TCP-aware Resource Allocation in OFDMA Based Wireless Networks

Toktam Mahmoodi, Vasilis Friderikos, Oliver Holland, Hamid Aghvami
Centre for Telecommunications Research, King's College London,
Strand, London WC2R 2LS, UK
{toktam.mahmoodi, vasilis.friderikos, oliver.holland, hamid.aghvami}@kcl.ac.uk

Abstract—**Various cross-layer optimizations have been recently proposed to improve the performance of TCP in wireless networks. In this paper, we propose a dynamic OFDMA based subcarrier/power allocation which aims to constrain the achievable rate on each flow proportionally to the TCP throughput of the corresponding flow. Numerical investigations reveal that the proposed approach significantly increases fairness among TCP flows operating over end-to-end paths of different characteristics.**

I. INTRODUCTION

Next generation wireless technologies, such as Long Term Evolution (LTE) [1] and IEEE 802.16e [2], specify Orthogonal Frequency Division Multiple Access (OFDMA) as their access method. In OFDMA based wireless networks the available bandwidth is divided into multiple orthogonal subcarriers, allowing users to transmit simultaneously through allocating different subsets of the available subcarriers to different users. The implied problem of subcarrier allocation, and transmission power and rate distribution among users has been a prominent area of research over the past few years.

Much of the previous research work has concentrated on maximizing the aggregate data rate subject to the available transmitted power or achievable Bit Error Rate (BER) constraints [3]. Clearly such a rate maximization problem is more relevant for data centric transmissions (elastic applications). On the other hand, a resource allocation problem which addresses the minimization of overall power consumption in the interests of energy efficiency, with a minimum rate constraint has been considered in [4]. This concept is of more interest for applications requiring a fixed data rate. None of the above approaches, however, explicitly satisfy fairness among users.

Alternative formulations do, however, consider fairness, either by prioritization using the weighted sum rate method [5], or by introducing proportional rate constraints [6]. A Max-Min approach is presented in [7], in which fairness is considered by maximizing the lowest achieved data rate among the user set. The research presented in [8] addresses proportional fairness in OFDMA resource allocations, based on the Nash bargaining solution. Although the issue of fairness with respect to the allocated data rate is investigated over the local link in some of the literature, aspects of fairness in the actual end-to-end data transmission are not sufficiently addressed.

In this paper, we specifically address the above question on end-to-end achievable rate for different users by including the theoretical upper bound of the Transmission Control Protocol (TCP) throughput in resource allocation problem in OFDMA. To this end, we formulate an optimization problem for resource allocation in OFDMA networks which attempts to achieve the proportional data rate among flows with respect to the corresponding flows' TCP theoretical throughput. Considering that the theoretical TCP throughput is the optimal throughput that each end-to-end path can achieve, the rationale of the proposed approach is to distribute the resources more optimally, and fairer among the competing TCP flows.

The remainder of this paper is as follows. In Section II, we review the related literature on TCP-aware resource allocation algorithms, and highlight the novelty of the approach presented in this paper. Section III includes the description of our models and expressions. In section IV, our proposed resource allocation problem is formulated, followed by its solutions. Section V presents simulation scenarios to evaluate the performance of the proposed algorithm, and proceeds to study the effect of our scheme on fairness among TCP flows and discuss the implications of important results. This paper concludes in Section VI.

II. TCP-AWARE RESOURCE ALLOCATION ALGORITHMS

TCP [9], is the prominently used transport layer protocol to achieve reliable end-to-end data transfer in the Internet. TCP, however, exhibits a number of shortages when the underlying medium detracts from the reliable, wired media which it was originally designed to serve [10]. In particular, in the case of TCP over wireless links, the random losses of such links cause it to erroneously assume that these losses are due to congestion, and therefore reduce its transmission rate unnecessarily. Such factors, in addition to differences in the way different flavors of TCP react to random wireless losses, mean that-particularly over wireless links-there can be considerable unfairness among TCP flows [11].

A thorough overview of cross-layer design for resource allocation algorithms in third generation wireless networks is given in [12], where TCP over CDMA is also addressed. TCP-aware resource allocation algorithms over a CDMA network are studied in [13], the objective being to optimize throughput. The novel proposed algorithm in that paper uses information from the TCP state machine (slow start or congestion avoidance) to allocate the data rate more appropriately at the

wireless link. In the context of IEEE 802.16, reference [14] proposes a TCP-aware allocation algorithm which estimates the bandwidth demand based on the long-term data rate, and allocates resources accordingly.

Closely related to the above works, in this paper we propose a TCP-aware resource allocation algorithm for OFDMA based access networks. We define a resource allocation problem based on channel conditions, which also provides fair allocation among TCP flows. To do this, we define an optimization problem with the objective of maximizing the sum rate, constrained to the weighted proportional rate, with respect to the optimal TCP throughput.

Our major contributions in this paper are as follows. We present a resource allocation problem which provides a proportional data rate among users, based on the characteristics of end-to-end TCP flows. Unlike available solutions in the literature, we use the closed form expression of TCP throughput [15] as a means of TCP-awareness in allocations, and in contrast with existing TCP-aware resource allocation techniques, we focus on OFDMA-based systems. To the best of our knowledge, this is the first research work which addresses this problem. We then study the performance of the proposed method, not only from the TCP point of view, but also from the wireless link side. We also show that our approach can provide significantly fairer allocations among end-to-end flows compared with the existing methods which attempt to allocate OFDMA subcarriers fairly [5][6][7][8]. The derived solution for our proposed resource allocation problem is mainly based on the approach used in [6].

III. SYSTEM MODEL

We assume there are n active TCP flows, all in congestion avoidance phase, and there are m subcarriers to the base station. Let, for flow i, the rate on subcarrier j be r_{ij}. We assume each user has a single TCP flow, therefor the achievable rate by user i is

$$R_i = \sum_{j=1}^{m} a_{ij} r_{ij}, \qquad (1)$$

$$a_{ij} = \begin{cases} 1 & \text{if subcarrier } j \text{ is assigned to user } i, \\ 0 & \text{if subcarrier } j \text{ is not assigned to user } i. \end{cases} \qquad (2)$$

The channel gain of user i in the subcarrier j is denoted by G_{ij}. With the thermal noise power, σ^2, the ith user's received signal to noise ratio (SNR) on subcarrier j is thus denoted as,

$$\gamma_{ij} = \frac{p_{ij} G_{ij}}{\sigma^2}. \qquad (3)$$

where p_{ij} is the allocated power to flow i on subcarrier j.

For each user, adaptive modulation provides the desired rate in the allocated subcarrier. Moreover, BER is expressed based on the adaptive M-array quadratic modulation (M-QAM) [16].

$$BER_{ij} \approx c_1 \, e^{-c_2 \frac{\gamma_{ij}}{2^{r_{ij}-1}}}. \qquad (4)$$

where $c_1 \approx 0.2$, $c_2 \approx 1.5$. Similar to [8] we assume a fixed and the same BER for all users in all subcarriers. Solving for r_{ij}, the achievable rate for user i on the jth subcarrier can be described as follows,

$$r_{ij} = w_j \log_2 \left(1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right) \quad bits/s. \qquad (5)$$

where $c_3 = -ln(BER/c_1)/c_2$ with $BER = BER_{ij} \;\; \forall i,j$, and w_j is the bandwidth of subcarrier j which is assumed to be equal for all subcarriers given by w.

We assume the slow-fading channel such that the channel is constant within each OFDM frame. The slowly time varying assumption is crucial since it is also assumed that perfect estimation of the subchannels is available for each user. Moreover, mobile users and the base station are synchronized, thus there is no inter-carrier interference.

As mentioned above, allocated subcarriers and power to user i defines the achievable data rate by that user. On the other hand, the theoretical end-to-end throughput, bounds the achievable data rate for each specific user. Moreover, the end-to-end throughput is not only affected by the BER of the channel, but also by the end-to-end Round Trip Time (RTT). In this paper, we assume TCP as our end-host transport layer, thus TCP throughput is the theoretical data rate which can be achieved by each flow.

As BER is a fixed value, TCP throughput is mainly affected by the diversity of RTT. TCP throughput of flow i is expressed as a closed-form function proportional to the probability of a packet in error, which can be driven directly from BER, and the end-to-end RTT [15].

$$B_i = MSS \cdot \frac{\frac{1-e_i}{e_i} + E[W_i]}{\overline{RTT}_i \left(\frac{1}{2} \cdot E[W_i] + 2 \right)}, \qquad (6)$$

where

$$E[W_i] = -\frac{1}{3} + \sqrt{\frac{8(1-e_i)}{3e_i} + \left(\frac{1}{3}\right)^2}. \qquad (7)$$

where it is assumed that every single transmitted packet is acknowledged, MSS is the TCP Maximum Segment Size, and e is the probability of a TCP packet in error. We assume that the bottleneck link is at the wireless link, thus e can be driven from the BER of wireless link. $\overline{RTT}$ is an average value of RTT; thus the instantaneous variation in the RTT causes by each single loss of acknowledgements does not affect the throughput. The throughput expression here is based on the well-used Reno TCP, although the proposed method in this paper is independent of the flavor of TCP, thus the throughput expression of any other TCP flavor such as NewReno can be replaced.

To calculate the TCP throughput in the wireless base station, where the resource allocation algorithm takes place, the value of RTT is required. Various methods are presented in the literature to estimate RTT either actively or passively at an interior point in the network. The passive measurement can be done based on the three-way handshake message [17], or by associating the data segment with the acknowledgement

that triggers the packet. TCP timing information can also be included in the Timestamp option of the TCP segment. Experiments show that 90% of the passive measurements are within 10% of the precise RTT value [17]. These methods are not computationally complex and can be easily implicit in the link-layer of the base station.

IV. RESOURCE ALLOCATION PROBLEM WITH PROPORTIONAL TCP THROUGHPUT CONSTRAINTS

We formulate the optimization problem to maximize downlink sum rate, but TCP fairness is assured by imposing a set of nonlinear constraints into the optimization problem. We propose to constrain the proportional downlink rate among users with respect to the TCP theoretical throughput. Therefore, our resource allocation function would be summarized in the following problem,

$$(\text{P1}) : \text{Maximize} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} w \log_2 \left(1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right),$$

$$\text{subject to:} \sum_{i=1}^{n} a_{ij} \leq 1, \qquad \forall j \tag{8}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} p_{ij} \leq P_T, \tag{9}$$

$$\frac{R_i}{B_i} = \frac{R_1}{B_1}, \qquad \forall i \in \{2, ..., n\} \tag{10}$$

$$p_{ij} \geq 0, \qquad \forall i, j \tag{11}$$

$$a_{ij} \in \{0, 1\}, \qquad \forall i, j \tag{12}$$

where P_T is the total available power at the base station. Constraints (12) and (11) ensure the correct assigned values for the subcarrier and power allocation matrices. Constraint (9) restrict the total available power at the base station, and constraint (8) imposes that every subcarrier is assigned to one and only one user. Moreover, constraint (10) provide fairness among TCP flows with maintaining proportional rate with respect to the TCP theoretical throughput for each user.

A. Optimal Subcarrier Allocation and Power Distribution

Ideally subcarrier and power should be assigned jointly to achieve the optimal solution. This problem address a mixed integer programming problem which pose a high computational complexity. To solve the problem optimally, discrete problem (P1) can be reformulated to a continuous optimization problem. Thus, constraint (12) can be rewritten as aij bounded in $[0, 1]$. Moreover, the objective function also needs to be rewritten as,

$$\sum_{i=1}^{n} R_i - M \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} (1 - a_{ij}). \tag{13}$$

where M is a relatively large value to ensure the integer assignment for a_{ij}.

The continuous reformulation of problem (P1) can be solved using well-known methods and also MATLAB optimization toolbox. In the other hand, by increasing the number of mobile users and the number of subcarriers, number of constraints of the optimization problem (P1) are also increased. Thus, solving the problem is computationally complex, and it can be prohibitive for the base station to solve this problem in real time. Therefore, we use the method presented in the literature to decouple the problem to the two separate problems. Firstly to allocate the subcarriers, and secondly distribute power into the allocated subcarriers. The first problem itself is an NP-hard combinatorial optimization problem with finite number of combinations.

B. Suboptimal Subcarrier Allocation and Power Distribution

We use the approach similar to [6] to decouple the optimization problem. In the subcarrier allocation it is assumed that power is equally distributed in all the subcarriers, therefore solution is suboptimal. The principle of the algorithm is for each user to allocate the subcarrier with the highest channel gain available. We detail a suboptimal algorithm, whose at the first iteration each user select the best available subcarrier, starting from the user with the highest value of B_i and continue in the order of their B_i values. Afterwards, at each iteration, the user with the lowest proportion of $\frac{R_i}{B_i}$ has the option to choose the subcarrier. Finally, Ω_i is the set of assigned subcarriers to user i.

Algorithm 1 Subcarrier Allocation Algorithm

---

a) Initialization
  1) Set R_i=0 and $\Omega_i = \phi$ for i=1 to n and C={1,2,...m}.
  2) Sort the users' index in the descending order of B_i.
b) for i=1 to n
  1) Find the subcarrier k satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C$.
  2) Let $\Omega_i = \Omega_i \cup \{k\}$ and $C = C - \{k\}$.
  3) Update R_i
c) while $C \neq \phi$
  1) Find user l satisfying $\frac{R_l}{B_l} < \frac{R_i}{B_i}$ for all $i \in \{1, ..., n\}$.
  2) For user l, find the subcarrier k satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C$.
  3) Let $\Omega_l = \Omega_l \cup \{k\}$ and $C = C - \{k\}$.

In the next step, to a certain subcarrier allocation, problem (P1) will be simplified into a maximization problem over continues variable p_{ij}.

$$(\text{P1}') : \text{Maximize} \sum_{i=1}^{n} \sum_{j \in \Omega_i} w_j \log_2 \left(1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right),$$

$$\text{subject to:} \sum_{i=1}^{n} \sum_{j \in \Omega_i} p_{ij} \leq P_T, \tag{14}$$

$$\frac{R_i}{B_i} = \frac{R_1}{B_1}, \qquad \forall i \in \{2, ..., n\} \tag{15}$$

$$p_{ij} \geq 0, \qquad \forall i, j \tag{16}$$

Ω_i is the set of assigned subcarrier to user i; Ω_{i_1} and Ω_{i_2} are mutually exclusive if $i_1 \neq i_2$.

The problem of power distribution among subcarriers, and the performance comparison between equal and optimal power distribution is well-studied in [18], in which it has been shown that equal power distribution can not generally result in near optimal solution. Therefore, to find the optimal solution for power distribution, we use similar approach to [8] and [6]. Thus, problem (P1′) can be solved writing the lagrangian dual function. Differentiating the lagrangian dual function with respect to p_{ij} and set the derivative to 0, power can be distributed with the same method presented in [6].

V. PERFORMANCE INVESTIGATION

In order to investigate the performance of the proposed scheme, we simulate an OFDMA system with 52 subcarriers (equal to the number of OFDM subcarriers in IEEE 802.11a). Hence, this scenario can be easily extended to the simulation scenarios with larger number of subcarriers. The rest of simulation parameters are similar to the ones used in [1]. The available bandwidth is 5MHz, maximum power at the base station is 43dBm, the target BER is 10^{-4}, and the average value of SNR is 15dB. Moreover, the thermal noise power, σ^2, is -107dBm (Johnson-Nyquist noise over 5MHz bandwidth).

The maximum segment size (MSS) of the TCP flows is set to the standard maximum transfer unit of an Ethernet which is 1460 bytes. We further assume that the end-to-end RTT for any of the TCP flows is a uniformly distributed random value in the range (10ms, 200ms). We perform 150 monte carlo simulations; thus each simulation round runs with a new set of random RTT values per TCP flow, and a new set of random SNR values per subcarrier.

After initializing values in step (a), the iterative steps (b) and (c) in Algorithm 1 are performed to allocate subcarriers. In order to optimally distribute power over the allocated set of subcarriers, problem (P1′) is solved using the MATLAB optimization toolbox.

Our benchmarks are the sum rate maximization problem, denoted by (BM1), and also the sum rate maximization with an equal rate constraint, denoted by (BM2). Clearly, power and subcarrier constraints of Equations (12)-(9) also apply to the benchmark problems. These two benchmarks represents the two extremes of the resource allocation schemes, (BM1) does not consider fairness in the allocations and aims only to achieve the maximum capacity on the link. On the other hand, (BM2) blindly provides fairness with equal rate allocation to all users.

A. Fairness Analysis

We use Jain's fairness index [19], denoted by FI, to measure fairness of our allocation scheme among TCP flows. This index is well-used as a quantitative measure of fairness in both wired and wireless networks. The index FI is 1 when
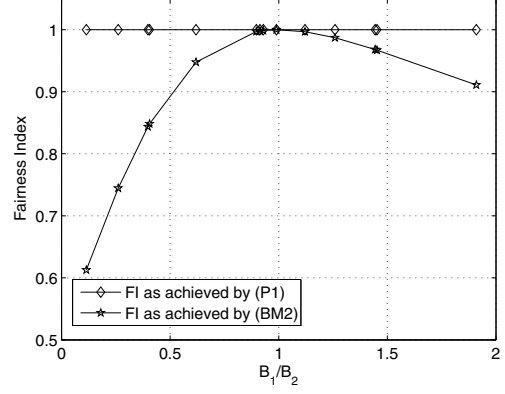

Figure 1. Fairness Index as achieved by solving resource allocation problems (P1), and (BM2) in two-user scenario Vs. the proportion of two TCP flows' optimal throughput.

there is a complete fair allocation.

$$FI = \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n \cdot \sum_{i=1}^{n} x_i^2}. \tag{17}$$

Assuming x_i is the data rate of user i, proportional to the optimal rate that can be achieved on the corresponding end-to-end path, then FI as described in (17) can be the measure of fairness among end-to-end flows. The optimal throughput for each TCP flow, is the theoretical TCP throughput defined by Equation (6) in Section III. Thus, in Equation (17), x_i can be replaced by R_i/B_i.

B. Numerical Results

In the first simulated scenario, two users are assumed, each user having a single TCP flow. Figure 1 presents the results of this scenario. This figure shows the ratio of two flows' TCP steady state throughput (B_1/B_2) on the x-axis, and FI as achieved by solving resource allocation problems (P1), and (BM2) on the y-axis. It can be seen that when the optimal throughput of the two TCP flows are close to each other (in other words when the ratio B_1/B_2 is close to 1), the FI achieved by both problems are almost identical. On the other hand, as B_1/B_2 diverge from one, the fairness index achieved by the proposed allocation method (i.e., problem P1) shows significant improvement.

In the second simulation scenario, the number of mobile users is increased from two to five, ten, fifteen, and twenty. The achieved fairness index by resource allocation problems (P1), (BM1), and (BM2) are plotted in Figure 2. Observed from this figure, the average increase in the fairness index is 40% from (BM1) to (P1), and 20% from (BM2) to (P1). In the scenario with larger number of mobile users competing over wireless channel, providing the fair distribution for the benchmark resource allocation schemes is more challenging, thus our proposed scheme enhance the fairness index more
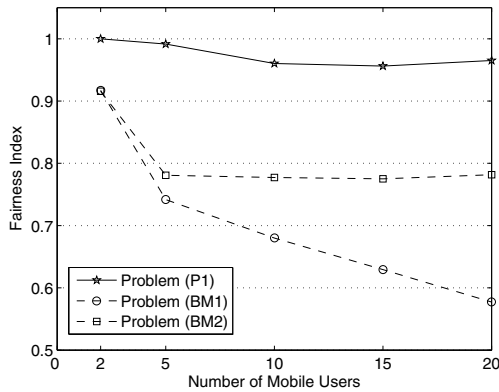
Figure 2. Fairness Index as achieved by solving resource allocation problems (P1), (BM1), and and (BM2) Vs. the number of mobile users.
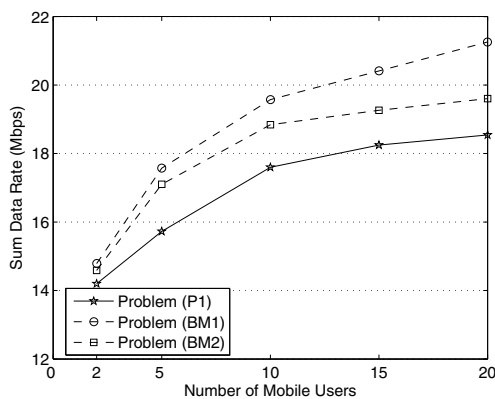

Figure 3. Sum data rate (Mbps) as achieved by solving resource allocation problems (P1), (BM1), and and (BM2) Vs. the number of mobile users.

significantly, e.g. this index is increased up to 65% in twenty-user scenario. Despite the increase in fairness among TCP flows, we expect the aggregated data rate on wireless link to be degraded. This effect can be seen in Figure 3, in which the aggregated throughput shows a decrement of 10% and 5% in average comparing the results of problem (P1) with (BM1) and (BM2) successively. The achieved sum data rate plotted in Figure 3, is based on the allocated data rate in wireless link which can be utilized by the user regarding its end-to-end capacity. Further observation shows that, maximum degradation in the sum rate is also in the twenty-user scenario comparing our proposed TCP-aware scheme with the pure rate maximization (BM1) that is 12.5% decrease in the overall throughput.

VI. CONCLUSIONS

We have presented a TCP-aware resource allocation algorithm, which attempts to maximize sum rate capacity while at the same time provide fairness among TCP flows. Our proposed resource allocation approach, considers the end-to-end performance of TCP flows and distributes data rates among users across the wireless link proportional to the theoretical TCP throughput. Numerical investigations reveal that the proposed resource allocation scheme can increase the fairness among TCP flows by up to 65%. Although in this study we have used the well known analytical model of TCP throughput, an interesting area of future work would be to investigate the performance of the proposed scheme in advanced network-level simulations.

VII. ACKNOWLEDGMENT

REFERENCES

[1] 3GPP Std. TR 25.814, *3rd Generation Partnership Project, Technical Specification Group Radio Access Network: Physical Layer Aspects for Evloved Universal Radio Access (UTRA)(Release 7)*, Sep. 2006.

[2] IEEE Std. 802.16e-2005, *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, Feb. 2006.

[3] I. C. Wong and B. L. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 962–971, Feb. 2008.

[4] Z. Han, Z. Ji, and K. J. R. Liu, "Power Minimization for Multi-Cell OFDM Networks Using Distributed Non-cooperative Game Approach," *Proc. IEEE GLOBECOM'04*, vol. 6, pp. 3742–3747, Nov. 2004.

[5] L. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: Asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, pp. 922–930, June 2004.

[6] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2726–2737, Nov. 2005.

[7] W. Rhee and J. M. L. Cioffi, "Increase in Capacity of Multiuser OFDM System using Dynamic Subchannel Allocation," *Proc. IEEE VTC Spring*, pp. 1085–1089, May 2000.

[8] Z. Han, Z. Ji, and K. J. R. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Trans. Commun.*, vol. 35, pp. 1366–1375, Aug. 2005.

[9] Information Science Institute, University of Southern California, CA, USA, "Transmission Control Protocol," *IETF RFC 793*, Sep. 1981.

[10] A. Chockalingam, E. Altman, J. Murthy, and R. Kumar, "Cross-layer design for optimizing TCP performance," *Proc. IEEE ICC'05*, vol. 5, pp. 3538–3542, May 2005.

[11] T. Mahmoodi, V. Friderikos, O. Holland, and A. H. Aghvami, "Cross-layer Optimization to Maximize Fairness among TCP Flows of different TCP Flavors," *Proc. IEEE GLOBECOM'08*, Nov. 2008.

[12] H. Jiang, W. Zhuang, and X. Shen, "Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond," *IEEE Commun.*, vol. 43, pp. 120–126, Dec. 2005.

[13] M. Ghaderi, A. Sridharan, H. Zang, D. Towsley, and R. Cruz, "TCP-Aware Channel Allocation in CDMA Networks," *IEEE Trans. Mobile Comput.*, vol. 8, pp. 14–28, Jan. 2009.

[14] S. Kim and I. Yeom, "TCP-aware Uplink Scheduling for IEEE 802.16," *IEEE Commun. Letters*, vol. 11, pp. 146–148, Feb. 2007.

[15] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation," *Proc. ACM SIGCOMM '98*, pp. 303–314, 1998.

[16] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, pp. 1561–1571, Sep. 2001.

[17] H. Jiang and C. Dovrolis, "Passive Estimation of TCP Round-Trip Times," *ACM Comp. Commun. Review*, vol. 32, pp. 75–88, July 2002.

[18] H.-W. Lee and S. Chong, "Downlink Resource Allocation in Multi-Carrier Systems: Frequency-Selective vs. Equal Power Allocation," *Proc. IEEE WoWMoM*, Jun. 2007.

[19] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems," *DEC Research Report TR-301*, Sep. 1984.