# Scheduling Both GBR and Non-GBR Uplink Bearers Based on Moving Average of Data Rate

Fang-Chang Kuo, Kuo-Chang Ting*, Chih-Cheng Tseng, Hwang-Cheng Wang, and Ming-Wei Chen

National Ilan University, Taiwan
kfc@niu.edu.tw
* Minghsin University of Science and Technology, Taiwan

*Abstract*—**Many resource allocation schemes have been proposed based on different criteria such as system throughput, fairness, transmission power, user priority and others. According to the 3GPP specifications, however, as long as a GBR (guaranteed bit rate) bearer is admitted, eNB has to allocate resource blocks (RBs) for this bearer to ensure its transmission right. At the same time, eNB also must distribute the remaining RBs to non-GBR bearers so as to achieve high radio resource utilization. However, 3GPP has not specified how to measure GBR. In this paper, we use exponentially weighted moving average (EWMA) to define the measurement of data rate and show that it is beneficial to the scheduling on both UE and eNB sides. We also present a scheduling scheme, AAG-2, which aims at ensuring QoS for all in-progress GBR sessions, while at the same time efficiently allocating RBs to non-GBR sessions.**

***Keywords-uplink scheduling, resouce allocation, EWMA, non-GBR, AAG-2***

## I. Introduction

Dynamic resource allocation (DRA) for LTE uplink bearers is an important task of radio resource management (RRM). The schemes for allocating resource blocks (RBs) to user equipment (UE) have been extensively studied. The schemes proposed in the early articles usually emphasize maximizing the system throughput of an eNB. For example, the Recursive Maximum Expansion (RME) and relevant schemes allocate free RBs to a UE which has best channel quality [1, 2].

Many researchers felt that fairness is also a very important criterion. As a result, some paradigms have emerged. A resource chunk (RC) paradigm algorithm treat RC, which is a fixed number of RBs, as a basic unit of allocation to UEs. Any UE that is not scheduled in a transmission time interval (TTI) will have higher priority in subsequent TTIs [3-5]. One simple scheduling scheme of this paradigm is frequency domain round robin (RR), which allows RCs to be assigned to multiple UEs within one TTI in a cyclic manner [6]. The proportional fair (PF) paradigm allocates resources to users based on the ratio of their channel condition over their lifelong service rate [7, 8]. As a result, users with poor channel conditions will get some resources but fewer than those with better channel conditions. However, both of these paradigms do not take UE QoS requirements into account when forming the allocation matrix and still suffer problems related to fairness in terms of throughput. Some resource allocation schemes are concerned with the transmission power of UEs [9, 10]. Authors in [11] proposed a QoS-aware resource allocation paradigm for LTE uplink scheduling that gives greater advantage to UEs with high-priority data, while not starving other users. However, bearers in LTE are classified into GBR and non-GBR ones, rather than priorities among UEs.

The above-mentioned dynamic resource allocation (DRA) schemes are designed based on different criteria. We would like to briefly review some concepts pertaining to DRA in 3GPP specifications [12, 13] before presenting our scheme.

- The task of radio admission control (RAC) is to admit or reject the establishment requests for new radio bearers. The goal of RAC is to ensure high radio resource utilization and at the same time maintain proper QoS for in-progress sessions.
- DRA typically takes into account the QoS requirements associated with radio bearers, the channel quality information for UEs, interference situation, buffer status, etc.
- Each GBR (guaranteed bit rate) bearer is associated with throughput requirements expressed in terms of (a) GBR, which is the bit rate that can be expected to be provided, and (b) maximum bit rate (MBR), which is the maximum bit rate that can be expected to be provided.
- The throughput of non-GBR bearers in a UE is denoted in group by aggregate maximum bit rate (AMBR).

Based on these key concepts, we come up the following observations, which are often neglected in previous research:

- As long as a GBR bearer is admitted, eNB has to allocate sufficient RBs to meet the guaranteed data rate irrespective of the channel quality. However, 3GPP has not specified how to measure GBR, MBR and AMBR.
- Apart from allocating RBs for GBR bearers to ensure their QoS, eNB also must distribute the remaining RBs to non-GBR bearers in order to achieve high radio resource utilization.
- Buffer status report (BSR) also provides important information for DRA.

In this paper, we present an algorithm that uses exponentially weighted moving average (EWMA) to define the measurement of GBR and facilitate the operation of DRA. We also present a scheduling scheme, AAG-2 (Allocate As Granted, version 2), which employs BSR to prevent the waste of RBs. AAG-2 aims at ensuring QoS for all in-progress GBR sessions, while at the same time efficiently distributing RBs to non-GBR sessions so as to improve resource utilization.

In the following, we describe the use of EWMA for data rate measuring in Section II. Section III briefly describes the operation of AAG scheme, which is used as a module in our

This paper is supported by MOST of Taiwan under grant no. 100-2221-E-197-019

QSHINE 2015, August 19-20, Taipei, Taiwan

DOI 10.4108/eai.19-8-2015.2260881

proposed scheme AAG-2 that is presented in Section IV. The performance evaluation of AAG-2 is presented in Section V. We list the benefits of using EWMA in section VI and draw conclusions in section VII.

## II. EWMA: Method to Measure Data Rate

In order to facilitate DRA in LTE, we adopt EWMA to define the data rate of a bearer.

Let $R_{m,GBR}^{grant}$ be the GBR that is granted by the RAC for the *m*-th UE ( $UE_m$ ). Because the time granularity for allocating RBs is a TTI, which is one millisecond, the data rate can be alternatively expressed as the number of bits transmitted per TTI, $B_{m,GBR}^{grant}$, where

$$B_{m,GBR}^{grant}\left(\text{bit}/\text{TTI}\right)=R_{m,GBR}^{grant}\left(\text{bps}\right)*10^{-3}\left(\text{s}/\text{TTI}\right). \qquad (1)$$

For example, the data rate $R_{m,GBR}^{grant}$ =1M bps translates to $B_{m,GBR}^{grant}$ =1K bit/TTI. In average, eNB should allocate RBs to $UE_m$ such that $B_{m,GBR}^{grant}$ bits can be transmitted during one TTI so as to meet the guaranteed data rate $R_{m,GBR}^{grant}$.

Let $\bar{B}_m\left(n\right)$ denote the average number of bits that has been sent after the *n*-th TTI. Then, according to the definition of EWMA, we have

$$\bar{B}_m\left(n\right)=\left(1-\alpha\right)\bar{B}_m\left(n-1\right)+\alpha B_m\left(n\right), \qquad (2)$$

where $B_m\left(n\right)$ denotes the number of bits transmitted during the *n*-th TTI and $\alpha\in\left(0,1\right)$ is the weighting factor. By combining (1) and (2), we can obtain the EWMA data rate that $UE_m$ has achieved up to the *n*-th TTI. $\bar{B}_m\left(n\right)$ should not be smaller than $B_{m,GBR}^{grant}$ if eNB wishes to keep the EWMA data rate of $UE_m$ no less than $R_{m,GBR}^{grant}$. As a result, in the *n*-th TTI, the minimum number of bits that $UE_m$ is entitled to transmit is expressed as

$$\frac{B_{m,GBR}^{grant}-\left(1-\alpha\right)\bar{B}_m\left(n-1\right)}{\alpha}. \qquad (3)$$

Eq. (3) means that eNB should allocate RBs to support the transmission of this number of bits to ensure the QoS of $UE_m$.

## III. Previous Work – AAG Scheduling Scheme

In section IV we will present a novel scheme AAG-2 that not only ensures QoS for GBR bearers but also offers a certain level of service for non-GBR bearers at the same time. AAG-2 builds on AAG which was presented in [14, 15]. For sake of completeness, we briefly describe the operation of AAG in the follows.

The key idea of AAG is to allocate RBs to UEs according to priority metrics and the guaranteed bit rates that are granted by RAC. AAG uses EWMA to calculate the average number of bits that have been sent in a TTI.

Based on EWMA, (3) describes the number of bits that $UE_m$ is allowed to transmit. For our purpose, it is refined as

$$B_m^{plan}\left(n\right)=\min\left\{\max\left[\frac{B_{m,GBR}^{grant}-\left(1-\alpha\right)\bar{B}_m\left(n-1\right)}{\alpha},0\right],L_m\left(n-1\right)\right\}. \qquad (4)$$

The superscript “plan” means that eNB should plan to allocate RBs for the transmission of the number of bits. The newly introduced term $L_m\left(n-1\right)$ represents the total queue length of $UE_m$. This term, which is obtained through BSR, can prevent the waste of RBs if there are not so many bits pending in the buffer. Although a UE will not update BSR in every TTI, eNB can guess the up-to-date value by subtracting the number of bits that has been scheduled for transmission. The parameter is then updated with the actual value whenever eNB receives a new BSR.

In order to ensure the throughput of every UE, AAG allocates RBs to UEs according to the descending order of the priority metric defined as

$$P_m\left(n\right)=\frac{B_{m,GBR}^{grant}-\bar{B}_m\left(n-1\right)}{B_{m,GBR}^{grant}}. \qquad (5)$$

The metric reflects the ratio of the shortage of transmitted data rate relative to the granted data rate.

AAG updates (4) and (5) for every UE in each TTI. When allocating RBs for a UE, AAG first selects a free RB with highest channel quality. If this RB alone cannot satisfy the requirement indicated by (4), AAG searches for and includes free adjacent RBs on both sides. AAG scheme takes into account the constraints that RBs allocated to a UE must be contiguous and use the lowest modulation-coding scheme (MCS) available. The latter constraint is also known as robust MCS constraint. Simulation results showed that AAG can obtain better packet delay, system throughput, and UE satisfaction in comparison with RME and RR schemes in simulated scenarios [15].

## IV. Proposed Scheme: AAG-2

### A. *Motivation*

Apart from GBR bearers, a UE can also establish non-GBR bearers. Although the quality of non-GBR traffic is not guaranteed, eNB will allocate RBs for the UE as long as the QoS of GBR traffic is satisfied. However, AAG was devised to schedule for GBR traffic only. Thus, it is necessary to design a scheme capable of scheduling for both GBR and non-GBR bearers.

Consider the general scenario where some UEs establish both uplink GBR and non-GBR bearers. To allocate RBs in a TTI, one possibility is that eNB allocates RBs to UEs to meet the QoS requirement of all the GBR bearers and then distributes the rest of RBs to UEs which need to transmit non-GBR traffic. With this procedure, however, it is quite possible that the RBs allocated to a UE are not contiguous. This is not allowed for LTE uplink transmission.

To tackle this issue, we present a modified dual-stage AAG scheme, called AAG-2, which is capable of (1) satisfying the requirement of GBR bearers, (2) making efficient use of remaining RBs to transmit non-GBR traffic, and (3) meeting the contiguous constraint on RB allocation.

## B. Operation of AAG-2

For $\mathrm{UE}_m$, we define the total granted bit rate of all the admitted GBR bearers as

$$R_{m,GBR}^{grant} = \sum_{j=1}^{h} R_{m,j,GBR}^{grant} , \tag{6}$$

where $R_{m,j,GBR}^{grant}$, in bps, is the granted bit rate of the $j$-th GBR bearer and the superscript "grant" means that the data rate is granted by RAC. Besides, we denote the total data rate that is requested by both GBR and AMBR bearers as

$$R_{m,MIX}^{grant} = R_{m,GBR}^{grant} + R_{m,AMBR}^{grant} , \tag{7}$$

where $R_{m,AMBR}^{grant}$ corresponds to the demand of all the non-GBR bearers of $\mathrm{UE}_m$.

Similar to the manipulation in (1), each data rate is converted to the number of bits to be sent in a TTI as follows:

$$B_{m,GBR}^{grant} = R_{m,GBR}^{grant} * 10^{-3} \tag{8}$$

$$B_{m,MIX}^{grant} = R_{m,MIX}^{grant} * 10^{-3}. \tag{9}$$

Recall that eNB has to ensure the QoS of GBR bearers while making good use of the remaining RBs so as to achieve high resource utilization. As a result, if the system is lightly loaded, eNB tries to allocate $B_{m,MIX}^{grant}$ bits in average for every $\mathrm{UE}_m$ so as to meet all of the data rate expectations. On the contrary, if the system is heavily loaded, eNB should firstly satisfy the $B_{m,GBR}^{grant}$ for GBR bearers and then try to transmit non-GBR traffic.

The operating principle of AAG-2 is trying to allocate RBs to meet $B_{m,MIX}^{grant}$ for the UEs whose priority metrics are in the top $x$% of all UEs, and then allocate the remaining RBs to meet the requirement of $B_{m,GBR}^{grant}$ for the other UEs. The method for allocating RBs in each TTI is based on the AAG scheme described in section III.

It is not easy to determine a suitable value for $x$. If it is too small, RBs cannot be efficiently used and less non-GBR traffic is served. On the contrary, with too large $x$, some UEs may not get enough RBs to guarantee the QoS of GBR bearers. As a result, AAG-2 dynamically adjusts the value of $x$ as described in the following.

Define

$$S_{GBR}(n) = \frac{1}{K} \sum_{m=1}^{K} s_m(n-1) \tag{10}$$

as the average satisfaction ratio associated with the GBR traffic of all $K$ UEs when eNB is scheduling for the $n$-th TTI. The $s_m(n-1)$ is the GBR queue length of $\mathrm{UE}_m$ that is predicted by eNB. Initially $x$ is set to zero and then dynamically adjusted as follows:

$$x = \begin{cases} \max\left(0, \min\left(100, x + \Delta x\right)\right) & \text{if} \quad S_{GBR}(n) > S_{th} \\ \max\left(0, \min\left(100, x - \Delta x\right)\right) & \text{if} \quad S_{GBR}(n) \le S_{th}, \end{cases} \tag{11}$$

where $S_{th}$ denotes the threshold of average satisfaction ratio and $\Delta x$ is the step size for adjustment.

The flowchart of AAG-2 is illustrated in Figure 1. As we can see, the first stage tries to meet the requirements of both GBR and non-GBR traffic for the UEs ranked in the top $x$% of priority values, which corresponds to $\lceil K \cdot x\% \rceil$ UEs. The second stage tries to ensure the GBR requirement for the remaining UEs.

Because AAG-2 embeds AAG as a module for allocating RBs, it inherits the characteristics of AAG such as:

- UE with the highest shortage ratio is served first, rather than the one with best channel quality. As long as a GBR bearer of a UE is admitted, it will be granted proper right to transmit, even if its channel quality is poor. Of course, the system can drop a connection if the channel efficiency is not satisfied.
- BSR is taken into consideration to avoid wasting RBs.

# V. PERFORMANCE EVALUATION

## A. Simulation Setup

We evaluate the performance as a function of load by varying the number of UEs admitted by eNB. Table I summa-

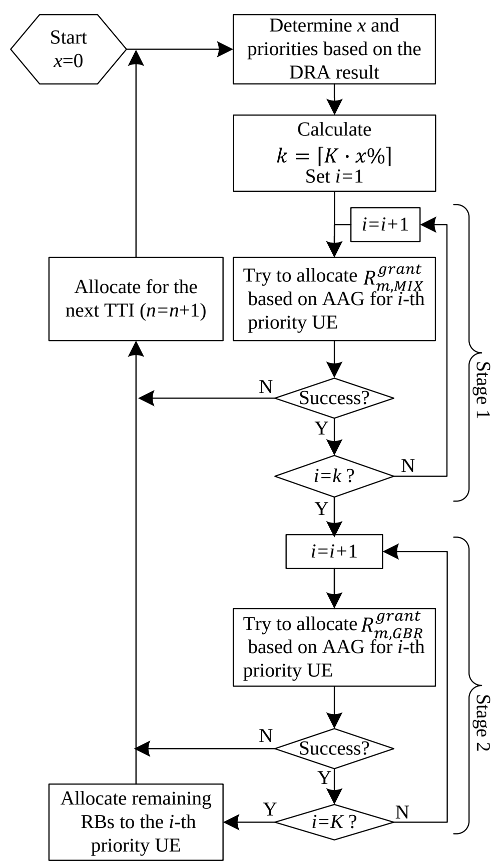


Figure 1 Flowchart of AAG-2

rizes the simulation parameters. Truncated Pareto distribution is assumed for the Near Real Time Video (NRTV) [16] with a long term mean rate of 1.663 Mbps, as shown in Table II. However, if the data rate is averaged by EWMA, the value would fluctuate over time. For example, when there is a large burst of data during the transmission, the EWMA value at that moment would be higher than the long term mean rate. If we set the $R_{m,GBR}^{grant}$ in (6) as the long term mean rate, part of the burst would suffer long delay. As a protective measure, we set $R_{m,GBR}^{grant}$ as 1.993 Mbps, which is about 20% higher than the long term mean rate. Accordingly we obtain $R_{m,MIX}^{grant}$ =3.5 Mbps by (7).

The functionality of AAG in an environment with random channel quality has been verified in [14, 15]. In order to focus our attention on how the system capacity is shared among GBR and non-GBR traffic in AAG-2, the channel quality of all RBs is chosen to be uniform, with an MCS index of 28. This corresponds to a transport block size (TBS) index of 26, which in turn means that 712 bits can be transmitted by a single RB. As indicated in the TBS table [17] which shows TBS as a function of MCS and number of RBs, each RB can transmit slightly more bits in average when several RBs are combined.

TABLE I. SIMULATION PARAMETERS

| | |
|---|---|
| System bandwidth | 20MHz |
| Number of RBs | 100 (all RBs can be used for user data) |
| Numbers of UEs in the cell | 10, 20, 30, 40 |
| Simulation duration | 60 seconds |
| $\Delta x$ | 1, 5, 10, 15, 20 |
| Channel quality | MCS index = 28 (TBS index = 26) for all RBs |
| Bearers of each UE | one GBR bearer; one non-GBR bearer |
| GBR traffic pattern | Near Real Time Video (NRTV), Truncated Pareto distribution, Long term average data rate: 1.663Mbps (see Table II for more details) |
| Non-GBR traffic pattern | real e-mail traffic, Long term average data rate: 1.507Mbps |
| $\alpha$ of EWMA | 0.01 |
| Threshold of average satisfaction ratio $S_{th}$ | 95% |
| Sending Sequence in UEs | GBR traffic is sent first, followed by non-GBR traffic |
| BSR update | eNB can get the update BSRs whenever it want to schedule for the next 10 TTI |
| MAC buffer sizes for GBR and non-GBR bearers | Infinite buffer size without feedback flow control |

TABLE II. PARAMETERS OF NRTV TRAFFIC

| Inter-arrival time between the beginning of consecutive frames | Number of packets (slices) in a frame | Packet (slice) size | Inter-arrival time between packets (slices) in a frame |
|---|---|---|---|
| 100ms (10fps) | 8 | Truncated Pareto: min = 512 bytes max = 1024 bytes Shape parameter = 1.2 | Truncated Pareto: min = 2.5 ms max = 12.5 ms Shape parameter = 1.2 |

## B. Numerical Results

Figure 2 shows the respective total throughput of GBR bearers and non-GBR bearers corresponding to different system loads and $\Delta x$ . Figure 3 shows the average throughput of each bearer. The utilization rate of RBs corresponding to various system loads is shown in Figure 4. If there are 50 UEs, the system would be overloaded with GBR traffic. This is not allowed if RAC works properly. Consequently, we set the heaviest load at 40 UEs.

Figure 2 and Figure 3 reveal that the throughput of the GBR traffic is always the same as the GBR input traffic. That is to say, the throughput of GBR traffic is guaranteed. Besides, when eNB is burdened with more GBR traffic, it always allocates sufficient RBs for the GBR traffic whereas fewer RBs are devoted to non-GBR traffic.

A packet waiting in the MAC buffer may be divided into several frames for transmission. As a result, the delay of a packet is usually defined as the duration from the time it arrives at the buffer until it is completely sent out. Figure 5 shows that the average packet delay of GBR packets is no more than 9.5 ms even when eNB is very heavily loaded with 40 UEs. As we can see in Figure 2, when the system is loaded with 30 or 40 UEs, the summation of the GBR and non-GBR throughputs saturates at about 70 Mbps, which is less than the traffic from both GBR and non-GBR bearers. As a result, only the throughput of GBR bearers can be ensured at the expense of non-GBR traffic as shown in Figure 3. Therefore, the packet delay of non-GBR traffic increases dramatically as shown in Figure 6. Indeed, the delay of non-GBR traffic would continue to increase if we extend the simulation time because a greater number of non-GBR packets would be cumulated in the buffer.

With regard to $\Delta x$ , Figure 2 and Figure 3 reveal that it has

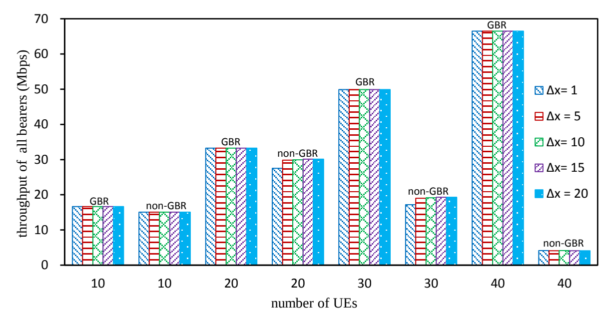


Figure 2 Throughput of all bearers vs. number of UEs

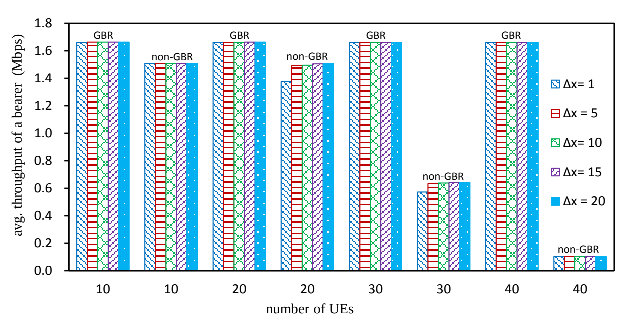


Figure 3 Average throughput of a bearer v.s. number of UEs

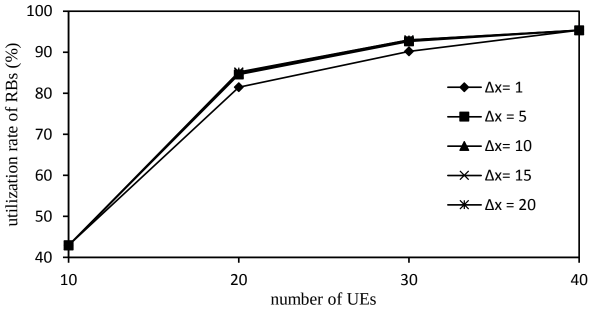


Figure 4 Utilization rate of RBs v.s. number of UEs

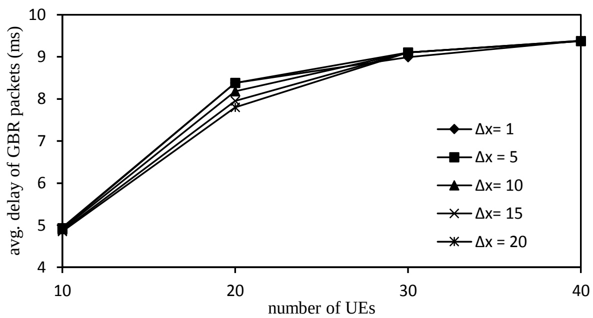


Figure 5 Average delay of GBR packets vs. number of UEs

no impact on the throughput of non-GBR traffic when eNB is very lightly loaded with only 10 UEs because all of the traffic can be easily scheduled. It also has no impact when eNB is very heavily loaded with 40 UEs because in that case eNB seldom allocates RBs for non-GBR traffic. However, when the load is moderate, e.g., with 20 or 30 UEs in the system, by using larger values of $\Delta x$, eNB can improve RB utilization and serve more non-GBR traffic.

For a real system, in order for the RAC to decide whether to accept a new bearer or to upgrade service for existing bearers, real time observation of system performance is very important.

## VI. Benefits from Using EWMA

Based on EWMA, we express in (3) the number of bits that a UE is allowed to transmit GBR traffic. The corresponding priority metric is expressed in (5). As a result, the longer duration a UE keeps silent, the larger burst it can transmit with higher priority. In view of this, there are at least two merits to adopting EWMA for data rate measuring.

- It can flexibly accommodate both variable bit rate (VBR) and constant bit rate (CBR) traffic:

  For a VBR bearer, such as video or voice bearer, when there is no traffic pending for transmission, it can accumulate the transmission right and then transmit in a burst at a later moment. If a UE always has some data to transmit, it can transmit in a style similar to a CBR session.

- It can improve the utilization and capacity of eNB:

  The examples of GBR traffic listed in QCI table are conversational voice, video, real time gaming and non-conversational video. All of them are VBR in nature. It is not economical to allocate resources throughout an entire session in accordance with the peak data rate because a lot of RBs would be wasted. Take downlink transmission as examples, some schemes are proposed to statistically multiplex H.264/AVC video traffic so that more sessions can be accommodated [18-22]. The performance is usually measured in statistical multiplexing gain (SMG), which is defined as ($1–A/B$), where $A$ is the aggregate bandwidth required to satisfy a given QoS requirement for all VBR streams and $B$ is the sum of peak rates of all individual streams. In our view, it is also important to improve SMG for uplink transmission. With the help of EWMA, we don't need to set GBR as high as the peak data rate of a VBR traffic. It is more reasonable to set it as a value between the peak data rate and the long term mean rate. For example, we set it as 20% higher than the long term mean rate for the simulation. As we can see in Figure 2, when 40 UEs are present, the GBR NRTV bearers result in a throughput of about 66.5 Mbps with average packet delay as short as 9.4 ms. Besides, it still offers 4.1 Mbps throughput for non-GBR traffic. The total throughput is about 70.6 Mbps. We did not measure the peak data rate of the NRTV traffic because there is no commonly accepted definition. However, if the peak data rate of the NRTV traffic is 2.5 Mbps and eNB wants to provide each GBR bearer with 2.5 Mbps, then only about 28 (i.e., 70.6/2.5) bearers can be accommodated. It is quite inefficiency as compared with accommodating 40 UEs. This is the reason we claim that the utilization and capacity of eNB can be improved by employing EWMA. How to set proper values of GBR for a VBR session and how to evaluate the SMG of uplink scheduling will be explored in future studies.

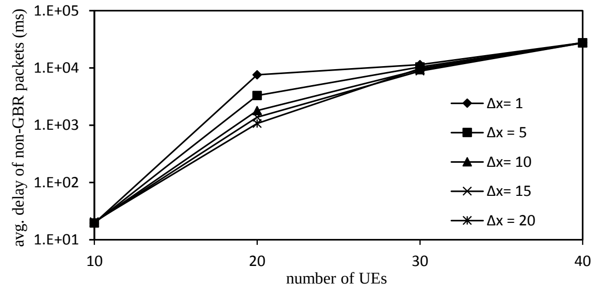


Figure 6 Average delay of non-GBR packets v.s. number of UEs

## VII. Conclusion and Discussion

In this paper, we use EWMA to define the measurement of GBR to facilitate the operation of DRA. Then, in order to simultaneously schedule both GBR and non-GBR bearers, we present the AAG-2 scheme, which employs AAG as the scheduling module that aims at ensuring the throughput of GBR sessions regardless of channel quality. Simulation results show that the proposed scheme can always provide GBR sessions with sufficient throughput and short delay, while at the same time achieving high resource utilization by efficiently providing RBs to non-GBR sessions.

Apart from GBR, another data rate constraint associated with a GBR bearer is MBR as mentioned in section I. Any traffic exceeding the specified MBR will be discarded through rate policing. In the future we would like to define MBR based on EWMA to facilitate rate policing.


## ACKNOWLEDGMENTS

The authors would like to thank Hua-Chun Kao, the master's degree student of National Ilan University, who helped develop the first version of the simulation program.


## REFERENCES


[1] L. Ruiz de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen, "Channel-aware scheduling algorithms for SC-FDMA in LTE uplink," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2008.

[2] F. Liu, X. She, L. Chen, and H. Otsuka, "Improved Recursive Maximum Expansion Scheduling Algorithms for Uplink Single Carrier FDMA System," in *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, 2010, pp. 1-5.

[3] J.-H. Noh and S.-J. Oh, "Distributed SC-FDMA Resource Allocation Algorithm Based on the Hungarian Method," in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, 2009, pp. 1-5.

[4] O. Nwamadi, X. Zhu, and A. Nandi, "Dynamic subcarrier allocation for single carrier - FDMA systems," in *16th European Signal Processing Conference*, 2008, pp. 1-5.

[5] O. Delgado and B. Jaumard, "Scheduling and resource allocation for multiclass services in LTE uplink systems," in *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2010, pp. 355-360.

[6] O. Iosif and I. Banica, "LTE uplink analysis using two packet scheduling models," in *19th Telecommunications Forum (TELFOR)*, 2011, pp. 394-397.

[7] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional Fair Frequency-Domain Packet Scheduling for 3GPP LTE Uplink," in *IEEE INFOCOM*, 2009, pp. 2611-2615.

[8] F. D. Calabrese, P. H. Michaelsen, C. Rosa, M. Anas, C. U. Castellanos, D. L. Villa, *et al.*, "Search-Tree Based Uplink Channel Aware Packet Scheduling for UTRAN LTE," in *IEEE Vehicular Technology Conference, VTC Spring 2008*, pp. 1949-1953.

[9] F. Kaddour, E. Vivier, L. Mroueh, M. Pischella, and P. Martins, "Green Opportunistic and Efficient Resource Block Allocation Algorithm for LTE Uplink Networks," *IEEE Transactions on Vehicular Technology, in press*.

[10] M. Kalil, A. Shami, and A. Al-Dweik, "QoS-Aware Power-Efficient Scheduler for LTE Uplink," *IEEE Transactions on Mobile Computing, in press,* 2014.

[11] H. Safa, W. El-Hajj, and K. Tohme, "A QoS-Aware Uplink Scheduling Paradigm for LTE Networks," in *IEEE 27th International Conference on Advanced Information Networking and Applications, AINA*, 2013, pp. 1097-1104.

[12] 3GPP, "TS 23.401 v. 13.2.0," 2015.

[13] 3GPP, "TS 36.300 v. 12.4.0," 2015.

[14] F.-C. Kuo, H.-C. Wang, C.-C. Tseng, K.-C. Ting, and P.-E. Liu, "Call admission control based resource allocation scheme for LTE uplink," in *15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2012, pp. 554-558.

[15] F.-C. Kuo, H.-C. Wang, K.-C. Ting, C.-C. Tseng, and P.-E. Liu, "Robust LTE uplink scheduling based on call admission control," in *International Conference on ICT Convergence (ICTC)*, 2012, pp. 548-552.

[16] 3GPP, "TR 25.892 v. 2.0.0," 2004.

[17] 3GPP, "TS 36.211 v. 12.5.0," 2015.

[18] P. V. Pahalawatta, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Content-Aware Resource Allocation for Scalable Video Transmission to Multiple Users Over a Wireless Network," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2007, pp. 853-856.

[19] G. Liebl, M. Kalman, and B. Girod, "Deadline-aware scheduling for wireless video streaming," in *IEEE International Conference on Multimedia and Expo, ICME*, 2005, pp. 1-4.

[20] M. Andrews, "Probabilistic end-to-end delay bounds for earliest deadline first scheduling," in *19th IEEE INFOCOM, Joint Conference of the IEEE Computer and Communications Societies*, 2000, pp. 603-612.

[21] C. Lou and L. Qiu, "QoS-Aware Scheduling and Resource Allocation for Video Streams in e-MBMS Towards LTE-A System," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1-5.

[22] Z.-L. Zhang, J. Kurose, J. D. Salehi, and D. Towsley, "Smoothing, statistical multiplexing, and call admission control for stored video," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1148-1166, 1997.