User Throughput Estimation for the PF Scheduling Algorithm under MIMO Channel Environments

Jeongsoo Choi, Yong-Hwa Kim, Jong-Ho Lee, Saewoong Bahk, and Seong-Cheol Kim

Abstract—In cellular systems, user throughput estimation is needed for cell-site selection, handover design, etc. This letter introduces long-term average user throughput estimation methods for the proportional fair (PF) scheduling algorithm under multiple-input multiple-output (MIMO) channel environments. From the observation that the current data rate divided by the long-term average throughput of each user is roughly around a similar level, we propose estimation methods for both single user MIMO (SU-MIMO) and simple multiuser MIMO (MU-MIMO) scheduling scenarios. Since our throughput estimation methods do not depend on the statistics of other users’ data rates, they can be easily implemented in practical systems regardless of user data rate distribution.

Index Terms—Throughput estimation, MIMO channel, proportional fair scheduling.

I. INTRODUCTION

SINCE mobile devices have become widespread, and the demand for mobile data bandwidth has increased explosively, the next generation wireless network standards have widely adopted multiple-input multiple-output (MIMO) techniques. Even though these promising physical layer schemes provide a reliable and high capacity link between a base station (BS) and users, sharing the available resources is still a crucial concern.

In order to balance throughput and fairness among users, Kelly proposed the proportional fair (PF) criterion [1]. The PF scheduling algorithm serves each user according to the amount of past received-data and the instantaneous data rate of each user. Due to its low implementation complexity and remarkable performance gain, various properties have been explored in the literature, including convergence [2], optimality [3], [4], throughput analysis [5], [6], etc.

The motivation of this work comes from the consideration of recent cellular environments, where each user tries to associate with a BS providing the best received signal strength (RSS). However, the association based only on the RSS information results in the loads between neighboring BSs being unbalanced. For this reason, estimating the long-term achievable throughput from each BS and associating with the BS providing the highest long-term throughput are the key factors for system performance improvement. Thus, the accurate estimation of user throughput is important in improving system performance.

In MIMO channels, the statistics of user data rates are widely varying with transmission schemes, receiver types, number of antennas, correlation between antennas, etc. Instead of providing analytical throughput estimation formulae for various system configurations, we propose to use a practical approach. Our scheme monitors the variation of user data rates during a short time period (i.e. a few hundreds of time slots), and then estimates the achievable throughput from the monitored samples.

According to the property studied in [5] and [6] where the data rate divided by long-term average throughput for each user fluctuates around a similar level, we propose a throughput estimation method for a single user MIMO (SU-MIMO) scheduling case where every data stream is transmitted into a single user. And extend this result to a simple multiuser MIMO (MU-MIMO) scheduling case. We also confirm that the long-term average user throughput approximation does not depend on the statistics of other users’ data rates.

Throughout this letter, we denote $E[\cdot]$ as the expectation operator, $[^H]$ as the conjugate transpose, $\mathbb{C}^{m \times n}$ as the set of $m \times n$ complex matrices, and $\mathbf{I}_n$ as the $m \times m$ identity matrix. Also $\mathbf{H}^{-1}$ and $[\mathbf{H}]_{m,n}$ are the inverse matrix and the $(m,n)$-th element of matrix $\mathbf{H}$, respectively.

II. MODEL DESCRIPTION

Consider a MIMO channel with a BS equipped with $n_t$ transmit antennas and $N$ users equipped with $n_r$ receive antennas each. Users share the downlink channel using the time multiplexing scheme and as a result, a single user is serviced at each time slot. It is assumed that users have the perfect channel state information (CSI) whereas the BS does not. Each user measures the obtainable data rate and feeds back the channel quality indicators (CQIs) to the BS, and the BS serves users according to PF policy. The received signal vector $\mathbf{y}_i \in \mathbb{C}^{n_r \times 1}$ for user $i$ is given by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i,$$

where $\mathbf{H}_i \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix for user $i$, including both large-scale behaviors (i.e. pathloss and shadowing) and small-scale fading. The transmitted signal vector $\mathbf{x} \in \mathbb{C}^{n_t \times 1}$ has the total transmit power constraint as $\mathbb{E}[|\mathbf{x}|^2] \leq P$, and $\mathbf{n}_i \in \mathbb{C}^{n_r \times 1}$ is a spatially white noise vector for user $i$ with the covariance matrix of $\mathbb{E}[\mathbf{n}_i \mathbf{n}_i^H] = N_0 \mathbf{I}_{n_r}$.

Assume that $n_t \leq n_r$ and the BS transmits $n_t$ independent data streams with equal power. The statistics of user data

Manuscript received June 19, 2012. The associate editor coordinating the review of this letter and approving it for publication was H. Viswanathan.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0000162).

J. Choi, S. Bahk, and S.-C. Kim are with the Department of Electrical Engineering and INMC, Seoul National University, Seoul, Korea (e-mail: {sic722, sbahk, samkim}@snu.ac.kr).

Y.-H. Kim is with the Division of Maritime Electronic and Communication Engineering, Mokpo National Maritime University (MMU), Mokpo, Korea (e-mail: yhkim@mmu.ac.kr).

J.-H. Lee is with the Department of Electronic Engineering, Gachon University, Seongnam-si, Gyeonggi-do, Korea (e-mail: jongho.lee@gachon.ac.kr).

Digital Object Identifier 10.1109/WCL.2012.072012.120455
rates vary according to the receiver type even under a same environment. The data rate of user \( i \) using an optimal receiver is [7]
\[
r_i = \log_2 \det(\mathbf{I}_{n_r} + \frac{P}{n_t n_0} \mathbf{H}_i \mathbf{H}_i^H).
\] (2)

As a suboptimal approach, the data streams from the BS can be decoded independently when a linear receiver such as zero-forcing (ZF) receiver or minimum mean squared error (MMSE) receiver is used. The data rate using a linear receiver is given as
\[
r_i = \sum_{m=1}^{n_t} \log_2(1 + \gamma_{i,m}),
\] (3)
where \( \gamma_{i,m} \) is the signal to noise ratio (SNR) of user \( i \) from the \( m \)-th transmit antenna given by
\[
\gamma_{i,m}^{ZF} = \frac{P}{n_t n_0 |\mathbf{H}_m^H \mathbf{H}_i|^2},
\] (4)
for ZF receiver, and
\[
\gamma_{i,m}^{MMSE} = \frac{1}{\left(\frac{1}{n_t n_0} |\mathbf{H}_m^H \mathbf{H}_i|^{-1}\right)} - 1
\] (5)
for MMSE receiver [8].

Regardless of receiver types used in practical systems, the average throughput of user \( i \) until time slot \( t \) is expressed as
\[
\theta_i(t) = \frac{1}{t} \sum_{\tau=1}^{t} r_i(\tau) I_i(\tau),
\] (6)
where \( r_i(\tau) \) is the data rate of user \( i \) at time slot \( t \) and \( I_i(\tau) \) is the indicator function of the event that the BS chooses user \( i \) to serve at time slot \( \tau \) (i.e., \( I_i(\tau) = 1 \) if user \( i \) is serviced at time slot \( \tau \) and 0 otherwise). According to PF scheduling policy, the indicator function of user \( i \)'s activity at time slot \( t \) is defined as
\[
I_i(t) = \begin{cases} 
1 & \text{when } i = \arg \max \theta_i(t-1) \\
0 & \text{otherwise}
\end{cases}
\] (7)
Suppose that the average throughput of each user converges to a stationary value under the PF scheduling algorithm. Let \( \theta_i \) denotes the convergence value of the average throughput of user \( i \), i.e., \( \theta_i(t) \rightarrow \theta_i \) as \( t \rightarrow \infty \). Then, we find the long-term average throughput of each user in the next section.

### III. ESTIMATION OF LONG-TERM AVERAGE THROUGHPUT

#### A. SU-MIMO PF Scheduling

In the SU-MIMO scheduling scheme, every transmit antenna at the BS is devoted to a single user at each time slot. The collection of data rates for user \( i \) given by \( \{r_i(t), t = 1, 2, \ldots \} \) can be modeled as a stationary discrete-time stochastic process, where each element is independent and has an identical distribution with random variable \( R_i \). By the definitions of \( \theta_i(t) \) and \( \theta_i \), the stationary value \( \theta_i \) can be expressed as
\[
\theta_i = \mathbb{E}[\theta_i] = \lim_{t \rightarrow \infty} \mathbb{E}[\theta_i(t)] = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^{t} R_i I_i(\tau)
\] (8)
\[
= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^{t} \mathbb{E}[R_i I_i(\tau)].
\] (9)
For a sufficiently large \( t \), long-term throughput \( \theta_i(t - 1) \) converges to \( \theta_i \) and as a consequence, the collection of preference metrics \( \{u_i(t) = r_i(t)/\theta_i(t-1), t = 1, 2, \ldots \} \) can be approximated as a stationary process where each element is independent of each other and has an identical distribution with random variable \( U_i = R_i/\theta_i \). Since the scheduling choice is made in a dependent manner, \( \mathbb{E}[R_i I_i(\tau)] = \mathbb{E}[R_i | U_i > u_j(\tau), \forall j \neq i] \), where \( \mathbb{E}[\cdot|\cdot] \) is the conditional expectation operator. As \( \tau \) approaches infinity, we have
\[
\lim_{\tau \rightarrow \infty} \mathbb{E}[R_i I_i(\tau)] = \mathbb{E}[R_i | U_i > U_j, \forall j \neq i].
\] (9)
It is known that if a sequence of complex number \( \{s_n\}_{n=1}^\infty \) converges to \( s \), then the sequence of their arithmetic mean \( \frac{1}{n} \sum_{i=1}^{n} s_i \) also converges to \( s \) [9]. Combining (8) and (9), we obtain the integral form of \( \theta_i \) as
\[
\theta_i = \mathbb{E}[R_i | U_i > U_j, \forall j \neq i]
\] (10)
\[
= \int_0^\infty \theta_i u f_{U_i}(u) \prod_{k \neq i} F_{U_k}(u) du,
\] (11)
where \( f_X(\cdot) \) and \( F_X(\cdot) \) represent the probability density function (PDF) and the cumulative density function (CDF) of the random variable \( X \). Equation (11) is derived by taking the expectation of \( R_i = \theta_i u \) where the dummy variable \( u \) varies from 0 to infinity under the condition of \( U_j < U_i = u, \forall j \neq i \).

To make the analysis simple, we assume that all the random variables \( U_i \) have a similar distribution as \( F_{U_i}(u) \sim F_{U_j}(u), \forall i \neq j \) [6]. Fig. 1 illustrates an example of CDF for the random variables \( U_k, \forall k \), for the \( 2 \times 2 \) SU-MIMO scheduling scenario with an optimal receiver and \( N = 10 \). The detailed system parameters are presented in Section IV. It shows that a user with minimum mean rate 18Mbps has a bit wider fluctuation ratio of its instant rate to the average rate compared to another user with maximum mean rate 120Mbps. However, all the random variables \( U_i \) fluctuate around the similar interval, and it leads us to use the assumption that

\footnote{The equality between CDFs holds under the symmetric scenario (i.e. a user with mean rate 10Mbps has 1Mbps fluctuation whereas the other user with mean rate 3Mbps have 0.3Mbps fluctuation).}
Therefore we can estimate the long-term average user throughput as from the CDF of data rates and the number of users.

\[ F_{U_i}(u) \approx F_{U_j}(u), \forall i \neq j, \] with some marginal error even under the asymmetric scenario.

With this observation, we can rewrite equation (10) as

\[ \theta_i \approx \int_0^{\infty} \theta_i f_{U_i}(u) (F_{U_i}(u))^{N-1} du. \] (11)

Refer to [5], [6] for a justification of the above derivation. Using the linear relationship between \( R_i \) and \( U_i \), the CDF of \( U_i \) is expressed as \( F_{U_i}(u) = F_{R_i}(\theta_i u) \). Substituting this relationship into (11) and changing the dummy variable as \( z = F_{U_i}(u) \), we have

\[ \theta_i \approx \int_0^{1} F_{R_i}^{-1}(z) z^{N-1} dz, \] (12)

where \( F_{R_i}^{-1}(\cdot) \) represents the inverse function of \( F(\cdot) \). Hence, the long-term average user throughput is determined by the CDF of the data rate distribution and the number of users.

To reduce complexity, we introduce a linear approximation process. Let us define \( G(z) = F_{R_i}^{-1}(z) \) and then the first order Taylor series expansion of \( G(z) \) is given as

\[ G(z) \approx G(z_0) + G'(z_0)(z - z_0), \] \( z_0 \in (0, 1) \). Substituting this equation into (12), we have \( z_0 \) dependent approximation result, which is

\[ \theta_i(z_0) = \frac{1}{N} G(z_0) + \frac{1}{N+1} G'(z_0) - \frac{z_0}{N} G'(z_0). \] (13)

It is clear that equation (13) varies with \( z_0 \), and choosing an appropriate one is crucial in reducing the linear approximation error. As we can see in (12), the contribution of \( G(z) \) to the integral grows as \( z \) goes to one, and we should pay more attention to the interval near one. For almost every fading scenario, the PDF of data rate distribution has a bell shaped curve (i.e. see [8] for linear receivers). Let the maximum of \( f_{R_i}(\cdot) \) be attained at \( \tau_i \), and then, the CDF of data rate distribution is concave on the interval \( (\tau_i, \infty) \). Hence, we can conclude that \( G(z) \) is convex on the interval \( (F_{R_i}(\tau_i), 1) \), which is the main interval of interest.

For a convex function, its linear approximation always results in a smaller value than the original function and as a consequence, \( \theta_i(z_0) \) in (13) is smaller than \( \theta_i \) in (12) regardless of \( z_0 \). This means that the gap between \( \theta_i(z_0) \) and \( \theta_i \) is minimized when \( \theta_i(z_0) \) is maximized. To find \( z_0 \), which maximizes \( \theta_i(z_0) \), we take a partial derivative of (13) with respect to \( z_0 \) and then have

\[ \frac{\partial \theta_i(z_0)}{\partial z_0} = G''(z_0) \left( \frac{1}{N+1} - \frac{z_0}{N} \right), \] (14)

where \( G''(\cdot) \) represents the second order derivative function.

From (14), we can conclude that the gap between \( \theta_i \) and \( \theta_i(z_0) \) is minimized when \( z_0 = \frac{N}{N+1} \), and \( \theta_i \) is simply expressed as

\[ \theta_i \approx \theta_i(z_0) = \frac{1}{N} F_{R_i}^{-1} \left( \frac{N}{N+1} \right). \] (15)

Therefore we can estimate the long-term average user throughput from the CDF of data rates and the number of users.

**B. MU-MIMO PF Scheduling**

In the MU-MIMO scheduling scheme, each data stream can be transmitted to a different user simultaneously. Since the CSI is not available at BS side, precoding or power allocation is not considered in this work. At each time slot, users report the set of obtainable data rates from all of independently decoded data streams, and the BS assigns each transmit antenna to a single user according to PF scheduling policy. Let \( r_{im}^m(t) \) denote the data rate of the \( m \)-th data stream for user \( i \) and then, the average throughput of user \( i \) until time slot \( t \) is given as

\[ \theta_i(t) = \frac{1}{t} \sum_{\tau=1}^{t} \sum_{m=1}^{n_t} r_{im}^m(\tau) I_i^m(\tau), \] (16)

where \( I_i^m(\tau) \) is the function to indicate the relation between the \( m \)-th transmit antenna and user \( i \) at time slot \( t \) defined by

\[ I_i^m(\tau) = \begin{cases} 1 & \text{when } i = \arg \max_k r_{im}^m(\tau) \\ 0 & \text{otherwise.} \end{cases} \] (17)

The collection of data rates from the \( m \)-th transmit antenna to user \( i \) given by \( \{r_{im}^m(t), t = 1, 2, \ldots\} \) can be modeled as a stochastic process that consists of independent and identical distributed (i.i.d.) random variables. Without any precoding or power allocation scheme, the data streams from each transmit antenna are assumed to have a same distribution owing to the symmetry property. Thus, for the simple MU-MIMO scheduling scenario, we can express the long-term average user throughput as the summation of throughput from each transmit antenna as follows.

\[ \theta_i \approx \frac{1}{N} \sum_{m=1}^{n_t} F_{R_i}^{-1} \left( \frac{N}{N+1} \right). \] (18)

**C. Implementation Issues**

To estimate the long-term user throughput using (15) and (18), the distribution of data rates should be known. Several works provide the CDF of data rates when each element in MIMO channels has i.i.d. or semi-correlated Rayleigh distribution [8], [10]. In practical systems, the BS keeps track of the variation of data rate of user \( i \) during \( N_s \) time slots instead of using the analytical results, and considers \( F_{R_i}^{-1}(\frac{N}{N+1}) \) as the \( \lceil \frac{N}{N+1} \rceil \)-th minimum value from the measured data rate set, where \( \lceil \cdot \rceil \) represents the ceiling operation.

**IV. PERFORMANCE EVALUATION**

Consider a single cell MIMO downlink scenario. We set the cell coverage as 2km, the total transmit power as 46dBm, the center frequency as 2000MHz, the system bandwidth as 10MHz, and the interval of time slot as 10ms. The height of the BS is set as 15m and as a consequence, the pathloss of an urban environment is expressed as

\[ L = 128.1 + 37.6 \log_{10}(d[\text{km}]), \] where \( d \) is the distance between the BS and a user. The log-normal shadowing with the standard deviation of 8dB is also considered. Each component in MIMO channels is generated using Jake’s fading model, assuming the speed of each user is 3km/h. Other parameters not shown in here follow the 3GPP LTE (Long Term Evolution) specification.

\( ^2 \)Also users can estimate their long-term average throughputs if the number of active users is provided.
The performance metric in this letter is the relative error to the actual long-term average throughput. To obtain the actual long-term throughput, we execute simulations over 10,000 time slots where the average throughput of each user is considered to be converged. It is assumed that the large-scale behavior is stationary during our simulation run, and only the small-scale fading yields variation in MIMO channels. As mentioned in Section III-C, we monitor the first $N_s$ data rates and estimate the long-term throughput using (15) and (18). Moreover, each result is obtained by averaging 100 different simulation initializations.

Fig. 2 shows the relative error to the actual throughput versus the number of users under $2 \times 2$ MIMO scheduling scenarios, including both SU-MIMO with three different types of receivers and MU-MIMO with two different types of receivers. In almost every case, our proposed methods yield accurate results for a wide range of users in number, where the relative error to the actual throughput is less than 6% when $N_s = 100$, and 4% when $N_s = 300$. Hence, the estimation error is reduced when there are abundant observed samples.

In order to verify the performance of the proposed method extensively, we consider various SU-MIMO scheduling scenarios. Fig. 3 shows the CDF of the relative error to the actual throughput when $N = 10$ and $N_s = 200$ for both $2 \times 2$ and $2 \times 4$ MIMO antenna configurations with an optimal receiver, respectively. Also it shows the results when the transmit antenna correlation is 0.5 and zero. Finally, in the mixed case, each user randomly chooses a receiver among the optimal receiver, ZF receiver and MMSE receiver. For every case, our method yields very accurate results, where the average error is less than 4%. Furthermore, we can also recognize that our proposed methods yield lower error with the number of receiver antennas, followed by a smaller amount of fluctuation in data rate.

V. Conclusion

In this letter, we estimated the long-term average user throughput for the PF scheduling algorithm under MIMO environments. It was shown that our estimation results are highly accurate with an average error of less than 4% for almost every case when there are abundant data rate samples. Our throughput estimation methods are useful for practical use, because they perform well regardless of antenna correlation and configurations. We expect that a wireless network can be designed very well by applying our user throughput estimation methods.

REFERENCES