

Cross Layer Designs for OFDMA Wireless Systems with Heterogeneous Delay Requirements

David Shui Wing Hui*, Vincent Kin Nang Lau** and Wong Hing Lam*

*Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

**Department of Electrical and Electronic Engineering, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong

Email: swdhui@eee.hku.hk, ceknlau@ee.ust.hk, whlam@eee.hku.hk

Abstract—This paper investigates a cross layer scheduling scheme for OFDMA wireless system with heterogeneous delay requirements. Unlike most existing cross layer designs which take a decoupling approach, our design considers both queueing theory and information theory in modeling the system dynamics. The cross layer design is formulated as an optimization of total system throughput, subject to individual user's delay constraint and total base station transmit power constraint. The optimal scheduling algorithm for the delay-sensitive cross layer optimization is to dynamically allocate radio resources based on users' channel state information, source statistics and delay requirements. Specifically, optimal power allocation was found to be multilevel water-filling where urgent users have higher water-filling levels, while optimal subcarrier allocation strategy is shown to be achievable by low complexity greedy algorithm. Simulation results also show the proposed jointly optimal power and subcarrier allocation policy can provide substantial throughput gain with all delay constraints being satisfied.

I. INTRODUCTION

Because of the robust performance of Orthogonal Frequency Division Multiplexing (OFDM) over frequency selective channel, OFDM can support high speed data transmission through multiple subcarriers and it has been proposed as the modulation and multiple access schemes for next generation networks such as IEEE 802.16 wireless broadband access system.

Conventional multiuser OFDM system, e.g. OFDM-FDMA and OFDM-TDMA, with fixed subcarrier allocation scheme only allows a single user to transmit on all of the subcarriers or a fixed subset of subcarriers [1] and thus fails to exploit the inherent multi-user diversity in the time varying wireless channel. OFDMA with cross layer scheduling exploits this multi-user diversity, by carefully assigning multiple users to transmit simultaneously on the different subcarriers for each OFDM symbol with optimal power and rate allocations. With the dynamic subcarrier and power allocation in OFDMA systems, the overall system throughput can be increased significantly. There are considerable amount of existing works on cross layer scheduling design for OFDMA systems (see for example, [2, 3, 4] and references therein). However, in most works, the cross layer designs were based on a decoupled approach where the source statistics and the queue dynamics were ignored from the physical layer information theoretical models. In other words, these cross layer designs were targeted for delay insensitive applications only. The authors in [5] attempt to design cross layer schedulers that can incorporate

both the source statistics and queue dynamics into a simple On-Off physical layer model. In [6, 7], the authors proposed a cross layer algorithm based on combined information theory [8] and queueing theory [9] to minimize the average system delay. In [10], the authors consider cross layer scheduling through utility based optimization on optimizing delay related utility. The authors in [11] proposed a heuristic urgency based allocation policy for Multiuser MISO wireless system with zero forcing beam-forming strategy. Yet, it is not clear how good the proposed heuristic allocation policy performs compared with the optimal performance. Furthermore, in all these designs, they are targeted for systems with homogeneous users only.

In this paper, we shall focus on the cross layer scheduling design for OFDMA systems which consider the effect of the source statistics and queue dynamics in addition to the information theoretical physical layer model. We consider delay-sensitive systems with mixed traffics having heterogeneous delay requirements. Compared to the conventional assumption on homogeneous traffic in cross layer design, it would be a more realistic assumption on traffic nature of next generation wireless network which requires simultaneous heterogeneous multimedia transmission of different delay requirements. Unlike [4] which simplified the physical layer into a simple ON-OFF model, we consider a more sophisticated information theoretical model to capture the performance of the OFDMA physical layer. The delay-sensitive cross layer scheduling problem is formulated into a mixed convex and combinatorial optimization problem. Based on the optimization framework, we deduce the jointly optimal power and subcarrier allocation strategy that satisfies the heterogeneous delay requirements. We found that the optimal power allocation strategy is given by multi-level water-filling where users with tighter delay constraint will have a higher "water-level". The optimal user selection (or sub-carrier allocation) strategy is greedy in nature which has a linear complexity with respect to number of users.

This paper is organized as follows. Section II describes the system model, including channel model, multiuser physical layer model, source model and MAC layer model. Section III presents the formulated optimization problem and the corresponding power and subcarrier allocation policy. Simulation results are studied in Section IV and a conclusion is given in Section V.

II. SYSTEM MODEL

This section outlines the OFDMA system model which is the basis of the resource allocation problem formulated in Section III. The conceptual system architecture of cross-layer scheduling model for multiuser downlink OFDM scheduler is shown in Figure 1. At the beginning of each scheduling slot, a central resource scheduler in the base station is fed with queue state information (QSI) and channel state information (CSI) about the mobile users. The resource scheduler made a scheduling decision based on this information and passes the resource allocation scheme to the OFDM transmitter. According to this subcarrier, power and rate allocation scheme, the transmitter picks up the appropriate number of bits from different users for all subcarriers to form an OFDM symbol. The update process of status information of all users and also the decision process are made once every time slot. The subcarrier allocation and power allocation decision made by the Base Station transmitter is assumed to be announced to individual mobile user through a separate control channel with the help of MAC layer protocol. The following assumptions are also made throughout the presentation of this paper. We assumed the power and subcarrier allocation process is centralized at Base Station. Perfect channel state information is available at the transmitter (CSIT) and receiver (CSIR). It is also assumed that transmission rate chosen from a continuous set is realizable and perfect channel coding on each subcarrier can be performed according to the channel characteristic. We further assumed an infinite buffer for the arrival packets.

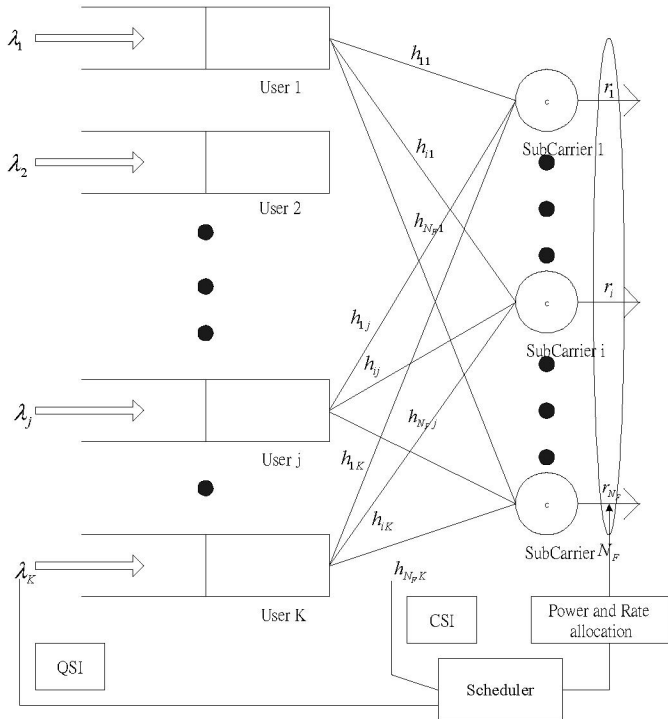


Figure 1. Cross-layer scheduling model under Conceptual Channel Model for OFDMA system with heterogeneous users

A. Downlink Channel Model

We consider an OFDMA system with quasi-static fading channel within a scheduling slot t_s (2ms). This is a reasonable

assumption for users with pedestrian mobility where the coherence time of the channel fading is around 20ms or more. It is also assumed each subcarrier has bandwidth much less than coherent bandwidth of the channel so that there is only flat fading within each subcarrier. The channel fading between different users and different subcarriers is modeled as i.i.d. complex Gaussian distribution with zero mean and unit variance. Specifically, let i denotes the subcarrier index and j denotes the user index, then the received symbol Y_{ij} at the j -th mobile user on the i -th subcarrier is given by

$$Y_{ij} = h_{ij}X_{ij} + Z_{ij} \quad (1)$$

where X_j is the OFDM symbol from the base station to the j -th mobile user, h_{ij} is the complex channel gain of the i -th subcarrier for the j -th mobile which is zero mean complex Gaussian with unit variance and Z_{ij} is the zero mean complex Gaussian noise with variance σ_z^2 . The average transmit power allocated from the base station to user j through subcarrier i is given by $p_{ij} = E[|X_{ij}|^2]$. We define a subcarrier allocation strategy $S_{N_F \times K} = [s_{ij}]$, where $s_{ij} = 1$ when user j is selected for subcarrier i , otherwise $s_{ij} = 0$. We also impose a constraint on the average total transmit power of the base station

$E[\sum_{j=1}^K \sum_{i=1}^{N_F} s_{ij} p_{ij}] \leq P_{TOT}$, where P_{TOT} is the available average power in the base station and also the individual power constraint $p_{ij} \geq 0$ if $s_{ij} = 1$ and $p_{ij} = 0$ if $s_{ij} = 0$.

B. Multi-user Physical Layer Model for OFDMA Systems

In order to decouple the problem to be formulated in this paper from specific implementation of coding and modulation schemes, we consider information theoretical capacity as the abstraction of the multi-user physical layer model. Shannon's capacity can be achieved by random codebook and Gaussian constellation at the base station. The maximum achievable data rate r_{ij} convey from base station to user j through subcarrier i during the fading block is given by the maximum mutual information between X_{ij} and Y_{ij} given the perfect CSIT h_{ij} if $s_{ij} = 1$. Thus the maximum achievable data rate r_{ij} (bits/s/Hz) can be expressed as

$$r_{ij} = s_{ij} \log_2 \left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_z^2} \right). \quad (2)$$

We also represent the transmission rate in matrix form by $R_{N_F \times K} = [r_{ij}]$ with each individual matrix element equal to r_{ij} . In fact, using turbo codes or LDPC codes, the achievable data rate could approach Shannon's capacity to within 0.05dB.

C. Source Model

In this paper, we assumed packets come into each user j 's buffer according to an independent Poisson process with rate λ_j and each packet has fixed size F (bits) for simplicity. Furthermore, we consider the scenario with heterogeneous

mobile user applications. The nature of each user application is characterized by all the K tuples $[\lambda_j, T_j]$, where λ_j is the average arrival rate to user j and T_j is the delay constraint requirement on the average delay $E[w_j]$ by the user j. Users with heavier traffic load will have a higher λ_j and more delay sensitive application will have stringent delay requirements T_j .

D. MAC Layer Model

The MAC layer is responsible for the cross-layer scheduling channel resource at every fading block. At the beginning of every frame, the base station estimates the channel matrix $H_{N_F \times K} = [h_{ij}]$ through the reverse channel from the mobile users as well as collecting the queue state information $Q_K = [q_j]$, where $\chi = (H_{N_F \times K}, Q_K)$ forms the system state which characterizes the system dynamics. Based on the CSIT $H_{N_F \times K}$ obtained, the scheduler determines the subcarrier allocation policy $S_{N_F \times K}$, the power allocation policy $P_{N_F \times K}$ and the corresponding rate allocation policy $R_{N_F \times K}$ for the selected users. The scheduling results are then broadcasted on the downlink common channels to all mobile users before the subsequent downlink packets transmissions at the scheduled rates.

III. PROBLEM FORMULATION

In this section, we formulate an OFDM resource allocation problem based on the system model presented in Section II. Our objective is to achieve maximum average total system throughput $E(\sum_{i=1}^{N_F} \sum_{j=1}^K r_{ij})$ while maintaining OFDMA physical layer constraints on subcarrier selection, average transmission power constraint and average delay $E[w_j]$ constraints. The optimization problem is formally formulated as follows:

$$\begin{aligned} & \max_{S,P} E\left(\sum_{i=1}^{N_F} \sum_{j=1}^K s_{ij} \log_2\left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_z^2}\right)\right) \\ & \text{subject to (C1): } s_{ij} \in \{0, 1\} \\ & \text{(C2): } \sum_{j=1}^K s_{ij} = 1 \\ & \text{(C3): } p_{ij} \geq 0 \\ & \text{(C4): } E\left[\sum_{i=1}^{N_F} \sum_{j=1}^K s_{ij} p_{ij}\right] \leq P_{TOT} \\ & \text{(C5): } E[w_j] \leq T_j \\ & \text{(C6): } \sum_{i=1}^{N_F} s_{ij} \log_2\left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_z^2}\right) \times t_s \times \frac{BW}{N_F} / F \leq q_j \quad \forall \chi, i, j \end{aligned} \quad (3)$$

where the expectation is taken over all system dynamics, t_s is the duration of the encoding frame, and BW is the total Bandwidth of the OFDM system, F (bits) is the size of each packet, q_j is the instantaneous queue length of user j buffer (in terms of number of packet), and the instantaneous system state χ is assumed to be known.

In the optimization problem (3), constraints (C1) and (C2) are used to ensure only one user can occupy a subcarrier i at one time, (C3) is used to ensure transmit power would only take positive value, (C4) is the average total power constraint, (C5) is the average delay constraint where $E[w_j]$ is the system time (including waiting time and service time) of user j and (C6) is used to ensure that the maximum number of scheduled packet for the user j to be less than the queue length q_j of user j's buffer. As we will assume there are always packets to be scheduled out, we will ignore the constraint (C6) in the subsequent discussion.

A. Relationship between scheduled data rate and delay parameters

The optimization problem in (3) is not directly solvable. For the ease of the subsequent analysis, a relationship between the minimum requirement on average scheduled data rate $E[r_j]$ (bits/s) of user j and the user j's characteristic tuples $[\lambda_j, T_j]$ is first formulated in order to transform the constraint (C5) to a traceable form.

Lemma 1: A necessity and sufficient condition for constraint (C5) could be obtained by Pollaczek-Khinchin formula [9]:

$$E[X_j] + \frac{\lambda_j E[X_j^2]}{2(1 - \lambda_j (E[X_j]))} \leq T_j \quad (4)$$

when there is always a subcarrier available for any particular user j (which is a practical assumption in OFDMA system since number of subcarrier N_F is usually much greater than number of user K).

From Lemma 1, and with the recognition that $E[X_j^2] \geq (E[X_j])^2$ and $E[X_j^2] \geq F / E[r_j]$, the constraint (C5) can be simplified to a constraint on scheduled data rate r_j of user j which directly relate to user characteristic tuple $[\lambda_j, T_j]$, and also packet size F as follows:

Corollary 1: A necessity condition for the constraint (C5) is

$$E[r_j] \geq \rho_j(\lambda_j, T_j, F) \quad (5)$$

$$\text{where } \rho_j(\lambda_j, T_j, F) = \frac{(2T_j \lambda_j + 2) + \sqrt{(2T_j \lambda_j + 2)^2 + 8T_j \lambda_j}}{4T_j} F,$$

λ_j is the arrival rate of user j, T_j is the average delay requirement of user j and F is the size of each packet, and r_j is the varying scheduled data rate (bits/s) of user j depending on the CSIT realization $H_{N_F \times K}$.

IV. SCHEDULING STRATEGIES

The optimization problem (3) is a mixed integer and convex optimization problem which is commonly solved by separating the integer programming part (which only involves discrete integer variables $\{s_{ij}\}$) from the convex programming part (which only involves continuous variables $\{p_{ij}\}$) in the following manner. For each possible subcarrier allocation

policy $S_{N_F \times K}$, we compute the optimal power allocation and the corresponding user data rates $(r_{11}, \dots, r_{N_F K})$. Based on the computed rate vector $(r_{11}, \dots, r_{N_F K})$, the instantaneous total

system throughput $\sum_{i=1}^{N_F} \sum_{j=1}^K r_{ij}$ can be evaluated. We can evaluate

the total throughput for all different cases by enumerating all possible subcarrier allocation policy $S_{N_F \times K}$ and the one that gives the largest throughput will be the optimal solution. However, based on the exhaustive search approach for $\{s_{ij}\}$, the total search space is given by N_F^K , which is not feasible for moderate N_F . In this section, we shall illustrate that the search for $\{s_{ij}\}$ can be decoupled between the N_F subcarriers and hence, only with complexity $N_F \times K$ only. By replacing the constraint (C5) with the necessary condition presented in Corollary 1, optimization problem (3) can be reformulated into the following traceable form:

$$\begin{aligned} \max_{S, P} E \left[\sum_{j=1}^K \sum_{i=1}^{N_F} s_{ij} \log_2 \left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_z^2} \right) \right] \\ \text{subject to (C1): } s_{ij} \in \{0, 1\} \\ \text{(C2): } \sum_{j=1}^K s_{ij} = 1 \\ \text{(C3): } p_{ij} \geq 0 \\ \text{(C4): } E \left[\sum_{j=1}^K \sum_{i=1}^{N_F} s_{ij} p_{ij} \right] \leq P_{TOT} \\ \text{(C5): } E \left[\sum_{i=1}^{N_F} s_{ij} \log_2 \left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_z^2} \right) \right] \geq \rho'_j(\lambda_j, T_j, F) \end{aligned} \quad (6)$$

where $\rho'_j(\lambda_j, T_j, F) = \rho_j(\lambda_j, T_j, F) / (BW / N_F)$.

This optimization problem (6) is also a mixed integer and convex optimization problem. In order to make the problem even more traceable, we remove constraint (C1) to let s_{ij} further relaxed to be real number (between 0,1) (a fractional value of s_{ij} refers to time sharing of the subcarrier) and setting $\tilde{p}_{ij} = p_{ij} s_{ij}$, we can reformulate the optimization problem in (6) as a convex maximization problem. Using Lagrange Multiplier techniques, the following Lagrangian could be obtained:

$$\begin{aligned} L = \sum_{j=1}^K \sum_{i=1}^{N_F} s_{ij} \log_2 \left(1 + \frac{\tilde{p}_{ij} |h_{ij}|^2}{s_{ij} \sigma_z^2} \right) - \mu \left(\sum_{j=1}^K \sum_{i=1}^{N_F} \tilde{p}_{ij} - P_{TOT} \right) \\ + \sum_{j=1}^K \gamma_j \left(s_{ij} \log_2 \left(1 + \frac{\tilde{p}_{ij} |h_{ij}|^2}{s_{ij} \sigma_z^2} \right) - \rho'_j \right) + \sum_{i=1}^{N_F} \phi_i \left(\sum_{j=1}^K s_{ij} - 1 \right) \end{aligned} \quad (7)$$

After finding the KKT condition through this Lagrangian, we get the following optimal power and subcarrier allocation stated in Theorem 1.

A. Optimal Subcarrier Allocation and Optimal Power Allocation

Theorem 1: Given the CSIT realization h_{ij} , the optimal subcarrier allocation $S_{opt} = [s_{ij}^*]$ can be decoupled between N_F subcarriers and is given by:

$$\begin{aligned} \text{FOR } i = 1 : N_F \\ j^* = \arg \max_{j \in \{1, 2, \dots, K\}} \left(\log_2 \left(\frac{(1 + \gamma_j) |h_{ij}|^2}{\mu \sigma_z^2} \right) \right)^+ - \mu \left(\frac{(1 + \gamma_j)}{\mu} - \frac{\sigma_z^2}{|h_{ij}|^2} \right)^+ \\ s_{ij}^* = \begin{cases} 1, & j = j^* \\ 0, & \text{otherwise} \end{cases} \\ \text{END} \end{aligned} \quad (8)$$

The corresponding optimal power allocation $P_{opt} = [P_{ij}^*]$ is given by:

$$P_{ij}^* = \begin{cases} \left(\frac{(1 + \gamma_j)}{\mu} - \frac{\sigma_z^2}{|h_{ij}|^2} \right)^+, & \forall s_{ij}^* = 1 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where μ , $(\gamma_1, \dots, \gamma_K)$ are the Lagrange multipliers which satisfy all the KKT conditions and characterize the jointly optimal power and subcarrier allocation policy.

In Theorem 1, the optimal subcarrier allocation strategy $S_{opt} = [s_{ij}^*]$ in (8) is shown to be implementable by a greedy algorithm with linear complexity of K , and the optimal power allocation $P_{opt} = [P_{ij}^*]$ expressed in (9) can be interpreted as a multi-level water-filling strategy. It means that those users with urgent packets have to transmit at higher power level $\frac{(1 + \gamma_j)}{\mu}$,

while non-urgent users (i.e. those users with average delay strictly less than delay deadline) are allocated with the same power level $\frac{1}{\mu}$. It is noted that the corresponding scheduled

data rate would be $r_{ij} = s_{ij}^* \log_2 \left(1 + \frac{P_{ij}^* |h_{ij}|^2}{\sigma_z^2} \right)$.

B. Minimal power required for QoS support

It should be remarked that a feasible power allocation and subcarrier allocation policy that ensure satisfactions of all user requirements may not exist.

Given all K tuples $[\lambda_j, T_j]$, where λ_j is the average arrival rate to user j and T_j is the delay constraint requirement by user j , under the jointly optimal power and subcarrier allocation policy, the minimum power required to support the delay constraints of all users are given by P_{min} which is calculated as follows:

$$\begin{cases} P_{\min} = E \left[\sum_{i=1}^{N_r} \sum_{j=1}^K s_{ij}^* \left(\frac{(1+\gamma_j)}{\mu} - \frac{\sigma_z^2}{|h_{ij}|^2} \right)^+ \right] \\ E \left[\sum_{i=1}^{N_r} s_{ij}^* \log_2 \left(\frac{(1+\gamma_j) |h_{ij}|^2}{\mu \sigma_z^2} \right) \right] = \rho'_j, \forall j \end{cases} \quad (10)$$

where

$$\rho'_j(\lambda_j, T_j, F) = \left(\frac{(2T_j\lambda_j + 2) + \sqrt{(2T_j\lambda_j + 2)^2 + 8T_j\lambda_j}}{4T_j} F \right) / (BW / N_F)$$

and $\mu, (\gamma_1, \dots, \gamma_K)$ are the Lagrange multipliers characterizing the jointly optimal power and subcarrier allocation policy. It should be noted that with this only barely sufficient total transmit power P_{\min} , delay constraints of all users are active; on the other hand, when the transmit power is less than P_{\min} , at least one delay constraint of a user cannot be satisfied for any jointly optimal power and subcarrier allocation policy. Some numerical examples on the minimum required power will be shown in Section V.

V. SIMULATION RESULT

In this section, we present the simulation results using Monte Carlo simulation to illustrate the performance of the proposed cross layer scheduler for OFDM system with mixed traffic having heterogeneous delay requirements.

A. Simulation Model

Mobile users are assumed to be homogeneous in terms of path loss. We also consider arriving packets are of fixed size of 80 bits, and the arrival process to each mobile user's buffer is an independent Poisson process with arrival rate is 0.5 packets per time slot. The total system bandwidth is 80 kHz. Each time slot lasts for 2ms. And we also assumed the channel noise has unit variance.

B. Simulation Results

Consider the simulation model in Section V A, and with number of subcarrier $N_F = 4$ and number of users $K = 5$ in which only user 1 and user 2 has delay constraint specified by T_0 (unspecified users are assumed to be able to tolerate a delay up to 1000 time slots), it can be seen from Figure 2 that with the delay constrained users being more urgent, the minimum power required to support all delay constraints of the user would increase, and at the same time the system throughput would become smaller under same provision of average transmit power.

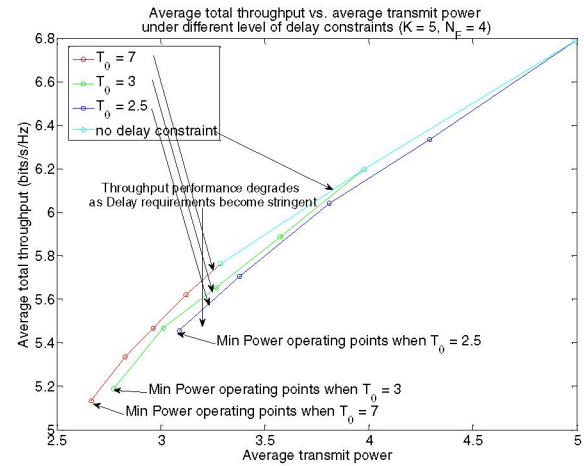


Figure 2. Average total throughput vs average transmit power under different delay constraint of user 1 and 2, $T_0 = 2.5, 3, 7$ time slots

Figure 3 illustrates the scheduling performance with respect to the effect of number of user K with only user 1 and user 2 has delay constraint. It can be seen that the system throughput would increase with K because of the gain provided by improved multiuser diversity, i.e. it is more likely to select user with good channel conditions when more users are available. However the gain of the average total system throughput from multiuser diversity with users being delay constrained is less than that with users having no delay constraint as seen in Figure 3. Besides, the minimum power required to support the same delay constraints of user 1 and user 2 would also increase when number of users K in the system increases. In practice, more users would also induce more overhead in feeding back the channel estimation information to base station.

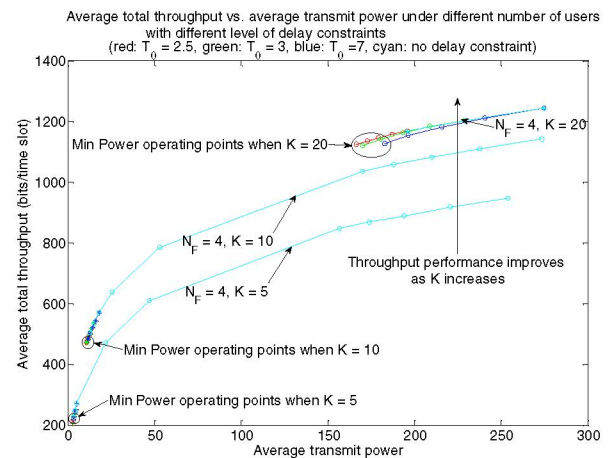


Figure 3. Average total throughput vs average transmit power under different number of user, $K = 5, 10, 20$

With simulation setting changed to $N_F = 5$, and $K = 5$ and maintaining delay constraints of user 1 and user 2 to be $T_0 = 2.5$. Figure 4 illustrates the scheduling performance under different schedulers similar to those results shown in [2]. It can be seen that system throughput is maximally achieved by the jointly optimal Dynamic Subcarrier Allocation (DSA) algorithms and Adaptive Power Allocation algorithm (APA).

In particular, the throughput was increased mostly by using optimal DSA algorithm, while the gain provided by optimal APA is comparatively less significant when optimal DSA algorithm is present. It can also be observed that the minimum power required to support the delay constraints of all users is the least for the case of jointly optimal APA and DSA.

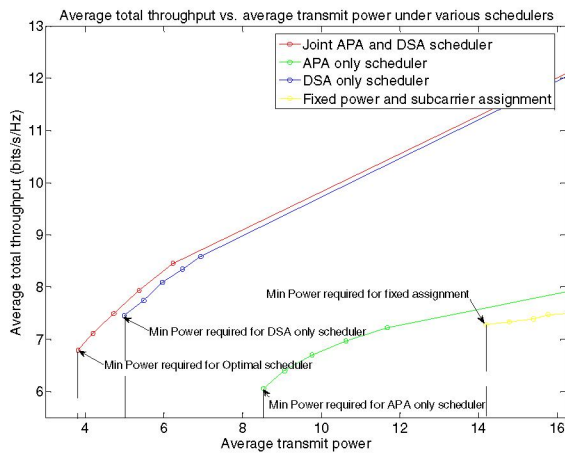


Figure 4. Average total throughput vs average transmit power under different schedulers

Upon the verification of delay constraints satisfaction of the scheduling algorithm, we consider an OFDMA system with 4 subcarriers and 5 users again. But the arrival process to each mobile user's buffer in this case is an independent Poisson process with different arrival rate for different users. We assumed user 1 and user 2 has delay constraint of $T_1 = 4$ and $T_2 = 2$ time slots respectively, and their incoming packet arrival rate is 0.3 packets per time slot. Other users are assumed to have no delay constraint. We tested the effect of varying arrival rate of the background traffic (consisting of unclass users), while keeping the same total transmit power constraint $P_{TOT} = 3$. It is shown in Figure 5 that delay constraints of the delay constrained users (user 1 and user 2) are satisfied, independent of the background traffic loading, with the only scarification of the delay performance of those delay insensitive users.

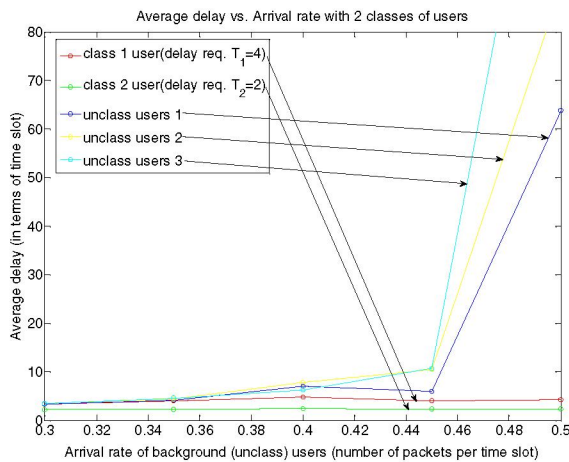


Figure 5. Average delay vs arrival rate of background users' traffic, in which user 1 and user 2 has delay constraint of $T_1 = 4$ and $T_2 = 2$ respectively

VI. CONCLUSION

In this paper, we have presented a cross-layer scheduler for OFDMA systems with heterogeneous delay requirements. The cross layer design problem is formulated as an optimization problem with consideration of the source statistics, queue dynamics as well as the CSIT in the OFDMA systems. The optimal power allocation and subcarrier allocation solutions are obtained based on the optimization framework. The proposed cross layer scheduler offers a nice balance of maximizing throughput and providing QoS (delay) differentiation of the mixed heterogeneous users. We also investigated the minimum power required for finding a feasible scheduler that can satisfy all delay requirements. From the simulation results, it was also shown that substantial throughput gain is achieved by jointly optimal power and subcarrier allocation policy and all users' delay constraints are satisfied.

REFERENCES

- [1] Rohling, H., Gruneid, R, "Performance comparison of different multiple access schemes for the downlink of an OFDM communication system", in *Proc. IEEE. Vehicular Technology Conf.*, pp. 1365 – 1369, 1997.
- [2] G. Song and Y. (G.) Li, "Cross-layer Optimization for OFDM wireless network—Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614-624, Mar. 2005.
- [3] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation", *IEEE J. Sel. Areas Commun.*, Vol. 17, no. 10, Oct. 1999.
- [4] M. Ergen, S. Coleri, and P. Varaiya, "QoS Aware Adaptive Resource Allocation Techniques for Fair Scheduling in OFDMA Based Broadband Wireless Access Systems," *IEEE Trans. On Broadcasting*, Vol. 49, no. 4, Dec 2003.
- [5] S. Kittipiyakul and T. Javidi, "Resource Allocation in OFDMA: How Load-Balancing Maximizes Throughput When Water-Filling Fails", *UW Technical Report*, UWETR-2004-0007.
- [6] E. M. Yeh and A. S. Cohen, "Information Theory, Queuing, and Resource Allocation in Multi-user Fading Communications," in *Proc. of the 2004 Conference on Information Sciences and Systems*, Mar. 2004.
- [7] E. M. Yeh, "Multiaccess and Fading in Communication Networks." Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, 2001.
- [8] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [9] D. Bertsekas and R.G. Gallager, *Data Networks*, 2nd Ed., Prentice-Hall, 1992.
- [10] G. Song, Y. Li, L. Cimini, Jr. and H. Zheng, "Joint Channel-aware and Queue-aware Data Scheduling in Multiple Shared Wireless Channels," in *Proc. IEEE Wireless Comm. and Networking Conf. (WCNC)*, pp. 1922 – 1927, Mar. 2004.
- [11] V. K. N. Lau, M. L. Jiang, S. Liew and O. C. Yue, "Performance Analysis of Downlink Multi Antenna Scheduling for Voice and Data Applications" in *Proc. 42nd Allerton Conf. Comm., Control and Comp.*, Sept. 2004.