-of-line packet delay, spectral efficiency of UE</td>
<td>Good scheduling performance</td>
<td>complex</td>
</tr>
<tr>
<td>LOG rule</td>
<td>Head-of-line packet delay, spectral efficiency of UE</td>
<td>Good scheduling performance</td>
<td>complex</td>
</tr>
</tbody>
</table>

where

$$\bar{D}_i = \frac{1}{N_a} \sum_{i=1}^{N_a} \alpha_i D_i$$ (8)

and $N_a$ is the number of active real time DL flows, $\alpha_i$ is the weighting parameter, $D_i$ is the head-of-line delay, $d_i^k(t)$ is the expected data rate for $i$th user at time $t$ on $k$th RB, $R(t-1)$ is the average throughput up to time slots $t-1$. PF algorithm handles non real time flows.

**Exponential rule:** This algorithm has been presented in [24]. This algorithm is enhancement of EXP/PF. Its priority metric is given as

$$p_i^k = \arg \max_j \left( b_i \exp \left( \frac{a_i D_i}{c + \sum_{j \neq i} \alpha_j D_j} \right) \Gamma_i^k \right) \quad 1 \leq i, j \leq N$$ (9)

where $a_i$, $b_i$, and $c$ are optimal parameter set for the system requirements, and $N_a$ is the number of active real time DL flows, $D_i$ is the head-of-line delay, $\Gamma_i^k$ is the spectral efficiency of $i$th user on $k$th RB and $N$ is the number of users. EXP rule takes into account the overall network status, because that the delay of the considered UE is normalized over the sum of experienced delays of all UEs. In [25] author proposes an interesting procedure based on cooperative game theory that performs resource sharing based on EXP rule with virtual token mechanism.

**Logarithmic rule:** This algorithm has been also presented in [24]. This algorithm is similar to that of EXP rule but it uses logarithmic function of delay to calculate the scheduling metric.

$$p_i^k = \arg \max \left( b_i \log \left( c + a_i D_i \right) \Gamma_i^k \right) ; 1 \leq i \leq N$$ (10)

where $b_i$, $c$ and $a_i$ are tunable parameters, $D_i$ is the head-of-line delay, $\Gamma_i^k$ is the spectral efficiency of $i$th user on $k$th RB and $N$ is the number of users. optimum throughput and fairness can be achieved by taking suitable values for $b_i$, $c$ and $a_i$. In [26], the paper considers the design of multiuser opportunistic packet scheduler for UEs sharing a time-varying wireless channel based on LOG rule. It is concluded that a scheduler optimized for the overall system performance is likely to be more robust to changes in the traffic and channel statistics than the one optimized for the worst case. Table 2 summarizes the comparison of main scheduling algorithms used in wireless networks including LTE.

Apart from above mentioned scheduling algorithms many algorithms are proposed in literature. Some considers buffer status, some considers energy savings and some presents the combinations of algorithms.

**Delay based algorithms:** In [27], a cross layer algorithm is proposed as an optimization problem to minimize the average delay under the constraints of the transmission power and block error rate requirements. In [28], the author associates a delay function to each data packet. That is, longer the delay higher the probability of the packet to be allocated. Delay based prioritized scheduling approach is proposed in [29]. In this paper, the algorithm first orders the users depending on the remaining time before deadline expiration. Once the urgent users were identified the frequency allocation step is performed in order to transmit the head of line packet. Frequency-time scheduling approach with delay constraints is proposed in [30] to handle streaming services like multimedia services. The QoS for this kind of service can be attained at the expense of overall system capacity. The proposed algorithm tries to balance the QoS of service without losing much cell capacity.

**Power based algorithms:** Transmit power based scheduling algorithm is proposed in [31], where the scheduling metric is based on the ratio of transmit power per bit and allocates resources. A resource allocation problem is formulated in [32], with constraints on power, rate and delay. The analytical design is used to evaluate the system performance under different network parameters. Imperfect channel state information is considered in problem formulation. Based on the matrix analysis the formulated problem is transformed to optimization problem. The solution obtained was complex and considered as impractical one. So a heuristic
solution is proposed with lower complexity to achieve acceptable system performance.

**CQI feedback based algorithms:** In LTE, CQI feedback enables the scheduler to achieve multiuser diversity gain by allocating resources to UEs with good channel condition leading to unfairness. A multiuser scheduling method is proposed in [33] which is based on fairness utility function and aims to provide resources to the cell edge users also. A priority based approach is presented in [34] where the priority sets are created based on CQI of each bearer and classified as GBR and non-GBR set. By using FDPS, GBR set is allocated RBs and the remaining RBs are allocated to non-GBR set based on PF algorithm.

Service based algorithms: A QoS aware scheduling algorithm for downlink transmission with emphasis for support in overload state is presented in [35]. In this method multiple real time traffic metrics are processed through array of ranking functions and then they are multiplexed to aggregate ranking function. This ranking function is utilized in scheduling function. A different approach is followed in [36], where authors develop two level framework that guarantees delay to real flows. At upper level, discrete control theory is applied to every frame to calculate the total amount of data of real time flows which is having delay constraints. At lower level PF metric is used to satisfy the non real time flows. It is concluded that this algorithm is suitable for real time video flows.

In a similar manner [37] designed an algorithm which consists of three phases. In frequency domain, a method is developed to utilize the RBs in an effective manner. In time domain, a method is proposed for predicting the packet delays for different applications in the queue. Then with the calculated results, a method is proposed which rearrange the transmission order and discard the packets which do not meet the delay requirements. It is claimed that this method is suitable for real time services. In [38] dynamic scheduler for VoIP is proposed. It facilitates priority during only VoIP can be allocated. For QoS aware packet scheduler, delivering packets within the pre-fixed expiration is the fundamental characteristics. To satisfy the QoS requirements, [39] proposed a method called Token Bucket Scheduler (TBS). TBS utilizes instantaneous DL channel Signal to Interference plus Noise Ratio (SINR) and QoS information while allocating RB to real time services.

**Queue based algorithms:** A cross layer design approach is presented in [40], which makes use of both users’ queue states and channel states in allocation of resources. This method has low computational complexity since it considers users’ minimum data rate and target bit error rate as QoS parameters. Similar to this approach the algorithm proposed in [41] allocates resources according to CQI feedback from the physical layer and considers buffer status of user to avoid buffer overflow to guarantee fairness among users.

**QoE based Algorithms:** In [42] the paper deals with the human perceived quality maximization, when scheduling multimedia traffic. Work has been carried out in two levels. First, the Quality-of-Experience (QoE) aware scheduling problem is formulated as a suitable Markov Decision Process (MDP) model. Then to solve the problem, a simple heuristic rule is designed. This paper claims that QoE experienced by users outperforms most scheduling techniques.

**IV. PERFORMANCE ANALYSIS OF MAIN SCHEDULING STRATEGIES**

From user’s point of view, the important characteristic of a network related to its performance is its QoS. QoS is analyzed by means of many parameters like goodput, delay, Packet Loss Ratio (PLR), fairness index etc. In this section, the main scheduling algorithms used in LTE network are analyzed by using an open source simulator, LTEsim [43]. The modules present in LTEsim are explained in [44].

A. Simulation Scenario

For performance analysis, a single cell scenario with fixed eNB is considered where the users are uniformly distributed among the cell. Two user’s speed (30 knmph, 120 knmph) are considered for analysis using random mobility model [45]. Each user receives one video flow, one VoIP flow and one Best Effort (BE) flow at the same time. The buffer at the scheduler is considered to be infinite i.e., the packet loss is not due to the buffer overflow.

Video service has 128 kbps source rate produced by H.264 coder. For VoIP, G.729 voice flows are used with source rate of 8 kbps. Finally, best effort traffic corresponds to the ideal source that always has packet to send. In real time services, the maximum delay should be in the range of 100-200 ms. Accordingly, the target delay is set to 0.1 s. The main simulation parameters used for analysis is shown in Table.3

<table>
<thead>
<tr>
<th>TABLE III. SIMULATION PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Carrier frequency</td>
</tr>
<tr>
<td>Frame structure</td>
</tr>
<tr>
<td>Channel model</td>
</tr>
<tr>
<td>Path loss</td>
</tr>
<tr>
<td>Penetration loss</td>
</tr>
<tr>
<td>Mobility model</td>
</tr>
<tr>
<td>Speed of UE</td>
</tr>
<tr>
<td>UE application flow</td>
</tr>
<tr>
<td>Maximum delay</td>
</tr>
<tr>
<td>Simulation duration</td>
</tr>
<tr>
<td>Traffic model</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

B. Results and Discussion

As a result of simulation, the performance analysis is performed in terms of average goodput, PLR and fairness index by increasing the number of users with the step of ten. A scheduling algorithm should balance its allocation method between fairness and throughput maximization in order to guarantee at least minimum data rate to the cell edge users. The fairness index for video, VoIP and BE service is shown in Figs 5.
For real time flows, it is observed that all the schedulers except PF show comparable level of fairness closer to 0.85 then decrease with increasing users. Since PF algorithm goes on balancing between the data rate and fairness, it shows lower fairness index for real time flows. The fairness index experienced by the non-real time flows is lesser than that of real time flows because of its low priority.

Regarding real time flows PLR is the standard metric used to evaluate the QoS offered by the network. Figs. 6 and 7 show the PLR achieved for video and VoIP flow respectively.

It is observed that VoIP flows experience considerably smaller PLRs than video. This is due to the fact that VoIP flows having a lower source bit rate get higher priority from the schedulers. It is also noted that as the user speed increases, PLR also increases for all investigating algorithms.

For video flows, the PLR is higher for PF algorithm when compared with other algorithms. In low mobility scenario, PLR is about 0.2-1% for lesser number of users and increases with increasing users. PF algorithm drops 40% of forwarded packets and other algorithms used for analysis, drop 5-10% of packets for maximum number of users used in this simulation. When the speed is increased to 120 kmph PF algorithm drops 90% of forwarded packets and other algorithms drops 50-55% of forwarded packets for maximum number of users used in this simulation. For VoIP flows with low mobility EXP rule algorithm gives good performance of 0-0.03% of packet drop and other algorithm achieves 0.04-0.08% of packet drop for maximum number of users used in the simulation. For the speed of 120 kmph EXP rule gives 0.5-1% packet loss and other algorithm achieves 2-3% packet loss for maximum number of users used in this simulation.

The average goodput for video, VoIP and BE flows with two different user speeds are shown in Figs. 8, 9, 10. Average goodput is defined as the rate of useful bits successfully transmitted during the entire simulation.
Average goodput is limited by packet loss and delay. From the figures it is observed that goodput decreases with increasing users. The rate of reduction is higher in case of PF algorithm. In high mobility scenario, users are affected by lower average goodput which affects the service quality for real time services. It is observed that EXP rule algorithm gives better performance when compared with other algorithms for real time services. For non real time services, all algorithms provide similar performance.

V. FUTURE CHALLENGES

Though LTE overcomes the performance of current 3G systems, it is not suitable for future data traffic requirements which have been given in IMT-Advanced [46]. For fulfilling the requirements of IMT-Advanced, LTE-Advanced is introduced by 3GPP. To meet the target requirements LTE-Advanced is standardized by many technologies like carrier aggregation, enhanced multi-antenna support, Coordinated Multi-Point (CoMP) transmission techniques, relaying and Heterogeneous Networks (HetNet) deployment [47].

In order to utilize the wider bandwidth up to 100 MHz and keeping backward compatible with LTE, a carrier aggregation scheme is proposed. Carrier aggregation consists of LTE component carriers (CCs) so that the devices can be able to use a greater amount of bandwidth. A device capable of carrier aggregation has one primary CC and one or more secondary CCs. Carrier aggregation may be of contiguous or non-contiguous. An eNB can use up to 5 adjacent channels of 20 MHz to increase the network capacity [48]. In a scenario where both LTE and LTE-Advanced users are present, the RRM at MAC layer should be able to differentiate among these users. In particular, LTE-Advanced users should be assigned multiple CCs while LTE users should be assigned single carrier. Scheduling for LTE-Advanced users with carrier aggregation can be done in two ways: same carrier scheduling and cross carrier scheduling.
In same carrier scheduling, separate PDCCH is used for each CC and in cross carrier scheduling, common PDCCH is used for multiple CCs. The challenge of designing the scheduler lies in allocating CCs and resources to the users.

CoMP is contemplated as an essential technique to alleviate inter-cell interference and to improve the cell-edge performance in LTE-Advanced networks. It makes use of multiple transmit and receive antennas from multiple eNBs to enhance the signal quality and to decrease the interference. The resource allocation for this technique needs to be synchronized among different eNBs. CoMP in downlink can be classified into two schemes. Coordinated Scheduling and/or Beam-Forming (CS/CB) and Joint Processing (JP). In CS/CB, the signaling and resource allocation to a single user are performed from the serving cell. However the scheduling is dynamically coordinated between the cells. In that manner, the interference between different transmissions can be decreased. The scheduler at eNB makes its decisions independently but additional information about other user’s channel condition is needed to perform more optimal scheduling. In JP, signaling and resource allocation to a single user is concurrently performed from multiple transmissions to optimize cell edge performance. This scheme is more challenging because the scheduling decisions must be exchanged over the backhaul. So the scheduling algorithms ensuring eNB synchronization need to be designed [49].

With the introduction of low power eNBs like micro, pico, femto, the demand for different scheduling methods has been increased. Combining macro cells with low power nodes will overcome the problem of coverage holes which improves the capacity in hot spot areas. The low power nodes would be able to serve the limited number of UEs with less coverage area. All the scheduling methods which were previously discussed can be used for heterogeneous networks by taking into account, the interference produced by macro cells [50]. The method of simulating LTE femtocells using open source simulator is presented in [51] which shows that throughput is maximized with the introduction of low power nodes.

To increase the capacity of a cellular network, large number of base stations have to be deployed which is very expensive. For this purpose LTE-Advanced introduced a technique called relaying. eNB can forward the information through relay node to remote user with high data rate, thereby reducing the deployment cost. The source eNB from which the relay node receives the signal is called donor eNB and the link between the donor eNB and relay node is called backhaul link. The link between the relay node and the end user is called access link. Since the control and data transmissions are carried out through backhaul link and then access link, the scheduler in the donor eNB should take into account the additional link delay also. The resource allocation algorithm for relaying may be centralized or distributed. In centralized resource allocation, the donor eNB is in charge of doing resource allocation to both end user and relay node. In distributed resource allocation, relay node is incharge of doing similar resource allocation as done by donor eNB with a constraint that it has limited resources. The combination of OFDMA with relaying techniques provides increased opportunities for cost-effective and high-performance networks. To make use of such opportunities requires intelligent RRM algorithms [52].

VI. CONCLUSION

The 3GPP LTE standards aim to achieve revolutionary data rate, spectral flexibility with seamless mobility and enhanced QoS over the entire IP network. In this paper, a study of downlink scheduling algorithms for LTE networks has been carried out. The most important RRM task is performed by the packet scheduler who distributes radio resources among UEs in efficient way. This paper identifies the strength and weakness of well known algorithms in the downlink LTE system and the key aspects that should be taken into account when designing a new algorithm. It can be inferred that PF algorithm is more suitable for non real time services because it does not account packet delays during its decision making. On the other hand the other schedulers such as M-LWDF, EXP/PF, EXP rule and LOG rule are better choice for real time services. Combinations of algorithms can also be carried out to find the optimal solution according to the network requirements. These algorithms can also be used to LTE-Advanced with some modifications according to the techniques used.

REFERENCES


UTRAN; Medium Access Control (MAC) protocol specification (Release 9), 3GPP TS 36.321.
[39] Liqun Zhao, Yang Qin, Maohe MA, Xiaoxiong Zhong, Li Li,” QoS Guaranteed Resource Block Allocation Algorithm in LTE Downlink” IEEE International


S. Fouziya Sulthana received her B.E degree in Electronics and Communication from Bharadidasan University, Tamilnadu in 2002 and M.Tech degree in electronics and Communication from Pondicherry University, Puducherry in 2008. She is student member of IEEE and IEICE. She worked as Assistant Professor in private engineering college. She is currently a PhD student in Pondicherry University. Her current research focuses on Radio resource Management in broadband networks.

Dr. R. Nakkeeran received B.Sc. degree in Science and B.E degree in Electronics and Communication Engineering from Madras University in 1987 and 1991 respectively and M.E degree in Electronics and Communication Engineering (with diversification in Optical Communication) from Anna University in 1995. He received PhD degree from Pondicherry University in 2004. Since 1991, he has been working in teaching profession. Presently, he is Associate Professor in the Department of Electronics Engineering, School of Engineering and Technology, Pondicherry Central University. He is life member of IETE, ISTE, OSI and fellow of IE (I). Also, he is member of IEEE, OSA, IEICE and SPIE. He has published around 175 papers in National and International Conference Proceedings and Journals. He has co-authored a book, published by PHI. His areas of interest are Optical Communication, Networks, Antennas, Electromagnetic Fields and Wireless Communication.