Quality of service oriented scheduling algorithms for a cross-layer downlink wireless system

Deepali Arora and Panajotis Agathoklis
Department of Electrical and Computer Engineering
University of Victoria
Victoria, B.C., CANADA, V8W 3P6
Email: {darora, pan}@ece.uvic.ca

***Abstract*— A downlink wireless system is considered where a single base station serves two users simultaneously based on the information available at the PHY and MAC layers. Two new quality of service (QoS) oriented scheduling algorithms are proposed that consider both channel state and direction of arrival information available at the PHY layer and queue length information available at the MAC layer to schedule users. The performance of the proposed algorithms is compared with existing scheduling algorithms, that consider either the channel state information, delay information or the combination of the two, in terms of SINR and exceedance probability of delay and queue length. The results obtained indicate that the proposed algorithms perform better than existing algorithms.**

I. Introduction

The current demand on wireless systems to support both the real time traffic such as audio and video and data traffic such as web browsing, data messaging and file transfer require high quality of service (QoS) guarantees at both the physical (PHY) and the medium access control (MAC) layers. Most studies assess the QoS issues at the PHY and MAC layers separately. The MAC layer designers usually view the PHY layer as black box and focus on improving the QoS by designing efficient queuing and scheduling algorithms. On the other hand, the PHY layer designers focus on ensuring the minimum QoS requirement assessed in terms of system capacity or signal to interference and noise ratio (SINR) by minimizing the effect of fading and interference using diversity or signal processing techniques (e.g., beamforming) and seldom take into account the effect of higher layer requirements. Design of efficient communication systems that satisfy minimal QoS requirements at both PHY and MAC layers require jointly addressing their issues in an integrated framework.

The QoS is usually characterized at the PHY layer by signal to noise ratio (SNR) in case of single user case or signal to interference and noise ratio (SINR) for multiple users and by delay at the MAC layer. For example, scheduling algorithms presented in [1]-[6] consider only the PHY layer QoS issues. These algorithms consider the channel state information (CSI) available at the PHY layer to schedule users at the MAC layer with the aim of maximizing SNR/SINR. The performance of algorithms designed specifically for single user case [5]-[6], however, degrades significantly in the presence of multiple users due to interference. The algorithms that do focus on maximizing SINR [1]-[4] in the presence of multiple users suffer from increased computational complexity associated with finding the best combination of users that may be served simultaneously. Recently [7] proposed a computationally inexpensive scheduling algorithm (discussed later) that serves multiple users simultaneously while minimizing interference they cause on each other, thus maximizing SINR. None of these algorithms consider the effect of delay at the MAC layer. Scheduling algorithms that consider both the channel state information from the PHY layer and delay information from the MAC layer [10]-[13] are also available for both single user [10]-[12] and multiple users [13].

In this paper, we present two new scheduling algorithms that serve multiple users for downlink wireless system that are an extension of the algorithm proposed in [7]. The proposed algorithms considers the CSI, angular location of mobile users around the base station and queue information to schedule the users while trying to minimize interferences from co-channel users (to maximize SINR) and minimize the delay for users in each queue. This paper is organized as follows. The system model and the proposed scheduling algorithms are described in section 2. Comparisons are also performed with existing scheduling algorithms which are also briefly described in the same section. Numerical results are presented in section 3 and finally conclusion in section 4.

II. System model

A multi-user downlink model is considered whose MAC layer contains N queues which receive packets destined for their respective end mobile users. The packet inter-arrival times are assumed to be Poisson distributed with an average arrival rate of λ packets per time step for all queues. The average arrival rate for each queue is thus $\frac{\lambda}{N}$ packets/time step. We assume that only two users may be simultaneously served and thus the maximum packet departure rate (μ) is 2 packets/time step. Of course, when total number of packets at the front of each of the N queues is less than two then only the available packets are serviced. When there are more then two packets waiting for service the unserviced packets experience delay. For the stability of the queues it is essential that $\rho = \frac{\lambda}{\mu}$ should be less than 1, i.e., average packet arrival rate must be less than the average packet departure rate. It is assumed that the base station has both the channel state and direction of arrival (DoA) information available at the PHY

layer based on the incoming pilot signals in an uplink. The CSI can be obtained using several existing techniques such as the one proposed in [8]. The DoA information used to establish the angular location of the mobile users around the base station can be obtained using an efficient DoA estimation technique proposed recently [9]. Each of the arriving packets is associated with a corresponding channel via which it is serviced and the angular location of the user to which the packet is destined is assumed to be randomly distributed between -90^o and $+90^o$ around the base station. The instantaneous SNR of each channel is characterized by its amplitude response h which is assumed to be Rayleigh distributed. For simplicity, the buffer capacity of the queues is assumed to be infinite.

For the two users m and n that are scheduled to be serviced in a given time step the instantaneous SINR (Γ) are given by

$$\Gamma_m = \frac{|h_m|^2}{\sigma_{gn}^2 + |h_n|^2 I} \quad \text{and} \quad \Gamma_n = \frac{|h_n|^2}{\sigma_{gn}^2 + |h_m|^2 I} \tag{1}$$

where h_m and h_n are the amplitude responses of the channels via which the users are simultaneously served, σ_{gn}^2 is the noise and I is the interference scalar that varies between 0 and 1 and determines how much interference user n casts on user m. We assume that $I_{mn} = I_{nm} = I$ which implies that both users cast same relative interference on each other although, of course, the absolute amount of interference depends on the channel amplitude response of each user. Equation (1) for user m may be simplified as follows

$$\Gamma_m = \frac{\frac{|h_m|^2}{\sigma_{gn}^2}}{1 + \frac{|h_n|^2 I}{\sigma_{gn}^2}} = \frac{\gamma_m}{1 + \gamma_n I} \tag{2}$$

where γ_m is the instantaneous SNR for user m, and similarly for user n. Equation (2) shows that when user m and n do not cast interference on each other ($I = 0$) then $\Gamma = \gamma$.

Following [7], the interference scalar is estimated on the basis of a beam that is used for beamforming when the two users are served simultaneously. The beam is approximated as

$$I = e^{-|\theta|/c} \tag{3}$$

where θ is the angular separation and c is a parameter that is surrogate for, and inversely proportional to, the number of antenna elements. The higher (lower) the value of c, the wider (narrower) the beam, and more (less) energy is dissipated in other directions. The beam pattern allows to determine leakage of power in the direction of simultaneously served co-channel users based on the angular separation between the users.

The objective of scheduling algorithms is to choose users m and n using criterion such that SINR is maximized and/or delay is minimized, etc. SINR is maximized by choosing users with high values of h and/or by choosing users that are far apart (low I) such that they cause minimal interference on each other. Delay is minimized by giving preference to packets that arrived earlier and/or by preferentially serving longest queues. A scheduler at the MAC layer may use one or all the four available pieces of information to make a scheduling decision. These four pieces of information are CSI and DoA at the PHY layer, and the current delay associated with each packet and queue length information at the MAC layer. The scheduling algorithms analyzed here are briefly explained below and use one or all four pieces of information. Depending on how packets arrive in the N queues the number of packets waiting for service (K) can vary from 0 to N.

A. Scheduler proposed in [7]

The scheduler proposed in [7] schedules packets for two users for simultaneous service by selecting the first user m within a group of K available packets waiting for service on the basis of their corresponding instantaneous signal to noise (γ) ratio

$$m = arg \max_{i=1,2,\ldots K} [\gamma_i] \tag{4}$$

The second user n is selected in such a way such that it is farthest away from user m in an angular sense.

$$n = arg \max_{j=1,2,\ldots K; j \neq m} [\theta_{jm}] \tag{5}$$

where θ_{mj} is the angular separation between mobile users m and j. In this algorithm the selection of user m attempts to maximize SNR while selection of user n attempts to reduce interference thus attempting to maximize SINR for the combined users. Although this scheduling algorithm yields improvement over traditional Greedy and Round-Robin scheduling algorithms [7] it does not take into account the delay or queue length. The algorithms B and C discussed next are modifications of algorithm A and explicitly take queue length into account for scheduling users.

B. SNR preferred modification of scheduling algorithm A

In this modification of scheduling algorithm A, the first user m is chosen in a similar way as in algorithm A, i.e. the user with the highest SNR γ (equation 4). The second user is selected from the queue with largest number of packets waiting and attempts to minimize the delay.

$$p = arg \max_{i=1,2,\ldots K; i \neq m} [q_i] \tag{6}$$

where q_i is the queue length of the i^{th} queue. In case there is more than one queue with same number of packets waiting to be serviced, then the queue containing the user farthest away from user m is selected for service.

$$n = \begin{cases} p & \text{if } L(p) = 1 \\ arg \quad \max_{j=1,2,..,L(p)} [\theta_{mp(j)}] & \text{if } L(p) > 1 \end{cases} \tag{7}$$

where $L(p)$ is the length of p, $L(p) = 1$ when p is a scalar and $L(p) > 1$ when p is a vector. In this algorithm the selection of user m aims at maximizing SNR, while the selection of user n serves the dual purpose of minimizing both delay and/or interference.

C. Queue preferred modification of scheduling algorithm A

In this modification of scheduling algorithm A the first user is chosen from the queue containing largest number of packets

$$p = arg \max_{i=1,2,...K}[q_i] \tag{8}$$

In case there is more than one queue containing same number of packets then preference is given to the queue containing packet associated with the highest SNR value. Thus,

$$m = \begin{cases} p & \text{if } L(p) = 1 \\ arg \ \max_{j=1,2,..,L(p)}[\gamma_{p(j)}] & \text{if } L(p) > 1 \end{cases} \tag{9}$$

The second user n is selected such that it is farthest away from user m in an angular sense as in scheduling algorithm A (equation 5).

Both algorithms B and C are extensions of algorithm A and aim at minimizing the delay and maximizing the SINR respectively in two different ways. Algorithm B requires that the SNR condition (equation 4) of algorithm A is used for scheduling the first user while algorithm C requires that the minimum interference condition (equation 5) of algorithm A is used for scheduling the second user. The performance of the proposed scheduling algorithms B and C will be compared with existing scheduling algorithms that are briefly discussed next.

D. Greedy scheduling algorithm

The greedy algorithm [2] prioritizes and pairs users according to channel conditions characterized by their instantaneous signal to noise ratio (γ_i). Users m and n with best channel conditions (highest values of γ_i) out of K users waiting to be serviced are scheduled for transmission in each time step as in the first step of algorithm A (equation 4). This algorithm attempts to maximize signal to noise ratio but does not take into account the delay or angular location of the mobile users around the base station.

E. Longest queue algorithm

The longest queue algorithm [14] always serves the users from the queues containing the largest number of packets. Users m and n with longest queues are scheduled for transmission in each time slot as follows

$$m = arg \max_{i=1,2,...K}[q_i] \tag{10}$$

$$n = arg \max_{i=1,2,...K; i \neq m}[q_i] \tag{11}$$

This scheme attempts to minimize delay but does not take into account the channel conditions of the users or their angular location around the base station.

F. Modified largest weighted delay first M-LWDF

This scheduling algorithm, based on [10], finds a delay weighted metric

$$\eta_i = d_i \gamma_i \tag{12}$$

where d_i is the head of line packet delay for queue i and γ_i is the instantaneous SNR. Users m and n with highest values of η_i are selected. This metric ensures that users with either higher SNR, delay or both are preferentially served.

G. Scheduling algorithm proposed in [11]

In the scheduling algorithm proposed in [11], the scheduling decision is based on the following metric

$$\xi_i = \gamma_i + \alpha d_i \tag{13}$$

where γ_i is the SNR of the i^{th} user, d_i its current delay and α is an arbitrary constant that determines the behavior of the scheduler from an essentially greedy scheduler (for small values of α) to a delay based scheduler (for large values of α). This scheduler takes either one or both the performance measures into consideration while making a scheduling decision based on the value of α.

III. Numerical results

Numerical simulations are carried out to analyze the effect of different scheduling algorithms on the performance of the downlink system composed of single base station capable of serving two users simultaneously using beamforming. It is assumed that the users are using single antenna systems with no diversity. The performance of the scheduling algorithms is assessed in terms of average SINR, a delay metric d_{95} and queue length metric q_{95}. d_{95} represents the delay in number of time steps that has an exceedance probability of 5%, that is

$$\text{prob}\,[d > d_{95}] \approx 0.05 \tag{14}$$

and similarly for q_{95} which represents the queue length whose exceedance probability is 5%. Scheduling algorithms that yield high values of average SINR and low values of d_{95} and q_{95} are considered superior. For the simulation presented here the value of ρ is taken to be 0.95, c is assumed equal to 90 (similar performance to 4 antenna elements), and σ^2_{gn} = 1.

A. Performance in terms of delay and queue length

Table I compares the performance of different scheduling algorithms in terms of d_{95} and q_{95} metrics. The number of queues N here is equal to 6. A value of α = 100 is used for algorithm G that essentially makes it a delay-based scheduler. As expected, algorithm G which is essentially a delay based scheduler yields best performance in terms of delay and queue length. The longest queue (Algo. E) and the two modifications of algorithm A (Algos. B and C) yield similar results and performs better than algorithms D and F. The scheduling algorithm proposed by [7] (Algo. A) which does not take delay or queue length into account yields similar performance as the M-LWDF [10] algorithm (Algo. F) and finally the greedy algorithm (Algo. D) which yields performance even lower than the algorithms A and F respectively.

Table I shows that scheduling algorithms that do take delay or a queue length into consideration while making the scheduling decision yield better performance compared to algorithms that do not take these criteria into consideration.

TABLE I

d_{95} AND q_{95} METRICS FOR DIFFERENT SCHEDULING ALGORITHMS. PARAMETER VALUES USED ARE ρ = 0.95, C= 90 AND σ^2_{gn} = 1

Algorithms	A	B	C	D	E	F	G
d_{95}	11.7	9.7	9.5	13.7	9.7	11.2	8.7
q_{95}	3.8	2.9	2.8	4.2	2.6	3.7	2.9

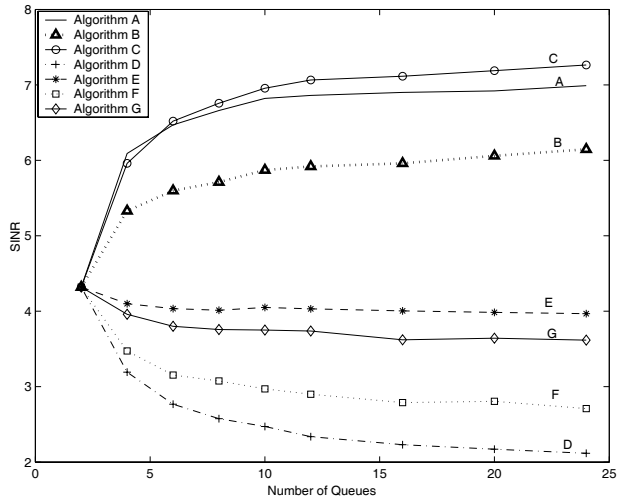

Fig. 1. Effect of number of queues on average SINR (dB) for different scheduling algorithms. Parameter values are ρ = 0.95, C= 90 and σ^2_{gn} = 1

B. Performance in terms of SINR

Figure 1 compares the performance of the scheduling algorithms in terms of SINR. Simulations are performed for number of queues N equal to 2 and higher upto 24. All algorithms yield same SINR for N = 2. This is because when number of queues equals two the scheduling algorithms do not have a choice in selecting the users and the available users are served. As the queue length increases, the choice in regard to which users may be served increases and the differences between the algorithms become prominent. Algorithm A and its modifications B and C perform better than other algorithms because they explicitly take into account the angular separation between mobile users while making a scheduling decision. The greedy algorithm yields low SINR because it chooses the users with the highest SNR regardless of angular separation between them and this leads to high interference between simultaneously served users. The delay (F and G) and queue length (E) based algorithms perform better than the greedy algorithm but yield poor performance relative to algorithms A, B and C because they do not focus on interference reduction. The performance of scheduling algorithms in terms of average SINR tends to remain constant for higher values of N.

In terms of overall performance algorithm C yields the highest SINR and its delay and queue length metrics are second only to the delay based scheduler (Algo. G). The other algorithms do not perform that well in terms of either SINR or delay and queue length.

IV. SUMMARY AND CONCLUSIONS

A downlink wireless system is considered where a single base station schedules two users simultaneously based on the information available at the PHY and MAC layers. The QoS is assessed in terms of SINR at the PHY layer and delay/queue length at the MAC layer. Two new quality of service oriented scheduling algorithms are proposed that consider both channel state information and direction of arrival (DoA) information of the users available at the PHY layer and delay information available at the MAC layer in making the scheduling decision. The performance of the proposed algorithms is also compared with the existing scheduling algorithms. The results obtained show that the proposed algorithm C performs better than the existing algorithms in terms of SINR and is the second best in terms of delay and queue length. The improvement in SINR in the proposed algorithm C is achieved by taking an additional PHY layer parameter, the DoA information, into consideration while scheduling the users.

ACKNOWLEDGMENT

The support of the Natural Sciences and Engineering Research Council of Canada is acknowledged.

REFERENCES

[1] O.S. Shin and K.B. Lee, "Antenna assisted round-robin scheduling for MIMO cellular systems," *IEEE Commun. Letters*, Vol. 7, No. 3, pp. 109-11, Mar. 2003.

[2] M. Airy, S. Shakkottai and R.W. Heath, "Spatially greedy scheduling in multi user MIMO wireless systems," *37th Asilomar Conference on Sig. Sys. and Comp.*, Vol. 1, pp. 982-986, Nov. 2003.

[3] D. Aktas and H.E. Gamal, "Multiuser scheduling for MIMO wireless systems," *IEEE 58th VTC 2003-Fall*, Vol. 5, pp. 1743 - 1747, Oct. 2003.

[4] L. Dong, T. Li and Y-Fang. Huang, "Opportunistic transmission scheduling for multiuser MIMO systems," *ICASSP 2003*, pp. 65-68, May 2003.

[5] T. Park, O.S. Shin and K. Bok Lee, "Proportional fair scheduling for wireless communication with multiple transmit and receive antennas," *IEEE Commun. Letters*, Vol. 7, No. 3, pp. 109-111, Mar. 2003.

[6] L.T. Berger, T.E. Kolding, J. Ramiro-Moreno, P. Ameigeiras, L. Schumacher, P.E. Mogensen, "Interaction of transmit diversity and proportional fair scheduling," *VTC 2003-Spring*, Vol. 4, pp. 2423-2427, Apr. 2003.

[7] D. Arora and P. Agathoklis, "Multiuser scheduling for downlink in multi-antenna wireless systems," *presented in IEEE Symp. on Circuits and Sys.*, held in Kobe Japan, Vol. 2, May 2005, pp. 1718-21

[8] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. on Veh. Tech.*, Vol. 40, pp. 686-693, 1991.

[9] N. Wang and P. Agathoklis, "A new high resolution and capacity based DoA estimation technique based on sub-array beamforming," *38th Asilomar conference on Sig., Sys., and Commun.*, Vol. 2, Nov. 2004, pp. 2345-49.

[10] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar and P. Whiting, "Providing Quality of Service over a shared wireless link" *IEEE Communications Mag.*", Feb. 2001, pp. 150-154.

[11] R. Srinivasan and J.S. Baras, " Understanding the trade offs between multiuser diversity gain and delay - an analytical approach" *VTC spring-2004*, Vol. 5, pp. 2543-47, 2004.

[12] P. Liu, R. Berry and M.L. Honig, " Delay sensitive packet scheduling in wireless networks" *IEEE WCNC.*, Vol. 3, Mar. 2003, pp. 1627-32.

[13] B. Zerlin and J.A. Nossek, "Cross-layer QoS management in scheduled multi-user systems" *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications.*, June 2005, pp. 520-24.

[14] F. Gebali, *Computer Communication Networks: Analysis and Design*, Northstar Digital Design, 3^{rd} ed, 2005.