Throughput and Fairness maximization in Wireless Networks

Alireza Bayesteh, Mehdi Ansari Sadrabadi, and Amir K. Khandani
Coding and Signal Transmission (CST) Lab. (www.cst.uwaterloo.ca)
Dept. of Elec. and Comp. Eng., University of Waterloo
Waterloo, ON, Canada, N2L 3G1
e-mail: {alireza, mehdi, khandani}@cst.uwaterloo.ca

***Abstract*— In this paper, a single-antenna broadcast channel with large (K) number of users is considered. It is assumed that all users have a hard delay constraint D. We propose a scheduling algorithm for maximizing the throughput of the system, while satisfying the delay constraint for all users. It is proved that by using the proposed algorithm, it is possible to achieve the maximum throughput and minimum delay in the network, simultaneously, in the asymptotic case of $K \to \infty$. We introduce a new notion of fairness in the network, called "*Minimum Average Throughput*", and prove that the proposed algorithm maximizes the *minimum average throughput* in a broadcast channel. Finally, the proposed algorithm is generalized for MIMO Broadcast Channels (MIMO-BC).**

I. Introduction

With the development of personal communication services, one of the major concerns in supporting data applications is providing quality of service (QoS) for all subscribers. In most real-time applications, high data rates and small transmission delays are desired. Most data-scheduling schemes proposed for current systems have concentrated on the system throughput by exploiting multiuser diversity [1]–[5]. In cellular networks, by applying multiuser diversity, the time-varying nature of the fading channel is exploited to increase the spectral efficiency of the system. It is shown that transmitting to the user with the highest signal to noise ratio (SNR) at a time provides the system with maximum sum-rate throughput [6]. The opportunistic transmission is proposed in Qualcomm's High Data Rate (HDR) system [2].

Although applying multiuser diversity through the scheme in [6] achieves the maximum system throughput, QoS demands, including fairness and delay constraints, provoke designing more appropriate scheduling schemes. The schemes that consider delay constraints have been studied extensively in [1], [7]–[13]. In [10], the authors propose an algorithm which keeps a balance between the capacity maximization, delay, and outage probability in a multiple access fading channel. The tradeoff between the average delay and the average transmit power in fading environments is analyzed in [7]. In [8], [11], authors propose scheduling metrics that combine multiuser diversity gain with the delay constraints. In [9], the scheduling scheme is designed based on maximizing the effective capacity [14] which is characterized by data rate, delay bound, and delay-bound violation probability triplet. The throughput-delay tradeoff of the multicast channel is analyzed for different schemes in a single cell system [13]. In [12], the delay is defined as the minimum number of channel uses that guarantees all the users successfully receive m packets. Reference [12] studies the statistical properties of the underlaying delay function. However, the delay constraint is assumed to be *soft*, meaning that this scheme aims to minimize the total *average* network delay and there is not any delay constraints for the individual users.

In this work, we consider a *hard* delay constraint D for each user. We define a dropping event as the event that there exists a user who does not meet the desired delay constraint. This delay constraint can be enforced by the application, or the physical limitations (e.g., buffer size). In this paper, we propose a scheduling scheme for maximizing the throughput of the system, while satisfying the delay constraint for all users. It is proved that by using the proposed algorithm, it is possible to achieve the maximum throughput and minimum delay in the network, simultaneously, in the asymptotic case of $K \to \infty$. We introduce a new notion of fairness in the network, called "Minimum Average Throughput", and prove that the proposed algorithm is capable of maximizing the *Minimum Average Throughput* in a broadcast channel.

The rest of the paper is organized as follows. In section II, the system model is introduced and the proposed algorithm is described. Section III is devoted to the asymptotic analysis of the proposed algorithm. Section IV described the generalization of the proposed algorithm for MIMO-BC, and finally, section V concludes the paper.

Throughout this paper, for any functions $f(n)$ and $g(n)$, $f(n) = o(g(n))$ is equivalent to $\lim_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| = 0$, $f(n) = O(g(n))$ is equivalent to $\lim_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| < \infty$, and $f(n) = \omega(g(n))$ is equivalent to $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty$.

II. System Model and Proposed Algorithm

In this paper, a downlink environment in which a single-antenna Base Station (BS) communicates with a large number (K) single-antenna users, is considered. We assume a homogeneous network, where the channel between each user and the BS is modelled as a zero-mean complex Gaussian

[1]This work is financially supported by Nortel and by matching funds from the federal government of Canada (NSERC) and province of Ontario (OCE).

random variable (Rayleigh fading). The received signal at the kth terminal can be written as

$$y_k = h_k x + n_k, \quad (1)$$

where x denotes the transmitted signal by the BS, which is assumed to be Gaussian with the power constraint P, i.e., $\mathbb{E}\{|x|^2\} \leq P$, $h_k \sim \mathcal{CN}(0,1)$ denotes the channel coefficient between the BS and the kth terminal, and $n_k \sim \mathcal{CN}(0,1)$ is AWGN. The BS serves one user during each frame. The size of the frames is assumed to be fixed. The channel coefficients are assumed to be constant for the duration of a frame (block fading model). The frame itself is assumed to be long enough to allow communication at rates close to the capacity.

It is assumed that the users have delay constraint D. In other words, the delay between two consequetive received packets should not be greater than the duration of D frames. Otherwise, the transmitted packet will be dropped. The *network dropping event*, denoted by $\mathscr{B}$, is defined as the event that dropping occurs for any user in the network. We define a parameter ν for each user, which denotes the *expiry countdown* of that user's packet, i.e., the remaining time to the expiration of the packet. ν is expressed in terms of an integer multiple of the frame length. At the end of each frame, the *expiry countdown* of each user is decremented by one, except the user who is served during that frame. For this user, the *expiry countdown* is set to D at the start of the next frame. Therefore, for all users $\nu \leq D$.

The proposed algorithm is described as follows:
Algorithm 1:

1) The BS chooses a threshold Θ, and sends it to all users.
2) Let us define

$$\mathcal{S} \triangleq \{k| \ |h_k|^2 \geq \Theta\}. \quad (2)$$

All users in $\mathcal{S}$ send a confirmation message to the BS.
3) Among the users in $\mathcal{S}$, the BS serves the one with the minimum ν (*expiry countdown*).

In the proposed algorithm, the threshold Θ is set to trade-off the average throughput vs. the fairness in the system. If Θ is chosen very large, then the scheduling tends to maximize the throughput. If Θ is chosen very small, the algorithm tends to the Round-Robin scheduling.

III. Asymptotic Analysis

In this section, we analyze the network dropping probability, denoted as $\Pr\{\mathscr{B}\}$, in terms of the number of users K, and the delay constraint D, for the proposed scheduling. We consider the asymptotic case of $K \to \infty$ and derive the condition for D such that $\Pr\{\mathscr{B}\} \to 0$. To this end, the probability mass function (pmf) of ν, denoted as $f_\nu(\nu)$, is characterized in terms of D, K, and Θ. First, we consider two special cases of the proposed algorithm:

A. Special Case I; $\Theta = 0$:

In this case, the user with the minimum *expiry countdown* is served. In other words, the quality of channel does not play any role in the scheduling. The set $\mathcal{S}$ which is defined in (2) is simply the set of all users.

Theorem 1 *For $\Theta = 0$, $f_\nu(\nu)$ can be obtained as follows:*

$$f_\nu(\nu) = \begin{cases} \frac{1}{K} & D-K+1 \leq \nu \leq D \\ 0 & \nu \leq D-K \end{cases}. \quad (3)$$

Proof - Refer to [15].

The above theorem implies that the pmf of ν is a step function which is only non-zero in the interval $[D-K+1, D]$. Since the probability of dropping for any given user can be expressed as $\sum_{l=-\infty}^{0} f_\nu(l)$, it follows from the above equation that for $D \geq K$, the dropping probability for each user is zero and as a result, the network dropping probability is zero.

This scheduling is exactly the Round-Robin scheduling, when the users are served based on a pre-determined order. One can observe that this scheduling is the most fair scheduling, as all the users have the same opportunity for being served, regardless of their channel quality. However, due to disregarding the effect of channel quality in the scheduling, the achievable throughput is not good.

B. Special case II; $\Theta = \max_k |h_k|^2$:

In this scheduling, $|\mathcal{S}| = 1$. In other words, the user with the best channel quality is served during each frame. This results in the conventional scheduling to exploit the multiuser diversity and achieves the maximum sum-rate throughput in the system [16].

Theorem 2 *For the Special Case II, $f_\nu(\nu)$ is equal to*

$$f_\nu(\nu) = \frac{1}{K}\left(1 - \frac{1}{K}\right)^{D-\nu} u(D-\nu), \quad (4)$$

where $u(.)$ denotes the unit step function.

Proof - Refer to [15].

Theorem 3 *For $K \to \infty$, the necessary and sufficient condition to have $\Pr\{\mathscr{B}\} \to 0$ for the special case II is*

$$D \sim K \log K + \omega(K). \quad (5)$$

Proof - Refer to [15].

The above theorem states that the minimum delay constraint in order to have small dropping probability in the network must scale as fast as $K \log K$. Compared to the Round-Robin scheduling (Case I), we have a factor of $\log K$ increase in the delay, which is due to ingnoring ν in the scheduling.

C. Proposed Algorithm; The general case:

In the previous sections, we have studied our proposed scheduling algorithm in two extreme cases, where one extreme focuses on achieving the maximum fairness, and the other extreme on achieving the maximum sum-rate throughput. In general, it is possible to have a trade-off between the fairness and throughput, by adjusting the threshold value. Now, the question is, whether or not, it is possible to simultaneously achieve the maximum throughput and the maximum fairness of the system. The following theorem shows this is indeed possible in the asymptotic case of $K \to \infty$:

Theorem 4 *Consider the proposed algorithm in the asymptotic case of $K \to \infty$. Then, for the values of Θ satisfying*

$$\log K - 2\log\log K < \Theta < \log K - 1.5\log\log K, \qquad (6)$$

one can simultaneously achieve:
I- Maximum Throughput:

$$\lim_{K\to\infty} \mathcal{C}_{sum} - \mathcal{R} = 0, \qquad (7)$$

in which $\mathcal{C}_{sum}$ denotes the sum-rate capacity of the broadcast channel and $\mathcal{R}$ denotes the achievable sum-rate of the proposed algorithm,
II- Minimum Network Delay:

$$\lim_{K\to\infty} \frac{D}{K} = 1, \qquad (8)$$

and III- Zero Network Dropping Probability:

$$Pr\{\mathscr{B}\} \to 0. \qquad (9)$$

Proof - The steps of the proof are as follows: in Lemma 1, we study the behavior of $f_\nu(\nu)$ and derive the difference equation satisfied by $f_\nu(\nu)$. In Lemma 2, we derive an explicit solution for this difference equation. Based on this solution, in Lemma 3, we give a sufficient condition, such that $\lim_{K\to\infty} \frac{D}{K} \to 1$ and $\Pr\{\mathscr{B}\} \to 0$ are satisfied simultaneously. Finally, the theorem is proved by deriving a lower-bound on the achievable sum-rate, based on the threshold level given in (6). For the proof of the lemmas, the reader is referred to [15].

Lemma 1 *Defining $D_0 = D - \sqrt{K} n_0(n_0 - 1)$, where $n_0 = 3(\log K)^2$, for $D_0 \le \nu \le D$, $f_\nu(\nu) \sim \frac{1}{K}[1 - o(1/K)]$, and for $\nu < D_0$, $f_\nu(\nu)$ satisfies the following difference equation:*

$$f_\nu(\nu) - f_\nu(\nu - 1) \;=\; \eta f_\nu(\nu)\left[1 - pF_\nu(\nu)\right]^{K-1}, \qquad (10)$$

where $p = e^{-\Theta}$, $\eta \triangleq \frac{p}{1-p}$, and $F_\nu(\nu)$ denotes the CDF of ν.

Lemma 2 *The solution to the difference equation (10), in the asymptotic case of $K \to \infty$, is equal to*

$$f_\nu(\nu) = \frac{\frac{\eta}{(K-1)p} e^{(K-1)p} e^{\eta(\nu - D_0)}}{1 + e^{(K-1)p} e^{\eta(\nu - D_0)}} \quad \nu < D_0. \qquad (11)$$

Lemma 3 *Setting $D_0 = \frac{p}{\eta}[K-1] + \frac{\log K}{\eta}$ yields $Pr\{\mathscr{B}\} \to 0$, while satisfying $\lim_{K\to\infty}\frac{D}{K} = 1$.*

Conditioned on $\mathscr{C}$, where $\mathscr{C}$ denotes the event that $|\mathcal{S}| > 0$, the gain of the selected user is above Θ. Hence, the achievable sum-rate can be lower-bounded by $\log(1 + P\Theta)\Pr\{\mathscr{C}\}$. Noting that $\mathcal{C}_{sum} \sim \log(1 + P\log K + O(\log\log K))$ [17], $\Theta > \log K - 2\log\log K$, and $\Pr\{\mathscr{C}\} \gtrsim 1 - e^{-(\log K)^{1.5}}$ [15], we have $\mathcal{C}_{sum} - \mathcal{R} \gtrsim O\left(\frac{\log\log K}{\log K}\right)$, which incurs $\lim_{K\to\infty} \mathcal{C}_{sum} - \mathcal{R} = 0$. Combining this with Lemma 3 completes the proof of Theorem 4.

■

Since $D = K$ is the smallest delay constraint in order not to have any dropping in the network, the above theorem simply implies that the proposed algorithm is capable of achieving the maximum throughput, while guaranteeing the minimum delay for all users, asymptotically. In other words, this scheme is capable of achieving the maximum throughput and fairness, simultaneously.

Corollary 1 *Consider a Broadcast system, where all the users have the buffer size of one and the arrival rate of the packets for the kth user is r_k. Let us define the "average throughput" of user k as*

$$\mathcal{T}_k \triangleq r_k \mathcal{R}_k, \qquad (12)$$

where $\mathcal{R}_k$ denotes the amount of information per channel use of each packet for this user. Then, for any scheduling scheme, any $\mathcal{R}_k$ supported by the channel (decoding error approaches zero), and for any chosen r_k, the necessary condition for $Pr\{\mathscr{B}\} \to 0$ is having

$$\mathcal{T}_{\min} \triangleq \min_k \mathcal{T}_k \quad \lesssim \quad \frac{\log\log K}{K}, \qquad (13)$$

which is achievable by the proposed algorithm.

Proof - Refer to [15].

In the above corollary, the *minimum average throughput*, denoted by $\mathcal{T}_{\min}$, is defined as the measure of performance. This measure is suitable for the real-time applications like transmitting sound and video, where the packets have certain bit-rates and certain arrival rates. Note that in the above corollary we have assumed that the users have the buffer size of one, which is a very restrictive assumption in the wireless networks. For the realistic scenarios, this constraint is more relaxed. However, since we have shown the optimality of our proposed scheduling for this assumption, it easily follows that this optimality holds for the realistic assumptions as well.

Computing $\mathcal{T}_{\min}$ for the two special cases of the proposed algorithm, i.e., maximum-throughput scheduling ($\mathcal{T}_{\min}^{MT}$) and Round-Robin scheduling ($\mathcal{T}_{\min}^{RR}$), yields,

$$\begin{aligned} \mathcal{T}_{\min}^{MT} &\sim \frac{\log\log K}{K\log K}, \\ \mathcal{T}_{\min}^{RR} &\sim \frac{1}{K}. \end{aligned} \qquad (14)$$

Therefore, the proposed algorithm outperforms these conventional scheduling algorithms by a factor of $\log K$ and $\log\log K$, respectively.

IV. Extension to the MIMO-BC

In this section, we assume that the BS has M antennas, while the receivers have single antennas. The main difference between this case and the previous case is that for Single-Input Single-Output Broadcast Channel (SISO-BC), serving the best user during each frame (TDMA) is optimal in terms of achieving the maximum throughput of the system [16], while in the MIMO-BC this is not the case. Therefore, we must apply some modifications in our proposed algorithm, in order to make it suitable for MIMO-BC.

The channel model for the kth user is assumed to be

$$y_k = \mathbf{h}_k \mathbf{x} + n_k, \tag{15}$$

where $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the transmitted signal with the power constraint $\mathbb{E}\{\mathbf{x}^H\mathbf{x}\} \leq P$, $\mathbf{h}_k \in \mathbb{C}^{1\times M} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ is the channel vector, $n_k \sim \mathcal{CN}(0,1)$ is AWGN, and y_k is the received signal by the kth user.
The proposed algorithm is as follows:
Algorithm 2:

1) Set the threshold Θ.
2) The BS selects M orthogonal unit vectors, denotes by $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_M$, randomly, and sends it to all users.
3) Among each of the following sets:

$$\mathcal{S}_m = \{k| \quad \text{SINR}_k^{(m)} > \Upsilon\}, \quad m = 1, \cdots, M, \tag{16}$$

the BS serves the user with the minimum *expiry countdown*. In the above equation, $\text{SINR}_k^{(m)} \triangleq \frac{\rho|\mathbf{h}_k\mathbf{\Phi}_m^H|^2}{1+\sum_{j\neq m}\rho|\mathbf{h}_k\mathbf{\Phi}_j^H|^2}$, in which $\rho \triangleq \frac{P}{M}$, is the received Signal to Interference plus Noise Ratio (SINR) on the mth transmitted beam, by the kth user.

As can be observed, this algorithm is a variant of Random-Beam-Forming scheme proposed in [17], where the *expiry countdown* is considered in the scheduling.

Theorem 5 *Using Algorithm 2, for the values of* Υ *satisfying*

$$\begin{aligned}\frac{P}{M}\left[\log K - (M+0.5)\log\log K\right] &< \Upsilon \\ &< \frac{P}{M}\left[\log K - (M+1)\log\log K\right], \end{aligned} \tag{17}$$

we have $\lim_{K\to\infty} \mathcal{C}_{sum} - \mathcal{R} = 0$*, and* $\lim_{K\to\infty} \frac{MD}{K} = 1$*, while satisfying* $Pr\{\mathscr{B}\} \to 0$.

Proof - Refer to [15].

Noting that $\lceil \frac{K}{M} \rceil$ is the minimum achievable delay in MIMO-BC (using Round-Robin, assuming that M users are served during each frame), it follows that the proposed algorithm achieves the maximum sum-rate and minimum network delay at the same time.

Defining the *minimum average throughput* as in (13), it is straightforward to show that for the proposed algorithm,

$$\mathcal{T}_{\min} \sim \frac{M\log\log K}{K}, \tag{18}$$

which is asymptotically the maximum achievable value in MIMO-BC

V. Conclusion

In this paper, a single-antenna broadcast channel with large (K) number of users is considered. It has been assumed that all users have hard delay constraint D. We have proposed an scheduling algorithm for maximizing the throughput of the system, while satisfying the delay constraint for all users. By characterizing the network dropping probability, in terms of K, D, and the threshold value in the algorithm, it has been shown that by using the proposed algorithm, it is possible to achieve the maximum throughput and minimum delay in the network, simultaneously, in the asymptotic case of $K \to \infty$. Moreover, We have introduced a new notion of fairness in the network, called "*Minimum Average Throughput*", and proved that the proposed algorithm maximizes the maximum *minimum average throughput* in a broadcast channel. Finally, the proposed algorithm is generalized for MIMO-BC, and shown to be optimum in the sense of achieving the maximum throughput and minimum delay in the network, simultaneously.

References

[1] P. Viswanath, D.N.C. Tse, R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[2] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, pp. 70–77, July 2000.

[3] A. Jalali, R. Padovani, and R. Pankaj , "Data throughput of CDMA/HDR: A high efficiency, high data rate personal wireless system," in *Proc. IEEE Vehicular Tech. Conference*, vol. 3, pp. 1854–1858, May 2000.

[4] X. Liu, E. K. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE JSAC*, vol. 19, pp. 2053–2064, Oct. 2001.

[5] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Networking*, vol. 13, pp. 636 – 647, June 2005.

[6] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," *IEEE ICC'95*, vol. 1, pp. 331 – 335, June 1995.

[7] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1135 – 1149, May 2002.

[8] R. Srinivasan and J. S. Baras, "Understanding the trade-off between multiuser diversity gain and delay - an analytical approach," *IEEE Vehicular Technology Conference*, vol. 5, pp. 2543 – 2547, May 2004.

[9] D. Wu and R. Negi , "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *IEEE Trans. Vehicular Techn.*, vol. 54, pp. 1198 – 1206, May 2005.

[10] I. Bettesh and S. Shamai, "A low delay algorithm for the multiple access channel with Rayleigh fading," *in Proc. IEEE Personal, Indoor and Mobile Radio Commun.*, vol. 3, pp. 1367 – 1372, Sept. 1998.

[11] M. Andrews, K. Kumaran, K. Ramanan, A.L. Stoylar, R. Vijayakumar and P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, pp. 150 – 154, Feb. 2001.

[12] Masoud sharif and Babak Hassibi , "A delay analysis for opportunistic transmission in fading broadcast channels," in *Proc. IEEE, INFOCOM*, vol. 4, pp. 2720 – 2730, March 2005.

[13] P.K. Gopala and H. El Gamal, "On the throughput-delay tradeoff in cellular multicast," *International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 2, pp. 1401 – 1406, June 2005.

[14] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 630 – 643, July 2003.

[15] A. Bayesteh and A. K. Khandani, "Throughput and fairness maximization in wireless networks." Technical Report, University of Waterloo.

[16] D. Tse, "Optimal power allocation over parallel gaussian broadcast channels." Unpublished.

[17] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side infonnation," *IEEE Trans. on Inform. Theory*, vol. 51, pp. 506–522, Feb. 2005.