Packet Level Scheduling schemes for Multi-User MIMO Systems with beamforming

Masoomeh Torabzadeh
Department of Computer Sciences
University of Quebec at Montreal (UQAM)
torabzadeh_khorasani.masoomeh@courrier.uqam.ca

Wessam Ajib
Department of Computer Sciences
University of Quebec at Montreal (UQAM)
ajib.wessam@uqam.ca

ABSTRACT
We investigate the packet-level scheduling for the downlink of multiple-input multiple-output (MIMO) multi-user systems using beamforming. We consider the traffic arrival process and different packet lengths. We tackle low-complex practical implementation that provides low average packet transmission delay and bit error rates (BER) to users. We propose a work-conserving scheduling scheme that considers different users guarantees (heterogeneous users). We implement and compare different MIMO schedulers at the packet level. Simulations show the low average packet transmission delay and bit error rate of our proposed scheduler.

Categories and Subject Descriptors
C.2.1 [Wireless Communication]: resource allocation.

General Terms
Algorithms, Performance, Design.

Keywords
Scheduling, MIMO systems, spatial multiplexing, beamforming.

1. INTRODUCTION
In multi-user multiple-input multiple-output (MIMO) system [1], a multi-antenna base station (BS) can transmit simultaneously different signals on the same frequency bandwidth to different mobile users (MSs) realizing a space division multiple access (SDMA). However, packet scheduling algorithm at the link layer has significant impact on the performance and it is critical to efficiently exploit the MIMO benefits. Using beamforming at the BS can eliminate the interference at the receiving units. In this case, all computations are done at the BS and we can afford dumb receivers [2]. Assuming perfect knowledge of the channel state information (CSI) at the BS, the optimum downlink beamforming is dirty paper coding (DPC) [3–4]. Sub-optimal solutions, such as zero-forcing beamforming (ZFBF), are preferable due to their low complexity compared to DPC [5]. ZFBF provides a good trade-off between performance and complexity [6].

On the other hand, different link level schedulers are proposed in the literature for packet transmission in wireless systems. The max-min fairness scheduler [7] transmits to the users with the smallest mean throughputs. In [8], the authors introduced a maximum carrier-to-interference ratio (max-C/I) scheduler that maximizes the system capacity without providing fairness. A fair scheduler, named proportional fairness, (PF) [9–10] promises a trade-off between the throughput maximization and user fairness. At each slot, the user experiencing the highest instantaneous rate with respect to its average rate is scheduled. If queues are not infinitely backlogged, then the algorithm has to define how to deal with empty queues [11]. With finite queues, different versions can be designed depending on which users are eligible for service and how the average throughput of each user is updated. None of the previous schedulers was designed for MIMO systems in function of different packet lengths and hence they are not able to provide both fairness and low average packet transmission delay.

In this paper, we develop and implement a novel packet scheduler to be used with ZF beamforming, called the beamforming packet-based scheduling (BF-PS), for the downlink of a MIMO multi-user system. The scheduler takes into account practical issues such as the lengths of packets, complexity of implementation, heterogeneous users and fairness while providing high system performance in terms of average user delay, throughput, and BER. We demonstrate by simulations that it outperforms “traditional” schedulers when traffic characteristics are considered.

The rest of this paper is organized as follows. In the next section, we discuss the system model and precoding. In section III, we present our scheduling algorithm and we provide the numerical results in section IV. The fairness is discussed in the fifth section and we conclude in section VI.

2. SYSTEM MODEL AND PRECODING
We consider the downlink of a MIMO multi-user system, where one N-antennas BS is communicating simultaneously with K single-antenna MSs ($K \leq N$). We assume that the channels between the BS and the users are invariant during each time-slot of length $T_s$ and change over the time-slots. For the SDMA system, the signal model at time $t$ can be given by:

$$\mathbf{y}(t) = \mathbf{H} \mathbf{w}(t) + \mathbf{n}(t)$$  

where $\mathbf{y}(t)$ is the $K \times 1$ vector of signals received by the $K$ users, $\mathbf{s}(t)$ is the $K \times 1$ vector of signals sent by the BS, $\mathbf{n}(t)$ is the $K \times 1$ vector of complex Gaussian noise elements experienced by the $K$ users (i.e. $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_K$), where $\mathbf{I}_K$ is the $K \times K$ identity matrix) and $\mathbf{H}$ is the $K \times N$ channel matrix between the BS and users. We assume Rayleigh fading scenario, so the elements of each row of $\mathbf{H}$ corresponding to user $k$, $[\mathbf{h}](n) (n=0, 1, \ldots, N-1)$, are circularly symmetric zero mean unit variance complex Gaussian random variables. The $N \times K$ beamingforming matrix $\mathbf{W}$ contains the $K$ weight vectors for the users, $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \ldots \mathbf{w}_K]$. Assuming perfect channel knowledge at the BS, $\mathbf{W}$ can be given by [12]

$$\mathbf{W} = \mathbf{H}^H (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{D}_u \mathbf{D}_b$$  

where $\mathbf{D}_u$ and $\mathbf{D}_b$ are diagonal matrices containing the elements of $\mathbf{H}$ with corresponding to user $k$, $[\mathbf{h}](n) (n=0, 1, \ldots, N-1)$, are circularly symmetric zero mean unit variance complex Gaussian random variables.
where $(.)^H$ denotes the conjugate transpose, $D_\beta = \text{diag}(\beta_1, ..., \beta_K)$ contains the power allocation factors. $D_\alpha = \text{diag}(\alpha_1, ..., \alpha_K)$ captures the effect of equivalent channels.

Applying this, the received signal for user $k$ is

$$y_k(t) = \alpha_k \beta_k s_k(t) + n_k(t)$$

(4)

The interference between co-transmitted packets is mitigated but the effect of simultaneous transmission is remained in $\alpha_k$ and it increases if users’ channels are highly correlated.

There are various power allocation schemes [12]. Since the transmitter has complete knowledge of CSI, power can be allocated to provide minimum sum BER (MSB), or to guarantee fair equal rate and BER (ERB) to the users. The capacity provided by MSB approaches that of ERB at high SNR. Consequently, we assume that BS forces an equal SNR (hence equal rate and BER) to all users. Therefore, as $\alpha_k \beta_k$ would be the same for all users, the SNR for each user $k$ ($k = 1, ..., K$) will be defined as [13]

$$\gamma_k = \frac{P}{\sigma^2 \times \text{tr}((HH^H)^{-1})_k}$$

(5)

where $P_T$ is the total transmission power. Assuming a bandwidth of $B$, the rate of user $k$ can be calculated as

$$r_k = B \times \log_2 (1 + \gamma_k)$$

(6)

For BPSK modulation and using the usual approximation of $Q(x)$ [13 – 14], the BER can be approximated by

$$BER_k = \frac{1}{2} \exp\left(-\frac{|\beta_k|^2 \zeta_k^2}{\sigma^2}\right)$$

(7)

where $\zeta_k, k = 1, ..., K$, are the real nonnegative singular values of $H$.

### 3. PACKET SCHEDULERS

An example of the MAC transmission scheme using our proposed schedulers is shown in Fig. 1. We assume that BS has full CSI. Each time-slot contains some scheduling epochs. At the first epoch, the scheduler selects the $K$ users that maximize a pre-defined utility function. The head of line (HOL) packets of these users are scheduled. Selected users have the same transmission rate. Since packets have different lengths, the transmission of some packets will finish before the others. In this case, the scheduler moves to the next scheduling epoch to select another user(s) (in Fig. 1, when user 5 finishes transmitting at epoch 1; user 8 is selected to replace him at epoch 2). The new user (s) is (are) selected so that the new set of users maximizes the utility function. This procedure is repeated to select as many packets as possible for serving in the time-slot. Using packet fragmentation allows having no unused fractions of time at the end of each time-slot. Note that a packet may be served with various transmission rates since during its transmission; the set of served users and consequently the packet transmission rates may be modified.

After that, the BS broadcasts the scheduling results including the selected users and their allocated times. Following, the scheduled packets are transmitted to MSs. In the aforementioned MAC scheduling framework, we can define different schedulers with various utility functions. In the following sub-sections, we describe how to implement some important packet schedulers in this framework. We provide a low complexity packet scheduler to overcome the complexity of the exhaustive search. The scheduler also provides the quality of service to achieve high performance.

#### 3.1 Proportional fairness (PF)

We implement the PF scheduler in our packet-based MAC framework. At each scheduling epoch, the scheduler has to find the set of users that maximize the PF utility function, keeping the previously selected users whose packets are not completely transmitted. The users’ rate has to be higher than a given threshold $r_{Thr}$. The details of the PF algorithm are given below.
1. Initialize $\Theta = \{\text{users whose packets were fragmented in the previous time-slot}\}$; set the flag $E = 0$; and let the time $F$ equal the start of the current time-slot.

2. Select the set of users $U_F$ ([15]) to be served at $F$ such that

$$P = \arg \max_\Theta \Delta \Theta,$$

subject to $\Theta \subset U_S$, $U_s \subset \Omega$, and $r_k \geq r_{thr}, \forall k \in U_s$.

3. Check $U_F$ and $E$.

   3.1. If $U_F = \emptyset$ and $E = 0$ then fragment the HOL packets of users of $\Theta$; update $\Theta = \emptyset$, set $E = 1$, and go to step 2.

   3.2. Else, if $U_F = \emptyset$ and $E = 1$ then, go to step 6.

   3.3. Otherwise, update $\Theta = U_F$, set $E = 0$ and go to step 4.

4. Update the average throughput, $R_k$, of user $k$ as

$$R_k = \begin{cases} \frac{(T-1)R_k + r_k}{T}, & \text{if } k \in U_F, \\ \frac{(T-1)R_k}{T}, & \text{if } k \notin U_F \text{ and } k \in \Omega, \\ R_k, & \text{elsewhere} \end{cases}$$

5. Let $\Psi$ be the set of users with minimum finish time of their HOL packets defined as:

$$\Psi = \arg \min_{k \in \Psi} F_k$$

where $F_k$ is the finish time of the HOL packet of user $k$. Note that $F_k = F_i, \forall i, j \in \Psi$ and let $F = F_c$.

   5.1. If $F < T$, then update $\Theta = \Theta \setminus \Psi$ and go to step 2 (new epoch with new set of users).

   5.2. Else, if $F = T$, the HOL packets of users belonging to $\Theta \setminus \Psi$ will be fragmented at the end of time-slot.

   5.3. Otherwise, the HOL packets of users belonging to $\Theta$ will be fragmented at the end of time-slot.

6. Go to step 1 for the next time-slot.

### 3.2 Max-C/I

To implement max-C/I [8] scheduler in our framework, we develop a similar algorithm to the one detailed above but we remove step 4 and (8) in step 2 will be replaced by:

$$P = \arg \max_\Theta \Delta \Theta,$$

subject to $\Theta \subset U_S$, $U_s \subset \Omega$, and $r_k \geq r_{thr}, \forall k \in U_s$.

### 3.3 Beamforming packet-based scheduler (BF-PS)

BF-PS is a low complexity scheduler. It selects the users whose channels are both the best and the most orthogonal [16]. It also considers the user queue length to improve the performance. We assume heterogeneous users with different required service share and we define $w_k$ as the weight of user $k$ proportional to its service share. BF-PS defines the following metric of channel quality for user $k$

$$\hat{h}_k = \frac{h_k}{\mu_k}$$

where $\mu_k$ is the time average of $h_k$ norm, $\|h_k\|$, over a past window of size $t_e$ time-slots. It is updated as

$$\mu_k = (1 - \frac{1}{t_e})\mu_k - \frac{1}{t_e} \|h_k\|$$

We describe the BF-PS scheduler in details as follows.

1. Initialize $\Theta = \emptyset$, set $E = 0$; then fragment the HOL packets of users of $\Theta$; update $\Theta = \emptyset$, set $E = 1$, and go to step 2.

2. Else, if $E = 1$ then, go to step 8.

3. Otherwise, select the set of users $\Theta_{max}$ to be served at time $F$, set $E = 0$, and go to step 7.

6. Check $\Theta_{max}$ and $E$.

   6.1. If $\Theta_{max} = \emptyset$ and $E = 0$, then fragment the HOL packets of users of $\Theta$; update $\Theta = \emptyset$, set $E = 1$ and go to step 3.

   6.2. Else if $\Theta_{max} = \emptyset$ and $E = 1$, then go to step 8.

   6.3. Otherwise, select the set of users $\Theta_{max}$ to be served at time $F$, set $E = 0$, and go to step 7.
7. Compute \( \mathcal{P} = \arg \min_{\Theta \in \mathcal{P}} F_i \) as in (10) and update \( F = F_i, \forall i \in \mathcal{P} \).

7.1. If \( F < T_s \), then update \( \Theta = \Theta = \Theta_{\text{max}} \), \( C_{\text{max}} = 0 \), \( \Theta_{\text{max}} = \phi \), and go to step 3 for the next epoch.

7.2. Else, if \( F = T_s \), the HOL packets of users belonging to \( \Theta \backslash \mathcal{P} \) will be fragmented at the end of time-slot.

7.3. Otherwise, the HOL packets of users belonging to \( \Theta \) will be fragmented at the end of time-slot.

8. Go to step 1 for the next time-slot.

3.4 G-MSR
We implement greedy max-sum-rate (G-MSR) scheduling algorithm [16] in our framework. We use a similar algorithm to the one described above for BF-PS scheduler but we remove steps 1, and we set the parameter \( \gamma \) to zero. If there are enough backlogged users, G-MSR selects \( N \) users at each epoch, while BF-PS can select 1, 2 or \( N \) users depending on the provided system capacity.

4. SIMULATION EXPERIMENTS

4.1 System parameters and user behaviours
Our simulations assume \( N = 4, P_T / \sigma^2 = 16 \text{ dB}, B = 5 \text{ MHz}, T_S = 2.4 \text{ ms}, T = 2000 \text{ packets and } t_c = 500 \text{ time-slots. Our simulation duration is 28.8 s = 12000 time-slots.}

The traffic model considers that the overall data volume of each mobile user is broken down into HTTP (73%), Napster (9%), e-mail (6%), UDP (4%), FTP (2%), and other TCP applications (6%). UDP traffic is ignored since it is a non-real time traffic generated during active periods. Each of the flows is generated in the packet level with a log normal inter-arrival time distribution \((\mu, \sigma)\), where \( \mu \) is the mean and \( \sigma \) is the standard deviation. For example, to provide the total traffic rate of 56.612 Mbps for 8 users, \((\mu, \sigma) = (-10.2, 2.5229)\) for HTTP, \((-6.3335, 1.8768)\) for Napster, \((-7.2096, 2.3447)\) for e-mail, and \((-6.1089, 2.3280)\) for FTP. We use discrete distribution between 40 and 1500 bytes for the packet size as in [17].

4.2 System performance
Fig. 2 shows the average system throughputs (sum of users’ throughputs) under various traffic loads. We notice that BF-PS scheduler outperforms the G-MSR and is very close to the max-C/I and PF schedulers. As expected, max-C/I provides slightly better throughput at high traffic loads.

Fig. 3 shows that BF-PS scheduler achieves the lowest average user packet delay whereas G-MSR provides the highest. Fig. 4

Figure 2. System throughputs under various traffic loads

Figure 3. Average delay of users under various traffic loads

Figure 4. Average delay of each user under 91.384 Mbps traffic load

Figure 5. Average user rates under 91.384 Mbps traffic load
shows the average packet transmission delay of different users for 91.384 Mbps traffic load. BF-PS scheduler achieves the lowest average delay as well as more fairness among users as user delays are relatively close together. We show the average user transmission rates in Fig. 5. As we can see the user rates using BF-PS is the highest. G-MSR is not able to provide high user rates because it tends to serve simultaneously 4 users if possible, while BF-PS is more flexible. This also can affect the BER of users.

In Fig. 6 and Fig. 7, we show average user BERs (computed by (7)) under various traffic loads and 91.384 Mbps load, respectively. They show that BF-PS can achieve low BER like PF and max-C/I, while G-MSR provides higher BER.

![Figure 6. Average BER of users under various traffic loads](image1)

![Figure 7. Average BER of users under 91.384 Mbps traffic load](image2)

5. CONCLUSION
We tackled the problem of low complexity implementation of packet scheduling for MIMO multi-user systems using zero-forcing beamforming, by taking into consideration the traffic arrival process with different packet lengths. Simulation results considering traffic characteristics demonstrate the performance obtained by using our proposed packet based beamforming scheduler in terms of delay, throughput, system capacity, bit error rate, and fairness.

6. REFERENCES