Development of a Prioritized Scheduling Algorithm for Congestion Management in Long Term Evolution

Kuboye B. M, Adewale O. S., Alowolodu O. D. and Damilola E. O.

Abstract—These instructions give you guidelines for preparing pap The surfacing of a Long Term Evolution (LTE) in a mobile networks environment was welcomed by the public due to its promising high speed and capacity in both voice and internet services. The number of subscribers to this network continues to increase daily thereby giving rise to congestion on the scarce resources available for allocation to the subscribers. Scheduling is one of the methods that that can be employed to mitigate the congestion on the network; as a result, several scheduling algorithms on the management of Radio Blocks (RBs) in LTE Networks were studied. Thereafter, a new prioritized algorithm was designed and simulated using Simulink in MATLAB. The results were compared to Best CQI and Round Robin. The proposed algorithm showed promising statistics in comparison with the Best CQI and Round Robin algorithms in terms of throughput and resource block allocation fairness.

***Index Terms*—Prioritized Scheduling, Algorithm, Congestion Management, Long Term Evolution.**

I. INTRODUCTION

The numbers of internet users continue to increase daily, especially those accessing the Iinternet through Smartphones and tablets. More devices are coming up and more users utilize data services through various applications. According to LiveU White-Paper (2014) the increasing number of internet users gives rise to congestion. LTE networks provide tens of megabits of bandwidth and yet will not solve congestion problem since the content delivery applications are also becoming more and more bandwidth starving. Rysavy research (2015) stated that congestion occurs when demands for network resources are greater than network capacity. Analogous situations are highways with too many cars or water supplies with too many users. Sometimes, the topography sometimes impact the performance of cellular connectivity. Subscribers in the lowlands may sometimes not receive clear signals as subscribers in the hilly areas or near equal height to the base-station.

Manuscript received Dec 16, 2015 and accepted January 10 , 2016
B. M Kuboye is with the Federal University of Technology, Akure, Ondo State, Nigeria. Phone: 234-08033865251 kubonline@yahoo.co.uk; bmkuboye@futa.edu.ng
O. S.Adewale is with the Federal University of Technology, Akure, Ondo State, Nigeria e-mail: adewale@futa.edu.ng.
O. D, Alowolodu is with the Federal University of Technology, Akure, Ondo State, Nigeria e-mail: odalowolodu@futa.edu.ng.
E.O. Damilola graduated from the department of Computer Science at the the Federal University of Technology, Akure, Ondo State, Nigeria.

Congestion is the unavailability of network to the subscriber at the time of making a call or a demand for a service. It is the situation where no free path can be provided for an offered call (Syski, 1986). That is, it is a situation in which a subscriber cannot obtain a connection to the wanted service immediately (Kuboye, 2010). Generally, the network does not degrades during traffic congestion, it is the applications that are served by the network that do not respond as expected and therefore, the user-experience falls below expectation (Lucente, 2012). Degraded performance can range from slow response to requests for information to loss of data which can manifest itself as distorted video and unintelligible audio. In designing a management algorithm for congestion, the urgency and importance of requests should be greatly considered beside other important factors. Therefore, the objective of this work is to develop a prioritized scheduling model for congestion management in LTE.

II. LONG TERM EVOLUTION (LTE)

LTE is a specification for cellular 4G standards. It is based on the GSM/EDGE and UMTS/HSPA network technologies to increase the capacity and speed of data connections using a different radio interface together with core network improvements (LTE, 2015). LTE network motivations include increasing the capacity and speed of wireless data using newly developed Digital Signal Processing (DSP) techniques and modulations. LTE network architecture was also redesigned and simplified to an IP-based system with significant reduction of transfer latency compared to the 3G architecture. LTE was first proposed by NTT DoCoMo of Japan in 2004 and was officially started in 2005, and its service was launched publicly in Oslo and Stockholm on December 14, 2009 by Telia Sonera.

A. LTE Network Architecture

The high-level network architecture of LTE comprises the User Equipment, evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and evolved Packet Core (EPC). Please refer to [4] for the The major components in LTE.

The User Equipment (UE) is a device for accessing the LTE network by the subscribers. It can be a USB modem or a mobile that is LTE compliant. The UE allows the user to have access to services provided by the LTE network. The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) has the element known as the eNode-B. eNode-B is the first contact of a UE while trying to communicate to the LTE network. It manages the entire radio resources for a particular area and allocates them to mobiles. User Equipment connects to the eNode-B

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

through a LTE-Uu interface. The Evolved Packet Core (EPC) is the core of the network and is a flat all-IP-based core network that communicates through 3GPP radio access (UMTS, HSPA, HSPA+, LTE) and non-3GPP radio access such as WiMAX and WLAN. It manages handover procedures within and between both access types. The access flexibility to the EPC makes it attractive to the operators since it enables them use a single core to support different services (Akyildiz, 2010).

The architecture of Evolved Packet Core (EPC) is grouped into two main planes known as the user plane and the control plane. Mobility Management Entity (MME) forms the core of the control plane while serving gateway (S-GW) forms the core of the user plane. The S1 interface connects the eNode-B to the MME and S-GW. Please refer to [5] for the The Evolved Packet Core (EPC) (The core network).

The Home Subscriber Server (HSS) is the central database with the purpose of having information about all the subscribers on the network. The serving gateway (S-GW) acts as a router and forwards data between the base station and the Packet Data Network (PDN) gateway. The PDN Gateway (P-GW) provides connectivity between the UE and external PDNs. It performs policy enforcement, packet filtering for each user, charging support, lawful Interception, and packet screening. Each packet data network is identified by an Access Point Name (APN) that enables subscriber has access to the network. The MME serves as the termination point for ciphering and integrity protection for Non- Access Stratum (NAS) signaling. MME. also handled the security key management and provision of control plane function for mobility between LTE and other access networks.

B. LTE Downlink Frame Structure and Resource Blocks

LTE transmission is segmented into frames, each frame consists of 10 subframes and each subframe is further divided into two slots each 0.5ms, making the total time for one frame equivalent to 10ms. Each timeslot on the LTE downlink system consists of 7 OFDM symbols. The very flexible spectrum allows LTE system to use different bandwidths ranging from 1.4 MHz to 20 MHz where higher bandwidths are used for higher LTE data rates. The physical resources of the LTE downlink can be illustrated using a frequency-time resource grid (Please refer to [8] for the Illustration of a scheduling block in LTE downlink). A Resource Block (RB) has duration of 0.5msec (one slot) and a bandwidth of 180 kHz (12 subcarriers). It is straightforward to see that each RB has 84 resource elements in the case of normal cyclic prefix and 72 resource elements in the case of extended cyclic prefix (Habaebi et al, 2013; Mohammed et al, 2013).

C. Traffic Class Prioritization

Traffic class prioritization is a quality of service (QoS) mechanism that categorizes assigned data traffic priority values according to their class of service (CoS) and these values are used by the network scheduler to effectively manage resource usage on the network during congestion periods. It identifies the different classes of data traffic and manages the available resource to effectively handle the traffic demand.

The utilized traffic classes are Real Time Conversational, Real-time Streaming, Near Real-Time interactive and non Real-time background. Delay sensitivity is the main feature that differentiates these traffic classes. The Real-time class encompasses traffic that has high delay sensitivity while the non-Real-time traffic class has the lowest delay sensitivity (Dushyanth, 2006). Real time traffic belongs to either conversational or streaming class. Conversational class traffic include telephony speech, voice over IP, video conferencing while Streaming class traffic are streamed video or audio (Omotoye et al, 2014). Non-real time traffic belongs to either interactive or background class.

TABLE I
TRAFFIC CLASSES AND CHARACTERISTICS
(SOURCE: OMOTOYE ET AL, 2014)

Traffic Class	Characteristics	Priority	Examples
Real-time Conversational (T_1)	Delay sensitive, Low latency and Jitter	High	Voice and video calls
Real-time Streaming (T_2)	Limited tolerance to loss, low latency	Medium	Audio and Video streaming
Near Real-time Interactive (T_3)	Error-sensitive, best effort, low packet loss.	Normal	Web Browsing
Non Real-time Background (T_4)	Best effort, loss Tolerant	Low	e-mail and file transfer

They are both delay insensitive but require high throughput and less error rate. Interactive class traffic includes web browsing, database retrieval while examples background traffic are telemetry and e-mailing. The table I shows the different traffic classes, their characteristics, the priority and examples.

Prioritization of service can be carried out by using a different admission technique for each class of service. A strict admission policy can be implemented for class/service with lower priority. The priority level is used to allocate resources to the classes/services. A request is rejected if it is discovered that reserved resources for its service/class are not enough.

III. REVIEW OF RELATED WORKS

Meader and Schmid (2012) discussed user-plane congestion management, that is, managing congestion at the user side. Congestion in mobile networks, congestion scenarios, solution components in congestion management and LTE Quality of Service' features and limitations were discussed. The paper explains how the congestion can be detected and signaled and how applicable management policy can be applied. The paper then concludes that congestion management solutions rely on smart mitigation mechanisms which need new metrics to characterize end-user congestion, lightweight signaling of congestion occurrence and location, application and QoE aware traffic management and control loop to react timely on congestion. The implementation of the smart mitigation mechanisms proposed is not discussed in details. Ahmad et al (2012) proposed a model depending on the iterative server mechanism. In this model, when any client sends a request, the client listener will listen to the client request and then send notification to the main server about the requested client, afterwards, the main server instructs the client server to generate a new temporary server for the requested client. The new temporary server set up for this client is referred to as 'iterative server'. The client makes use of this iterative server and once the task has been completed the iterative server is automatically destroyed.

Bahreyni et al (2014) proposed a fairness aware downlink scheduling algorithm which can help manage and reduce congestion in LTE networks. In this algorithm, it is assumed that each e-nodeB receives channel feedback information in form of CQI (Channel Quality Indicator) feedback matrix. The matrix size equals to number of UEs by number of Resource Blocks in each Transmission Time Interval (TTI). This model gives preference to those users which uses less bandwidth than others. Also, it evenly distributes the resources among the users during each TTI.Kanagasundaram and Kadhar-Nawal (2013) proposed an algorithm for scheduling real time services in LTE networks. This algorithm considers both the resource block allocation and scheduling process. The resource block allocation considers the instantaneous data rate and the average data rate. This scheme allocates the resources that are required to perform a real time connection, but, if the resources are busy, then, the user connection is scheduled using a lower level scheduler. The scheduler has a timer based on which user connections are updated. In this scheduling period, the available resources are assigned to user. In the RB allocation, it will allocate the resources that are required to perform the real-time connection. A time constant is used to update the average rate that is allocated to the users and Q(t) factor is used to support different class of QoS. If the resources are busy then the approach will schedule the user connection using the lower level of the scheduler. The scheduler has a timer based on the user connections that are updated. In the scheduling period, the available resources are assigned to the user. The main advantage of this method is in the assignment of the reserved RBs to real-time users so that average spectrum efficiency and average cell throughput can be improved. This approach does not consider the importance of urgency attached to individual calls and this can lead to wastage of resources and time.

Yifeng (2009) proposed a congestion management solution specific to the eNode-B. He developed an Active Queue Management (AQM) scheme, for managing traffic in the eNode-B. He advised that AQMs be implemented at the eNode-B rather than the User Equipment (UE) side because, implementing AQMs at the UE side does not guarantee good performance. The AQMs' targets are to control the queue length in a way to control the queue length of all the user equipments, reduce end-to-end delays and reduce probabilities for buffer overflow or underflow**.** Amokrane (2011) proposed the use of Machine Type Communication (MTC) to control congestion over LTE networks. MTC is also referred to as Machine to Machine (M2M). M2M applications are those that involve machines or devices through a network without human intervention. M2M applications can remotely configure machines, collect data from machines, process collected data to make decisions and send notifications in unusual situations. These M2M applications can be used to monitor a network, collect data from the network and then make decisions whether to trigger congestion management solutions and policies. Makara et al (2012) proposed an optimized scheduling approach that exploits multiuser diversity by considering each user's instantaneous downlink conditions and QoS information when distributing resources. They propose an approach towards the management of resources in the LTE downlink that fully exploits multiuser diversity

IV. THE PROPOSED PRIORITIZED SCHEDULING ALGORITHM

This work proposed prioritized scheduling algorithm to manage congestion on the LTE network. In this proposed priority based algorithm, traffic will be assigned resources based on their classes and priority levels. There is a queue for each traffic type and traffic belonging to a class joins the queue meant for that class. The algorithm is explained thus.

i There is a queue for each traffic class, where Real-time Conversational traffic class with high priority is T_1, Real-time Streaming with medium priority is T_2, Near Real-time Interactive with normal priority is T_3 and Non Real-time Background is T4 as shown in Fig. 1. Traffic belonging to a class joins the queue meant for that class. Each queue has a limit that is based on the number of radio blocks available.

ii The threshold is set for each traffic class T_1, T_2, T_3 and T_4. A threshold here is the maximum time a user can spend on the network before equal priority user can preempt it in a situation where there is no RB to occupy. The threshold only work for equal priority user since lower priority user will automatically be preempted by higher priority user.

iii. On each queue, request is ordered according to their Channel Quality Indicator. The CQI value at the time of admission of each request is based on the distance of the user equipment from the Base-Station as specified in the Table II.

iv Assign available resource blocks to the traffics demand. The assignment will be based on the priority scheme specified in Table I. That is, $T_1 > T_2 > T_3 > T_4$, Traffic of higher priority as defined by the traffic class pre-empts traffics of lower priority or lower traffic class.

v. If the traffic to be serviced by the resource blocks did not see any free RB to use, then, if it is of equal priority with any traffic occupying the RB, then the threshold limit is used to pre-empt old requests in order to admit new ones, that is, pre-empt any old traffic that has reached the set threshold in order to admit a new request into the queue.

vi Move it to the next traffics on the queue (Q_T) for scheduling in the next round.

$T_1 \rightarrow Q_1$,
$T_2 \rightarrow Q_2$,
$T_3 \rightarrow Q_3$,
$T_4 \rightarrow Q_4$

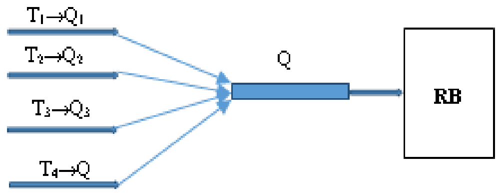

Fig. 1. Resource block assignment queue

TABLE II
CQI BASED ON DISTANCE FROM BASE-STATION

Distance (m)	MCS Index	CQI
100	28	15
2000	27	13
2400	25	12
2600	23	11
3100	20	10
3500	16	9
4000	14	8
4600	12	7
6000	9	6
7000	7	5

V. RESULTS AND ANALYSIS

The proposed prioritized scheduling algorithm flowchart is shown in Fig 2 and is simulated using MATLAB and Simulink. Simulink is a block diagram environment for multi-domain simulation and model-based design (Math Works, 2015). The Simulink is integrated with MATLAB, enabling users to incorporate MATLAB algorithms into models and export simulation results to MATLAB for further analysis. MATLAB is ideal for the modeling and simulating the LTE network, as well as implement and test the new proposed prioritized scheduling algorithms on the LTE network. The data parameter that was used for the simulation are listed in Table III.

Table III
Simulation Parameters

PARAMETERS	VALUE
Channel Bandwidth	1.4 - 20MHz
Mod Type	2
SNR (dB)	12
Number of RBs (numRBs)	30
Number of SubFrames	Ceil (numRBs/2)
Total users	[6, 5, 8, 7]
Scheduler	PSA, Best CQI, Round Robin

The performance of the Proposed Prioritized Scheduling Algorithm (PSA) is evaluated and compared to Round Robin (RR) and Best CQI. Results of two different algorithms (PSA, Round Robin and Best CQI) using the LTE metrics of Resource Block allocation, fairness and throughput of network are discussed thus.

A. *System throughput*

System throughput is the total number of bits successfully transmitted over the air interface from the UE up to the eNode-B over the total simulation time (AlQahtani and AlHassany, 2014). Equation 1. shows the formula used for the throughput.

$$Throughput = \frac{B}{Tsim} \quad \ldots\ldots\ldots 1$$

where

B = the total amount of received bits and

Tsim = the total simulation time.

B. *Total RBs allocated*

It refers to the sum of the number of resource blocks assigned to the traffic classes from the available resource blocks. Total RBs allocated is represented in equation 2.

$$Total\ RBs = \sum Rt_1 + Rt_2 + Rt_3 + Rt_4 \quad \ldots\ldots .2$$

where

Rt_1 = number of resource blocks allocated to traffic class 1
Rt_2 = number of resource blocks allocated to traffic class 2
Rt_3 = number of resource blocks allocated to traffic class 3
Rt_4 = number of resource blocks allocated to traffic class 4

C. *Fairness*

Fairness is the measure of equal share of the RBs among UEs of the same class. In this case, we are considering whether the traffic classes get allocated fair share of the network resources. This is referred to as allocation fairness. One of the most famous formula for fairness is the Jain's fairness index shown in equation 3 (Chisung and Dong, 2007; Lin and Laurie, 2008; AlQahtani and AlHassany, 2014). It is given as:

$$f(R_1, R_2, \ldots, R_k) = \frac{\left[\sum_{k=1}^{K} R_k\right]^2}{k\sum_{k=1}^{K}(R_k)^2} \quad \ldots\ldots\ldots 3$$

where there are K UEs in the LTE system and R_k is the number of RBs given to UEi.

The PSA has the highest throughput of 20.8055 Mbps compared to Round Robin with 19.7415 Mbps and Best CQI with 0.66386 Mbps. This means that there is high number of successfully transmitted bits over the air interface from the UE up to the eNodeB, and there is less degradation of service. Users' job are done in time and the risk of congestion occurring is minimized. Fig. 2 shows the throughput for the different scheduling algorithms.

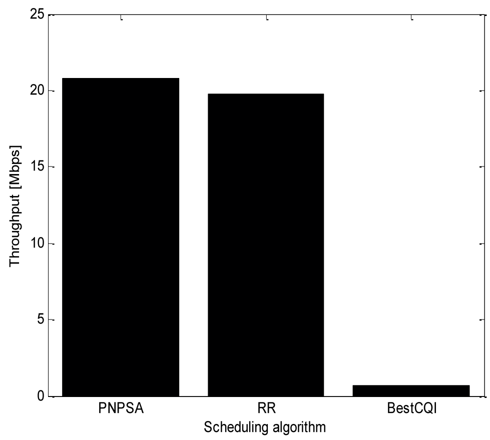

Fig. 2. System throughput comparison between the scheduling algorithms

In terms of allocated RBs, the PSA has a total of 29 (T1=12, T2=9, T3=6, T4=3) allocated RBs according to

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the priority attached to the traffic classes such that $T_1 > T_2 > T_3 > T_4$. The Round Robin allocates 30 (T1=8, T2=8, T3=7, T4=7) RBs. This algorithm tries to satisfies all traffic classes by allocating almost the same RBS to all traffic classes. It therefore leads to wastage and under-utilization of network resources as users who require more RBSs are assigned almost the same RBs as users that do not need as much. The starved users stay too long on the queue and this leads to congestion. The best CQI algorithm allocates RBs based on the channel quality and the total RBs allocated = 30 (T1=3, T2=9, T3=11, T4=7). It allocates resources based on the proximity of the UE to the e-NodeB, regardless of the traffic class the UE request belongs to. As shown in Fig. 3, traffic T_3 with normal priority is assigned the highest RB by the Best CQI algorithm while, traffic T_1 with high priority is assigned the lowest resource block. This ultimately leads to congestion in the network and a serious degradation of service enjoyed by users. The PSA performs best among the three algorithms in terms of resource block allocation.

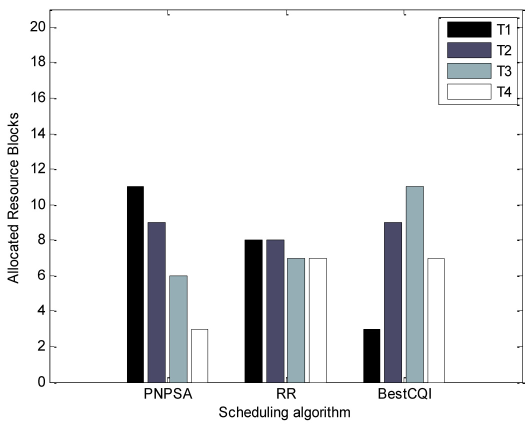

Fig 3. Resource block allocation comparison between the algorithms based on traffic

In terms of allocation fairness to the traffic classes, the allocation fairness index shows that PSA has 0.83433 while Round Robin has 0.77586. Fig. 4 shows that the PSA performs better in terms of fairness to all the traffic classes and followed by Round Robin.

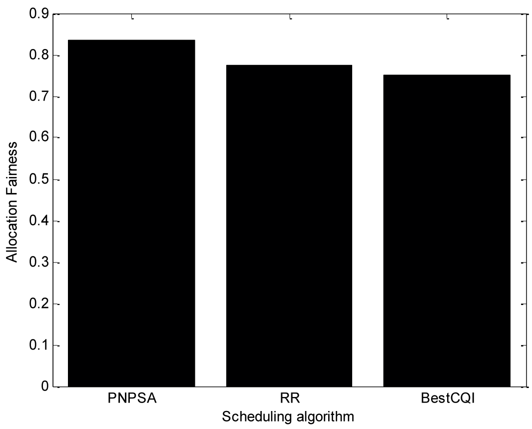

Fig. 4. Allocation fairness comparison between the scheduling algorithms

VI. CONCLUSION

In this work, a prioritized scheduling algorithm (PSA) for Long Term Evolution (LTE) Network was developed. User requests were classified into different traffic classes based on the nature of the request and each class was assigned a priority value. The priority is based on the urgency of the requests. The simulations showed that the proposed method perform well than the Round Robin that is popularly used in LTE Networks.

REFERENCES

[1]. Rysavy P. (2015). Enabling Innovation and Customer Experience. LTE Congestion Management. Rysavy Research LLC.

[2]. LiveU. (2014). Beating Cellular Congestion. Xtender White-Paper.

[3]. LTE. (2015) Retrieved from www.3gpp.org/technologies/keywords-acronyms/98-lte

[4]. Fayssal Bendaoud, Marwen AbdennebiFedoua Didi (2014). “Survey On Scheduling And Radio Resources Allocation In Lte”. International Journal of Next-Generation Networks (IJNGN) Vol.6, No.1, March 2014

[5]. LTE Network Architecture (2013) Retrieved from http://www.tutorialspoint.com/lte/lte_network_architecture.htm

[6]. Akyildiz I. F., Gutierrez-Estevez D. M., Reyes E. C. (2010) “The evolution to 4G cellular systems: LTE-Advanced”, Physical Communication 3 (2010) 217–244. www.elsevier.com/locate/phycom

[7]. Habaebi M.H., Chebil J., Al-Sakkaf A.G., Dahawi T.H. (2013). Comparison between Scheduling Techniques in Long-Term Evolution, IIUM Engineering Journal Vol. 14 No. 1.

[8]. Mohammed A.A., Moshen M.T., Mohamed E. (2013). LTE QoS Dynamic Resource Limitation and Queue Stability Constraints. International Journal of Computer Networks & Communications. Vol. 5.

[9]. Lucente C. (2012). Priority and Pre-emption mechanisms in LTE Broadband Communications Networks. Technology Advisory Group, volume 4.

[10]. Syski R. (1986). Introduction to Congestion Theory in Telephone Systems. Elsevier Science Publishers B. V.

[11]. Kuboye B.M. (2010). Optimization models for minimizing congestion in Global System for Mobile Communications (GSM) in Nigeria. Journal Media and Communication Studies Vol. 2(5), pp. 122-126.

[12]. Dushyanth B. (2006). QoS in Cellular Networks. http://wwww.cse.wustl.edu/jain

[13]. Frenzel L.E. (2008). Principles of Electronics Communication Systems. Mcgraw Hill Publishers. 3rd Edition. Available at http://www.amazon.com/Principles-Electronics-Communication-Published-Education/dp/B008DMCPU0

[14]. Omotoye K.O., Akogbe M.A., Olaniyi O. M., Inyiama H.C. (2014). Congestion Management on GSM Networks using Traffic Class Prioritization. International Journal of Engineering Research & Technology (IJERT) Vol. 3, Issue 12.

[15]. Maeder A., Schmid S. (2012). Traffic Management Solutions for User-Plane Congestion in Mobile Networks. NEC Laboratories Europe.

[16]. Ahmad K., Kumar S., Shekhar J., (2012). Network Congestion Control in 4G Technology through Iterative Server. International Journal of Computer Science Issues, Vol. 9, Issue 4.

[17]. Bahreyni M.S., Naeini V.S. (2014). Fairness Aware Downlink Scheduling Algorithm for LTE Networks. Journal of Mathematics and Computer Science

[18]. Kanagasundaram K., Kadar Nawal G.M. (2013). Optimized Resource Allocation and Scdeuling for Real-Time Services in LTE Networks. Journal of Theoretical and Applied Information Technology. Vol. 58 No.1.

[19]. Makara J. and Neco V. (2012). Downlink Packet Scheduling for Variable Bitrate Traffic in LTE Networks.

[20]. http://www.satnac.org.za/proceedings/2011/papers/Network_Engineering/187.pdf

[21]. Tan, Yifeng (2009). Active Queue Management for LTE uplink in the eNode-B. Masters Thesis, Dept of Communication Engeneering Helsinki university of technology .http://urn.fi/URN:NBN:fi:aalto-201203071276.

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[22]. Amokrane Ahmed (2011) Congestion control in the context of Machine Type Communications in 3GPP LTE Networks, ENS Cachan, Brittany Extension.

[23]. MathWorks. (2015). LTE System Toolbox. www.mathworks.com/products/lte-systems/

[24]. AlQahtani and AlHassany, (2014). Performance Modelling and evaluation of novel Scheduling Algorithms for LTE Networks. Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), January Edition, Volume 4.

[25]. Chisung B., and Dong C. (2007). Fairness-Aware Adaptive Resource Allocation Scheme in Multihop OFDMA System, Communication letters IEEE, 11 (2007) 134-136.

[26]. Lin X. and Laurie C. (2008). Improving fairness in relay-based access networks, in ACM MSWIM, (2008) 18-22.

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)