A new Channel and QoS Aware Scheduler to enhance the capacity of Voice over LTE systems

Biljana Bojovic and Nicola Baldo

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)

Av. Carl Friedrich Gauss 7, Castelldefels (Barcelona), Spain

{bbojovic, nbaldo}@cttc.es

Abstract—**In this paper we propose a novel MAC scheduling mechanism for the downlink of LTE systems based on a channel and QoS aware algorithm which performs joint time and frequency scheduling. The proposed scheduler aims at maximizing system radio resource utilization while providing QoS requirements to classes of traffic that have very tight QoS requirements in the terms of bit rate and delay, e.g. VoIP, online gaming and video conferencing.The performance of the proposed scheme is evaluated on voice traffic by using the E-Model to measure the perceived QoS. We carried out performance evaluation by simulation to compare the behavior of the proposed scheduler with state of the art schedulers in different network conditions. Results show that, in realistic scenarios in which the channel quality varies over time and frequency, the proposed scheduler significantly outperforms the state of the art solutions in terms of provided QoS and system capacity.**

I. Introduction and Related Work

Voice over LTE (VoLTE) is foreseen to become the dominant solution for the provisioning of voice services over 4G systems in the upcoming years [1]. While in previous mobile network technologies, such as 3G, voice traffic is conveyed over dedicated circuit-switched networks, in VoLTE system it is transferred over packet-switched networks along with other data traffic, such as for example best-effort HTTP traffic. Voice traffic has very tight Quality of Service (QoS) requirements, such as bit rate and delay. In order to support the QoS requirements for different type of services, LTE already provides the possibility of setting up different bearers within the Evolved Packet System (EPS), each being associated with a different traffic flow and mapped to its specific QoS requirements [2]. Vendor-specific QoS solutions are then expected to be deployed in order to fulfill these requirements.

In this paper, we focus on such QoS solutions related to the downlink part of the LTE radio interface, which is based on the Orthogonal Frequency Division Multiple Access (OFDMA). OFDMA allows a fine-grained dynamic allocation of radio resources both in the time domain (TD) and frequency domain (FD). This task is often referred to as Dynamic Packet Allocation or Packet Scheduling [3], and resides in the base station, which is called Evolved Node B (eNB) using the LTE terminology. The design of efficient Packet Scheduling algorithms is left open for LTE eNB manufacturers to come up with advanced solutions that are envisioned to become key product differentiation factors. Considering VoLTE services in particular, a good Packet Scheduling solution is required to include a radio resource allocation mechanism that is aware of the QoS requirements as well as of the channel conditions,

in order to maximize the voice capacity, i.e., the number of voice flows that can be served by the eNB with guaranteed QoS requirements.

Due to the increasing popularity of the LTE technology systems worldwide, there has been a growing interest in the design of LTE Packet Scheduling algorithms, and several downlink scheduling algorithms have already been proposed in the scientific literature. A very recent and abundant survey on the topic is provided in [4]. However, most of the scheduling algorithms mentioned in this survey, such as Round Robin, Proportional Fair, Maximum Throughput, Throughput to Average and Blind Equal Throughput, actually are not QoS-aware, and hence are not suitable for VoLTE systems. For this reason, we do not consider such schedulers in the present paper. Instead, we focus on the most relevant recently proposed QoS-aware LTE downlink scheduling algorithms.

A first category of such algorithms includes those that are aiming at satisfying the delay requirement of real-time traffic, such as the scheme proposed in [5] which prioritizes the data flows to be scheduled based on the Head-of-line (HOL) delay parameter. A downside of this approach is that it does not take into consideration the variable channel conditions; in particular, in realistic scenarios in which the presence of fast and frequency-selective fading is expected, assigning radio resources based only on the HOL metric often results in the selection of lower modulation and coding schemes, which is spectrally inefficient and thus does not allow to achieve a high voice capacity.

Among the channel-aware approaches, we cite the Token Bank Fair Queue (TBFQ) scheduler [6], which is a queue- and channel-aware scheduling algorithm which attempts to maintain fairness among users. TBFQ is based on the leaky-bucket principle, and it is mainly designed to support bursty traffic, by assigning a higher amount of resources to the users that have more data in the queues. This feature of the TBFQ approach is not adequate for voice traffic, since it is characterized by small packet sizes and low expected queue fill levels. Furthermore, TBFQ does not explicitly take into account the delay requirements.

A better candidate for voice traffic is the Priority Set Scheduler (PSS) [7], which is a channel-aware scheduler that aims at guaranteeing a predefined bit rate to each user. This algorithm has a very good performance because it successfully combines TD and FD scheduling in order to achieve a higher spectral efficiency and increase the overall system capacity. Regarding the QoS support, the main drawback of this scheduler is that

it only considers the Guaranteed Bit Rate (GBR) parameter specified within the EPS bearer. This means that delay sensitive classes of traffic, such as voice, video and gaming, may suffer poor quality even if their GBR requirement is satisfied. This limits the application of this scheduler to delay insensitive traffic.

As a step forward in this research line, in this paper we propose a new LTE downlink scheduling algorithm called Channel and QoS Aware (CQA) scheduler. The QoS parameters that it considers are the HOL and the GBR parameters. The CQA scheduler performs the scheduling according to different criteria in the TD and FD, in order to achieve a high spectral efficiency while at the same time taking care of satisfying the delay requirements of the traffic.

The main contributions of this paper are:

1) the definition of the novel CQA scheduling algorithm;
2) its performance evaluation by means of network simulations in a VoLTE scenario, in comparison with state-of-the-art QoS-aware LTE downlink schedulers.

II. User-perceived QoS of voice calls

According to the ITU-T E-model [8] the quality of a voice call can be estimated by calculating the R factor, which we denote as R_f:

$$R_f = Ro - Is - Id - Ie_{eff} + A \, , \qquad (1)$$

where Ro is the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise; Is are impairments simultaneous to voice signal transmission, such as too loud speech level, non-optimum sidetone, quantization noise; Id are impairments caused by delay and echo effects; Ie_{eff} represents impairments caused by low bit-rate codecs and impairment due to randomly distributed packet losses; finally, the advantage factor A allows for compensation of impairment factors when the user benefits from other types of access to the user. A user is satisfied with the QoS of voice call when the R_f is greater than a threshold [8]. In order to provide satisfactory R_f, the scheduling mechanism should optimize all the metrics at the MAC layer that affect QoS of voice calls. Those metrics are: MAC layer throughput, MAC layer queuing delay and packet losses caused by buffer overflows. Taking this into account, we propose an algorithm that aims at simultaneously:

- minimizing delay by giving priority to the user with greater HOL delay
- maximizing MAC layer throughput by improving radio resource utilization
- allocating to each user the amount of radio resources that is necessary to achieve the guaranteed bit-rate specified by the GBR parameter in the EPS bearer

III. Proposed scheduling algorithm

The CQA scheduler is based on joint TD and FD scheduling, which has been shown in many studies to be more efficient approach than only TD or FD scheduling; an example of such performance comparison for LTE system can be found in [9].

The proposed algorithm runs every transmission time interval (TTI) which is equal to 1 ms. The TTI is the smallest resource unit in the time domain. In the FD the smallest resource unit is resource block (RB) which forms RB groups (RBGs). Depending on the system settings, such as bandwidth and type of allocation, one RBG can contain different number of RBs. The smallest resource unit that our scheduling algorithm assigns is one RBG.

In the TD, at each TTI, the CQA scheduler selects from all the users $j = 1, ..., N$ those that did not yet reached the maximum bit rate (MBR)[1] and groups them by HOL delay calculating the metric m_{td} in the following way:

$$m_{td}^j(t) = \lceil \frac{d_{HOL}^j(t)}{g} \rceil \, , \qquad (2)$$

where $d_{HOL}^j(t)$ is the current value of HOL delay of flow j, and g is a grouping parameter that determines granularity of the groups, i.e. the number of the flows that will be considered in the FD scheduling iteration. The grouping is used to select the most urgent flows, i.e., with the highest value of HOL delay, and to enforce the scheduling mechanism to consider those flows in the following FD scheduling iteration. A low value for g reduces the users diversity, thus decreases scheduler's gains in the FD; on other hand, it gives more importance to the d_{HOL} metric in the scheduling. This can be useful in network scenarios in which all users have relatively good channel conditions and the fast fading is negligible. On the contrary, a high value for g increases the users diversity, thus increases FD gains, but d_{HOL} has less impact in scheduling decisions. However, this parameter should be set up according to the network capacity and the expected average number of the users in the system. The groups of flows selected by TD iteration are forwarded to FD scheduling starting from the flows with the highest value of the m_{td} metric until all RBGs are assigned in the corresponding TTI.

In the FD, for each RBG $k = 1, ..., K$, the CQA scheduler assigns the current RBG to the user j that has the maximum value of the FD metric which we define in the following way:

$$m_{fd}^{(k,j)}(t) = d_{HOL}^j(t) \cdot m_{GBR}^j(t) \cdot m_{ca}^{k,j}(t) \, , \qquad (3)$$

where $m_{GBR}^j(t)$ is calculated as follows:

$$m_{GBR}^j(t) = \frac{GBR^j}{\overline{R^j}(t)} = \frac{GBR^j}{(1-\alpha) \cdot \overline{R^j}(t-1) + \alpha \cdot r^j(t)} \, , \quad (4)$$

where GBR^j is the bit rate specified in EPS bearer of the flow j, $\overline{R^j}(t)$ is the past averaged throughput that is calculated with a moving average, $r^j(t)$ is the throughput achieved at the time t, and α is a coefficient such that $0 \le \alpha \le 1$. In (3) the purpose of the d_{HOL} and m_{GBR} metrics is to provide to all flows the same level of QoS regarding delay and GBR by prioritizing the flows that have higher HOL delay and the flows which ratio of GBR^j to $\overline{R^j}(t)$ is larger. For example, if the GBR is achieved, but not also the MBR,

---

[1]MBR represents the maximum allowed bit rate and it is specified in EPS bearer of the flow.

$GBR^j < \overline{R^j}(t) < MBR^j$, to the flow j will be assigned lower priority in scheduling since $m_{GBR} \leq 1$. The purpose of $m_{ca}^{(k,j)}(t)$ is to add channel awareness to the system in order to maximize the capacity by assigning the resources to the flows that can use them more efficiently. For $m_{ca}^{(k,j)}(t)$ we consider two different metrics: $m_{pf}^{(k,j)}(t)$ and $m_{ff}^{(k,j)}(t)$. The m_{pf} is the Proportional Fair metrics which is defined as follows:

$$m_{pf}^{(k,j)}(t) = \frac{R_e^{(k,j)}}{\overline{R^j}(t)} \ , \tag{5}$$

where $R_e^{(k,j)}(t)$ is the estimated achievable throughput of user j over RBG k calculated by the Adaptive Modulation and Coding (AMC) scheme that maps the channel quality indicator (CQI) value to the transport block size in bits. The CQI value is the indication of the data rate which can be supported by the channel, taking into account the signal to interference plus noise ratio (SINR) and the characteristics of the UE's receiver [2]; this value is reported by UE to eNB for each RB as part of channel state information (CSI) reporting procedures which are defined in [10]. We consider the m_{pf} metric as good channel awareness metric since it aims at simultaneously achieving the fairness among flows and maximizing system capacity by prioritizing the users that have suffered lower channel quality and the users that have extremely good instantaneous channel quality; we denote the CQA scheduler that uses this channel awareness metric CQA_{PF}. The other channel awareness metric that we consider is m_{ff} which is proposed in [7] and it represents the frequency selective fading gains over RBG k for user j and is calculated in the following way:

$$m_{ff}^{(k,j)}(t) = \frac{CQI^{(k,j)}(t)}{\sum_{k=1}^{K} CQI(t)^{(k,j)}} \ , \tag{6}$$

where $CQI^{(k,j)}(t)$ is the last reported CQI value from user j for the k-th RBG. We consider this metric as good channel awareness metric since it aims at increasing the overall system capacity by prioritizing users that can use available resources more efficiently. We name the CQA scheduler that uses this channel awareness metric CQA_{FF}.

IV. PERFORMANCE EVALUATION

A. Description of the scenarios

To evaluate the proposed scheduler we simulate a typical outdoor scenario in which N UEs are attached to a single eNB. All users perform voice over IP (VoIP) calls and have corresponding GBR EPS bearers set up in the EPS. We consider a single cell scenario, thus inter cell interference is not considered in this work. The users are randomly distributed in a square area around the macro cell. We consider two channel scenarios that are based on different channel models:

- Simplified channel model: the UEs are static ($v = 0 \ km/h$) and no model for time and frequency selective fading is used. Thus, in this scenario the SINR perceived by the UEs remains unchanged during the simulation. Even if the simulation scenario based

TABLE I.   SIMULATION PARAMETERS

Macro cell transmission power	40 dBm
Bandwidth	6 RB (1.4 MHz)
Downlink/Uplink carrier frequency	2120/1930 MHz
RBs per RB group (RBG)	1
Pathloss model	Cost 231 (Hata Model PCS Extension)
Fading loss model	Pedestrian EPA model 3 km/h
Adaptive modulation and coding scheme	PiroEW2010
Users mobility model	Steady state random waypoint
Simulation area	1560x1560 m^2
Voice EPS bearer settings	GBR, conversational voice, 80 kb/s
Voice codec	G711
TTI	1 ms
Number of UEs in system	1-20
Replications of each simulation setup	10
Simulation time	8 s

on this model does not represent a realistic LTE system, we consider that using this model can help correlation understanding the performance of different schedulers for different channel conditions.

- EPA model: the UEs are mobile ($v = 3 \ km/h$) and Extended Pedestrian A model (EPA) is used to simulate fading with values of model parameters defined in [11]. Due to the UE's mobility along with the fading model, the quality of channel varies over time and frequency, thus the scenario that is based on EPA model can be considered as a realistic LTE scenario.

B. Simulation setup

We use the LTE-EPC network simulator (LENA) [12] to carry out the performance evaluation. To simulate the performance of the state of the art algorithm we used the implementation of the PSS scheduler provided by [13], considering both versions of PSS scheduler: PSS_{pf} and PSS_{coita}. We implemented in the LENA simulator the HOL scheduler according to [5] and both versions of the proposed CQA scheduler that we described in Sec. III of this paper. The simulation parameters are shown in Table I and the system configuration is as follows.

The macro cell is connected via the PDN Gateway (PGW) to the internet and for each UE a separate remote peer was placed in the internet and connected to the gateway of the LTE network (the PGW) via a separate point to point link with overprovisioned bandwidth in order to simulate end-to-end performance of voice calls. For the path loss model we adopt the COST-231 path loss model [14], which is a common choice to simulate macro cell outdoor scenarios. The size of the square area is chosen such that, given the path loss model and the other network parameters, we have a wide range of SINR values, which we verified by observing the CQI values reported by the UEs to be in range $[1, 15]^2$. The tunable parameter of CQA scheduler g was determined empirically to give the best throughput performance for CQA scheduler for this network scenario. Its value is constant for both scenarios and equal to 300. While in simplified scenario we use a constant position mobility model, in the EPA scenario

---

[2]Note that CQI 1 correspond to a UE using the lowest MCS but still connected to the eNB, whereas CQI 15 correspond to a UE using the highest MCS.
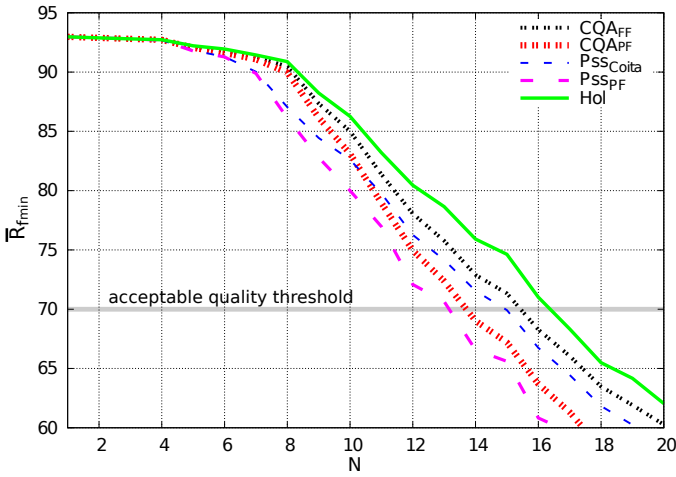
Fig. 1. QoS performance comparison by using $R_{f\,min}$ metric in the static scenario

Fig. 2. System throughput performance comparison in the static scenario

we simulate the mobility of UEs by using steady state random waypoint mobility model [15]. The initial distribution of UEs is the same in both scenarios. We simulate the fast fading model described by EPA model using a Rayleigh multi-path fading model and we set the model parameters by using values defined in [11] for the EPA model. For each experiment setup we run 10 independent replications of each experiment which gives different position topologies.

C. Results

The performance of the proposed CQA scheduler and the state of the art schedulers is evaluated in the terms of QoS satisfaction and system performance. To measure QoS satisfaction we use the R-factor R_f that we described in Sec. II. According to the E-model [8] the users are satisfied with QoS of voice call when $R_f \geq 70$. We use this threshold to evaluate if QoS of calls is satisfied. We consider the minimum R_f value $R_{f\,min}$ over all the users in the system as the strictest metric to evaluate the performance of schedulers. We aim to evaluate the average number of users that each scheduling algorithm can support while providing to all the users satisfactory QoS. For this purpose, we define the metric $\overline{R_{f\,min}}$ which represents the average $R_{f\,min}$ over the set of independent replications of simulations. We denote the number of users for which $\overline{R_{f\,min}} = 70$ as $\overline{N_{QoS}}$, which represents the voice capacity of the system.

1) Simplified channel model: In Fig. 1 we show the $\overline{R_{f\,min}}$ for all the schedulers. The HOL scheduler achieves on average the highest number of satisfied users $N_{QoS} = 16$. In this scenario the schedulers that are channel aware cannot benefit from the UE diversity, thus their performance degrades. The schedulers that use the proportional fair metric, such as CQA_{PF} and PSS_{PF}, perform worst, while the schedulers based on the m_{ff} metric, such as CQA_{FF} and PSS_{coita}, perform slightly better.

In Fig. 2 we show the system throughput which is calculated as the average total VoIP throughput at the application layer over all simulations having the same number of users. In a scenario with VoIP calls the application layer throughput is significantly lower than the MAC throughput because of
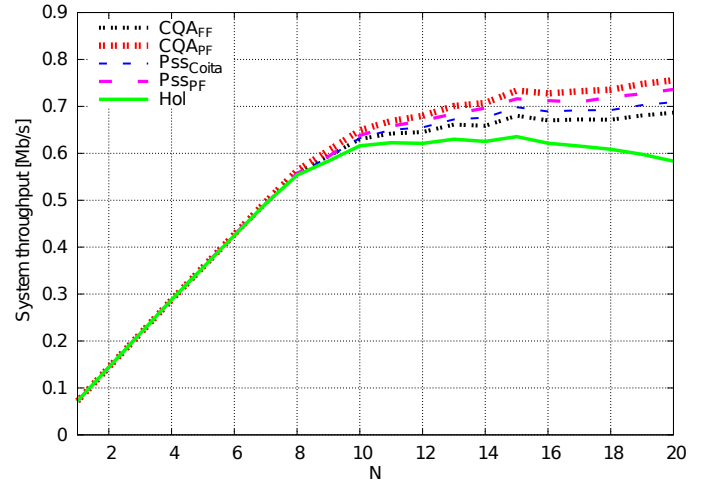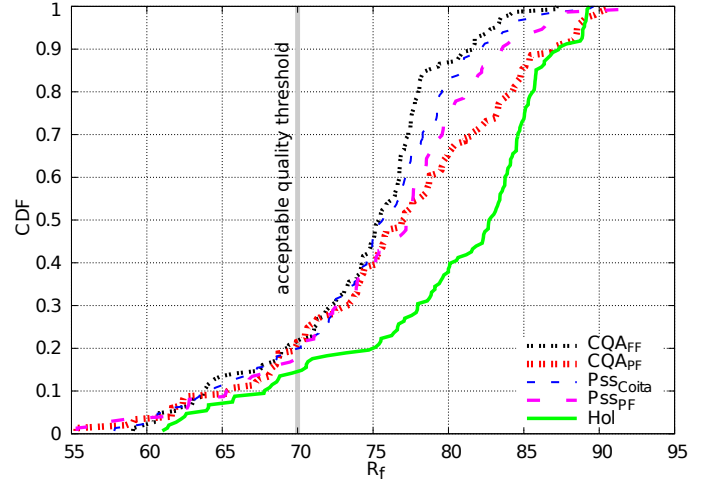

Fig. 3. Cumulative distribution function of R_f in the static scenario when $N = 15$

$RLC + PDCP + IP + RTP$ overhead; moreover the transport blocks assigned to the user is is often greater than the VoIP packet. Because of this, a portion of resources that are assigned to the user is wasted. Also it is important to notice that channel quality varies among UEs, and the UEs achievable rate depends on AMC, so the system throughput is significantly lower than the peak LTE throughput that is often advertised for the given bandwidth. From Fig. 1 and Fig. 2 we notice that the HOL scheduler achieves good performance in the terms of the QoS while has poor system throughput performance. In fact, a low throughput performance is expected for the HOL scheduler since it has low radio resource utilization efficiency. On the other hand, CQA_{PF} and PSS_{PF} have the highest throughput performance, but the worst QoS.

In Fig. 3 we show the cumulative distribution function of R_f. The HOL scheduler has the lowest probability to have unsatisfied VoLTE users, and this probability is 5% lower comparing to other schedulers. We notice from the same figure that the second scheduler with the lowest CDF is CQA_{PF} scheduler and we argue that this is because it achieves high throughput gains (See. Fig. 2), which reduces delays and
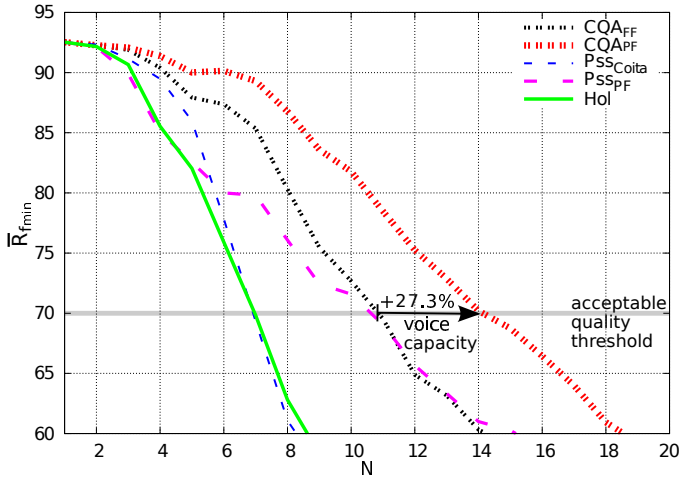
Fig. 4. QoS performance comparison by using $R_{f\,min}$ metric in the EPA scenario

Fig. 5. System throughput performance comparison in the EPA scenario

packet losses due buffer overflow.

2) EPA model: We evaluate the performance of all schedulers in the scenario in which UEs are moving, and fast frequency selective fading is present. From Fig. 4 we notice that the proposed CQA_{PF} scheduler achieves significantly better performance regarding the provided QoS than the other schedulers. The CQA_{PF} scheduler achieves increase in VoIP capacity up to 27% compared to CQA_{FF} and Pss_{PF} gains; and up to 100% compared to the Pss_{coita} and the HOL schedulers. We explain these performance gains by the use of m_{fd} metric from (3) which balances the delay and GBR requirements with the capacity maximization objective. On the other hand, we explain the low performance of the HOL scheduler by the fact that it is not leveraging fast fading and users diversity. The Pss_{PF} and the CQA_{FF} schedulers perform similarly which can be explained by the fact that CQA_{FF} is gaining more performance by being HOL delay aware, while Pss_{PF} is gaining higher performance by being channel aware. We notice that the schedulers that are using the Proportional Fair metric, i.e., CQA_{PF} and Pss_{PF}, achieve much higher performance gains than the schedulers that are using m_{ff} metric.

In Fig. 5 we show the application layer system throughput performance for the EPA scenario for the case $N = 15$. From the figure we notice that the proposed CQA_{PF} scheduler achieves significant gains in the terms of system throughput comparing to all the other schedulers. We again notice that the schedulers that use the m_{pf} metric achieve higher performance.

In Fig. 6 we show the cumulative distribution function of R_f for EPA scenario for the case $N = 15$. We notice that the CQA_{PF} scheduler has 25% higher probability to have satisfied users than the Pss_{PF} scheduler, around 80% higher than the CQA_{FF} scheduler, and approximately 90% higher than the Pss_{coita} and HOL schedulers. Finally, we conclude that in a realistic scenario with fast and frequency selective fading, schedulers which use Proportional Fair metric, such as CQA_{PF} and Pss_{PF}, significantly outperform schedulers based on m_{ff} metric, such as CQA_{FF} and Pss_{coita}.
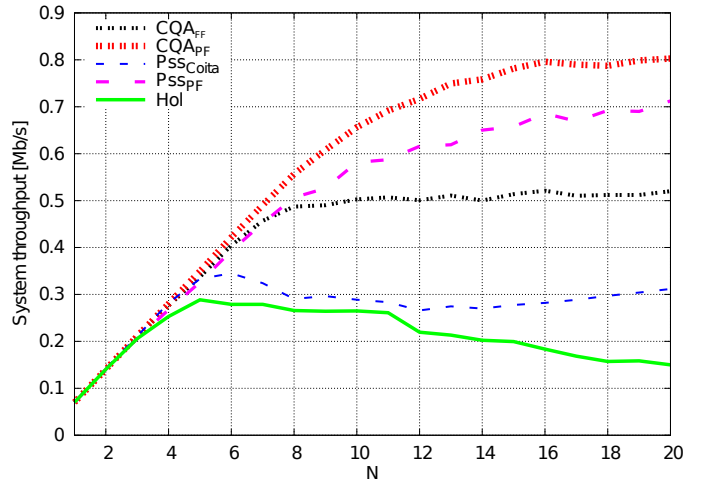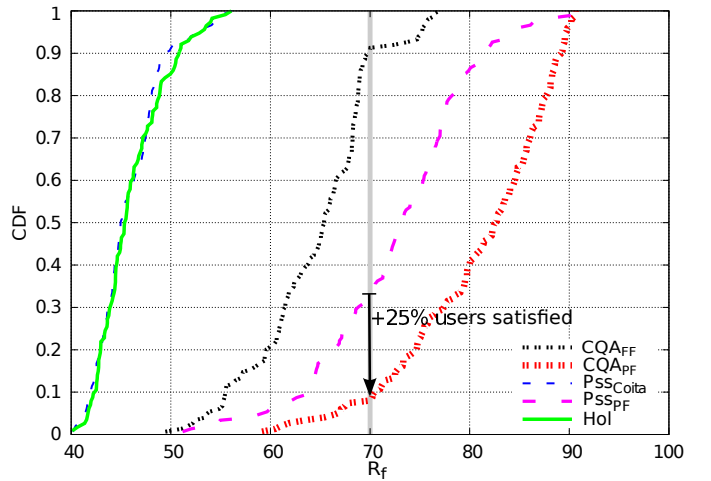

Fig. 6. Cumulative distribution function of R_f in the EPA scenario when $N = 15$

V. CONCLUSIONS

In this paper we proposed a novel scheduling algorithm that is both channel aware and QoS aware. Our CQA scheduling algorithm aims to enhance the VoLTE capacity. We proposed two different version of this algorithm CQA_{PF} and CQA_{FF}. We carried out performance evaluation by simulation and compared our solution with the state of the art scheduling algorithms: the PSS and the HOL delay scheduler. Results show that, in a realistic pedestrian scenario in which fast fading is present, our CQA scheduler gains approximately to 27% of VoLTE capacity compared to the PSS scheduler and almost 100% compared to the HOL scheduler.

REFERENCES

[1] M. Poikselkä, H. Holma, J. Hongisto, J. Kallio, and A. Toskala, *Voice over LTE (VoLTE)*. John Wiley & Sons, 2012.

[2] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution (From Theory to Practice)*. John Wiley & Sons, 2009.

[3] 3GPP TS 36.300 version 11.7.0 Release 11, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," 2013.

[4] F. Capozzi, G. Piro, G. L.A, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, 2013.

[5] I. Ahmed, L. Badia, N. Baldo, and M. Marco, "Design of a Unified Multimedia-Aware Framework for Resource Allocation in LTE Femtocells," in *Proceedings of International Symposium on Mobility Management and Wireless Access (MobiWac)*, 2011.

[6] F. Bokhari, W. Wong, and H. Yanikomeroglu, "Adaptive Token Bank Fair Queuing Scheduling in the Downlink of 4G Wireless Multicarrier Networks," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[7] G. Mongha, K. Pedersen, I. Kovács, and P. Mogensen, "QoS Oriented Time and Frequency Domain Packet Schedulers for The UTRAN Long Term Evolution," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[8] *ITU-T G.107, The E-model: a computational model for use in transmission planning*, 2011.

[9] K. C. Beh, S. Armour, and A. Doufexi, "Joint time-frequency domain proportional fair scheduler with harq for 3gpp lte systems," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[10] 3GPP TS 36.213 version 11.4.0 Release 11, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 2013.

[11] 3GPP TS 36.104 version 11.6.0 Release 11, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," 2013.

[12] "LTE-EPC Network Simulator (LENA)," Last time accessed: December 2013. [Online]. Available: http://networks.cttc.es/mobile-networks/software-tools/lena/

[13] D. Zhou, N. Baldo, and M. Miozzo, "Implementation and Validation of LTE Downlink Schedulers for ns-3," in *Proceedings of Workshop on ns-3 (WNS3)*, 2013.

[14] *Digital Mobile Radio: COST 231 View on the Evolution Towards 3rd Generation Systems*. Commission of the European Communities, L-2920, Luxemburg, 1989.

[15] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," *IEEE Trans. Mobile Comput.*, vol. 3, pp. 99–108, 2003.