

Cross Layer Optimization for Fairness Balancing Based on Adaptively Weighted Utility Functions in OFDMA Systems

Jianwei Wang, Timo Korhonen, Yuping Zhao

Abstract—Cross layer optimization based on utility functions has been recently studied extensively, meanwhile, numerous types of utility functions have been examined in the corresponding literature. However, a major drawback is that most utility functions take a fixed mathematical form or are based on simple combining, which can not fully exploit available information. In this paper, we formulate a framework of cross layer optimization based on Adaptively Weighted Utility Functions (AWUF) for fairness balancing in OFDMA networks. Under this framework, a two-step allocation algorithm is provided as a sub-optimal solution, whose control parameters can be updated in real-time to accommodate instantaneous QoS constrains. The simulation results show that the proposed algorithm achieves high throughput while balancing the fairness among multiple users.

Keywords—OFDMA, Fairness, AWUF, QoS.

I. INTRODUCTION

The unique characteristics of wireless channels create many challenging issues for efficient resource management. Cross layer design is becoming increasingly important for improving the performance of wireless networks [1], [2]. By simultaneously optimizing the network control across multiple layers, cross-layer design can substantially increase network capacity, reduce interference and power consumption [3], [4].

An effective trade-off among spectral efficiency, fairness, and QoS requirement is always desired in wireless resource allocation. The issues of efficient and fair resource allocation have already been well studied in economics analysis, where utility functions are used to quantify the benefit of usage of certain resources. Similarly, utility theory can also be used in communication networks to evaluate the degree to which a network configuration satisfies specific service requirements of each user's applications, rather than in terms of system-centric quantities like throughput, outage probability, packet drop rate, and power [6]. The rationale of utility pricing structures is to map the resource (bandwidth, power, etc.) or performance criteria (data rate, delay, etc.) into the corresponding utility or price values and to optimize the established utility pricing function. An important aspect of this paper is that we strive to search total system utility, i.e., summation of the utility for each individual user, as the primary performance measurement. Utility theory, on the other hand, provides

means to formulate the relations between user experience-level information and various network performance matrices.

Various types of utility functions have been addressed in recent literature. The problem of maximum utility resource management for wired networks was first addressed by Shenker in [6]. One of the earliest utility related wireless management was revealed by Bianchi et al., who proposed a utility-fair scheduling algorithm that achieves equal utility to all users [8]. The efforts by Gao explored a maximizing global utility scheme while considering channel quality variations [9]. A similar utility maximizing issue for best effort traffic was addressed by Jiang in [7], in which log-type and exponential-type utility functions as well as simple hybrid types were considered. However, most of the utility functions designed in literature are simple summations of individual utility elements. In this paper, we propose an adaptively weighted utility function (AWUF) as an optimization objective for OFDMA networks, which makes a better use of queuing information available. A two-step sub-optimal solution is also provided under this cross layer framework.

Orthogonal frequency division multiplexing (OFDM) has been identified as one of the most promising schemes for broadband wireless networks, e.g., digital video broadcasting DVB-T (EN 300744), wireless LAN (802.11), and WiMAX (802.16). The scheme of transmitting high speed data via multiple, parallel low-rate streams presents excellent characteristic for flexible and cost-efficient implementation and high performance over frequency selective fading channel. Here we address an OFDMA system where each user occupies a subset of subchannels, and each subchannel is assigned exclusively to only one user at any time. The advantage of OFDMA over OFDM-TDMA and OFDM-CDMA is the high suppression of intra-cell interference. This avoids the need of CDMA type multi-user detection meaning reduction or even cancellation of near-far effect [5]. In this paper, we apply the proposed AWUF as the optimization objective in OFDMA networks and evaluate the performance of our two-step solution.

The rest of this paper is organized as follows. Section II presents the system model. Section III introduces our proposed AWUF and formulates the optimization problem. A two-step sub-optimal solution is provided in section IV and the simulation results are presented in V. The summarizing conclusions are given in Section VI.

Jianwei Wang is with the Satellite and Wireless Communications Lab, Peking University, China, email: wjw.pku@gmail.com

Timo Korhonen is Adjunct Professor in the Communications Lab, Helsinki University of Technology, Finland, email: Timo.Korhonen@hut.fi

Yuping Zhao is Professor in the Satellite and Wireless Communications Lab, Peking University, China, email: Yuping.Zhao@pku.edu.cn.

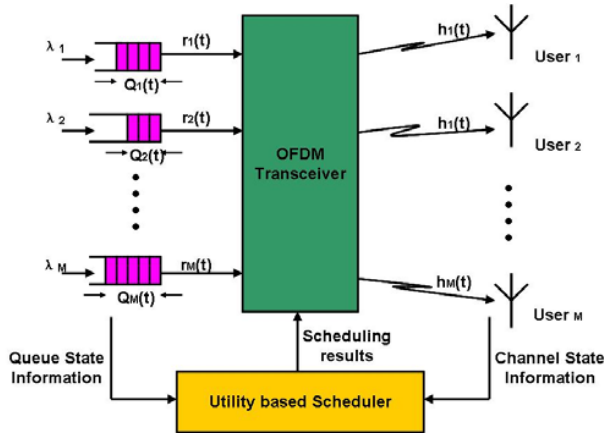


Fig. 1. The System Model

II. SYSTEM MODEL

In this section, the framework of cross layer design is described for the analytical modeling tools addressed, including channel model, rate control and frequency/power allocations. The structure for our OFDMA system to be cross layer optimized is shown in Fig. 1. We consider single cell in OFDMA networks. Assume that we have M users within the range of the base station (BS). The BS keeps a queue for each user and applies cross layer scheduling for data transmission.

A. Fading Channel Model

Let us assume a fading multipath channel h_i for each user (Fig. 1). The complex baseband expression for the impulse response of the wireless channel for the i th user is:

$$h_i(t, \tau) = \sum_{k=1}^L \alpha_{k,i}(t) \delta(\tau - \tau_{k,i}), \quad (1)$$

where L is the number of multi-path rays, $\alpha_{k,i}(t)$ is the amplitude of the k th path and $\tau_{k,i}$ is the corresponding delay of the k th path. The channel impulse response in frequency domain is obtained as:

$$H_i(f, t) = \int_{-\infty}^{\infty} h_i(t, \tau) e^{-j2\pi f \tau} d\tau = \sum_{k=1}^L \alpha_{k,i}(t) e^{-j2\pi f \tau_{k,i}}. \quad (2)$$

Hence, the SNR of the i th user at frequency f is [13]:

$$\gamma_i(f) = \frac{p(f) |H_i(f)|^2}{N_i(f)}, \quad (3)$$

where $p(f)$ is the transmission power density and $N_i(f)$ is the noise power density function of user i .

We assume a slowly fading subchannel that maintains the channel state constant within OFDM symbol's time. Keeping the system aware of the channel state is in practise an important point to consider: Pilot symbols and feedback channels are required in frequency division duplexed systems. In time division duplexed systems with symmetrical up and down links a practical arrangement could be to measure the uplink channel

state and use this to estimate the downlink channel. However, in practical systems, using separate frequency bands for uplink and downlink, channel symmetry might not always apply.

B. Adaptive Techniques in Cross Layer Design

1) *Rate Control*: Rate control techniques can efficiently utilize channel state information and improve system spectral efficiency [10], [11]. Adaptive rate control strives to approach the Shannon's channel capacity by ensuring an acceptable BER for all the subchannels. X.Qiu et al.[12] provide an exact expression for achievable data rate while satisfying a certain BER value:

$$T_i(f) = \log_2(1 + k\gamma_i(f)), \quad (4)$$

where $T_i(f)$ is the achievable throughput of the i th user at frequency f and $\gamma_i(f)$ is the SNR of the i th user at the corresponding frequency f . Parameter k is a constant for a specific target BER value, eg.

$$k = \frac{1.5}{-\ln(5BER)}. \quad (5)$$

The parameter k is also named as "SNR gap" as it indicates the difference between the SNR needed to achieve a certain data transmission for a practical system and the theoretical limit.

2) *Adaptive Subchannel Allocation and Power Allocation*: Let us now consider the framework of cross layer design in global energy constrained OFDMA networks with a known channel states of all users. For optimization of channel allocations we need to address two important questions:

- First, for a specific subchannel, which user should it be allocated for?
- Second, how much power should we assign to the allocated user?

The answer to the first question is the adaptive subchannel allocation (ASA) and the answer to the second question is just the adaptive power allocation (APA) technique. Some algorithms have been proposed recently to jointly optimize ASA and APA [13], [14] that we will apply here to get a deeper insight to the problem.

III. PROBLEM FORMULATION

As shown in Fig. 1, the scheduling algorithm at the MAC layer is modeled as an optimization problem with respect to some physical layer constraints as well as for application of QoS constraints. At every timeslot, the scheduling algorithm has to optimize the rate allocation $\mathbf{r} = (r_1, \dots, r_M)$ as well as power allocation $\mathbf{p} = (p_1, \dots, p_M)$ for all the M users based on the observation results of the current channel state information (CSI) from the physical layer and the queue state information (QSI) from the application layer. Considering the input and output of scheduling algorithm, we can formulate the cross layer optimization as follows:

A multi-user wireless network can be modeled as a dynamic system with states evolution given by the following:

- *Channel State Information*: The channel fading states $\tilde{\mathbf{H}}(t) = [H_1(t), \dots, H_M(t)]$, where H_i is the channel fading state between BS and the i th user.

- **Queue State Information:** The queue state information contains two sets – the queue length and the queue delayed time:
 - The queue length state denotes the number of untransmitted packets in the buffer of users and is indicated by: $\mathbf{Q}(t) = [Q_1(t), \dots, Q_M(t)]$.
 - The queue delayed time is the delay time of the last arrived packet in the queue and is indicated by: $\tilde{\mathbf{T}}(t) = [T_1(t), \dots, T_M(t)]$

The objective of the optimization problem is:

$$(\mathbf{r}, \mathbf{p}) = \operatorname{argmax} \left\{ U(\tilde{\mathbf{H}}(t), \tilde{\mathbf{Q}}(t), \tilde{\mathbf{T}}(t)) \right\}. \quad (6)$$

This refers to the fact that we should get an optimized data rate vector and the power allocation which maximizes the global utility. In this case, equation (6) is subject to the following constraints:

$$\bigcup_{i=1}^M D_i \subseteq [0, B], \quad (7)$$

$$D_i \cap D_j = \emptyset \quad i \neq j \text{ and } i, j = 1, 2, \dots, M, \quad (8)$$

$$\frac{1}{M} \int_0^B p(f) df \leq 1, \quad (9)$$

$$p(f) \geq 0. \quad (10)$$

where D_i is the subchannel allocation for the i th user, $p(f)$ is the power allocation over the bandwidth B .

Under the above general framework of cross layer optimization, from the user's QoS point of view, we propose our Adaptively Weighted Utility Function (AWUF) as:

$$U(\mathbf{r}) = \sum_{i=1}^M \omega_i(t) \log(r_i + 1). \quad (11)$$

where $\omega_i(t)$ is an adaptive weight to user i . The constant value 1 inside the logarithmic function is intended to ensure that the user i receives zero utility when its allocated data rate r_i is zero [7]. The adaptive weight $\omega_i(t)$ takes both $Q_i(t)$ and $T_i(t)$ into account and hence, has a general form:

$$\omega_i(t) = \frac{F[Q_i(t)] + G[T_i(t)]}{2}, \quad (12)$$

where $F[Q_i(t)]$ quantifies the queue length effect and $G[T_i(t)]$ quantifies the queue delay effect. Both $F[Q_i(t)]$ and $G[T_i(t)]$ should be monotonically increasing to enable unique optimization result. What is more, these two functions can be updated in real-time following instantaneous variations of QoS constraints. E.g., stringent delay limits requires larger $G[T_i(t)]$ while small buffer size leads to a bigger $F[Q_i(t)]$. Therefore, AWUF schema provides flexible control of specified QoS.

IV. SUB-OPTIMAL SOLUTION

In this section, we discuss the solutions to the AWUF based cross layer optimization presented above. For simplicity, we assume:

$$F[Q_i(t)] = \exp[\alpha Q_i(t)] - 1, \quad (13)$$

$$G[T_i(t)] = \exp[\beta T_i(t)] - 1. \quad (14)$$

where α and β are the scaling factor corresponding to different QoS constrains. Both (13) and (14) satisfy that none priority will be obtained if $Q_i(t)$ and $T_i(t)$ are zero value. So, the objective utility function (11) can be described by:

$$U(\mathbf{r}) = \sum_{i=1}^M \frac{(e^{\alpha Q_i(t)} - 1) + (e^{\beta T_i(t)} - 1)}{2} \cdot \log(r_i + 1). \quad (15)$$

that should be solved following constraints (7)-(10).

To our best knowledge, it is the first time that a combined priority considering both queue length and queue delay is introduced in target utility function as described in (15). Considering solving the above, linear program solvers are not well-suited due to the fact that the objective function is nonlinear. Thus we have developed a sub-optimal algorithm addressing both weighted channel allocation and water-filling simultaneously.

A. Weighted Channel Allocation

Let us first consider the channel state and queue state information for subchannel allocation, that we call the first step of the algorithm. We assume the transmit power density $p(f) = 1$ for each user. Also, assume a certain subchannel j allocated for the i th user. Now, the available data rate will be:

$$r(i, j) = \int_0^{B_j} \omega_i(t) \log_2(1 + k\gamma_i(f)) \cdot \quad (16)$$

The subchannel will then be allocated for the m th user by following (15) as:

$$m = \operatorname{arg max}_{i=1, \dots, M} \{r(i, j)\}, \quad (17)$$

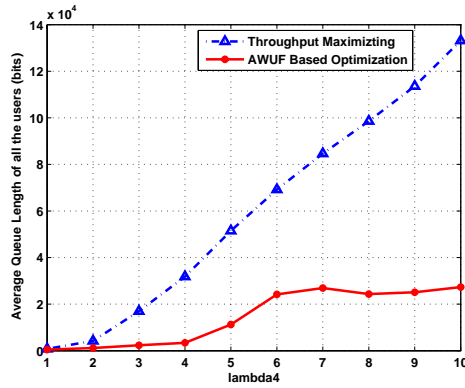
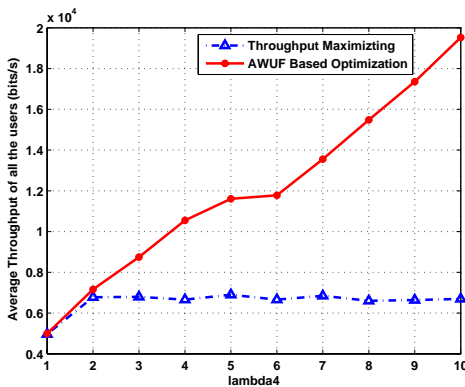
where m is the user who has the largest $r(i, j)$ over all the M users.

B. Water-filling On Weighted Channel Gain Matrix

After the first step of subchannel allocation, we get the channel gain matrix $\{\mathbf{H}_{select}\}$, where the matrix element is the selected channel gain $H_m(t)$ inherited from the first step that we should implement the power allocation. The well-know optimal power allocation in OFDM system follows the water-filling principle. However, in our framework of AWUF based optimization, we should also consider the queue state information. So, we should do water-filling based on weighted channel gain matrix \mathbf{H}_w , where the matrix element is weighted to be $\omega_m(t) * H_m(t)$. Therefore, as a result, we perform water-filling power allocation on $\{\mathbf{H}_w\}$ instead of $\{\mathbf{H}_{select}\}$.

V. SIMULATION RESULTS

In this section, the simulation results of the proposed AWUF based cross layer optimization are provided. We applied the two-step sub-optimal solution in our system of Fig. 1 and focused on the performance of balancing fairness under unbalanced traffic among different users.

Fig. 2. Average queue length vs. λ_4 Fig. 3. Average throughput vs. λ_4

A. Fairness Balancing

In this subsection, we evaluate the performance of the AWUF based cross layer optimization comparing to the Throughput Maximizing (TM) schema, whose objective function is $U(\mathbf{r}) = \sum_{i=1}^M r_i$ [7]. The simulation parameters are listed in Table I. We fix α, β , data arrival rates $\lambda_1 \sim \lambda_3$ and vary λ_4 to examine the fairness balancing ability of the two algorithms. The average queue length and throughput of all the users are shown in Fig. 2 and Fig. 3. From Fig. 3, we can see that the average throughput of TM algorithm keeps steady when λ_4 is bigger than 2. The reason is that the TM algorithm doesn't consider the queue state information and cannot increase the throughput of user 4 although λ_4 is much larger when compared to $\lambda_1 \sim \lambda_3$. As a result, the average queue length of TM increases in Fig. 2. On the other hand, our proposed AWUF based solution applies the queue state information and enhances the transmission possibility of user 4 when λ_4 grows. So, the AWUF based solution highly reduces average queue length in Fig. 2 and increases average throughput in Fig. 3.

B. Control Parameters α and β

From equation (15), we can see that the effect of α and β is quite the same. Here, we take β as an example to discuss

TABLE I
SIMULATION PARAMETERS FOR FIG. 2 AND FIG. 3

Subchannel bandwidth	10 KHz
Number of subchannels	32
Number of users	4
Data arrival rate of User 1 (λ_1)	0.1(pkts/s)
Data arrival rate of User 2 (λ_2)	0.5 (pkts/s)
Data arrival rate of User 3 (λ_3)	1 (pkts/s)
Data arrival rate of User 4 (λ_4)	varies from 1~10 (pkts/s)
Packet length	128 bits
Target BER	1.0×10^{-5}
BER gap k , eqn.(4)	0.1515
α	1
β	1

the effects of the two controlling parameters. According to our analysis in the previous section, both α and β can be adaptively updated in order to accommodate current QoS constrains, e.g., β should be increased to satisfy stringent delay constrain while making the α bigger should decrease the buffer size. We will note that this controlling scheme of α and β in AWUF based optimization leads to different performance under different traffic loads.

Fig. 4 shows the performance of controlling β when traffic load is low. The data arrival rate vector $[\lambda_1, \lambda_2, \lambda_3, \lambda_4]$ is set to $[0.1, 0.5, 1, 2.5]$. In this low traffic load scenario, we fix all the other parameters and vary β , which means changing the delay constrain level. We can see that the average queue length decreases in Fig. 4(a) and the average throughput increases in Fig. 4(b) while β is increased. The reason is that when we enhance the delay requirement by increasing β , the transmission possibility of those users who buffer a larger queue increases. However, when we fix $[\lambda_1, \lambda_2, \lambda_3, \lambda_4]$ to a higher traffic load as $[1, 5, 12, 18]$, performance is completely reverse as shown in Fig. 5. The reason is that when the traffic load is high, most of the users keep a large queue and demand for transmission all the time. Therefore, an intelligent way to enhance average throughput is to perform channel-aware adaptation by neglecting queue state information.

VI. CONCLUSION

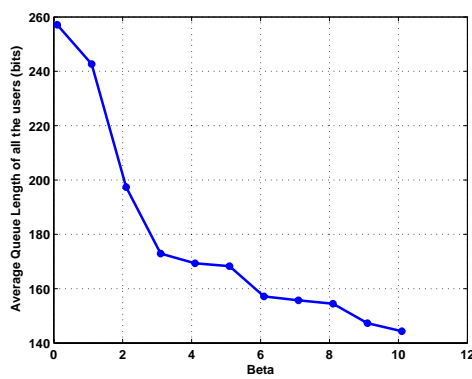
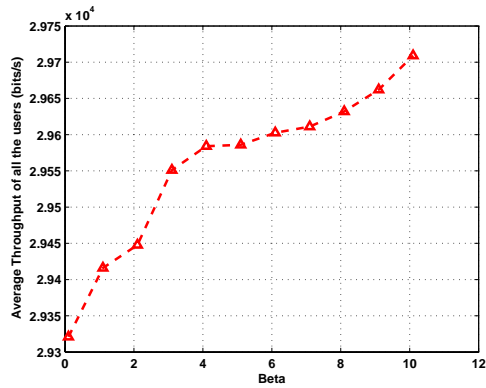
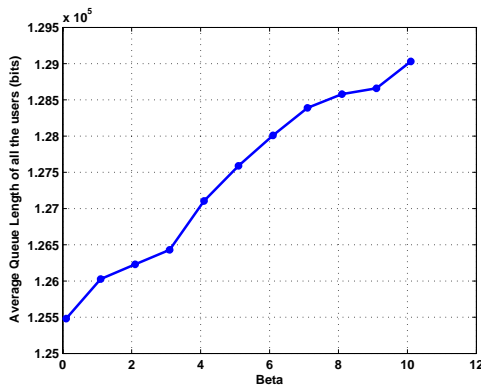
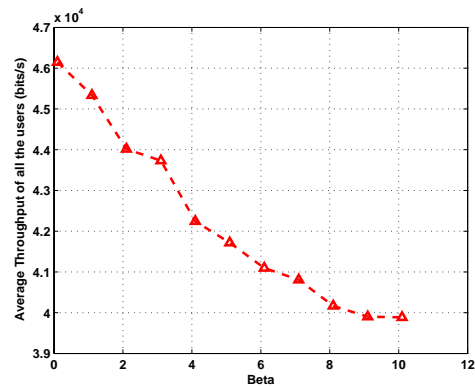
In this paper, we have proposed a cross layer optimization framework based on Adaptively Weighted Utility Function (AWUF) and provided a sub-optimal solution. The suggested AWUF design makes good use of available information including CSI and QSI. Comparing to the traditional TM algorithm, the AWUF scheduling can achieve much higher efficiency in fairness balancing as demonstrated by our simulations. In addition, the suggested weighted control has the potential to follow the instantaneous QoS requirement in real-time enabling better adaptation to both channel and user conditions.

ACKNOWLEDGMENT

This research work is financially supported by the Chinese National High-Tech 863 Project (No. 2006AA01Z278).

REFERENCES

- [1] S. Toumpis; A. J. Goldsmith; "Capacity regions for wireless Ad Hoc networks," IEEE Transactions on Wireless Communications, vol. 2, no. 4, pp. 736-748, July 2003.

(a) Average Queue Length vs. β (b) Average Throughput vs. β Fig. 4. Performance of β control when the traffic load is low with $[\lambda_1, \lambda_2, \lambda_3, \lambda_4]=[0.1, 0.5, 1, 2.5]$ (a) Average Queue Length vs. β (b) Average Throughput vs. β Fig. 5. Performance of β control when the traffic load is high with $[\lambda_1, \lambda_2, \lambda_3, \lambda_4]=[1.5, 12, 18]$

- [2] M. Johansson; L. Xiao; "Scheduling, routing and power Allocation for fairness in wireless networks," in IEEE VTC - Spring, Milan, Italy, May 2004.
- [3] M. Chiang; "To layer or not to layer: balancing transport and physical layers in wireless multihop networks," in Proceedings of IEEE INFOCOM, Hong Kong, March 2004.
- [4] X. Lin; N. B. Shroff; "Joint rate control and scheduling in multihop wireless networks," in Proceedings of the IEEE Conference on Decision and Control, Paradise Island, Bahamas, December 2004.
- [5] Li, Z.; Latva-Aho, M.; "Nonblind and semiblind Space-Frequency multi-user detection for multirate MC-CDMA systems," Signal Processing, Volume 54, Issue 11, Nov. 2006
- [6] S. Shenker; "Fundamental design issues for the future internet," IEEE JSAC, vol. 13, no. 7, Sept. 1995, pp. 1176-88.
- [7] Jiang, Z.; Ge, Y.; Li, Y.; "Max-utility wireless resource management for best-effort traffic," Wireless Communications, IEEE Transactions on Volume 4, Issue 1, Jan. 2005 Page(s):100-111
- [8] G. Bianchi; A. T. Campbell; R. R.-F. Liao; "On utility-fair adaptive services in wireless networks," in Proc. 6th Int. Workshop Quality of Service, Napa, CA, May 18-20, 1998, pp. 256-267.
- [9] X. Gao; T. Nandagopal; V. Bharghavan; "Achieving application level fairness through utility-based wireless fair scheduling," in Proc. IEEE GLOBECOM, Nov. 2001, pp. 3257-3261.
- [10] Goldsmith; A. J.; "Variable-rate variable-power MQAM for fading channel," IEEE Trans. Commun., vol. 45, pp. 1218-1230, Oct. 1997.
- [11] Nanda, S.; Balachandran, K.; Kumar, S.; "Adaptation techniques in wireless packet data services," IEEE Commun. Magazine, pp. 54-64, Jan. 2000.
- [12] X. Qiu; K. Chawla; "On the performance of adaptive modulation in cellular systems", IEEE Trans. Commun., vol. 47, no. 6, pp. 884-895, Jun. 1999.
- [13] Guocong Song; Ye Li; "Cross layer optimization for OFDM wireless networks-Part I: Theoretic Framework," IEEE transaction on wireless communication vol.4. NO. 2, 2005;
- [14] Guocong Song; Ye Li; "Cross layer optimization for OFDM wireless networks-Part II: Algorithm Development," IEEE transaction on wireless communication vol.4. NO. 2, 2005;

Jianwei Wang Jianwei Wang is currently a PhD candidate in Peking University, China. From October, 2006 to July, 2007, he has been selected by China Government to study in Helsinki University of Technology (HUT) which is financially supported by Chinese Scholarship Council (CSC). His research interests include OFDM and OFDMA systems, WiMax communication systems, UWB communication, RFID systems, wireless communication and wireless networks.

Timo Korhonen Dr Timo O. Korhonen is currently Adjunct Professor of Telecommunications Systems engineering in Communications Laboratory, Helsinki University of Technology (HUT), Finland. Prior this, he worked in TKK/Communications Laboratory as Associate Professor in 1999-2006. His research interests include telecommunications technology, intelligent transport systems; high-tech product development process; project, organizational and strategic management; user centric design and structured innovation processes. He has worked as Technical Program Committee member in various international conferences of IEEE. Also, he works as a reviewer for several international journals.

Yuping Zhao Yuping Zhao received her Master degree from Northern Jiaotong University, Beijing China in 1986, and Doctor of Science degree from Helsinki University of Technology, Finland in 1999. She has been working with Railway Communications and Signaling Company, China; Nippon Data System Company, Japan; Nokia Research Center, Finland. Currently she is a professor at Peking University. Dr. Zhao's main research fields are physical layer and MAC layer of wireless communication systems, which include Coding and Modulations, Mobile Radio Channel estimations, OFDM and MIMO communication Systems, Radio Resource Management, wireless sensor networks, RFID systems.