Long-term Fairness in Multi-cell Networks Using Rate Predictions

Hatem Abou-zeid*, Hossam S. Hassanein*, and Nizar Zorba†
*Queen's University, Canada, h.abouzeid@queensu.ca, hossam@cs.queensu.ca
†Qatar Mobility Innovations Center, P.O.Box 210531, Doha, Qatar, nizarz@qmic.com

Abstract—**Providing fair service to users in cellular networks is of significant importance to mobile operators. However, the geographic variability of network coverage and irregularity of user demand in space and time has made consistent fair service provisioning a challenging problem. Therefore, vehicular users traversing small cells experience significant service fluctuations. In this paper we investigate how rate predictions of mobile users can be utilized to improve long-term fairness. We use the α-fair utility function to formulate a predictive long-term resource allocator that improves fair user service over multiple cells. The α-fair utility provides the flexibility to obtain the desired long-term tradeoff between fairness and throughput. Numerical results indicate promising fairness-throughput gains over schemes that do not incorporate user rate predictions.**

I. Introduction

The continuous increase of global mobile traffic is a fundamental challenge to mobile network operators. Furthermore, studies have shown that traffic is highly uneven across geographical locations and changes significantly during the day [1]. Cellular network coverage and signal strength are also highly dependent on the geographical location as reported in radio map databases [2]. These two factors of variable network availability, and variable spatio-temporal user demand lead to challenges in providing long-term service fairness to mobile users [3]. This is because while some users may head to favourable network locations, others may move towards congestion zones or locations with poor signal quality. This results in significant variability in long-term service fairness.

Predicting user mobility can however be of significant use by enabling the estimation of future user channel gains due to the correlation between location and signal strength [2]. With such information, long-term planning of resource allocation both within a single cell, and across multiple cells can be possible. Mobility predictions are particularly plausible for users in public transportation or vehicles on highways. Studies have also shown that peoples' daily routes exhibit a high degree of temporal and spatial regularity, with people following particular routes to and from frequently visited places [4]. This faciliates a database of user profiles that can further improve prediction accuracy.

The notion of long-term fairness over multiple cells was discussed in [3], where user Quality of Service (QoS) history in previous cells is exploited to make allocations in the current cell. Such a scheme improves long-term fairness by prioritizing users that had poor service in previously traversed cells. However, it does not exploit rate predictions to make long-term plans. The use of mobility and rate *predictions* to improve network functions and user QoS have been investigated in a limited number of recent works. In [5], Ali et al. employ rate predictions to increase system data rate of a single-cell but do not consider cooperation between multiple cells, or formulate different network objectives, and assume an infinite backlog traffic model only. Yao et al. demonstrate the use of radio maps in video streaming in [6], where video servers proactively switch to the transmission rates predicted to be supportable at the user, and therefore improve TCP rate control and throughput. In a recent related work [7], we illustrated how user mobility predictions can be exploited to increase network throughput and to provide max-min fairness. As max-min fairness provides a fairness performance limit but is generally not a practical fairness objective, a more general predictive fair allocator is needed. Towards this end, in this paper we present a generic optimal long-term fairness Resource Allocation (RA) formuation based on the α-fair utility function [8]. This allows variable degrees of fairness and enables the formulation of a predictive proportation fair allocator as a special case. Through extensive simulations with practical vehichle mobility, we demonstrate the gains of the predictive fair allocators compared to schemes that do not exploit user rate predictions.

This paper is organized as follows. In the next section, we present our system models, followed by the formulation of the optimal long-term fair allocation problem in Sec. III. Numerical results are presented and discussed in Sec. IV. Finally, we conclude our work in Sec. V.

II. System Description

A. System Model

We consider a network with a Base Station (BS) set $\mathcal{M}$ of M BSs and a user set $\mathcal{N}$ with N users. An arbitrary BS is denoted by $j \in \mathcal{M}$ and a user by $i \in \mathcal{N}$. Users are associated to BSs based on the strongest received signal. The set $\mathcal{U}_{j,t}$ contains the indices of all the users associated with BS j at time t.

A highway scenario covered with three BSs is considered as shown in Fig. 1. To provide realistic vehicular mobility we use the SUMO traffic simulator [9] to generate vehicle movement traces. The achieved date rate by each user is dependant on the

This work was made possible by a *National Priorities Research Program* (NPRP) grant from the Qatar National Research Fund (Member of Qatar Foundation).

978-1-4799-0724-3/13/$31.00 ©2013 IEEE

path loss model $\mathrm{PL}(d) = 128.1 + 37.6 \log_{10} d$ where the BS-user distance d is in km [10]. The data rate is computed using Shannon's equation with SNR clipping at 20 dB to account for practical modulation orders. Therefore, a user i at time t, will have a feasible data rate

$$\hat{r}_{i,t} = B \log_2(1 + P_{\text{rx}}/N_o B) \quad (1)$$

where P_{rx} is the predicted received power based on the predicted BS-user distance d, while N_o and B are the noise power spectral density and the transmission bandwidth respectively. Future user link rates are assumed to accurately known for a duration of T seconds, which we call the *prediction window*. This is illustrated in Fig. 2 for a user travering the indicated highway of Fig. 1.

B. Resource Sharing Model

Time t is divided in equal slots of duration Δt, during which $\hat{r}_{i,t}$ is assumed to be constant. A typical value for Δt is 1 s for vehicle speeds up to 20 m/s, during which average channel gain is not significantly changed. During the allocation interval Δt, the BS resources can be shared in arbitrary ratios among the users. We define the resource sharing factor $x_{i,t} \in [0, 1] : i \in \mathcal{N}, t = \{1, 2, \ldots, T\}$ as the fraction of time during each slot t that the BS bandwidth is assigned to user i. The user received data is therefore $\hat{r}_{i,t} x_{i,t}$. Fig. 2 illustrates an example of $\hat{r}_{i,t}$ and $\hat{r}_{i,t} x_{i,t}$ for a user traversing the indicated highway in Fig. 1. We can see that the user predicted rate is expected to increase followed by a decrease, and then another increase. This is based on the prediction of the user motion across the multiple BSs. A sample allocation, that indicates the proportion of the predicted channel rate planned to be allocated to that user is represented by the bars in the figure. Therefore, $x_{i,t}$ is the optimization variable used to define long-term user allocations over multiple BSs.

C. User Traffic

The data traffic requested by user i at time t is denoted by $D_{i,t}$, and the total data requested during T is denoted by D_i. We consider two cases for user traffic: (i) full buffer traffic, where $D_{i,t} \to \infty \; \forall\, i, t$, and (ii) file download traffic, where D_i is finite $\forall\, i$. Full buffer traffic is used to illustrate the limits of the fairness performance. A summary of the commonly used notation in this paper is provided in Tab. I.

D. Reference Resource Allocators

To provide a performance reference we consider the following resource allocators that do not incorporate rate predictions.

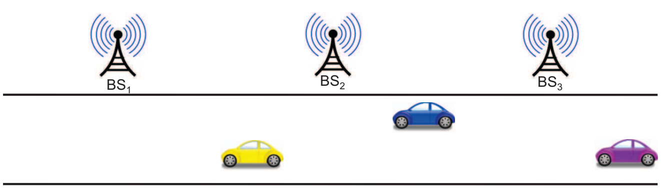


Fig. 1. Considered highway scenario with vehicular mobility.

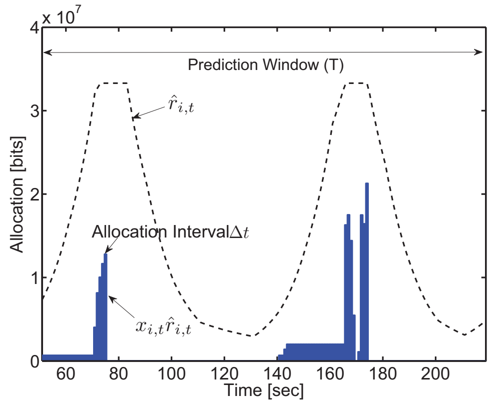


Fig. 2. Example of the predictive RA made to a user.

1) Max-rate Allocation: In Maximum Rate (MR) allocation, the user with the highest data rate $\hat{r}_{i,t}$ is granted full channel access, i.e. $x_{i,t} = 1$. This maximizes the network throughput but makes no effort to serve users fairly.

2) Equal Share Allocation: In Equal Share (ES) allocation, the BS air time (resource) is shared equally among the users at each time slot t. If there are $N_{j,t}$ users associated with BS j at time t (i.e. $N_{j,t} = |\mathcal{U}_{j,t}|$), then $x_{i,t} = 1/N_{j,t}$ for each user $i \in \mathcal{U}_{j,t}$, and the rate received by each is $\hat{r}_{i,t}/N_{j,t}$.

III. Predictive α-Proportional Fair RA

In this section we formulate predictive resource allocation that achieves long-term α-fairness over multiple cells. Predictions of user rates are assumed to be available at a central coordinating BS which plans user allocations for all the cooperating BSs.

A. α-Proportional Fairness

In resource allocation, achieving fairness among users generally comes at the cost of reducing system throughput. There are several notions of what constitutes a fair allocation [11]. For example, max-min fairness ensures that the minimum data rate that any user receives is maximized. In other words, an allocation is said to be max-min fair if any rate r_i cannot be increased without decreasing some r_j which is smaller than or equal to r_i. This measure gives absoulte priority to fairness over system throughput. Proportional fair RA [12] on the other hand provides a trade-off between fairness and throughput and its usefulness in scheduling has been of significant interest in literature and industry. To achieve an arbitrary degree of the fairness-throughput tradeoff, a generalization of proportional fair and max-min fair allocation was presented in [8]. This is known as the α-proportional fair allocation where the parameter α controls the degree of fairness in the allocation. α-proportional fair allocation is based on defining the following utility function:

$$\phi_\alpha = \begin{cases} \frac{x^{(1-\alpha)}}{(1-\alpha)}, & \text{if } \alpha \geq 0, \alpha \neq 1, \\ \log x, & \text{if } \alpha = 1, \end{cases} \quad (2)$$

and thereafter solving the optimization problem:

$$\underset{\mathbf{r}}{\text{maximize}} \quad \sum_{i=1}^{N} \phi_\alpha(r_i) \qquad (3)$$
$$\text{subject to:} \quad \mathbf{r} \in \mathcal{S},$$

where $\mathbf{r}$ is the user allocation vector and $\mathcal{S}$ is the feasible region of $\mathbf{r}$. Note that when $\alpha \to 0$, α-proportional fairness is reduced to maximum throughput allocation, and when $\alpha \to 1$, proportional fairness is achieved from the logarithmic utility. It has also been shown that when $\alpha \to 2$, potential delay minimization is obtained and when $\alpha \to \infty$, max-min fairness is achieved [8]. More details on the characterization of the generalized weighted α-proportional fairness allocation have been studied in [13].

B. Predictive α-Proportional Fairness

In order to plan an allocation that achieves long-term α-proportional fairness over multiple cells, the optimization problem in (3) can be formulated as follows (for $\alpha \neq 1$):

$$\underset{\mathbf{x}}{\text{maximize}} \quad \sum_{i=1}^{N} \frac{\left(\sum_{t=1}^{T} \hat{r}_{i,t}\, x_{i,t}\right)^{(1-\alpha)}}{1-\alpha} \qquad (4)$$
$$\begin{aligned}
\text{subject to:}\quad & \text{C1:} \quad \sum_{i \in \mathcal{U}_{j,t}} x_{i,t} \leq 1, && \forall\, j, \Delta t \\
& \text{C2:} \quad \sum_{t=1}^{T} \hat{r}_{i,t}\, x_{i,t} \leq D_i, && \forall\, i \\
& \text{C3:} \quad x_{i,t}\, \hat{r}_{i,t} \geq \text{MBR}, && \forall\, i, \Delta t \\
& \text{C4:} \quad 0 \leq x_{i,t} \leq 1 && \forall\, i, \Delta t.
\end{aligned}$$

Note that the inner summation over time in the objective is introduced as we are interested in maximizing the utility achieved from the total data a user receives during the prediction window T. The predicted user rate at every time slot is $\hat{r}_{i,t}$, and the optimization variable $x_{i,t}$ is determined such that the total utility over all users is maximized. The optimization problem is coupled over multiple BSs as the total user data allocation is computed over all the BSs the user traverses during T. It is therefore solved centrally at the coordinating BS.

The constraint C1 expresses the resource sharing limitation at each BS. It ensures that the sum of the resource sharing factors of all users associated with a BS j is equal to 1 at every time slot. C1 is applied at each BS, and therefore the subset of users $i \in \mathcal{U}_{j,t}$ is needed for every time slot. This is predicted based on the predictions of user-BS distance with time. C2 limits the amount of data assigned to each user during T to the total amount request. Note that this constraint also couples user data allocations made over multiple BSs. In C3, an optional Minimum Bit Rate (MBR) is defined to provide a lower limit of user allocation for each time slot. This is to ensure that while considering the long-term utility, the per-slot user application needs can also be met. Finally, C4 provides the bounds for the resource sharing factor.

TABLE I
FREQUENTLY USED NOTATIONS

Symbol	Description
N	Number of users in the network
M	Number of BSs in the network
T	Prediction window: time duration over which user routes are predictable [s]
t	Current time slot: each slot has a duration of 1 s
α	Fairness control parameter
$\mathcal{U}_{j,t}$	Set of the indices of users associated with BS$_j$ at time t
$\hat{r}_{i,t}$	Predicted rate that user i can receive at time t [bits/s]
$x_{i,t}$	Fraction of BS resource assigned to user i at time t
$D_{i,t}$	Traffic requested by user i at time t [bits]
D_i	Total traffic requested by user i during $T\,s$ [bits]
R_i	Total traffic received by user i during $T\,s$ [bits]

We refer to this formulation as the Predictive α-Proportional Fairness (PPF) which is a convex optimization problem as the constaints are linear, and the objective function is an increasing, strictly concave, and continuously differention function on the open interval $(0, \infty)$ for $\alpha > 0, \alpha \neq 1$. The problem can also be equivalently represented as follows by introducing additional optimization variables R_i that denote the total data assigned to each user:

$$\underset{\mathbf{x},\mathbf{R}}{\text{maximize}} \quad \sum_{i=1}^{N} \frac{R_i^{(1-\alpha)}}{1-\alpha} \qquad (5)$$
$$\begin{aligned}
\text{subject to:}\quad & \text{C1, C2, C3, C4} \\
& \text{C5:} \quad \sum_{t=1}^{T} \hat{r}_{i,t}\, x_{i,t} = R_i \qquad \forall\, i.
\end{aligned}$$

For the case when $\alpha = 1$, the resulting PPF problem becomes:

$$\underset{\mathbf{x},\mathbf{R}}{\text{maximize}} \quad \sum_{i=1}^{N} \log R_i \qquad (6)$$
$$\text{subject to:} \quad \text{C1, C2, C3, C4, C5},$$

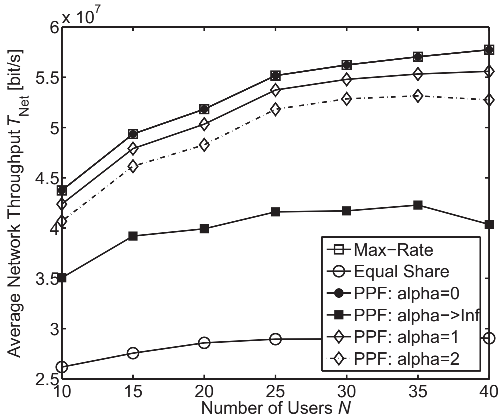
which we refer to as predictive proportional fairness. This is also a convex problem due to the concave log utility function.

IV. Simulations and Numerical Results

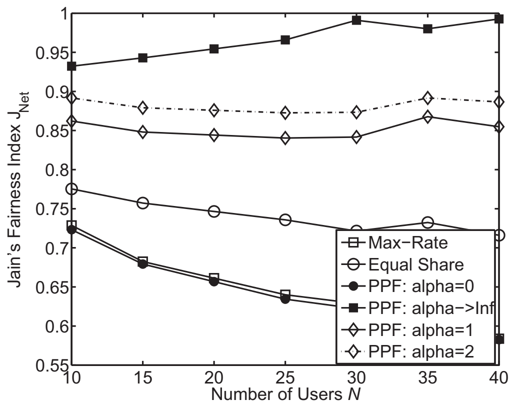
In this section, we present the numerical results of the predictive α-proportional fairness and illustrate the potential gains of incorprating rate predictions in resource allocation planning over multiple cells.

A. Simulation Set-up

We evaluate PPF for the highway illustrated in Fig. 1 where the inter-BS distance is set to 1 km. BS transmit power is 40 W and the bandwidth is 5 MHz. Vehicle mobility traces are generated using SUMO for vehicles entering the highway at a rate of 1 vehicle per second. For file download traffic, the



(a) Network throughput



(b) Jain's fairness index

Fig. 3. Network throughput T_{Net} and Jain's Fairness Index J_{Net} vs. number of users N for full buffer traffic.

required files are 1 Gbits. The prediction window T is 200 s and MBR is set to zero as no minimum rate is required for a file download. Simulations are repeated 50 times, and (5) and (6) are solved to obtain average values of the metrics defined below.

B. Performance Metrics

- T_{Net}: the average downlink network throughput calculated as the sum of the average data rate of all the users.
- J_{Net}: Jain's fairness index [14] for user throughput and is computed as $(\sum_{i=1}^{N} \mathrm{Th}_i^2/N\sum_{i=1}^{N} \mathrm{Th}_i^2$, where Th_i is the average user throughput during T. Fairness is computed based on the total data a user receives during T since we are interested in long-term fairness.

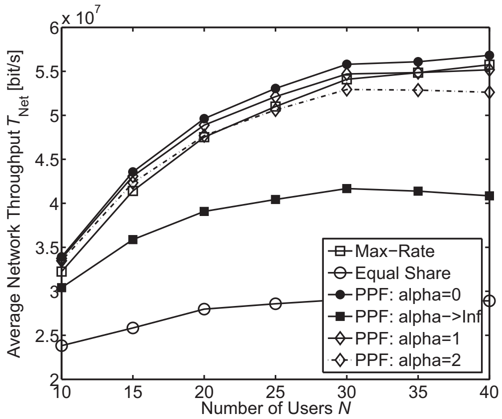
C. Full Buffer Traffic

Fig. 3 shows the results of PPF for full buffer user traffic. When $\alpha = 0$ throughput is maximum but fairness is severely affected as illustrated in Fig. 3(a) and Fig. 3(b) respectively. On the other hand when $\alpha \to \infty$, the α-proportional fairness performs as a max-min allocator resulting in the converse behavior of a very high fairness and a low throughput. The results for $\alpha = 1$ (predictive proportional fairness) and $\alpha = 2$ demonstrate the usefulness of the α-proportional fair utility and provide a good trade-off between fairness and throughput.

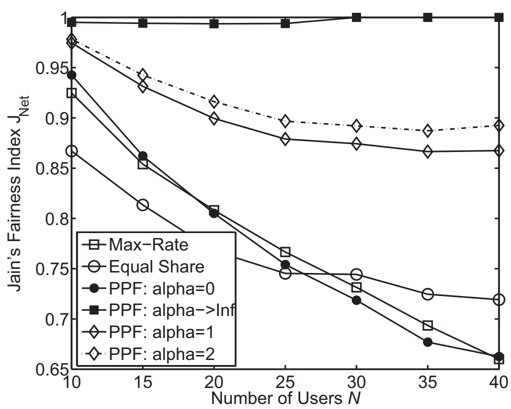
It is worth pointing that the base-line equal rate allocator performs poorly in both throughput and fairness. This is due to its lack of knowledge of the rates users will experience, and it therefore cannot make opportunitic allocations or planned transmission delays to improve network throughput and user service.

D. File Download Traffic

In Fig. 4 we present the results when users have finite traffic requests, which is the more practical case. Here we can see that at low load network performance is not very different for the α-proportional fairness variants. However, there still is a considerable performance gains over the non-

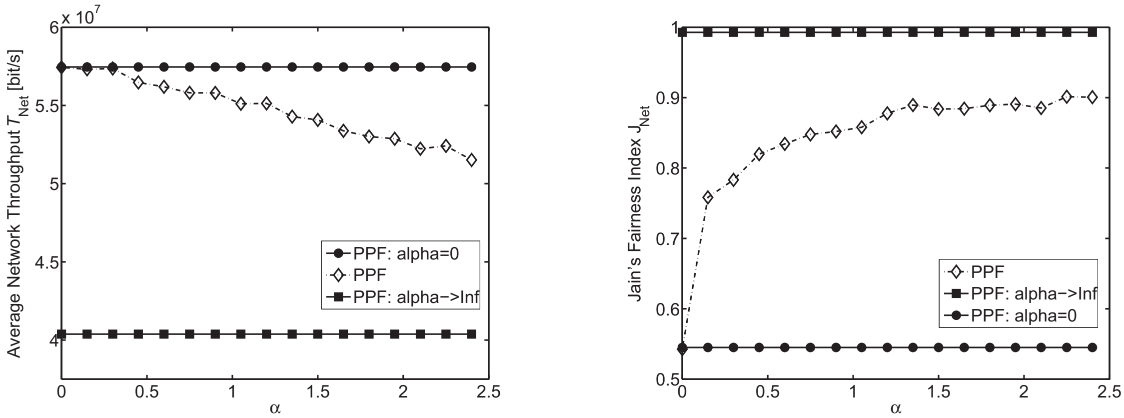


(a) Network throughput



(b) Jain's fairness index

Fig. 4. Network throughput T_{Net} and Jain's Fairness Index J_{Net} vs. number of users N for file download traffic.



(a) Network throughput

(b) Jain's fairness index

Fig. 5. Network throughput T_{Net} and Jain's Fairness Index J_{Net} vs. α for 40 users and file download traffic.

predictive equal rate allocation scheme. As the load increases, the behavior tends to follow the full buffer traffic scenario (which is expected). The benefits derived from setting α to 1 or 2 are also apparent from the satisfactory throughput-fairness trade-off.

E. Effect of α

Fig. 5 provides the throughput and fairness results for varying values of α. We can see that as α increases from 0 to 1, a significant fairness improvement is achieved with a relatively small throughput loss. This indicates that predictive proportional fairness ($\alpha = 1$) can provide a good operating point. As α increases further, the rate of increase in fairness decreases while throughput begins to decrease considerably, thereby indicating that this is not a desirable operating direction. Fig. 5 also illustrates the extent by which predictive max-min fairness sacrifices throughput to achieve fairness.

V. Conclusion

In this paper, we presented predictive α-proportional fairness resource allocation that uses rate predictions to achieve long-term fairness. The problem is formulated using the α-fair utility function for a network of multiple cells, and is solved centrally for the cooperating BSs. Numerical results of the formulated optimization problems demonstrate the usefulness of the α utility in achieving a desirable throughput-fairness trade-off. Future work includes investigating performance in the presence of prediction errors, as well as decentralized predictive algorithms that can be implemented in real-time at the base stations.

References

[1] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, pp. 882–890, May 2011.

[2] OpenSignal, "The opensignal, project homepage." http://opensignal.com/, 2013. Link verified on Feb. 15th, 2013.

[3] H. Abou-zeid, H. Hassanein, S. Valentin, and M. Feteiha, "Lookback scheduling for long-term quality-of-service over multiple cells," in *Proc. IEEE IWCMC 2013, QoS and QoE in Wireless Commun./Networks Wksp*, July 2013.

[4] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.

[5] S. H. Ali, V. Krishnamurthy, and V. C. M. Leung, "Optimal and approximate mobility-assisted opportunistic scheduling in cellular networks," *IEEE Trans. Mobile Comput.*, vol. 6, pp. 633–648, June 2007.

[6] J. Yao, S. Kanhere, and M. Hassan, "Improving QoS in high-speed mobility using bandwidth maps," *IEEE Trans. Mobile Comput.*, vol. 11, pp. 603–617, Apr. 2012.

[7] H. Abou-zeid, H. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *submitted to IEEE GLOBECOM 2013*, Dec. 2013.

[8] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, pp. 556–567, Oct. 2000.

[9] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo - simulation of urban mobility: An overview," in *Proc. Third Int. Conf. on Advances in System Simulation (SIMUL 2011)*, pp. 63–68, Oct. 2011.

[10] 3GPP, "LTE;E-UTRA; radio frequency system scenarios," Technical Report TR 36.942 V11.0.0, 3GPP, 2012.

[11] S. Huaizhou, R. Venkatesha, E. Onur, and I. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys and Tutorials*, vol. 15, pp. 1–20, May 2013.

[12] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, 1997.

[13] M. Uchida and J. Kurose, "An information-theoretic characterization of weighted α-proportional fairness," in *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, pp. 1053–1061, Apr. 2009.

[14] 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." Tech. Rep. TR-301, DEC, sep 1984.