A Novel Packet Scheduling with Channel-Aware Algorithm for Multi-Service Flow in the LTE Network

Ida Nurcahyani[#], Sisdarmanto Adinandra[#]

[#]Electrical Engineering Department, Universitas Islam Indonesia, Yogyakarta, 55584, Indonesia
E-mail: ida.nurcahyani@uii.ac.id, s.adinandra@uii.ac.id

Abstract—**Scheduling algorithm is one of many ways to increase system capacity and Quality of Service (QoS) in a telecommunication network. Designing a scheduling algorithm for a wireless network is challenging because its channel condition changes randomly according to the user's position. This paper investigated the performance of our proposed scheduling algorithm when managing multi-service flows in the downlink LTE network. The algorithm is the improvement of Proportional Fairness (PF) algorithm. The Extended-PF algorithm is a channel-aware scheduling algorithm that exploits users' channel condition information for calculating scheduling matrix in every scheduling process. Our simulation results showed that our proposed algorithm was able to provide better system's spectral efficiency.**

Keywords— **scheduling algorithm; QoS; Extended-PF; channel-aware**

I. INTRODUCTION

Mobile telecommunication technology nowadys has momentum in improvement and outgrows along with the proliferation of services demanded by customers. Mobile data becomes the main reason for the insignificant increase in the development of mobile communication and bandwidth consumption [1]. After years of voice and short message services' domination in mobile traffic, today's multimedia services turn out to be the most demanded service by mobile communication users [2].

To answer this service demand, 4G Long Term Evolution (LTE) technology from 3rd Generation Partnership Project (3GPP) has emerged with better capability in providing system QoS and spectrum efficiency than its predecessor. Using Orthogonal Frequency Division Multiplexing (OFDM) as a core technology, 3GPP LTE is now an international standard which is designed to increase system capacity and quality of broadband mobile services. The idea to divide data streams with higher data rate into smaller data streams with a lower data rate and then transmit it in parallel ways is proven in providing higher transmit rate. The multi-carrier modulation scheme is also proven in reducing the symbol duration so that inter-symbol interference (ISI) can be avoided.

Mobile broadband users will be nearly 50% of the global device by 2020 [3]. Increasing users will result in higher capacity demand. Since bandwidth is a limited resource in the telecommunication networks, bandwidth scarcity will be the main problem shortly. Several methods to increase system capacity, such as using smaller cells, increasing available bandwidth, and improving telecommunication technology have been used so that the available capacity closes to the theoretical channel capacity [4], have been tested by network operators. Hence, the need to satisfy their customer by providing anywhere and anytime communication with high transfer rate becomes a priority of the network providers.

The solution to increase system capacity and data transfer rate by using smaller cell will involve the utilization of a home eNodeB as additional hardware, which unfortunately results in extra costs. Moreover, home eNodeB that is part of closed subscriber group (CSG) can only control single cell so that the X2 interface used in communicating with other cells is not available yet [4]. Another solution with adding the available bandwidth will create another challenge as the bandwidth is limited and its use is strictly controlled. The solution of using a scheduling algorithm to increase system capacity has attracted many researchers since it is cheaper and more straightforward. Scheduling algorithm in LTE is left entirely to the network provider so that network planning can be implemented flexibly. The objectives of a scheduling algorithm are to provide better QoS and spectrum efficiency while maintaining fairness among users in the system. Designing such an algorithm is very challenging, and it usually takes many parameters, such as channel condition, service type, QoS target, and others, while maintaining low computational complexity. In LTE, time domain and frequency domain can be exploited in the scheduling process flexible.

A study revealed to maximize resource allocation by implementing the combination of Genetic Algorithm (GA) and Simulated Annealing (SA) when scheduling users in the frequency domain [5]. Modified-Earliest Due Date (M-EDD) is used to achieve higher bit rate for real-time video service [6]. Users with better channel condition and earliest due date become the priority of this scheduling algorithm. Traffic-based queue-aware is proposed [7] using discrete-time Markov-modulated Poison algorithm. Another solution to improve real-time service flows by using resource block preserver is proposed [8]. The algorithm employs two layer scheduling mechanism by combining multiple sub-frames for real-time services. These algorithms show excellent performance in system QoS and fairness. However, the algorithm's behavior in managing both real-time and non-real-time services was not investigated further. In this paper, we try to analyze our proposed algorithm in managing real-time and non-real-time services. The algorithm is an improvement of the fundamental Proportional Fairness (PF) algorithm.

The rest of the paper is organized as follows. In the next section, the LTE QoS and Scheduling process, our proposed algorithm, and the FLS and Exp-Rule as comparison algorithms are discussed briefly. Section III is the simulation process, result, and analysis. Finally, Section IV is the conclusion.

II. Materials and Methods

3GPP improves LTE network structure by making new standards in radio access network (RAN) and Core Network. RAN is an air interface that connects users and eNodeB. Essentially, LTE is an improvement from previous technology in the radio access network side. The air interface of a universal mobile telecommunication system (UMTS) consists of two nodes, which are NodeB and radio network controller (RNC). In LTE, the air interface is simplified so that it consists of only one logical node (evolved-NodeB/eNodeB) that also has RNC's abilities. This means that eNodeB is not only responsible for network air interface, but also managing resource management, interference, and users' mobility.

Furthermore, it ensures that every service provided by the network meets its QoS requirement [9]. This evolved air interface conveys challenges in controlling Radio Resource Management (RRM). The RRM primary responsibility is to guarantee that radio resources can be utilized efficiently. Additionally, it also controls the link adaptation technique and accommodates the users' QoS requirements according to their class of service [10]. The implemented algorithm in RRM is designed to maximize radio interface cross-layer functions.

LTE layers were designed to operate at all packet-system. This means that LTE has no dependency to the connection-oriented circuit switch system anymore. Moreover, the IP based Evolve Packet Core (EPC) in LTE can support the connectivity between 3GPP and non-3GPP standards. Thus, seamless roaming can be supported among user equipment to any radio standards in LTE.

Within LTE, various services can run at the same time. This is due to the ability of LTE to support multi-QoS with different bearers [11]. A user can make voice calls, browse a web page, and download a file at the same time. A real-time service as voice call constrains a tight delay and jitter requirement while packet loss tolerant. In contrast, non-real time services such as web browsing and file transfer do not necessarily require a tight delay and jitter guaranty but packet loss intolerant. By using multi-bearer, services with certain bit rate and QoS requirement will acquire some portion of the system resources that permanently provided to transmit. On the other hand, there is no dedicated channel for non-real time service. It means that non-real time service will be transmitted if there is a free portion of system resource.

QoS control concept in LTE is the core element of the evolved packet system (EPS). Every service is assigned to its QoS class. A bearer is designated only to one service (QoS) class. It separates services with different parameters. In UE, packet filters for uplink transmission; and gateway for downlink transmission read every packet attribute that passes through them and chooses its dedicated bearer. Packet flows in the same bearer receive similar treatment. QoS bearer parameters are QoS class identifier (QCI), allocation retention priority (ARP), maximum bit rates (MBR/AMBR), and guaranteed bit rate (GBR-optional) [12]. Using this bearer, a dedicated resource can be reserved for a particular condition. Fig. 1 indicates the EPS bearer process in LTE network.
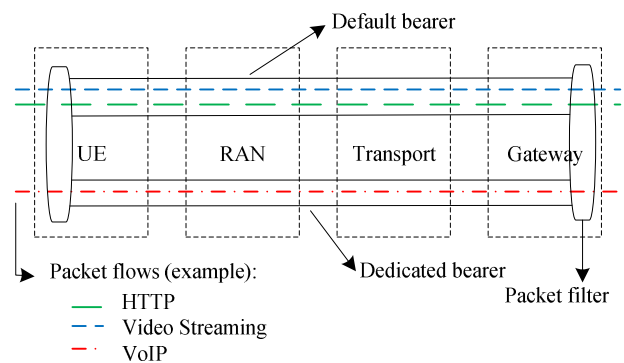

Fig. 1 EPS bearer

Each bearer in eNodeB reads the Class Identifier (QCI) and Allocation Retention Priority (ARP) that each service has passed through it. The QCI value is an agreement between network vendors so that service characteristics among vendors and operators can be standardized. This service identity is read, then each service is handled according to its priority, packet delay budget, and acceptance packet loss values [13]. The scheduler in eNodeB to select which user should be transmitted first uses the values of these packet delay budget and priority.

There are several phases in LTE downlink scheduling and transmission scheme. Fig. 2 shows how the scheduling process in LTE is done. The first phase involves the process of sending scheduling commands by eNodeB to the User Equipment (UE). This scheduling command is sent through the Physical Downlink Control Channel (PDCCH). The scheduling command information includes data amount, resource block allocation, and modulation scheme [4]. Once the UE receives this scheduling command information, eNodeB will send data via Downlink Shared Channel (DL-SCH) and Physical Downlink Shared Channel (PDSCH).

The data transmitted occupies a transport block with the duration of 1 Transmission Time Interval (TTI) in 1 msec. The UE will send Hybrid-Acknowledgement (Hybrid-ACK) if the data is received. This Hybrid-ACK is delivered via Physical Uplink Shared Channel (PUSCH). The eNodeB will switch to a new transport block to send further data after receiving this Hybrid-ACK.
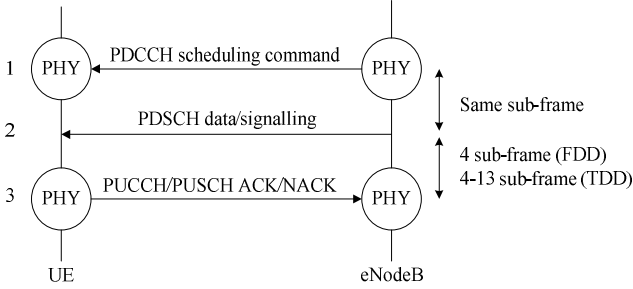

Fig. 2 LTE downlink scheduling process [4]

A. Frame Level Scheduling Algorithm

For every scheduling process, Frame Level Scheduling (FLS) algorithm calculates the amount of real-time data that flows into the queue, then estimates the transmission time required by each data [14]. The algorithm utilizes channel sharing mechanism on the network by allocating resource block according to the constraint delay needed by each queue. If there are N traffic flows in the system, the channel available then will be divided equally among these flows.

For every flow, the packets that are waiting to be sent are stored in a queue. The FLS algorithm calculates transmission time needed by every queue in each frame. At the beginning of the k-frame of the n-flow, we have a starting time $t_{k,i}$, which is considered as a constant sampling instant. The frame duration is calculated as $\Delta t(k) = t_{k,i+1} - t_{k,i}$. This value can also be considered as a sampling time. For calculating the queue length, we have the following equation:

$$q_i(k+1) - q_i(k) = d_i(k) - u_i(k) \qquad (1)$$

where $q_i(k)$ is the i-th queue length at time t_k, and $q_i(k+1)$ is i-th queue length at time t_{k+1}, the value of $u_i(k)$ describes the data size that is sent during the k-th frame, and $d_i(k)$ is the sum of data in the queue of k-th frame, $d_i(k)$ also represents the behavior of how the source is filling the queue for each k-th frame.

To maintain bounded delay for each packet, the control law should follow the equation $u_i(k) = h_i(k) * q_i(k)$, the operation $*$ is a discrete time convolution process. This means, to transmit $u_i(k)$ the amount of data at k-the frame, the signal in $q_i(k)$ has to be filtered through a time-invariant linear filter with pulse response $h_i(k)$. In LTE, the eNodeB always notes how many queues each traffic flow has in every transmission time. The equation can be simplified by:

$$u_i(k) = q_i(k) + \sum_{n=2}^{M_i} Q c_i(k) \qquad (2)$$

where

$$Q = [q_i(k-n+1) - q_i(k-n+2) \\ -u_i(k-n+1)] \qquad (3)$$

$c_i(k)$ is a real finite coefficient, and M_i is the length of pulse response.

B. Exponential Rule Algorithm

The queuing model of the Exponential Rule (Exp-Rule) algorithm was designed to be able to maximize the channel allocation of users with varying channel condition [15]. In Exp-Rule, N a single server serves input flows. Every flow has its queue. The channel is operated in discrete time. We have t a time slot and M channel state, which every time slot is associated with a different channel state. This channel state is fixed for every time slot. Every channel state has a data rate vector μ_i^m. For a given time slot t, the channel is in m state and all services in this time slot are allocated to the i queue.

The system's random process behavior is described in:

$$S = (S(t), t = (0,1,2,\cdots)) \qquad (4)$$

where

$$S(t) = \left\{ \left(W_{i1}(t), \cdots, W_{iQ_i(t)}(t), i = 1, \cdots, N; m(t) \right) \right\} \qquad (5)$$

$Q_i(t)$ Is the i-type queue length at a time t, and $W_{i1}(t)$ is the current delay of the k-th type i customer present in the system at the time t, with $W_i(t) = 0$ if $Q_i(t) = 0$. The Exp-Rule algorithm is throughput optimal in every scheduling process if a queue is chosen based on the value of:

$$i\epsilon i(S(t)) = \arg\max_i \gamma_i \mu_i(t) \exp\left(\frac{a_i Q_i(t)}{\beta + [\overline{Q}(t)]^\eta} \right) \qquad (6)$$

where $\mu_i(t) \equiv \mu_i(t)$ and $\overline{Q}(t) = (1/N) \sum_i a_i Q_i(t)$, and waiting time to be served is equal to:

$$i\epsilon i(S(t)) = \arg\max_i \gamma_i \mu_i(t) \exp\left(\frac{a_i W_i(t)}{\beta + [\overline{W}(t)]^\eta} \right) \qquad (7)$$

Where $\overline{W}(t) = (1/N) \sum_i a_i W_i(t)$. With the value of $\beta, \eta \epsilon (0,1)$, $\gamma_i, a_i, t \epsilon N$. The value of γ_i in a lossless condition is 1.

C. Extended Proportional Fairness Algorithm

The Proportional Fairness (PF) was created to manage system channel with elastic traffic flows [16]. The elastic traffic is a service with the ability to adjust its rate according to the channel condition. The system channel that is shared among elastic traffic users within the network must meet its stability and fairness values. By controlling the stability condition between supply and demand, the near optimum throughput value can be achieved.

Stabilizing the desired rate in the network should meet the following logarithmic optimization equation:

$$\max \sum_{m \epsilon M} \phi_m \log s_m \qquad (8)$$

which means in order m-th user at the t time want to get flow rate s_m, the throughput that should be met is ϕ_m. This equation results in Pareto optimal condition, hence when scheduling a user, the vector rate change s_m^+ of m-th user should meet the following equation:

$$\sum_{i \in M} \frac{s_m^* - s_m}{s_m} \leq 0 \qquad (9)$$

Using this equation, system stability can be achieved if the following matrix equation selects every user:

$$m = \arg\max_m \frac{s_m}{S_m} \qquad (10)$$

Where S_m is the m-th user's average data rate in the previous time slot. This equation results in a good trade-off between system throughput and fairness, but when implementing it in real time services result in high packet delay violation [17] since the delay is not considered in this matrix calculation.

To optimize the long-term received rate, the scheduling matrix becomes [18]:

$$\arg\max_{p, w} \prod_{m \in M} \left(\alpha_m + \frac{(1 - \alpha_m) w_m \log\left(1 + \frac{p_m}{\delta_m w_m}\right)}{S_m} \right) \qquad (11)$$

where w_m is bandwidth allocation, p_m is the distributed power from eNodeB, δ_m is the noise function, and α_m is a time window parameter.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
eNodeB	1
Cell radius	1 km
System bandwidth	10 MHz
UE number	10, 20, 30, 40, 50, 60
UE speed	30 km/h
UE distribution and mobility	Random
Propagation model	Macrocell urban
Service	BE, Video, and VoIP
Comparative algorithm	FLS and Exp-Rule

To investigate our algorithm's performance in managing real time and non-real time services, we generated Best Effort (BE), Voice over IP (VoIP), and streaming video services for every user in the system at the same time. The simulation was performed in a single cell while users were moving around cell randomly. The radius of the cell was 1 km. eNodeB was placed in the center of the cell with 10 MHz bandwidth for uplink and downlink transmissions. The parameters of the simulation are presented in Table 1.
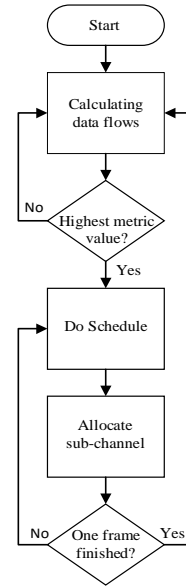

Fig. Three eNodeB scheduling process

The scheduling process of our algorithm is depicted in Fig. 3. The scheduling algorithm in LTE works based on channel sensitivity concept. In every TTI, eNodeB calculates data flows that have to be transmitted. Then, it reads the users' CQI feedback and MAC queue length for each flow. eNodeB then selects the highest metric value and allocates the sub-channel for it. This process is repeated until data flows equal to zero.

III. RESULTS AND DISCUSSION

Fig. 4 shows the cell spectral efficiency of the three algorithms. Cell spectral efficiency shows how the spectrum is utilized efficiently. The result denoted that our proposed algorithm was able to provide higher spectral efficiency compared to the FLS and Exp-Rule. This result proves that in the same system, our algorithm provided greater capacity compared to the other simulated algorithms. This was caused by the scheduling matrix calculation in the Extended-PF taking into account the channel condition information. Users who are moving around cell within a certain speed will experience changes in their channel condition according to their movement. The Extended-PF algorithm can adapt rapidly to this change so that in each scheduling process, it can maximize the channel allocation to the user to ensure a long-term received rate.
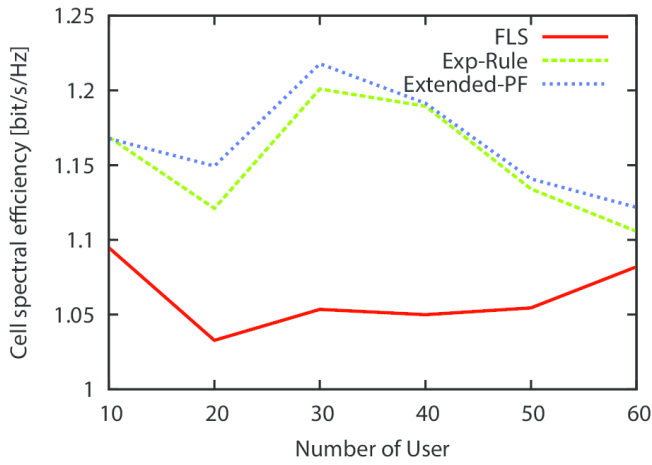
Fig. 4 Cell spectral efficiency

Fig. 6 Throughput result of VoIP service

Fig. 5 and Fig. 6 indicate the total throughput of real-time video and VoIP services. The VoIP throughput for three algorithms tended to show similar value. This was due to the VoIP characteristic, which was a golden class service. VoIP generally has a dedicated channel due to the constant transfer and smaller bit rate. In contrast, Fig. 7 and Fig. 8 points the delay values for video and VoIP services. The proposed algorithm provided lowest delay value for VoIP service. The delay results in the simulation were the end-to-end delay that had been affected by processing, queue, and transmission delays. The Extended-PF algorithm could provide a better delay guarantee for VoIP service. On the other hand, the FLS algorithm showed better performance in managing video service compared to the other algorithms. This was due to the ability of the upper-level algorithm calculated the estimated delay constraint for real-time services so that the delay constraint was not violated.
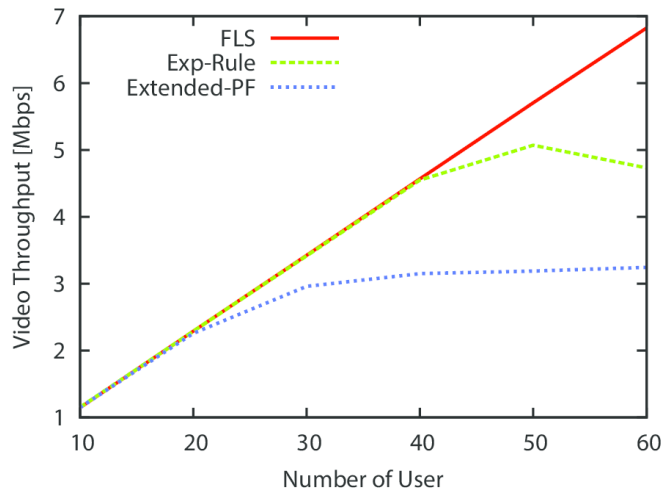

Fig. 7 Delay result of video service

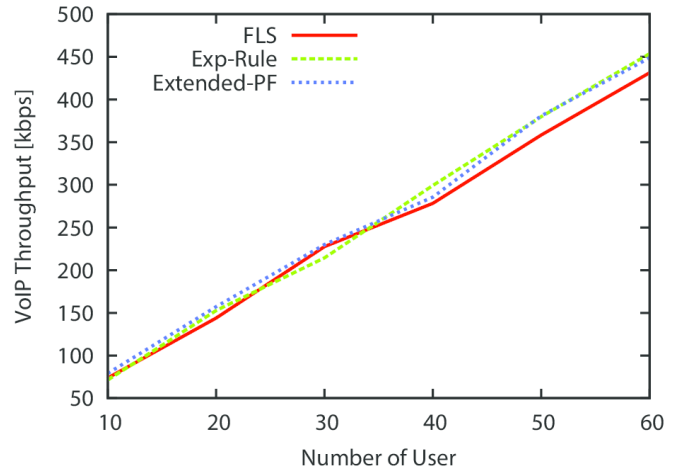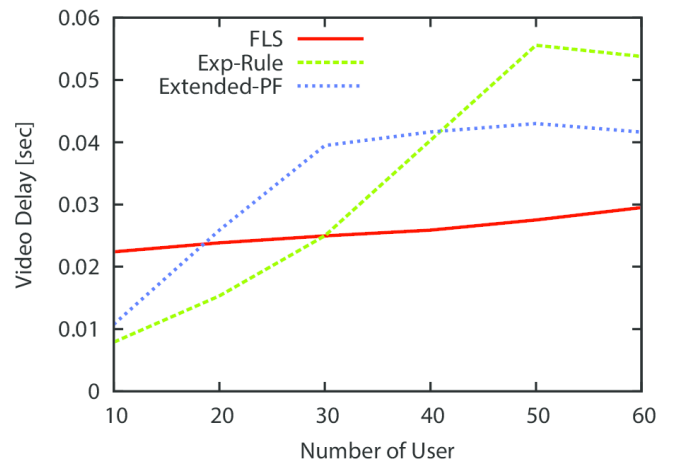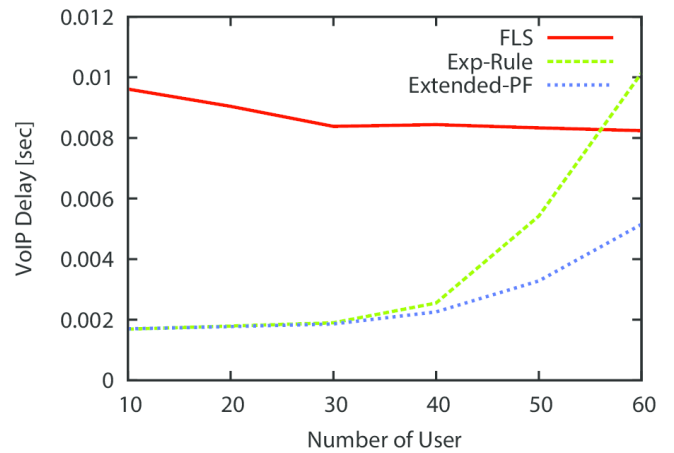Fig. 5 Throughput result of video service

Fig. 8 Delay result of VoIP service

Figs. 9 and 10 represent throughput and PLR values for BE service respectively. Non-real time service is sensitive to packet loss. Poor channel condition and congestion can cause packet loss. On the contrary, this service was not affected by delay and jitter. Furthermore, this service was burst so that throughput could be varied. Generally, BE service will get resource allocation after the real-time services. The graph shows the proposed algorithm performed better than the other two algorithms in managing non-real

753

time service. This was because of the fundamental characteristic of Proportional Fairness algorithm that provided resource allocation for every flow in the system. The BE service got the part of the spectrum at one scheduling process.
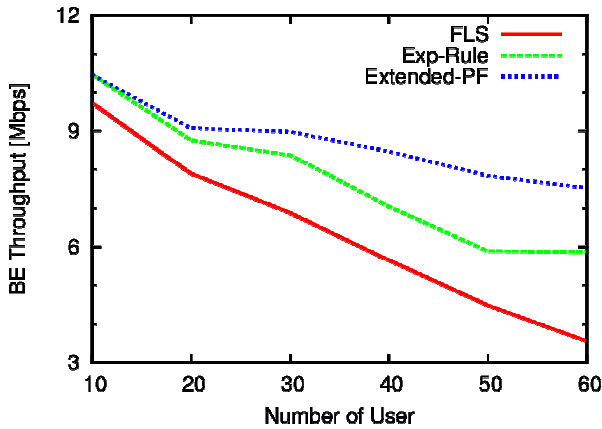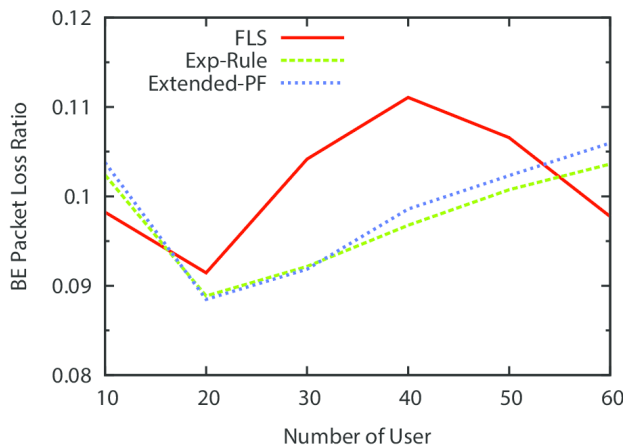

Fig. 9 Throughput value of BE service

Fig. 10 Packet Loss Ratio (PLR) value of BE service

Fairness is considered one of the critical network performance parameters. It indicates how system resource is shared among packet flows in the network. This fairness value also determines how a node in a network is given an opportunity to utilize the available resource blocks; therefore, its QoS requirements can be met. Fairness value ranges from 0 to 1, with 1 denotes the highest fairness and 0 is the lowest. Figs. 11, 12, and 13 demonstrate the fair values of all scheduling algorithms in each service.
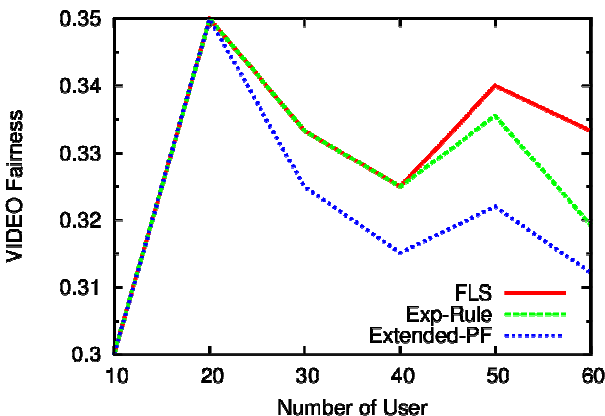

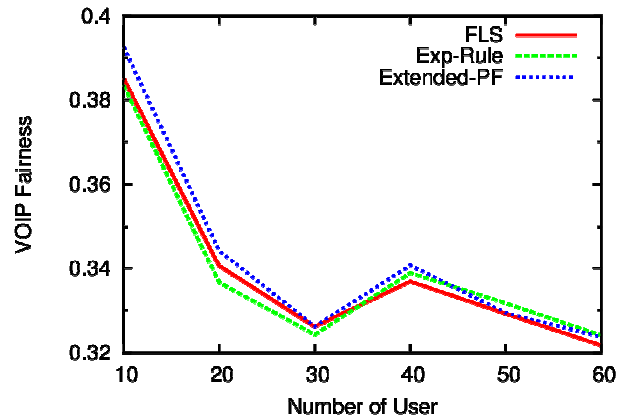Fig. 11 Fairness value of video service

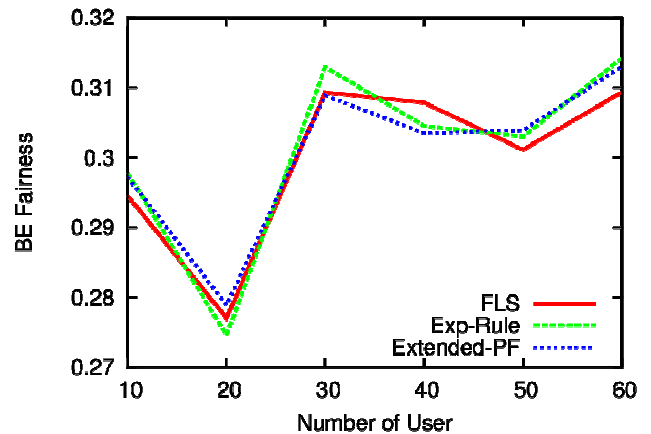Fig. 12 Fairness value of VoIP service

Fig. 13 Fairness value of BE service

The fair value of a node was shared between all services. Thus, the value of fairness for each service revolved around the value of 3. For video service, FLS indicated the best fairness value among others. In contrast, the proposed algorithm provided better fairness value in VoIP service. Since real-time services have higher priority than non-real time service, the allocation of resource blocks is prioritized for them. This resulted in a lower fair value of BE service compared to video and VoIP services. The BE fairness values for Exp-Rule and Extended-PF show similar values; their performance was better than the FLS algorithm.

IV. CONCLUSIONS

This paper addressed the performance of our proposed scheduling algorithm in managing triple play services. The FLS and Exp-Rule algorithms were used as parallel algorithms to investigate how well our algorithms in controlling system resource for every service. The simulation results showed that Extended-PF algorithm could provide better cell spectral efficiency. Furthermore, it also showed lower delay value for VoIP service and lower PLR value for BE service. This was due to the ability of the algorithm to adapt quickly to the changing of users' channel condition.

REFERENCES

[1] "ITU: Measuring the Information Society Report 2016," 2016.

[2] "Cisco Visual Networking Index: Forecast and Methodology 2015-2020," 2016.

[3] T. J. Barnett, A. Sumits, S. Jain, and U. Andra, "Cisco Visual Networking Index (VNI) Update Global Mobile Data Traffic Forecast," pp. 2015–2020, 2016.

[4] C. Cox, *An Introduction to LTE*. West Sussex: John Wiley & Sons, 2014.

[5] M. E. Aydin, R. Kwan, and J. Wu, "Multiuser Scheduling on the LTE Downlink with Meta-Heuristic Approaches," *Phys. Commun.*, vol. 9, pp. 257–265, 2013.

[6] M. I. Elhadad and M. Abd-elnaby, "Resource Allocation for Real-Time Services Using Earliest Due Date Mechanism in LTE Networks," in *Fourth International Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC)*, 2016, pp. 9–12

[7] W. Ke, L. I. Xi, J. I. Hong, and M. A. Ze-wen, "Traffic-Based Queue-Aware Scheduling for 3GPP LTE System,"*The Journal of China Universities of Post and Telecommunication*, vol. 21, no. April, pp. 63–68, 2014.

[8] S. Ghassan, A. Ali, M. D. Baba, and M. A. Mansor, "Resource Block Preserver Scheduling Algorithm for VoIP Applications in LTE Networks," pp. 146–150, 2015.

[9] M. Sauter, From GSM to LTE-Advanced: an Introduction to Mobile Networks and Mobile Broadband, 2nd Edition. West Sussex: Wiley, 2014.

[10] K. I. Pedersen, T. E. Kolding, F. Frederiksen, I. Z. Kovács, D. Laselva, and P. E. Mogensen, "An Overview of Downlink Radio Resource Management for UTRAN Long-Term Evolution," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 86–93, 2009.

[11] S. Palat and P. Godin, "Network Architecture," in *LTE-The UMTS Long Term Evolution: From Theory to Practice*, S. Sesia, I. Toufik, and M. Baker, Eds. West Sussex: John Wiley & Sons, 2011, pp. 25–55.

[12] H. Ekström, "QoS Control in The 3GPP Evolved Packet System," *IEEE Communications Magazine*, vol. 47, no. 2, pp. 76–83, 2009.

[13] M. Baker and T. Moulsley, "Downlink Physical Data and Control Channels," in *LTE-The UMTS Long Term Evolution: From Theory to Practice*, S. Sesia, I. Toufik, and M. Baker, Eds. West Sussex: John Wiley & Sons, 2011, pp. 189–214.

[14] G. Piro and L. Grieco, "A Two-Level Scheduling Algorithm for QoS Support in The Downlink of LTE Cellular Networks," *Wireless Conference (EW), 2010 Europe,* 2010.

[15] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," *Teletraffic Science Engineering*, vol. 4, pp. 793–804, 2001.

[16] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[17] I. Nurcahyani, I. W. Mustika, and Selo, "Performance Analysis of Packet Scheduling Algorithm for Video Service in Downlink LTE," in *2014 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS)*, 2014, November, pp. 52–57.

[18] I. W. Mustika and I. Nurcahyani, "Proportional Fairness with Adaptive Bandwidth Allocation for Video Service in Downlink LTE," in *2015 IEEE International Conference on Communication, Networks and Satellite (COMNESTAT)*, 2015, pp. 54–59.