Low Complexity Cross-Layer Scheduling and Resource Allocation for VoIP in 3G LTE

Richard Musabe and Hadi Larijani
Dept. Computer, Communication, and Interactive Systems (CCIS)
School of Engineering and Built Environment
Glasgow Caledonian University, Glasgow, Scotland, UK.
E-mails:richard.musabe@gcu.ac.uk, h.larijani@gcu.ac.uk

***Abstract*—3G Long Term Evolution (LTE) is an emerging and promising technology that aims at providing broadband ubiquitous Internet access and improving multimedia services. This is achieved through streamlining the system for packet services, since LTE is an all Internet Protocol(IP) based network. The fact that 3G LTE is a packet based network brings about some improvements in the form of higher bit rates, lower latencies, and a variety of service offerings. However, some technical challenges are expected to arise when voice traffic is transmitted over an LTE network. This has become an interesting area of research and different types of resource management schemes have been developed, which are quite challenging and complex. In this paper, we have projected the voice packet scheduling and resource allocation problem as a constrained optimization problem. Our optimization objective is formulated using channel state information such as, transmission rate at the physical layer as well as the queuing state information like queue length at the MAC layer. We provide the algorithmic implementation of the obtained solution and also investigate the performance, complexity, and fairness of our proposed cross-layer scheduling algorithm under different conditions such as, VoIP delay, packet loss, etc. We compared it with other algorithms in literature such as, proportional fair (PF) and exponential proportional fair (EXP-PF). Based on the numerical and simulation analysis, we found that our proposed algorithm performed better than PF and EXP-PF in most cases.**

***Keywords*—LTE, Scheduling Schemes, VoIP, Complexity, Utility Function.**

I. INTRODUCTION

In this paper, we investigate the performance, complexity, and fairness of our proposed cross-layer scheduling algorithm for Voice over Internet Protocol (VoIP) in 3G LTE. This work is an extension of the analysis done in [1]. 3G LTE was identified by the third generation partnership project (3GPP) as the preliminary version of next generation wireless communication systems because of its high data rates [2]. 3G LTE technology provides a maximum 100Mbps downlink and 50Mbps uplink while using 20 MHz bandwidth [3]. In the downlink physical layer, LTE uses Orthogonal Frequency-Division Multiple Access (OFDMA) radio technology to meet the LTE requirements for spectrum flexibility and enables cost-efficient solutions for wide carriers with high peak rates. In the uplink, LTE uses a pre-coded version of OFDMA known as Single-Carrier Frequency-Division Multiple Access (SCFDMA), in order to compensate for a drawback with normal OFDMA of a high Peak-to-Average-Power Ratio (PAPR) [4].

Wireless technology has expanded from *voice only*: to high-speed data, multimedia applications, and wireless Internet [5]. LTE requirements for high data rates are achieved by the fact that this technology is only designed for packet switched networks (PSN): hence, there is no need for the circuit switched mode. However, this design brings with it more technical challenges especially for voice services. VoIP services are both delay and packet loss sensitive. The biggest challenge of VoIP over LTE is to deliver Quality of Service (QoS). Normally users would expect voice with the same quality as that provided by circuit switched networks. However, traffic delivered over PSNs are subject to delay and packet loss [6]. A major issue with VoIP over LTE is that 3G LTE adopts a different method of resource transmission from other cellular systems like Code Division Multiple Access (CDMA). 3G LTE uses Physical Resource Blocks (PRB) as its transmission unit. PRBs can be defined as the basic unit with both frequency and time aspects [7]. Basically, the base station of 3G LTE, known as eNodeB has a fixed number of available PRBs according to their allocated bandwidth and it is supposed to assign PRBs repeatedly at every Transmission Time Interval (TTI) [3].

Our contributions in this paper are:

- Formulating the problem of scheduling and resource allocation using utility function optimization by extending the proposed approaches in [8][9] to include VoIP metrics
- Present a mathematical model for our extended version of problem formulation
- Use this technique to theoretically analyze the performance, complexity, and fairness of our proposed algorithm in [10] based on transmission rate, queue delay, and queue length parameters
- Through numerical and simulations analysis, we studied the performance, complexity, and fairness of the proposed algorithms
- Based on the numerical and simulation analysis, our proposed algorithm performed better than other algorithms proposed in literature in most cases

The rest of the paper is organized as follows: Section II discusses the related work, Section III analyzes VoIP QoS in 3G LTE, Section IV discusses the the system model, where we discuss the general problem formulation and the extended version of the problem formulation. Section V gives the summary of our proposed algorithm, PF, and EXP-PF. Section VI presents the simulation details where we discuss the PRB characteristics and scenario setup, Section VII presents the results analysis, which include numerical, performance, complexity, and fairness analysis of our proposed algorithm. Section VIII reviews the main conclusions and future work.

II. Related Work

Different techniques have been introduced in the literature to overcome the challenges faced when real time traffic is transmitted over an LTE network.

In [8], Jianwei Huang *et al.* addressed the gradient-based scheduling and resource allocation problem for the downlink OFDM system. They considered various practical features such as, integer tone allocation, different sub-channelization schemes, maximum SNR constraint per tone, self noise due to channel estimation errors, and phase noise. During each time slot, a sub-set of users are scheduled and the available tone and transmission power is allocated to them. Using the gradient based approach, they reduced this problem into an optimization problem, which can be solved in each time slot. Using the dual formulation, they were also able to give an optimal algorithm for this problem when multiple users can time share each tone. Their approach motivated us to further address the problem of gradient-based scheduling and resource allocation. We specifically focused on VoIP packets when transmitted over an LTE network. In our approach, we considered various parameters provided in channel state information such as, transmission rate. We also used the parameters provided at the Medium Access Control (MAC) layer (i.e., queue length) and the VoIP QoS requirements (such as, delay parameters).

In [11], Yaacoub *et al.* proposed two low complexity heuristic algorithms. The complexity of both algorithms was analyzed. The first algorithm had a linear complexity in the number of users and quadratic complexity in the number of resource blocks. The second algorithm had a linear complexity in both the number of users and resource blocks. It was shown that good results could be achieved by the proposed linear complexity algorithm (second algorithm). It was also shown through simulations that the maximization of total throughput leads to a higher cell throughput, although considering the logarithm of throughput as a utility function ensures proportional fairness, and thus constitutes a tradeoff between throughput and fairness.

In [12], Zhao *et al.* investigated two fairness criteria with regards to adaptive resource allocation for uplink OFDMA systems. These two criteria were Nash Bargaining Solution (NBS) fairness and proportional fairness (PF). These two criteria can provide attractive tradeoffs between total throughput and each user's capacity. Using Karush-Kuhn-Tucker (KKT) condition and iterative methods, two effective algorithms were designed to achieve NBS fairness and proportional fairness respectively. Through simulation results, NBS fairness criteria showed better performance in total capacity but the BS could not control the rate ratio because it only depends on the channel state of the users. PF criteria can provide a controllable rate ratio regardless of the channel condition for each user. However, to achieve the hard fairness, the system capacity degrades sharply.

With all these techniques introduced in the literature, there are still some challenges when real-time traffic such as, voice is transmitted over an LTE network. This is mostly due to the fading channels of wireless links and the delay and packet loss sensitive voice characteristics. So in this work, we extended the work in [8], where the problem of scheduling and resource allocation was formulated using the utility function optimization approach. We introduce the VoIP metrics to this approach and determine the resource allocation for VoIP users instead of power allocation, which was considered in [8]. Voice packet scheduling has some particular requirements such as, minimum end-to-end delay requirements, subchannel or subcarrier allocation constraints, etc.

We have projected the voice packet scheduling and resource allocation problem as a constrained optimization problem.This optimization objective aims at maximizing the expected total utility under different constraints. We implemented an algorithm for the proposed solution and analyzed its performance, complexity, and fairness. We then compared it with other algorithms in [13]. The simulation results were generated using the open source LTE system simulator called LTE-SIM [13]. It models different uplink and downlink scheduling strategies in multicell/multiuser environments: taking into account user mobility, radio resource optimization, frequency reuse techniques, the adaptive modulation, and coding (AMC). It is important to analyze the QoS requirements for voice when transmitted over an LTE network. This will give us an idea of what voice quality the end user can expect during the VoIP call.

III. VoIP QoS Analysis in LTE

A. VoIP in LTE Traffic and protocols

Conversation VoIP traffic in LTE can be assumed as the two state Markov model with a suitable voice activity factor (VAF). Different open source Codecs can be used in LTE but the most popular codec according to [2] is Adaptive Multirate (AMR). This codec provides 32-bytes voice payload in every 20 milliseconds while talking and 7-bytes payload every 160 millisecond while silent. The VoIP protocol stack, which utilizes the real-time transport protocol (RTP) is encapsulated to the user datagram protocol (UDP), which is in turn carried by IP. The use of all these protocols brings the total header size to 40 bytes for IPv4 header or a 60-bytes for IPv6 header. The overhead brought about by these headers causes serious degrading in the spectral efficiency supporting VoIP traffic in LTE. So to solve this problem, an efficient and robust header compression (ROHC) technique is used. This technique solves the overhead problem by minimizing the size of the IP/UDP/RTP headers to as little as 2 or 4 bytes using IETF RFC 3059 [7][14].

B. VoIP End-to-End Delay and Capacity

The main characteristic of voice traffic is strict delay requirements [15], according to [16] the allowed maximum mouth-to-ear delay for voice is 250ms. This delay requirements includes the assumption that the delay for the core network is approximately 100ms. The tolerable delay for radio link control (RLC), medium access control (MAC) buffering, scheduling, and detection should be strictly less than 150ms, this has been depicted in Figure 1 [2]. If we take into account that both end users are LTE users, then we can assume that the tolerable delay for buffering and scheduling is less than 80ms. To better account for unpredictability in network end-to-end delays, 3GPP performance evaluation has also chosen the delay of 50ms from eNodeB to UE [15]. Packets will be dropped when packet error and packet delay exceeds the target latency while VoIP traffic is transmitted over an LTE network. This may not affect the voice quality if the packet loss is less than outage threshold [15]. The outage limit means that packet error rate (PER) of VoIP users is kept below 2%. This gives us the actual limit that the maximum VoIP capacity for LTE is limited by the outage limit, which is described in TR 25.814 [7] and was later updated in R1-070674 [17]. We can finally describe VoIP capacity in LTE as the maximum number of VoIP users that can be supported without exceeding a given threshold and at least 95% of total VoIP users should meet the above described outage limits [2].

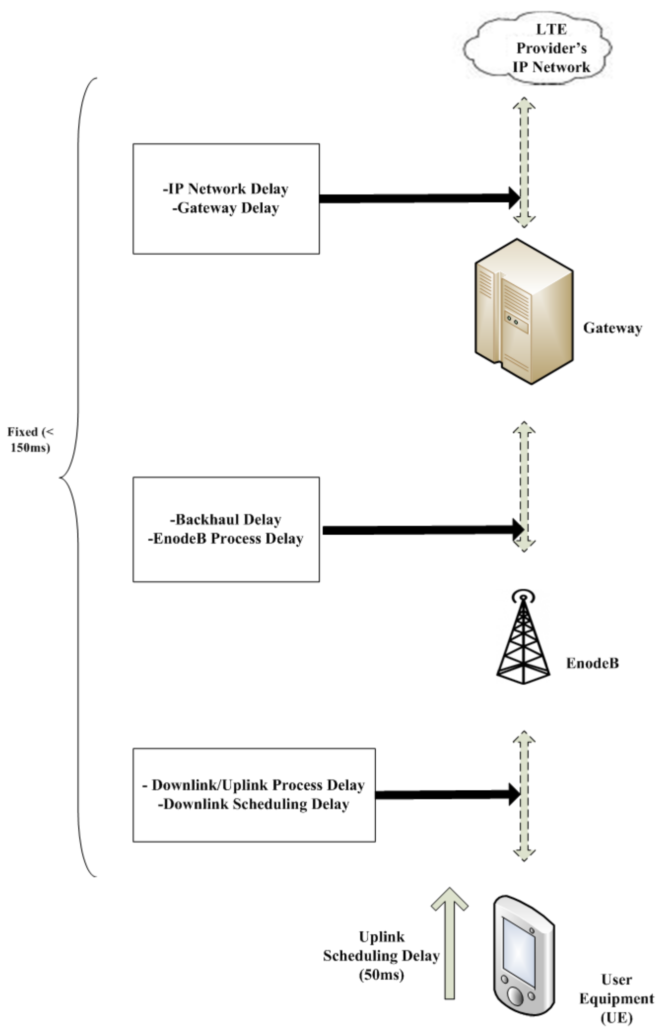

Figure 1 – VoIP End-to-End Delay Components in LTE

IV. SYSTEM MODEL

A. General Problem Formulation

In [8], the authors considered the downlink transmissions in an OFDM cell with base station and number of users. The authors considered K to be the maximum number of available users such that the number of users range from 1 to K, i.e., $k = \{1, ..., K\}$. So in every time slot, scheduling and resource allocation decision was done by choosing the rate vector $r_t = (r_{1,t}, ..., r_{K,t})$ from $R_{e_t} \subseteq R_+^K$, where e_t is the time varying channel state information available at time t. In short, the general problem is to find $r_t \epsilon R(e_t)$ that can maximize the system utility function $U(W_t) := \sum_{i=1}^{K} U_i(W_{i,t})$, where $U_i(W_{i,t})$ is the increasing concave utility function of user i's average throughput $W_{i,t}$, up to time t.

B. Our Extended Version of the Problem Formulation

1) utility function: Before extending the problem formulation in [8] to include VoIP metrics and other parameters, let us first describe the utility function. Utility functions can be useful in cross layer optimization as they can map network resources utilized by users into real numbers. The utility function can also indicate the level of satisfaction of the user, which in turn helps in the balancing the efficiency and fairness between the users. In 3G LTE, such as most wireless communication technologies, consistent transmission rate is the main factor that can determine the level of satisfaction of the user. So if we take m_j to be the transmission rate vector, then its utility $U(m_t)$ should be a nondecreasing function of the transmission rate m_j.

We adopted the utility function calculation from [9] and used it for the transmission rate m_t as:

$$U(m_j) = X_j\{\frac{1}{1 + e^{-p_j(m_j - R_j)}} - Y_j\} \quad (1)$$

With

$$X_j = \frac{1 + e^{p_j R_j}}{e^{p_j R_j}} \quad (2)$$

and

$$Y_j = \frac{1}{1 + e^{p_j R_j}} \quad (3)$$

where $U(m_j)$ is the utility function of user j with respect to their transmission rate. p_j is the priority tag assigned to VoIP users. R_j is the available resource blocks. X_j and Y_j are constants used to normalize the utility function.

2) Optimization Problem Formulation: The main aim of this problem formulation is to map the network resources of each user to their corresponding utility values. After that, the established utility function is optimized. Let K indexed by j, be the maximum number of available users such that $j \in \{1, ..., K\}$. If we consider the utility function of user j to be $U_j(.)$, then if user j has the transmission rate as m_j, we can say that the utility of user j is $U_j(m_j)$. Again if we let Q_l to be the length of user j's queue and Q to be the total number of queues for user j. Q is index by i, so $i \in \{1, ..., Q\}$. Then, the total utility function of user j is calculated from the utility function of its queue.

$$total_{utility} = Q_l * U_j(p_j R_j m_j) \quad (4)$$

where $U_j(p_j R_j m_j)$ can be equal to $U(m_j)$ in equation (1). If we take all the user j's queues in the network, then

$$Total_{utility} = \sum_{i=1}^{Q} Q_l * U_j(p_j R_j m_j) = \sum_{i=1}^{Q} Q_l * U_j(m_j) \quad (5)$$

So, our problem is to find a VoIP user that can maximize the total utility with respect to transmission rate and user queue values:

$$Max \sum_{i=1}^{Q} total_{utility} = Max \sum_{i=1}^{Q} Q_l * U_j(m_j) \quad (6)$$

However, the fact that we are dealing with VoIP application means that we need some constraints to control its QoS requirements. So, the above optimization objective equations should be subject to: $m_j \leq NC$ and $Q_d \leq D_{max}$. Where NC is the total available network capacity, Q_d is the queuing delay, and D_{max} is the maximum allowable mouth to ear delay.

Having seen that the main metrics in our problem formulation procedure are transmission rate (m_j), and queue length (Q_l), it is very important to know how these two metrics are obtained. This is described below.

3) Finding the Transmission Rate (m_j): During every transmission process, the user sends its instantaneous achievable signal to noise ratio (SNR) to its eNodeB. This value keeps on changing depending on different factors like mobility, selective fading channels, etc. So according to [9], user j's transmission rate at time t can be calculated as:

$$m_j(t) = \frac{n_{bits}}{symbols} * \frac{n_{symbols}}{slot} * \frac{n_{slots}}{TTI} * \frac{n_{subcarriers}}{RB} \quad (7)$$

where n_{bits}, $n_{symbols}$, n_{slots}, and $n_{subcarriers}$ are respectively the number of bits, number of symbols, number of slots, and number of subcarriers according to the PRB characteristic described earlier. TTI is the time transmission interval and RB is the resource blocks. These PRB characteristics are affected by path loss and fading channels but its values are kept constant for the entire PRB transmission time. According to [18], the channel gain of user j on a PRB at time t as a function of loss is calculated as:

$$CN_{gain_j}(t) = 10^{\frac{pathloss}{10}} * 10^{\frac{fading}{10}} \quad (8)$$

It should be noted that both $pathloss$ and $fading$ are measured in dB scale. Using this channel gain (CN_{gain}), the user knows the instantaneous SNR to send to eNodeB. Again according to [19], this SNR can be calculated as a function of CN_{gain}:

$$SNR_j(t) = \frac{P_{total} * CN_{gain}(t)}{R(N_o + I)} \quad (9)$$

Where P_{total} is the power with, which the eNodeB transmits, R is the total number available PRBs, I is neighboring interference, and N_o is the thermal noise measure.

4) Finding the Queue Length (Q_j): In order to obtain queue length metric, we adopted the queuing method in the LTE-SIM simulator. In this method, different traffic generators were developed, these generated packets that are transported by a dedicated radio bearer at the application layer. Using the application class, we were able to generate the packets and deliver them to the network. Once the packets reach the network, they are forwarded to the user-plane protocol stack to add protocol headers. Then, the packets are placed in the queue by the MAC queue class at the MAC layer before being sent to the destination. The MAC queue object has a counter, which increases or decreases when the packet is inserted or removed from the queue respectively.

Let $Q_j[l]$ be the amount of packets in user j's queue at time T_s. So, if the base station serves user j at rate $r_j[n]$ in time slot n, then user j's queue length at time $(n+1)T_s$, is expressed as:

$$Q_j[n+1] = Q_j[n] - r_j[n]T_s + a_j[n] \quad (10)$$

where $a_j[n]$ is the amount of arrival bits during time slot n.

5) Solution Approach: In order to solve our optimization problem in equation (6) that will maximize the network utility, we used the dual decomposition approach with lagrange multipliers. Solving this equation determines the user to be scheduled and assigned resource blocks according to the transmission rate (m_j) and queue values (Q_l, Q) parameters subject to $m_j \leq NC$ and $Q_d \leq D_max$ constraints. Writing up the optimization problem as a lagrange dual function, it becomes:

$$L(m_j, Q_d, \lambda, \mu) = \begin{array}{l} \sum_j U(m_j) + \lambda(NC - \sum_j m_j) + \\ \mu(D_{max} - \sum_j Q_d) \end{array} \quad (11)$$

The corresponding dual function can be written as:

$$L(\lambda, \mu) = MAX_{m_j, Q_d} L(m_j, Q_d, \lambda, \mu) \quad (12)$$

The inequality constraints in the optimization problem are put under consideration by augmenting the objective function with a weighted sum of the constraint function. Therefore, λ is called the lagrangian multiplier associated with $m_j \leq NC$ constraint and μ is the lagrangian multiplier associated with $Q_d < D_{max}$ constraint.

If we divide the objective function above into $|\lambda|$ and $|\mu|$ separate subproblems, then each subproblem can be solved separately if the values of λ and μ are known. the objective function of the dual problem then becomes:

$$D(\lambda) = MAX_{m_j \varepsilon NC} L(NC, \lambda) \quad (13)$$

and

$$D(\mu) = MAX_{Q_d \varepsilon Q_{max}} L(Q_{max}, \mu) \quad (14)$$

V. Summary of the Scheduling Algorithms

In this section, we are going to investigate how three different scheduling algorithms assign resources to their users in order to maximize their utility function. These algorithms are: our proposed algorithm, Proportional Fair algorithm (PF), and Exponential/Proportional Fair algorithm (EXP/PF).

A. Our Proposed Algorithm

Detailed explanations of this algorithm can be found in [10]. In our proposed algorithm, the scheduler assigns resources once every TTI and based on the user's current transmission rate (m_j), queuing delay (Q_d), and queue length (Q_l).

1) Summary of our Proposed Algorithm: The proposed algorithm is based on the modifications to the problem formulation in [8] and utility calculation in [9]. We introduced new parameters such as, transmission rate (m_j), queuing delay (Q_d), and queue length (Q_l). The first part of our algorithm is to determine the scheduling order for VoIP users. This can be done by ordering the users according to their decreasing sequence of their queuing delay (Q_d), and queue length (Q_l). Once the scheduling order is determined, the resource allocation is done by taking each user and determining the parameters that can maximize the utility of transmission rate (m_t). In order not to starve other applications in the network, we used the adaptive method proposed in [3].

This method provides limits to our proposed scheduling algorithm, which is adaptively changed between a pre-specific minimum and maximum according to the ratio of dropped packets. Higher drop ratio means that there are many ongoing VoIP calls, and hence it is necessary to increase the limits to allow more consecutive TTIs to be dedicated to VoIP calls. On the other hand, low drop ratio implies that QoS of VoIP calls are satisfied at decent levels, and thus it is safe to reduce the duration of the algorithm and serve other applications in the network.

2) Steps Involved in our Proposed Algorithm:

- **Step 1**: Determine the procedure of inserting users/packets into their queues
- **Step 2**: Scheduling starts at every TTI
- **Step 3:** Find out if there are any VoIP users/packets in the queues
- **Step 4:** If there are VoIP users/packets in the queue, apply our proposed algorithm and go to the next step otherwise apply the normal scheduling algorithm and exit
- **Step 5:** Arrange the VoIP users according to their decreasing values of their queuing delay (Q_d), and queue length (Q_l).
  Then, initialize $j = 1$, $m_{ext} = m_{max}$ and $Q_{ext} = Q_{max}$. Where m_{ext} and Q_{ext} are the extra/remaining transmission rate and queue length values at each stage
- **Step 6:** Determine if the successive counts of our proposed algorithm are not greater than the provided adaptive limit
- **Step 7:** If it is not greater than the limits then go to the next step, otherwise apply the normal algorithm, i.e., default algorithm such as, FIFO and exit
- **Step 8:** Find the parameter that maximizes the utility function for VoIP user j with $m_j \leq m_{ext}$ and $Q_j \leq Q_{ext}$
- **Step 9**: Schedule this VoIP user
- **Step 10:** Reduce m$_{ext}$ by m_j and Q$_{ext}$ by Q_j respectively
- **Step11:** If more resources blocks(RBs) and VoIP users exist, as well as $m_{ext} > 0$, $Q_{ext} > 0$ then set $j = j+1$ and repeat from step 8. If any of the three checks fails, then exit

The algorithm flow chart is presented in the Figure 2 below.

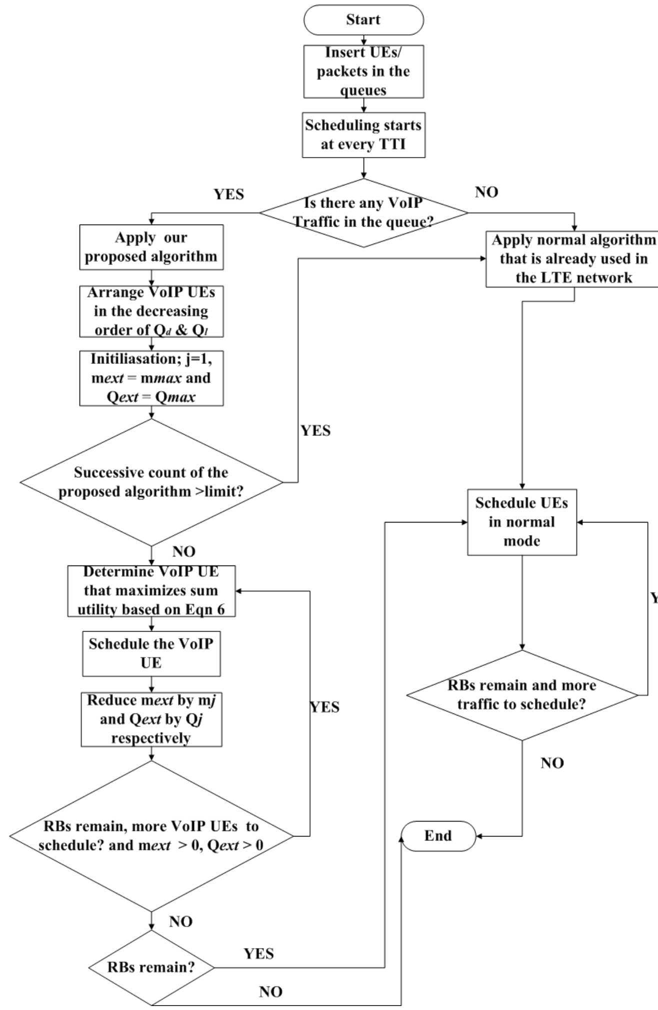

Figure 2 – Algorithm Flow Chart

B. Summary of Proportional Fair Algorithm (PF)

This scheduler was developed in [13][20]. Its main aim is to maximize the total network utility so that it can improve the network throughput and to guarantee fairness among flows. It assigns radio resources taking into account both the experienced channel quality and the past user throughput [21]. This scheduler uses a metric, which is defined as the ratio between the instantaneous available data rate and the average past rate with reference to the j-th flow in the i-th flow subchannel. This can be depicted in equation (15) below obtained from [13].

$$m_{i,j} = \left(\frac{r_{i,j}}{R_{i,j}}\right) \tag{15}$$

where $m_{i,j}$is the transmission rate, $R_{i,j}$ is the estimated average data rate, and $r_{i,j}$ is the instantaneous available data rate. $r_{i,j}$ is computed by the AMC module in LTE-Sim while considering the channel quality indicator (CQI) feedback that the UE hosting the j-th flow has sent for the i-th subchannel. It should also be noted that i and j are sub channel flows.

C. Summary of Exponential/Proportional Fair Algorithm (EXP-PF)

This scheduler was also developed in [13][20][22]. Its main aim is to increase the priority of real-time flows with respect to non-real-time flows, where their head-of-line packet (first packet in the queue) delay is very close to the delay threshold [21]. Its metrics were computed in [13] using the following equations.

$$m_{i,j} = exp\left(\frac{\alpha_i D_{HOL,i} - X}{1+\sqrt{X}}\right) \tag{16}$$

The variable X in equation (16) can be obtained from equation (17) below

$$X = \frac{1}{N_{r,t}} \sum_{i=1}^{N_{r,t}} \alpha_i D_{HOL,i} \tag{17}$$

where $N_{r,t}$ is the number of active downlink real-time flows. α_i in equation (17) can be described as the maximum probability that delay $D_{HOL,i}$ of the head-of-line packet exceeds the delay threshold. If we consider the packet threshold to be T_i, then α_i in equation (17) can be calculated as follows:

$$\alpha_i = -\frac{log_2 \alpha_i}{T_i} \tag{18}$$

Equations (17) and (18) proposed in [13], calculates the average total of the entire down link real time flows based on the probability that the first packet to be transmitted in the queue exceeds the delay threshold. This helps to prioritize down link real time flows.

VI. Simulation Details

A. PRB Characteristics

Before we go into the details of our simulation setup, lets first introduce the characteristics of PRBs, described as transmission resources in LTE. LTE systems can be analyzed both in time and frequency planes. The time plane is divided into 1 ms TTI, which consists of two slots of 0.5 ms to form 1 ms sub frames, where each sub frame contains 7 OFDMA symbols. In each TTI there are 14 OFDMA symbols, where 2 symbols out of 14 are reserved for uplink pilot transmission, while the other 12 symbols are used for data and control information transmission. TTIs can be defined as the minimum allocation unit in the time domain [23]. If we consider the frequency plane, the minimum allocation unit is the PRB, where each PRB contains 12 subcarrier of 15 KHz bandwidth each. The number of OFDMA symbols in a resource block depends on a cyclic prefix being used. All these can be depicted in Figure 3. It must be noted that VoIP packets must be transmitted per TTI and they can occupy one or more PRBs [6]. The amount of data bits that can be transmitted by one PRB depends on the link between the eNodeB and the user mobile terminal. This is due to the fact that 3G LTE uses adaptive modulation and coding (AMC), in order to change the modulation and coding schemes depending on the wireless link conditions.

B. LTE-Sim

In order to evaluate our proposed algorithm we used LTE-Sim simulation software. LTE-Sim is an open source

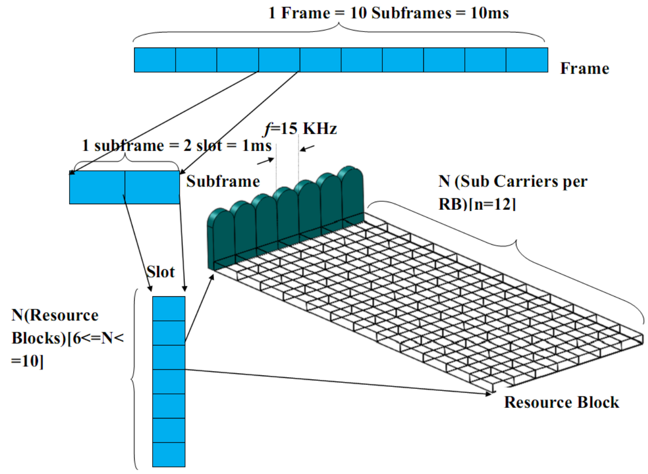

Figure 3 – Resource Grid

software, which is used to simulate LTE networks. It was developed in [13] and it is freely available under the GPLv3 licence. LTE-Sim is written in C++ using the object-oriented concept as an event driven simulator. This simulation software has all the important aspects of LTE networks notably the Evolved Universal Terrestrial Radio Access (E-UTRAN) and the Evolved Packet System (EPS). It maintains both single cell/multiuser and multiple cell/multiuser network topologies. This simulation software also supports different features, i.e., QoS management, user mobility, handover procedures, frequency reuse techniques, etc. Four different traffic generators are implemented and the management of data radio bearer is supported in this simulation software. The network nodes developed in this software are: User Equipment (UE), evolved Node B (eNB) and Mobility Management Entity/Gateway (MME/GW). Other features developed in LTE-Sim include: AMC scheme, channel quality indicator feedback, and some well known scheduling algorithms such as, PF and EXP-PF.

C. Scenario Setup

Our network topology is made up of a set of cells and different network nodes, which include: the eNodeB, mobility management/gateway (MME/GW), and user equipments (UEs). All the simulations were run in a three tier diamond-pattern macro scenario with 19-3-sector sites, which had a total of about 57 cells. The channel model used is the propagation loss channel model with channel realization. Most of the simulation parameters are presented in the Table 1. VoIP flows are generated by the traffic generator in LTE-SIM called VoIP application, which generates G.729 voice flows. The voice flow has been modelled with an ON/OFF Markov chain. The ON period is exponentially distributed with a mean value of 3s and the OFF period has a truncated exponential probability density function with an upper limit of 6.9s as well as an average value of 3s [24]. During the ON period, the source sends 20 bytes sized packets every 20 ms, which implies that the source data rate is 8 kb/s, on the other hand during the OFF period the rate is zero because the presence of voice activity detector is assumed. Three different scheduling algorithms were used in all simulation scenarios, these were: our proposed algorithm as well as EXP-PF and PF developed in [13]. Simulations were run for a number of iterations and in every iteration the seed number was updated. This was done in order to analyze the accuracy and the confidence interval of our simulation results.

TABLE I – Simulation Parameters

Simulation Parameters	Values
Bandwidth	5MHz
PRB structure	12subcarrier and 2subframes
TTI	1msec
Number of available PRBs	25
Number of sectors	3
Simulation time	1000 TTIs
Cyclic prefix	Normal
Number of Iterations	5
Scheduling Algorithms	Our Proposed Algorithm, EXP-PF, and PF
Cell radius	1 Km

VII. Results Analysis

A. Numerical Analysis

In this subsection, we present the numerical analysis to compare the analysis of our proposed scheduling algorithm with that of EXP-PF and PF proposed in the literature. The main reason for comparing our proposed scheduling algorithm with these two scheduling algorithms is that, they used the same PRBs allocation as ours and have similar simulation parameters except that they apply different metrics. They were also used as the benchmark scheduling algorithms in the LTE-Sim simulator. This made our comparison more feasible. Regarding the numerical analysis, we compared the level of VoIP user satisfaction (utility level) with the packet loss ratio, VoIP delay, and transmission rate. First let us analyze system utilization of all the three schedulers.

From the utilization point of view, as the transmission rate increases the bandwidth utilization also increases. This can be seen in Figure 4. This is mainly due to the fact that as the transmission rate increases, more users in the network are scheduled hence utilizing more bandwidth.

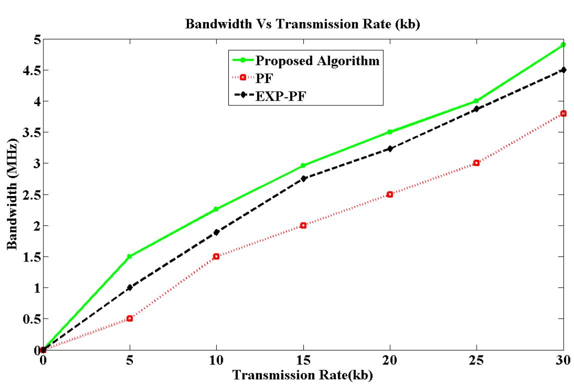

Figure 4 – Bandwidth (MHz) Vs Transmission Rate(kb)

From Figure 5, which shows the comparison between the utility levels and PLR, we can see that the VoIP user satisfaction levels were dropping as the PLR increased. PLR is the rate at, which VoIP packets are dropped during voice traffic transmission. So if more VoIP packets get dropped, the utility levels also start to fall. However, we note that there are some differences in the three schedulers. When PF and EXP-PF are used, the utility levels are lower than those of our proposed algorithm. This due to the fact that, with PF and EXP-PF schedulers, when there are high concurrent real-time flows, the probability of discarding packets for deadline expiration increases [11]. However, with our proposed algorithm, we do not calculate the deadline expiration factor for VoIP packets, it employs a simple method of scheduling users based on simple metrics, as well as the availability of the resource blocks.

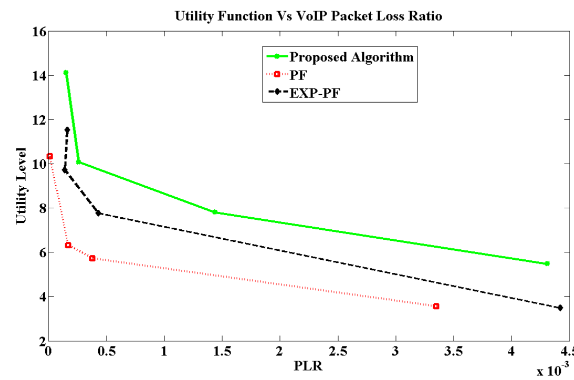

Figure 5 – Utility Function Vs PLR

Figure 6 compares the utility levels with VoIP delay. Again the utility levels decreases as the VoIP delay increases. When PF and EXP-PF are used, there is more reduction in the satisfaction levels than when our proposed algorithm is used. Our proposed algorithm employs a simple method of allocating resource blocks and scheduling VoIP user, which is less affected by high load factor as compared to the other two scheduling algorithm that employ packet deadline expiration procedure, which is highly affected by high load factor.

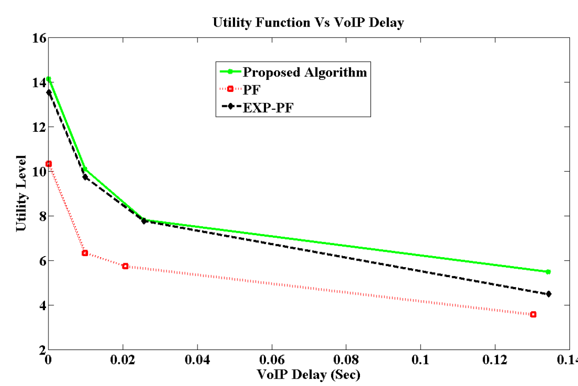

Figure 6 – Utility Function Vs VoIP Delay

Figure 7 compares the utility levels with transmission rates (m_j). The transmission rate was measured using the report provided by the user to eNodeB and was calculated using equations (7), (8), and (9) The better the transmission rate, the higher the levels of user satisfaction. As it can be seen in Figure 7, the utility levels are increasing as the m_j increases for all the three schedulers. Again, our proposed algorithm managed higher utility levels than EXP-PF and PF.

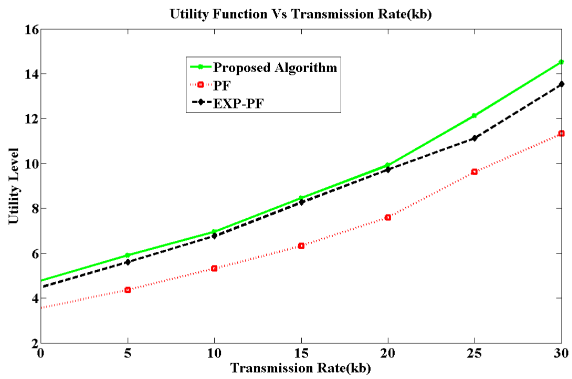

Figure 7 – Utility Function Vs Transmission Rate(kb)

B. *Performance Analysis Based on VoIP Users*

Regarding the performance analysis, we analyzed the same three VoIP metrics, which were throughput, VoIP delay, and packet loss ratio against the number of VoIP users. These metrics were compared with those of EXP-PF and PF scheduling algorithms. We analyzed user throughput for all three schedulers, while gradually increasing the number of VoIP users. As it can be seen in Figure 8, throughput decreased as the number of VoIP users increased in all algorithms. As the number of VoIP users increases, they overutilized the link and hence reduced the channel quality. This results in the VoIP packets being dropped as the number of VoIP users are increased, which led to less utilization of PRBs hence reducing the the total throughput achieved by VoIP users. As VoIP packets are small packets, many packets are needed to fully utilize the available PRBs. However, as congestion increased in the network, this led to VoIP packets being dropped, which led to less utilization of the available PRBs. But once again, our proposed algorithm obtained better throughput than the other two scheduling algorithms.

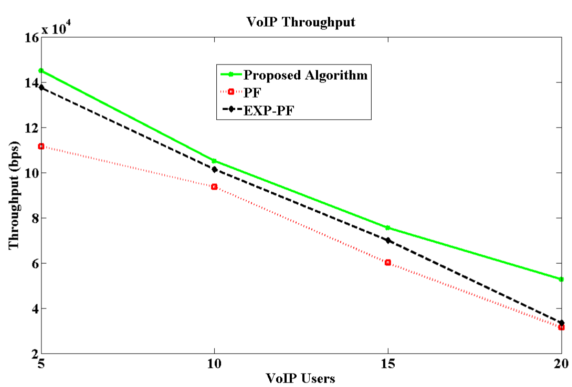

Figure 8 – Throughput Vs VoIP Users

We also analyzed VoIP delay while gradually increasing the number of VoIP users. This is shown in Figure 9. The VoIP delay is plotted on the Y axis in seconds as we increased the number of users steadily to twenty. As it can be seen, VoIP delay increased as the number of VoIP users increased. This is mainly due to the fact that as the number of users increase, they overutilize the link and cause congestion in the network. This will affect the transmission rate (m_j) and queue length (Q_l) metrics. This results in delay for VoIP packets. Even though there was an increase in VoIP delay for all the three schedulers, there are some differences in the three schedulers. When PF and EXP-PF are employed, the VoIP delay increases higher than of our proposed algorithm.

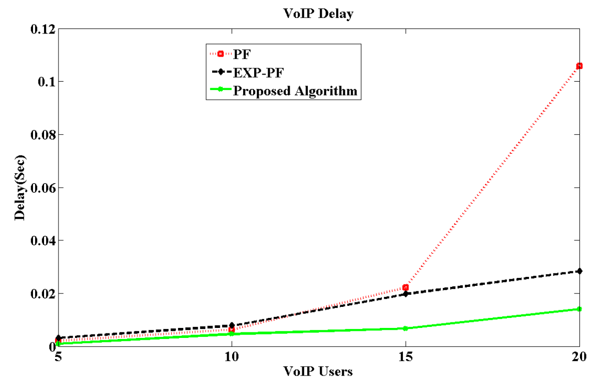

Figure 9 – Delay Vs VoIP Users

The packet drop ratio is analyzed and plotted on the Y axis as we increased the number of VoIP users steadily to the maximum of twenty users. Figure 10 shows the packet loss ratio for VoIP flows. As it can be seen, VoIP PLR increased as the number of VoIP users increased. This is mainly due to similar factors that affect VoIP delay. Even though there was an increase in VoIP PLR for all the three schedulers, we note some differences in the three schedulers. When PF and EXP-PF was used, the VoIP PLR increases more than that of our proposed algorithm. This is due to the fact that, with PF and EXP-PF schedulers, when there are high concurrent real-time flows, the probability of discarding packets for deadline expiration increases [8].

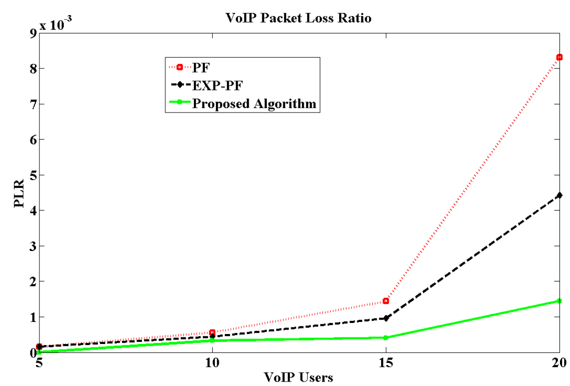

Figure 10 – PLR Vs VoIP Users

C. *Complexity Analysis*

Our proposed algorithm performs the scheduling operation after searching the user that can maximize utility function based on transmission rate (m_j) and queue length (Q_l) metrics. Therefore, the complexity to schedule the first user is

$O(KN)$, this will be the complexity for the first iteration. The complexity to schedule the second user is $O((K-1)N)$ and so on. In our algorithm, the number of iterations depends on the number of users K. As there are K iterations, the overall algorithm complexity can approximately be expressed as:

$$O(KN+(K-1)N+..+2N+N) = O(\frac{K(K+1)}{2}N) \approx O(K^2N) \quad (19)$$

Where $O(K^2N)$ implies that there is a second order complexity in the number of users based on m_j, Q_l metrics and there is also linear complexity in resource blocks N. This is due to the fact that there is no search done on the resource blocks, any available resource block is assigned to the user with highest metric. So for real complexity implementation of this algorithm, you only need to apply the possible values of N and K in the equation (19) to determine where it is efficient.

If we compare our algorithm to algorithm 1 in [11] that has a linear complexity in relation to the number of user and quadratic complexity in relation to the number of resource blocks, i.e., $O(N^2K)$, it is clear that our algorithm will only outperform it when the number of users are low since it will perform less iterations however when the number of users increases, algorithm 1 in [11] performs better.

D. Fairness Analysis

The fairness aspect is introduced mainly to solve the resource starvation problem, where users close to the base station are allocated more resources and edge users generally suffer from resource starvation [11]. Fairness can be described as a loose concept, which implies that all users are allocated equal amount of resources in order to meet QoS requirements. From the fairness point of view, we computed *jain* fairness index, which can be found in [25]. We compared our algorithm with PF and EXP-PF developed in [13]. Their fairness and complexity context constitutes an extension to algorithms described in [11][12] and they are also the bench mark schedulers in the simulator that we used. We analyzed the fairness index of all the scheduling schemes. As seen in Figure 11, fairness index decreased as the number of users increases. The fairness index of our proposed algorithm is higher than that of PF but lower than EXP-PF. It should be noted that the main advantage of our proposed algorithm is to improve the QoS of voice traffic when transmitted over an LTE network. At the same time, it reduces the negative impact, which may be caused by the introduction of the new algorithm on the entire systems performance. However, when we consider fairness analysis, EXP-PF out performs our proposed algorithm due to the following reason: EXP-PF employs the fairness concept in [11], which uses the algorithmic utility function that is associated with proportional fairness of the utility based optimization. This helps it in achieving a better fairness factor.

VIII. Conclusion and Future Work

In this paper, we analyzed the problem of scheduling and resource allocation for VoIP in 3G LTE. We also investigated the performance, complexity, and fairness of our proposed cross-layer scheduling algorithm for VoIP in 3G LTE. We projected the voice packet scheduling and resource allocation

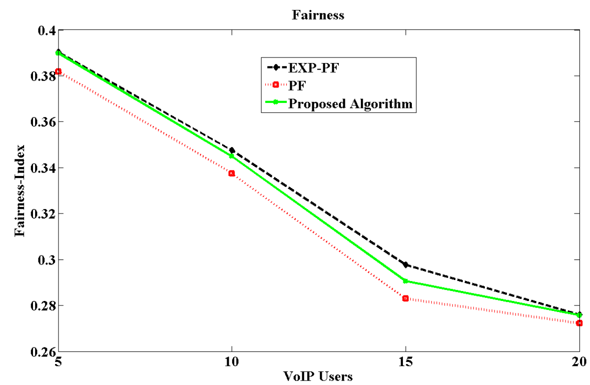

Figure 11 – Fairness Vs VoIP Users

problem as a constrained optimization problem. We solved this problem using a dual optimization approach with the goal of maximizing the expected total utility function under different constraints. Finally, we provided the algorithmic implementation of the obtained solution and also studied the performance of the proposed algorithm under different conditions and compared it with other algorithms in the literature, i.e., PF and EXP-PF.

Unlike other algorithms, which are time consuming and very complex, our proposed algorithm uses a metric maximization procedure to assign resource blocks to VoIP users. The main metrics used being queue length and transmission rate, this procedure makes our proposed algorithm less complex and it is executed in a short time.

Regarding complexity, our proposed algorithm performs better when the number of users is small since it schedules users after searching the user with highest utility metrics based on (m_j), (Q_l) and the search goes on for all available users. So for a small number of users, the search iterations done are less and hence the better performance. Our proposed algorithm performed better than PF but slightly poorer than EXP-PF.

Regarding numerical analysis, we compared the level of VoIP user satisfaction (utility level) with the packet loss ratio, VoIP delay, and transmission rate. These metrics were compared for all three scheduling algorithms, i.e., our proposed algorithm, EXP-PF, and PF. Our proposed scheduling algorithm performed better than the other two scheduling algorithms.

Regarding the performance analysis, we analyzed the same three VoIP metrics, which are throughput, VoIP delay, and packet loss ratio against the number of VoIP users. These metrics were compared with those of EXP-PF and PF scheduling algorithms. Again, our proposed algorithm out performed the other two algorithm.

Regarding fairness, the fairness index of our proposed algorithm is higher than that of PF but lower than EXP-PF. This is mainly due to the fact that EXP-PF uses the algorithmic utility function that is associated with proportional fairness of the utility based optimization. This helps in achieving a better fairness factor.

In future work, we will try to employ different tests such as, real life scenarios (existing networks) in order to analyze the difficulties and additional cost that would be required. It would also help us to analyze the practicability of our results and to make them more reliable. Our proposed algorithm can also be extended to other real-time applications, i.e., video, as well as extending it to latest LTE Advanced technology standards.

References

[1] R. Musabe and H. Larijani, “Complexity and Fairness Analysis of a new Scheduling Scheme for VoIP in 3G LTE,” Eighth Advanced International Conference on Telecommunications (AICT), pp. 34-39, May 2012.

[2] S. Y. Kim, “An Efficient Scheduling Scheme to Enhance the Capacity of VoIP Services in Evolved UTRA Uplink,” EURASIP journal of Wireless Communications and Networking, pp. 1-9, March 2008, doi:10.1155/2008/732418.

[3] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, “Mac Scheduling Scheme for VoIP Traffic Services in 3G LTE,” IEEE 66th Vehicular Technology Conference, pp. 1441-1445, October 2007, doi: 10.1109/VETECF.2007.307, .

[4] Rohde and Schwarz, “UMTS long term evolution (LTE) Technology introduction,” A report by Rohde and Schwarz, pp. 1-30, March 2007.

[5] M. C. Chuah and Q. Zhang, “Introduction to Wireless Communications,” US: Springer, 2006.

[6] S. Saha and R. Quazi, “Priority-Coupling- A Semi-Persistent MAC Scheduling Scheme for VoIP Traffic on 3G LTE,” ConTEL, 10th International Conference on Telecommunications, pp. 325-329, August 2009.

[7] 3GPP, “Physical layer aspects for Evolved UTRA,” 3GPP Technical report 25.814, version 7.1.0, pp. 1-135, September 2006.

[8] J. Huang, V. Subramanian, R.Agrawal, and R. Berry, “Downlink scheduling and resource allocation for OFDM systems,” IEEE Transactions on Wireless Communications, Vol.8, pp. 288-296, January 2009.

[9] S. Ali and M. Zeeshan, “A Utility Based Resource Allocation Scheme with Delay Scheduler for LTE Service-Class Support,” IEEE Wireless Communications and Networking Conference: MAC and Cross-Layer Design, pp. 1460-1465, 2012.

[10] R. Musabe, H. Larijani, B. Stewart, and T. Boutaleb, “A New Scheduling Scheme for Voice Awareness in 3G LTE,” IEEE Computer Society, Sixth International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 1-8, December 2011.

[11] E.Yaacoub, H. Al-Asadi, and Z. Dawy, “Low Complexity Scheduling Algorithms for LTE Uplink,” IEEE Symposium on Computers and Communications (ISCC), pp. 266-270, July 2009, doi: 10.1109/ISCC.2009.5202296.

[12] Y. Zhao, L. K. Zeng, G.Xie, Y.A. Liu, and F. Xiong, “ Fairness based resource allocation for uplink OFDMA systems,” Journal of China universities of post and telecommunications, pp. 50-55, Vol. 15, No. 2, June 2008.

[13] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, “Simulating LTE Cellular Systems: An Open-Source Framework,” IEEE Transactions on Vehicular Technology, pp. 1-16, Vol. 60, No. 2, February 2011.

[14] IETF, “Attribute list extension for the service location protocol,” pp. 1-7, IETF RFC 3059, February 2001.

[15] T. Henttonen, K. Aschan, J. Puttonen, and et al, “Performance of VoIP with Mobility in UTRA Long Term Evolution,” Vehicular Technology Conference, pp. 2492-2496, May 2008, doi: 10.1109/VETECS.2008.549.

[16] ITU-T, “One way transmission time,” ITU-T recommendation G.114, pp. 1-20, February 2003.

[17] TSG-RAN1, “LTE physical layer framework for performance verification,” Technical Report, 3GPP TSG-RAN1 R1-070674, March 2007.

[18] X. Qiu and K. Chawla, “performance of adaptive modulation in cellular systems,” IEEE Transactions on Communications, pp. 884-895, Vol.47, June 1999, doi: 10.1109/26.771345.

[19] H.Holma and A.Toskala, “WCDMA for UMTS: HSPA Evolution and LTE,” John Wiley, fourth edition, 2007.

[20] H. Ramli, R. Basukala, K. Sandrasegaran, and R. Patachaianand, “Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system,” 9th IEEE Malaysia International Conference on Communications (MICC), pp. 815-820, December 2009.

[21] G. J. Choi and S. Bahk, “Cell-throughput Analysis of the proportional fair scheduler in the single-cell environment,” IEEE Transaction Vehicular Technology, pp. 766-778, Vol.56, April 2007.

[22] R. Basukala, H. Mohd Ramli, and K. Sandrasegaran, “Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system,” First Asian Himalayas International Conference on Internet, AH-ICI, pp. 1-5, Nov 2009.

[23] Roke, “LTE eNodeB MAC Schedular,” Technical Report, Roke Manor Research Ltd, Siemens, January 2009.

[24] C. Chuah and R. H. Katz, “Characterizing packet audio streams from internet multimedia applications,” Proc. IEEE, ICC, pp. 1199-1203, August 2002.

[25] R. Jain, “The Art of Computer Systems Performance Analysis,” New York, Wiley, 1991.