Utility-Based Adaptive Radio Resource Allocation in OFDM Wireless Networks with Traffic Prioritization

Mohammad Katoozian, Keivan Navaie, *Member, IEEE*, and Halim Yanikomeroglu, *Member, IEEE*

Abstract—In this letter, a joint transmit scheduling and dynamic sub-carrier and power allocation method is proposed to exploit multi-user diversity in downlink packet transmission in an OFDM wireless network with mixed real-time and non-real-time traffic patterns. To balance efficiency and fairness and to satisfy the QoS requirements of real-time users, we utilize a utility-based framework and propose a polynomial-time heuristic algorithm to solve the formulated optimization problem. The distinguishing feature of the proposed method is that it gives in one shot, the transmission scheduling, the sub-carriers assigned to each user, and the power allocated to each sub-carrier, based on a fair and efficient framework while satisfying the delay requirements of real-time users.

Index Terms—Fairness, multi-user diversity, OFDM networks, radio resource allocation, utility function.

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 multi-user OFDM wireless networks, channel characteristics for different users are almost mutually independent; the sub-carriers experiencing deep fading for some users may not be in a deep fade for others. Therefore, each sub-carrier could experience "good" channel conditions for some users; hence using adaptive resource allocation in appropriate time-scales, multi-user diversity gain [1] could be achieved. Moreover, in OFDM, adaptive power allocation can be implemented for each sub-carrier for further performance improvement.

Achieving efficiency while keeping a certain level of fairness is a crucial issue when allocating fluctuating radio 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.

Utility theory can be used in communication networks to evaluate the degree to which a network satisfies service requirements of users' applications [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 power allocation in CDMA downlink for voice and data applications has been proposed in [5]. Utility-based downlink resource allocation for non-real-time traffic in CDMA/TDMA cellular networks has been investigated in [6] as well.

Optimal power allocation with dynamic sub-carrier assignment in OFDM networks has been studied in [7], [8]. In [7], the authors have investigated optimal resource allocation in multi-user OFDM systems to minimize the total transmission power while satisfying a minimum rate for each user. Scheduling and resource allocation in OFDM networks including sub-carrier and power allocation has been studied in [9] as well. In [10], [11], resource allocation and packet scheduling in OFDMA systems in considered and a scheduler is proposed to distribute radio resources among users according to their individual QoS requirements and traffic class. A proportional rate adaptive resource allocation method for OFDM systems is proposed in [12] where sub-carrier and power allocation are carried out sequentially to reduce the complexity, and an optimal power allocation procedure is derived, through which proportional fairness is achieved. In [13], an adaptive resource allocation method is proposed for cellular OFDM systems which includes dynamic cell selection.

Utility-based optimization in OFDM networks, including dynamic sub-carrier and power allocation to balance fairness and efficiency by jointly optimizing the physical and medium access control (MAC) layer, has been proposed in [14], [15]. Utilizing utility-based optimization results in data rate adaptation over the sub-carriers with better corresponding sub-channel conditions. This improves throughput while simultaneously ensuring 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 [15], in this letter we have considered an OFDM 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, transmission scheduling has been added to the joint dynamic sub-carrier assignment and adaptive power allocation 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 [15].

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 consider a utility-function that is a function of the user's experienced delay, the corresponding channel quality and the user's delay requirements. 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 letter, 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 then solved using a novel heuristic algorithm.

Simulation results indicate a significant performance improvement in terms of the decreased delay variance amongst users, at the same time gaining over 85% of the achievable throughput and satisfying real-time users' delay requirements.

II. SYSTEM MODEL

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-real-time) and real-time traffic. Best-effort traffic has no specific QoS requirements, but real-time 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 bits.

Maximum total base-station transmit power is P Watts and the maximum transmit power of a sub-carrier is P_s Watts. The set of S sub-carriers, indexed by i, used in the network is shown by $\mathbb{S} = \{SC_1, SC_2, \ldots, SC_S\}$. Rate adaptation is used with K pre-selected modulation schemes (i.e., K possible bit-rates), indexed by k, which are determined by the bandwidth of each 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:

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

where $\rho_{ij}(t)$ is the signal-to-noise ratio (SNR) of user j over SC_i when the transmission power is unity; $\rho_{ij}(t)$ 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 [14]: $\beta = \frac{1.5}{-\ln(5*\text{BER})}$. Equation (1) clearly indicates that more transmission power is required when either 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 $i = 1, \ldots, S$, $j = 1, \ldots, N$, and $k = 1, \ldots, K$. Obviously, due to the frequency-selective nature of the wireless channel being considered, users will experience deep fading (i.e., low $\rho_{ij}(t)$) for some sub-carriers, and good channel conditions (i.e., high $\rho_{ij}(t)$) for 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 multi-user diversity.

For a user j, we associate a utility-function $u_{ijk}(t)$ 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. The utility-function is a function of $r_k(t)$, $\theta_j(t)$, and $\alpha_j(t)$:

$$u_{ijk}(t) = \mathcal{F}\big(r_k(t), \theta_j(t), \alpha_j(t)\big), \tag{2}$$

where $r_k(t)$ is the transmission rate and $\theta_j(t)$ is a normalized indicator of user j's experienced delay such that

$$\theta_j(t) \triangleq \frac{\tau_j(t) - \bar{\tau}(t)}{\bar{\tau}(t)}. \tag{3}$$

In the above, $\tau_j(t)$ is the amount of time user j has spent in queue up to time t, and $\bar{\tau}(t)$ is the average experienced delay over all users in queue at time t. The parameter $\alpha_j(t)$, is the QoS coefficient related to user j. The value of this parameter is higher for users with more strict QoS requirements, giving them a higher priority for transmission. Note that in (3), we utilize a normalized version of delay in order to model proportional fairness in the experienced delay.

In this letter, we consider a utility-function, $\mathcal{F}\big(r_k(t), \theta_j(t), \alpha_j(t)\big)$, which is an increasing function of the wireless link quality (i.e., achievable data-rate $r_k(t)$) and the experienced normalized delay, $\theta_j(t)$, for a specific user. Using $\alpha_j(t)$, different utility-functions are defined for best-effort and real-time users in order to give priority to real-time users approaching their delay deadlines. Therefore, the utility-function is a transmission priority metric for each user over all sub-carriers and rates. Thus, real-time 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 consider a NULL user $j = 0$ for which $u_{i0k}(t) = 0$ and $p_{i0k}(t) = 0$ for all sub-carriers and modulation schemes. We then define the total network utility at time t, $U(t)$, as the network performance indicator:

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

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

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

If $x_{i0k}(t) = 1$, sub-carrier i is assigned to the NULL user at time t, meaning no transmission is scheduled on this sub-carrier over this time-slot.

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 real-time users, delay requirements of these users are reflected in their corresponding utility-functions by means of the QoS coefficient 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, $U(t)$. 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/or the overall power available at the base-station. In the following, based on the different constraints that can be applied to solve this optimization problem, we define two different problems based on the possible power constraints.

A. Constrained Total Base-station Transmit Power

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 matrix, $\chi_x(t)$, in (5) should be less than or equal to P. According to this limitation, we define Problem $\mathcal{O}1$ as follows:

Problem $\mathcal{O}1$:

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

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

$$\sum_{j=0}^{N}\sum_{k=1}^{K} x_{ijk}(t) = 1 \qquad i = 1,\ldots,S \tag{8}$$

$$x_{ijk}(t) \in \{0,1\} \tag{9}$$

where (7) is the total power constraint at the base-station, and (8) and (9) ensure that each sub-carrier is assigned to only one user. The fact that $x_{ijk}(t)$ may be zero for some user j over all i, k shows that transmission scheduling is implicitly handled in this formulation. A constraint-relaxed version of this problem has been examined in [9]. Also, in [10], an optimization problem similar to $\mathcal{O}1$ has been considered, and to reduce complexity, the problem is reduced to a linear one by equally distributing the transmit power among sub-carriers.

Solving the optimization problem $\mathcal{O}1$, yields in finding the assignment indication matrix, $\chi_x(t)$, which indicates the users scheduled for transmission and the amount of resources allocated to each user to maximize the total network utility. The number of packets scheduled for transmission for each user is not one of the optimization variables, however by solving the optimization problem and specifying the scheduled users and their corresponding modulation scheme, the number of packets per user is also obtained. Note that here we consider a *static setup* in which our main objective is to maximize the total network utility in *each time slot* without taking the future utility into consideration.

Next, we show that $\mathcal{O}1$ is a Multiple-Choice Knapsack Problem (MCKP).

Definition MCKP [16]: Suppose that there are S classes $N_1, N_2, \ldots, N_S$ of items to be packed in a knapsack with capacity P. Each item $j \in N_i$ has a profit u_{ij} and a weight p_{ij}, and the problem is to choose one item from each class such that the profit sum is maximized without having the weight sum exceed P.

We consider each sub-carrier as a class of $K(N + 1)$ items corresponding to $N + 1$ users including NULL over K bit-rates to be packed in a knapsack of capacity P. Each item in the knapsack resembles a user and one of the K possible modulation schemes. The profit of each item in the knapsack is the corresponding utility-function $u_{ijk}(t)$ and the required resource (weight) is $p_{ijk}(t)$. The objective of the optimal downlink transmission in Problem $\mathcal{O}1$ is to choose exactly one user/modulation (i.e., item) for each sub-carrier (i.e., class) to maximize the total network utility, subject to the total power constraint. Therefore, Problem $\mathcal{O}1$ is an MCKP instance, which is NP-Hard.

B. Constrained Base-station and Sub-carrier Power

Since the constraint in Problem $\mathcal{O}1$ is set on the total base-station transmit power, there is no supervision on the power of individual sub-carriers. This may lead to allocating large amounts of power to a few sub-carriers. To resolve this problem, we set a limit on the maximum amount of power assigned to each sub-carrier. This leads to Problem $\mathcal{O}2$.

Problem $\mathcal{O}2$:

$$\max_{\chi_x(t)} \quad U(t), \tag{10}$$

$$\text{s.t.} \quad \sum_{j=0}^{N}\sum_{k=1}^{K} p_{ijk}(t)x_{ijk}(t) \leq P_s \quad i = 1,\ldots,S \tag{11}$$

$$\sum_{i=1}^{S}\sum_{j=0}^{N}\sum_{k=1}^{K} p_{ijk}(t)x_{ijk}(t) \leq P \tag{12}$$

$$\sum_{j=0}^{N}\sum_{k=1}^{K} x_{ijk}(t) = 1 \qquad i = 1,\ldots,S \tag{13}$$

$$x_{ijk}(t) \in \{0,1\} \tag{14}$$

where P_s in (11) indicates the maximum power available to each sub-carrier and P in (12) is the total power constraint at the base-station. Furthermore, (13) and (14) ensure that each sub-carrier is assigned to at most one user.

Problem $\mathcal{O}2$ is a modified MCKP which additionally applies a limitation on the weight of each item in the knapsack, i.e., items which have a weight exceeding the maximum allowed weight per item are not chosen. Note that adding sub-carrier power constraint in (11) to the NP-Hard problem $\mathcal{O}1$, reduces the computational complexity by limiting the

number of possible assignments. The amount of reduction in the computational complexity depends on $SP_s - P$, and in the case where the amount of $SP_s - P$ is negative, the problem reduces to a simple assignment problem. However, our work addresses the case where $P \ll S.P_s$

This formulation is the most comprehensive, taking into account both of the existing resource constraints. In the following section we propose a heuristic algorithm for solving Problem $\mathcal{O}2$.

IV. HEURISTIC ALGORITHM

The proposed heuristic algorithm for solving Problem $\mathcal{O}2$ uses the Lagrange Multipliers approach [17] as a basis for the approximation of the optimization solution. The Lagrange Multipliers Theorem has been used before in [6] to solve a constrained optimization problem in CDMA networks. Since we have to solve $\mathcal{O}2$ for each time slot t, hereafter we drop the time index t for brevity. Utilizing the Lagrange Multipliers approach, the solution of the constrained optimization problem $\mathcal{O}2$ is the solution of the following unconstrained optimization problem

$$\max \left\{ \sum_{i=1}^{S} \sum_{j=0}^{N} \sum_{k=1}^{K} (u_{ijk} - g_i p_{ijk}) x_{ijk} \right\} \quad (15)$$

where $g_1, \ldots, g_S$ are S non-negative Lagrange multipliers. The obvious solutions are $x^*_{ijk} = 1$ if $u_{ijk} - g_i p_{ijk} > 0$, and 0, otherwise. However, these solutions do not necessarily satisfy (13).

The Lagrange Multipliers approach might result in more than one solution, among them the one which satisfies (13) is the optimal solution. As a matter of fact, if $\{g_i\}_{i=1}^{S}$, are known, the solution of $\mathcal{O}2$ can be obtained easily. If these multipliers are computed so that $P - \sum_{i=1}^{S} \sum_{j=0}^{N} \sum_{k=1}^{K} p_{ijk} x^*_{ijk} \geq 0$, then the solution satisfies (12), thus is feasible.

To efficiently compute the Lagrange multipliers, we first find the sub-carrier with the most offending resource constraint violation, and exchange the user/modulation assigned to the sub-carrier with the user/modulation which results in the least Lagrange multiplier increase. Meanwhile, to satisfy (11), only user/modulation pairs not violating the sub-carrier power constraint are considered for exchange. The proposed algorithm, shown in Fig. 1, has a polynomial-time computational complexity which can be shown has a complexity order of $O(SN^2K^2)$.

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 $u_{ijk}(t)$ and $p_{ijk}(t)$, and the output is the assignment matrix, $\chi_x(t)$. At time t, each user measures the channel QoS $\rho_{ij}(t)$ over each sub-carrier and sends them to the base-station. Users' experienced delay, $\tau_j(t)$, and QoS coefficient, $\alpha_j(t)$, are also available at the base-station. Therefore, $u_{ijk}(t)$ and $p_{ijk}(t)$ 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.

```
Initial Assignment
s1.
    max_{jk}{u_{ijk}}|p_{ijk} ≤ P_s, x_{ijk} ← 1, for all i
Adjustment Body
s2.1. Reset the Lagrange multipliers:
    g_i ← 0, for all i
s2.2. Compute the base-station resource
constraint violation:
    π ← ∑_{i=1}^{S} ∑_{j=0}^{N} ∑_{k=1}^{K} p_{ijk}x_{ijk} − P

while π > 0
    s2.3. Determine the maximum assigned
    power (to be changed):
    for all {i,j,k : x_{ijk}(t) = 1}
        i*j*k* ← arg max_{ijk}{p_{ijk}}
    s2.4. Find the new assignment:
        for j = 0 : N
            for k = 1 : K
            Δ_{jk} ← (u_{i*j*k*} − u_{i*jk} − g_{i*}(p_{i*j*k*} − p_{i*jk}))
            end
        end
    J*K* ← arg min_{jk}{Δ_{jk}}, for all j, k
    s2.5. Make the new assignment:
        g_{i*} ← g_{i*} + Δ_{J*K*}
        x_{i*j*k*} ← 0, , x_{i*J*K*} ← 1
        π ← π − p_{i*j*k*} + p_{i*J*K*}
end
```

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

ρ_{ij} for $i=1,\ldots,S$, $j=1,\ldots,N$
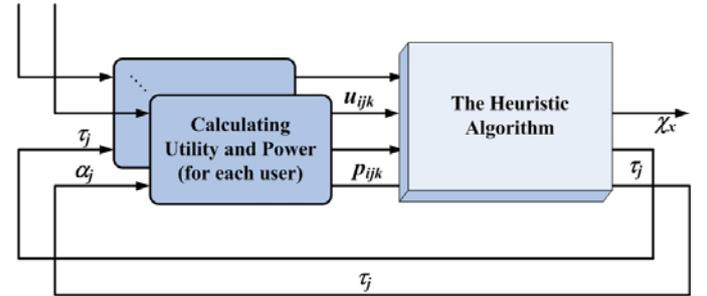sent from active users in the network

Fig. 2.   System block diagram.

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. 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. It is assumed that each OFDM sub-carrier is subject to flat Rayleigh fading plus path loss and shadowing. The shadowing is constant over the whole bandwidth. The simulation parameters are presented in Table I.

First of all, we tested the heuristic algorithm against an exact algorithm as far as the memory requirements of the exact algorithm allowed us. Comparative results showed that while the heuristic is faster than the exact algorithm in all cases and this gap widens exponentially when increasing the number of

TABLE I
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
Modulation schemes	BPSK, QPSK, 16QAM, 64QAM
Packet size (L)	128 Bytes
Average SNR	15 dB
BER	10e-3
T_C	10 ms
T_S	10 ms
Propagation loss exponent	4

users/sub-carriers, the accuracy of the heuristic according to the tests is higher than 97%.

For a user j transmitting over SC_i at rate k, similar to our previous work [6], we define the following utility-function:

$$u_{ijk}(t) = \alpha_j(t) r_k(t) \exp(\theta_j(t)). \tag{16}$$

Note that the utility function $u_{ijk}(t)$ in (16) is defined so as to equalize the effect of delay, throughput and QoS constraints in allocating radio resources and assigning the priority of transmission to each user over each time interval. This utility function has been chosen similar to the exponential rule [18], however issues regarding the possibility of stabilizing the system using this exponential utility remain open to be discussed. The utility function used in this letter is not the only possible choice, and other variations of this function could have also been used (e.g., [2]).

Half of the users are assumed to have real-time (voice) traffic with maximum allowable delay of 50 ms. The rest of the users are best-effort users. Real-time packets with τ_j exceeding the delay deadline are dropped and discarded. The QoS coefficient, which is applied to provide a higher utility for voice users compared to best-effort users, is defined as:

$$\alpha_j(t) = \begin{cases} 1 & \text{Best-effort users} \\ T_j + 1 & \text{Real-time users} \end{cases} \tag{17}$$

where $T_j = \tau_j / T_S$ is the number of time-slots user j has spent in the system with data in queue up to time t. This coefficient can be relaxed or strengthened for other real-time applications with weaker or stronger delay requirements.

For comparison, we consider three different systems. In *System I*, the utility-function in (16) along with the QoS coefficient in (17) is used for all users. In *System II*, the utility-function in (16) with $\alpha_j(t) = 1$ is used so the delay requirements of real-time users is not taken into account. Resource allocation in *System III* is based on maximizing the total throughput; therefore fairness and delay requirements are not considered and the utility-function is equal to $r_k(t)$.

Fig. 3 illustrates the overall throughput of the three discussed systems versus packet arrival rate. This figure indicates that *System III* which ignores fairness and delay issues, and *System II* which ignores delay requirements, achieve a 10-15% better overall throughput compared to *System I*. It is also seen that the throughput difference between *System I* and the other two systems is larger for low packet arrival rates. This,
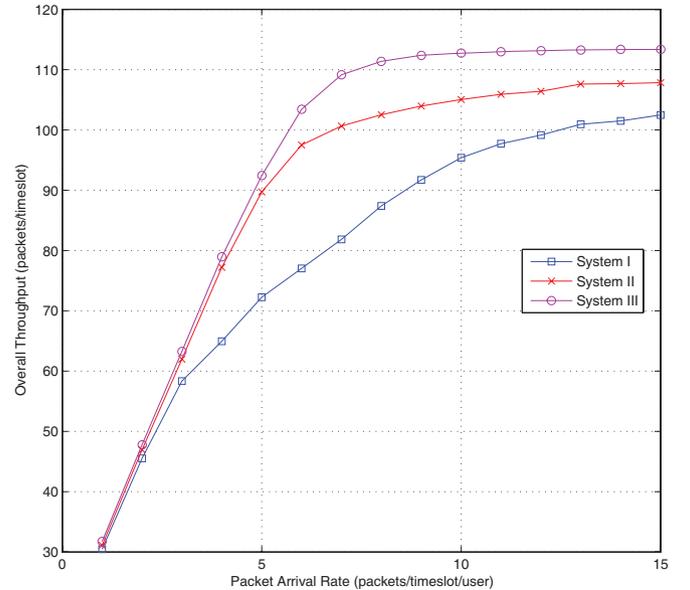

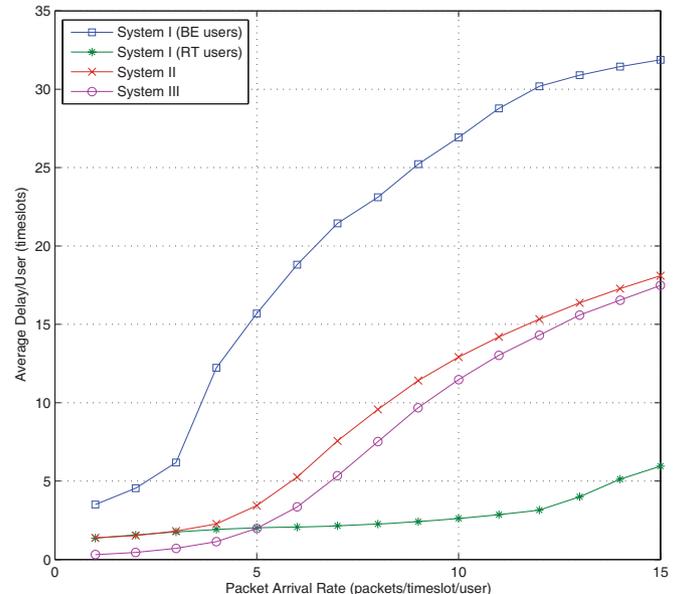Fig. 3. Overall throughput *vs.* packet arrival rate.

Fig. 4. Average delay/user *vs.* packet arrival rate.

in fact is the region where we can observe the influence of the scheduling mechanism; giving priority to real-time users to meet their delay deadlines which results in lower bandwidth efficiency. At very low packet arrival rates, the deadlines are easily met, and at high arrival rates, lots of packets are dropped and the scheduler isn't doing any good for real-time users, so the lower difference is justifiable.

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 best-effort users compared to the other two systems, however, the average delay of real-time 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. If sub-carrier allocation is based solely on maximizing the
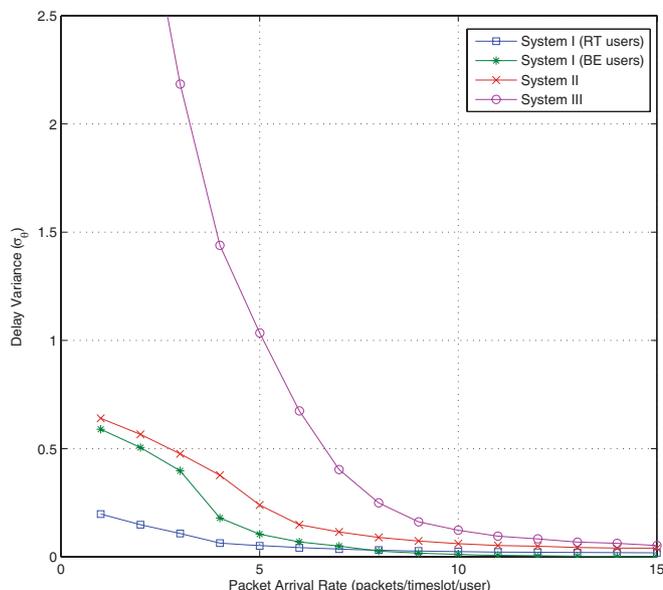
Fig. 5.　Delay variance *vs.* packet arrival rate.

throughput, resources may not be allocated over a long period of time to BE users with poor SNR, which leads to an increase in the delay variance between users. To prevent such cases, the proposed algorithm implements proportional fairness among users by including a delay factor in the utility function, with the goal of equalizing the delay experienced by all users. To measure our success in equalizing the delay among users, we have considered the variance of the delay experienced by all users as the metric. It can be observed that *System I* and *System II* yield in higher overall fairness by having a lower delay variance; zero delay variance meaning equal delay for all users, thus absolute fairness. It is seen that the mere 10-15% 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 letter, we have presented a joint transmit scheduling and dynamic sub-carrier and power allocation scheme to exploit multi-user 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% of the total achievable throughput and satisfying the delay requirements of real-time users.

REFERENCES

[1] P. Viswanath, D. Tse, and R. Larioa, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 12771294, June 2002.

[2] K. Navaie, D. Y. Montuno, and Y. Q. Zhao, "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 Trans. Telecommun.*, vol. 8, pp. 33-37, Jan. 1997.

[4] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 18831892, Oct. 2001.

[5] P. Liu, M. Honig, and S. Jordan, "Forward link CDMA resource allocation based on pricing," in *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 3, pp. 1410-1414, Chicago, IL, USA, Sept. 2000.

[6] K. Navaie and H. Yanikomeroglu, "Optimal downlink resource allocation for non-real-time traffic in cellular CDMA/TDMA networks," *IEEE Commun. Lett.*, vol. 10, pp. 278-280, Apr. 2006.

[7] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multi-user OFDM with adaptive sub-carrier, bit, and power allocation," *IEEE J. Selected Areas Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.

[8] W. Rhee and J. M. Cioffi, "Increase in capacity of multi-user OFDM system using dynamic sub-carrier allocation," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, pp. 10851089, Tokyo, Japan, May 2000.

[9] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," in *Proc. 40th Annual Conf. Inform. Sciences and Systems (CISS)*, pp. 1272-1279, Princeton, NJ, Mar. 2006.

[10] L. Yanhui, W. Chunming, Y. Changchuan, and Y. Guangxin, "Downlink scheduling and radio resource allocation in adaptive OFDMA wireless communication systems for user-individual QoS," in *Proc. World Academy of Sci., Eng. and Technol.*, vol. 12, pp. 221-225, Mar. 2006.

[11] L. Yanhui, Z. Lizhi, Y. Changchuan, and Y. Guangxin, "Adaptive radio resource allocation for multiple traffic OFDMA broadband wireless access system," in *Proc. World Academy of Sci., Eng. and Technol.*, vol. 12, pp. 231-236, Mar. 2006.

[12] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems," in *Proc. IEEE Global Telecommun. Conf. 2003*, vol. 1, pp. 337-341, Dec. 2003.

[13] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1566-1575, Sept. 2004.

[14] G. Song and Ye (Geoffrey) Li, "Cross-layer optimization for OFDM wireless networks-Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614-624, Mar. 2005.

[15] G. Song and Ye (Geoffrey) Li, "Cross-layer optimization for OFDM wireless networks-Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625-634, Mar. 2005.

[16] M. Moser, P. Jokanovic, and N. Shiratori, "An algorithm for the Multidimensional Multiple-choice Knapsack Problem," *IEICE Trans. Fundamentals*, vol. E80-A, no. 3, pp. 582-589, Mar. 1997.

[17] 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.

[18] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *American Mathematical Society Translations*, ser. 2, Y. M. Suhov, ed., 2002.