Throughput and Fairness maximization in Wireless Downlink Systems

Alireza Bayesteh, Mehdi Ansari Sadrabadi, and Amir K. Khandani

Coding & Signal Transmission Laboratory(www.cst.uwaterloo.ca)

Dept. of Elec. and Comp. Eng., University of Waterloo

Waterloo, ON, Canada, N2L 3G1

Tel: 519-884-8552, Fax: 519-888-4338

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 maximum fairness in the network, simultaneously, in the asymptotic case of $K \to \infty$. We introduce a new performance metric 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. Finally, the proposed algorithm is generalized for MIMO Broadcast Channels (MIMO-BC), and shown to achieve the maximum throughput and fairness, simultaneously, in the asymptotic case of $K \to \infty$.

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) 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],

Financial supports provided by Nortel, and the corresponding matching funds by the Federal government: Natural Sciences and Engineering Research Council of Canada (NSERC) and Province of Ontario: Ontario Centres of Excellence (OCE) are gratefully acknowledged.

[7]–[21]. In [7], the authors propose an algorithm which maintains a balance between the throughput 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 [8]. In [9], [10], authors propose scheduling metrics that combine multiuser diversity gain with the delay constraints. In [11], the scheduling scheme is designed based on maximizing the effective capacity [22] 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 [12]. This trade-off has been obtained for more general network topologies in [13]. In the static random network with n nodes, the results of [13] show that the optimal tradeoff between throughput T_n and delay D_n is given by $D_n = \Theta(nT_n)$. They also show that the same result is achieved in random mobile networks, when $T_n = O(1/\sqrt{n \log n})$. The first studies on achieving a high throughput and low delay in ad-hoc wireless networks are framed in [4], [14], and [15]. This line of work is further expanded in [13], [16], [17] by using different mobility models such as the random walk and the Brownian mobility models. Neely and Modiano [17] consider the delay-throughput tradeoff only for mobile ad-hoc networks. They investigate the delay characteristics by using the redundant packets transmission through multiple paths. In [18], the authors have proposed and compared different scheduling achemes based on the users' channel qualities and their remaining job times, in the downlink of a MIMO wireless cellular packet data system in fast and slow channel variation scenarios. In [19], the authors have analytically characterized the scheduling gain achieved by opportunistic schedulers with both single-user and multi-user multiplexing, and showed that the average delay grow double-exponentially with the overall throughput, with any opportunistic (single-user time-sharing or multi-user multiplexing) scheduling. In [20], the authors consider a wireless downlink communication system, where the channels are characterized by frequency-selective fading, modeled as a set of M parallel block-fading channels, and a frequency-flat distance-dependent path loss. They compare delay-limited systems (which impose hard fairness) with variable-rate systems (which impose proportional fairness), in terms of the achieved system spectral efficiency C (bit/s/Hz) versus E_b/N_0, and find simple iterative resource allocation algorithms that converge to the optimal delay-limited throughput for orthogonal (FDMA/TDMA) and optimal (superposition/interference cancellation) signaling. In the limit of large K and finite M, the authors find closed-form expressions for C as a function of E_b/N_0 and show that in this limit, the optimal allocation policy consists of letting each user transmit on its best subchannel only.

In [21], the delay is defined as the minimum number of channel uses that guarantees all n users successfully receive m packets. Reference [21] 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 paper, we consider a *hard* delay constraint D for each user, which is enforced by the application or physical limitations (e.g. buffer size). We define a dropping event as the event that there exists a user who does not meet the desired delay constraint. We propose a scheduling scheme for maximizing the throughput of the system, while satisfying the delay constraint for all users. The proposed scheduling algorithm works based on setting a threshold on the channel gain of the users and among the users with channel gains above the threshold, the user with the minimum *Packet Expiry Countdown*s (PED), which is defined as the remaining time to the expiration of that users' packet, is served. By doing asymptotic analysis, it is proved that by selecting the threshold level properly, the proposed scheduling algorithm achieves the maximum throughput, maximum fairness, and minimum delay in the network, simultaneously, in the asymptotic case of $K \to \infty$. The analysis is based on characterizing the probability mass function of PED in terms of K, D , and the threshold value, and evaluating the network dropping probability accordingly. It is also demonstrated that the Round-Robin (RR) scheduling, which focuses on maximizing the fairness and minimizing the delay in the network, and Multi-User Diversity (MUD) scheduling, which focuses on maximizing the throughput in the system, are two extreme cases of the proposed algorithm, where the former suffers from the weak performance in terms of throughput and the latter increases the network delay by a factor of $\log K$. Moreover, we have introduced a new notion of performance in the network, called "Average Throughput", which is defined as the product of the packet arrival rate and the amount of information per channel use in each packet, and proved that the proposed algorithm maximizes the *Minimum Average Throughput* in a broadcast channel. Finally, it is demonstrated that the proposed scheduling outperforms the conventional multiuser diversity scheduling and Round-Robin scheduling in terms of the *Minimum Average Throughput*, by factors $\log K$ and $\log \log K$, respectively. The proposed algorithm is also generalized to MIMO Broadcast Channels (MIMO-BC) by modifying the Random Beam-Forming scheme proposed in [24]. It is shown that the proposed algorithm is capable of achieving the maximum throughput, maximum fairness, and minimum delay, simultaneously, in the asymptotic case of $K \to \infty$. Moreover, it maximizes the *Minimum Average Throughput* in a MIMO-BC.

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 describes the generalization of the proposed algorithm for MIMO-BC, and finally, section V concludes the paper.

Throughout this paper, the norm of the vectors are denoted by $\|.\|$, the Hermitian operation is denoted by $(.)^H$. Notation "$\log$" is used for the natural logarithm, and the rates are expressed in *nats*. $\mathrm{RH}(.)$ represents the right hand side of the equations. For any given functions $f(N)$ and $g(N)$, $f(N) = O(g(N))$ is equivalent to $\lim_{N\to\infty} \left|\frac{f(N)}{g(N)}\right| < \infty$, $f(N) = o(g(N))$ is equivalent to $\lim_{N\to\infty} \left|\frac{f(N)}{g(N)}\right| = 0$, $f(N) = \Omega(g(N))$ is equivalent to $\lim_{N\to\infty} \frac{f(N)}{g(N)} > 0$, $f(N) = \omega(g(N))$ is equivalent to $\lim_{N\to\infty} \frac{f(N)}{g(N)} = \infty$, and $f(N) = \Theta(g(N))$ is equivalent to $\lim_{N\to\infty} \frac{f(N)}{g(N)} = c$, where $0 < c < \infty$. Also, $f(N) \sim g(N)$ is equivalent to $\lim_{N\to\infty} \frac{f(N)}{g(N)} = 1$, $f(N) \gtrsim g(N)$ is equivalent to $\lim_{N\to\infty} \frac{f(N)}{g(N)} \geq 1$, $f(N) \cong g(N) + \odot(h(N))$ is equivalent to $f(N) - g(N) \sim \odot(h(N))$, where $\odot(.)$ can be any of the notations O, o, ω, Ω, or Θ. Moreover, $f(N) \gtrsim= g(N) + \odot(h(N))$ is equivalent to $f(N) - g(N) \gtrsim \odot(h(N))$, where $\odot(.)$ can be any of the notations O, o, ω, Ω, or Θ. Finally, $f(N) \approx g(N)$ means that $f(N)$ is approximately equal to $g(N)$, i.e., if we replace $f(N)$ by $g(N)$ in the equations, the results still hold.

II. System Model and Proposed Algorithm

A. System Model, Assumptions, and Definitions

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

$$y_k = h_k x + n_k, \tag{1}$$

where x denotes the transmitted signal by the BS, which is assumed to be Gaussian with the power constraint P, i.e., $\mathbb{E}\{|x|^2\} \leq P$ [1], $h_k \sim \mathcal{CN}(0,1)$ denotes the channel coefficient between the BS and the kth terminal, and $n_k \sim \mathcal{CN}(0,1)$ is AWGN. We assume that block coding for error free transmission is performed over frames, where the information content of a frame is called packet. In addition, we assume that the frame length is constant (unit of time), while the information content of a frame can potentially vary depending on the capacity of the corresponding channel realization. As we will see later, the proposed method results in almost equal information content (packet length in bits) for all the frames. It is also assumed that *only one user* is served during each frame. The channel coefficients are assumed to be constant for the duration of a frame, and change independently at the start of the next frame (block fading model). The frame itself is assumed to be long enough to allow communication at rates close to the capacity. This model is also used in [21] and [24].

It is assumed that the users have delay constraint D. In other words, the delay between two consecutive 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 for the user which 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 $\nu \leq D$ (Fig. 1). Since

[1]Note that the power constraint here is *per frame*, i.e, is independent of the channel realizations.

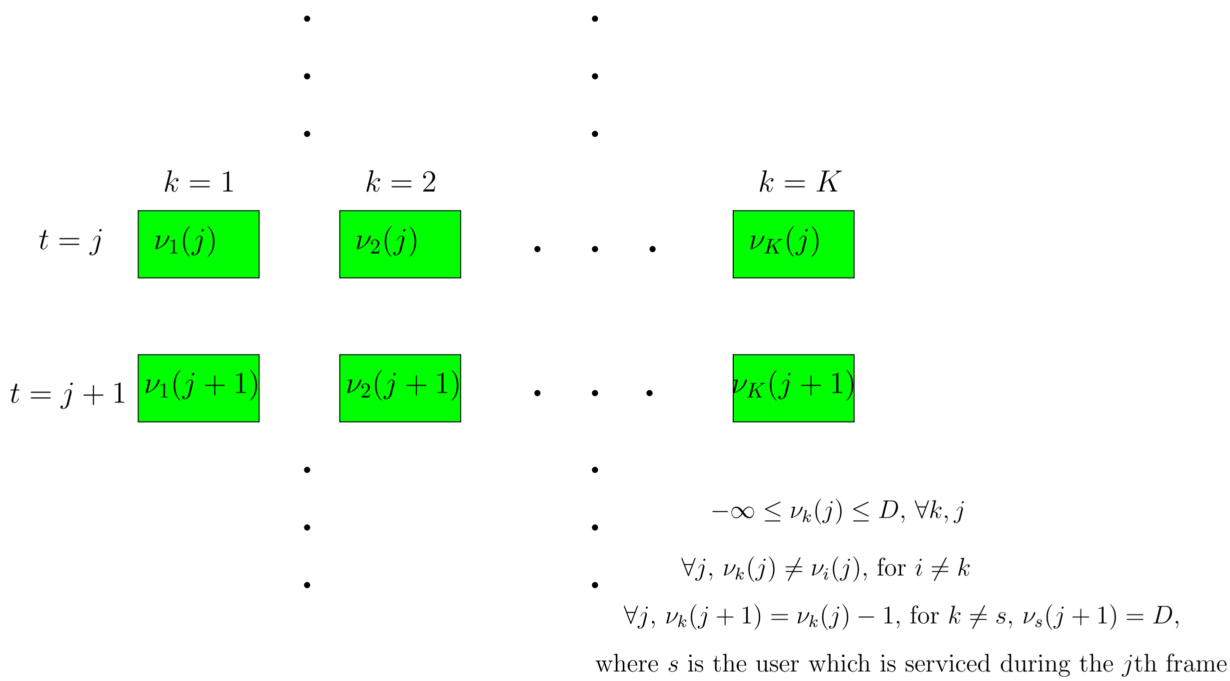

Fig. 1. A Schematic figure for the *expiry countdown*.

the channel model is independent block fading, and the network topology and the proposed scheduling algorithm are symmetric with respect to the users, it can be easily shown that there exists a steady state for the system (no matter what the initial state is), in which the statistical behavior of the users' expiry countdowns is independent of the time index. In the steady state, since in each frame only one user is served by the transmitter, the expiry countdown of the users are distinct in each time. All the results derived in this paper are based on the assumption that the system is in the steady state.

In this paper, we are interested in maximizing the *throughput* and *fairness* in the network. First, we give the definitions of *throughput* and *fairness*:

Definition 1 *The* ***throughput*** *is defined as the average sum-rate of the system, when the average is computed over all the channel realizations.*

Definition 2 *Consider a scheduling $\mathscr{S}$. Then, the* ***Fairness Factor*** *(FF) for this scheduling is defines as*

$$FF(\mathscr{S}) \triangleq \frac{D_{\min}(\mathscr{S})}{K}, \tag{2}$$

where $D_{\min}(\mathscr{S})$ denotes the minimum value of D such that $Pr\{\mathscr{B}\} \to 0$, using scheduling $\mathscr{S}$.

Definition 3 *A scheduling $\mathscr{S}$ is said to achieve the maximum fairness, if $FF(\mathscr{S}) = 1$* [2].

[2]This definition is motivated by the fact that for Round-Robin scheduling (which is known to be the most fair scheduling), $D_{\min} = K$.

B. Proposed Scheduling Algorithm

The proposed scheduling algorithm is described as follows:

Algorithm 1:

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

$$\mathcal{S} \triangleq \{k|\ |h_k|^2 \geq \Theta\}. \tag{3}$$

   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 throughput vs. the fairness in the system. If Θ is chosen to be very large, then the scheduling tends to maximize the throughput. If Θ is chosen to be very small, the algorithm tends to maximize the fairness in the network.

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 ν 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 (3) is simply the set of all users.

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

$$f_\nu(\nu) = \begin{cases} \frac{1}{K} & D-K+1 \leq \nu \leq D \\ 0 & \nu \leq D-K \end{cases}. \tag{4}$$

Proof - Let us define $\nu_{\min}(t) \triangleq \min_{k\in\mathcal{S}} \nu_k(t)$, where $\nu_k(t)$ denotes the *expiry countdown* for the kth user at time t. We have

$$\begin{aligned} \Pr\{\nu_{\min}(t)=l\} &\stackrel{(a)}{=} \sum_{k=1}^{K} \Pr\{\nu_k(t)=l, \nu_i(t)>l, i\neq k\} \\ &\stackrel{(b)}{=} K\Pr\{\nu_1(t)=l, \nu_2(t)>l, \cdots, \nu_K(t)>l\} \\ &= K\Pr\{\nu_1(t)=l\}\Pr\{\nu_2(t)>l, \cdots, \nu_K(t)>l|\, \nu_1(t)=l\}, \end{aligned} \tag{5}$$

where (a) follows from the fact that as in each channel use only one user is served, the random variables $\nu_i(t)$'s are distinct in each time slot t, and (b) results from the symmetry between the users. We have

$$\Pr\{\nu_2(t) > l, \cdots, \nu_K(t) > l | \, \nu_1(t) = l\} = 0, \text{ for } \quad l > D - K + 1, \tag{6}$$

which results from the fact that for $l > D - K + 1$, there are at most $K - 2$ possible choices for each of $\nu_i(t)$, $i = 2, \cdots, K$, and since $\nu_i(t)$ are distinct, the assignment problem has no solution. Moreover, we can write,

$$\Pr\{\nu_k(t) = l - 1\} \;\; = \;\; \Pr\{\nu_k(t-1) = l, \mathscr{X}_k^C(t-1)\}, \tag{7}$$

where $\mathscr{X}_k(t-1)$ represents the event that user k is served during the $(t-1)$th frame, and $\mathscr{X}_k^C(t-1)$ denotes the complement of $\mathscr{X}_k(t-1)$. Since we are studying the behavior of the system in its steady state condition, it follows that the statistical properties of $\nu_k(t)$ and $\mathscr{X}_k(t-1)$ are independent of the time index. Hence, we can drop the time index in the above equation and write

$$\begin{aligned}
\Pr\{\nu_k = l - 1\} \;\; &= \;\; \Pr\{\nu_k = l, \mathscr{X}_k^C\} \\
&= \;\; \Pr\{\nu_k = l\} \left(1 - \Pr\{\mathscr{X}_k | \nu_k = l\}\right) \\
&= \;\; \Pr\{\nu_k = l\} \left(1 - \Pr\{\nu_{\min} = l | \nu_k = l\}\right). \qquad (8)
\end{aligned}$$

Combining (5) and (8), and noting that $\Pr\{\nu_k = l\} = f_\nu(l)$ and $\Pr\{\nu_{\min} = l | \nu_k = l\} = \Pr\{\nu_2 > l, \cdots, \nu_K > l | \nu_1 = l\}$ (by the symmetry), we have

$$f_\nu(l-1) = f_\nu(l) - f_\nu(l) \Pr\{\nu_2 > l, \cdots, \nu_K > l | \nu_1 = l\}. \tag{9}$$

Substituting (6) in (9), we get

$$f_\nu(l) = f_\nu(l-1), \text{ for } \qquad D - K + 2 \leq l \leq D. \tag{10}$$

Since during each frame, exactly one user is served, there is always one user with *expiry countdown* equal to D in the system. In other words,

$$\Pr\left\{ \bigcup_{k=1}^{K} (\nu_k = D) \right\} = 1. \tag{11}$$

Since the events $\nu_k = D$, $k = 1, \cdots, K$, are mutually exclusive, it follows that

$$\begin{aligned}
\sum_{k=1}^{K} \Pr\{\nu_k = D\} \;\; &= \;\; 1 \\
\Rightarrow f_\nu(D) \;\; &\stackrel{(a)}{=} \;\; \frac{1}{K}, \qquad (12)
\end{aligned}$$

where (a) comes from the fact that $\Pr\{\nu_k = D\}$ is the same for all k, and is equal to $f_\nu(D)$. Combining (10) and (12), we have

$$f_\nu(l) = \frac{1}{K}, \qquad D - K + 1 \leq l \leq D. \tag{13}$$

Since $\sum_{l=-\infty}^{D} f_\nu(l) = 1$, from the above equation it follows that

$$f_\nu(l) = 0, \qquad l \leq D - K, \tag{14}$$

which completes the proof of Theorem 1.

■

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 ($FF = 1$), 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. More precisely, it can be easily shown that the achievable throughput of this scheduling scales as $\Theta(1)$.

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 [23].

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

$$f_\nu(\nu) = \frac{1}{K}\left(1 - \frac{1}{K}\right)^{D-\nu} u(D - \nu), \tag{15}$$

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

Proof - Similar to (8), we can write

$$\begin{aligned} f_\nu(l-1) &= f_\nu(l)\left(1 - \Pr\{\mathscr{X}_k | \nu_k = l\}\right) \\ &\stackrel{(a)}{=} f_\nu(l)\left(1 - \Pr\{\mathscr{X}_k\}\right) \\ &\stackrel{(b)}{=} f_\nu(l)\left(1 - \frac{1}{K}\right), \end{aligned} \tag{16}$$

where (a) comes from the fact that the selection of users is performed regardless of the value of their *expiry countdown*. (b) results from the fact that the fading process is block-wise independent, and as a result, the probability that the channel norm of any user is the highest during a frame is $\frac{1}{K}$. From the above equation, the pmf of ν can be written as

$$f_\nu(l) = f_\nu(D)\left(1 - \frac{1}{K}\right)^{D-l}, \quad l \leq D. \tag{17}$$

From (16) and noting that $\sum_{l=-\infty}^{D} f_\nu(l) = 1$, we have $f_\nu(D) = \frac{1}{K}$. Hence,

$$f_\nu(l) = \frac{1}{K}\left(1 - \frac{1}{K}\right)^{D-l} u(D-l), \tag{18}$$

where $u(.)$ denotes the unit step function. Hence, the pmf of ν follows the exponential distribution with the parameter $1 - \frac{1}{K}$.

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

$$D \cong K \log K + \omega(K). \tag{19}$$

Proof - *Sufficient Condition:* Using (18), the dropping probability for a user k, denoted as $\Pr\{\mathscr{B}_k\}$, can be written as

$$\begin{aligned} \Pr\{\mathscr{B}_k\} &= \sum_{l=-\infty}^{0} f_\nu(l) \\ &= \sum_{l=-\infty}^{0} \frac{1}{K}\left(1 - \frac{1}{K}\right)^{D-l} \\ &= \left(1 - \frac{1}{K}\right)^{D} \\ &\sim e^{-\frac{D}{K}}. \end{aligned} \tag{20}$$

The network dropping probability ($\Pr\{\mathscr{B}\}$) can be written as $\Pr\{\bigcup_{k=1}^{K} \mathscr{B}_k\}$. Using the union bound for the probability, we have

$$\begin{aligned} \Pr\{\mathscr{B}\} &\leq \sum_{k=1}^{K} \Pr\{\mathscr{B}_k\} \\ &\overset{(20)}{\sim} K e^{-\frac{D}{K}} \\ &= e^{-\frac{D - K\log K}{K}}. \end{aligned} \tag{21}$$

Hence, having $D \cong K \log K + \omega(K)$ guarantees $\Pr\{\mathscr{B}\} \to 0$.

Necessary Condition: We can write

$$\Pr\{\mathscr{B}\} = 1 - \Pr\left\{\bigcap_{k=1}^{K} \mathscr{B}_k^C\right\}. \tag{22}$$

The dropping event for the kth user, $\mathscr{B}_k$, is equivalent to $\nu_k \leq 0$. Hence, the above equation can be written as

$$\begin{aligned}
\Pr\{\mathscr{B}\} &= 1 - \Pr\{\nu_1 > 0, \cdots, \nu_K > 0\} \\
&= 1 - \Pr\{\nu_1 > 0\} \prod_{k=2}^{K} \Pr\{\nu_k > 0 | \nu_1 > 0, \nu_2 > 0, \cdots, \nu_{k-1} > 0\} \\
&= 1 - \Pr\{\nu_1 > 0\} \prod_{k=2}^{K} \left(\frac{\sum_{\substack{(a_1,\cdots,a_{k-1}) \\ 1 \leq a_i \leq D}} f_{\nu_1,\cdots,\nu_{k-1}}(a_1, \cdots, a_{k-1})}{\Pr\{\nu_1 > 0, \cdots, \nu_{k-1} > 0\}} \times \right. \\
&\quad \left. \Pr\{\nu_k > 0 | \nu_1 = a_1, \nu_2 = a_2, \cdots, \nu_{k-1} = a_{k-1}\} \right) \qquad (23) \\
&\overset{(a)}{=} 1 - \Pr\{\nu_1 > 0\} \prod_{k=2}^{K} \left(\frac{\sum_{\substack{(a_1,\cdots,a_{k-1}) \\ 1 \leq a_i \leq D}} f_{\nu_1,\cdots,\nu_{k-1}}(a_1, \cdots, a_{k-1})}{\Pr\{\nu_1 > 0, \cdots, \nu_{k-1} > 0\}} \times \right. \\
&\quad \left. \Pr\{\nu_k > 0 | \nu_k \notin \{a_1, a_2, \cdots, a_{k-1}\}\} \right) \\
&= 1 - \Pr\{\nu_1 > 0\} \prod_{k=2}^{K} \left(\frac{\sum_{\substack{(a_1,\cdots,a_{k-1}) \\ 1 \leq a_i \leq D}} f_{\nu_1,\cdots,\nu_{k-1}}(a_1, \cdots, a_{k-1})}{\Pr\{\nu_1 > 0, \cdots, \nu_{k-1} > 0\}} \times \right. \\
&\quad \left. \frac{\Pr\{\nu_k > 0\} - \sum_{i=1}^{k-1} f_{\nu_k}(a_i)}{1 - \sum_{i=1}^{k-1} f_{\nu_k}(a_i)} \right) \qquad (24) \\
&\overset{(b)}{\geq} 1 - \prod_{k=1}^{K} \Pr\{\nu_k > 0\} \\
&\overset{(20)}{=} 1 - \left[1 - \left(1 - \frac{1}{K} \right)^D \right]^K \\
&\overset{(c)}{\geq} 1 - e^{-K\left(1-\frac{1}{K}\right)^D}, \qquad (25)
\end{aligned}$$

where (a) follows from the fact that the only dependency among ν_k's is that they are distinct random variables, (b) results from the fact that $\frac{\Pr\{\nu_k > 0\} - \sum_{i=1}^{k-1} f_{\nu_k}(a_i)}{1 - \sum_{i=1}^{k-1} f_{\nu_k}(a_i)} \leq \Pr\{\nu_k > 0\}$, and (c) results from the fact that $(1-x)^n \leq e^{-nx}$, $\forall n > 0, x < 1$. It follows from the above equation that in order to have $\Pr\{\mathscr{B}\} \to 0$, we must have $e^{-K\left(1-\frac{1}{K}\right)^D} \to 1$, which incurs $K\left(1 - \frac{1}{K}\right)^D \to 0$. Since $K \to \infty$, we can write

$$\begin{aligned}
K\left(1 - \frac{1}{K}\right)^D &= K e^{D \log\left(1-\frac{1}{K}\right)} \\
&\sim K e^{-\frac{D}{K}(1+O(1/K))} \\
&\sim e^{-\frac{D-K\log K}{K}(1+O(1/K))}. \qquad (26)
\end{aligned}$$

Hence, $K\left(1 - \frac{1}{K}\right)^D \to 0$ is equivalent to $\frac{D-K\log K}{K} \to \infty$, which incurs $D \cong K \log K + \omega(K)$. This completes the proof of Theorem 3.

■

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 Fairness Factor (or equivalently, a factor of $\log K$ increase in the network delay), which is due to ignoring ν in the scheduling [3]

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*

$$\log K - 2 \log\log K < \Theta < \log K - 1.5 \log\log K, \tag{27}$$

one can simultaneously achieve:
I- Maximum Throughput:

$$\lim_{K\to\infty} \mathcal{C}_{\text{sum}} - \mathcal{R} = 0, \tag{28}$$

in which $\mathcal{C}_{\text{sum}}$ denotes the maximum achievable sum-rate in the broadcast channel and $\mathcal{R}$ denotes the achievable sum-rate of the proposed algorithm, and
II- Maximum Fairness:

$$\lim_{K\to\infty} \frac{D}{K} = 1, \quad \textit{while } \textbf{\textit{Pr}}\{\mathscr{B}\} \to 0 \ (\textit{or equivalently}, \lim_{K\to\infty} FF = 1). \tag{29}$$

Proof - The steps of the proof are as follows: in Lemma 1, we study the behavior of $f_\nu(l)$ and derive a difference equation satisfied by $f_\nu(l)$. In Lemma 2, we derive an explicit solution for this difference equation. Based on this solution, in Lemma 3, we present a sufficient condition such that the conditions $\lim_{K\to\infty} \frac{D}{K} \to 1$ 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 (27).

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

$$f_\nu(l) - f_\nu(l-1) \;=\; \eta f_\nu(l)\left[1 - pF_\nu(l)\right]^{K-1}\left[1 + O(1/\sqrt{K})\right], \tag{30}$$

[3]It should be noted that this scheduling is *long-term fair*, i.e., all the users are equally served over a long period of time. However, with our definition of fairness (which can be called *short-term* fairness), this scheduling is away from the maximum fairness by a factor of $\log K$.

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

Proof - Similar to (8), we have

$$f_\nu(l-1) = f_\nu(l)\left(1 - \Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\}\right), \tag{31}$$

where $\nu_{\min} = \min_k\{\nu_k | k \in \mathcal{S}\}$. Having the fact that

$$p \triangleq \Pr\{k \in \mathcal{S}\} = e^{-\Theta}, \tag{32}$$

which is resulted from the exponential distribution for $|h_k|^2$ (as a result of the Complex Gaussian distribution for h_k), and the independence between the users' channels, it follows that $|\mathcal{S}|$ is a Binomial random variable with parameters (K, p). As a result, we have

$$\begin{aligned}
\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\} &= \sum_{n=1}^{K} \Pr\{\nu_{\min} = l, k \in \mathcal{S}, |\mathcal{S}| = n | \nu_k = l\} \\
&= \sum_{n=1}^{K} \Pr\{|\mathcal{S}| = n, k \in \mathcal{S} \; | \nu_k = l\} \Pr\{\nu_{\min} = l | \, \nu_k = l, |\mathcal{S}| = n, k \in \mathcal{S}\} \\
&\overset{(a)}{=} \sum_{n=1}^{K} \Pr\{|\mathcal{S}| = n, k \in \mathcal{S}\} \Pr\{\nu_{\min} = l | \, \nu_k = l, |\mathcal{S}| = n, k \in \mathcal{S}\} \\
&= \sum_{n=1}^{K} \binom{K-1}{n-1} p^n (1-p)^{K-n} \Pr\{\nu_{\min} = l | \, \nu_k = l, |\mathcal{S}| = n, k \in \mathcal{S}\} \\
&= \sum_{n=1}^{K} \binom{K-1}{n-1} p^n (1-p)^{K-n} \times \\
&\quad \Pr\{\nu_i > l, i \in \mathcal{S}, i \neq k | \, \nu_k = l, |\mathcal{S}| = n, k \in \mathcal{S}\},
\end{aligned} \tag{33}$$

where (a) comes from the fact that the events $|\mathcal{S}| = n$ and $k \in \mathcal{S}$ are independent of the event $\nu_k(t) = l$. In fact, the event $\nu_k(t) = l$ is a function of $\{h_k(j)\}_{k=1}^{K}$, $j < t$, while the events $|\mathcal{S}(t)| = n$ and $k \in \mathcal{S}(t)$ are functions of $\{h_k(t)\}_{k=1}^{K}$, and because of the independent block fading assumption, are independent of $\{h_k(j)\}_{k=1}^{K}$, $j < t$, and consequently independent of $\nu_k(t) = l$.

To evaluate the right hand side of the above equation, we need to find the following probability:

$$\Pr\{\nu_i > l, i \in \mathcal{S}, i \neq k | \, \nu_k = l, |\mathcal{S}| = n, k \in \mathcal{S}\}, \tag{34}$$

which is, by symmetry, equal to

$$\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l | \, \nu_n = l\}, \tag{35}$$

noting that $\nu_k(t)$ and $h_k(t)$ are independent random variables. An upper-bound on this probability can be given as bellow:

$$\begin{aligned}
\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l |\, \nu_n = l\} &= \Pr\{\nu_1 > l |\, \nu_n = l\} \times \\
&\quad \prod_{i=2}^{n-1} \Pr\{\nu_i > l |\, \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\} \qquad (36) \\
&\overset{(a)}{\leq} \left[\Pr\{\nu_i > l |\, \nu_n = l\}\right]^{n-1} \\
&\overset{(b)}{=} \left[\frac{G_\nu(l)}{1 - f_\nu(l)}\right]^{n-1}, \qquad (37)
\end{aligned}$$

where (a) follows from (24), in which we have shown that $\Pr\{\nu_i > l |\, \nu_1 > l, \cdots, \nu_{i-1} > l\} \leq \Pr\{\nu_i > l\}$, and by following the same approach we can show $\Pr\{\nu_i > l |\, \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\} \leq \Pr\{\nu_i > l |\, \nu_n = l\}$, and (b) results from the fact that the only dependency between ν_i and ν_n is that they are distinct, and hence $(\nu_i > l |\, \nu_n = l)$ is equivalent to $(\nu_i > l |\, \nu_i \neq l)$, with the probability of $\frac{G_\nu(l)}{1-f_\nu(l)}$, where $G_\nu(l) \triangleq 1 - F_\nu(l)$.

In order to lower-bound $\Pr\{\nu_i > l |\, \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\}$, we need to derive an upper-bound on $f_\nu(l)$. Since $f_\nu(l)$ is an increasing function of l (from (31)), it follows that

$$f_\nu(l) \leq f_\nu(D), \quad \forall l. \qquad (38)$$

However, unlike the previous cases, $f_\nu(D) \neq \frac{1}{K}$. This results from the fact that using the proposed algorithm in the general case, it is probable that no user is served. Defining the event $\mathscr{X}(t) \triangleq \bigcup_{k=1}^{K} \mathscr{X}_k(t)$ as the event of serving at least one user in frame t, we have

$$\begin{aligned}
\Pr\{\mathscr{X}(t)\} &= \Pr\{|\mathcal{S}(t)| > 0\} \\
&= 1 - \prod_{k=1}^{K} \Pr\{|h_k|^2 < \Theta\} \\
&= 1 - \left(1 - e^{-\Theta}\right)^K. \qquad (39)
\end{aligned}$$

Noting that $\log K - 2\log\log K < \Theta < \log K - 1.5\log\log K$, we have $\frac{(\log K)^{1.5}}{K} < e^{-\Theta} < \frac{(\log K)^2}{K}$, and hence, $\left(1 - e^{-\Theta}\right)^K \lesssim e^{-(\log K)^{1.5}}$. Moreover, $\Pr\{\mathscr{X}(t)\}$ in terms of $f_\nu(D)$ can be written as

$$\begin{aligned}
\Pr\{\mathscr{X}(t)\} &= \sum_{k=1}^{K} \Pr\{\nu_k(t+1) = D\} \\
&= K f_\nu(D), \qquad (40)
\end{aligned}$$

where the first line comes from the distinction of ν_k's and the second line follows from the symmetry between the users and dropping the time index. Combining (39) and (40) yields,

$$f_\nu(D) \cong \frac{1}{K}\left[1 - \left|O\left(e^{-(\log K)^{1.5}}\right)\right|\right], \qquad (41)$$

which is less than $\frac{1}{K}$. Combining (38) with the above equation yields

$$f_\nu(l) \leq \frac{1}{K}, \quad \forall l. \tag{42}$$

Similar to (23) and (24), we can write

$$\begin{aligned}
\Pr\{\nu_i > l |\, \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\} &= \frac{\sum_{\substack{(a_1,\cdots,a_{i-1})\\ l<a_k\leq D}} f_{\nu_1,\cdots,\nu_{i-1},\nu_n}(a_1,\cdots,a_{i-1},l)}{\Pr\{\nu_1 > l,\cdots,\nu_{i-1}>l,\nu_n=l\}} \times \\
&\quad \Pr\{\nu_i > l |\, \nu_1 = a_1, \cdots, \nu_{i-1} = a_{i-1}, \nu_n = l\} \\
&= \frac{\sum_{\substack{(a_1,\cdots,a_{i-1})\\ l<a_k\leq D}} f_{\nu_1,\cdots,\nu_{i-1},\nu_n}(a_1,\cdots,a_{i-1},l)}{\Pr\{\nu_1 > l,\cdots,\nu_{i-1}>l,\nu_n=l\}} \times \\
&\quad \Pr\left\{\nu_i > l |\, \nu_i \notin \{a_1,\cdots,a_{i-1},l\}\right\} \\
&= \frac{\sum_{\substack{(a_1,\cdots,a_{i-1})\\ l<a_k\leq D}} f_{\nu_1,\cdots,\nu_{i-1},\nu_n}(a_1,\cdots,a_{i-1},l)}{\Pr\{\nu_1 > l,\cdots,\nu_{i-1}>l,\nu_n=l\}} \times \\
&\quad \frac{\Pr\{\nu_i > l\} - \sum_{k=1}^{i-1} f_{\nu_i}(a_k)}{1 - \sum_{k=1}^{i-1} f_{\nu_i}(a_k) - f_{\nu_i}(l)} \\
&\geq \frac{\sum_{\substack{(a_1,\cdots,a_{i-1})\\ l<a_k\leq D}} f_{\nu_1,\cdots,\nu_{i-1},\nu_n}(a_1,\cdots,a_{i-1},l)}{\Pr\{\nu_1 > l,\cdots,\nu_{i-1}>l,\nu_n=l\}} \times \\
&\quad \left(\Pr\{\nu_i > l\} - \sum_{k=1}^{i-1} f_{\nu_i}(a_k)\right) \\
&\overset{(a)}{\geq} G_\nu(l) - \frac{i-1}{K},
\end{aligned} \tag{43}$$

where (a) follows from the fact that $f_{\nu_i}(a_k) \leq \frac{1}{K}$, $\forall a_k$ (equation (42)). From the above equation and (36), $\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l |\, \nu_n = l\}$ can be lower-bounded as

$$\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l |\, \nu_n = l\} \geq \prod_{i=0}^{n-2} \left(G_\nu(l) - \frac{i}{K}\right). \tag{44}$$

Using the above equation, and defining $n_0 \triangleq 3(\log K)^2$ and $D_0 \triangleq D - \sqrt{K} n_0(n_0-1)$, a lower-bound on $\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l |\, \nu_n = l\}$ is given as,

$$\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l |\, \nu_n = l\} \geq g(n,l), \tag{45}$$

where

$$g(n,l) \triangleq \begin{cases} \prod_{i=0}^{n-2} \left(G_\nu(l) - \frac{i}{K}\right) & l < D_0 \text{ and } n \leq n_0 \\ 0 & \text{Otherwise.} \end{cases} \tag{46}$$

As we will see later, the form in (46) is more convenient to carry out our subsequent derivations.

From (33), (35), (36), and (37), an upper-bound on $\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\}$ can be obtained as follows:

$$\begin{aligned}
\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\} &\leq \sum_{n=1}^{K} \binom{K-1}{n-1} p^n (1-p)^{K-n} \left(\frac{G_\nu(l)}{1-f_\nu(l)} \right)^{n-1} \\
&= \eta \sum_{n=1}^{K} \binom{K-1}{n-1} p^{n-1} (1-p)^{K-n+1} \left(\frac{G_\nu(l)}{1-f_\nu(l)} \right)^{n-1} \\
&= \eta \sum_{n=0}^{K-1} \binom{K-1}{n} \left(\frac{p\, G_\nu(l)}{1-f_\nu(l)} \right)^{n} (1-p)^{K-n} \\
&= \eta \left(\frac{p\, G_\nu(l)}{1-f_\nu(l)} + 1 - p \right)^{K-1} \\
&\overset{(a)}{<} \eta \left(p\, G_\nu(l) \Big(1 + \frac{2}{K}\Big) + 1 - p \right)^{K-1} \\
&= \eta \left(1 - p\, F_\nu(l)\right)^{K-1} \left(1 + \frac{2p\, G_\nu(l)}{K(1-pF_\nu(l))} \right)^{K-1} \\
&\overset{(b)}{\leq} \eta \left(1 - p\, F_\nu(l)\right)^{K-1} \left(1 + \frac{2p\, G_\nu(l)}{K(1-p)} \right)^{K-1} \\
&\sim \eta \left(1 - p\, F_\nu(l)\right)^{K-1} e^{\frac{2p\, G_\nu(l)}{1-p}} \\
&\overset{(c)}{\cong} \eta \left(1 - p\, F_\nu(l)\right)^{K-1} [1 + O(p)],
\end{aligned} \tag{47}$$

where $\eta \triangleq \frac{p}{1-p}$. (a) comes from the facts that $\forall l$, $f_\nu(l) \leq \frac{1}{K}$ (equation (42)), and for x sufficiently small, $\frac{1}{1-x} < 1 + 2x$, (b) results from $F_\nu(l) \leq 1$, and (c) follows from the fact that since $\log K - 2 \log\log K < \Theta < \log K - 1.5 \log\log K$, we have $\frac{(\log K)^{1.5}}{K} < p = e^{-\Theta} < \frac{(\log K)^2}{K}$, which implies that $p \sim o(1)$.

Moreover, from (33), (35), and (45), a lower-bound on $\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\}$, for $l < D_0$, is given as follows:

$$\begin{aligned}
\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\} &\geq \sum_{n=1}^{K} \binom{K-1}{n-1} p^n (1-p)^{K-n} g(n,l) \\
&= \sum_{n=1}^{n_0} \binom{K-1}{n-1} p^n (1-p)^{K-n} \prod_{i=0}^{n-2} \left(G_\nu(l) - \frac{i}{K} \right) \\
&= \sum_{n=1}^{n_0} \binom{K-1}{n-1} p^n (1-p)^{K-n} G_\nu(l)^{n-1} \prod_{i=0}^{n-2} \left(1 - \frac{i}{K G_\nu(l)} \right).
\end{aligned} \tag{48}$$

By repeated application of (31) and using (47) to upper-bound $\Pr\{\nu_{\min}=l, k\in\mathcal{S}|\nu_k=l\}$, we obtain

$$
\begin{aligned}
f_\nu(D)-f_\nu(D_0) &\leq \sum_{l=D_0}^{D}\eta f_\nu(l)\left(1-pF_\nu(l)\right)^{K-1}[1+O(p)]\\
&\overset{(a)}{\leq} \frac{\eta(D-D_0+1)}{K}\left(1-p+p\frac{D-D_0}{K}\right)^{K-1}[1+O(p)]\\
&\leq \frac{\eta(D-D_0+1)}{K}e^{-(K-1)p\left(1-\frac{D-D_0}{K}\right)}[1+O(p)], \qquad (49)
\end{aligned}
$$

where (a) comes from the fact that $f_\nu(l)\leq\frac{1}{K}$ and as a result $F_\nu(l)\geq 1-\frac{D-l}{K}$, which implies that $F_\nu(l)\geq 1-\frac{D-D_0}{K}$ for $l\geq D_0$. Having the facts that $D-D_0\sim 9\sqrt{K}(\log K)^4$ and $\log K-2\log\log K<\Theta<\log K-1.5\log\log K$, which results in $\frac{(\log K)^{1.5}}{K}<p<\frac{(\log K)^2}{K}$, and $\eta=\frac{p}{1-p}\sim p$, the right hand side of the above equation can be upper-bounded as

$$
\text{RH}(49) \lesssim \frac{9(\log K)^6}{K^{3/2}}e^{-(\log K)^{1.5}}. \qquad (50)
$$

Substituting in (49) and using (41), noting that $e^{-(\log K)^{1.5}}\sim o(1/K)$, we obtain

$$
f_\nu(D_0)\cong\frac{1}{K}\left[1+o(1/K)\right]. \qquad (51)
$$

Since $f_\nu(l)$ is an increasing function of l, it follows from the above equation that

$$
f_\nu(l)\cong\frac{1}{K}\left[1+o(1/K)\right], \quad \forall l, D_0\leq l\leq D. \qquad (52)
$$

The above equation incurs that for $l<D_0$, $G_\nu(l)\gtrsim\frac{D-D_0}{K}=\frac{n_0(n_0-1)}{\sqrt{K}}$. As a result, $\prod_{i=0}^{n-2}\left(1-\frac{i}{KG_\nu(l)}\right)$ in (48) can be lower-bounded as

$$
\begin{aligned}
\prod_{i=0}^{n-2}\left(1-\frac{i}{KG_\nu(l)}\right) &\overset{(a)}{\gtrsim} \prod_{i=0}^{n_0-2}\left(1-\frac{i}{\sqrt{K}n_0(n_0-1)}\right)\\
&\overset{(b)}{\approx} \prod_{i=0}^{n_0-2}e^{-\frac{i}{\sqrt{K}n_0(n_0-1)}}\\
&= e^{-\frac{(n_0-1)(n_0-2)}{2\sqrt{K}n_0(n_0-1)}}\\
&\cong 1+O\left(1/\sqrt{K}\right), \qquad (53)
\end{aligned}
$$

where (a) follows from the fact that $n\leq n_0$, and (b) results from the fact that as $i<n_0$, $\frac{i}{\sqrt{K}n_0(n_0-1)}\ll 1$, which implies that $1-\frac{i}{\sqrt{K}n_0(n_0-1)}\approx e^{-\frac{i}{\sqrt{K}n_0(n_0-1)}}$. Moreover, similar to (47), we can write $\Psi\triangleq\sum_{n=1}^{n_0}\binom{K-1}{n-1}p^n(1-p)^{K-n}G_\nu(l)^{n-1}$ as

$$
\begin{aligned}
\Psi &= \eta\left[\left(1-pF_\nu(l)\right)^{K-1}-\sum_{n=n_0}^{K-1}\binom{K-1}{n}p^n(1-p)^{K-n}G_\nu(l)^n\right]\\
&\geq \eta\left[\left(1-pF_\nu(l)\right)^{K-1}-\sum_{n=n_0}^{K-1}\binom{K-1}{n}p^n(1-p)^{K-n}\right]\\
&\overset{(a)}{\approx} \eta\left[\left(1-pF_\nu(l)\right)^{K-1}-Q\left(\frac{n_0-(K-1)p}{\sqrt{(K-1)p(1-p)}}\right)\right], \qquad (54)
\end{aligned}
$$

where (a) results from the Gaussian approximation for a Binomial distribution with parameters (n, p), when $np \to \infty$. Noting $n_0 = 3[\log K]^2$ and $p < \frac{[\log K]^2}{K}$, it follows that $n_0 \geq 3(K-1)p$. Substituting in the above equation, and having the fact that $Q(x) \approx \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$ for large enough x, the right hand side of the above equation can be lower-bounded as

$$\text{RH (54)} \;\; \geq \;\; \eta \left[(1 - pF_\nu(l))^{K-1} - e^{-2(K-1)p} \right]. \tag{55}$$

Having the facts that $(1 - pF_\nu(l))^{K-1} \sim e^{-(K-1)pF_\nu(l)} \geq e^{-(K-1)p}$, RH (55) can be lower-bounded as

$$\begin{aligned} \text{RH (55)} \;\; &\geq \;\; \eta \left(1 - pF_\nu(l)\right)^{K-1} \left[1 - e^{-(K-1)p}\right] \\ &\overset{(a)}{\cong} \;\; \eta \left(1 - pF_\nu(l)\right)^{K-1} \left[1 + O(1/K)\right], \end{aligned} \tag{56}$$

where (a) follows from the fact that as $p > \frac{(\log K)^{1.5}}{K}$, we have $e^{-(K-1)p} \sim O(1/K)$. Combining (48), (53), (54), (55), and (56), we have

$$\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\} \;\; \gtrsim\!= \;\; \eta \left(1 - pF_\nu(l)\right)^{K-1} \left[1 + O\left(1/\sqrt{K}\right)\right], \tag{57}$$

for $l < D_0$. Combining (47) and (57), noting that $p \sim o(1/\sqrt{K})$, yields

$$\Pr\{\nu_{\min} = l, k \in \mathcal{S} | \nu_k = l\} \;\; \cong \;\; \eta \left(1 - pF_\nu(l)\right)^{K-1} \left[1 + O\left(1/\sqrt{K}\right)\right], \tag{58}$$

for $l < D_0$. Substituting in (31), we have

$$f_\nu(l) - f_\nu(l-1) \cong \eta f_\nu(l) \left(1 - pF_\nu(l)\right)^{K-1} \left[1 + O\left(1/\sqrt{K}\right)\right], \quad l < D_0. \tag{59}$$

Moreover, for $D_0 \leq l \leq D$, from (52), we have $f_\nu(l) \cong \frac{1}{K}\left[1 + o(1/K)\right]$, which completes the proof of Lemma 1.

■

Lemma 2 *The solution to the difference equation (30), in the asymptotic case of $K \to \infty$, is*

$$f_\nu(l) \sim \frac{\frac{\varphi}{(K-1)p} e^{(K-1)p} e^{\varphi(l-D_0)}}{1 + e^{(K-1)p} e^{\varphi(l-D_0)}} \quad l < D_0, \tag{60}$$

for some $\varphi \cong \eta \left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$.

Proof - Rewriting (30), we have

$$\begin{aligned} f_\nu(l) - f_\nu(l-1) \;\; &\cong \;\; \eta f_\nu(l) \left(1 - pF_\nu(l)\right)^{K-1} \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\overset{(a)}{\cong} \;\; \eta f_\nu(l) e^{-(K-1)pF_\nu(l)[1+O(p)]} \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\cong \;\; \eta f_\nu(l) e^{-(K-1)pF_\nu(l)} \left[1 + O(Kp^2)\right] \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\overset{(b)}{\cong} \;\; \eta f_\nu(l) e^{-(K-1)pF_\nu(l)} \left[1 + O\left(1/\sqrt{K}\right)\right] \quad l < D_0, \end{aligned} \tag{61}$$

where (a) comes from the fact that $(1+x)^n \sim e^{xn[1+O(x)]}$ for $x \sim o(1)$, and (b) results from the fact that $p < \frac{[\log K]^2}{K}$ and as a result, $Kp^2 \sim o\left(1/\sqrt{K}\right)$.

Now, consider the following differential equation:

$$x'(u) = \varphi x(u) e^{-(K-1)pX(u)} \quad u < D_0, \tag{62}$$

with the boundary conditions: $x(-\infty) = X(-\infty) = 0$, and $X(D_0) = 1 - \frac{D-D_0}{K}$, in which u is a continuous variable, and $X(u) = \int_{-\infty}^{u} x(t)dt$, and $\varphi \cong \eta \left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$. Writing the Taylor series for $x(u-1)$ about u, we have

$$x(u) - x(u-1) = x'(u) + \sum_{n=2}^{\infty} \frac{(-1)^{n+1} x^{(n)}(u)}{n!}. \tag{63}$$

For the second derivative of (62), we have

$$\begin{aligned} x''(u) &= \varphi x'(u) e^{-(K-1)pX(u)} - \varphi(K-1)px(u)^2 e^{-(K-1)pX(u)} \\ &= \varphi x'(u) e^{-(K-1)pX(u)} - (K-1)px'(u)x(u). \end{aligned} \tag{64}$$

From the above equation, noting that with the given boundary conditions for the differential equation in (62), we have $e^{-(K-1)pX(u)} \leq 1$ (which follows from the facts that $x'(u) \geq 0$ and $x(u) \geq 0$, which incurs $X(u) \geq 0$), and $x(u) \leq \frac{1}{K}$ (which follows from solving (62) with the boundary condition $X(D_0) = 1 - \frac{D-D_0}{K}$), it is easy to see that $|x''(u)| < \varphi |x'(u)|$. Similarly, we can show that $|x^{(n)}(u)| < 2^{n-1}\varphi^n |x'(u)|$. Substituting in (63), noting that $\varphi \sim \eta \sim p < \frac{[\log K]^2}{K}$, yields

$$\begin{aligned} x(u) - x(u-1) &\cong x'(u)[1 + O(\varphi)] \\ &\overset{(a)}{\cong} \varphi x(u) e^{-(K-1)pX(u)} \left[1 + O\left(\varphi\right)\right] \\ &\overset{(b)}{\cong} \eta x(u) e^{-(K-1)pX(u)} \left[1 + O\left(1/\sqrt{K}\right)\right] \quad u < D_0, \end{aligned} \tag{65}$$

where (a) comes from (62) and (b) follows from the facts that $\varphi \cong \eta \left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$ and $\varphi \sim O(1/\sqrt{K})$. We also have

$$\begin{aligned} X(u) &\overset{(a)}{=} \sum_{v=-\infty}^{u} [X(v) - X(v-1)] \\ &\overset{(b)}{=} \sum_{v=-\infty}^{u} \left[x(v) + \sum_{n=1}^{\infty} \frac{(-1)^n x^{(n)}(v)}{(n+1)!}\right] \\ &\overset{(c)}{\sim} \sum_{v=-\infty}^{u} x(v) \left[1 + O(\varphi)\right], \end{aligned} \tag{66}$$

where (a) results from the fact that $X(-\infty) = 0$, (b) follows from writing the Tailor series for $X(v-1)$ about v, and (c) comes from the the fact that $|x'(v)| \leq \varphi x(v)$, $\forall v$ (62), and also $|x^{(n)}(v)| < 2^{n-1}\varphi^n |x'(v)|$, demonstrated earlier. Defining $Z(u) \triangleq \sum_{v=-\infty}^{u} x(v)$ and using the above equation and (65), we have

$$\begin{aligned} x(u) - x(u-1) &\cong \eta x(u) e^{-(K-1)pX(u)} \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\cong \eta x(u) e^{-(K-1)pZ(u)[1+O(\varphi)]} \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\overset{(a)}{\cong} \eta x(u) e^{-(K-1)pZ(u)} \left[1 + O(Kp^2)\right] \left[1 + O\left(1/\sqrt{K}\right)\right] \\ &\overset{(b)}{\cong} \eta x(u) e^{-(K-1)pZ(u)} \left[1 + O\left(1/\sqrt{K}\right)\right], \end{aligned} \tag{67}$$

where (a) results from the fact that $\varphi \sim p$, and (b) follows from the fact that $p < \frac{[\log K]^2}{K}$ and as a result, $Kp^2 \sim o\left(1/\sqrt{K}\right)$ (similar to (b) in (61)). The above equation incurs that the solution of (62) also satisfies (61). More precisely, for any value of l, $l < D_0$, there exists a φ such that $\varphi \cong \eta\left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$, and $f_\nu(l) \sim x(l)$, where $f_\nu(l)$ is the solution of (61) and $x(l)$ is the solution of (62) at $u = l$. This suggests us to solve the differential equation (62), instead of the difference equation (61), assuming the same boundary conditions. The boundary conditions are $x(-\infty) = f_\nu(-\infty) = 0$ and $X(D_0) = F_\nu(D_0) = 1 - \frac{D-D_0}{K}$. The second condition comes from the fact that $f_\nu(l) \approx \frac{1}{K}$, for $l \geq D_0$.

By taking the integral from both sides of (62), we obtain

$$x(u) = -\frac{\varphi}{(K-1)p} e^{-(K-1)pX(u)} + c. \tag{68}$$

Noting that $X(-\infty) = x(-\infty) = 0$, $c = \frac{\varphi}{(K-1)p}$. Substituting $e^{-(K-1)pX(u)}$ by $\frac{x'(u)}{\varphi x(u)}$ from (62), we come up with the following differential equation:

$$\frac{x'(u)}{\varphi x(u)\left[1 - \frac{(K-1)p}{\varphi} x(u)\right]} = 1, \tag{69}$$

which can be solved as follows:

$$\begin{aligned} \frac{x'(u)}{x(u)} + \frac{\frac{(K-1)p}{\varphi} x'(u)}{1 - \frac{(K-1)p}{\varphi} x(u)} &= \varphi \\ \Rightarrow \ln \frac{x(u)}{1 - \frac{(K-1)p}{\varphi} x(u)} &= \varphi u + b, \end{aligned} \tag{70}$$

where b is the constant of the integration, to be determined by the other boundary condition. Solving the above equation, $x(u)$ can be written as

$$x(u) = \frac{Ae^{\varphi u}}{1 + \frac{A(K-1)p}{\varphi} e^{\varphi u}}, \tag{71}$$

where $A = e^b$. Using (68) and (71), we have

$$X(u) = \frac{1}{(K-1)p} \log\left(1 + \frac{A(K-1)p}{\varphi} e^{\varphi u}\right). \tag{72}$$

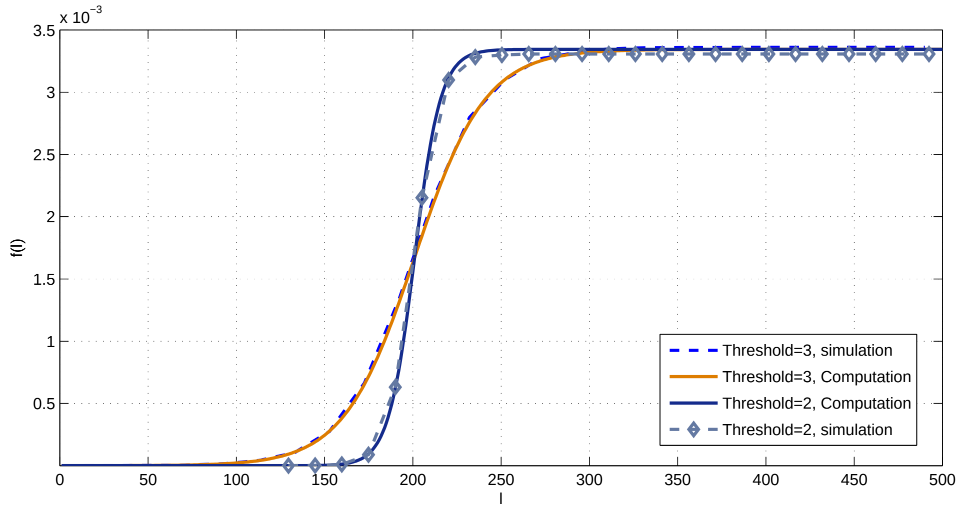

Fig. 2. $f_\nu(l)$ for the proposed method with $\Theta = 2, 3$; $D = 500, K = 300$, comparison between simulation and computation.

Applying the condition $X(D_0) = 1 - \frac{D-D_0}{K}$ yields

$$\begin{aligned} A &= \frac{\varphi}{(K-1)p}\left[e^{(K-1)p(1-\frac{D-D_0}{K})} - 1\right] e^{-\varphi D_0} \\ &\approx \frac{\varphi}{(K-1)p} e^{(K-1)p - \varphi D_0}, \end{aligned} \tag{73}$$

where the second line comes from the facts that $(K-1)p \gg 1$ (since $p > \frac{(\log K)^{1.5}}{K}$) and $p(D - D_0) \ll 1$ (since $p < \frac{(\log K)^2}{K}$ and $D - D_0 \sim 9\sqrt{K}(\log K)^4$). Substituting A in (71), we have

$$x(u) \sim \frac{\frac{\varphi}{(K-1)p} e^{(K-1)p} e^{\varphi(u-D_0)}}{1 + e^{(K-1)p} e^{\varphi(u-D_0)}}. \tag{74}$$

One can easily check that $x(D_0) \sim \frac{1}{K}$, which is consistent with (51). Combining (74) with the fact that $f_\nu(l) \sim x(l)$, Lemma 2 easily follows.

■

Although the derived analytical pmf in (74) is valid in the asymptotic regime of $K \to \infty$, figure 2 shows that the analytical expression in (74) indeed works for finite number of users. In this figure, $f_\nu(l)$ is depicted for the proposed scheduling algorithm with the threshold values of 2 and 3, assuming $K = 300$ and $D = 500$. As can be observed, the curves derived by simulation almost follow the curves derived by computation of $f_\nu(l)$ from (74).

Figure 3 shows the plots of $f_\nu(l)$ for different values of threshold Θ. The plots of $f_\nu(l)$ for the Round-Robing scheduling and the maximum-throughput scheduling are also given for comparison. It is observed that as the value of threshold decreases, $f_\nu(l)$ merges to that of Round-Robin scheduling, while by increasing the threshold value, it merges to that of the maximum-throughput scheduling.

Lemma 3 *Setting* $D_0 = \frac{p}{\varphi}(K-1) + \frac{\log K}{\varphi}$, *for some* φ *such that* $\varphi \cong \eta\left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$, *yields Pr*$\{\mathscr{B}\} \to 0$, *while satisfying* $\lim_{K\to\infty} \frac{D}{K} = 1$.

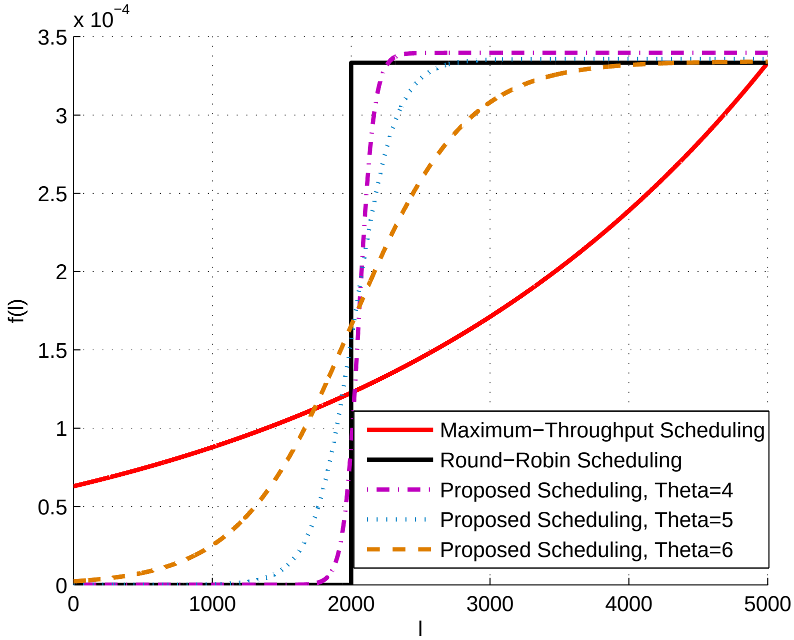

Fig. 3. Comparison of $f_\nu(l)$ for the proposed method with $\Theta = 4, 5, 6$, the Round-Robin scheduling ($\Theta = 0$), and the maximum throughput scheduling; $D = 5000, K = 3000$.

Proof - We have seen earlier that the dropping probability for each user is equal to $F_\nu(0)$. Using the union bound for the probability, it follows that having $F_\nu(0) \sim o(\frac{1}{K})$ guarantees Pr$\{\mathscr{B}\} \to 0$. Using (72) and (73), we have

$$F_\nu(0) \sim X(0) \quad = \quad \frac{1}{(K-1)p} \log \left(1 + e^{(K-1)p - \varphi D_0}\right), \tag{75}$$

for some $\varphi \cong \eta \left[1 + O\left(\frac{1}{\sqrt{K}}\right)\right]$. From the above equation, the condition $F_\nu(0) \sim o\left(\frac{1}{K}\right)$ can be equivalently written as

$$e^{(K-1)p - \varphi D_0} \sim o(p).$$

It can be easily verified that having $D_0 = \frac{p}{\varphi}(K-1) + \frac{\log K}{\varphi}$, results in $e^{(K-1)p} e^{-\varphi D_0} = \frac{1}{K}$, which satisfies the above condition (since $\frac{1}{K} \sim o(p)$). Furthermore, since $\Theta < \log K - 1.5 \log\log K$, it follows that $\varphi \sim \eta \sim p > \frac{[\log K]^{1.5}}{K}$, which incurs that $\frac{\log K}{\varphi} \lesssim \frac{K}{\sqrt{\log K}}$. Combining this with the facts that $\lim_{K \to \infty} \frac{p}{\varphi} = 1$ and $D \cong D_0 + 9\sqrt{K}[\log K]^4$ (which follows from the definition of D_0), we have $\lim_{K \to \infty} \frac{D}{K} = 1$. This completes the proof of Lemma 3.

■

The achievable sum-rate of the proposed algorithm can be lower-bounded as follows:

$$\begin{aligned}\mathcal{R} &= \mathcal{R}_{\mathscr{X}}\Pr\{\mathscr{X}\} + \mathcal{R}_{\mathscr{X}^C}\Pr\{\mathscr{X}^C\} \\ &\geq \mathcal{R}_{\mathscr{X}}\Pr\{\mathscr{X}\} \\ &\overset{(a)}{\geq} \log(1+P\Theta)\Pr\{\mathscr{X}\} \\ &\overset{(39)}{\geq} \log(1+P\Theta)\left[1 - \left|O\left(e^{-(\log K)^{1.5}}\right)\right|\right]. \end{aligned} \tag{76}$$

where $\mathcal{R}_{\mathscr{X}}$ and $\mathcal{R}_{\mathscr{X}^C}$ denote the achievable sum-rate conditioned on $\mathscr{X}$ and $\mathscr{X}^C$, respectively, and $\mathscr{X}^C$ (complement of $\mathscr{X}$) is defined as the event that $|\mathcal{S}| = 0$. In the above equation, (a) follows from the fact that conditioned on $\mathscr{X}$, the channel gain of the selected user is greater than Θ, and hence, the achievable sum-rate is lower-bounded by $\log(1+P\Theta)$.

From the above equation and noting the facts that $\mathcal{C}_{\text{sum}} \sim \log(1 + P\log K + O(\log\log K))$ [24], and $\Theta > \log K - 2\log\log K$, we have

$$\begin{aligned}\mathcal{C}_{\text{sum}} - \mathcal{R} &\lesssim O\left(\frac{\log\log K}{\log K}\right) \\ \Rightarrow \lim_{K\to\infty} \mathcal{C}_{\text{sum}} - \mathcal{R} &= 0. \end{aligned} \tag{77}$$

Combining the above equation with Lemma 3 completes the proof of Theorem 4.

■

Remark 1- 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 scheduling algorithm is capable of achieving the maximum throughput and minimum network delay, simultaneously.

Remark 2- Assume that the information data delivered to the users are put in *packets*, which are stored in the transmitter buffer and each packet is mapped to a coded frame, consisting of n channel uses, and transmitted over the channel (Fig. 4). Assume that the Packet Arrival Rate (PAR) for user k to be fixed and equal to r_k (measured as the number of arrived packets per unit time, i.e., one frame duration) and the amount of information in each packet of that user to be $n\mathcal{R}_k$. In order to have arbitrary small outage probability, $\mathcal{R}_k$, $k = 1, \cdots, K$, must be inside the capacity region of the underlying broadcast channel, which implies that $\mathcal{R}_k \leq \mathcal{C}_{\text{sum}}$, $\forall k$. Moreover, in order to have arbitrarily small dropping probability in the network, the vector consisting of the PAR of the users, denoted by $\mathbf{r} = (r_1, \cdots, r_K)$, must be inside the *stability region* of the network [25]. More specifically, for $r_1 = r_2 = \cdots = r_K = r$, this condition reduces to $r \leq \frac{1}{K}$ [4]. From this discussion, it follows that the maximum r and $\mathcal{R}_k$, $k = 1, \cdots, K$, in order not to have any dropping or outage in the network scale as $\frac{1}{K}$ and $\mathcal{C}_{\text{sum}}$, respectively. The above theorem states that the proposed scheduling is capable of achieving the maximum values of r and $\mathcal{R}_k$, $k = 1, \cdots, K$, simultaneously. In other words, the proposed algorithm reaches the boundary of the *capacity region* and

[4]Note that this is based on the assumption that at each frame, only one user is served.

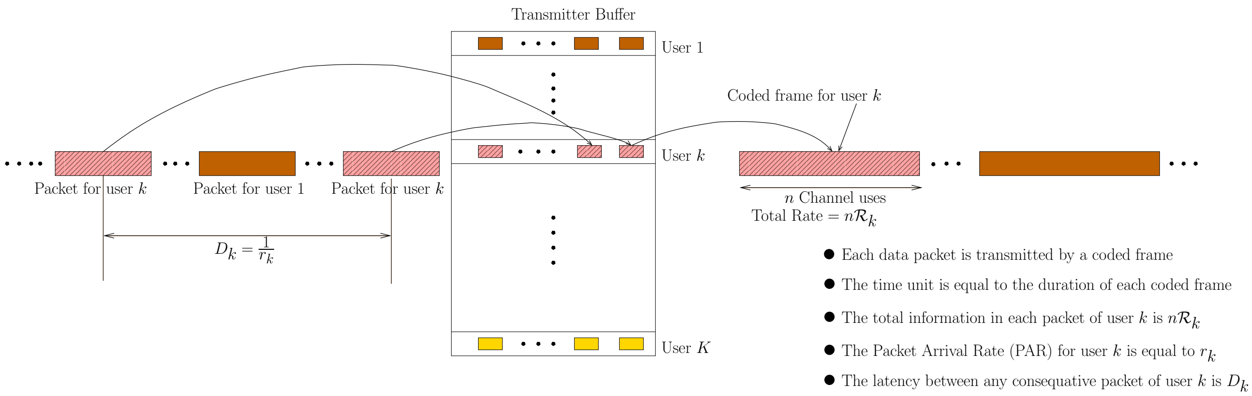

Fig. 4. A Schematic figure for the transmission of data packets over the broadcast channel

stability region of the network, simultaneously. The following corollary illustrates this fact from a different perspective:

Corollary 1 *Consider a Broadcast system illustrated in Fig. 4, where the transmitter has the buffer size of one packet for each user and the Packet Arrival Rate (PAR) for the kth user is r_k and the amount of information in each packet for user k is $n\mathcal{R}_k$. Let us define the "average throughput" of user k (normalized per channel use) as* [5]

$$\mathscr{T}_k \triangleq r_k \mathcal{R}_k. \tag{78}$$

Then, for any scheduling scheme, any rate vector $\mathbf{R} = (\mathcal{R}_1, \cdots, \mathcal{R}_K)$ supported by the channel (decoding error approaches zero), and for any PAR vector $\mathbf{r} = (r_1, \cdots, r_K)$, the necessary condition for Pr$\{\mathscr{B}\} \to 0$ is having

$$\mathscr{T}_{\min} \triangleq \min_k \mathscr{T}_k \quad \lesssim \quad \frac{\log \log K}{K}, \tag{79}$$

which is achievable by the proposed algorithm.

Proof - *Necessary Condition* - Consider a long interval of time T. Defining $\mathcal{A}_k(t)$ as the indicator variable taking one when the user k is served during the frame t, and taking zero otherwise, we have

$$\sum_{k=1}^{K} \mathcal{A}_k(t)\mathcal{R}_k \leq \mathcal{C}_{\text{sum}}, \quad \forall t, 1 \leq t \leq T. \tag{80}$$

The above equation comes from the fact that the rates $(\mathcal{R}_1, \cdots, \mathcal{R}_K)$ must be supported by the channel. Taking the summation with respect to t, we can write

$$\sum_{t=1}^{T} \sum_{k=1}^{K} \mathcal{A}_k(t)\mathcal{R}_k \leq \mathcal{C}_{\text{sum}} T. \tag{81}$$

[5]This definition is motivated by the fact that there is a time delay of $\frac{1}{r_k}$ between two consecutive packets of user k, and as a result, the average amount of information per channel use delivered to user k is equal to $r_k\mathcal{R}_k$.

Since $\Pr\{\mathscr{B}\} \rightarrow 0$, the arrival rate of the packets must be less than or equal to their service rate, over a long period of time, almost surely. In other words, $\sum_{t=1}^{T} \mathcal{A}_k(t) \gtrsim Tr_k$, $\forall k, 1 \leq k \leq K$. Substituting in the above equation yields

$$\sum_{k=1}^{K} \mathscr{T}_k = \sum_{k=1}^{K} r_k \mathcal{R}_k \lesssim \mathcal{C}_{\text{sum}} \stackrel{(a)}{\sim} \log(P \log K), \tag{82}$$

where (a) comes from [24]. Combining (78) and (82), yields

$$\begin{aligned} \mathscr{T}_{\min} &\leq \frac{\sum_{k=1}^{K} \mathscr{T}_k}{K} \\ &\lesssim \frac{\log\log K}{K} + \frac{\log P}{K} \\ &\sim \frac{\log\log K}{K}. \end{aligned} \tag{83}$$

Sufficient Condition - Consider the proposed algorithm, with the condition of Theorem 4, i.e., $\log K - 2\log\log K < \Theta < \log K - 1.5\log\log K$. It is realized from Lemma 3 that selecting $r_k = \frac{1}{D}$ for all users, where D is obtained as follows:

$$D = \frac{p}{\varphi}(K-1) + \frac{\log K}{\varphi} + 9\sqrt{K}[\log K]^4,$$

guarantees $\Pr\{\mathscr{B}\} \rightarrow 0$. Furthermore, the channel can support the rate

$$\mathcal{R}_k = \log\left[1 + P(\log K - 2\log\log K)\right],$$

with probability $\Pr\{\mathscr{X}\}$ (which is almost equal to 1 from (39)), for all users. Hence,

$$\begin{aligned} \mathscr{T}_{\min} &\geq \frac{\log\left[1 + P(\log K - 2\log\log K)\right]}{D} \\ &\sim \frac{\log\log K}{K}. \end{aligned} \tag{84}$$

■

In the above corollary, the *minimum average throughput*, denoted by $\mathscr{T}_{\min}$, is defined as the measure of performance. The average throughput itself can be interpreted as the average amount of information (per channel use) delivered to a user over a long period of time. This measure is suitable for the real-time applications, where the packets have certain amount of information 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 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 more relaxed assumptions, as well.

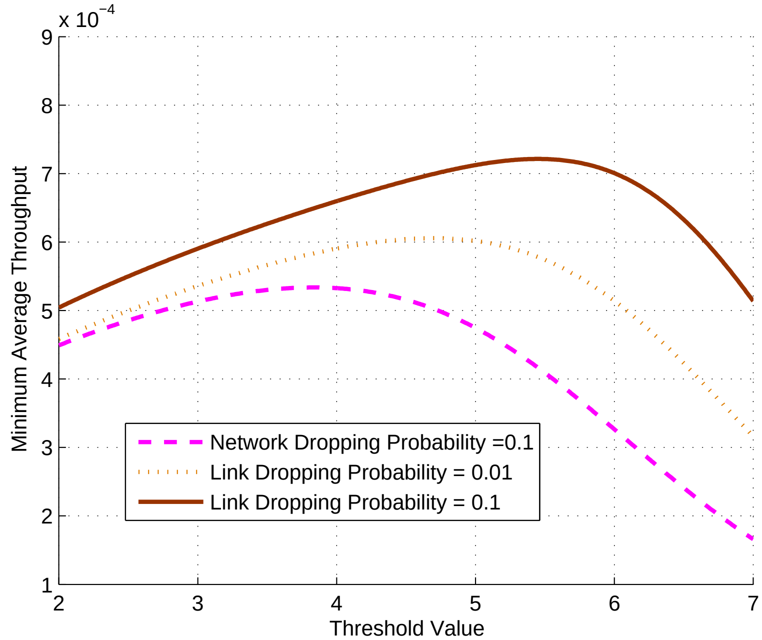

Fig. 5. *Minimum average throughput* vs. the threshold value, $K = 3000$, $P = 0dB$.

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

$$\mathscr{T}_{\min}^{MT} \sim \frac{\log\log K}{K\log K},$$
$$\mathscr{T}_{\min}^{RR} \sim \frac{1}{K}. \tag{85}$$

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

The above corollary states that the proposed scheduling scheme maximizes the *minimum average throughput* of the system while making the network dropping probability arbitrarily small in the asymptotic regime of $K \to \infty$, for all the threshold values in the interval $[\log K - 2\log\log K, \log K - 1.5\log\log K]$. However, for finite number of users, it is not possible to simultaneously maximize the *minimum average throughput* and make the network dropping probability zero. In fact, for a given constraint on the dropping probability, the *minimum average throughput* will be a function of the threshold value, which is desired to be maximized. Figure 5 shows the plots of the *minimum average throughput* versus the threshold value, for different assumptions on the link and network dropping probabilities. The number of users K is set to 3000 and the SNR value P is set to $0\,dB$. As can be observed, for each plot, there is an optimum threshold value for which the *minimum average throughput* is maximized. Moreover, by making the constraint on the dropping probability more restrictive, the optimum threshold value decreases.

IV. EXTENSION TO THE MIMO-BC

So far, we have assumed that the transmitter and the receivers are all equipped with single antennas. In this section, we assume that the transmitter has M antennas, while the receivers have single antennas. The main difference between this case and the previous case is that for SISO-BC, serving one user at each time (TDMA) is optimal in terms of achieving the maximum throughput of the system [23], while in the MIMO-BC, this is not the case. Therefore, we must apply some modifications to our proposed algorithm, to make it suitable for MIMO-BC.

A. System Model and Proposed Algorithm

The channel model for the kth user is assumed to be

$$y_k = \mathbf{h}_k \mathbf{x} + n_k, \tag{86}$$

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.

Algorithm 2:

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

$$\mathcal{S}_m = \{k| \quad \text{SINR}_k^{(m)} > \Upsilon\}, \quad m = 1, \cdots, M, \tag{87}$$

   the BS serves the user with the minimum *expiry countdown*. In the above equation, $\text{SINR}_k^{(m)} \triangleq \frac{\frac{P}{M}|\mathbf{h}_k\mathbf{\Phi}_m^H|^2}{1+\sum_{j\neq m}\frac{P}{M}|\mathbf{h}_k\mathbf{\Phi}_j^H|^2}$ 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 [24], where the *expiry countdown* is considered in the scheduling.

B. Asymptotic Analysis

In this section, we analyze the performance of the proposed algorithm in the asymptotic case of $K \to \infty$. Similar to the SISO case, it is interesting to investigate the possibility of achieving the maximum throughput and fairness of the system, simultaneously, which is performed in the following theorem:

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

$$\frac{P}{M}\left[\log K - (M+1)\log\log K\right] < \Upsilon < \frac{P}{M}\left[\log K - (M+0.5)\log\log K\right], \tag{88}$$

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

Proof - Using the same approach as in the proof of Theorem 4, we first derive $f_\nu(\nu)$ in terms of K, D, and Υ. Consider the following sets:

$$\mathcal{S}'_m \triangleq \left\{k \,\middle|\, k\in\mathcal{A}_m,\ \text{SINR}_k^{(m)}>\Upsilon\right\},\quad m=1,\cdots,M, \tag{89}$$

where $\mathcal{A}_m \triangleq \{k|\ |\mathbf{h}_k\mathbf{\Phi}_m^H|^2 > |\mathbf{h}_k\mathbf{\Phi}_j^H|^2, \forall j\neq m\}$. For simplicity of analysis, we assume that the step 3 of Algorithm 2 works based on $\mathcal{S}'_m$ instead of $\mathcal{S}_m$. It is obvious that $\mathcal{S}'_m\subset\mathcal{S}_m$. However, since $\sum_{m=1}^{M}|\mathbf{h}_k\mathbf{\Phi}_m^H|^2=\|\mathbf{h}_k\|^2<\log K+O(\log\log K)$, with probability one [24], it follows that having $\text{SINR}_k^{(m)}>\Upsilon$, where $\Upsilon\sim\beta\frac{P}{M}\log K$ and $\beta>\frac{1}{2}$, yields $k\in\mathcal{A}_m$. This implies that for the values of Υ satisfying (88), we have $\mathcal{S}'_m=\mathcal{S}_m$, with probability one. Rewriting (8), we have

$$f_\nu(l-1)=f_\nu(l)\left(1-\Pr\{\mathscr{X}_k|\nu_k=l\}\right). \tag{90}$$

$\Pr\{\mathscr{X}_k|\nu_k=l\}$ can be written as follows:

$$\begin{aligned}\Pr\{\mathscr{X}_k|\nu_k=l\} &\overset{(a)}{=} \Pr\{\mathscr{X}_k,k\in\mathcal{S}'|\nu_k=l\}\\ &= \sum_{m=1}^{M}\Pr\{\mathscr{X}_k,k\in\mathcal{S}'|\nu_k=l,\mathscr{F}_m\}\Pr\{\mathscr{F}_m|\nu_k=l\}\\ &\overset{(b)}{=} \sum_{m=1}^{M}\Pr\{\mathscr{X}_k,k\in\mathcal{S}'|\nu_k=l,\mathscr{F}_m\}\Pr\{\mathscr{F}_m\}\\ &\overset{(c)}{=} \frac{1}{M}\sum_{m=1}^{M}\Pr\{\mathscr{X}_k,k\in\mathcal{S}'_m|\nu_k=l,\mathscr{F}_m\}\\ &\overset{(d)}{=} \Pr\{\mathscr{X}_k,k\in\mathcal{S}'_m|\nu_k=l,\mathscr{F}_m\},\end{aligned} \tag{91}$$

where $\mathcal{S}'\triangleq\bigcup_{m=1}^{M}\mathcal{S}'_m$, and $\mathscr{F}_m\triangleq\{k\in\mathcal{A}_m\}$. In the above equation, (a) results from the fact that $\mathscr{X}_k\subseteq(k\in\mathcal{S}')$, in order words, the necessary condition for user k to be served is being in any of the sets $\mathcal{S}'_m$, $s=1,\cdots,M$. (b) results from the independence of the events $\nu_k=l$ and $\mathscr{F}_m$ [6]. (c) follows from the fact that conditioned on $\mathscr{F}_m$, i.e. $k\in\mathcal{A}_m$, $k\in\mathcal{S}'$ incurs $k\in\mathcal{S}'_m$, and also the fact that $\Pr\{\mathscr{F}_m\}=\frac{1}{M}$. (d) follows from the symmetry between the terms $\Pr\{\mathscr{X}_k,k\in\mathcal{S}'_m|\nu_k=l,\mathscr{F}_m\}$, $m=1,\cdots,M$.

We have

$$\begin{aligned}\Pr\{\mathscr{X}_k,k\in\mathcal{S}'_m|\nu_k=l,\mathscr{F}_m\} &\overset{(a)}{=} \sum_{n=1}^{K}\sum_{s=n}^{K}\Pr\{\mathscr{X}_k,k\in\mathcal{S}'_m,|\mathcal{S}'_m|=n,|\mathcal{A}_m|=s\ \big|\ \nu_k=l,\mathscr{F}_m\}\\ &\overset{(b)}{=} \sum_{n=1}^{K}\sum_{s=n}^{K}\Pr\{|\mathcal{A}_m|=s|\mathscr{F}_m\}\Pr\left\{k\in\mathcal{S}'_m,|\mathcal{S}'_m|=n\ \big|\ |\mathcal{A}_m|=s,\mathscr{F}_m\right\}\\ &\times\ \Pr\{\mathscr{X}_k|\nu_k=l,|\mathcal{S}'_m|=n,|\mathcal{A}_m|=s,k\in\mathcal{S}'_m\}\end{aligned} \tag{92}$$

[6]In fact, $\mathscr{F}_m(t)$ is a function of $\{\mathbf{h}_k(t)\}_{k=1}^{K}$, while the event $\nu_k(t)=l$ is a function of $\{\mathbf{h}_k(j)\}_{k=1}^{K}$, $j<t$. Since the channel model is assumed to be independent block fading, the independence of $\nu_k=l$ and $\mathscr{F}_m$ easily follows.

In the above equation, (a) follows from the fact that $\mathcal{S}'_m \subset \mathcal{A}_m$, and hence $s \geq n$. (b) results from the facts that the events $|\mathcal{A}_m| = s$ and $k \in \mathcal{S}'_m$ are independent of $\nu_k = l$ (As explained in the footnote), and $k \in \mathcal{S}'_m$ is a subset of $\mathscr{F}_m$.

$\Pr\{|\mathcal{A}_m| = s|\mathscr{F}_m\}$ can be computed as

$$\begin{aligned}\Pr\{|\mathcal{A}_m| = s|\mathscr{F}_m\} &= \frac{\Pr\{|\mathcal{A}_m| = s, k \in \mathcal{A}_m\}}{\Pr\{k \in \mathcal{A}_m\}} \\ &\overset{(a)}{=} M\binom{K-1}{s-1}\left(\frac{1}{M}\right)^s\left(\frac{M-1}{M}\right)^{K-s}, \end{aligned} \tag{93}$$

where (a) follows from the facts that $\Pr\{k \in \mathcal{A}_m\} = \frac{1}{M}$, and $|\mathcal{A}_m|$ is a Binomial random variable with parameters $(K, \frac{1}{M})$. In order to compute $\Pr\left\{k \in \mathcal{S}'_m, |\mathcal{S}'_m| = n \big| |\mathcal{A}_m| = s, \mathscr{F}_m\right\}$, we first compute $q \triangleq \Pr\{k \in \mathcal{S}'_m|\mathscr{F}_m\}$ as follows:

$$\begin{aligned} q &= \frac{\Pr\{k \in \mathcal{S}'_m, k \in \mathcal{A}_m\}}{\Pr\{k \in \mathcal{A}_m\}} \\ &\overset{(a)}{=} \frac{\Pr\{k \in \mathcal{S}_m, k \in \mathcal{A}_m\}}{\Pr\{k \in \mathcal{A}_m\}} \\ &= Mp\Pr\{k \in \mathcal{A}_m|k \in \mathcal{S}_m\}, \end{aligned} \tag{94}$$

where $p \triangleq \Pr\{k \in \mathcal{S}_m\} = \frac{e^{-\frac{M\Upsilon}{P}}}{(1+\Upsilon)^{M-1}}$ [24]. In the above equation, (a) results from the fact that $(k \in \mathcal{S}'_m) = (k \in \mathcal{S}_m)\bigcap(k \in \mathcal{A}_m)$. Note that as $\Pr\{k \in \mathcal{A}_m|k \in \mathcal{S}_m\} \approx 1$, it follows that $q \approx Mp$. Having q from the above equation, we can write

$$\Pr\left\{k \in \mathcal{S}'_m, |\mathcal{S}'_m| = n \big| |\mathcal{A}_m| = s, \mathscr{F}_m\right\} = \binom{s-1}{n-1} q^n (1-q)^{s-n}. \tag{95}$$

Substituting $\Pr\{|\mathcal{A}_m| = s|\mathscr{F}_m\}$ and $\Pr\left\{k \in \mathcal{S}'_m, |\mathcal{S}'_m| = n \big| |\mathcal{A}_m| = s, \mathscr{F}_m\right\}$ from (94) and (95), and noting that conditioned on $|\mathcal{S}'_m| = n$, $\mathscr{X}_k$ is independent of $|\mathcal{A}_m| = s$, RH (92) can be written as

$$\begin{aligned}\text{RH}(92) &= \sum_{n=1}^{K}\sum_{s=n}^{K} M\binom{K-1}{s-1}\left(\frac{1}{M}\right)^s\left(\frac{M-1}{M}\right)^{K-s}\binom{s-1}{n-1}q^n(1-q)^{s-n} \times \\ &\quad \Pr\{\mathscr{X}_k|\nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\} \\ &= M\left(\frac{M-1}{M}\right)^K \sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{1-q}\right)^n \Pr\{\mathscr{X}_k|\nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\} \times \\ &\quad \sum_{s=n}^{K}\binom{K-n}{s-n}\left(\frac{1-q}{M-1}\right)^s \\ &= M\left(\frac{M-1}{M}\right)^K \sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{1-q}\right)^n \Pr\{\mathscr{X}_k|\nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\} \times \\ &\quad \left(\frac{1-q}{M-1}\right)^n\left[1 + \frac{1-q}{M-1}\right]^{K-n} \\ &= M\sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\Pr\{\mathscr{X}_k|\nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\}. \end{aligned} \tag{96}$$

As can be observed, the above equation is very similar to (33), and by a similar argument we can show that

$$\Pr\{\nu_i > l, i = 1, \cdots, n, i \neq k | \nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\} \leq \Pr\{\mathscr{X}_k \mid \nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\}$$
$$\leq \Pr\{\nu_i \geq l, i = 1, \cdots, n, i \neq k \mid \nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\}. \quad (97)$$

In the above equation, the first inequality results from the fact that having $\nu_i > l$, $i \neq k$, implies that the kth user has the minimum *expiry countdown* among $\mathcal{S}'_m$, and hence, will be selected. The second inequality follows from the fact that the kth user must have the minimum expiry countdown in $\mathcal{S}'_m$ in order to be selected, i.e., no user in $\mathcal{S}'_m$ should have a smaller *expiry countdown*. Noting the symmetry of the problem with respect to the users and the fact that the events $\nu_i > l$ (or $\nu_i \geq l$) are independent of $|\mathcal{S}'_m| = n$ and $k \in \mathcal{S}'_m$, the upper bound can be written as $\Pr\{\nu_1 \geq l, \cdots, \nu_{n-1} \geq l | \nu_n = l\}$, which is by the chain rule equal to

$$\Pr\{\nu_1 \geq l, \cdots, \nu_{n-1} \geq l | \nu_n = l\} = \Pr\{\nu_1 \geq l | \nu_n = l\} \times \prod_{i=2}^{n-1} \Pr\{\nu_i \geq l | \nu_1 \geq l, \cdots, \nu_{i-1} \geq l, \nu_n = l\}. \quad (98)$$

Consider the following probability:

$$\Pr\{\nu_i = l_1 | \nu_j = l_2\}, \quad i \neq j. \quad (99)$$

For $l_1 = l_2$, the above probability can be upper-bounded as

$$\Pr\{\nu_i = l_1 | \nu_j = l_1\} \leq f_\nu(l_1). \quad (100)$$

The above inequality comes from the fact that $\Pr\{\nu_i = l_1, \nu_j = l_1\} \leq \Pr^2\{\nu_i = l_1\} = f^2_\nu(l)$, which is shown in Appendix A. A brief explanation of this would be, there are $M(M-1)$ possibilities for the users i and j to be selected in the same frame (since there are M possibilities for assigning each of them to any of the beams and they can not be assigned to the same beam), while in the term $\Pr^2\{\nu_i = l_1\}$ all the M^2 possibilities are encountered.

Also, for $l_1 \neq l_2$, we have

$$f_\nu(l_1) \leq \Pr\{\nu_i = l_1 | \nu_j = l_2\} \leq \frac{f_\nu(l_1)}{1 - f_\nu(l_2)}. \quad (101)$$

To prove the above equation, first we note that the ratio $\frac{\Pr\{\nu_i = l_1 | \nu_j = l_2\}}{f_\nu(l_1)}$ is the same for all $l_1 \neq l_2$. In other words, the condition $\nu_j = l_2$ scales the probabilities of the outcomes $\nu_i = l_1$ by the same value for $l_1 \neq l_2$

in the conditional sample space. To establish (101), let us denote $x \triangleq \frac{\Pr\{\nu_i=l_1|\nu_j=l_2\}}{f_\nu(l_1)}, l_1 \neq l_2$. We have

$$\begin{aligned} \sum_{u\neq l_2} \Pr\{\nu_i = u|\nu_j = l_2\} + \Pr\{\nu_i = l_2|\nu_j = l_2\} &= 1. \\ \Rightarrow \sum_{u\neq l_2} f_\nu(u)x + \Pr\{\nu_i = l_2|\nu_j = l_2\} &= 1 \\ \Rightarrow x = \frac{1 - \Pr\{\nu_i = l_2|\nu_j = l_2\}}{1 - f_\nu(l_2)}. \end{aligned} \quad (102)$$

Therefore,

$$\begin{aligned} \Pr\{\nu_i = l_1|\nu_j = l_2\} &= f_\nu(l_1)x \\ &= \frac{f_\nu(l_1)\left[1 - \Pr\{\nu_i = l_2|\nu_j = l_2]\right\}}{1 - f_\nu(l_2)}. \end{aligned} \quad (103)$$

Using (100) and the fact that $\Pr\{\nu_i = l_2|\nu_j = l_2\} \geq 0$, (101) easily follows.

Using (100) and (101), the upper-bound in (98) can be further upper-bounded as

$$
\begin{aligned}
\Pr\{\nu_i \geq l | \nu_1 \geq l, \cdots, \nu_{i-1} \geq l, \nu_n = l\} &= \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \Pr\{\nu_i \geq l \,|\, \nu_1 = a_1, \cdots, \nu_{i-1} = a_{i-1}, \nu_n = l\} \\
&= \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \left[\Pr\{\nu_i \geq l, \mathscr{Y} \,|\, \mathscr{Q}\} + \Pr\{\nu_i \geq l, \mathscr{Y}^C \,|\, \mathscr{Q}\}\right] \\
&= \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \Big[\Pr\{\mathscr{Y} \,|\, \mathscr{Q}\}\Pr\{\nu_i \geq l \,|\, \mathscr{Y}, \mathscr{Q}\} + \\
&\quad \Pr\{\mathscr{Y}^C \,|\, \mathscr{Q}\}\Pr\{\nu_i \geq l \,|\, \mathscr{Y}^C, \mathscr{Q}\}\Big] \\
&\leq \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \Big[\Pr\{\mathscr{Y} \,|\, \mathscr{Q}\} + \Pr\{\nu_i \geq l \,|\, \mathscr{Y}^C, \mathscr{Q}\}\Big] \\
&\overset{(a)}{\leq} \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \left[\sum_{k=1}^{i-1} \Pr\{\nu_i = a_k\} + \Pr\{\nu_i = l\} + \Pr\{\nu_i \geq l \,|\, \mathscr{Y}^C, \mathscr{Q}\}\right] \\
&\overset{(b)}{\leq} \sum_{\substack{(a_1,\cdots,a_{i-1}) \\ l \leq a_1 \leq D, \cdots, l \leq a_{i-1} \leq D}} \frac{f_{\nu_1,\cdots,\nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 \geq l, \cdots, \nu_{i-1} \geq l | \nu_n = l\}} \times \\
&\quad \left[\sum_{k=1}^{i-1} f_\nu(a_k) + f_\nu(l) + \frac{\Pr\{\nu_i \geq l\} - \sum_{k=1}^{i-1} f_\nu(a_k) - f_\nu(l)}{1 - \sum_{k=1}^{i-1} f_\nu(a_k) - f_\nu(l)}\right] \\
&\overset{(c)}{\leq} \frac{Mi}{K} + G_\nu(l-1).
\end{aligned} \tag{104}
$$

where $\mathscr{Y} \triangleq \bigcup_{k=1}^{i-1}\{\nu_i = \nu_k\} \bigcup \{\nu_i = \nu_n\}$ and $\mathscr{Q} \triangleq \{\nu_1 = a_1, \cdots, \nu_{i-1} = a_{i-1}, \nu_n = l\}$. In the above equation, (a) results from (100), which incurs that $\Pr\{\mathscr{Y}|\mathscr{Q}\} \leq \sum_{k=1}^{i-1} \Pr\{\nu_i = a_k\} + \Pr\{\nu_i = l\} = \sum_{k=1}^{i-1} f_\nu(a_k) + f_\nu(l)$, (b) results from (101), noting that conditioned on $\mathscr{Y}^C, \mathscr{Q}$, the points $a_1, \cdots, a_{i-1}, l$ are excluded from the sample space. (c) results from: (i) upper-bounding $f_\nu(a_k)$, $k = 1, \cdots, i-1$, and $f_\nu(l)$ by $\frac{M}{K}$, which is due to the facts that $f_\nu(l) \leq f_\nu(D)$ and $f_\nu(D) = \Pr\{\mathscr{X}_k\} \leq \frac{M}{K}$, where $\Pr\{\mathscr{X}_k\}$ is the probability that user k is being selected in a frame [7], and (ii) upper-bounding $\frac{\Pr\{\nu_i \geq l\} - \sum_{k=1}^{i-1} f_\nu(a_k) - f_\nu(l)}{1 - \sum_{k=1}^{i-1} f_\nu(a_k) - f_\nu(l)}$

[7] In fact, $\Pr\{\mathscr{X}_k\} \leq \frac{M}{K}$ follows from the union bound on the probability. More precisely, denoting $\mathscr{X}_k^{(m)}$ as the event that user k is assigned to beam m, using the same argument as in the SISO case, one can show that $\Pr\{\mathscr{X}_k^{(m)}\} \leq \frac{1}{K}$, and hence, $\Pr\{\mathscr{X}_k\} = \Pr\{\bigcup_{m=1}^{M} \mathscr{X}_k^{(m)}\} \leq \frac{M}{K}$.

by $\Pr\{\nu_i \geq l\} = G_\nu(l-1)$.

Using the above equation and (98), the upper bound in (97) can be upper-bounded as

$$\Pr\{\nu_1 \geq l, \cdots, \nu_{n-1} \geq l | \nu_n = l\} \leq \prod_{i=1}^{n-1} \left(G_\nu(l-1) + \frac{Mi}{K} \right). \tag{105}$$

Moreover, to lower-bound the lower bound in (97), we first lower-bound $\Pr\{\nu_i > l | \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\}$ as follows:

$$\begin{aligned}
\Pr\{\nu_i > l | \nu_1 > l, \cdots, \nu_{i-1} > l, \nu_n = l\} &\geq \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad \Pr\{\nu_i > l \mid \nu_1 = a_1, \cdots, \nu_{i-1} = a_{i-1}, \nu_n = l\} \\
&= \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad \Big[\Pr\{\mathscr{Y} \mid \mathscr{Q}\} \Pr\{\nu_i > l \mid \mathscr{Y}, \mathscr{Q}\} + \\
&\quad \Pr\{\mathscr{Y}^C \mid \mathscr{Q}\} \Pr\{\nu_i > l \mid \mathscr{Y}^C, \mathscr{Q}\} \Big] \\
&\geq \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad \Pr\{\mathscr{Y}^C \mid \mathscr{Q}\} \Pr\{\nu_i > l \mid \mathscr{Y}^C, \mathscr{Q}\} \\
&\overset{(a)}{\geq} \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad \Pr\{\mathscr{Y}^C \mid \mathscr{Q}\} \Pr\{\nu_i > l\} \\
&= \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad (1 - \Pr\{\mathscr{Y} \mid \mathscr{Q}\}) \Pr\{\nu_i > l\} \\
&\overset{(b)}{\geq} \sum_{\substack{(a_1, \cdots, a_{i-1}) \\ l < a_1 \leq D, \cdots, l < a_{i-1} \leq D}} \frac{f_{\nu_1, \cdots, \nu_{i-1}}(a_1, \cdots, a_{i-1} | \nu_n = l)}{\Pr\{\nu_1 > l, \cdots, \nu_{i-1} > l | \nu_n = l\}} \times \\
&\quad \left(1 - \sum_{k=1}^{i-1} f_\nu(a_k) - f_\nu(l) \right) \Pr\{\nu_i > l\} \\
&\overset{(c)}{\geq} G_\nu(l) - \frac{Mi}{K},
\end{aligned} \tag{106}$$

where (a) results from (101) which implies that $\Pr\{\nu_i > l \mid \mathscr{Y}^C, \mathscr{Q}\} \geq \Pr\{\nu_i > l\}$, (b) follows from (100), which incurs that $\Pr\{\mathscr{Y} | \mathscr{Q}\} \leq \sum_{k=1}^{i-1} f_\nu(a_k) + f_\nu(l)$. Finally, (c) results from the fact that $f_\nu(\nu) \leq \frac{M}{K}$, and writing $\Pr\{\nu_i > l\}$ as $G_\nu(l)$.

Using the above equation, the lower-bound in (97) can be lower-bounded as

$$\Pr\{\nu_1 > l, \cdots, \nu_{n-1} > l | \nu_n = l\} \geq \prod_{i=1}^{n-1} \left(G_\nu(l) - \frac{Mi}{K} \right). \quad (107)$$

Similar to the approach used in the SISO case, by defining $n_0 = 3(\log K)^2$ and $D_0 = D - \sqrt{K} n_0 (n_0 - 1)$, first we show that for $D_0 \leq l \leq D$, we have $f_\nu(l) \sim \frac{M}{K}$. For this purpose, by repeated application of (90), and using (91), (92), (96), (97), and (105), we have

$$f_\nu(D) - f_\nu(D_0) \leq \sum_{l=D_0+1}^{D} \mathcal{W}_l, \quad (108)$$

where $\mathcal{W}_l \triangleq M \sum_{n=1}^{K} \binom{K-1}{n-1} (\frac{q}{M})^n \left(1 - \frac{q}{M}\right)^{K-n} \prod_{i=1}^{n-1} \left(G_\nu(l-1) + \frac{Mi}{K}\right)$. In Appendix B, it has been shown that $\mathcal{W}_l$ is upper-bounded as $M \frac{(\log K)^2}{K} e^{-(\log K)^{1.5}}$, which implies that

$$\begin{aligned} f_\nu(D) - f_\nu(D_0) &\leq (D - D_0) M \frac{(\log K)^2}{K} e^{-(\log K)^{1.5}} \\ &\sim 9M \frac{(\log K)^6}{\sqrt{K}} e^{-(\log K)^{1.5}} \\ &\sim o\left(e^{-(\log K)^{1.5}}\right). \end{aligned} \quad (109)$$

Moreover, $f_\nu(D)$ can be written as $\Pr\{\mathscr{X}_k\}$ [8], which denotes the probability that user k is selected in a frame. This probability can be expressed as $\Pr\{\bigcup_{m=1}^{M} \mathscr{X}_k^{(m)}\}$, where $\mathscr{X}_k^{(m)}$ denotes the event that the kth user is assigned to the mth beam. Defining $\mathscr{X}^{(m)} \triangleq \bigcup_{k=1}^{K} \mathscr{X}_k^{(m)}$, which is the probability that the mth beam is assigned to some user, we have

$$\begin{aligned} \Pr\{\mathscr{X}^{(m)}\} &= 1 - \Pr\{|\mathcal{S}'_m| = 0\} \\ &= 1 - (1 - \Pr\{k \in \mathcal{S}'_m\})^K \\ &\overset{(a)}{=} 1 - \left(1 - \frac{q}{M}\right)^K \\ &\cong 1 - e^{-Kq/M} \\ &\overset{(b)}{\geq} 1 - e^{-(\log K)^{1.5}}, \end{aligned} \quad (110)$$

where (a) follows from the definition of q in (94), and (b) results from the fact that $\frac{q}{M} \sim p > \frac{(\log K)^{1.5}}{K}$, Having the fact that the events $\mathscr{X}_k^{(m)}$, $k = 1, \cdots, K$ are mutually exclusive, i.e., beams can not be assigned to multiple users simultaneously, we have

$$\begin{aligned} \Pr\{\mathscr{X}^{(m)}\} = \sum_{k=1}^{K} \Pr\{\mathscr{X}_k^{(m)}\} &\geq 1 - e^{-(\log K)^{1.5}} \\ \Rightarrow \Pr\{\mathscr{X}_k^{(m)}\} &\geq \frac{1}{K} \left(1 - e^{-(\log K)^{1.5}}\right), \end{aligned} \quad (111)$$

[8]More precisely, $f_{\nu_k(t)}(D) = \Pr\{\mathscr{X}_k(t-1)\}$, where the the time index are removed due to the steady state condition.

where the second line results from the symmetry between the users. Moreover, since the sets $\mathcal{S}'_m$, $m = 1, \cdots, M$ are disjoint, it follows that the events $\mathscr{X}_k^{(m)}$, $m = 1, \cdots, M$ are mutually exclusive. Therefore, using the above equation,

$$\Pr\{\mathscr{X}_k\} = \sum_{m=1}^{M} \Pr\{\mathscr{X}_k^{(m)}\} \geq \frac{M}{K}\left(1 - e^{-(\log K)^{1.5}}\right). \tag{112}$$

Combining the above equation with (109), it follows that

$$f_\nu(l) \cong \frac{M}{K}\left[1 + o\left(Ke^{-(\log K)^{1.5}}\right)\right], \quad D_0 \leq l \leq D. \tag{113}$$

In other words, in the interval $[D_0, D]$, $f_\nu(l)$ is almost constant.

In the region $l < D_0$, by defining the following functions:

$$g_u(n,l) = \begin{cases} \prod_{i=1}^{n-1}\left(G_\nu(l-1) + \frac{iM}{K}\right), & n \leq n_0 \\ 1 & n > n_0 \end{cases}, \tag{114}$$

and

$$g_l(n,l) = \begin{cases} \prod_{i=1}^{n-1}\left(G_\nu(l) - \frac{iM}{K}\right), & n \leq n_0 \\ 0 & n > n_0 \end{cases}, \tag{115}$$

where $n_0 = 3(\log K)^2$, using the equations (97), (105), and (107), it follows that

$$g_l(n,l) \leq \Pr\{\mathscr{X}_k | \nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\} \leq g_u(n,l), \tag{116}$$

where $\Pr\{\mathscr{X}_k | \nu_k = l, |\mathcal{S}'_m| = n, k \in \mathcal{S}'_m\}$ is the probability we need to find in order to compute $\Pr\{\mathscr{X}_k|\nu_k = l\}$ in (96). From the above equation, $\Pr\{\mathscr{X}_k|\nu_k = l\}$ can be upper-bounded as follows:

$$\begin{aligned}
\Pr\{\mathscr{X}_k|\nu_k = l\} &\leq M\sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n} g_u(n,l) \\
&\overset{(a)}{=} \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n} g_u(n+1,l) \\
&= \eta\sum_{n=0}^{n_0}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\prod_{i=1}^{n}\left(G_\nu(l-1)+\frac{iM}{K}\right) + \\
&\quad \eta\sum_{n_0+1}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n} \\
&= \eta\sum_{n=0}^{n_0}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n} G_\nu(l-1)^n\prod_{i=1}^{n}\left(1+\frac{iM}{KG_\nu(l-1)}\right) + \\
&\quad \eta\sum_{n_0+1}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n},
\end{aligned} \tag{117}$$

where $\eta = \frac{q}{1-\frac{q}{M}}$. In the above equation, (a) results from taking the terms $\frac{\frac{q}{M}}{1-\frac{q}{M}}$ outside the summation and make a change of variable $n-1$ to n. Since $f_\nu(l) \sim \frac{M}{K}$ for $D_0 \leq l \leq D$, it follows that $G_\nu(D_0) \sim$

$\frac{M(D-D_0)}{K} = \frac{Mn_0(n_0-1)}{\sqrt{K}}$, which implies that $G_\nu(l-1) \geq \frac{Mn_0(n_0-1)}{\sqrt{K}}$, for $D_0 \leq l \leq D$. Therefore, the term $\prod_{i=1}^{n}\left(1+\frac{iM}{KG_\nu(l-1)}\right)$ can be written as

$$\begin{aligned}
\prod_{i=1}^{n}\left(1+\frac{iM}{KG_\nu(l-1)}\right) &\leq \prod_{i=1}^{n}\left(1+\frac{i}{\sqrt{K}n_0(n_0-1)}\right)\\
&\overset{(a)}{\approx} 1+\sum_{i=1}^{n}\frac{i}{\sqrt{K}n_0(n_0-1)}\\
&= 1+\frac{n(n+1)}{2\sqrt{K}n_0(n_0-1)}\\
&\overset{(b)}{\sim} 1+O\left(\frac{1}{\sqrt{K}}\right),
\end{aligned} \tag{118}$$

where (a) results from the fact that as $i \leq n_0$, $\frac{i}{\sqrt{K}n_0(n_0-1)} \ll 1$, and (b) follows from $n \leq n_0$. Having the above equation, RH (117) can be written as

$$\begin{aligned}
\text{RH (117)} &\cong \eta\sum_{n=0}^{n_0}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}G_\nu(l-1)^n\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]+\\
&\quad \eta\sum_{n_0+1}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\\
&\leq \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}G_\nu(l-1)^n\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]+\\
&\quad \eta\sum_{n_0+1}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\\
&\overset{(a)}{\cong} \eta\left[1-\frac{q}{M}F_\nu(l-1)\right]^{K-1}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]+\eta Q\left(\frac{n_0-(K-1)\frac{q}{M}}{\sqrt{(K-1)\frac{q}{M}\left(1-\frac{q}{M}\right)}}\right)\\
&\overset{(b)}{\leq} \eta e^{-(K-1)\frac{q}{M}[F_\nu(l)-f_\nu(l)]}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]+\eta e^{-2(K-1)\frac{q}{M}}\\
&\overset{(c)}{\cong} \eta e^{-(K-1)\frac{q}{M}F_\nu(l)}\left[1+O\left(\frac{1}{\sqrt{K}}\right)+e^{-(K-1)\frac{q}{M}}\right]\\
&\overset{(d)}{\cong} \eta e^{-(K-1)\frac{q}{M}F_\nu(l)}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right].
\end{aligned} \tag{119}$$

In the above equation, (a) follows from approximating the tale of the Binomial random variable with the Gaussian $Q(.)$ function. In deriving (b), we first approximate $\left[1-\frac{q}{M}F_\nu(l-1)\right]^{K-1}$ by $e^{-(K-1)\frac{q}{M}F_\nu(l-1)} = e^{-(K-1)\frac{q}{M}[F_\nu(l)-f_\nu(l)]}$, which follows from $q \ll 1$. Using the fact that as $\frac{P}{M}[\log K-(M+1)\log\log K] < \Upsilon < \frac{P}{M}[\log K-(M+0.5)\log\log K]$, we have $\frac{q}{M} < \frac{(\log K)^2}{K}$, which implies that $n_0 > 3(K-1)\frac{q}{M}$, and also the fact that for $x \gg 1$, $Q(x) < e^{-x^2/2}$, $Q\left(\frac{n_0-(K-1)\frac{q}{M}}{\sqrt{(K-1)\frac{q}{M}\left(1-\frac{q}{M}\right)}}\right)$ is upper-bounded as $e^{-2(K-1)\frac{q}{M}}$. (c) results from the facts that: (i) as $f_\nu(l) \leq \frac{M}{K}$, we have $e^{(K-1)\frac{q}{M}f_\nu(l)} \cong 1+O(q) \cong 1+O\left(\frac{1}{\sqrt{K}}\right)$, and

ii) since $F_\nu(l) \le 1$, $e^{-(K-1)\frac{q}{M}F_\nu(l)} \ge e^{-(K-1)\frac{q}{M}}$, and as a result, $e^{-2(K-1)\frac{q}{M}} \le e^{-(K-1)\frac{q}{M}} e^{-2(K-1)\frac{q}{M}F_\nu(l)}$. Finally, (d) follows from the fact that $e^{-(K-1)\frac{q}{M}} \sim o\left(\frac{1}{\sqrt{K}}\right)$, which is due to the fact that $q > M\frac{(\log K)^{1.5}}{K}$.

Similar to (117) and (119), a lower-bound for $\Pr\{\mathscr{X}_k|\nu_k = l\}$ can be given as follows:

$$\begin{aligned}
\Pr\{\mathscr{X}_k|\nu_k = l\} &\ge M\sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n} g_u(n,l) \\
&= \eta\sum_{n=0}^{n_0}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\prod_{i=1}^{n}\left(G_\nu(l)-\frac{iM}{K}\right) \\
&\cong \eta e^{-(K-1)\frac{q}{M}F_\nu(l)}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]. \qquad (120)
\end{aligned}$$

Comparing (119) and (120), it follows that

$$\Pr\{\mathscr{X}_k|\nu_k = l\} \cong \eta e^{-(K-1)\frac{q}{M}F_\nu(l)}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]. \qquad (121)$$

Substituting in (91), we reach the following difference equation in the region $l < D_0$:

$$f_\nu(l) - f_\nu(l-1) \sim \eta f_\nu(l) e^{-(K-1)\frac{q}{M}F_\nu(l)}\left[1+O\left(\frac{1}{\sqrt{K}}\right)\right]. \qquad (122)$$

Comparing the above equation with (30), it is realized that the above difference equation is the same as the difference equation obtained in the SISO case, with the difference in replacing K by $\frac{K}{M}$, and p by q. Therefore, all the results stated in Lemmas 2-4 are valid for the MIMO case, by substituting K by $\frac{K}{M}$, which completes the proof of Theorem 5.

■

In fact, algorithm 2 basically separates the MIMO-BC into M “virtual” SISO-BCs by assigning the users to the beam for which the maximum SINR is attained. Therefore, the analysis of $f_\nu(l)$ is similar to the case of SISO-BC, discussed in the previous section. However, there are two main differences: i) In SISO-BC, all the users are always served by the same transmitter, while in MIMO-BC the users are switched independently between the virtual transmitters, from frame to frame. This causes $\nu_1, \cdots, \nu_K$ (The packet expiry countdown of the users) not to be necessarily distinct. However, we have shown in the proof of Theorem 5 that this does not affect the analysis. ii) The sizes of the virtual SISO-BCs ($\mathcal{A}_m$) are not fixed. In fact, $|\mathcal{A}_m|$, $m = 1, \cdots, M$, are Binomial random variables with parameters $(K, \frac{1}{M})$. Using Gaussian approximation for the Binomial distribution, we can write

$$\Pr\left\{\frac{K}{M}(1-\epsilon) < |\mathcal{A}_m| < \frac{K}{M}(1+\epsilon)\right\} \approx 1 - 2Q\left(\frac{\frac{K}{M}\epsilon}{\sqrt{\frac{K}{M}(1-\frac{1}{M})}}\right). \qquad (123)$$

Setting $\epsilon \triangleq \sqrt{\frac{2(M-1)\log K}{K}}$, and using the approximation $Q(x) \approx \frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}$ for $x \gg 1$, the above equation can be written as

$$\Pr\left\{\frac{K}{M}\left(1-\sqrt{\frac{2(M-1)\log K}{K}}\right) < |\mathcal{A}_m| < \frac{K}{M}\left(1+\sqrt{\frac{2(M-1)\log K}{K}}\right)\right\} \sim 1 - o\left(\frac{1}{K}\right). \qquad (124)$$

Therefore, with probability one, the size of the sets $\mathcal{A}_m$ scales as $\frac{K}{M}\left[1 - O\left(\sqrt{\frac{\log K}{K}}\right)\right]$. Following the above discussions, MIMO-BC can be considered as M parallel SISO-BCs, each serving approximately $\frac{K}{M}$ users. The network dropping event ($\mathscr{B}$) can be considered as the union of the dropping events for the SISO sub-channels, denoted by $\mathscr{B}_m$, $m = 1, \cdots, M$. From the union bound for the probability, we have

$$\begin{aligned} \text{Pr}\{\mathscr{B}\} &\leq \sum_{m=1}^{M} \text{Pr}\{\mathscr{B}_m\} \\ &= M\text{Pr}\{\mathscr{B}_m\}, \end{aligned} \tag{125}$$

where the second line comes from the symmetry between the events $\mathscr{B}_m$. Following the steps of proof for Theorem 4, and setting $\frac{[\log K]^{1.5}}{K} < p < \frac{[\log K]^2}{K}$ and $D = \frac{p}{\varphi}\frac{K}{M} + \frac{\log K}{\varphi} + 9\sqrt{K}[\log K]^4$, guarantees $\text{Pr}\{\mathscr{B}_m\} \to 0$, and hence, $\text{Pr}\{\mathscr{B}\} \to 0$. Note that as $p \sim \frac{e^{-\frac{M\Upsilon}{P}}}{(1+\Upsilon)^{M-1}}$ [24], the condition $\frac{[\log K]^{1.5}}{K} < p < \frac{[\log K]^2}{K}$ incurs that

$$\frac{P}{M}\left[\log K - (M+1)\log\log K\right] < \Upsilon < \frac{P}{M}\left[\log K - (M+0.5)\log\log K\right]. \tag{126}$$

Noting that $\mathcal{C}_{\text{sum}} \sim M\log(1 + \frac{P}{M}\log K + O(\log\log K))$ [24], it follows that $\lim_{K\to\infty} \mathcal{C}_{\text{sum}} - \mathcal{R} = 0$.

■

Theorem 5 implies that the proposed scheduling algorithm is capable of achieving the maximum sum-rate throughput, while guaranteeing $\lim_{K\to\infty} \frac{MD_{\min}}{K} = 1$, where $D_{\min}$ is the minimum value of D such that $\text{Pr}\{\mathscr{B}\} \to 0$. Noting that $\lceil \frac{K}{M} \rceil$ is the minimum value of D in MIMO-BC to have $\text{Pr}\{\mathscr{B}\} \to 0$, (using Round-Robin scheduling, assuming that M users are served during each frame), it follows that the proposed scheme achieves the maximum sum-rate and maximum fairness in the network, simultaneously.

Defining the *minimum average throughput* as in (79), it is straightforward to show that for the proposed algorithm,

$$\mathscr{T}_{\min} \sim \frac{M\log\log K}{K}, \tag{127}$$

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 a 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 maximum fairness in the network, simultaneously, in the asymptotic case of $K \to \infty$. Moreover, we have introduced a performance measure 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 maximum fairness in the network, simultaneously, in the asymptotic case of $K \to \infty$.

APPENDIX A; PROOF OF (100)

From the definition of $\nu_i(t)$, we have

$$\begin{aligned}
\Pr\{\nu_i(t) = l_1, \nu_j(t) = l_1\} &= \Pr\left\{\nu_i(\psi) = D, \nu_j(\psi) = D, \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l), \bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l)\right\} \\
&\overset{(a)}{=} \Pr\left\{\mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l), \bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l)\right\} \\
&= \Pr\left\{\mathscr{X}_i(\psi-1)\right\} \Pr\left\{\mathscr{X}_j(\psi-1) \,|\, \mathscr{X}_i(\psi-1)\right\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l) \,\middle|\, \mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1)\right\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\} \\
&\overset{(b)}{=} \Pr\left\{\mathscr{X}_i(\psi-1)\right\} \Pr\left\{\mathscr{X}_j(\psi-1) \,|\, \mathscr{X}_i(\psi-1)\right\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l) \,\middle|\, \mathscr{X}_i(\psi-1)\right\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\} \\
&\overset{(c)}{=} \Pr\{\nu_i(t) = l_1\} \Pr\left\{\mathscr{X}_j(\psi-1) \,|\, \mathscr{X}_i(\psi-1)\right\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\}, \qquad (128)
\end{aligned}$$

where $\psi \triangleq t - D + l_1$. In the above equation, (a) comes from the fact that the events $\nu_i(\psi) = D$ and $\mathscr{X}_i(\psi - 1)$ are equivalent [9]. (b) results from the fact that conditioned on $\mathscr{X}_i(\psi-1)$, $\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)$ is independent of $\mathscr{X}_j(\psi-1)$ [10]. Finally, (c) follows from writing $\Pr\left\{\mathscr{X}_i(\psi-1)\right\} \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l) \,\middle|\, \mathscr{X}_i(\psi-1)\right\}$ as $\Pr\{\nu_i(t) = l_1\}$. For computing $\sigma \triangleq \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_i(\psi-1), \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\}$, we have

$$\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\} = \sigma\mu + \sigma^*(1-\mu), \qquad (129)$$

[9] In fact, if we have $\mathscr{X}_i(\psi-1)$, i.e., the user i is served in the $(\psi-1)$th frame, in the next frame its expiry countdown will be set to D. In other words, $\mathscr{X}_i(\psi-1)$ results in $\nu_i(\psi) = D$. By a similar argument one can conclude that $\nu_i(\psi) = D$ results in $\mathscr{X}_i(\psi-1)$. Therefore, this two events are equivalent.

[10] In fact, since in each frame M users are served with probability one, conditioned on $\mathscr{X}_i(\psi-1)$, there are $M-1$ other users which are served in the same frame. Since the rest of users are all the same for the ith user (because of the homogeneity of the network), it follows that the condition $\mathscr{X}_j(\psi-1)$ does not change the conditional probability $\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l) \,\middle|\, \mathscr{X}_i(\psi-1)\right\}$.

where $\sigma^* \triangleq \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \bigcap_{l=\psi-1}^{t} \mathscr{X}_i^C(l)\right\}$ and

$$\mu \triangleq \Pr\left\{\mathscr{X}_i(\psi-1) \,\middle|\, \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\}.$$

From the above equation, σ can be written as

$$\begin{aligned}
\sigma &= \frac{\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)\right\} - (1-\mu)\sigma^*}{\mu} \\
&\overset{(a)}{\leq} \frac{\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\} - (1-\mu)\sigma^*}{\mu} \\
&\overset{(b)}{\leq} \frac{\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\} - (1-\mu)\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \mathscr{Z}_j\right\}}{\mu} \\
&\overset{(c)}{\approx} \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\},
\end{aligned} \tag{130}$$

where $\mathscr{Z}_j$ denotes the event that user j is excluded from the network, and hence is never served. (a) comes from the fact that the event $\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l)$ reduces $\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\}$. (b) results from the fact that $\sigma^* \geq \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \mathscr{Z}_j\right\}$, which is due to the fact that excluding the jth user from the network, increases the chance of user i to be served during each frame and as a result, reduces the conditional probability $\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \bigcap_{l=\psi-1}^{t} \mathscr{X}_i^C(l)\right\}$. (c) follows from the fact that as $K \to \infty$, the effect of excluding the user j from the network on the conditional probability $\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\}$ is negligible. In other words,

$$\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1), \mathscr{Z}_j\right\} \approx \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\}.$$

Substituting σ from the above equation in the right hand side of (128) yields

$$\begin{aligned}
\Pr\{\nu_i(t)=l_1, \nu_j(t)=l_1\} &\leq \Pr\{\nu_i(t)=l_1\}\Pr\{\mathscr{X}_j(\psi-1) \mid \mathscr{X}_i(\psi-1)\} \times \\
&\quad \Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_j^C(l) \,\middle|\, \mathscr{X}_j(\psi-1)\right\} \\
&\overset{(a)}{=} \Pr\{\nu_i(t)=l_1\}\Pr\{\nu_j(t)=l_1\}\frac{\Pr\{\mathscr{X}_j(\psi-1) \mid \mathscr{X}_i(\psi-1)\}}{\Pr\{\mathscr{X}_j(\psi-1)\}} \\
&\overset{(b)}{\approx} \Pr\{\nu_i(t)=l_1\}\Pr\{\nu_j(t)=l_1\}\frac{M-1}{M},
\end{aligned} \tag{131}$$

where (a) follows from the fact that $\Pr\{\mathscr{X}_j(\psi-1)\}\Pr\left\{\bigcap_{l=\psi}^{t} \mathscr{X}_i^C(l) \,\middle|\, \mathscr{X}_i(\psi-1)\right\} = \Pr\{\nu_i(t)=l_1\}$, and (b) results from the fact that $\Pr\{\mathscr{X}_j(\psi-1)\} \sim \frac{M}{K}$ (which we have shown earlier in the paper in (112)) and also $\Pr\{\mathscr{X}_j(\psi-1) \mid \mathscr{X}_i(\psi-1)\} \sim \frac{M-1}{K}$. The latter is due to the fact that conditioned on $\mathscr{X}_i(\psi-1)$, the network can be considered as a $(K-1)$-user broadcast channel, in which $(M-1)$ beams are to be assigned to $(M-1)$ users. Hence, the probability of assigning a beam to a randomly selected user is $\frac{M-1}{K-1} \approx \frac{M-1}{K}$. From (131), (100) easily follows.

Appendix B

For upper-bounding the right hand side of (108), we use the fact that

$$G_\nu(l-1) \leq \frac{M(D-l+1)}{K}, \tag{132}$$

which follows from the fact that $f_\nu(l) \leq \frac{M}{K}$, $\forall l$, and consequently, $G_\nu(l-1) = \sum_{\nu=l}^{D} f_\nu(\nu) \leq \frac{M(D-l+1)}{K}$. Having the above equation, RH (108) can be upper-bounded as follows:

$$\begin{aligned}
\text{RH (108)} &\leq M\sum_{n=1}^{K}\binom{K-1}{n-1}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\prod_{i=1}^{n-1}\left(\frac{M(D-l+1)}{K}+\frac{Mi}{K}\right)\\
&= \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\prod_{i=1}^{n}\left(\frac{M(D-l+1)}{K}+\frac{Mi}{K}\right)\\
&= \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\left(\frac{M(D-l+1)}{K}\right)^n\prod_{i=1}^{n}\left(1+\frac{1}{D-l+1}i\right)\\
&\overset{(a)}{\leq} \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\left(\frac{M(D-l+1)}{K}\right)^n\prod_{i=1}^{n}(1+i)\\
&= \eta\sum_{n=0}^{K-1}\binom{K-1}{n}\left(\frac{q}{M}\right)^n\left(1-\frac{q}{M}\right)^{K-n}\left(\frac{M(D-l+1)}{K}\right)^n(n+1)!\\
&\overset{(b)}{=} \eta\left(1-\frac{q}{M}\right)^K\sum_{n=0}^{K-1}\frac{(K-1)!}{K^n(K-n-1)!}(n+1)\left[(D-l+1)\eta\right]^n\\
&\overset{(c)}{\leq} \eta\left(1-\frac{q}{M}\right)^K\sum_{n=0}^{K-1}(n+1)\left[(D-l+1)\eta\right]^n\\
&\overset{(d)}{\leq} \eta\left(1-\frac{q}{M}\right)^K\frac{1}{\left[1-(D-l+1)\eta\right]^2}\\
&\sim \eta\left(1-\frac{q}{M}\right)^K\\
&\overset{(e)}{\leq} \frac{M(\log K)^2}{K}e^{-(\log K)^{1.5}},
\end{aligned} \tag{133}$$

where $\eta = \frac{q}{1-\frac{q}{M}}$. In the above equation, (a) follows from the fact that $D-l+1 \geq 1$ (since $l \leq D$). (b) follows from writing $\binom{K-1}{n}$ as $\frac{(K-1)!}{n!(K-n-1)!}$ and canceling out $n!$ by $(n+1)!$, which leaves the term $n+1$ in the numerator. (c) results from the fact that $\frac{(K-1)!}{(K-n-1)!} = (K-1)(K-2)\cdots(K-n) \leq K^n$, which leads to having $\frac{(K-1)!}{K^n(K-n-1)!} \leq 1$. (d) follows from upper-bounding the sum $\sum_{n=0}^{K-1}(n+1)\left[(D-l+1)\eta\right]^n$ by an infinite sum $\sum_{n=0}^{\infty}(n+1)\left[(D-l+1)\eta\right]^n$ which equals to $\frac{1}{[1-(D-l+1)\eta]^2}$, noting that since $D-l \leq D-D_0 \leq 9\sqrt{K}(\log K)^4$ and $\eta \sim q \sim Mp \leq \frac{(\log K)^2}{K}$ [11], we have $(D-l+1)\eta \ll 1$. Finally, (e) results from upper-bounding $\eta \sim Mp$ by $\frac{M(\log K)^2}{K}$, which is explained in the footnote, and also approximating

[11] As it is shown in the paper, since $\frac{P}{M}(\log K - (M+1)\log\log K) < \Upsilon < \frac{P}{M}(\log K - (M+0.5)\log\log K)$, we have $p = \frac{e^{-M\Upsilon/P}}{(1+\frac{M\Upsilon}{P})^{M-1}} < \frac{(\log K)^2}{K}$.

$\left(1-\frac{q}{M}\right)^K$ by $e^{-\frac{Kq}{M}} \sim e^{-Kp}$ which is upper-bounded by $e^{-(\log K)^{1.5}}$, which is due to the fact that as $\Upsilon < \frac{P}{M}\left(\log K - (M+0.5)\log\log K\right)$, $p = \frac{e^{-M\Upsilon/P}}{\left(1+\frac{M\Upsilon}{P}\right)^{M-1}} > \frac{(\log K)^{1.5}}{K}$.

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] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, “Optimal throughput-delay scaling in wireless networks - part I: The fluid model,” *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2568–2592, June 2006.

[14] N. Bansal and Z. Liu, “Capacity, delay and mobility in wireless ad-hoc networks,” *in Proc. IEEE INFOCOM*, April 2003, pp. 1553–1563.

[15] S. Toumpis and A. J. Goldsmith, “Large wireless networks under fading, mobility, and delay constraints,” *in Proc. IEEE INFOCOM*, March 2004, pp. 609–619.

[16] G. Sharma L. Xiaojun, R. R. Mazumdar, and N. B. Shroff, “Degenerate delay-capacity tradeoffs in ad-hoc networks with brownian mobility,” *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2777–2784, June 2006.

[17] M. J. Neely and E. Modiano, “Capacity and delay tradeoffs for ad hoc mobile networks,” *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 1917–1937, June 2005.

[18] M. Airy, S. Shakkottai and R. Heath, “Limiting queuing models for scheduling in multi-user MIMO systems,” *IASTED Conference on Communications, Internet and Information Technology*, Scottsdale, AZ, November 17–19, 2003.

[19] Manish Airy, Sanjay Shakkottai and Robert W. Heath Jr, “Scheduling for the MIMO Broadcast Channel: Delay-Capacity Tradeoff,” Preprint.

[20] Giuseppe Caire, Ralf R. Muller and Raymond Knopp, “Hard Fairness Versus Proportional Fairness in Wireless Communications: The Single-Cell Case,” *IEEE Trans. on Information Theory*, vol. 53, no. 4, pp. 1366–1385, April 2007.

[21] 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.

[22] 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.

[23] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels." Unpublished.

[24] 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.

[25] D. Bertsekas and R. Gallagher, *Data Networks.* Prentice Hall, Englewood Cliffs, NJ, 2nd edition, 1991.