Downlink Packet Scheduling for Variable Bitrate Traffic in LTE Networks

Joel Makara, Neco Ventura
Department of Electrical Engineering
University of Cape Town, Private Bag X3, Rondebosch 7700
Tel: +27 21 650 2813, Fax: +27 650 3465
email: {jmakara, neco}@crg.ee.uct,ac,za

Abstract – In recent years, there has been a sharp increase in the demand for real-time multimedia applications from mobile subscribers. Voice and video are the most common forms of network traffic used in these applications and to optimise space-quality ratios, variable bitrate coding techniques are used to encode them. Even with the advent of high speed access network technologies such as the long term evolution (LTE), there is a need for efficient management of radio resources to ensure that the quality of service (QoS) requirements for these applications is met. We present an optimised scheduling approach that exploits multiuser diversity by considering each user's instantaneous downlink conditions and QoS information when distributing resources. Simulation results show that our proposed scheme outperforms other dynamic resource allocation approaches with regard to guaranteeing quality of service and overall system throughput.

Index Terms— LTE, VBR, Quality of Service, Real-time traffic

I. INTRODUCTION

Over recent years, the telecommunication industry has seen a sharp rise in the demands for services that are increasingly intensive on network resources from mobile subscribers. Global mobile network traffic is expected to increase seven-fold by 2014 [1]. To address this growing need for network resources from mobile users, standardisation bodies are continuously designing mobile broadband networks that offer improved performance over the available frequency spectrum that is very much scarce and highly competed for [2,3].

The Third Generation Partnership Project (3GPP) developed the Long Term Evolution (LTE) standard as a successor to previous commercially successful cellular standards under its development. 3GPP cellular standards are the world's most widely used with more than 3.5 billion subscribers who account for 89.5% of the global mobile population [3]. The LTE radio access technology is optimised for packet-switched transmission and offers downlink data rates of up to 300Mbps at 20MHz of bandwidth. The standard uses orthogonal frequency division multiple access (OFDMA) as its channel access scheme due to immunity from frequency selective fading and robustness from inter-symbol interference while also offering high data rates.

Trends in mobile data traffic also show that multimedia applications such as online gaming, web television and video calling account for the largest majority of global data traffic [1]. This is illustrated in Figure 1 below;

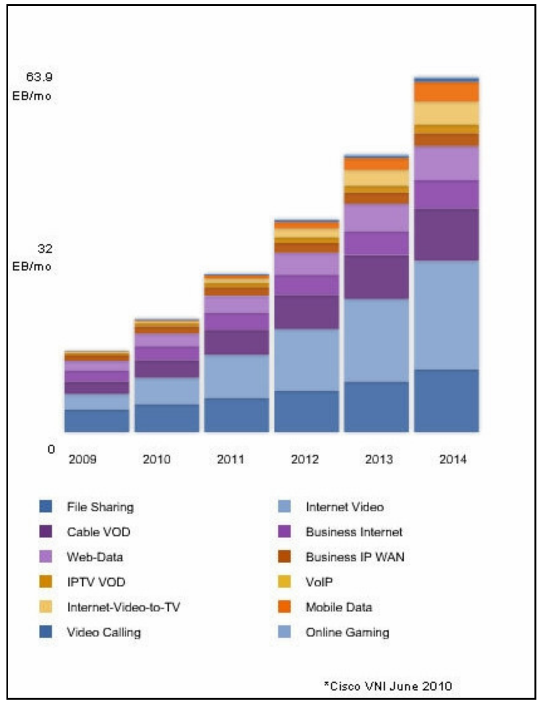

Fig1: Projections on Global Data Traffic

These multimedia applications make use of real time video and voice content which is usually encoded using variable bitrate (VBR) techniques to optimise space-to-quality ratios. With VBR encoding, simple parts of either voice or video data are sampled at lower rates while complex parts are sampled at higher rates to achieve high content quality where it matters. This produces a smaller output file compared to the product of a constant bitrate encoder. To ensure that the quality of service requirements for real time VBR data are met, packets must be received before the application's playback deadline expires and packet loss must be minimised by keeping it below the application's required threshold. This is achieved through efficient scheduling mechanisms.

Along with guaranteeing that application QoS requirements are met, an efficient scheduler must ensure that available bandwidth is used efficiently through increasing overall system throughput and ensuring that all users are allocated radio resources in fairness. In the LTE's network architecture, where the number of nodes is minimised to reduce capital and operating expenditure, radio resource management functions such as scheduling are performed by the enhanced node B (eNodeB) which is the only node interfacing the user and the core network.

The remainder of this paper is organised as follows: Section II discusses some related work in this field by looking at real time scheduling in the LTE and similar OFDMA systems. Section III gives an overview of our proposed system model while the fourth section discusses the simulation and the results obtained from it. Finally, Section V concludes the paper.

II. REAL TIME SCHEDULING IN OFDMA SYSTEMS

There is a substantial amount of research in the area of scheduling in LTE and similar OFDMA systems to support real time applications. Two categories of scheduling algorithms have been presented in literature; these are static and dynamic resource allocation schemes.

Static scheduling algorithms allocate either all the subcarriers to one user at a scheduling instant, or dedicate a sub-band to a user for the entire scheduling duration. These algorithms are more suited for traditional multiple access schemes, such as Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), as they do not exploit the time-varying link qualities for each user on each of the many OFDMA sub-channels [5]. Although static algorithms have low computational complexity, they lack in exploiting multiuser diversity and run the risk of wasting system resources over transmission on poor channels.

Dynamic algorithms take each user's instantaneous channel conditions into consideration when making a scheduling decision; this is also known as multiuser diversity. The LTE's specification facilitates channel quality indicators (CQIs) that are fed back from a user terminal to the eNodeB reporting sub-channel link quality. In [5] a delay-prioritized approach is presented to ensure that quality of service is guaranteed for delay sensitive applications. The three-staged approach first determines the user whose head of line packet has the shortest time before approaching its delay threshold, then further assigns that user the best resource block from a list of available resource blocks based on the user's CQI values. In the third step, the list of available RBs is then updated by removing the assigned block.

Another dynamic approach towards providing applications with their quality of service requirements is studied in [9]. This approach employs the popular modified largest weighted deadline first (M-LWDF) scheduling algorithm. The scheduler, at all times aims to maintain that the probability of a head-of-line packet missing its playback deadline, is kept below a predetermined threshold δ. Packet's that are closest to their deadline are given highest priority and are transmitted on the best resource block available.

The dynamic scheduling algorithms presented above offer the advantages of guaranteeing application quality of service and improving spectral efficiency by scheduling each user on the resource block where the user experiences the best channel quality. They however do not present practical solutions as only real time traffic streams are considered whereas in reality network traffic is characterised by a diverse mixture of both real-time and non-real-time traffic. The schemes also fall short in ensuring that all users are served in fairness. Users that are farthest from the eNodeB may have weak CQI values in all the available subcarriers causing these users to be starved of resources.

III. SYSTEM MODEL

We propose an approach towards the management of resources in the LTE downlink that fully exploits multiuser diversity. Our proposed scheme, the VBR-optimised Scheduler (VOS), is related to the work covered in [4] - [9] but differs in that it makes provision for non-real-time traffic flows. This is achieved by introducing a reservation of bandwidth for both real-time VBR and non-real-time streams based on the LTE operator's network usage statistics. The goals of our scheme are to ensure that quality of service is guaranteed for real-time flows, and that all users are scheduled in fairness, while resource utilisation is optimised through increased overall system throughput. In this context, application QoS is defined by the delay bound before which a packet should be delivered to the receiver [4].

The basic unit for resource allocation in LTE networks is the resource block which spans 180 KHz in the frequency domain and a transmission time interval (TTI) of 0.5ms. Our approach divides the scheduling problem into two processes; resource block allocation and resource block assignment. The allocation process determines the number of blocks the eNodeB will use to transmit data to the user, while the assignment procedure maps the physical blocks to each user.

A. *Resource Block Allocation*

An LTE system with L resource blocks per TTI is assumed where K mobile users are serviced by an eNodeB. Of these users, N are each running an application that has an active connection to a server on a packet data network connected to the eNodeB. The following parameters are defined for our system:

R_i: The average data rate achieved by the i^{th} user at a time t when a scheduling decision is to be made.

r_i: The minimum required data rate for the i^{th} user at t.

τ_i: The playback delay threshold for the i^{th} user; maximum allowable time before the user's head of line packet in its queue can be delivered.

$L_{VBR}(i)$: number of resource blocks allocated to the i^{th} real-time variable bit rate user in one TTI.

$L_{nVBR}(i)$: number of resource blocks allocated to a non-real-time user.

E_i: the effective data rate of the i^{th} user computed from the CQI values of all the subcarriers (as if the user was allocated the whole of the system's available band).

The number of resource blocks allocated to real time VBR users is then is determined as:

$$L_{VBR}(i) = \left(\frac{\propto}{\propto + \beta}\right) \left\lfloor \frac{\frac{1}{\tau_i}}{\sum_{n=1}^{N}\left(\frac{1}{\tau_n}\right)} \cdot \frac{\left(\frac{r_i}{R_i}\right)}{\sum_{n=1}^{N}\left(\frac{r_i}{R_i}\right)} \right\rfloor L \qquad (1)$$

The parameters α and β are constants assigned by the network's operator. They are selected such that the ratio α:β represents the amount of real-time to non-real-time traffic that generally flows through the network. These values may be based on observations of network usage or market research.

For the determination of the number of resource blocks allocated to non-real-time users L_{nVBR}, a modification on the algorithm presented in [6] is employed. For non real-time traffic, the following rule is applied in the determination of resource blocks allocated.

$$L_{nVBR}(j) = \left(\frac{\beta}{\propto + \beta}\right) \left\lfloor \frac{E_j}{\sum_{n=1}^{N}(E_n)} \cdot \frac{\left(\frac{r_j}{R_j}\right)}{\sum_{m=1}^{N}\left(\frac{r_j}{R_j}\right)} \right\rfloor L \qquad (2)$$

Both the rules in (1) and (2) have components that are rounded down and this may leave some resource blocks unallocated. To improve the system's overall throughput, the remaining blocks are allocated to the user that has the highest CQI value in each of the blocks.

B. *Resource Block Assignment*

The assignment of resource blocks is performed such that the highest overall system throughput is achieved. Of the available blocks, each user is assigned those in which it experiences the best channel conditions.

The assignment procedure works by initially constructing an *L x K* matrix containing values representing the relative channel conditions of each user on each sub-channel as illustrated below:

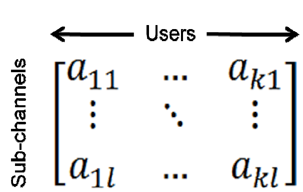

Fig 2: Assignment matrix of relative user channel quality per sub-channel

The values a_{kl} for each user on each sub-channel are calculated based on the comparison of CQI value of each user on each channel, and a comparison of each channel's CQI for each user as follows:

$$a_{lk} = \frac{CQI_k^{(l)}}{\frac{1}{L}\sum_{i=1}^{L} CQI_k^{(i)}} \cdot \frac{CQI_k^{(l)}}{\frac{1}{K}\sum_{j=1}^{K} CQI_j^{(l)}} \qquad (3)$$

The assignment matrix is sorted in descending order of the value a_{kl} implying that the user and sub-channel combination with the highest value occupy position (1,1). The sub-channel in this combination is assigned to the corresponding user then removed from the assignment matrix. This process is done iteratively until all resource blocks are assigned. The matrix is then repopulated in the next transmission time interval after the allocation rule has been applied.

IV. PERFORMANCE EVALUATION

In this section, we present simulation details for our proposed VOS scheme, the results obtained from simulation and their discussion.

A. *Simulation Setup*

The popular discrete event simulator, Network Simulator 2 was used to simulate our LTE system. The simulator was used with the Enhanced UMTS Radio Access Network Extension (EURANE) patch. The eNodeB was emulated as a hybrid of the radio network controller and the base station nodes. The three algorithms to be compared were incorporated into the NS environment by modifying the queuing procedures in EURANE's UMTS module.

Table 1 summarises the parameters configured in the simulations:

Table1: Summary of Simulation Parameters

Carrier Frequency	2.1 GHz
Channel Bandwidth	5 MHz
Propagation Model	Two Ray Ground Reflection
Scheduling Interval	1 ms
VBR Packet Size	50 bytes
VBR Average Rate	70 Kbps
VBR Peak Rate	120 Kbps
Non-VBR Packet Size	50 bytes
Non-VBR Average Rate	70 Kbps

From the above specified parameters, the resulting traffic model for one of the VBR users is illustrated in Figure 3.

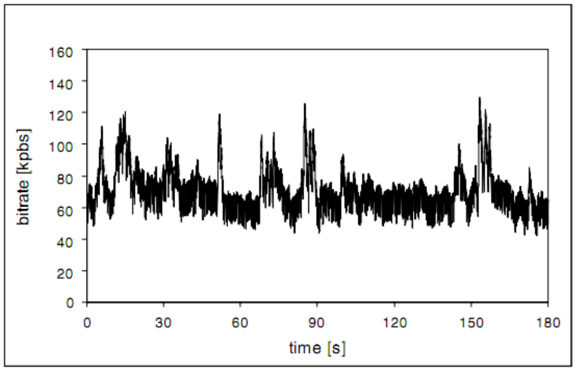

Fig 3: Sample VBR Source

B. *Performance Metrics*

The performance of our proposed VOS scheduling algorithm is evaluated based on the degree of guaranteeing quality of service to real-time users and overall system throughput. The overall system throughput is the total amount of data transmitted to all users throughout the simulation duration. The evaluation of application quality of service is determined by the average delay experienced by each real time packet from the time it was queued up in the eNodeB to the time it was delivered to its user.

C. *Results*

In this section we present the simulation results of our proposed solution. Its performance is compared to the modified largest weighted deadline first algorithm presented in [4] and the delay prioritised scheduling algorithm in [5].

Figure 4 illustrates the overall throughput that the system achieves under different scheduling algorithms. We notice that the throughput increases as the number of users in the system increasing. This is owing to the fact that more data is transmitted across the network. There will however be a threshold to the throughput as the eNodeB's transmission capacity is reached. This will result in packet loss due to the buffer overflow in the eNodeB.

The most significant observation is that our proposed VBR-optimised scheduling algorithm achieves greater throughput values than the other scheduling algorithms. This is attributed to the fact that the algorithm always selects the combination of the user and sub-carrier that best utilise the available radio resources.

Figure 5 shows the mean system delay of the VOS, DPS and the M-LWDF algorithms with respect to the number of users in the network. We notice that for cases where there are 70 or fewer mobile users in the system, the algorithms perform equally well; the mean delay is kept to below 2ms. The average delay increases significantly as the number of users increases but the VOS scheme performs better than the two other algorithms. This improved performance is due to the VOS prioritising packets whose playback deadline is approaching.

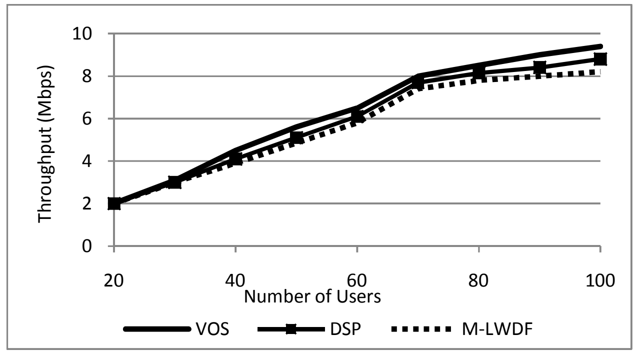

Figure 4: Throughput vs. Number of Users

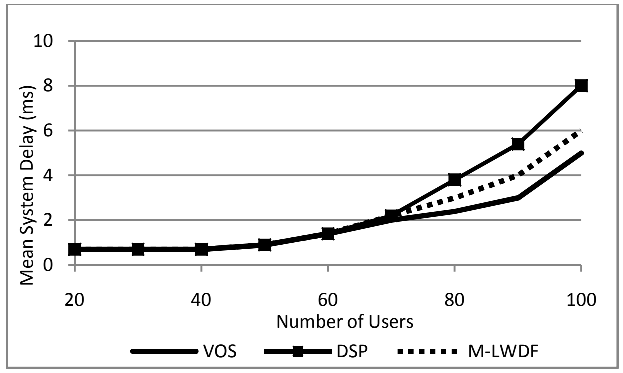

Figure 5: Mean System Delay vs Number of Users

V. CONCLUSION

In this paper, we have presented a scheme that is optimised for offering improved quality of service for a diverse mix of traffic including real-time VBR traffic in the downlink of LTE networks. The proposed scheme aims to satisfy application quality of service requirements and improve overall system throughput by exploiting multiuser diversity. This is achieved by minimising the average delay experienced by real-time packets in the network and scheduling users in the sub-channels where they experience the best link quality, implying higher data rates. Our proposed VBR-optimised scheme performs better than the well known modified largest weighted delay first algorithm and the delay prioritised scheme. The VOS achieved higher overall system throughput and lower average delay for real-time packets transmitted through the network.

REFERENCES

[1] "Cisco Visual Networking Index Forecast Widget," http://www.ciscovni.com/vni_forecast/advanced.html, April 2011

[2] E. Dahlman et al., 3G Evolution: HSPA and LTE for Mobile Broadband, Academic Press, 2nd ed., Oxford, England, 2007.

[3] H. Holma and A. Toskala, LTE for UMTS: OFDMA and SC-FDMA Based Radio Access, John Wiley & Sons, 1st ed., New York, USA, 2009.

[4] A. Khattab, and K. Elsayed, "Opportunistic scheduling of delay sensitive traffic in OFDMA-based wireless networks," *International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pp. 278-288, June 2006.

[5] K. Sandrasegaran, H. Mohd Ramli, & R. Basukala, "Delay-Prioritized Scheduling (DPS) for Real Time Traffic in 3GPP LTE System," *IEEE Wireless Communications and Networking Conference,* pp.1-6, April 2010.

[6] M. Assaad and A. Mourad, "New Frequency-Time Scheduling Algorithms for 3GPP-LTE-like OFDMA Air Interface in the Downlink", *Vehicular Technology Conference.* pp. 1964-1969, May 2008.

[7] B. Sadiq, R. Madan, and A. Sampath, "Downlink Scheduling for Multiclass Traffic in LTE," *EURASIP Journal on Wireless Communications and Networking,* pp 9-9, March 2009.

[8] J. Gross, H. Karl, F. Fitzek, and A. Wolisz, "Comparison of heuristic and optimal subcarrier assignment algorithms," *Proc. of Intl. Conf. On Wireless Networks (ICWN),* pp. 249-255, June 2003.

[9] L. Xiantao, L. Guangyi, W. Ying; Z. Ping, "Downlink Packet Scheduling for Real-Time Traffic in Multi-User OFDMA System," *Vehicular Technology Conference,* pp. 1-5, September 2006

BIOGRAPHY

Joel Makara received his B. Eng. degree from the National University of Lesotho in 2009. He is presently studying towards an MSc degree at the University of Cape Town. His research interests are related to quality of service in next generation networks.