QoS-Based Fairness-Aware Downlink Scheduling in LTE-Advanced

Ying-Hong Wang
Dept. of Computer Science and Information Engineering, Tamkang University, Taiwan, R.O.C.
inhon@mail.tku.edu.tw

Chih-Hsiao Tsai
Dept. of Information Technology, Takming University of Science and Technology, Taiwan, R.O.C.
chtsai2104@gmail.com

Hao-Yi Huang
Dept. of Computer Science and Information Engineering, Tamkang University, Taiwan, R.O.C.
maxiory@hotmail.com

Yi-Jia Tong
Dept. of Computer Science and Information Engineering Tamkang University, Taiwan, R.O.C.
esc0928@hotmail.com

摘要

近年來，台灣的無線上網的需求迅速的增加。因用戶大多是使用視訊通話或是觀看影音串流等需要高頻寬等需求，在無線資源有限的情況下如何去滿足大多數人的需求成為一個非常重要的議題。在無線資源管理中，排程器的設計是其中最主要的部份。故本論文提出一個能盡量滿足大多數使用者需求的排程機制，此機制能先判斷連線的種類是否需要即時的服務，在逾時之前將需求作處理。若是使用者處於收訊不好的環境，則其需求有較大的機會因避免浪費系統資源而被放棄掉，且飢餓(Starvation)跟公平性的問題也都被正視，不會有連線長期佔有資源與等待甚久不被服務的情況以期達到服務品質的保證。實驗結果將證明本文所提之方法會在執行時間效率上優於傳統之作法。

Abstract

Recent years, the demands of wireless networks in Taiwan grow fast because of users almost take a cam call or watch video streams on youtube. How to fulfill most users' demand under limited wireless internet resources situation becomes an important issue. In wireless networks resource management, the design of schedulers is the most important part. This paper proposed a scheduling scheme that meets most users' demands .This scheme distinguishes whether this connection is real-time or not and dealing it before it time out. If a user is located a cell edge, his demands may be rejected because of avoiding waste system resource. Besides Starvation and Fairness problems are also considered. There are no connections occupy resource for a long time or connections wait for a long time to be served to meet QOS(the quality of service).The experiment results will proof our method is efficiency than others.

Keywords: scheduling, Quality of Service, fairness

1. Introduction

With the improvement of mobile phone, People download Apps and watch stream videos on mobile phone. The huge data amount transmission depends on wireless radio resource. Compare with wire network, wireless radio resource is rare.

According to wireless radio resource is rare, it's important to allocation wireless radio resource efficiently. In this paper, We allocation wireless radio allocation by users' requests like email, stream, and file download. The past research focused on fairness and throughput until 3G was proposed, Quality of Service (QoS) started to be concern

In this thesis, we proposed a downlink packet scheduling focus on the following parameter:

- Channel Quality Indicator(CQI)
- Type of user requests
- Dynamic priority coordinating
- Fairness

The research is expected to achieve the following goals :

- Guarantee QoS
- Maximize Throughput
- Provide Fairness

2. Related Work

In the current fourth-generation communication system techniques, including by the 3GPP (Third Generation Partnership Project) research and development of LTE (Long Tern Evolution) and WIMAX (Worldwide Interoperability for Microwave Access), LTE technology can result in the original base station technology upgrades, system operators do not have to pay the additional cost of erecting the relationship between base stations, LTE technology has made most of the industry's support. With the release of Release10 also has evolved to LTE-Advanced, making communication transmission theoretical value can reach 500Mbps uploading and downloading 1Gbps.LTE-A system consists of Evolved Packet Core Networks(EPC). And Evolved Universal Terrestrial Radio Access Network

(E-UTRAN), composed to handle scheduling the eNB which is located in the E-UTRAN architecture shown in Fig.1.

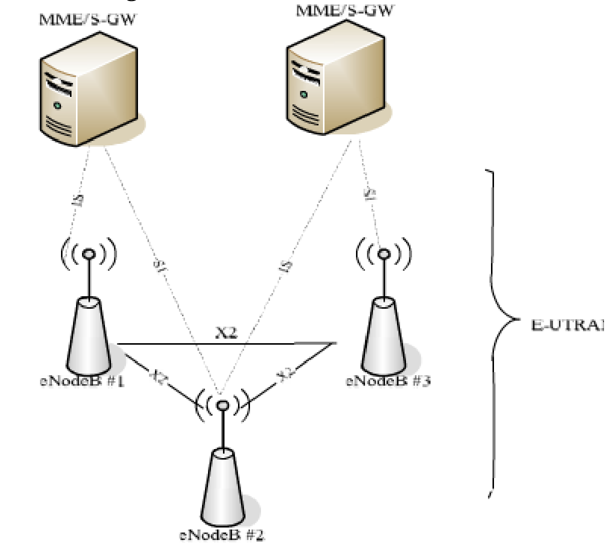

Fig 1 E-UTRAN architecture

Quality of Service (QoS), from as early as the IEEE 802.16 standard [1] have been proposed, according to the needs of different users, providing different levels of service. In LTE-A, although the parameters are different, but the concept is similar. When the system load is high, can be managed through negotiation that way, as far as possible to meet most of the needs of QoS, in this section will be LTE-A, do a brief introduction of QoS.

In LTE-A [2] in the established connection is divided into two categories, namely, Guaranteed Bit Rate (GBR), and Non-Guaranteed Bit Rate (non-GBR). In GBR classes generally require lower latency (Delay) and packet loss rate (Packet Error Loss Rate), but will have a higher priority. Shown in Fig.2.

QCI	Resource type	Priority	Packet delay budget	Packet error loss rate	Example services
1	GBR	2	100 ms	10^{-2}	Conversational voice
2		4	150 ms	10^{-3}	Conversational video (live streaming)
3		3	50 ms	10^{-3}	Real time gaming
4		5	300 ms	10^{-6}	Non-conversational video (buffered streami
5	Non-GBR	1	100 ms	10^{-6}	IMS signaling
6		6	300 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2 sharing, progressive video, etc.)
7		7	100 ms	10^{-3}	Voice Video (live streaming) Interactive gaming
8		8	300 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2 sharing, progressive video, etc.)
9		9			

Fig 2 LTE QoS parameter

OS Shin and K. Lee in [3] proposed Round Robin (RR) algorithm, which features provide the most complete fairness, which features provide the most complete fairness, each working in a unit of time have the same chance of being selected. In packet scheduling, it does not consider the quality of the user's channel (channel quality Indicator, CQI), non-discriminatory provision of services. Although the benefit is provided between the user absolutely fair, but because when the user is in poor channel quality of service, it may lead to resources being wasted throughput. Shown in Fig.3, although the CQI UE1 is better, but the chance of the two is scheduled to be equal.

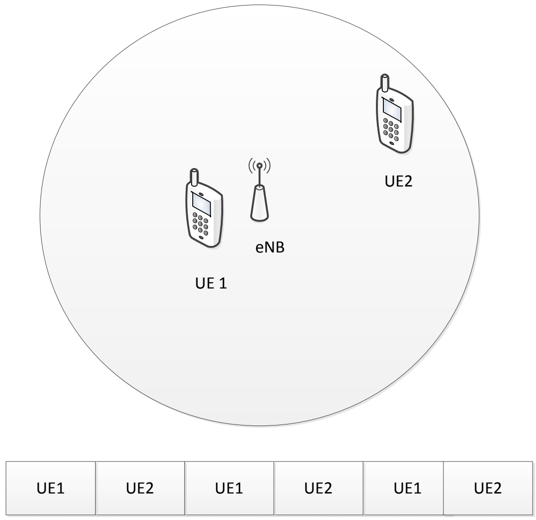

Fig 3 RR scheduling

Yujia Wang in [4] proposed Maximum Carrier Interference (MAX C / I) algorithm concept always select the best channel quality user has to provide services, and these users are usually from the base station closer. If only two users are waiting to be scheduled, the channel quality will always be a better user selected poor user almost no chance of being selected. Thus, the algorithm can get the maximum throughput, but lost fairness. Therefore algorithms are often used as the upper limit of measurement throughput. Shown in Fig.4, UE1 has a good CQI, the scheduling get a lot of opportunities.

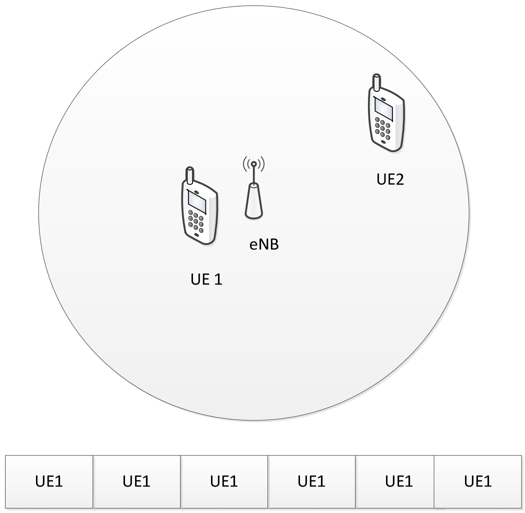

Fig 4 MAX C/I scheduling

Andrews, M. in [5] mentioned Modified Largest Weighted Delay First (M-LWDF) this algorithm, it is improved by re-PF algorithm. Compared to the PF is, M-LWDF the user can tolerate delay time into account, so that users who need immediate service can have greater access to services, although it looks to improve the PF's shortcomings, but delay is not required for non-real-time users, but to get the opportunity to schedule reduced. Ju Yong Lee in [6] proposed a modified version of the M-LWDF M-LWWF, the average rate for a relatively low priority users to improve and reduce the packet loss rate.

Wenyu Li in [7] Priority update granularity (PUG) algorithm dynamically calculated for each unit time by the user density and latency requirements, redistribution of resources. The paper does not reduce the throughput to improve the fairness of the premise.

Bin Liu in [8] proposed M-EDF-PF algorithms to PF fairness and EDF on latency and throughput fairness requires balance, and in the LTE downlink system for realtime service with good throughput

3. Robot-Based Deployment Mechanism

QoS-Based Fairness-Aware Downlink Scheduling, (QFS). LTE-A is based on quality of service to meet the requirements of the design, build, when demand will assign a priority number, the smaller the number the higher the priority behalf. If demand can't meet the QoS requirements, QFS will dynamically adjust its priorities in order to improve the situation.

QFS possess and maintain fairness guaranteeing QoS characteristics, in order to achieve this feature paper presents three directions:

- Provide better QoS.
- Maximize Throughput
- Providing Fairness among User Equipments(UEs)

In QFS, using three queues are arranged according to priority needs, These queues are GBR queue, Non-GBR queue and Urgent queue. In this among Urgent queue has the highest priority, when there is a demand assigned to Urgent queue, the highest priority are processed. GBR queue has the second highest priority, if Urgent queue waiting to be scheduled no demand, QFS will start to demand GBR queue scheduling. Non-GBR queue then has the lowest priority in the schedule if QFS finished the previous two queue only when there is a surplus of resources for the Non-GBR demand for queue scheduling. On the demand to establish when, QFS will be based on demand in the LTE QCI standard classification to be divided into classes and Non-GBR GBR class and assigned to the corresponding queue inside.

In GBR queue, the priority needs of the judgment is based on their channel quality conditions (Channel Quality Indicator, CQI) to distinguish heights, channel quality conditions is provided by the user's device information to the eNB. CQI is higher, the better channel conditions, it represents the higher the priority, QFS according to the priority needs to allocate resources to the level of demand.

In the Non-GBR queue, the demand is based on the priority level of CQI values to make judgment on Urgent queue and queue GBR needs are dealt with, there are resources in the remaining words, Non-GBR queue demand the highest priority will be scheduled. Fig.5 is a schematic classification requirements.

Additionally, a user may simultaneously GBR and Non-GBR demand, that in order to provide users with fairness, each user needs only be scheduled again until the needs of each user are scheduled over so far. When each user scheduling resources remaining after the first time there, QFS will continue to provide a mechanism according to the original schedule for the second time or more. This is to allow the user also has CQI low chance of being scheduled, and will not let the user have better CQI occupy most of the resources.

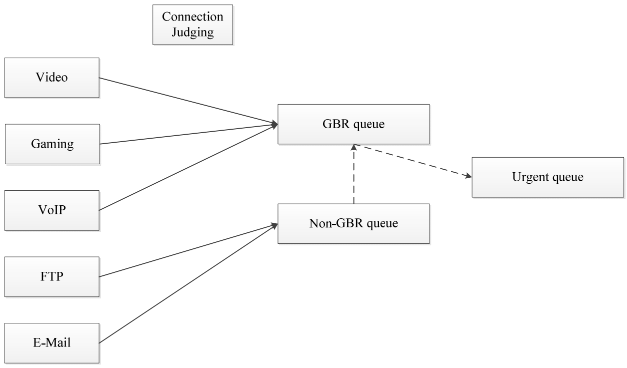

Fig 5 request classification

In GBR queue, if there can not be the next frame timeout occurs scheduling needs, QFS will move Urgent queue such demand, so that these needs in the next frame arrives priority to be scheduled to reduce the likelihood of timeouts.

If such demand is huge cause insufficient resources in Urgent queue scheduling all needs, as such needs to be completed within the specified time, then the retransmission timeout is meaningless, so are not scheduled to demand will be discarded (Drop).

As Urgent queue in order to avoid the occurrence of timeout has its urgency, and therefore Urgent queue scheduling needs to be included in the calculation principle of fairness is not being scheduled times.

In the Non-GBR queues in order to avoid the occurrence of hunger, all Non-GBR demand has an additional parameter called Starvation Value, if there is not a transmission time interval is scheduled to if this parameter plus one, when this parameter is greater than a given value, the representative hunger is about to happen, in order to avoid this phenomenon, QFS such needs will move GBR queue first, making such demand GBR queue has the highest priority. Regardless of its priority transmission time interval after a few will not be surpassed until the scheduled date.

QFS will calculate at a transmission time interval, three queues upper and lower bounds of the demand for bandwidth and available bandwidth according to the size of the comparison value, to determine the three kinds of allocation of resources, the following definitions of the various parameters of this section.

1. B^{all}_{Urgent}: In Urgent queue all the needs are most urgent needs to be scheduled, so all the required bandwidth requirements are required to be taken into account.

2. B^{min}_{GBR}: In GBR queue, GBR needs to ensure that every frame needs to be a minimum data transfer rate to B^{min}_{GBR} represents a transmission time interval GBR queue minimum required bandwidth.

3. B^{max}_{GBR}: GBR queue represents a transmission time interval all the required bandwidth.

4. B^{all}_{NGBR}: In Non-GBR queue, Non-GBR needs not guarantee the data transfer rate to B^{all}_{NGBR} representatives of Non-GBR queue all the necessary bandwidth.

5. $B^{max}_{request}$: to $B^{max}_{request}$ represents a transmission time interval of three queues that will require maximum bandwidth, $B^{max}_{request} = B^{max}_{GBR} + B^{all}_{NGBR} + B^{all}_{Urgent}$.

6. $B^{min}_{request}$: to $B^{min}_{request}$ represents a transmission time interval within three queues that will be the minimum required bandwidth, $B^{min}_{request} = B^{all}_{Urgent} + B^{min}_{GBR}$.

QFS allocate bandwidth based on the total allocated bandwidth to be compared with the size, can know what kind of situation should be based on bandwidth allocation.

Situation1: $B^{available} > B^{max}_{request}$ (1)

If you meet the formula (1), represents the currently available bandwidth to meet all needs, in this case, QFS will be the first in the queue to meet the needs of Urgent, then all the requirements GBR queue bandwidth, and finally the Non-GBR queue where all the requirements of bandwidth.

Situation2: $B^{max}_{request} > B^{available} > B^{min}_{request}$ (2)

If you meet the formula (2) cases, represents the currently available bandwidth at the maximum and the minimum required bandwidth being, QFS will first meet the needs of Urgent queue, then queue to meet the smallest data GBR guaranteed rate, that is B^{min}_{GBR}, remaining resources will be based on priority queues continue to meet the needs of GBR.

Situation3: $B^{min}_{request} > B^{available}$ (3)

If you meet the three-case represents a heavy burden on the current system, even the minimum requirement no way to meet, QFS will first try to schedule queues in Urgent needs in order to avoid time-out, if there are any remaining bandwidth can be allocated , and then try to meet the needs of the smallest GBR queue information guaranteed rate. Fig.6 is a flowchart of bandwidth allocation

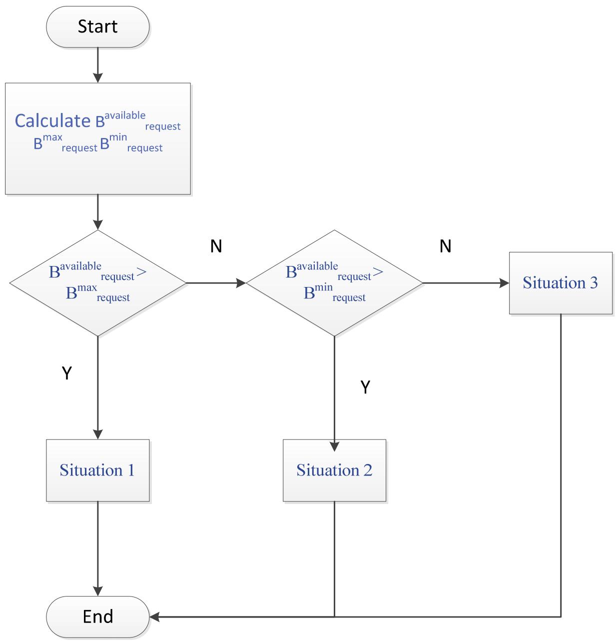

Fig 6 flowchart of bandwidth allocation

4. Simulation and Analysis

We use the LTE simulator[9] to simulate LTE network environment, and increase the user to simulate through the metropolitan area increased burden on the base station, manned by the number in the system increases, while whether the proposed method is based on Chapter III ensure QoS in latency and throughput, and to ensure fairness. Representative both scheduling and other methods, can know from section 2-3, MAX C / I can be used as the limit system throughput, therefore, in addition to the analog system throughput will MAX C / I algorithm join comparison. The remaining simulation are compared with the M-LWDF, because M-LWDF algorithm is more representative in recent years.

Table 1 simulation parameter

Parameter	Value
Carrier frequency	2 GHz
System bandwidth	20 MHZ
SubCarries per RB	12
UE numbers	10,20,30,40,50,60
TTI	1ms
Starvation Value	5(5ms)
GBR data	242 kb/s
Non-GBR data	300 kb/s

Fig.7 and Fig.8 is the throughput simulation among QPS, M-LWDF and MAX C/I.

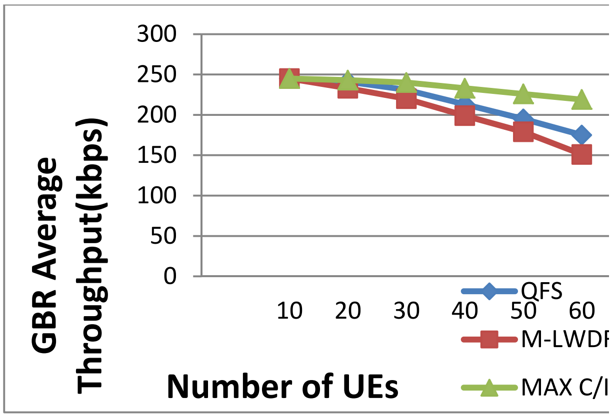

Fig 7 GBR throughput

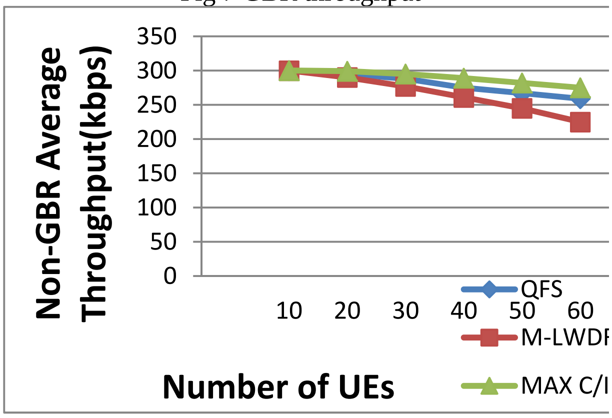

Fig 8 Non-GBR throughput

Can clearly be seen from Fig.7 and Fig.8 of the proposed algorithm in the QFS GBR throughput slightly worse than the M-LWDF algorithm, it is because QFS going timeout to allow users to get the opportunity to schedule, but this user may be in a bad channel conditions which led to decreased throughput, while the Non-GBR throughput also have the same situation, in order to avoid hunger, resulting in decreased throughput, but compared with the M-LWDF not too different .

Fig.9 is the packet delay simulation.

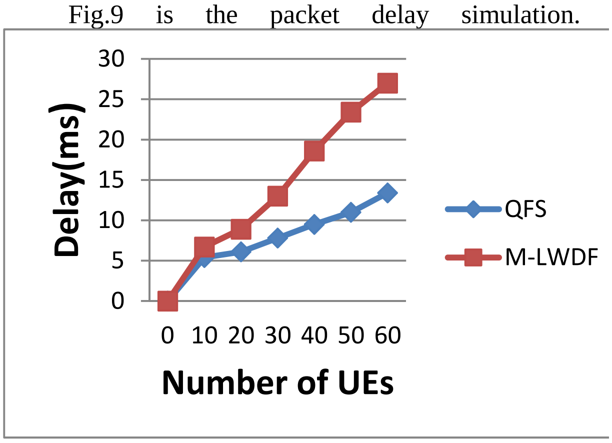

Fig 9 packet delay simulation

Can be observed from Fig.9, when the user more for a long time, M-LWDF delay rising faster, it is because M-LWDF on request timeout handling is not as QFS to instant, QFS found if requested in the next if a transmission time interval will be scheduled over the circumstances, this requirement will be moved Urgent queue, it has a very big opportunity for the next transmission time interval is scheduled in order to avoid time-out, therefore, can be delayed rise compared to M-LWDF come slowly.

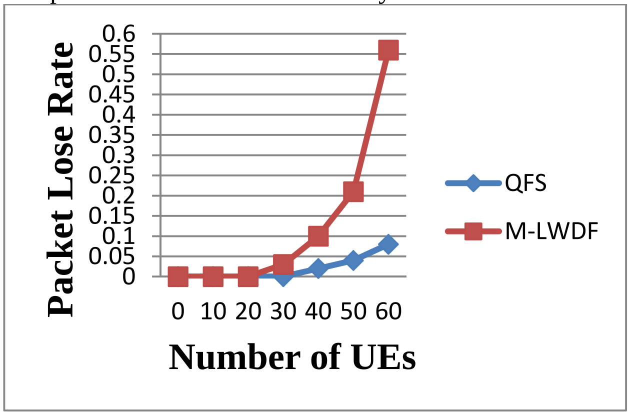

Fig 10 packet loss rate simulation

From Figure 10 it can be observed before the user is less than 30, because the two algorithms are tolerable packet delay will be taken into account, therefore, use less, the loss rate can be maintained at a low, but as Figure 16, QFS requirements can be reduced due to the timeout rate, so it can also depress loss rate, while M-LWDF because the processing time out and did not come immediately QFS, consequential loss rate is therefore worse than the QFS.

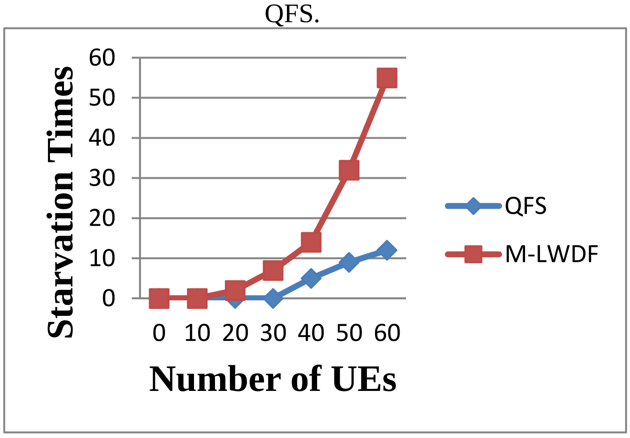

Fig 11 staration times simulation

Can be seen from Figure 11, QFS mechanism because there Starvation Value mechanism, after five transmission time interval after it is yet to be scheduled, will increase its priority, so the opportunity is scheduled to increase, resulting in less than 30 before the user does not starvation. In contrast M-LWDF because of the lack of non-real-time requirements of fairness, if immediate substantial increase in demand, the non-real-time requirements difficult to get the chance to be scheduled, which led to a substantial increase in the starvation value

5. Conclusion and Future Work

We proposed QoS-Based Fairness-Aware Downlink Scheduling, under the framework of

LTE-A system to dynamically adjust the priority needs of its various services and access to quality assurance, to avoid the occurrence of hunger and provide the fairness among users in the simulation results, quality of service parameters for LTE, such as delay and packet loss rate compared to other algorithms have encountered a good performance.

Uplink and downlink in the network and is considered to be one, so in future studies will uplink situation into account, and in smart phones, power has always been an important issue, if the phone continue to send and receive data, or at listening conditions, unable to cope with the current battery technology, how to sleep in the schedule by adding a mechanism to lose power but not receiving important messages, will be the focus of future research.

References

[1] IEEE Standard 802.16 Working Group, *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air interface for Broadband WirelessAccess Systems* (P802.16Rev2/D3).Feb.2008.

[2] Mehdi Alasti, Behnam Neekzad, Clearwire, Jie Hui, Rath Vannithamby, Intel Labs, "Quality of Service in WiMAX and LTE Networks," *IEEE Communications Magazine*, pp.104 – 111, May 2010.

[3] O.S.Shin and K. Lee, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Commun*. Letter, vol. 7, pp. 109-111, March 2003

[4] Yujia Wang, Xue Yunfeng, Chaogang Yu, "Scheduling algorithm for distributed station based on particle swarm optimization," in *Proc. 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pp. 4292 – 4295, 2011

[5] Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, SCHEDULING IN A QUEUING SYSTEM WITH ASYN-CHRONOUSLY VARYING SERVICE RATES, Probability in the *Engi-neering and Informational Sciences*, vol. 18, no.02, pp. 191- 217, 2004

[6] Ju Yong Lee, Sorour S. ,Valaee, S. , "Dynamic parameter adaptation for M-LWDF/M-LWWF scheduling" in *Proc. IEEE Transactions on Wireless Communications*, vol.11, no. 3, pp. 927-937, March 2012.

[7] Wenyu Li ; Li Jin ; Zhongfang Wang ; Lin Zhang ; Yu Liu, "A dynamic Priority update Granularity adjustment algorithm based on user density and service delay requirement in LTE downlink scheduling" in *Proc. 2012 3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 122-127, 2012.

[8] Bin Liu ,Hui Tian ,Lingling Xu, "An Efficient Downlink Packet Scheduling Algorithm for Real Time Traffics in LTE Systems" in Proc .*Communications and Networking Conference (CCNC)*,2013IEEEObjectIdentifier:10.1109/CCNC.2013.6488471 Publication Year: 2013, Page(s): 364 – 369.

[9] G. Piro "LTE-Sim - the LTE simulator," Available online at: http://telematics.poliba.it/LTE-Sim