On the Influence of Packet Scheduling on the Trade-Off between System Spectral Efficiency and User Fairness in OFDMA-Based Networks

Emanuel B. Rodrigues[1], Michael L. Walker[2], and Fernando Casadevall[1]

[1] Universitat Politècnica de Catalunya (UPC),
Departament de Teoria del Senyal i Comunicacions (TSC),
Barcelona, Spain
{emanuel,ferranc}@tsc.upc.edu
[2] Università degli Studi di Padova,
Dipartimento di Ingegneria dell'Informazione,
Padova, Italy
michaellee.walker@studenti.unipd.it

Abstract. System spectral efficiency and user fairness are crucial aspects for resource allocation in multi-user OFDM-based cellular networks. This work intends to investigate the influence of the performance of packet scheduling algorithms on the trade-off between these two objectives in scenarios with non real-time and real-time services. By means of system-level simulations, we were able to create a didactic map of the relation between these two aspects and propose ways to exploit this trade-off efficiently.

Keywords: Packet Scheduling, spectral efficiency, fairness, OFDMA.

1 Introduction

The wireless shared channel in cellular networks is a medium over which many Mobile Terminals (MTs) compete for resources. In such a scenario, spectral efficiency and fairness are crucial aspects for resource allocation. From a cellular operator perspective, it is very important to use the channel efficiently because the available frequency spectrum is scarce and the revenue must be maximized. From the users' point of view, it is more important to have a fair resource allocation so that they can meet their Quality of Service (QoS) requirements and maximize their satisfaction. The time-varying nature of the wireless environment, coupled with different channel conditions for different MTs, poses significant challenges to accomplishing these goals. In general, these objectives cannot be achieved simultaneously and an efficient trade-off must be achieved. In recent years, Radio Resource Management (RRM) has been envisaged as one of the most efficient techniques to achieve a desirable trade-off among these two conflicting objectives in cellular multi-carrier systems.

Many next generation wireless systems are based on Orthogonal Frequency Division Multiple Access (OFDMA), which provides a high degree of flexibility

that can be exploited by RRM algorithms. There are different sources of diversity in an OFDMA-based system, such as time, frequency and multi-user diversities. Thus, it is possible to dynamically allocate subsets of sub-carriers for different MTs, and to adapt the Modulation and Coding Scheme (MCS) and power for each sub-carrier according to the instantaneous channel conditions. Multi-carrier Packet Scheduling (PS) is an RRM strategy that assigns the sub-carriers to the users based on priority functions that can take into account channel- and user-related information, such as channel gains, QoS metrics, buffer occupancy, etc.

Some works in the literature tried to find an efficient trade-off between system capacity and fairness in OFDMA networks based on cross-layer optimization [1,2,3], utility theory [4] or both of them [5]. The only works that have used a well defined methodology for analyzing the fairness were [1] and [4], while the others made the evaluation implicitly comparing QoS metrics. The advantage of [1] compared to the others is that it used an intuitive and easy way to assess the fairness among the users by means of a fairness index. The present paper extends the work of [1], emphasizing and explicitly showing the influence of PS algorithms on the trade-off existent between spectral efficiency and fairness on OFDMA cellular systems considering scenarios with Non-Real Time (NRT) or Real Time (RT) services.

The paper is organized as follows. Section 2 presents the system modelling. In section 3, we show the PS algorithms studied in this contribution, while section 4 depicts the simulation results. Finally, the conclusions are drawn in section 5.

2   System Model

The considered scenario is a single cell with hexagonal shape. We consider a network with one transmitter (Base Station (BS)) and J receivers (MTs). The transmitted Orthogonal Frequency Division Multiplexing (OFDM) signal is time-slotted, where in every time slot at most one user can be served over each sub-carrier.

The considered environment is Typical Urban (TU) [10] where each user experiences independent transmit conditions. The channel has a frequency-selective Rayleigh fading, with the channel coherence time such that each sub-carrier experiences only flat fading. It is assumed that the channel fading rate is slow enough so that the frequency response does not change during a Transmission Time Interval (TTI). Each user also experiences shadowing with log-normal distribution. A perfect knowledge of the Channel State Information (CSI) at the transmitter side is assumed, with no signaling overhead transmitted. The signal strength at the receiver side depends on the path-loss calculated by: $L = 128.1 + 37.6 \log_{10} d$, where d is the distance to the BS in km. It was assumed that the MTs remained stationary, hence there is no need to implement any handover scheme.

Regarding the power allocation strategy, we assume that the total BS transmission power is equally divided among all sub-carriers. The bit allocation on each sub-carrier is determined using the modified Shannon's capacity model [5]:

$c_{j,k} = \log_2 (1 + \Gamma p_k \rho_{j,k})$, where $c_{j,k}$ is the achievable throughput of user j over sub-carrier k, p_k is the transmit power allocated at sub-carrier k, $\rho_{j,k}$ is the Signal-to-Noise Ratio (SNR) of user j at sub-carrier k, and Γ is the SNR gap given by $\frac{1.5}{-\ln 5BER}$ [5] (the target Bit Error Rate (BER) was 10^{-6}). Since we used discrete modulations QPSK, 16-QAM and 64-QAM, we made an appropriate integer quantization of $c_{j,k}$. Assuming that a sub-carrier set $\mathcal{K}_j$ is assigned to user j, its transmission rate is calculated as $r_j = \sum_k r_{j,k} = \sum_k c_{j,k} \cdot \Delta f$, where Δf is the sub-carrier bandwidth and $k \in \mathcal{K}_j$.

Regarding the traffic models, all NRT users are assumed to have an infinite amount of data to transmit during the whole simulation run (full-buffer model). As an example of RT service we considered Voice over IP (VoIP). This model follows an ON-OFF pattern, where each ON and OFF period duration are exponentially distributed. During the ON period, the transmitter generates one packet with fixed size of 32 bytes every voice frame. In the case of the 3rd. Generation Partnership Project (3GPP) Adaptive Multirate (AMR) voice codec, the frame duration is 20 ms. If any packet arrives at the receiver with a delay higher than 100 ms, it is discarded. In order to simulate a scenario with only VoIP users, we would need to consider a huge amount of users so that our high capacity OFDMA system becomes loaded, which would be unfeasible in terms of computational cost. In order to solve this problem we made two assumptions: i) consider 100% of voice activity, and ii) decrease the packet inter-arrival time to 2 ms. These two assumptions can be justified if we consider that each real VoIP user has an associate cluster of C virtual users that will be responsible to generate the traffic. These virtual users are ideally located in the same position of the real user and the difference in their propagation gains is assumed negligible.

3   Packet Scheduling Algorithms

In the following, the formulation of the PS techniques studied in this paper is presented. Sections 3.1 and 3.2 present the algorithms suitable for NRT and RT services, respectively.

3.1   Non Real-Time Services

In order to study the trade-off between system capacity and user fairness in a scenario with a NRT service, we evaluated the Max-Rate (MR), Max-Min Fairness (MMF) and Proportional Fairness (PF) PS algorithms, whose mathematical formulations are presented in equations 1, 2 and 3, respectively. The MT j^* is chosen to transmit on the kth sub-carrier in TTI n if it satisfies the condition given by the corresponding equation:

$$j^* = \arg\max_j \left\{ r_{j,k}[n] \right\}, \quad \forall j \tag{1}$$

$$j^* = \arg\max_j \left\{ \frac{1}{T_j[n-1]} \right\}, \quad \forall j \tag{2}$$

$$j^* = \arg\max_j \left\{ \frac{r_{j,k}[n]}{T_j[n-1]} \right\}, \quad \forall j \tag{3}$$

where $r_{j,k}[n]$ is the achievable data rate of the jth MT on the kth sub-carrier in TTI n and $T_j[n-1]$ is the average throughput of the jth MT calculated up to TTI $n-1$. The throughput of the jth MT is averaged using a Simple Exponential Smoothing (SES) filtering, as indicated in equation 4.

$$T_j[n] = (1-\lambda) \cdot T_j[n-1] + \lambda \cdot r_j \tag{4}$$

where r_j is the instantaneous achievable data rate of the jth MT and λ is the filtering constant.

The MR scheduling policy was firstly presented in [6]. With this strategy, a specific sub-carrier is assigned to the MT with the best channel quality (i.e. the highest achievable bit-rate) on that sub-carrier. This scheduling policy provides the maximum cell throughput at the expense of lower throughput-based fairness, since the MTs in bad fading conditions would not be chosen for transmission.

The MMF PS gives priority to the MT that has experienced the worst through-put so far [7]. In this way, in terms of throughput, it is the most fair criterion possible, since all MTs will have approximately the same throughput in the long-term. This high fairness will be achieved at the expense of low spectrum efficiency, caused by the MTs with poor channel quality. One can notice that in equation 2, the achievable bit-rate on the sub-carriers is not present. Therefore, in each TTI, all the sub-carriers will be assigned to the chosen MT j^*, giving to MMF a Time Division Multiple Access (TDMA) behavior. This fact is enforced by considering that in the model assumed in this work the NRT user always has data to transmit.

Finally, the PF strategy takes into account both the instantaneous channel conditions and the average throughput of the MTs [8]. In this way, it is a trade-off between the spectral efficiency and throughput-based fairness achieved by MR and MMF.

3.2   Real Time Services

In order to investigate the trade-off in an OFDMA system with RT service, we chose to evaluate the PF, Delay-Based First In First Out (D-FIFO) and Modified Largest Weighted Delay First (M-LWDF) PS algorithms, whose priority policies are presented in equations 5, 6 and 7, respectively. The MT j^* is chosen to transmit on sub-carrier k at TTI n if it satisfies the condition given by the corresponding equation:

$$j^* = arg\max_j \left\{ \frac{r_{j,k}[n]}{T_j[n-1]} \right\}, \quad \forall j \tag{5}$$

$$j^* = arg\max_j \left\{ D_j^{hol}[n] \right\}, \quad \forall j \tag{6}$$

$$j^* = arg\max_j \left\{ \frac{r_{j,k}[n]}{T_j[n-1]} \cdot D_j^{hol}[n] \right\}, \quad \forall j \tag{7}$$

where $D_j^{hol}[n]$ is the delay experienced by the Head Of Line (HOL) packet of user j at TTI n, while $r_{j,k}[n]$ and $T_j[n-1]$ are defined and calculated as in section 3.1.

Among the three algorithms presented in this section, PF is expected to be the most unfair in terms of delay, and the most efficient in terms of resource usage, because it is not directly influenced by the delay metric and takes into account the channel quality. PF was chosen instead of MR because the former is expected to perform better than the latter for delay-sensitive RT services.

D-FIFO gives priority to the MT with the highest HOL packet delay. In this way, in terms of delay, it is the most fair criterion possible. Like the MMF criterion, D-FIFO also does not use CSI of the sub-carriers, ignoring the frequency diversity offered by the OFDMA system. Therefore, D-FIFO assumes a TDMA behavior, giving to the user chosen for transmission at TTI n the right to transmit over all the sub-carriers. This strategy is not efficient in the resource usage, so it is expected to provide lower system throughput.

The M-LWDF algorithm [9] is a trade-off between PF and D-FIFO, since it should provide intermediate delay-based fairness and intermediate resource usage efficiency.

4   Simulation Results

Similar to [1], the present work uses a well defined methodology to evaluate the fairness among the users based on the definition of the user and system fairness indexes. This paper extends the work in [1] proposing fairness indexes suitable for RT services. Our definition of fairness is based on QoS, so the fairness indexes for NRT and RT services are based on session throughput and delay, respectively.

To calculate the fairness of the system, we first have to define the fairness indexes related to the NRT and RT users, as presented in equation 8 below.

$$\phi_j^{NRT}[n] = \frac{T_j[n-1]}{T_j^{req}}; \qquad \phi_j^{RT}[n] = \frac{D_j[n]}{D_{req}} \qquad (8)$$

where T_j^{req} is the throughput requirement of the jth NRT user and D_{req} is the maximum allowable delay for the RT service. $D_j[n]$ is calculated through a SES filtering as shown in equation 9 below.

$$D_j[n] = (1-\tau) \cdot D_j[n-1] + \tau \cdot \overline{D_j^{inst}}[n] \qquad (9)$$

where $\overline{D_j^{inst}}$ is the instantaneous mean delay of the packets present in the buffer of user j at TTI n.

The fairness indexes $\phi_j^{NRT}[n]$ and $\phi_j^{RT}[n]$ are calculated at each TTI and for each user depending on his type of service. The overall fairness index related to the system is defined as:

$$\Phi^{NRT}[n] = \frac{\left(\sum_{j=1}^{J} \phi_j^{NRT}[n]\right)^2}{J \cdot \sum_{j=1}^{J} \left(\phi_j^{NRT}[n]\right)^2}; \qquad \Phi^{RT}[n] = \frac{\left(\sum_{j=1}^{J} \frac{1}{\phi_j^{RT}[n]}\right)^2}{J \cdot \sum_{j=1}^{J} \left(\frac{1}{\phi_j^{RT}[n]}\right)^2} \qquad (10)$$

where J is the number of MTs in the cell, and $\phi_j^{NRT}[n]$ and $\phi_j^{RT}[n]$ are calculated in equation 8. Notice that $\frac{1}{J} \leq \Phi \leq 1$. A perfect fair allocation is achieved when $\Phi = 1$, which means that the throughput or delay experienced by the MTs are equally proportional to their throughput/delay requirements (all user fairness indexes are equal). The worst allocation occurs when $\Phi = \frac{1}{J}$, which means that all sub-carriers were allocated to only one MT. In this way, the overall fairness indexes $\Phi^{NRT}[n]$ and $\Phi^{RT}[n]$ are intuitive and easy ways to assess the fairness among the users in their respective service classes.

Table 1. Simulation parameters

Parameter	Value	Unit
Number of cells	1	-
BS transmission power	5	W
Cell radius (R)	500	m
Number of sub-carriers (K)	192	-
Sub-carrier bandwidth (Δf)	15	KHz
Transmission Time Interval	0.5	ms
Shadowing standard deviation	8	dB
Noise power per sub-carrier	-123.24	dBm
BER requirement	10^{-6}	-
Modulation schemes	QPSK, 16-QAM, 64-QAM[a]	-
Throughput requirement (T_j^{req})	1.4	$Mbps$
Delay Requirement (D_{req})	100	ms
Cluster load for RT model (C)	10	-
Throughput filtering constant (λ)	0.1	-
Delay filtering constant (τ)	0.1	-
Fairness factor filtering constant	0.1	-
Simulation time span	15	s
Number of realizations for each point	10	-

[a] Only QPSK and 16-QAM were used for the simulations with RT services.

The parameters used in the simulation campaign are depicted in table 1. We divide the presentation of the results in two parts: NRT and RT in sections 4.1 and 4.2, respectively. In order to evaluate the efficiency in the resource usage, we use the total cell throughput metric. The QoS-based fairness metric presented in the graphics is the mean system fairness index: the values of the fairness indexes $\Phi^{NRT}[n]$ and $\Phi^{RT}[n]$ were calculated using equation 10, filtered with an exponential filter and averaged over the whole simulation. In order to extend the fairness analysis, we equally separate the users in two groups of same size, inner and outer users, based on path loss and shadowing. The former are the ones closest to the BS (inner zone of the cell) that experience better channel conditions, while the latter are those that are far from the BS (outer zone of the cell).

4.1   Non Real-Time Services

In Fig. 1(a), the throughput-based mean system fairness index is shown. As
expected, MMF is the most fair algorithm, while MR is the most unfair. PF
showed an intermediate behavior, confirming to be a trade-off. The performance
of PF was closer to MMF than to MR, with a fairness index around 0.9.

Fig. 1(b) depicts the total cell throughput, which is a measure of the system
spectral efficiency. As expected, MR is the algorithm with the highest perfor-
mance, while MMF is not able to exploit efficiently the available resources. Again,
PF presented an intermediate performance. Looking at Figs. 1(a) and 1(b), one
can clearly see the conflicting objectives of capacity and fairness maximization,
and how MR and MMF are able to achieve one objective in detriment of the
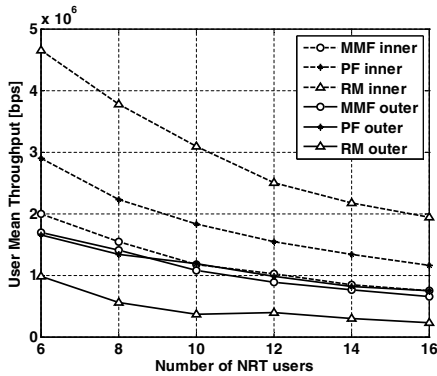other. PF is an exception because it is able to find a trade-off.

The mean throughput of the users classified as inner (good) and outer (bad)
users is shown in Fig. 1(c). This analysis is very important to give us an insight
about the distribution of fairness in the system. It is interesting to focus on the
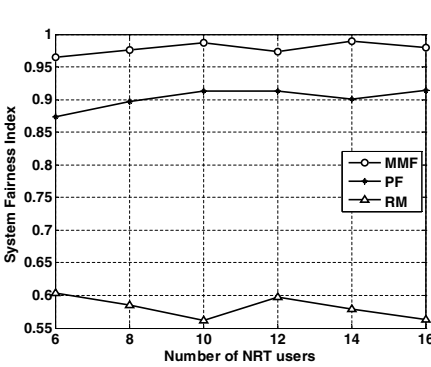
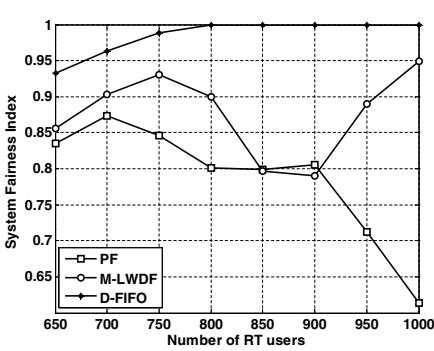(a) Throughput-based fairness
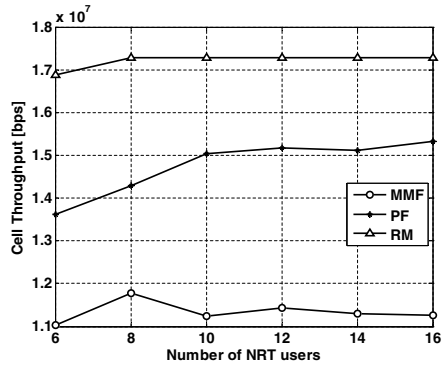
(b) Cell throughput

(c) Mean throughput of inner and outer
users

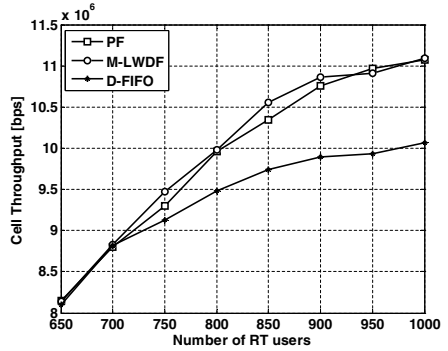**Fig. 1.** Trade-off analysis for NRT services

difference between the inner and outer curves within the same algorithm in Fig.
1(c). For instance, if we consider MR, the difference between the throughput
curves is the biggest. This means that MR assigns more priority to inner users
than to outer users, producing a sensible decrease in fairness (see Fig. 1(a)). On
the other extreme, MMF is the algorithm with the smallest difference between
inner and outer curves, which proves that it is the most fair algorithm is terms
of user throughput. Since PF is a trade-off, it presents an intermediate behavior.

### 4.2   Real Time Services

The delay-based mean system fairness index is shown in Fig. 2(a). D-FIFO is
the algorithm with the highest fairness because it gives strict priority to the
users with higher queuing delay. On the other hand, since PF does not take
into account the delay, it presents the lowest values of the fairness index, which
decreases monotonically when the number of users increases. M-LWDF takes
into consideration in its scheduling policy both the channel quality and the
delay, which provokes a varying behavior of the fairness in the considered range

(a) Delay-based fairness

(b) Cell throughput
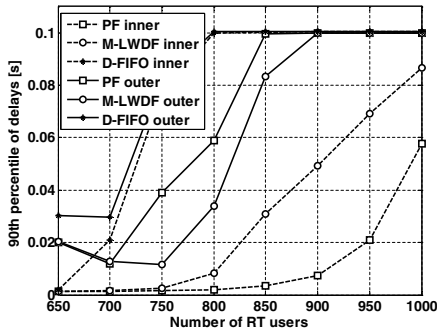
(c) 90th percentile of the delays of the
inner and outer users

Fig. 2. Trade-off analysis for RT services

of loads. As expected, the values of the M-LWDF fairness index were inside the range of values of D-FIFO and PF.

In Fig. 2(b) the cell throughput is given. As expected, the best and worst performance were presented by PF and D-FIFO, respectively. This is explained by the fact that the former gives more importance to the channel quality, and so the resources are used more efficiently, while the latter only takes the delay information into account, which may lead to an inefficient resource allocation. A surprising result was obtained for M-LWDF, since it performs as good as PF, providing high cell throughput. Looking at Figs. 2(a) and 2(b), one can conclude that M-LWDF can achieve a good trade-off between resource efficiency and user fairness in a scenario with RT users.

Fig. 2(c) presents the 90th percentile of the packet delays of inner and outer users. Comparing Figs. 2(c) and 2(a), one can see that the difference in performance between the two groups using the same PS algorithm has an intrinsic relation with the fairness. In the case of PF, the difference between the curves is the greatest, which indicates that PF is giving considerably more priority to the inner users than it does for the outer ones, providing the lowest fairness. For D-FIFO the difference between the curves is the smallest, so the scheduler is giving almost the same priority to both groups, yielding the highest fairness. Since M-LWDF is a trade-off algorithm, it has an intermediate behavior. Although D-FIFO takes into account the delay, it is the one that presents the highest 90th percentile of the packet delays. This shows that the fact of not exploiting the OFDMA diversities is not beneficial in terms of QoS. Furthermore, when the system load increases, it causes the system to become stuck, i.e. the majority of the packets are discarded because they have a delay greater than 100 ms. This can be seen in Fig. 2(a), where the fairness index of D-FIFO is 1 for a system load higher than 800 users. This higher fairness is provided at the expense of very poor performance in terms of QoS.

Table 2. Relations between PS algorithms, spectral efficiency and fairness

	NRT services	RT services
RM	High spectral efficiency and low throughput-based fairness	-
MMF	Low spectral efficiency and high throughput-based fairness	-
PF	Trade-off between efficiency and fairness for NRT services	-
PF	-	High spectral efficiency and low delay-based fairness
D-FIFO	-	Low spectral efficiency and high delay-based fairness
M-LWDF	-	Trade-off between efficiency and fairness for RT services

5   Conclusions

In this work we have evaluated PS algorithms in an OFDMA-based system suitable for NRT or RT services, and investigated their influence on the trade-off between the efficient usage of the resources and the fairness among the users. Analyzing the simulation results we can draw a didactic map of the relations between the two objectives mentioned above, which can be seen in table 2.

We can conclude that there are PS algorithms that are able to find this trade-off, such as PF and M-LWDF for the cases of NRT and RT services, respectively. However, these algorithms are only able to provide a static trade-off. It would be interesting to propose PS algorithms able to provide a dynamic and controllable trade-off according to the cellular operator's objectives. This is currently being developed as the next step of this investigation.

References

1. Rodrigues, E.B., Casadevall, F.: Adaptive Radio Resource Allocation Framework for Multi-User OFDM. In: IEEE 69th Vehicular Technology Conference - VTC Spring, Barcelona, Spain (2009)
2. Hou, H., Zhou, W., Zhou, S., Zhu, J.: Cross-Layer Resource Allocation for Heterogeneous Traffics in Multiuser OFDM Based on a New QoS Fairness Criterion. In: IEEE 66th Vehicular Technology Conference - VTC Fall, pp. 1593–1597 (2007)
3. Shuang, W., Youjun, G., Xuelin, G., Hui, T., Ping, Z.: Packet Scheduling for Multimedia Traffics in Downlink Multi-User OFDM Systems. In: International Conference on Wireless Communications, Networking and Mobile Computing - WiCOM, pp. 1–4 (2006)
4. Lei, H., Zhang, L., Zhang, X., Yang, D.: A Packet Scheduling Algorithm Using Utility Function for Mixed Services in the Downlink of OFDMA Systems. In: IEEE 66th Vehicular Technology Conference - VTC Fall, pp. 1664–1668 (2007)
5. Song, G., Li, Y.G.: Cross-Layer Optimization for OFDM Wireless Networks - Part I: Theoretical Framework and Part II: Algorithm Development. IEEE Transactions on Wireless Communications 4(2), 614–634 (2005)
6. Jang, J., Lee, K.B.: Transmit Power Adaptation for Multiuser OFDM Systems. IEEE Journal on Selected Areas in Communications 21(2), 171–178 (2003)
7. Ameigeiras, P.J.G.: Packet Scheduling and Quality of Service in HSDPA, Ph.D. dissertation, Aalborg University, Aalborg, Denmark (2003)
8. Kelly, F.: Charging and Rate Control for Elastic Traffic. European Transactions on Communications 8, 33–37 (1997)
9. Andrews, M., Kumaran, K., Ramanan, K., Stolyar, A., Whiting, P., Vijayakumar, R.: Providing Quality of Service over a Shared Wireless Link. IEEE Communications Magazine 39(2), 150–154 (2001)
10. 3GPP.: Deployment Aspects, TR. 25.943 v8.0.0 (2008)