Cross layer scheduling for real-time traffic in multiuser MIMO-OFDMA systems

SUN Qiao-yun (✉), TIAN Hui, DONG Kun, ZHANG Ping

Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

A novel cross layer scheduling algorithm is proposed for real-time (RT) traffic in multiuser downlink multiple-input multiple-output orthogonal frequency division multiple access (MIMO-OFDMA) wireless systems. The algorithm dynamically allocates resources in space, time and frequency domain based on channel state information (CSI), users’ quality of service (QoS) requirements and queue state information (QSI). To provide higher data rate and spectrum efficiency, adaptive modulation and coding (AMC) is employed. The proposed algorithm can improve cell throughput and increase the number of users that can be supported while guaranteeing users’ QoS requirements and fairness among all users. Simulation results indicate that the proposed algorithm can achieve superior performance.

Keywords MIMO-OFDMA, cross layer, QoS, AMC

1 Introduction

The forthcoming wireless communication systems are expected to provide high transmission rate while guaranteeing required QoS of mobile users. To meet these requirements, orthogonal frequency division multiplexing (OFDM) and multi-input multi-output (MIMO) are two candidate key technologies [1–2]. MIMO technology can achieve high throughput by multiplexing gain and increase reliability by spatial diversity gain. OFDM is an efficient and low-complexity technique to combat frequency selective fading caused by multipath fading channels. By combining both techniques, MIMO-OFDM can offer both robustness and high throughput by enjoying different degrees of freedom (space, frequency, time, etc.) [3].

In multiuser MIMO-OFDM systems, one of the most important issues is how to dynamically allocate spatio-temporal-spectral resources according to the changing environments, QoS requirements and buffer status, to improve the overall system performance and ensure individual QoS requirement. At present, many scheduling algorithms in MIMO-OFDM systems have been studied [4–5]. However, these algorithms only consider the largest one or two eigenmode subchannels, and do not consider users’ QoS requirements or users’ buffer status. They cannot exploit spatial resources efficiently or satisfy as many users as possible. Proportional fairness (PF) is a famous scheduling algorithm which not only enhances the throughput but also guarantees fairness among users [6]. However, it only considers the channel condition, thus it cannot well suit RT traffics.

In this article, a novel cross layer scheduling algorithm is proposed for real time traffic in MIMO-OFDM systems, in which the users’ QoS requirements, queuing states observed at the MAC layer, and channel state information (CSI) observed at the PHY layer are considered. All the non-zero eigenmode subchannels are exploited to transmit data and the adaptive modulation and coding (AMC) is employed on each eigenmode subchannel, which can greatly improve the throughput of MIMO-OFDM systems. The proposed scheduling algorithm can dynamically allocate resources in space, time and frequency domain. As a result, the spatio-temporal-frequency resources of the system are utilized much more efficiently and more users can be supported. Meanwhile, users’ QoS requirements are met and fairness among them is guaranteed. Simulation results also indicate that the proposed algorithm can improve the cell throughput and increase the number of users that can be supported under the condition that the users’ QoS requirements are satisfied and long-term fairness is ensured.

The remainder of the article is organized as follows. In
2 System model

The downlink multiuser MIMO-OFDM system model is shown in Fig. 1. Equipped with \( N \) subcarriers, \( M_T \) transmits antennas at the base station (BS) and \( M_R \) receives antennas for each of \( K \) users served by the BS. At the transmitter, the data packets from the users are fed into a subcarrier allocation block, which allocates bits from different users to different subcarriers based on the algorithm proposed in this article. It is assumed that the CSI is known at the receiver and the transmitter, and the channel changes little during the transmission. The precoding is done on every subcarrier by beamforming based on SVD of the channel matrix of every subcarrier. Then the precoded symbols are sent to the inverse fast-Fourier-transformation (IFFT) module to conduct OFDM modulation and then cyclic prefix (CP) is added to every OFDM symbol. At last, the data symbols of every antenna are transmitted. At the receiver, the received signal is demultiplexed into \( J \) substreams. The transmitted symbol vector of \( J \) substreams is denoted as \( d = [d_1, d_2, ..., d_J]^T \). The corresponding channel matrix is \( H_{k,n} \) where 0 is the noise vector with i.i.d. complex Gaussian entries each having variance \( \sigma^2 \).

At the receiver, by decoding the received symbol vector \( r \) with \( \left(u_{k,n}^t\right)^H \), one can obtain the received data symbol of the \( j \)th substream:

\[
y_j = \left(u_{k,n}^t\right)^H r = \left(u_{k,n}^t\right)^H (H_{k,n}X + n) = \left(s_{k,n}^t(v_{k,n}^t)\right)^H \left(\sum_{j=1}^{J} v_{k,n}^t \sqrt{p_j d_j}\right) + \left(u_{k,n}^t\right)^H n = \left(s_{k,n}^t(v_{k,n}^t)\right)^H \left(\sum_{j=1}^{J} v_{k,n}^t \sqrt{p_j d_j}\right) + \left(u_{k,n}^t\right)^H n
\]  

(4)

From Eq. (4), one can see that by precoding the transmitted symbol vector at the transmitting antennas with the singular vector \( v_{k,n}^t \) and by decoding the received symbol vector at the receiver using singular vector \( u_{k,n}^t \), up to \( M \) parallel single-input single-output (SISO) eigenmode subchannels are constructed. Among different eigenmode subchannels, no crosstalk happens when CSI is known perfectly at the transmitter and the receiver. In this article, CSI is assumed to be known perfectly at the receiver and the transmitter and the channel changes little during transmission.

The signal to noise ratio (SNR) of user \( k \) on the \( j \)th eigenmode subchannel of subcarrier \( n \) can be expressed as:

\[
\text{SNR}_{k,n}^{j} = \frac{p_j \left(s_{k,n}^t\right)^2}{N_0 B}
\]  

(5)

where \( N_0 \) is the noise power spectrum density, and \( B \) is the subcarrier bandwidth.

According to the SNR on every eigenmode subchannel,
different modulation and coding scheme (MCS) of AMC is used. The mapping between the MCS and the required SNR threshold is shown in Table 1 [7]. It is assumed that no packet error happens if the SNR threshold is guaranteed. That is, packet loss only happens when the packet waiting time exceeds the maximum allowable delay.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Adaptive transmission schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>SNRReq/dB</td>
</tr>
<tr>
<td>MCS 1</td>
<td>2.2</td>
</tr>
<tr>
<td>MCS 2</td>
<td>5.2</td>
</tr>
<tr>
<td>MCS 3</td>
<td>7.6</td>
</tr>
<tr>
<td>MCS 4</td>
<td>10.9</td>
</tr>
<tr>
<td>MCS 5</td>
<td>14.5</td>
</tr>
</tbody>
</table>

According to the SNR\(_{k,n}\) and the MCS, one can obtain the number of bits mapped on the \(j\)th eigenmode subchannel of subcarrier \(n\) with user \(k\). Assume that the number of the bits is denoted as \(R_{k,n,j}\), and the number of bits of user \(k\) on subcarrier \(n\) is:

\[
R_{k,n} = \sum_{j=1}^{J} R_{k,n,j} \tag{6}
\]

Let \(\rho_{k,n}\) be the subcarrier allocation indicator. \(\rho_{k,n} = 1\) implies that subcarrier \(n\) is assigned to user \(k\). Otherwise, \(\rho_{k,n} = 0\). To guarantee that the subcarrier \(n\) is occupied by only one user, let \(\sum_{n=1}^{N} \rho_{k,n} = 1\) for all \(n\). Then the data rate of user \(k\) is denoted by:

\[
R_k = \sum_{n=1}^{N} \rho_{k,n} R_{k,n} = \sum_{n=1}^{N} \sum_{j=1}^{J} \rho_{k,n,j} R_{k,n,j} \tag{7}
\]

3 Algorithm description

The structure of the proposed scheduler is shown in Fig. 2.

![Fig. 2 The structure of the cross layer scheduler](Image)

On arriving at the BS, the packets from different users are buffered in separate queues, which are assumed to have infinite lengths. Within one queue, packets are served in a first-in first-out (FIFO) order. Across the queues, packets are served according to the proposed cross layer packet scheduling algorithm. The cross layer scheduling algorithm allocates resources dynamically based on users’ QoS requirements, queuing states observed at the MAC layer, and CSI observed at the PHY layer. In what follows, the priority metric \(\mu_i(t)\) of the proposed cross layer scheduling algorithm is described in detail.

3.1 The CSI factor

The subchannels may experience different attenuation with different users. A channel quality indicator (CQI) is utilized to provide CSI from user terminals to the base station scheduler. A novel scheduling strategies should be proposed to adaptively transmit data and dynamically assign wireless resources based on CSI. The key idea is to choose a user with good channel conditions to transmit packets. Taking advantage of the independent channel variation across users, the scheduling can substantially improve the network performance through multiuser diversity.

The famous channel-aware-only scheduling is PF [6]. It utilizes asynchronous channel variation to improve the overall system throughput while guaranteeing fairness among users. In the proposed algorithm, the channel condition is considered based on the following equation:

\[
f_i(CQI_i(t)) = \frac{\min(R_i(t), Q_i(t)/T_i)}{R_i(t)} \tag{8}
\]

where \(R_i(t)\) is the current data rate that the BS can support, \(\bar{R}_i(t)\) is the average rate received by user \(i\) over a window of appropriate size, \(Q_i(t)\) denotes the number of untransmitted bits in the queue of user \(i\) at time \(t\) and \(T_i\) is the length of the time slot. \(R_i(t)\) represents the current channel condition of user \(i\). \(\bar{R}_i(t)\) is the rate corresponding to the mean fading level of user \(i\). Both \(R_i(t)\) and \(\bar{R}_i(t)\) are time-varying with adaptive modulation depending on the channel condition. For each user \(i\), the average rate is updated in each timeslot by

\[
\bar{R}_i(t+1) = \frac{1}{t_i} \bar{R}_i(t) + \frac{1}{t_i} R_i(t) \tag{9}
\]

For a user who is not currently receiving service, \(R_i(t) = 0\). Users who have no data to send can also have their average rate updated.

3.2 The QoS factor

Two QoS parameters, packet loss rate (PLR) and packet delay are considered. The PLR occurring in real time traffic is defined as the sum of the packet error rate (PER) resulting from channel impairments and packet dropping rate (PDR) calculated from packets exceeding the maximum allowable delay \(W_{\max}\) in each traffic. It is assumed that the channel condition is well estimated and predicted, thus the PER is ignored. The PLR should be less than a certain determined
threshold, i.e., \(PLR_{req,i} \leq PLR_{th,i}\), for user \(i\).

As for the packet delay, \(W_i(t)\) denotes the head-of-line (HOL) packet delay in the queue of user \(i\) and \(W_{max,i}\) is the maximum allowable delay. In case the HOL packet delay for the RT traffic exceeds \(W_{max,i}\), the packet will be dropped out, and thus the corresponding QoS requirement may not be met. In this case, Eq. (11) should be guaranteed.

\[W_i(t) < W_{max,i}\] 

Therefore, the QoS factor can be represented by:

\[f_i (QoS_i(t)) = \frac{PLR_{req,i}}{PLR_{req,i} \cdot W_i(t) / W_{max,i}}\] 

### 3.3 The QSI factor

The queue status observed at the MAC layer is an important factor that affects the performance of the scheduler. The QSI factor can be given by:

\[f_i (QSI_i(t)) = \frac{Q(t)}{\bar{Q}(t)}\] 

where \(\bar{Q}(t) = (1/K) \sum_i Q(t)\). This factor indicates that the longer the queue length of the user \(i\), the higher priority user \(i\) has.

### 3.4 The cross layer scheduler

Therefore, for every subchannel at time slot \(t\), the priority metric of the proposed cross layer scheduler is given as follows:

\[\mu_i(t) = f(CQI_i(t), QoS_i(t), QSI_i(t)) = f_i (CQI_i(t)) \cdot f_i (QoS_i(t)) \cdot f_i (QSI_i(t)) = \frac{\min(R(t), Q(t)/T_i)}{R(t)} \cdot \frac{PLR_i \cdot W_i(t) / W_{max,i}}{PLR_{req,i} / Q(t)}\] 

The proposed cross layer algorithm is a jointly channel, QoS and queue-aware scheduling scheme. For each subchannel, the user with the highest priority metric is preferentially served at each scheduling slot, i.e., the MAC scheduler calculates a priority metric per user on each subchannel and chooses a user with the maximal value of the priority metric for service on the subchannel. When user \(i\) has larger current packet delay, higher packet loss rate and channel quality relative to its average level, and longer queue length, it will have greater chance of being scheduled.

### 4 Simulation parameters

#### 4.1 Traffic model

For the system level simulation, the video streaming model is considered. A video streaming is modeled as a variable bit rate class characterized by Pareto distribution. The modeling is composed of continuous video frames, where each video frame is divided into a fixed number of video packets [8]. Moreover, the size of a video packet is determined by Pareto distribution. The specific parameters of a video streaming model whose generation rate is 32 kb/s are listed in Table 2, and an example of video streaming model is shown in Fig. 3. The maximum allowable delay of video streaming is 100 ms and the requested packet loss rate is 0.01.

#### 4.2 System parameters

A radio cell with \(K\) active users and a centralized BS is considered. Inter-cell interference is not ignored. All the users are uniformly distributed in the cell. A summary of simulation parameters for system model is shown in Table 3.
To limit control signaling, the scheduling is not implemented at a subcarrier granularity. The minimum resource unit of scheduling is a time-frequency block corresponding to one time slot (1 ms) and 16 continuous subcarriers. Each resource block is used by only one user. Within each block, the same modulation and the same power allocation is used for the corresponding eigenmode subchannels on each of the 16 subcarriers. Estimation of the channel gain and SNR of the eigenmode subchannel is based on the average of all 16 subcarriers.

5 Simulation results

The performance of the proposed algorithm is compared with the PF algorithm in terms of cell throughput, average packet delay and packet loss rate.

Fig. 4 indicates that the cell throughput of the proposed scheduling is higher than PF. This is because the PF only considers the channel condition when the packets of users under bad channel condition exceed the maximum delay and gets lost. Fig. 4 also shows that the cell throughput of the proposed algorithm is increasing before the upper bound of the transmission capability is reached with the increase of the number of users. After the system load is beyond the upper bound of the system capability, the cell throughput is decreasing with the increase of the number of users. This is because the proposed algorithm guarantees user fairness and thus each user can gain fair chances. Consequently, every user has fewer chances of transmitting data, and thus their throughput is reduced. Besides, user fairness may also result in decrease of the cell throughput because users under bad channel conditions might be selected for transmission.

From Fig. 5, one can see that the proposed scheduling algorithm has the larger average packet delay. This is because the proposed algorithm considers more factors to transmit as many packets as possible successfully.
6 Conclusions

A novel cross layer scheduling is proposed for real time traffics in multiuser downlink MIMO-OFDMA systems. To obtain full MIMO capacity, AMC is adopted on every eigenmode subchannel. Subchannels are allocated dynamically based on CSI, users’ QoS requirements and the queuing state of each user. The proposed scheduling algorithm improves cell throughput significantly at the expense of increasing packet delay while satisfying the QoS requirements of the RT traffic. Besides, this algorithm also ensures fairness among users and increases the number of supported users.

Acknowledgements

This work was supported by the Hi-Tech Research and Development Program of China (2006AA01Z283, 2006AA01Z260), and the National Natural Science Foundation of China (60496312).

References

8. IEEE C802.16m-07/088r2. Draft IEEE 802.16m evaluation methodology document. 2007

(Editor: ZHANG Ying)