Rateless Code in QoS-based Multiuser Scheduling for MIMO Systems

Feng Wang 1,2, Zhifeng Zhao 1,2, and Honggang Zhang 1,2
1 York-Zhejiang Lab for Cognitive Radio and Green Communications
2 Department of Information Science and Electronic Engineering
Zhejiang University, Zheda Road 38, 310027 Hangzhou, China
Email: wfrabbit@zju.edu.cn, zhaozf@zju.edu.cn, honggangzhang@zju.edu.cn

Abstract—This paper proposes a new multiuser scheduling scheme for MIMO system that can simultaneously support various QoS requirements of different user groups while satisfying fairness among the same user group in a RRS (Round Robin Scheme) manner. To guarantee high transmission reliability, a cross-layer scheme is adopted by combining spatial diversity in the physical layer and systematic Raptor code in the MAC layer. The delay-sensitive users are scheduled with higher priority. When only one user is chosen at each time slot, spatial diversity is adopted to guarantee a low BER. Furthermore, if Nt users are chosen at each time slot, AA-RRS (Antenna Assisted Round Robin Scheme) is adopted to improve the system capacity, which is brought by the inherent multiuser diversity of the MIMO system. Simulation results demonstrate that the proposed cross-layer scheduling scheme can guarantee a much higher reliability than spatial diversity transmission with only a small increase of delay.

Keywords—MIMO; multiuser diversity; scheduling; Rateless code; systematic Raptor code

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology utilizes multiple antennas at the transmitter and/or the receiver in order to improve transmitting reliability and provide higher data rate. Spatial multiplexing can achieve much higher data rate by dividing the original data stream into multiple substreams and transmitting each substream through a different antenna [1] [2]. Spatial diversity can obtain much more reliability by sending the same data stream through all the antennas.

In cellular networks, independence of fading among different mobile users, called as multiuser diversity, can be exploited to increase the system capacity [3]. The scheduler at a base station (BS) can effectively allocate radio resource to various users in different channel conditions. Basically, fairness, capacity and QoS (quality-of-service) are three criteria which are mainly considered during scheduling. In [4], a round robin scheduling (RRS) scheme serves a number of users in a cyclic fashion regardless of the channel conditions and thus achieves fair sharing of channel accesses among those users. However, this scheme does not take advantage of the inherent multiuser diversity, resulting in a sub-optimal system capacity. O. S. Shin and K. B. Lee introduced an antenna-assisted round robin scheduler which achieved a higher capacity than RRS by adopting the multiuser diversity [5]. But, these two scheduling schemes do not consider the diverse QoS requirements of different users. When the BS transmits a series of packets, under the constraint of QoS requirement, we care about if the mobile user can receive all these packets in real time with high probability. Generally, in physical layer, spatial diversity can be adopted to assure a low BER (bit-error-rate) when the channel condition is not very hostile. But if the channel condition is much worse, spatial diversity can no longer guarantee a low BER and thus a high ratio of the transmitted packets may be lost.

Considering the various QoS requirements by different types of users, this paper proposes a novel multiuser data streams scheduling scheme based on systematic Raptor code to provide reliable service in real-time manner, meanwhile the fairness among the same type of users can be guaranteed as well. Systematic Raptor code, which has been applied to MBMS (Multimedia Broadcast and Multicast Services) and DVB-H (Digital Video Broadcasting-Handheld) [6] [7], is a particular type of rateless codes with the property that the original packets are included within a set of encoding packets. In our multiuser scheduling approach, the data stream of users with high reliability demand is pre-coded into two parts by using systematic Raptor code module, which are the original data stream and the repair data stream, respectively. Accordingly, the original data stream is transmitted with spatial diversity way to guarantee a low BER while the repair data stream is transmitted with spatial multiplexing way to guarantee a low delay and high system capacity. Therefore, the corresponding receivers can ideally recover the original source packets from any subsets of the encoding packets of the size equal to or only slightly larger than the number of the source packets [8] [9], and then feeds back a bit of acknowledgment to the BS. After the BS receives the acknowledgement bits from all the users with high reliability requirements, it stops transmission. The delay-sensitive users will be scheduled at a high priority. The users with both delay-constraints and high

This paper is supported by the Key Project of Education Department of Zhejiang Province (No. 20070247), the Key Project of the Office of Science and Technology of Zhejiang Province (No. 2008C01051-3), and the Key Project of the Ministry of Industry and Information Technology of China (“Research on Broadband Wireless Access Systems Architecture and Framework for Supporting Broadcasting Services” No.2009ZX03005-004).

CHINACOM 2010, August 25-27, Beijing, China
Copyright © 2011 ICST 978-963-9799-97-4
DOI 10.4108/chinacom.2010.19
reliability requirements will be served to get an appropriate tradeoff between them.

The rest of the paper is organized as follows. Section II provides the system and channel model of multiuser MIMO system. Section III gives an overview of two fairness-based scheduling schemes and proposes a novel QoS-aware scheduling scheme based on systematic Raptor code. In Section IV, we evaluate our scheduling scheme in terms of reliability and average delay through numerical simulations. Finally, Section V concludes the paper.

II. SYSTEM AND CHANNEL MODELS

We consider the downlink of a MIMO multiuser system, where the BS with $N_t$ antennas communicates with $U$ users and each user is equipped with $N_r$ antennas, as illustrated in Fig. 1. Each transmit antenna creates a spatial channel, resulting in $N_t$ channels in each time slot [3]. The channel between each user and the BS is assumed to be static during each time slot and varies independently over time slots.

The signal received by the user $u$ at time slot $t$ is given as

$$y_u(t) = H_u s(t) + n_u(t),$$

where $s(t)$ is a $N_r \times 1$ vector sent by the BS, $n_u(t)$ is a $N_r \times 1$ vector of additive white complex Gaussian noise experienced by user $u$, and $H_u$ is the $N_r \times N_t$ channel matrix between the BS and user $u$, which is given as

$$H_u(t) = \sqrt{\frac{SNR_0}{r_u}} \cdot 10^{\alpha/10} \cdot G_u(t),$$

where $SNR_0$ denotes the median signal to noise ratio, $r_u$ is the distance between the BS and the user $u$, $\alpha$ is the path loss exponent, and $S_u(t)$ is a real Gaussian random process with zero mean and variance of $\sigma_u^2$. The elements of the $N_r \times N_t$ matrix $G_u$ are independent complex Gaussian random variables with zero mean and unit variance and represent Rayleigh-distributed multi-path fading.

At the BS, the data streams for all the users are divided into four queues according to their QoS requirement priorities, such as delay-constraints and high reliability. These four queues are as follows

- $Q_1$: Both high reliability and low delay are guaranteed for the data in this queue.
- $Q_2$: Only high reliability is guaranteed for the data in this queue.
- $Q_3$: Only low delay is guaranteed for the data in this queue.
- $Q_4$: Neither high reliability nor low delay is guaranteed for the data in this queue.

The data streams within the same queue have the same QoS requirements and will be scheduled using the same scheme. Based on systematic raptor code, our goal is to guarantee the various QoS demands of each queue while maintaining the optimal system capacity. Moreover, the fairness among the same queue must be achieved. As described in Fig. 1, the data streams in $Q_1$ and $Q_2$ are pre-coded by the systematic Raptor coder module before transmitting. The section 1 contains the original data packets and is transmitted via spatial multiplexing. $Q_3$ and $Q_4$ are never scheduled until $Q_1$ and $Q_2$ are cleared up.

III. PACKET SCHEDULING SCHEME

A. Round Robin Scheduling (RRS)

One user among the $U$ users is selected at each time slot in a round robin fashion, and all the $N_r$ antennas in the BS are assigned to the selected user for this time slot [4]. The system capacity $C_{RRS}(t)$ for the given time slot may be expressed as

$$C_{RRS}(t) = \sum_{u=1}^{U} \log_2(1 + \gamma_{u,n}(t)).$$

In (3), $\gamma_{u,n}$ denotes the PD-SINR for the channel corresponding to the transmit antenna $n$ and the user $u$, which is given [10]

$$\gamma_{u,n}(t) = \frac{P_n |W_u^{(0)}| H_u^{(0)}|^2}{\sigma_n^2 N_t \sum_{m=1}^{N_r} |W_{u,m}^{(0)}|^2 + P_n \sum_{m=1}^{N_r} |W_{u,m}^{(0)}|^2}$$

$$W_u^{(t)} = H_u^{H}(t) \left[ H_u(t) H_u^{H}(t) + \frac{\sigma_n^2 N_t}{P_n} I_{N_t} \right].$$

Figure 1. Multiuser MIMO system.
where $P_r$ is the total received signal power, $\sigma_N^2$ is the noise power per receive antenna, and $W_u(t)$ is the MMSE (minimum-mean-square-error) weight matrix.

In this scheme, every user access $N_t$ spatial channels in $U$ time slots, thus the fairness among all the users can be achieved. However, it does not consider the multiuser diversity and results a sub-optimal system capacity.

B. Antenna-Assisted Round Robin Scheduling (AA-RRS)

This scheme consists of two steps [6]. In the first step, $N_t$ users are selected in each time slot in a Round Robin fashion. For example, in the case of $U=4$, $N_t=2$, in the first time slot, user group $\{u_1, u_2\}$ is selected and in the next time slot, user group $\{u_3, u_4\}$ is selected and so forth. In the second step, the selected users are mapped to each antenna. There are $N_t!$ possible one to one mappings between $N_t$ users and $N_t$ antennas, and an appropriate mapping can realize a diversity gain and achieve an optimal system capacity. For each time slot, every selected user calculates its SINR at the receiver and feeds it back to the BS. Let a sequence $\{u_{i_1}, u_{i_2}, \ldots, u_{i_{N_t}}\}$ represents a mapping with $u_{i_j}$ denoting the antenna $i$ assigned to the user $u_j$. For a given sequence, the system capacity is expressed

$$C_{AA-RRS}(t) = \sum_{n=1}^{N_t} \log_2 \left( 1 + \gamma_{n,u} (t) \right).$$

So the system capacity is calculated as

$$C_{AA-RRS}(t) = \max_{\{u_{i_1}, u_{i_2}, \ldots, u_{i_{N_t}}\}} C_{AA-RRS}(t) \left( \{u_{i_1}, u_{i_2}, \ldots, u_{i_{N_t}}\} \right)$$

The sequence that can achieve this capacity can be found using an exhaustive search over all the possible sequences.

C. QoS-Aware Scheduling Based on systematic Rateless Code

The former two algorithms are based on the system capacity and fairness, but do not consider the various QoS demands of different users. In some particular situations, some data streams may be more important and must be transmitted reliably. On the contrary, some delay-sensitive data streams must be transmitted as soon as possible. In our model, the BS divides the data streams for all the users into four queues before transmitting according to their QoS demands.

<table>
<thead>
<tr>
<th>Type of Queue</th>
<th>Coding</th>
<th>MIMO</th>
<th>Scheduling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue 1 ($Q_1$)</td>
<td>Raptor Code, STBC</td>
<td>VBLAST</td>
<td>Diversity Multiplex</td>
</tr>
<tr>
<td>Queue 2 ($Q_2$)</td>
<td>VBLAST</td>
<td>Multiplex</td>
<td>High-Priority</td>
</tr>
<tr>
<td>Queue 3 ($Q_3$)</td>
<td>Raptor Code, STBC</td>
<td>VBLAST</td>
<td>Diversity Multiplex</td>
</tr>
<tr>
<td>Queue 4 ($Q_4$)</td>
<td>VBLAST</td>
<td>Multiplex</td>
<td>Low-Priority</td>
</tr>
</tbody>
</table>

Each queue consists of more than one data streams, which will be transmitted to different users. The scheduler aims to provide fairness among the same queue and guarantee the QoS demand of different queues at the same time. As shown in Table 1, to guarantee a high reliability, we propose a cross-layer scheme. In the physical layer, we adopt the spatial diversity to achieve a lower BER instead of relying upon the spatial multiplex. In the MAC layer, each data stream in $Q_1$ and $Q_3$ is pre-coded by systematic Raptor code before transmitting. For the delay-sensitive users, $Q_1$ and $Q_3$ are transmitted in high priority. $Q_2$ and $Q_4$ are never scheduled until $Q_1$ and $Q_2$ have been cleared up.

![Figure 2. The transmitting process for $Q_1$ and $Q_3$.](image)

Every data stream in $Q_1$ and $Q_3$ is processed as in Fig. 2 whereas the data stream after the systematic Raptor code consists of two parts, which are the original data stream and the repair data stream. Thus $Q_i$ ($i = 1, 3$) consists of the original data streams $Q_{i,s}$ and the repair data streams $Q_{i,r}$. $Q_{i,s}$ is more important and is transmitted in spatial diversity pattern, which requires higher transmission performance by relying upon a STBC (Space-Time Block Code) module. We consider the probability $p$ at which the receiver can recover all the original packets. When the channel condition is good, the original part can be received at the mobile user with a high $p$. However, when the channel condition is bad, the spatial diversity is not enough to combat the channel fading and some repair packets must be transmitted. $Q_{i,r}$ is transmitted in spatial multiplex pattern, which needs a VBLAST (Vertical-Bell Laboratory layered Space-Time) module. For $Q_{i,r}$, a Diversity Round Robin Scheduling (D-RRS) is adopted. Accordingly, in each time slot, just one user is selected in the queue and the data is transmitted in a diversity way. In the next time slot to schedule this queue, another user is selected by the same way. This D-RRS can guarantee a high $p$ and fairness among the same queue but with some loss of capacity compared with the AA-RRS scheme. For $Q_{i,r}$, AA-RRS is used to guarantee the fairness among the same queue and take advantage of the multiuser diversity.
For the users in $Q_i$ and $Q_j$, the data is transmitted in AA-RRS scheme. Without the reliability-constrained, this scheme can provide an optimal capacity and also guarantee the fairness.

The scheduling algorithm for the four queues is illustrated as follows:

Scheduling Algorithm

Phase 1:

while $(Q_2 \neq \emptyset) \cap (Q_4 \neq \emptyset)$
for odd time slots
$u_i \in G_i$ is scheduled in D-RRS way
end for
for even time slots
$u_i \in G_i$ is scheduled in AA-RRS way
end for
end while

Phase 2:

$G_1^* = G_1$
for $u_i \in G_i$
if has received all the original packets
$u_i$ : sends $ACK \rightarrow BS$
$G_i^* = G_i - u_i$
end if
end for

while $G_1^* \neq \emptyset$
BS: $G_1 \in Q_1$, is scheduled in AA-RRS way
for $u_i \in G_1$
if $u_i$ has recovered all the original packets
$u_i$ : send $ACK \rightarrow BS$
$G_i = G_i - u_i$
end if
end for
end while

Phase 3:

for $Q_1$ and $Q_4$
the same as Phase 1

Phase 4:

for $Q_2$
the same as Phase 2

$G_i$ (or $G_j$) receives the original packets correctly in Phase 1 (or Phase 3), there is no need to transmit the repair packets for $u_i$ in Phase 2 (or Phase 4) and $u_j$ is taken from $G_i$ (or $G_j$). Let $G = G_i - u_i$ represents this modified user group. In Phase 2, when the BS receives an ACK from $u_i$, it stops transmitting the repair data for this user. Phase 3 is never gotten in until the BS has received the ACKs from all the users in $G_i$.

IV. SIMULATION RESULTS

To assess the performance of the proposed QoS-aware scheduling scheme, we consider the reliability and latency as metrics. We assume the number of antennas at the BS and the mobile users $N_b = N_m = 4$, the original packets for each user $K = 500$ and the packet length $L = 256$. Let $P_{s/STBC}$ and $P_{VBLAST}$ denote the BER and PER (packet error rate) when VBLAST is adopted. $P_{s/STBC}$ and $P_{p/STBC}$ represent the BER and PER when STBC is used. For a simple analysis, PER can be calculated as:

$$P_{p/VBLAST} = 1 - (1 - P_{s/VBLAST})^L.$$  (8)

$$P_{p/STBC} = 1 - (1 - P_{s/STBC})^L.$$  (9)

A. Reliability Analysis

The failure probability of Raptor codes with $K (> 200)$ number of source packets and $m$ number of received packets can be modeled by the following equation [11]:

$$P_{f}(m, K) = \begin{cases} 1, & \text{if } m < K \\ 0.85 \times 0.567^{m-K}, & \text{if } m \geq K \end{cases}.$$  (10)

For the users in $Q_i$ and $Q_j$, the probability $p$ that the receiver can recover all the $K$ original packets can be given as:

$$p = (1 - P_{s/STBC})^L + \sum_{m=0}^{M} \binom{M}{m} (1 - P_{s/VBLAST})^m P_{s/STBC} (1 - p_f(m, \lceil K \rceil)).$$  (11)

Proof: Please refer to the Appendix.

In Equation (11), the first part means that the mobile user decodes the original packets at first. If all these $K$ packets are recovered correctly, the receiver then feeds back one bit of acknowledgment to the BS and the BS will not transmit the repair packets. However, if the BS does not receive an acknowledgment bit in a certain time, it will transmit the repair packets and then the mobile user will recover the original packets in a systematic Raptor decoder, as expressed by the second part of the Equation (11). $M$ denotes the number of repair packets that the BS has transmitted.
Through simulation, the BER and PER of VBLAST and STBC in a certain SNR when \( N_T = N_R = 4 \) (we choose SNR=10 here) using BPSK can be given as:

<table>
<thead>
<tr>
<th>TABLE II. THE BER AND PER WHEN SNR=10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
</tr>
<tr>
<td>VBLAST</td>
</tr>
<tr>
<td>STBC</td>
</tr>
</tbody>
</table>

The PER of VBLAST \( P_{\text{VBLAST}} \) and STBC \( P_{\text{STBC}} \) is given by Equation (8) and (9).

\[
T_1 = \frac{2KU}{N_r} + \left( K - \frac{K}{N_r} \right) \frac{U}{N_r} .
\]

\[
T_2 = M \frac{U}{N_r} .
\]

Proof: Please refer to the Appendix.

\( T_1 \) is the time needed to transmit the original \( K \) packets in Phase 1 and \( T_2 \) is the time needed to transmit the \( M \) repair packets in Phase 2. So the total delay for user in \( Q_l \) is

\[
T_\Phi = T_1 + T_2 = \frac{2KU}{N_r} + \left( K - \frac{K}{N_r} \right) \frac{U}{N_r} + M \frac{U}{N_r} .
\]

For each user in \( Q_2 \), the latency can be given

\[
T_\Phi = \frac{2KU}{N_r} .
\]

Proof: Please refer to the Appendix.

Figure 3. The recover probability of spatial diversity, RRS (the same as AA-RRS) and the proposed algorithm (SNR = 10 dB, \( K = 500 \), \( N_T = N_R = 4 \)).

Figure 4. The average delay of \( Q_1 \) and \( Q_2 \).

From the Figure 4, we can see that compared with the average delay of users in \( Q_1 \), the delay of users in \( Q_2 \) is larger, which is because of using the spatial diversity and the transmission of the repair packets to guarantee a high reliability. This increase of delay is considered to be within the threshold of the mobile user. It can also be seen that the major increase of this delay is induced by the spatial diversity, which leads to a low data rate. The transmission of the repair packets has a weak influence on the increase of the delay, as the curve is almost flat. It can also be seen that a large \( U \) will lead to a much larger delay, which is easy to understand. So if there are lots of users with delay-constraints and reliability requirement, it will be necessary to choose a good tradeoff between these two demands.

V. CONCLUSIONS

Former research on MIMO multiuser scheduling schemes has mainly focused on the fairness and the system capacity, but
ignored the QoS demand of different users, especially the reliable transmission of data streams. In this paper, we proposed a new QoS-aware packet scheduling scheme for MIMO system by taking advantage of rateless code. The metrics we consider here to design this scheme are QoS demand, fairness and system capacity. The users are divided into four queues according to their QoS requirements. Different queues are transmitted with different way. For the delay-sensitive queues, the data streams are transmitted at a high priority. To guarantee the reliable transmission, a cross layer scheme is proposed. In the MAC layer, the original data streams are coded by the systematic Raptor code module into two parts to combat the packet loss. The first part, which contains the original data packets, is transmitted with spatial multiplex in the physical layer to guarantee a low BER. The second part, which contains the repair data packets, is transmitted with spatial diversity in the physical layer to guarantee a low BER. With this flexible scheduling scheme, the various QoS demands of different users can be satisfied. RRS and AA-RRS scheme are jointly adopted considering the particular situation to get an ideal tradeoff between the fairness and system capacity. Simulation results have verified that the proposed QoS-aware scheduling scheme based on rateless code provides much more reliable transmission compared with the traditional RRS and AA-RRS schemes, especially when the channel conditions are hostile.

VI. APPENDIX

A. Proof of equation (11)

From the law of large numbers (LLN), the number of the original packets are lost at average during the Phase 1 (Phase 3) for each user in \( G_i \) \( (G_j) \). So when the mobile user has received more than \( K_{p,SSBC} \) repair packets, it can start to decode the original packets. The probability of receiving \( m \) repair packets and decoding the original packets correctly is:

\[
\rho = \left( \frac{M}{m} \right) \left( 1 - P_{VBLAST} \right)^{m} P_{VBLAST}^{\left( M - m \right)} \left( 1 - P_{mK} \left( 1 - P_{p,SSBC} \right) \right)
\]

Summing all the possible \( \rho \), the recover probability can be given as:

\[
p = \left( 1 - P_{p,SSBC} \right)^{K} + \sum_{m=0}^{M} \left( \frac{M}{m} \right) \left( 1 - P_{VBLAST} \right)^{m} P_{VBLAST}^{\left( M - m \right)} \left( 1 - P_{mK} \left( 1 - P_{p,SSBC} \right) \right)
\]

B. Proof of equation (12) (13) (15)

In even time slots, \( N_j \) users in \( G_j \) are scheduled. At average, each user is chosen every \( \frac{U}{N_j} \) time slots. Thus the delay of user in \( G_j \) is

\[
T_{G_j} = \frac{2KU}{N_j}
\]

In Phase 1, the time before \( T_{G_i} \) is divided equally between \( G_1 \) and \( G_2 \). For each user in \( G_1 \), \( \frac{K}{N_j} \) packets have been transmitted during \( T_{G_i} \). After \( T_{G_i} \), the BS serves for \( G_j \) only and each user is scheduled every \( \frac{U}{N_j} \) time slots. So the time for users in \( G_j \) during Phase 1 is

\[
T_i = T_{G_i} + \left( \frac{K}{N_j} \right) \frac{U}{N_j}
\]

During Phase 2, the repair packets are transmitted in a AA-RRS way, so the time to transmit \( M \) packets is

\[
T_2 = \frac{MU}{N_j}
\]

REFERENCES

[4] “throughput simulations for MIMO and transmit diversity enhancements to HSDPA, 3GPP TSG RAN WG1,” Lucent Technol. TSGR1/17 (00) 1388.