

Optimal Utility-Based Resource Allocation for OFDM Networks with Multiple Types of Traffic

Mohammad Katoozian, Keivan Navaie
Electrical and Computer Engineering Department
Tarbiat Modares University, Tehran, Iran

Halim Yanikomeroglu
Systems and Computer Engineering Department
Carleton University, Ottawa, Canada

Abstract—This paper considers resource allocation in OFDM wireless networks with mixed real-time and non-realtime traffic patterns. We employ a utility-based framework to balance efficiency and fairness, while satisfying QoS requirements of real-time users. A utility function is assigned to each user over each sub-carrier; this indicates the transmission priority for the user/sub-carrier pair based on the achievable rate and experienced delay, subject to delay and power constraints. The proposed scheduler gives in one shot the sub-carrier and power allocation, plus transmission scheduling for each time-slot. Simulation results indicate that while satisfying real-time users' delay requirements, the proposed method achieves a significant performance improvement in terms of overall throughput and fairness.

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) divides an entire available spectrum into many orthogonal narrow-band sub-carriers (sub-channels) to deal with frequency-selective fading and support a high data rate. In an OFDM wireless network, different sub-carriers can be allocated to different users, which experience different corresponding channel conditions, to provide a flexible multiuser access scheme and to exploit multiuser diversity. Moreover, in OFDM, adaptive power allocation can be implemented for each sub-carrier for further performance improvement.

Optimal power allocation with dynamic sub-carrier assignment in OFDM networks has been studied in [1]. In that work, the authors have investigated optimal resource allocation in multiuser OFDM systems to minimize the total transmission power while satisfying a minimum rate for each user.

Achieving efficiency while keeping a certain level of fairness is a crucial issue when allocating fluctuating resources to flexible delay-tolerant services [2]. On one hand, spectral efficiency is evaluated in terms of the aggregate throughput, which is sometimes unfair to those users which experience poor channel conditions. On the other hand, absolute fairness may lead to low bandwidth efficiency. This becomes even more challenging for inflexible delay-sensitive services with lower delay and error tolerance and Quality-of-Service (QoS) requirements. Therefore, resource allocation in wireless networks with a mixed traffic nature requires an efficient tradeoff between efficiency and fairness, while satisfying the QoS requirements of real-time traffic.

Conventionally, utility functions are used to quantify the benefit of usage of certain resources. In communication networks, utility theory can be used to evaluate the degree to which service requirements of users' applications are met [3]. In wireless

networks, utility and pricing mechanisms were first proposed for resource allocation in the uplink of code division multiple access (CDMA) cellular networks [4]. Utility-based downlink resource allocation for non-realtime traffic in CDMA/TDMA cellular networks has been investigated in [5] as well.

Utility-based cross-layer optimization in OFDM networks, including dynamic sub-carrier assignment (DSA), adaptive power allocation (APA), and joint DSA and APA to balance fairness and efficiency by jointly optimizing the physical and medium access control (MAC) layer, has been proposed in [6], [7]. This results in data rate adaptation over the sub-carriers with better corresponding sub-channel conditions so as to improve throughput and simultaneously ensure an acceptable bit-error rate (BER) on each sub-carrier. However, considering infinite delay tolerance for all users has eased the optimization in this case.

Compared to [7], in this paper we have considered a network with multiple types of traffic and introduced a novel utility function which combines fairness and efficiency and is shown to satisfy the delay requirements of real-time traffic. Moreover, transmit scheduling has been added to the joint DSA and APA problem, and using a different approach we have proposed a low complexity heuristic algorithm to solve this optimization. Limitation on the sub-carrier transmit power, which in practice is a technical requirement, has also been considered in our approach which was not taken into account in [7].

In this paper, we propose a utility-based joint transmission scheduling, dynamic sub-carrier assignment and adaptive power allocation method which exploits multiuser diversity in an OFDM wireless network. Transmission scheduling indicates the users, and the number of packets for each user scheduled for transmission in each time frame. A mixed traffic scenario is considered with delay-sensitive and delay-tolerant services and the optimization is solved while satisfying QoS requirements of delay-sensitive users.

To balance efficiency and fairness, we employ a utility-based framework. In this framework, corresponding to each user being served on a particular sub-carrier, we define a utility function that is a function of the user's experienced delay and the corresponding channel quality. This utility function is defined so that the delay tolerance constraint of real-time users is also taken into account. The total network utility, which is the summation of the utilities of all served users at each time, is then considered as the network performance indicator. In this paper, the main objective is to maximize the total network utility. To do this,

we propose an optimization problem with base-station and sub-carrier power constraints, which is solved using a novel heuristic algorithm.

The distinguishing feature of our proposed method, compared to the previous works, is that while satisfying the QoS requirements of delay-sensitive users, it gives in one shot, the transmission scheduling, the corresponding sub-carriers assigned to each user and the amount of power allocated to each sub-carrier, based on a fair and efficient framework.

II. SYSTEM MODEL

The system parameters and notation used in this paper are shown in Table I.

We consider the downlink of an OFDM wireless network with one transmitter (base-station) and N receivers (users) indexed by j . Traffic is classified into best effort (non-realtime) and delay-sensitive (real-time) traffic. Best effort (BE) traffic has no specific QoS requirements, but real-time (RT) traffic is considered to have delay requirements, in which packets are dropped, thus lost, in case of violating the delay deadlines. Users' data is packetized into fixed length packets with length L . Maximum total base-station transmit power is P and maximum transmit power of a sub-carrier is P_s . The set of S sub-carriers, indexed by i , used in the network is shown by $\mathbb{S} = \{SC_1, SC_2, \dots, SC_S\}$. Rate adaptation is used with K modulation schemes (i.e., K possible bit-rates), indexed by k , which are determined by the bandwidth of each OFDM sub-carrier and the type of modulation used over SC_i at time t between the base-station and user j .

The system is time-slotted into T_S seconds slots, where $T_S \simeq T_C$ and T_C is the channel coherence time. Time-slots are indexed by t . The required power for transmission at rate k between the base-station and user j over SC_i at time t , $p_{ijk}(t)$, is determined by [6]:

$$r_k(t) = \log_2(1 + \beta p_{ijk}(t) \rho_{ij}(t)) \quad (1)$$

where $\rho_{ij}(t)$ is the signal-to-noise ratio (SNR) of user j over the i th sub-carrier when the transmission power is unity, i.e., $\rho_{ij}(t)$ indicates the quality of user j 's channel over this sub-carrier. The SNR gap, β , is a constant indicating the difference between the SNR needed to achieve a certain data transmission rate for a practical system and the theoretical limit, which is related to the targeted bit-error rate (BER) by [6]:

$$\beta = \frac{1.5}{-\ln(5 \cdot \text{BER})}. \quad (2)$$

Equation (1) clearly indicates that more transmission power is required either when transmitting at higher data rates, with worse channel conditions, or when requiring a higher level of transmission accuracy (i.e., lower BER).

For each active user in the network, a queue is associated in which packets are waiting to be transmitted by the base-station. It is assumed that in each time-slot, each user obtains and sends the channel conditions, $\rho_{ij}(t)$, over each sub-carrier to the base-station, so the network is able to obtain $p_{ijk}(t)$ for $i = 1, \dots, S$, $j = 1, \dots, N$ and $k = 1, \dots, K$. Obviously, and due to the frequency-selective nature of the wireless channel being

TABLE I
SYSTEM PARAMETERS

Parameter	Description
N	Number of users
S	Number of sub-carriers
K	Number of modulation schemes
$u_{ijk}(t)$	Sub-carrier i 's utility when assigned to user j with rate k
$r_k(t)$	Data rate when using modulation scheme k
$p_{ijk}(t)$	Power to send user j 's data over sub-carrier i with rate k
β	SNR gap (see (2))
ρ_{ij}	User j 's SNR over sub-carrier i
$\theta_j(t)$	Normalized experienced delay (see (3))
$\tau_j(t)$	Experienced delay by user j until time t (s)
$\bar{\tau}(t)$	Average experienced delay (s)
$U(t)$	Total network utility
P	Maximum base-station transmit power
P_s	Maximum sub-carrier transmit power

considered, users will experience deep fading (i.e., low $\rho_{ij}(t)$) for some sub-carriers, and good channel conditions (i.e., high $\rho_{ij}(t)$) on some other sub-carriers. Note that sub-carriers which experience deep fading for one user may not be in a deep fade for others, which creates the opportunity to exploit multiuser diversity.

In this paper, we consider a utility function which is an increasing function of the wireless link quality, data rate and delay for a specific user. For a user j , a utility function $u_{ijk}(t)$ is associated that indicates the profit earned by the network as a result of transmitting user j 's data at time t over SC_i with rate k . Different utility functions are defined for BE and RT users in order to give priority to RT users approaching their delay deadlines. The utility function for a BE user is a function of $r_k(t)$ and $\theta_j(t)$, where $\theta_j(t)$ is a normalized indicator of user j 's experienced delay up to time t such that

$$\theta_j(t) = \frac{\tau_j(t) - \bar{\tau}(t)}{\bar{\tau}(t)}. \quad (3)$$

In the above $\tau_j(t)$ is the amount of time user j has spent in the system until time t and $\bar{\tau}(t)$ is the average experienced delay over all users in queue at time t . Here we utilize a normalized version of delay in order to model delay fairness.

The utility function for a RT user, $u_{ijk}(t)$, is an increasing function of $\tau_j(t)$ and $r_k(t)$. The utility function indicates a transmission priority metric for each user over all sub-carriers and rates. Therefore, RT users close to their delay deadline are given a boost in their corresponding utility function in order to prevent violation of the delay requirement.

We define the total network utility at time t , $U(t)$, as the network performance indicator:

$$U(t) = \sum_{i=1}^S \sum_{j=0}^N \sum_{k=1}^K u_{ijk}(t) x_{ijk}(t) \quad (4)$$

where $x_{ijk}(t) \in \chi_x(t)$ is the assignment indicator; $x_{ijk}(t) = 1$ if sub-carrier i is assigned to user j with transmission rate k , and $x_{ijk}(t) = 0$, otherwise, and $\chi_x(t)$ is defined as

$$\chi_x(t) = \{x_{ijk}(t) | i = 1, \dots, S, j = 0, \dots, N, k = 1, \dots, K\}. \quad (5)$$

We also consider a NULL user $j = 0$ for which $u_{i0k}(t) = 0$ and $p_{i0k}(t) = 0$ for all i and k . If $x_{i0k}(t) = 1$, sub-carrier i is not assigned to any user at time t .

III. PROBLEM FORMULATION

The utility function, $u_{ijk}(t)$, serves as an optimization objective for packet transmission and is a function of the allocated network resources to each user as well as the Quality-of-Service experienced by that user. The earned profit modelled by the utility function also provides a priority metric for each user served by the base-station; the higher the value of a utility function, the higher the priority of transmitting to the corresponding user. Since there is no separate scheduling for RT users, delay requirements of these users shall be reflected in their corresponding utility functions in order to earn them a higher transmission priority when approaching the delay deadline.

In the utility-based framework, it is desirable to maximize the profit earned by the network; therefore our objective is to maximize the total network utility in (4). The optimization problem is constrained first of all, by the fact that one sub-carrier cannot be assigned to more than one user, although many sub-carriers may be assigned to one user. Yet, the crucial resource constraint is the maximum power allowed to be assigned to each sub-carrier and the overall power available at the base-station.

Let P be the maximum total base-station transmit power. Therefore, the total power required to send data to users, scheduled for transmission in each time interval according to the assignment indicator $x_{ijk}(t)$ in (4) should be less than or equal to P . In order to prevent allocating large amounts of power to a few sub-carriers, a constraint is set on the power assigned to each individual sub-carrier. This constraint makes sure that some sub-carriers don't eat up all the available power. According to these limitations, we define Problem \mathcal{O} as follows:

Problem \mathcal{O} :

$$\max_{x_x(t)} U(t), \quad (6)$$

$$\text{s.t.} \quad \sum_{j=0}^N \sum_{k=1}^K p_{ijk}(t)x_{ijk}(t) \leq P_s \quad i = 1, \dots, S \quad (7)$$

$$\sum_{i=1}^S \sum_{j=0}^N \sum_{k=1}^K p_{ijk}(t)x_{ijk}(t) \leq P \quad (8)$$

$$\sum_{j=0}^N \sum_{k=1}^K x_{ijk}(t) = 1 \quad i = 1, \dots, S \quad (9)$$

$$x_{ijk}(t) \in \{0, 1\} \quad (10)$$

where P_s in (7) indicates the maximum power available to each sub-carrier and P in (8) is the total power constraint at the base-station. Furthermore, (9) and (10) ensure that each sub-carrier is assigned to at most one user.

IV. HEURISTIC ALGORITHM

The proposed heuristic algorithm for solving Problem \mathcal{O} uses the Lagrange Multipliers approach [8] as a basis for the approximation of the optimization solution. This approach is a modification of the one already presented in [5]. Since we have to solve \mathcal{O} for each time-slot t , hereafter we drop the time index t for brevity.

```

Initial Assignment
s1.
    maxjk{uijk}|pijk ≤ Ps, xijk ← 1, for all i

Adjustment Body
s2.1. Reset the Lagrange multipliers:
    λi ← 0, for all i
s2.2. Compute the base-station resource
constraint violation:
    π ← ∑i=1S ∑j=0N ∑k=1K pijkxijk - P

while π > 0
    s2.3. Determine the maximum assigned
    power (to be changed):
    i*j*k* ← arg maxijk{pijk}
    s2.4. Find the new assignment:
    for j = 0 : N
        for k = 1 : K
            Δjk ← (ui*j*k* - uijk - λi(pi*j*k* - pijk))
        end
    end
    J*K* ← arg minjk{Δjk}, for all j, k
    s2.5. Make the new assignment:
    λi* ← λi* + ΔJ*K*
    xi*j*k* ← 0, xi*J*K* ← 1
    π ← π - pi*j*k* + pi*J*K*
end

```

Fig. 1. The Initial Assignment and Adjustment Body algorithm.

In [8], it is shown that the solution of the unconstrained optimization problem

$$\max \left\{ \sum_{i=1}^S \sum_{j=0}^N \sum_{k=1}^K (u_{ijk} - \lambda_i p_{ijk}) x_{ijk} \right\} \quad (11)$$

is the solution of the constrained optimization problem \mathcal{O} , where λ_i is Lagrange multiplier and $x_{ijk}^* = 1$ if $u_{ijk} - \lambda_i p_{ijk} > 0$, and 0, otherwise.

The Lagrange Multipliers approach might result in more than one solution, among them the one which satisfies (9) is the optimal solution. As a matter of fact, if $\lambda_i, i = 1, \dots, S$, are known, the solution of \mathcal{O} can be obtained easily. If these multipliers are computed so that $P - \sum_{j=0}^N \sum_{k=1}^K p_{ijk} x_{ijk}^* \geq 0$, then the solution satisfies (8), thus is feasible. The solution becomes optimal if

$$\sum_{i=1}^S \lambda_i (P - \sum_{j=0}^N \sum_{k=1}^K p_{ijk} x_{ijk}^*) = 0. \quad (12)$$

A. The Algorithm

The proposed algorithm has two main procedures which are executed consequently: the *Initial Assignment* and the *Adjustment Body* (Fig. 1).

The algorithm starts in the *Initial Assignment* by assigning the largest possible utility, not violating the maximum sub-carrier power constraint, among all users and data rates to each sub-carrier in (s1). At this point, the sub-carrier power constraints in (7) are fulfilled, but the overall base-station power constraint in (8) may be violated which will be corrected in the next procedure by adjusting the assignments.

After the completion of the *Initial Assignment*, adjustment of the previous assignments is done to hold the base-station total transmit power constraint in (8). The *Adjustment Body* starts by

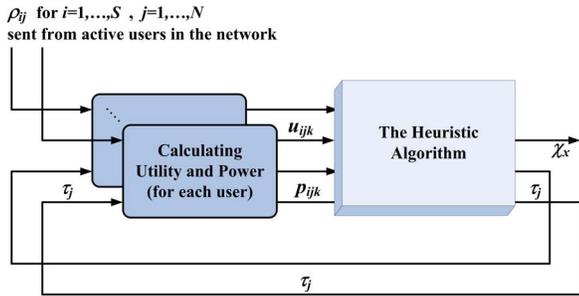


Fig. 2. System block diagram.

setting the Lagrange multipliers to zero in (s2.1) and finding the overall constraint violation in (s2.2). Next, in a while loop, while the constraint violation exists, the sub-carrier with the maximum assigned power is found in (s2.3). The increase in the Lagrange multiplier, resulting from exchanging the previously assigned user and data rate, by all other users and data rates is computed in (s2.4). Among all, the user and data rate which cause the least Lagrange multiplier increase, while still holding the sub-carrier power constraint in (7) are chosen, the exchange is done and the corresponding parameters are updated accordingly in (s2.5).

This new assignment minimizes the gap between the optimal solution and the previous assignment. The while loop in the *Adjustment Body* is repeated until the overall resource constraint in (8) is held.

B. Implementation and Computational Complexity

The overall system block diagram is shown in Fig. 2. The optimization is performed in a base-station which is transmitting to its active set of users. The inputs of the proposed algorithm are u_{ijk} and p_{ijk} , and the output is the assignment matrix χ_x . At the beginning of each time-slot, each user measures the channel qualities ρ_{ij} over each sub-carrier and sends them to the base-station. Users' experienced delay τ_j , and related queue length are also available at the base-station. Therefore, u_{ijk} and p_{ijk} are obtainable. The optimal assignment matrix is then found using the proposed algorithm. The base-station assignments and scheduling can be communicated to the corresponding users through signalling channels.

In the following we show that the proposed algorithm has a polynomial-time computational complexity.

In the *Initial Assignment*, (s1) has the complexity order of $O(NK)$. In the *Adjustment Body*, (s2.1) has the complexity order of $O(S)$, and (s2.2) has the complexity order of $O(SNK)$. In the while loop, (s2.3) and (s2.5) have the complexity order of $O(SNK)$ and $O(1)$, respectively. In (s2.4) there is one iteration for each user and modulation of the sub-carrier being changed its assignment, resulting in the total complexity order of $O(NK)$. In every iteration of (s2.4), one assigned user/rate is removed from one sub-carrier, thus, in the worst case the while loop is executed SNK times. This gives an overall complexity of $O(SN^2K^2)$ for the *Adjustment Body*. Thus the overall computational complexity order is $O(SN^2K^2)$.

TABLE II
SIMULATION PARAMETERS.

Parameter	Value
Base-Station transmit power (P)	10 W
Sub-carrier maximum transmit power (P_s)	1 W
No. of sub-carriers (S)	64
No. of active users (N)	40
No. of modulation schemes (K)	4
BER	10e-3
Channel coherence time (T_C)	10 ms
Time-slot duration (T_S)	10 ms
Delay deadline for voice users	50 ms
Propagation loss exponent	4

V. SIMULATIONS

We simulate a single cell network, containing one base-station which performs the optimization over a number of users. The number of active users in the network is N . Half of the users are assumed to have RT (voice) traffic with maximum allowable delay (delay deadline) of 50 ms. The rest of the users are BE users. RT packets with τ_j exceeding the delay deadline are dropped and discarded. Uniform user spatial distribution is considered, and for each user, fixed length packets are generated by a Poisson arrival process with an average rate of λ packets per second. For a BE user j , transmitting over SC_i at rate k , similar to our previous work [5], we define the following utility function:

$$u_{ijk} = r_k \exp(\theta_j). \quad (13)$$

Note that in (13), a packet transmitted on the i th sub-carrier at rate k is given a large utility either when the corresponding channel is good or user j experiences a bad relative delay status. Similarly, for a RT user j , transmitting over SC_i at rate k , we define the following utility function:

$$u_{ijk} = (\tau_j/T_S + 1)r_k \exp(\tau_j). \quad (14)$$

The $(\tau_j/T_S + 1)$ coefficient is applied to provide a higher utility for voice users compared to BE users. This coefficient can be relaxed or strengthened for other real-time applications with weaker or stronger delay requirements.

Based on the proposed scheme, parameters r_k , τ_j and θ_j are available in the BS. The heuristic algorithm proposed in the previous section is used to solve Problem \mathcal{O} as an approximation of the optimization solution. This simulation parameters are presented in Table II.

For comparison, we consider three different systems. In *System I*, the utility function in (14) is used for RT users and the utility function in (13) for BE users. In *System II*, the utility function in (13) is used for all users, so the delay requirements of RT users is not taken into consideration. Resource allocation in *System III* is based solely on maximizing the total throughput; therefore fairness and delay requirements are not considered and the utility function is equal to $r_k(t)$. All three systems utilize the heuristic algorithm mentioned in Section IV for solving the optimization problem.

Fig. 3 illustrates the overall throughput of the three discussed systems versus the packet arrival rate. The simulation is run for 60 consecutive time-slots and this process is iterated 100 times. At the end, the results are averaged for all iterations and time-slots and shown in Fig. 3. This figure indicates that *System III*

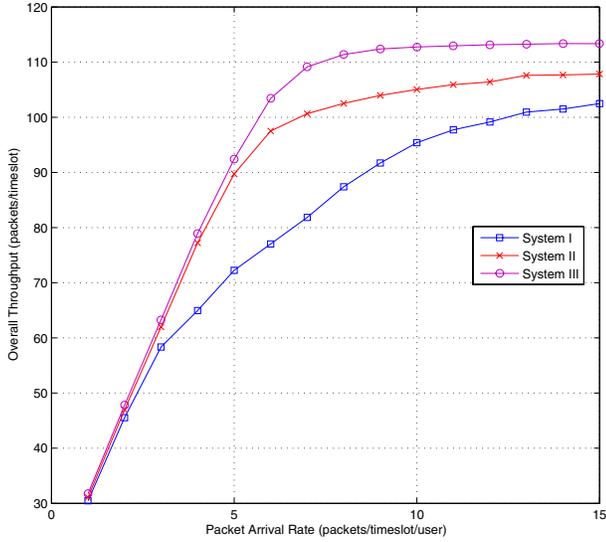


Fig. 3. Overall throughput vs. packet arrival rate

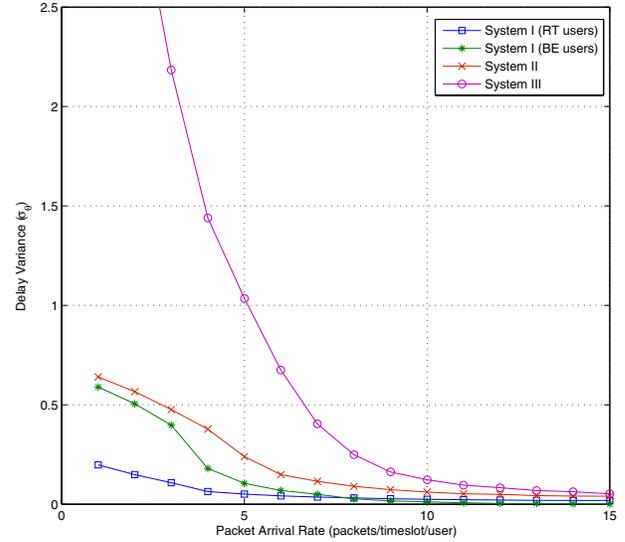


Fig. 5. Delay variance vs. packet arrival rate

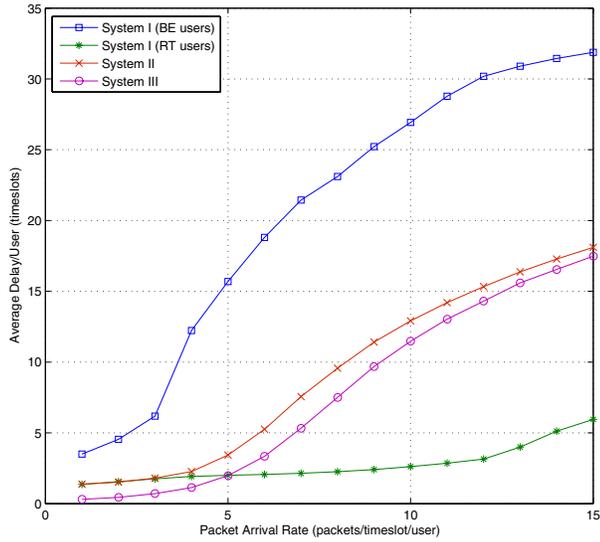


Fig. 4. Average Delay/user vs. packet arrival rate

which ignores fairness and delay issues, and *System II* which ignores RT delay requirements, achieve a 10-15 percent better overall throughput compared to *System I*.

In Fig. 4, the average delay per user is shown for the three systems. The figure shows that the proposed scheme (i.e. *System I*) has longer delay for BE users compared to the other two systems, however, the average delay of RT users were kept below the delay deadline (i.e. $5 \times T_S$).

As an indicator of fairness in Fig. 5, we compare the variance of the users' experienced delay, for the three systems. It can be observed that *System I* and *System II* yield in higher overall fairness by having a very low delay variance; zero delay variance would mean equal delay for all users, thus absolute fairness. It is interesting to see that the mere 10-15 percent decrease in throughput seen in Fig. 3 due to employing the utility-based framework is compensated by achieving fairness in allocating the network resources.

VI. CONCLUSION

In this paper, we have presented a joint transmit scheduling, sub-carrier assignment and power allocation scheme to exploit multiuser diversity in OFDM-based wireless networks with multiple types of traffic and QoS requirements. The utility-based framework is used to balance the efficiency and fairness of resource allocation. The proposed method gives in one shot, the transmission scheduling, the sub-carriers assigned to each user and the amount of power allocated to each sub-carrier, based on a fair and efficient framework. Simulation results indicate a significant performance improvement in terms of the decreased delay variance, therefore a high degree of fairness is achieved while maintaining over 85 percent of the total achievable throughput and satisfying the delay requirements of real-time users.

REFERENCES

- [1] C. Y. Wong, *et al.*, "Multiuser OFDM with adaptive sub-carrier, bit, and power allocation," *IEEE JSAC*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.
- [2] K. Navaie, *et al.*, "Fairness of resource allocation in cellular networks: A survey," *Resource Allocation in Next Generation Wireless Networks*, W. Li and Y. Pan, Eds. Nova Science Publishers, 2005.
- [3] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33-37, Jan. 1997.
- [4] C. U. Saraydar, *et al.*, "Pricing and power control in a multicell wireless data network," *IEEE JSAC*, vol. 19, no. 10, pp. 1883-1892, Oct. 2001.
- [5] K. Navaie and H. Yanikomeroglu, "Optimal Downlink Resource Allocation for Non-realtime Traffic in Cellular CDMA/TDMA Networks," *IEEE Communications Letters*, vol. 10, pp. 278-280, April 2006.
- [6] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks-Part I: Theoretical Framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614-624, March 2005.
- [7] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks-Part II: Algorithm Development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625-634, March 2005.
- [8] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operation Research*, vol. 11, no. 3, pp. 399-417, May-June 1963.