A Downlink Scheduling Approach for Balancing QoS in LTE Wireless Networks

Moustafa M. Nasralla and Maria G. Martini
Kingston University, London, UK
{M.Nasralla, M.Martini}@kingston.ac.uk

Abstract—**In this paper, we propose a strategy for resource allocation for different traffic classes at the Medium Access Control (MAC) layer of wireless systems based on Orthogonal Frequency Division Multiple Access (OFDMA), such as the recent Long-Term Evolution (LTE) wireless standard. In order to achieve inter-class fairness, we propose a modification of the Virtual Token Modified Largest Weighted Delay First (VT-M-LWDF) and Modified Largest Weighted Delay First (M-LWDF) rules. Through simulation, we show that the proposed scheduler introduces remarkable multi-objective improvement of the Quality of Service (QoS) performance parameters, i.e., Packet Loss Rate (PLR), average throughput, fairness index and system spectral efficiency, among different classes of traffic such as video, VoIP and best-effort.**

Index Terms—**Scheduling Algorithm, LTE, QoS, Real Time (RT) and non-Real Time (NRT) Services**

I. Introduction

In conjunction with the increasing growth of multimedia, Internet and NRT services, the LTE wireless standard has emerged to cope with these services efficiently. LTE has been introduced by the Third Generation Partnership Project (3GPP) as the next wireless technology after 3.5G (HSPA+) cellular networks. The system architecture of the 3GPP LTE system contains several base stations called LTE base station (eNodeB) where the packet scheduling process is performed along with other Radio Resource Management (RRM) tasks. LTE uses OFDMA in the downlink transmission mode and Single Carrier Frequency Division Multiple Access (SC-FDMA) in the uplink transmission mode. OFDMA extends the multi-carrier technology Orthogonal Frequency Division Multiplexing (OFDM) to provide a better and more flexible multiple access scheme. In other words, OFDMA splits the frequency band into multiple orthogonal sub-carriers. This helps improving the system capability to support high data rates, provide multi-user diversity, compact the Inter-Symbol-Interference (ISI) [1] [2] [3].

The QoS requirements must be satisfied by LTE systems by granting the admitted users the maximal balance of fairness and utilization of the service. There are two broad traffic classes: RT and NRT. These classes are categorised, in accordance with the 3GPP standardisation, into guaranteed bit rate (GBR) and non-guaranteed bit rate (non-GBR). Each radio bearer will be tagged with either GBR or non-GBR, depending on the QoS requirements of the flows they convey. Examples on the RT and NRT classes are video and Voice over IP (VoIP) services for the RT class and Constant Bit Rate (CBR) and best-effort services for the NRT class.

Due to the importance of the scheduling process at the MAC layer in LTE, various packet scheduling algorithms have been developed to support RT and NRT services, comprising the most commonly used Proportional Fair (PF), M-LWDF and Exponential Proportional Fair (EXP-PF) schedulers [4] [5] [6] [7]. With regard to the previously mentioned schedulers, each flow is assigned a priority value depending on specific measurements. Therefore, the bearer which carries the flow with the highest priority value will be scheduled first at the correspondent, Transmission Time Interval (TTI).

Because of the existence of different QoS requirements for RT and NRT classes, it is essential to design schedulers capable of dealing with diverse classes of traffic, as the contribution of designing a scheduler is left open by the 3GPP body to researchers and designers.

In this paper we propose a modification of the M-LWDF [6] and the VT-M-LWDF [8] scheduling algorithms.

The virtual token resource allocation algorithm, i.e., VT-M-LWDF, focuses on RT services whereas a minimum throughput has been guaranteed for NRT services. This scheduler is mainly based on the queue size of each RT flow‘s buffer, as this helps allocate most of the resources to video flows first, whereas NRT services are penalised.

The M-LWDF resource allocation algorithm is tailored to support multimedia services, considering that this scheduler outperforms other state-of-the-art schedulers in supporting video streaming services [4]. This scheduler mainly relies on the delay of the packets: the packets that violate the target delay of the service while waiting at the MAC buffer will be discarded. However, NRT services are scheduled using the classical PF rule.

The demand on both RT and NRT services is becoming significant and balanced in wireless networks. Hence, the proposed scheduler combines the previously mentioned components involved in the VT-M-LWDF and M-LWDF rules, i.e., the queue size of each flow and the packets delay are both considered in the proposed scheduler. As a result, this will help maintain the balance and satisfy both RT and NRT admitted users by providing a higher system spectral efficiency, higher system throughput, lower system PLR, guaranteeing a satisfactory level of fairness and supporting a higher number of subscribers.

The paper is organized as follows. Section II elaborates the system model, where we provide a description of the 3GPP

LTE system. Scheduling strategies are analysed in section III, where mathematical analysis and conceptual description are provided for the relevant schedulers and the proposed one. Section IV presents the performance evaluation results of the proposed scheduler compared to the benchmark schedulers, introduces the relevant performance metrics, the traffic models and the simulation environment. Finally, Section V concludes the paper.

II. System Model

The QoS in LTE is influenced by a significant number of factors such as number of resources available, channel conditions and the type of services (e.g., delay sensitive/insensitive). The resources allocated to a user are called physical resource blocks (PRB), equal to 180 KHz each in the frequency domain and to a slot of 0.5ms duration each in the time domain. The LTE standard comprises a flexible range of channel bandwidths; every bandwidth size is composed of different number of PRBs. The efficient use and proper selection of these bandwidths will improve the system efficiency as studied in [9].

The most relevant downlink parameters are set according to the LTE standard (i.e., resource allocation at every TTI equal to 1ms and composed of two time slots, whereas each time slot has 7 OFDM symbols spread over 12 consecutive sub-carriers and the FDD frame is composed of 10 TTIs).

In the system, Channel Quality Indicator (CQI) feedback should be reported to the eNodeB by the users using the uplink control messages over the Physical Uplink Shared Channel (PUSCH), as this CQI value represents the users‘ instantaneous channel quality at each PRB. The feedback mechanism adopted in our scenario is that the User Equipment (UE) sends a single CQI about every PRB in the correspondent channel bandwidth to the eNodeB. The Signal to Interference Noise Ratio (SINR) and CQI values are mapped in accordance with the mapping tables presented in [10]. As a result, the selected Modulation and Coding Scheme (MCS) guarantees a robust communication and good service delivery. Moreover, this decision ensures that the estimated block error rate (BLER) remains under the target BLER of 10% [10] [11].

The serving eNodeB's MAC layer controls the available PRBs by assigning them to the active flows which are competing for resources. At the MAC layer, packet schedulers are implemented. The scheduling decision is made based on various parameters such as channel conditions, Head-of-Line (HoL) packets delay, traffic types and buffer size.

NRT flows are transport control protocol (TCP) based and delay-tolerant services, i.e., delayed packets are not discarded and the buffer size increases. However, RT flows are delay sensitive services, where packets belonging to this kind of services are discarded, if the HoL packet delay exceeds the target delay.

III. Scheduling Strategies

Scheduling algorithms differ in the way they calculate the metric considered for the assignment of the resources to different users.

In most of the cases, scheduling algorithms require the knowledge of the average transmission data rate $\bar{R}_i(t)$ of each flow associated to the i-th user, and the instantaneous available data rate $r_{i,j}(t))$ for each sub-channel. As a result, when information about the performance guaranteed in the past for each flow is available, the system can perform fairness balancing.

The proposed algorithm is a result of a modification in the schedulers described below.

A. Proportional fair scheduler

The proportional fair (PF) scheduler assigns the PRBs by taking into account both the experienced channel quality and the past user average throughput [5]. This scheduler is very suitable to support NRT traffics. The goal of this scheduler is to improve the total network throughput and to guarantee fairness among the flows. Hence, the metric $W_{i,j}(t)$ in this scheduler is defined as the ratio between the instantaneous data rate of the i-th flow in the j-th sub-channel (i.e., $r_{i,j}(t)$) and the average data rate of the i-th flow (i.e., $\bar{R}_i(t)$). The following equation illustrates the metric used to represent the PF scheduler:

$$W_{i,j}(t) = \frac{r_{i,j}(t)}{\bar{R}_i(t)} \tag{1}$$

where $r_{i,j}(t)$ is computed based on the Adaptive Modulation and Coding (AMC) module, which is selected according to the CQI feedback received from the UE. This feedback represents the channel quality (e.g., SINR) of the j-th sub-channel associated to the the i-th flow. $\bar{R}_i(t)$ is the estimated average data rate.

B. M-LWDF Scheduler

The M-LWDF scheduler is developed to support multiple RT data users [6]. The scheduler assigns PRBs to different flows by considering the properties of the classical PF rule and the HoL packet delay for the RT flows. On the other hand, the PF scheduler is selected instead to support NRT data users. The following equation illustrates the metric used to represent the M-LWDF scheduler:

$$W_{i,j}(t) = \begin{cases} \alpha_i(t) * D_{HoL,i}(t) * (\frac{r_{i,j}(t)}{\bar{R}_i(t)}), & \text{if } i \in \text{RT} \\ \frac{r_{i,j}(t)}{\bar{R}_i(t)}, & \text{if } i \in \text{NRT} \end{cases} \tag{2}$$

where $\alpha_i(t)$ is the maximum probability that the HoL packet delay $D_{HoL,i}(t)$ (i.e., delay of the first packet to be transmitted in the queue) exceeds the target delay. Therefore, packets belonging to a RT service will be discarded if they violate the target delay while waiting at the MAC buffer. The definitions of $r_{i,j}(t)$ and $\bar{R}_i(t)$ are the same as the ones in III-A.

C. VT-M-LWDF Scheduler

The virtual token scheme presented in [8] is a modification of the M-LWDF rule, relying mainly on the parameters highlighted in III-B, except for the packet delay parameter. The VT-M-LWDF scheduler takes into consideration, in addition to the earlier parameters, the queue size parameter of each RT flow's buffer. This provides the scheduler an indication about the significance of the flow available in the buffer. Conversely, the M-LWDF scheduling decision is made with regard mainly to the packet delays.

The main goal of the VT-M-LWDF rule is to improve the QoS performance metrics for multimedia services, such as video and VoIP, and to maintain minimum throughput guarantee for NRT services. Furthermore, the latter impacts on the QoS requirements that belong to NRT services. Hence, there is a trade-off in the delivery of the QoS requirements between the RT and the NRT services. The following equation illustrates the metric used to represent the VT-M-LWDF scheduler:

$$W_{i,j}(t) = \begin{cases} \alpha_i(t) * Q_i(t) * (\frac{r_{i,j}(t)}{\bar{R}_i(t)}), & \text{if } i \in \text{RT} \\ \frac{r_{i,j}(t)}{\bar{R}_i(t)}, & \text{if } i \in \text{NRT} \end{cases} \quad (3)$$

where the definitions of $\alpha_i(t)$, $r_{i,j}(t)$ and $\bar{R}_i(t)$ parameters are the same as the ones defined earlier in III-B. $Q_i(t)$ refers to the queue size of the i-th flow at a particular scheduling epoch (t) when serving RT users. On the contrary, the PF scheduler is selected to make scheduling decisions when serving NRT users.

D. The Proposed Scheduler

Due to the voluminous and equal growth of multimedia services and NRT services, it is worth thinking of tailoring a downlink packet scheduling scheme that copes with both service classes simultaneously, delivers balanced QoS and utilizes the system radio resources efficiently. In other words, the users nowadays request not only video and VoIP services but also CBR, Internet browsing and best-effort services simultaneously.

The proposed scheduler is a modification of the M-LWDF and VT-M-LWDF schedulers described earlier in III-B and III-C respectively. This scheduler combines the effective components and the core principle of the aforementioned schedulers. The proposed scheduler adopts the consideration of the queue size and the packets delay parameters in the VT-M-LWDF and M-LWDF rules respectively, in order to improve the performance of the proposed scheduler when serving both RT and NRT classes compared to the benchmark schedulers. Hence, the major and effective components utilized in the proposed scheduler are the following:

- *Packets Delay* is the parameter that is obtained from the calculation of the HoL packet delay for every packet at every TTI. This parameter comes into play when the delay of the packet is approaching the target delay.
- *Queue Size* is obtained from the buffer that is attached to every flow which is carried over a unique bearer. This parameter helps the scheduler to be aware of the number of bits available on every buffer which will help quantify and prioritise the admitted users accordingly.

As a result, the proposed scheduler aims in particular at improving the performance metrics for video services and maintaining in general a satisfactory level of the performance metrics for the other services in the network simultaneously. The following equation illustrates the metric used to represent the proposed scheduler:

$$W_{i,j}(t) = \alpha_i(t) * D_{HoL,i}(t) * Q_i(t) * (\frac{r_{i,j}(t)}{\bar{R}_i(t)}), \text{for } i \in \text{RT/NRT} \quad (4)$$

where the definitions are the same as in III-B and III-C.

IV. Performance Analysis

The performance evaluation of the proposed scheduler is presented with respect to the VT-M-LWDF and the M-LWDF rules. Furthermore, this section introduces the simulation environment adopted in this paper, the traffic models, the relevant performance metrics and the numerical results.

A. Simulation Environment

We consider a single LTE cell, in which the users are uniformly distributed. In the centre of the cell, the eNodeB is positioned, whereas the users are modelled according to a random mobility model. The users' mobility is pedestrian with constant speed of 3 km/h. The distribution of the users, in terms of traffic classes, is considered in our scenario as follows: 40% of video users, 40% of VoIP users and 20% of CBR users. The proposed scenario is simulated using the LTE-Sim [11] simulator. This simulator provides the possibility to design various packet scheduling strategies at the eNodeB MAC layer depending upon the requirements needed.

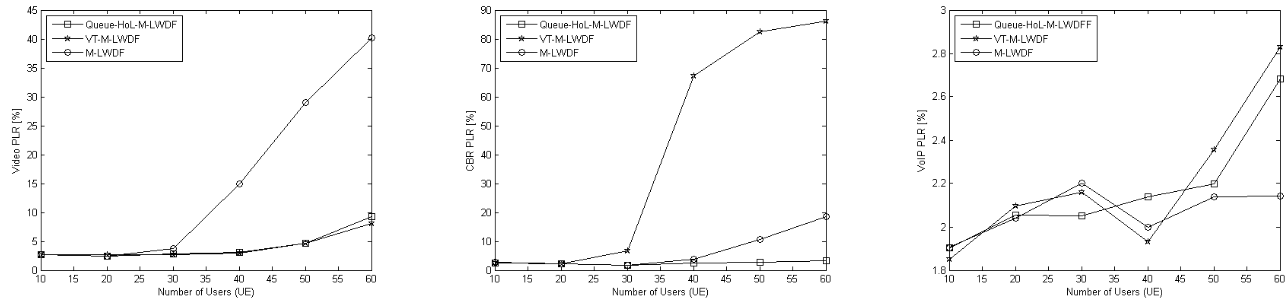
Simulation results of the relevant performance metrics are averaged over 10 simulation repetitions in order to obtain accurate and reliable results. The simulation parameters are reported in Table I.

TABLE I
Simulation parameters for LTE downlink system

PARAMETERS	VALUE
Bandwidth	10 MHz
Number of PRBs	50
Frame structure	FDD
Cell radius	1 km
E-UTRAN frequency band	1 (2.1 GHz)
Max delay	100 ms
Video bit-rate	242 kbps
CBR bit-rate	100 kbps
VoIP bit-rate	8.4 kbps
Flow duration	20 sec
Simulation time	30 sec
UE speed	3 km/h
Path loss/channel model	Typical Urban (Pedestrian-A propagation model)
Simulation repetitions per scheduler	10

B. Performance Metrics

The following four performance metrics are observed for the different scheduling strategies: *Packet loss ratio*, *average throughput*, *fairness index* and *system spectral efficiency*.



(a) Packet loss ratio for video flows (b) Packet loss ratio for CBR flows (c) Packet loss ratio for VoIP flows

Fig. 1. Packet loss ratio for Video, CBR and VoIP flows

The *packet loss ratio* parameter is calculated by dividing the difference between the transmitted and received packets with the number of transmitted packets. The *average throughput* parameter is calculated by dividing the number of successfully received bits with the duration of the flow. The *fairness index* parameter follows Jain's fairness index criterion [12]. The *system spectral efficiency* parameter is calculated by dividing the total number of received bits during simulation-time with the bandwidth size.

The packet loss ratio and the average throughput are user-oriented metrics, whereas the fairness index and the cell spectral efficiency are system-oriented metrics.

C. Traffic Model

The video flow is a trace-based application that sends packets based on realistic video trace files, which are available in [13]. The selected video flow is encoded at the rate of 242 kbps using the H.264 encoder. The maximum transmission unit is set to 500 bytes, as in [14].

The Voice flow is a bursty application that is modelled with an ON/OFF Markov chain. The generated VoIP flow is encoded at the rate of 8.4 kbps using the G.729 codec.

The CBR flow models a constant bit-rate application, where the packet size and the inter-arrival packet time are fixed. The selected CBR application is a best-effort flow and it has a constant bit-rate of 100 kbps, where the packet size and the inter-arrival packet time are set to 500 bytes and 0.04 s respectively.

D. Numerical Results

The numerical results show the relevant performance metrics defined in IV-B for the proposed scheduler with respect to VT-M-LWDF and M-LWDF as benchmark schedulers. Our proposed scheduler is demoted in the figures as Queue-HOL-M-LWDF.

The PLR for video, CBR and VoIP flows is depicted in Figure 1, where the degradation of the quality can be seen as the number of users increases. The proposed scheduler achieves remarkable balanced low PLR among the classes. The VT-M-LWDF achieves lower PLR for video flows but not for CBR flows, whereas the M-LWDF achieves lower PLR for CBR flows but not for video flows. The PLR for VoIP flows is less than 3% for the correspondent schedulers. The fluctuantion in Figure 1(c) is due to the nature of the VoIP traffic (bursty application) that is based on an ON/OFF Markov chain.

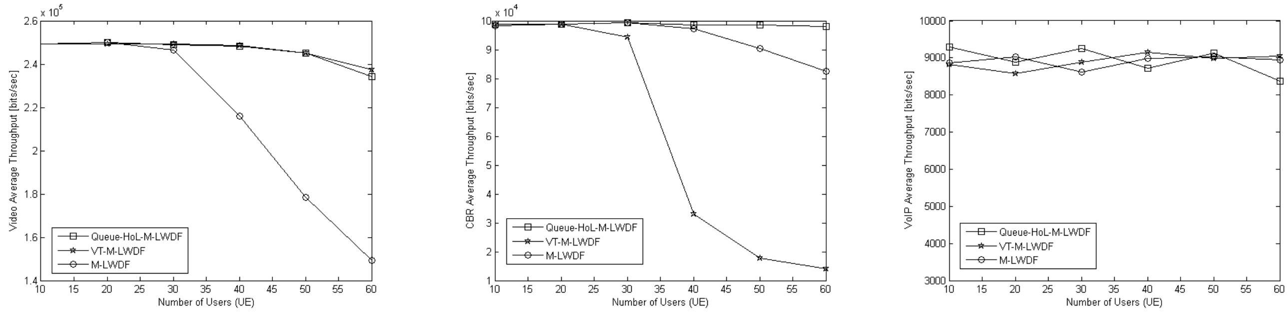
The average throughput for video, CBR and VoIP flows is depicted in Figure 2. The average throughput achieved by the proposed scheduler shows significant improvement for all the involved flows, even when the network is loaded with 60 users. The average throughput achieved by VT-M-LWDF is high for video flows but not for CBR flows, whereas M-LWDF achieves good average throughput for video flows when the network is loaded with 30 users and good average throughput for CBR flows when the network is loaded with 40 users. Furthermore, the average throughput drops as the number of users in the cell increases.

The fairness index parameter is depicted in Figure 3(a) and Figure 3(b) for video and CBR flows, respectively. This parameter shows how the system is serving users fairly by assigning a fair share of system resources. Hence, the figures show that the proposed scheduler achieves better fairness for different types of services.

The system spectral efficiency is depicted in Figure 3(c). The higher the spectral efficiency, the higher the channel utilization and the higher the throughput. It is evident from the figure that the proposed scheduler utilizes the channel more efficiently than VT-M-LWDF and M-LWDF. Furthermore, the cell is not saturated yet for the proposed scheduler and VT-M-LWDF, although the cell spectral efficiency is saturated for the M-LWDF scheduler when the cell is loaded with 60 users.

V. Conclusion

Providing an adequate QoS for different traffic classes over LTE networks is essential. In this paper, we proposed a downlink resource allocation algorithm for multi-traffic classes, e.g., video, VoIP, best-effort, etc., at the MAC layer in LTE. The proposed scheduler is based on a modification of the VT-M-LWDF and M-LWDF rules. We considered the queue size and the packet delay of the inter-class flows that are carried over various radio bearers. The proposed scheduler showed remarkable and balanced improvement for the different inter-class flows in terms of the following performance

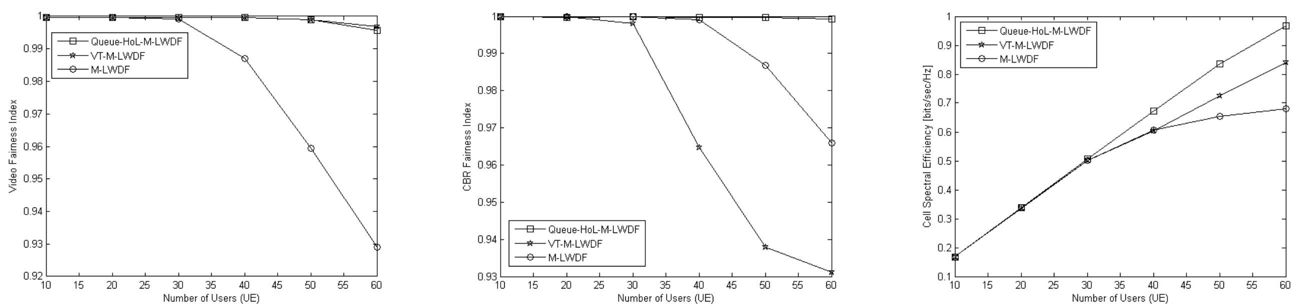


(a) Average throughput for a video flow

(b) Average throughput for a CBR flow

(c) Average throughput for a VoIP flow

Fig. 2. Average throughput for video, CBR and VoIP flows



(a) Fairness index for video flows

(b) Fairness index for CBR flows

(c) System spectral efficiency

Fig. 3. Fairness index for video and CBR flows in (a) and (b), and system spectral efficiency in (c)

metrics: PLR, average throughput, fairness among users and system spectral efficiency with respect to the VT-M-LWDF and M-LWDF rules. Moreover, the proposed scheduler aimed in particular at improving the performance metrics for video services and maintaining in general a satisfactory level of the performance metrics for the other services in the network simultaneously.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7) under grant agreement Nº 288502 (CONCERTO).

REFERENCES

[1] S. Sesia, I. Toufic, and M. Baker, "LTE - the UMTS long term evolution," *Wiley, UK*, 2009.
[2] W. Tong, E. Sich, Z. Peiying, and J. Costa, "True broadband multimedia experience," *IEEE Micro-wave Mag.*, vol. 9, no. 4, pp. 64–73, Aug. 2008.
[3] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, 2003.
[4] H. Ramli, R. Basukala, K. Sandrasegaran, and R. Patachaianand, "Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system," in *Proc. of IEEE Malaysia International Conf. on Comm., MICC*, Kuala Lumpur, Malaysia, 2009, pp. 815–820.
[5] J. G. Choi and S. Bahk, "Cell-throughput analysis of the proportional fair scheduler in the single-cell environment," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 766–778, Mar, 2007.
[6] P. Ameigeiras, J. Wigard, and P. Mogensen, "Performance of the M-LWDF scheduling algorithm for streaming services in HSDPA," in *IEEE Trans. Veh. Technol. Conf.*, vol. 2, Los Angeles, USA, Sep., 2004, pp. 999–1003.
[7] R. Basukala, H. M. Ramli, and K. Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system," in *IEEE F. Asian Himalayas Conf.*, vol. 2, Kathmandu, Nepal, Nov., 2009, pp. 1–5.
[8] M. Iturralde, A. Wei, T. Yahiya, and A. Beylot, " Performance study of multimedia services using virtual token mechanism for resource allocation in LTE networks." in *IEEE Vehicular Technology Conference (VTC)*, Sept. 2011, pp. 1–5.
[9] M. M. Nasralla, O. Ognenoski, and M. G. Martini, "Bandwidth scalability and efficient 2D and 3D video transmission over LTE networks," *IEEE ICC'13 - Workshop IIMC*, 2013.
[10] 3GPP Technical Specification 36.213, "Physical layer procedures (release 8)," www.3gpp.org.
[11] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: an open source framework," in *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, October, 2010, pp. 498–513.
[12] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *DEC Research Report TR-301*, 1984.
[13] http://trace.eas.asu.edu/, "H.264/AVC and SVC video trace library."
[14] C. Hewage, M. Martini, and N. Khan, "3D medical video transmission over 4G networks," in *4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Barcelona, Spain, October, 2011, pp. 26 – 29.