Performance Analysis of Scheduling Algorithms for Web QoE Optimization in Wireless Networks

Ianire Taboada, Jose Oscar Fajardo, Fidel Liberal

Department of Communications Engineering, University of the Basque Country

Urquijo s/n, 48013 Bilbao, Bizkaia (Spain)

Tel: (+0034) 946017352

E-mail: {ianire.taboada, joseoscar.fajardo, fidel.liberal}@ehu.es

Abstract

This paper deals with the problem of scheduling mobile Internet traffic over a single cell wireless downlink data channel. On the one hand, we compare the performance of different size-based or/and channel-aware scheduling disciplines in mean delay terms, concluding that an approach that combines both size- and channel-awareness is the best alternative, which guarantees acceptable uplink overhead due to channel information reports. On the other hand, we study the impact of different scheduling algorithms on user's satisfaction for web browsing service. As concluded, in order to achieve a trade-off between maximizing overall mean subjective quality and reducing uplink signaling overhead due to channel quality reports, a scheduling policy that benefits from opportunistic gains in combination with a size-aware tie-breaking rule gives the best results.

Keywords: CQI reporting rate, scheduling algorithm, tie-breaking rule, time-varying channel, web QoE optimization.

1. Introduction

Undoubtedly, Internet has become increasingly essential in our daily lives. Nowadays, due to the last decade technological developments in emerging wireless systems the provision of IP-based services is one of the main goals of modern cellular systems. Consequently, all this has led to a drastic growth of the demand for mobile web browsing service, which entails a higher demand of network bandwidth capacity. Thus, how a scheduler located in the base station assigns radio resources for this data flow transmission to mobile users becomes extremely important.

The traffic from web browsing applications is of elastic nature. A significant performance measure for such elastic flows is the delay. It is the scheduling algorithm used for sharing radio resources among users the major cause that affects the value of this metric. Since users do not care about delays of individual packets but only about the total time to transmit a file of a given size, flow-level models are the most appropriate to characterize the system at the time-scale where users experience the performance. Hence, finding the suitable scheduling policy that minimizes overall flow mean delay in the system is a challenging task.

The time-varying nature of wireless channel provides an opportunity to schedule flows when they see favorable instantaneous channel conditions, which is referred to as opportunistic or channel-aware scheduling [1]. Mobile users send this channel quality information to the scheduler located in the base station to use it when making scheduling decisions. Nevertheless, in order to not deteriorate system performance it is vital to guarantee an acceptable uplink overhead due to these channel quality reports [2], but the sending frequency reduction of these reports may ensure a suitable scheduling performance. Besides, size-based scheduling algorithms [3] [4] that favor short flows have been widely studied in mean delay optimization problems with constant channel capacity. Therefore, a resource allocation strategy that combines channel-awareness and size-awareness is a key source of performance enhancement for wireless data networks.

On the other hand, when there are several users that are the most suitable to transmit data, choosing the right tie-breaking rule is crucial [5]. These ties are usually broken randomly or using First In, First Out (FIFO), without considering size or channel information. In this way, in this paper we are interested in studying the performance of a channel-aware scheduling proposal with a size-aware tie-breaking rule.

Although most of resource allocation strategies for next generation mobile systems are driven by Quality of Service (QoS) parameters such as delay, subjective quality perception of end-users becomes essential for providers in order to achieve an acceptable user satisfaction. ITU-T defines Quality of Experience (QoE) [6] as "the overall acceptability of an application or service, as perceived by the end-user". Therefore, a QoE-driven resource allocation approach is a clear motivation to maximize end users' satisfaction. Even so, QoE-based resource allocation schemes have not been proposed so far in the literature. Utility functions that represent user satisfaction are used in [7] [8] [9]. This work focuses on dynamic radio resource management for web QoE optimization. In spite of this, since delay-based scheduling algorithms have received significant attention from research community over

years in conjunction with technological advances, interest in scheduling algorithms aimed at reducing overall mean delay remains. Then in this paper, instead of incorporating QoE-awareness in scheduling task, we study the performance of scheduling algorithms aimed at reducing mean flow delay not only in delay terms but also in QoE terms.

Thus, the aim of this paper is twofold. On the one hand, the paper considers a tie-breaking rule based on flow size to use in combination with a pure channel-aware policy, and it compares its mean delay performance with other size-based and/or channel-aware scheduling strategies in a time-varying channel context, taking into account channel quality information reporting rate. On the other hand, this work analyzes the impact of different scheduling algorithms on mean QoE for mobile web browsing service.

The rest of the paper is structured as follows. First of all, Section 2 highlights the related works. In Section 3 system model is described. The mean delay performance of different scheduling algorithms is studied in Section 4. Results inferred from subjective tests are used in Section 5 to compare the performance of scheduling disciplines in QoE terms. Finally, Section 6 gathers the main conclusions of this paper.

2. Related work

The analysis of mean delay minimization problem has been subject of many studies in recent years, both in the area of channel-aware scheduling algorithms and size-aware ones.

Referring to opportunistic scheduling, several channel-aware policies have been investigated in the literature, among the most known Best Rate and Proportional Fair [1]. However, flow-level models with time-varying service rates are extremely difficult to solve analytically, and to the best of our knowledge there is no mathematical resolution. [5], [10] and [11] formulate the problem of minimizing overall mean delay in a time-varying channel context as a Markov Decision Process (MDP). Due to the impossibility of solving it analytically, approximated solutions are derived, which result in sub-optimal heuristic index rules. Nevertheless, service time distributions considered in these works are not realistic, as they are far from the Pareto ones that model properly the size of Internet flows [12].

Concerning size-aware scheduling strategies, these can be classified into two main categories: those that need prior knowledge of flow sizes, and those that do not but which make use of size-related information. We restrict our study to the last category, denominated as non-anticipating, since they are more suitably applicable in most of current network systems. In this field, the most relevant work is the approach of Gittins [13] [14] [15] which based on attained services of jobs, proposes an index rule that always minimizes overall mean delay. For Decreasing Hazard Rate (DHR) service time distributions, such as Pareto, Least Attained Service discipline [16] is optimal, which in mean delay terms is equivalent to Gittins for this kind of service time distributions.

As concluded in [17], an Opportunistic Gittins rule that combines both size-awareness and channel-awareness seems to be the best option (but not optimal), which guarantees an acceptable uplink overhead due to channel quality reports. Nonetheless, in [17] ties are broken using FIFO, without considering size information. Following the philosophy of the

so-called Shortest Remaining Process Time Fastest Machine (SRPT-FM) [18] optimal anticipating policy, where the fastest machine is assigned to the shortest job, the idea of extrapolating this policy to a variable capacity single machine comes up. This way, based on this idea our proposal is to use a policy that maximizes opportunistic gains in the first iteration, and then use a tie-breaking rule based on flows' attained service.

As a result, although all the reviewed studies are related to the case study, none of them covers all the objectives proposed in this paper.

3. System model

In this section we describe our modeling assumptions in order to characterize the scheduling problem for downlink data traffic in a single cell of a wireless system.

3.1 Traffic model

We consider the transmission of downlink traffic from the base station to mobile users in a single cell context. We assume that traffic is of elastic nature representing web traffic. Each elastic flow is characterized by its size, X, this is, the total amount of bits to be transferred. Flow sizes are assumed to be independent and identically distributed random variables with the cumulative distribution denoted by $F_x(x) = P\{X \le x\}$, where $f_x(x)$ refers to the corresponding density function. Internet elastic traffic flow sizes are properly modeled by means of Pareto distributions. In concordance with this, we assume a Pareto service time distribution with shape parameter $\beta > 1$ and scale parameter $b > 0$ for all $x \ge 0$. This way,

$$f_x(x) = \frac{b \cdot \beta}{(1 + b \cdot x)^{\beta+1}} \tag{1}$$

To have a more realistic situation, instead of having a fixed number of users in the system, random arrivals are considered. We assume that elastic flows arrive at the base station according to a Poisson process with rate λ.

Another relevant parameter in this area is the hazard rate function, which is the rate of completing service being attained service x. This is defined by $h_x(x) = f_x(x) \Big/ \overline{F}_x(x)$, where $\overline{F}_x(x) = 1 - F_x(x)$. The resulting hazard rate for Pareto flow sizes is given in the following, which belongs to DHR function class.

$$h_x(x) = \frac{b \cdot \beta}{1 + b \cdot x} \tag{2}$$

On the other hand, mobile users' initial location in the cell follows a uniform distribution.

3.2 Channel model

Concerning channel characteristics, wireless link capacity, *R*, is variable over time. This time-varying behavior is due to fading degradations and user's mobility. In order to emulate channel time-varying capacity, Channel Quality Indicator (CQI) reports sent from mobile devices to the base station are used. These CQI traces are obtained from a system level radio access simulator [19]. A mapping from CQI values to rate values is shown in Table 1, extrapolated from [20] for a cell of 5 MHz.

Channel state is defined by CQI index, and channel state probabilities are presented in Table 1, both for static (p_1) and dynamic users (p_2). Thus, channel model covers two different rate distributions, one related to a scenario with no mobility and the other to a slowly moving users case. In the case of dynamic users, channel propagation is defined by Extended Pedestrian A model [21], and we consider that channel states for a user are independent and identically distributed in different time slots. However, as stated below, for the static case, the channel state, independent and identically distributed among users, is defined by the user initial position in the cell with the probabilities in Table 1, and it does not vary over time.

Table 1. CQIs and corresponding rates, and channel state probabilities in static and mobility (5 km/h) scenarios.

CQI index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rate (Mbps)	0	0.63	0.97	1.57	2.50	3.65	4.89	6.15	7.97	10.0	11.3	13.8	16.2	18.8	21.3	23.1
p_1	0	0.04	0.29	0.12	0.14	0.09	0.08	0.07	0.07	0.05	0.03	0.01	0.00	0	0	0
p_2	0.28	0.12	0.09	0.08	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.01	0.00	0.00	0.00	0

We define the overall mean and variance of channel capacity in the cell as follows:

- *Mean rate*$(\tilde{R}) = \mathrm{E}[\mathrm{E}[\tilde{R}_n]]$ (3)

- *Variance of rate*$(\tilde{R}) = \mathrm{E}[Variance[\tilde{R}_n]]$ (4)

Where R_n is the rate trace for user *n*, being $\tilde{R}$ the set of all users' mean rates.

As collected in Table 2, for moving users CQI variability is notable whereas for static ones channel capacity is almost invariable. On the other hand, higher rates are achieved on average when there is no user mobility.

Table 2. Channel transmission rate statistics for static and dynamic (5 km/h) scenarios.

Scenario	Rate statistic	
	Mean (Mbps)	*Variance (Mbps^2)*
Static	3.6798	0.0001
Dynamic	2.8290	2.8057

We define CQI burstlength as the consecutive number of TTIs with the same value of CQI. Fig. 1 shows a sample user's CQI burstlength evolution over time, where channel capacity time-varying behavior can be seen as well as CQI burstlength variability.
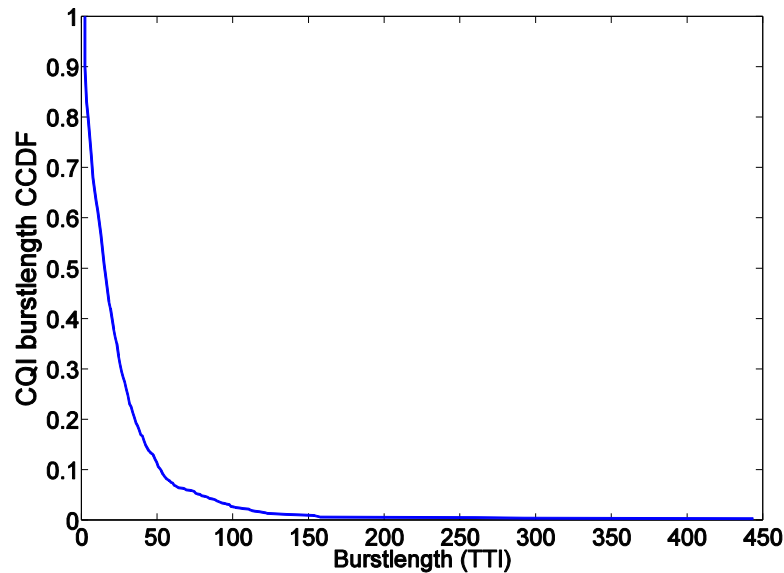

Figure 1. Distribution of CQI burstlengths.

3.3 Scheduling

The system works in discrete time slots in milliseconds order, called Transmission Time Interval (TTI). At the beginning of each time slot the scheduler makes decisions in order to choose the user to transmit. Server preemptive operation mode is considered, offering the possibility to stop giving resources to a user whose service requirement is incomplete.

Depending on the nature of the scheduling policy, the scheduler needs different kind of information in order to make decisions. Concerning size-aware disciplines, per flow attained service information must be saved and updated. Referring to channel-aware policies, channel state information is available to scheduler by means of CQI reports sent from mobile devices after a processing delay of milliseconds. The frequency of receiving channel state feedback depends on the selected CQI reporting rate (CRR), defined as the time period between CQI reports sent by mobile terminal. According to what is shown in Fig. 1, CRR would clearly affect system performance due to the heterogeneity of CQI burstlengths.

In order to study the performance of the system, different scheduling algorithms will be compared. Below we provide a brief description of the most relevant ones, which will be used later in simulations. Table 3 summarizes these disciplines categorization.

- Processor Sharing (PS): Round Robin in practice, where transmission time slots are assigned to each flow in equal portions and in circular order. No size nor channel information

is used at all.

- Least Attained Service (LAS): A non-anticipating size-aware policy that consists on serving the user with the least attained service.

- Gittins: This discipline gives service to the user with the highest Gittins index rule, which is the result of an efficiency function dependant on payoff/investment ratio. As proved in [13], if hazard rate is non-increasing, for a concrete attained service value the Gittins index value and the hazard rate value are the same. This way, for our case study in which we focus on Pareto distributed flow sizes, using Gittins index rule is equivalent to selecting the user with the highest hazard rate value in each decision slot.

- Proportional Fair (PF): A channel-aware policy that serves the user with the highest current transmission rate relative to its current average throughput, which guarantees fairness properties among users.

- Best Rate (BR): A channel-aware discipline that maximizes cell throughput, which consists on serving in each decision slot the user with the highest instantaneous rate.

- Opportunistic Gittins (OG): A mixed approach that takes advantage of both opportunistic gains and flow size properties. As derived in [22] the user with the highest modified Gittins index value is selected. This way, using Opportunistic Gittins index rule in each transmission slot the scheduler allows transmission to user j such that

$$j \in \mathrm{ar\,g\,max}_i \, R_i(t) \cdot h_x(R_i(t) \cdot A_i(t)) \tag{5}$$

Where $A_i(t)$ is the attained service related to user i at time t and $R_i(t)$ is the instantaneous channel capacity of user i at time t.

- BR-LAS: Following SRPT-FM philosophy extrapolated to single channel context, in order to provide a policy that takes the maximum benefit from channel characteristics, an algorithm that uses BR in the first iteration with LAS as tie-breaking rule is defined [18]. This is, the user in best channel conditions is served, and in case of tie, priority is given to the user with the least attained service.

Table 3. Categorization of scheduling algorithms.

Discipline	Property		
	Size-based	*Channel- aware*	*Tie-breaking rule*
PS	✘	✘	✘
LAS	✔	✘	FIFO
Gittins	✔	✘	FIFO
PF	✘	✔	FIFO
BR	✘	✔	FIFO
OG	✔	✔	FIFO
BR-LAS	✘	✔	LAS

4. Mean delay analysis

In this section, the mean delay performance of the scheduling algorithms described in section 3.3 is compared in the wireless paradigm defined in section 3. To that end, several MATLAB-driven simulations have been performed for different network conditions and scenarios. For each configuration 10000 seconds long simulations have been carried out.

The flow size distribution used in simulations is Pareto with parameters $\beta = 2$ and $b = 1.3 \cdot 10^{-6}$, and a mean value, *E[X]*, of 768000 bit (64 packets of 1500 byte). Flow arrival rate, λ, will determine network traffic load, ρ. We define traffic load as $\rho = \lambda \cdot E[S]$, being *E[S]* the mean service time, where $E[S] = E[X] \cdot E[\frac{1}{R}]$. In order to analyze scheduling behavior in different network conditions four network states are considered in our simulations: low load (ρ=0.25), low-medium load (ρ=0.5), medium load (ρ=0.75) and high load (ρ=0.95).

Concerning channel properties, CQI traces obtained from a system level radio access simulator [19] are used in our simulations, which will determine channel instantaneous rate per user in a more realistic fashion (see Table 1 and 2, and Fig. 1). Three scenarios are contemplated in this study. In the first two scenarios CQI reports are available to scheduler every TTI (1 ms). Static users are considered in the first scenario, whereas users in mobility in the second. In the third case study, several simulations are performed for different CRR values for dynamic users case.

Results for different scheduling policies are compared in mean delay ratio terms, relative to the novel OG policy proposed in [17] since it is the only one that combines both size- and channel-awareness in the first iteration of the algorithm. Thus, these normalized results are defined as $\frac{E[T_\pi]}{E[T_{OG}]}$, for π = LAS, Gittins, PS, BR, BR-LAS, PF. This way, ratios lower than one will show that the corresponding scheduling algorithm has better performance than OG in mean delay terms.

4.1 Scenario 1: Static users

In the first family of simulations static users configuration is considered, under network conditions and user characteristics already defined. Results collected for this scenario are summarized in Table 4 for different network loads.

Table 4. Mean delay ratio results for static scenario.

ρ	Mean delay ratio					
	LAS	*Gittins*	*PS*	*BR*	*BR-LAS*	*PF*
0.25	1.000	1.000	1.020	1.059	1.015	1.001
0.5	1.000	1.000	1.016	1.430	1.010	1.008
0.75	1.000	1.000	1.367	1.597	1.197	1.330
0.95	1.000	1.000	1.004	1.104	1.003	1.048

As concluded, almost identical results are obtained for LAS, Gittins and OG (with three decimal precision the same results), which are the best alternatives for static users setting. This clearly reflects that channel-awareness is not useful this case, which is in concordance with the fact that channel capacity variation in such scenario is almost null. Related to this, results of channel-aware disciplines are worse than size-based ones, which shows again the little importance of opportunistic gains for this situation. Moreover, quite similar results are achieved for PF and the non-channel-aware PS policy.

On the other hand, the performance improvement due to the new tie-breaking rule proposed is notable. BR in combination with a size-based tie-breaking rule gives better results than pure BR, which shows once again the importance of using size information when taking scheduling decisions for static users case.

4.2 Scenario 2: Users in mobility and ideal channel-awareness

Now we consider the case of users moving at 5 km/h in the cell, speed that corresponds to a user walking. In such situation, channel capacity notably varies over time. Results for this scenario are collected in Table 5.

The worst performance is given by pure size-aware policies, which considerably differ from OG. The non-channel-aware and non size-based PS discipline, although better than the previous ones, is also unsuccessful.

Undoubtedly, the most appropriate disciplines in this time-varying channel context are the pure channel-aware ones, which all of them behave better than OG. Quite similar results are achieved for both versions of BR and PF, being PF better for high ρ values.

Table 5. Mean delay ratio results for dynamic scenario.

ρ	Mean delay ratio					
	LAS	*Gittins*	*PS*	*BR*	*BR-LAS*	*PF*
0.25	350100	350000	1.5	0.99	0.98	0.98
0.5	4000	4000	50	0.70	0.69	0.68
0.75	3700	3700	100	0.69	0.68	0.67
0.95	2875	2800	105	0.71	0.69	0.57

Otherwise, no too much improvement is got when using LAS as tie-breaking rule instead of FIFO for BR discipline, which shows again that for the scheduler channel information is much more important than size information in such context. Nevertheless, when channel instantaneous capacity is not available for the scheduler in every TTI this behavior changes, just as described in next subsection.

4.3 Scenario 3: CQI reporting rate analysis

Now we consider the same scenario as in previous subsection, but considering that channel instantaneous information is less frequent and delayed for the scheduler. This is fundamental in current real networks, since massive signaling overhead in the uplink caused by CQI reports could negatively affect network performance. In such context, several simulations have been carried out for different CRR values from 1 ms to 100 ms. It is assumed that as long as no new CQI report is received from mobile terminal, the scheduler uses the last one received. Channel-aware policies are only studied here because channel-unaware ones are unaffected by CRR value.

The results of this section are presented in Table 6. Similar behavior is achieved for different CRR values for all the network loads considered, and that is why results for the worst network state are shown only.

Although the performance of both two BR versions and PF is superior to OG for CRR values lower than 20 ms, for network performance standpoint these reporting rates are completely unsuitable. For appropriate CRR values range, above 20 ms, OG is the best option, which guarantees a trade-off between scheduling performance and reducing CQI reports signaling uplink overhead.

Table 6. Mean delay ratio results for dynamic scenario for different CRR values (ρ=0.95).

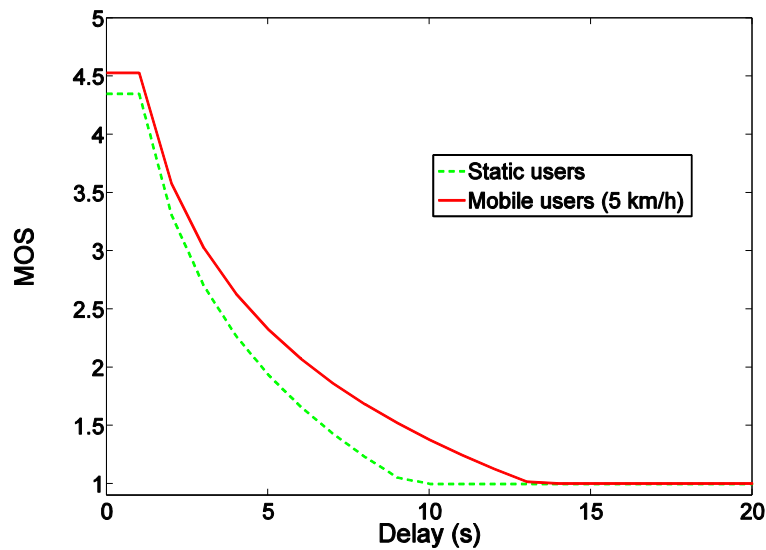| CRR (ms)	Mean delay ratio (ρ=0.95)		
	BR	BR-LAS	PF
1	0.71	0.69	0.54
5	0.62	0.61	0.59
10	0.97	1.02	0.81
20	1.39	1.29	1.51
30	1.38	1.23	1.41
50	1.17	1.11	1.29
100	1.19	1.15	1.31

Figure 2. Evolution of MOS vs. delay for static and dynamic users scenarios.

5. Subjective quality evaluation for mobile web browsing

Until now the performance of different scheduling algorithms in mean delay terms has been analyzed. In this section, we will study the impact of different scheduling disciplines on QoE for web browsing service in the wireless paradigm discussed in previous sections.

5.1 Subjective tests results

In this section, a mapping from delay to user's perception is derived in the scenarios previously analyzed. In order to do this, subjective tests results obtained in [23] are applied to our case of study.

In the carried out experiment in [23] 36 people were consulted. Participants were requested to search for a given web page, and then the result was displayed after a certain controlled delay. In our case, this delay would emulate flow transfer delay due to the multi-user resource allocation task. After each search presentation, subjects were asked to evaluate their satisfaction in Mean Opinion Score (MOS) scale of 1 (very poor quality) to 5 (excellent quality). In order to participants become familiar to the experiment instructions were e-mailed to them in advance, as well as a demo and training phase was provided before collecting their opinion. After the preparation phase, participants went through 3 rounds of 6 searches (of 49 randomized and emulated sessions).

On the other hand, in order to determine the effect of different network environments three delay scales were used representing fast, moderate and slow network contexts. We assume that results from fast network correspond to static users scenario, and those from moderate network to mobile users case. In Fig. 2 we present graphically MOS results concluded from subjective tests applicable to our study cases, where non-expert users are considered.

Obtained results are approximated by means of a logarithmic fitting, see expression (6). For static users case *a* and *b* coefficients are 1.5 and 4.35 respectively, whereas for dynamic users case 1.37 and 4.53. This formula will be used later to compute the overall mean MOS.

$$MOS = b - a \cdot \log(delay) \tag{6}$$

5.2 QoE analysis

We consider a wireless system resource allocation scheme aimed at maximizing the mean overall MOS. The objective function for such optimization problem is expressed as:

$$Average\ MOS(\tilde{D}) = \frac{1}{N} \cdot \sum_{n=1}^{N} MOS_n(delay_n) \tag{7}$$

where N is the total number of mobile users in the system, being $\tilde{D}$ the set of all users' delays, $delay_n$ the delay of user n and MOS_n the QoE metric of user n caused by $delay_n$ which is computed by (6).

Among all the elastic traffic present in the system we are only interested in this representing a web browsing search as in afore-described subjective tests. This way, flows smaller than 500 Kbyte are only taken into account for mean MOS computation.

In the following the impact of the scheduling algorithms described in section 3.3 on web browsing QoE is analyzed for the three scenarios discussed in section 4.

5.2.1 Static users case

Fig. 3 shows the average MOS for different network conditions and scheduling algorithms when there are no users in movement in the cell. The best results are achieved for the size-aware policies considered, both for pure size-based ones, such as LAS and Gittins, and the mixed approach OG. Nevertheless, BR-LAS policy gets very good performance in QoE terms, which comparing to BR a notable improvement is achieved by using LAS as tie-breaking rule.
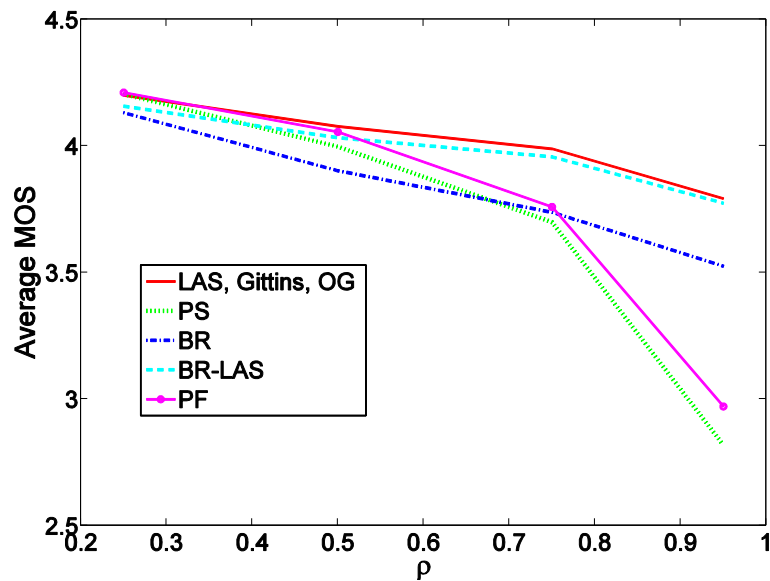
Figure 3. Average MOS comparisson for different scheduling policies in static users case.

Similar to the results concluded for mean delay metric for static users context, in order to maximize QoE in the cell, size information is also more important than channel awareness when taking scheduling decisions.

5.2.2 Users in mobility and ideal channel-awareness

From Fig. 4 we observe that when there is a complete knowledge of channel variations QoE performance behavior is alike the achieved in mean delay terms. Service is unaffordable for pure size-based policies, and quality degradation is appreciated for PS as traffic load increases. In general, pure channel-aware algorithms are the best alternatives.

5.2.3 CRR analysis

As seen in Fig. 5, contrary to mean delay optimization case, OG is not the best choice in QoE terms when CQI reports signaling limitation is taken into account.

For acceptable CRR values, BR policy is superior to the other channel-aware ones. Moreover, even better results are achieved for BR with the new tie-breaking rule proposal.

Therefore, in this context, BR-LAS is the best option, which guarantees a trade-off between QoE maximization and CQI signaling overhead minimization.
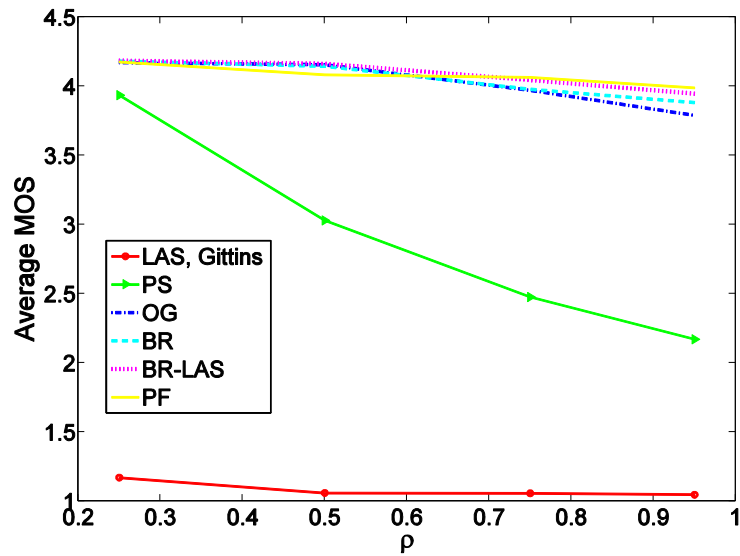
Figure 4. Average MOS comparisson for different scheduling policies for users moving at 5 km/h.
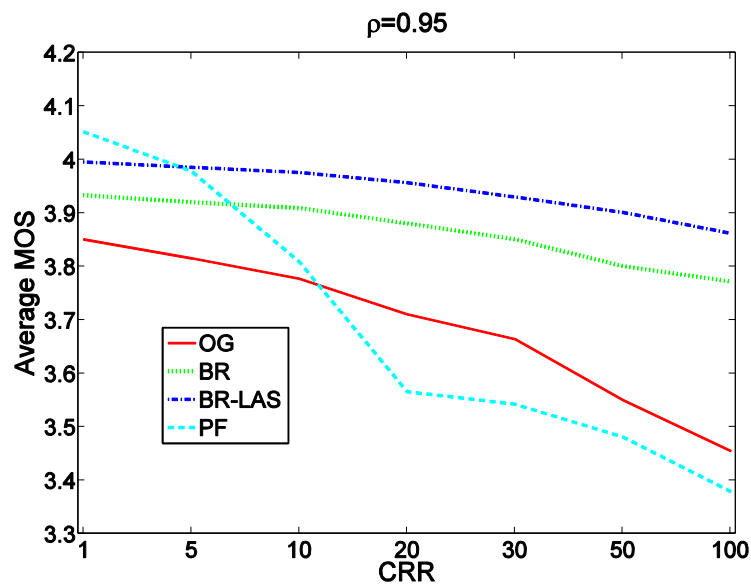

Figure 5. Average MOS comparisson for different scheduling policies using different CRR.

6. Conclusions

This paper deals with resource allocation problem for elastic traffic in wireless downlink channels. In such paradigm, the flow-level performance of scheduling algorithms aimed at minimizing delay in the cell is analyzed not only in mean delay terms but also on web browsing mean QoE terms.

A key aspect for improving scheduling mechanisms in a time-varying channel capacity context, consists on combining the use of opportunistic gains and short flows' priorization. This way, BR-LAS scheduling discipline is considered, which instead of taking into account

size and channel-awareness together (OG), it benefits from opportunistic gains combined with a size-based tie-breaking rule.

In order to compare the mean delay performance of the most known and novel scheduling strategies with our proposal, several simulations have been performed under different network load conditions and CQI reporting rates, for static and dynamic users cases. As discussed in section 5, pure size-aware scheduling policies are the most suitable when channel capacity is almost constant as when there are static users in the cell. Considering dynamic users (5 km/h) in the cell, OG is the best option which guarantees a trade-off between mean delay minimization and CQI signaling overhead reduction.

For quantifying mobile web QoE, a mapping from delay to user satisfaction is derived from subjective tests results and applied to our cases of study. As concluded in section 5, considering only interactive users, for static users scenario size-based policies give the best results in QoE terms. However, for dynamic users case, BR-LAS is the best option to maximize web browsing QoE under suitable CQI reporting rate conditions.

Therefore, this analysis will be very useful for network operators in order to choose the most appropriate scheduling strategy in each situation. This decision will depend on the type of users in the cell, as well as whether the objective is to minimize overall mean delay in the cell for all services or just to maximize QoE for a concrete service such as web browsing.

As an open issue, the way of sending CQI reports when reducing reporting rate could be considered. Instead of having at the scheduler delayed channel instantaneous information, channel state information averaged over a time period or other channel statistics should be sent by mobile terminal in order to improve system performance. In future work, it would be interesting to study the performance of the analyzed scheduling algorithms in multichannel contexts, as it happens in next generation networks, where multiple users can be served in parallel.

Acknowledgement

References

[1] Aalto S. and Lassila, P., "Flow-level stability and performance of channel-aware priority-based schedulers", 6th EURO-NF Conference on Next Generation Internet (IEEE NGI 2010). http://dx.doi.org/10.1109/NGI.2010.5534476

[2] Basukala, R., Ramli HAM, Sandrasegaran K. and Chen L., "Impact of CQI feedback rate/delay on scheduling video streaming services in LTE downlink", 12th IEEE International Conference on Communication Technology (IEEE ICCT), 2010. http://dx.doi.org/ 10.1109/ICCT.2010.5689046

[3] Avrachenkovt K., Ayesta U., Brown P. and Nyberg E., "Differentiation between short and long TCP flows: Predictability of the response time", 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM 2004). http://dx.doi.org/10.1109/INFCOM.2004.1356965

[4] Kleinrock L., *Queueing Systems: Volume 2: Computer Applications*, Vol. 82, 1976, John Wiley&Sons.

[5] Ayesta U., Erausquin M., Jonckheere M. and Verloop IM., "Scheduling in a random environment: stability and asymptotic optimality", *Arxiv* preprint arXiv:1101.5794, 2011.

[6] ITU-T, P.10/G.100 (2006) Amendment 1 (01/07): New Appendix I - Definition of Quality of Experience (QoE), 2007.

[7] Khan S., Thakolsri S., Steinbach E. and Kellerer, W., "QoE--based cross--layer optimization for wireless multiuser systems", 18th ITC Specialist Seminar on Quality of Experience, 2008.

[8] Thakolsri S., Khan S., Steinbach E. and Kellerer W., "QoE-driven cross-layer optimization for high speed downlink packet access", *Journal of Communications*, Vol. 4, Issue 9, pp. 669-680, 2009.

[9] Ameigeiras P., Ramos-Munoz J.J., Navarro-Ortiz J., Mogensen P. and Lopez-Soler J.M., "QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems", *Computer Communications*, Vol. 33. Issue 5, pp. 571-582, 2010, Elsevier. http://dx.doi.org/10.4304/jcm.4.9.669-680

[10] Ayesta U., Erausquin M. and Jacko P., "A modeling framework for optimizing the flow-level scheduling with time-varying channels", *Performance Evaluation*. Vol. 67, Issue 11, 2010, pp. 1014--1029. Elsevier Science Publishers BV. http://dx.doi.org/10.1016/j.peva.2010.08.015

[11] Ayesta U., Erausquin M. and Jacko P., "Resource Sharing in a Single Server with Time-Varying Capacity", 49th Annual Conference on Communication, Control and Computing (IEEE Allerton 2011) . http://dx.doi.org/10.1109/Allerton.2011.6120192

[12] Thompson K., Miller G.J. and Wilder R., "Wide-area Internet traffic patterns and characteristics". IEEE Network, Vol. 11, Issue 6, pp. 10--23. 1997. http://dx.doi.org/10.1109/65.642356

[13] Gittins J.C., Weber R. and Glazebrook K.D., *Multi-armed bandit allocation indices*, 2011, Wiley Online Library.

[14] Aalto S. and Ayesta U., "Optimal scheduling of jobs with a DHR tail in the M/G/1 queue", Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools 2008, ICST Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. http://dx.doi.org/10.4108/ICST.VALUETOOLS2008.4335

[15] Aalto S., Ayesta U. and Righter R., "Properties of the Gittins index with application to

optimal scheduling", *Probability in the Engineering and Informational Sciences*, Vol. 25, Issue 3, pp. 269--288, 2011, Cambridge Univ. Press. http://dx.doi.org/10.1017/S0269964811000015

[16] Yashkov SF., "On feedback sharing a processor among jobs with minimal serviced length", Technika Sredstv Svyazi. Ser. ASU, Vol. 2, pp. 51—62, 1978.

[17] Taboada I., Fajardo, J. O., Liberal, F. and Blanco B., "Size-based and channel-aware scheduling algorithm proposal for mean delay optimization in wireless networks", IEEE ICC 2012 – Workshop on Smart Communication Protocols and Algorithms year, June 2012.

[18] Pinedo M., "Scheduling: theory, algorithms, and systems", 2008, Springer Verlag.

[19] Ikuno J.C., Wrulich M. and Rupp M., "System level simulation of LTE networks", Proc. IEEE 71st Vehicular Technology Conference, 2010. http://dx.doi.org/10.1109/VETECS.2010.5494007

[20] TS-36.213 v9.3.0, Technical Specification Group Radio Access Network Evolved Universal Terrestrial Radio Access (E-UTRA), Physical layer procedures, 2010.

[21] TS-36521 v9.4.1, Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance testing, 2011.

[22] Lassila P. and Aalto S., "Combining opportunistic and size-based scheduling in wireless systems", ACM Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems, 2008. http://dx.doi.org/10.1145/1454503.1454558

[23] Ibarrola E., Liberal F., Taboada I. and Ortega R., "Web QoE Evaluation in Multi-Agent Networks: Validation of ITU-T G. 1030", 5th International Conference on Autonomic and Autonomous Systems, (ICAS 2009). http://dx.doi.org/10.1109/ICAS.2009.40

Copyright Disclaimer