Joint Channel and Delay Aware User Scheduling for Multiuser MIMO system over LTE-A Network

Jayeta Biswas[1] Ren Ping Liu[2] Wei Ni[2]

Iain B. Collings[2] Sanjay K. Jha [1]

[1] University of New South Wales, Australia
{jbiswas,sanjay}@cse.unsw.edu.au.com
[2] ICT Centre, CSIRO, Australia.
{ren.liu,wei.ni,iain.collings}@csiro.au

Technical Report
UNSW-CSE-TR-201227
October 2012

THE UNIVERSITY OF NEW SOUTH WALES

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

The increasing popularity of mobile video applications has generated huge demand for higher throughput and better quality of service (QoS) in wireless networks. Multiuser multiple input multiple output (MU-MIMO) is the most promising technology to boost the throughput in the Long Term Evolution - Advanced (LTE-A) networks. However, most existing LTE-A proposals achieve the MU-MIMO capacity gain by selecting orthogonal user sets without considering QoS requirement, such as delay. Such user selection may seriously degrade the QoS performance of delay sensitive applications, such as mobile video. In this paper, we propose a joint channel and delay aware user scheduling algorithm, which selects users based on both channel conditions and weighted delay. The weight factor is designed to capture the historic rate and maximum coding rate of a video stream. Simulation results show that the proposed algorithm reduces the average delay by up to 30% with a marginal 2% sacrifice of throughput, compared to previous work. It also reduces delay variations and improves fairness among the users.

1 Introduction

Prevailing high-speed mobile applications, such as multimedia streaming, IPTV and high-definition online gaming, are demanding higher bandwidth in wireless networks. A Cisco report [2] predicted that the demand for wireless data will increase 18-times from 2011 to 2016. A large part of the increase comes from video traffic. Multiuser multiple input multiple output (MU-MIMO), as one of the key technologies in the Long Term Evolution - Advanced (LTE-A) networks, promises high spectral efficiency and increased bandwidth [4]. In a LTE-A network, the base station, evolved NodeB (eNB), is able to transmit to several user equipment (UE) in the same time frequency slot, by exploiting multiuser spatial diversity using MU-MIMO technology. Although this simultaneous transmission increases the information theoretic capacity of the channel, the achievable user rate is limited by the spatial separation of the UEs. By choosing the spatially disjoint optimal set of users, throughput gain of the downlink transmission can be maximized. However, such user selection strategy may not satisfy the delay requirement of individual users. In this paper, we develop an efficient user selection algorithm that takes advantage of the throughput gain of MU-MIMO technology, and supports user delay requirements at the same time.

When a large number of users are present in a cell, the probability of getting a spatially separated user set becomes high. Therefore, the multiuser diversity becomes a form of selection diversity [20]. Determining an optimal user set requires an extensive search over the entire user set. In [10], brute force approach is used to get the optimal user set. This strategy is not scalable if a large number of users are served by a base station. A suboptimal algorithm called semi orthogonal user selection (SUS) is proposed in [20]. In [13], Razi et al. introduced a low complexity algorithm called greedy rate maximization (GRM) algorithm with the objective of maximizing the capacity region. This algorithm reduces the complexity of SUS algorithm. However, all these aforementioned studies only deal with the PHY layer constraints without considering the higher layer QoS of the users.

For non ideal channel state information, a near optimal opportunistic scheduling algorithm is proposed in [16]. In this opportunistic algorithm, the base station dynamically selects users so as to maximize a desired fairness utility measure that is a concave function of long term average transmission rate. In this work, fairness is considered only for transmission rate. A queue aware user grouping policy is investigated in [17]. Here, a throughput optimal objective function is established which selects a subset of users according to their queue states and geometric channel vector dependency. In [18], Torabzadeh et al. proposed a MU-MIMO user scheduling algorithm based on the head of line (HOL) packet length and the queue length. However, both [17, 18] don't consider QoS requirement of applications. In [19], another delay based rate scheduler is designed which uses the weighted waiting delay of each user's packet. In this algorithm, the number of users that can transmit simultaneously is limited by two. When large number of users need to be served, or more than 2 antenna is available, the complexity of the system will increase. For LTE-A downlink MU-MIMO mode, an approximation algorithm is proposed in [9] where the spatial channel gain and long-term throughput of the users are considered as the fairness criteria. However, other QoS parameters are not considered.

To implement the MU-MIMO for QoS sensitive application, we need to achieve two conflicting goals, i.e. delay constraint and system throughput. None of the previous scheduling schemes for MU-MIMO have considered the delay requirement and the traffic arrival characteristics. Although extensive research has been done to design a perfect precoder for implementing the MU-MIMO in LTE-A network [4], the selection of user set remains a issue [11]. If users are selected solely based on orthogonality, the overall system throughput will increase but the delay requirement of the users may not be satisfied.

Delay sensitive applications, such as mobile videos, have very distinct QoS requirement. There are several application layer mechanisms which focused on new video encoding to adapt to the available bandwidth. The adaptations to the wireless channel are done by switching the sending rate as well as adjusting the quality of the video[7]. For instance, if a user running video application is not scheduled for several scheduling period because of its channel condition, the adaptive mechanism will predict bandwidth fall and send low quality video even if it is selected for the next time period. Therefore, this adaptation will not work in MU-MIMO scenario.

To meet the delay requirement of the users we propose a cross layer user scheduling algorithm for Scalable video coding (SVC) coded video transmission over MU-MIMO system. The proposed algorithm selects the user based on three factors: 1) the estimated throughput, 2) HOL waiting delay of the flows, and 3) the traffic arrival rate. Since we are interested in delay sensitive video applications, we also consider the dependency among different temporal layer of the coded video frames.

The contributions of this paper are summarized below.

- We propose a delay based cross layer user selection mechanism, which is specifically designed for delay sensitive video applications to utilize the increased capacity gain of MU-MIMO system over LTE-A network.
- Unlike the sum rate maximization algorithm, our delay based algorithm selects user based on the weighted delay experienced by the users and the estimated throughput.
- The weight factor of the weighted HOL delay is explicitly designed to capture the effect of past average history of bit stream generation rate and the maximum supported decoding rate of the coded video.
- We validate our design by simulation. The performance of the algorithm is evaluated in terms of the average delay experienced by the user's application, delay fairness among the users and the achieved system throughput.
- The performance of the system is compared to the well known SUS [20] user selection algorithm and the queue length based algorithm [17]. Performance comparison results show that the proposed algorithm reduces the average delay by up to 30% with a marginal 2% sacrifice of throughput. It also reduces delay variations and improves fairness among the users.

The paper is organized as follows. In Section 2, we present the system model. In Section 3 we explicate the characteristics of the video content. In Section 4 we describe our proposed scheduling algorithm. The performance of the proposed controller is evaluated in Section 5. Our contributions are summarized in Section 6.

2 System Model

We consider a LTE-A network where a single eNB with N_T transmit antennas serve K users, each of them equipped with a single receive antenna. In this system, a video server is connected to the eNB through Evolved Packet Core (EPC). The video server will transmit precoded video stream with variable bit rate (VBR) to multiple users. Different users are characterized by different channel condition.

2.1 Physical Layer Model

The current version of LTE advanced, can support up to 8-by-8 MIMO configuration [4]. In case of LTE, the entire bandwidth is divided into 180 KHz physical resource block (RBs) each of 0.5ms duration with 6 or 7 symbol in time domain and 12 subcariers in the frequency domain [12]. In MU-MIMO mode, spatial dimension is added to the resource block. Thus N_T users can be assigned to the same resource block. The minimum allocation unit of a frame scheduled by the scheduler is a pair of resource block having length of one subframe (1ms). Therefore, at each transmission time interval (TTI) (1ms) for each resource block, multiple users are scheduled depending on the number of antennas N_T. Here, we assume the transmit power to be equally divided among the eNB antennas.

The received signal vector of user K at time t is expressed as

$$y_{k,n} = \boldsymbol{h}_{k,n}^{\dagger} \sqrt{\boldsymbol{P}_n} \boldsymbol{x}_n + z_{k,n}, \qquad k = 1, \ldots, K, \tag{2.1}$$

where $\boldsymbol{h}_{k,n}$ is the channel vector for the k-th user at subcarrier n, $\boldsymbol{x}_n$ is the transmitted symbol vector of the eNB on subcarrier n and $z_{k,n}$ is the additive complex Gaussian noise with variance $\sigma_{k,n}^2$ on subcarrier n at the k-th user and $\boldsymbol{P}_n = \mathrm{diag}(P_1, ..., P_k)$ is the transmission power for subcarrier n. Note that $\boldsymbol{x}_n$ is the output for all the users receiving the signal on subcarrier n. Thus, it is not subscripted with k. Here, $\boldsymbol{M}^{\dagger}$ denotes the conjugate transpose of $\boldsymbol{M}$. In this system model, we make use of the zero forcing (ZF) liner precoding [20] at the base station. The transmit spatial filter is represented by a weight vector $\boldsymbol{w}_{k,n}$ mapping the k-th user data symbol on subcarrier n, $b_{k,n}$, to the base station antenna array outputs. Given that K' users are selected for transmission, the output signal is given by [14]

$$\boldsymbol{x}_n = \sum_{k=1}^{K'} b_{k,n} \boldsymbol{w}_{k,n}, \ \forall \ n. \tag{2.2}$$

In each scheduling period, the ZF precoding matrix $\boldsymbol{W}_n(t)$ of the N_T base station antenna to the K' selected users is calculated by

$$\boldsymbol{W}_n(t) = \boldsymbol{H}_n(t) \big(\boldsymbol{H}_n^{\dagger}(t) \boldsymbol{H}_n(t) \big)^{-1},$$

2.2 MAC Layer Model

The medium access control (MAC) layer of LTE eNB is responsible for scheduling the arriving higher layer packets. To satisfy the QoS of users running different applications, a bearer is associated as an end to end connection identifier for

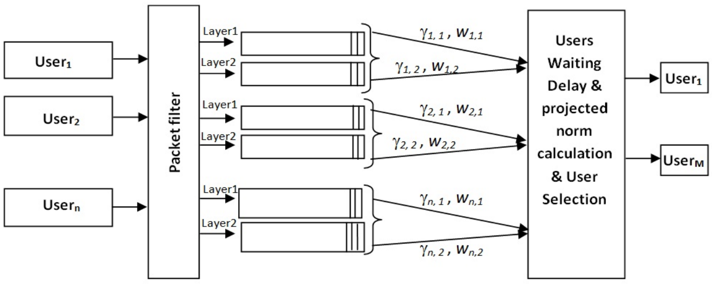

Figure 2.1: eNB User scheduler

each application requiring a particular scheduling consideration [5]. Each bearer is characterized by the combination of QoS class identifier and IP address of the user device. Hence, depending on the application requirement, multiple bearers can be associated with each user. At the eNB, there are separate queues for each bearer. A packet filter at the base station classifies the arriving packet and forwards it to an appropriate queue. Since, the capacity gain of MU-MIMO depends on the orthogonality of the users, the eNB scheduler needs to consider the user's channel vector at the MAC layer to make the scheduling decision. The realization of the user's channel vector extensively depends on the availability of perfect channel state information (CSI) at the eNB. Here, we assume that the base station has full knowledge of the channel information.

3 Video Content Characteristics

SVC is a video coding mechanism that encodes the video in such a way to adapt the various needs or preference of end users as well as to varying terminal capabilities or network conditions. There are three types of scalabilities [15]: spatial, temporal and quality. In temporal scalability mode of SVC, the reference frame (I or P) makes the base layer bit stream and the bi directionally predicted B frame makes the enhancement layer. Minimum quality of the video is guaranteed by timely delivery of the base layer bit stream. The successful decoding of the B frames depends on the previous and next I or P frame. For both of the layered bit stream, the ratio of bit stream generation rate and delivery rate is a relative measure of how much bit of the corresponding layer is pending for displaying. If the ratio for base layer bit is greater than that of the enhancement layer then the user will not have enough base layer frames to decode the already received enhancement layer frames. Therefore, at any time the ratio for enhancement layer needs to be at least equal to the ratio for base layer. The following equation represents this relationship.

$$\frac{R_b}{r_b} \leq \frac{R_e}{r_e} \tag{3.1}$$

where R_b and R_e are respectively the average bit stream generate rate (encoding rate) of the base layer and the enhancement layer at any time t. Accordingly, r_b

and r_e are respectively the bit stream delivery rate from the MAC layer queue, for base layer bit stream and enhancement layer bit streams.

4 User Scheduling

In our proposed user set selection mechanism, we not only consider channel correlation between the users to be selected, but take the intrinsic characteristics of SVC into account as well. At the base station, different layered video stream is mapped to different flow and forwarded to different queue so that they can be processed differently. The type of service field (TOS) of the IP packet header is used to indicate the frame type. Hence, the eNB packet filter can detect the frame type. The architecture of a downlink scheduler is shown in Fig. 2.1. At each scheduling period, a weight factor $\gamma_{k,f}$ is associated with each queue corresponding to flow f of user k to provide different priorities for different flow dynamically. In LTE, the EPS bearer are categorized into three major priority classes. These are guaranteed bit rate (GBR), non GBR and default bearer. For GBR video streaming application, when the corresponding queue of flow f of user k is non-empty, the weight factor $\gamma_{k,f}(t)$ at any scheduling period t is updated as

$$\gamma_{k,f}(t) = \frac{R_{k,f}(t) - r_{k,f}(t)}{R_{k,f}(t)}. \tag{4.1}$$

Here $R_{k,f}(t)$ and $r_{k,f}(t)$ denote respectively the average bit stream generate rate and the bit stream transmission rate from the MAC layer queue of flow f of user k at time t. For empty queue the weight factor is set to zero. By using (4.1), $\gamma_{k,f}(t)$ will be set dynamically within the range of [0,1] for each of the flow based on the relative measure of the bit stream generate rate and delivery rate. It ensures that the weight for the queue with more pending base layer frame will be higher than the queue with less pending enhancement layer frames. Most of the traditional schedulers select the weight factor as the queue length, to prevent queues from becoming very large. As this policy selects user with largest queue length, the expected delay for a user may be large even if the arrival rate is small. Hence, we consider the arrival rate to define the weight factor. At each scheduling period, the average bit stream transmission rate as well as the supportable decoding rate from the queue of flow f of user k , $r_{k,f}(t)$ is updated as

$$r_{k,f}(t+1) = \begin{cases} (1-\alpha)r_{k,f}(t) + \alpha r^*_{k,f}(t) & , k \in U(t) \\ (1-\alpha)r_{k,f}(t) & , k \notin U(t) \end{cases} \tag{4.2}$$

Here, $r^*_{k,f}(t)$ represent the actual bit stream transmission rate of flow f of the k-th user at the t-th scheduling period and $U(t)$ is the set of selected user. And $\alpha = \frac{1}{T_c}$ where T_c is the number of scheduling period which control the averaging window of the moving average. $r^*_{k,f}(t)$ is represented by

$$\begin{aligned} r^*_{k,f}(t) = &\min(y_{k,f}(t)/[y_{k,1}(t) + ... + y_{k,F}(t)] \\ &\times \log_2(1 + \frac{P}{|U(t)|}||\boldsymbol{g}_k||^2), y_{k,f}(t)). \end{aligned} \tag{4.3}$$

where $y_{k,f}(t)$ is the required transmission rate for flow f of user k at time t. Similarly the measure of arrival history as well as the average bit stream generate rate is updated as

$$R_{k,f}(t+1) = (1-\alpha)R_{k,f}(t) + \alpha R^*_{k,f}(t) \tag{4.4}$$

Where $R^*_{k,f}(t)$ represent the arrival rate to the queue of flow f of user k at the t-th scheduling period.

One UE may have multiple flows with different weight factor. The weight factor provides QoS to the video application so that each user can decode different layered frames properly and provide smooth playback. While selecting the users, the scheduler makes use of the HOL waiting delay $hol_{k,f}(t)$ of each flows and the weight factor corresponding to each flow. It ensures that the probability of selecting a user with high traffic arrival rate and long waiting delay is higher than a user with low traffic arrival rate and long waiting delay. The waiting delay of each user, $HOL_k(t)$ is obtained by the weighted sum of its HOL waiting delay of each flow instead of using the sum of HOL. The waiting delay of each user $HOL_k(t)$ is expressed as

$$HOL_k(t) = \sum_{f,k} \gamma_{k,f}(t) \times hol_{k,f}(t) \tag{4.5}$$

In MU-MIMO the transmission rate cannot be determined until all the users are selected. While selecting the i-th user, for each candidate user the transmission rate can be estimated from the channel vectors of the already selected users. Let $U(t)$ is the set of already selected users which is empty at the beginning. At each iterative step of the scheduling algorithm, one user is selected and added to the set $U(t)$. Let $\boldsymbol{h}_k$ is the channel vector of the the k-th user. The projected norm of the k-th user $\boldsymbol{g}_k$ with respect to the already selected user is determined by

$$\boldsymbol{g}_k = \boldsymbol{h}_k - \sum_{j \in U(t)} \frac{\boldsymbol{h}_k \boldsymbol{g}_j}{||\boldsymbol{g}_j||^2} \boldsymbol{g}_j. \tag{4.6}$$

When $U(t)$ is empty $\boldsymbol{g}_k = \boldsymbol{h}_k$. The achievable transmission rate for the k-th user is represented by $\log_2(1 + \frac{P}{|U(t)|} \times ||\boldsymbol{g}_k||^2)$. Where P is the total transmission power. Further details of projected norm and achievable transmission rate are given in [20]. For each scheduling period t the scheduler selects each user according to the waiting delay and the projected norm of the users. The selection of i-th user of $U(t)$ is given as

$$U_i(t) = \arg\max_k (HOL_k(t) \log_2(1 + \frac{P}{|U(t)|} ||\boldsymbol{g}_k||^2)). \tag{4.7}$$

subject to

$$\log_2(1 + \frac{P}{|U(t)|} ||\boldsymbol{g}_k||^2) \geq R_{min}. \tag{4.8}$$

Thus, the scheduler selects the user having largest product of the weighted waiting delay and the achievable rate. The achievable rate needs to be higher than the minimum rate requirement. Therefore, if the user has large waiting delay but estimated transmission rate is less than the minimum, it is not selected. This ensures that the transmission rate of the already selected users will not be less than the minimum value.

Table 5.1: Characteristics of the video source

Parameters	Matrix III trace
Resolution	352x288
Codec	MPEG-4 Part 2
Min Frame Size(Bytes)	8
Max Frame Size(Bytes)	36450
Mean Frame Size(Bytes)	3189.068
Display Pattern	IBBPBBPBBPBB
Group of Picture size	12
Frame rate(fps)	25

5 Performance Evaluation

5.1 Simulation Environment

Our simulation is based on MU-MIMO system over LTE network. We used OPNET Modeler and Matlab to develop a simulation model. The eNB is equipped with $N_T = 4$ transmit antenna. The antennas are configured as described in [8]. The total transmit power at base station is $P = 0.5W$. $K = 25$ users are randomly distributed around the base station. The received SNR for all user is not identical. Depending on the channel condition it varies between 5dB to 20dB. We have created different scenarios by varying the total number of users. Each user is equipped with a single receive antenna. We use 10 MHz Frequency Division Duplex (FDD) PHY profile of the LTE [1]. As a modulation scheme QPSK is used. OPNETs LTE module does not support MU-MIMO. Therefore, we integrate Matlab to do the antenna weight vector calculation and other complex calculation for the proposed scheduling algorithm.

We configured each UE to sequentially request for a video clip from a remote server. The server is connected with the eNB through EPC router. The traffic source is a MPEG-4 Matrix III movie trace [3]. The properties of the video trace are shown in table 5.1. Although each user is requesting the same video, they are served by the unicast flow rather than the broadcast service. In real life, they may use different videos but we use this method for comparison.

5.2 Simulation results

In this section, the performance of our proposed scheduling algorithm is evaluated for various simulation settings. We compare our delay based MU-MIMO scheduling algorithm with the SUS algorithm and the queue length based algorithm where the weight factor in (4.7) is chosen as the sum of queue length of each flow f of the k-th user.

Fig. 5.1 shows the average delay experienced by users for different number of users. The delay is averaged over 10 minute simulation runs. It illustrates that the proposed algorithm achieves lower delay than that of the SUS algorithm and the queue length based algorithm when the total number of users is large or the traffic intensity increases. On the other hand, when traffic load is not high, each of these algorithms experience almost same delay. When the number of users is small, each of the algorithms selects the same set of users most of the time. As

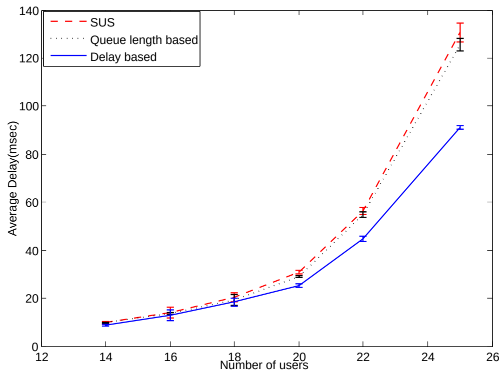

Figure 5.1: Average delay performance, illustrating the delay reduction for the delay based algorithm

Table 5.2: Throughput Performance

Number of users	10	14	20	25
Delay based throughput(Mbps)	4.61	7.0	9.4	12.45
Queue length based throughput(Mbps)	4.77	7.25	9.42	12.32
SUS throughput(Mbps)	4.87	7.4	9.63	12.57

the system is not fully loaded, the HOL packet does not experience much delay and does not cause much effect on user selection. For the queue length based algorithm, the delay reduction is not significant because it gives preference to the user with largest queue length at the recent scheduling period. Therefore, the delay is reduced for the users with higher traffic arrival rate but users with low traffic arrival rate may experience longer delay. Table 5.2 reveals the throughput resemblance of the delay based algorithm to the throughput optimal SUS and Queue length based algorithm. As the number of users increases, the multiuser diversity provides more options to the scheduler. Therefore, the throughput loss for our delay based algorithm becomes 1% when the total number of users is 25. On the other hand, the queue length based algorithm is seen to have 2% throughput loss for all load condition.

Fig.5.2 shows the delay of each individual user, when the total number of user in the system is 25. It is observed that, the best case user (index-5) having largest channel norm without any shadow fading is experiencing very short delay for both of the SUS and queue length based algorithm. On the other hand, the worst case user (index-12) having worst channel condition is facing long delay for SUS algorithm. However, both of the SUS and queue length based algorithms fail to provide fairness. As the delay based algorithm, considers the waiting delay of the flows and the bitstream transmission rate, all the users are experiencing reduced delay. Furthermore, no user is getting high delay penalty because of bad channel condition.

For delay sensitive video application, delay jitter is another factor that affects the user's satisfaction. We measure the end to end delay variation of each frame. Fig. 5.3 shows the average end to end delay variation of users. When the load becomes high, the delay based algorithm achieves smaller delay variation than

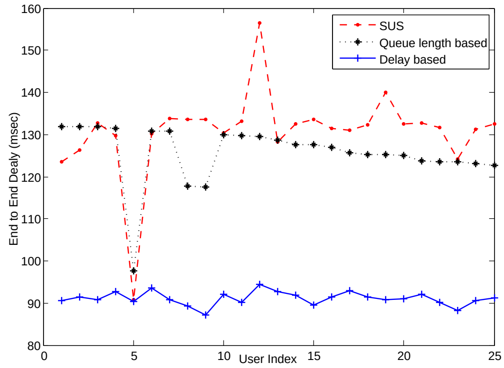

Figure 5.2: End to End delay of individual users. Traffic load Avg: 12.3Mbps, Max: 40Mbps, Min: 0.4Kbps

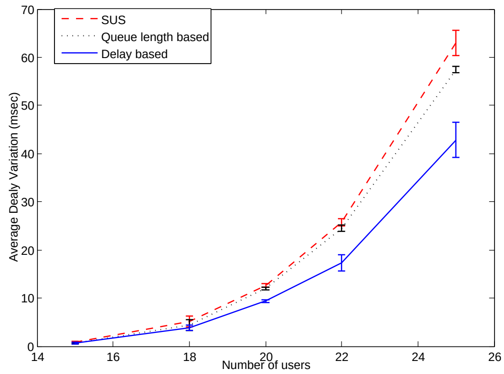

Figure 5.3: Average delay jitter

the other two algorithms. Even though some user have same average delay for each of the algorithms, the delay based algorithm achieves smaller delay variation. In SUS, the user set selected for transmission varies a lot due to orthogonality requirements. Therefore, the delay variation is larger compared to our new scheme. The same is true for queue length based algorithm as it does not take the past traffic arrival characteristics into account.

In order to measure the fairness of the proposed algorithm, we use the Jain fairness index [6]. This index is represented as

$$fairness\, index = \frac{(\sum_{i=1}^{K} v_i)^2}{K \times \sum_{i=1}^{K} v_i^2} \tag{5.1}$$

where v_i is the average delay experienced by the i-th user and K is the total number of users in the system. According to (5.1) the fairness index varies from 0 to 1. The ideal value is 1.

Fig. 5.4 shows the fairness among the users under various loads. In all load condition the delay based algorithm, achieves a fairness value that is close to 1. In SUS algorithm, the fairness index decreases gradually because the user's delay become more unbalanced when the load increases. Although, the queue

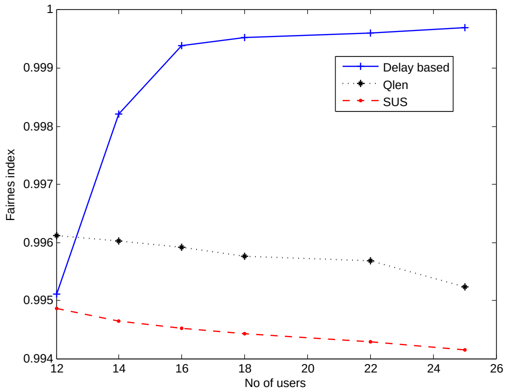

Figure 5.4: Fairness Index

length based algorithm has slightly higher fairness index compared to the sus algorithm, it is still lower than the delay based algorithm because user with low traffic arrival still needs to wait long time.

6 Conclusions

In this paper we have investigated the joint channel and delay aware scheduler design for MU-MIMO system over LTE network. For making the decision of selecting an user set, our proposed scheduler jointly consider the weighted waiting delay information from the MAC layer and the channel state information from the PHY layer. The simulation results exhibit that in our proposed algorithm, delay sensitive video application of each user achieves better delay performance without any throughput loss compared to the previously well-known MU-MIMO schedulers. Furthermore, the fairness of our proposed algorithm outperforms the aforementioned other algorithms.

Bibliography

[1] *3GPP Tech. Specif. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, 3GPP TS 36.101.*

[2] Cisco visual networking index: Global mobile data traffic forecast update, 20112016, 2012.

[3] G. Auwera, P. David, and M. Reisslein. Traffic characteristics of H.264/AVC variable bit rate video, march 2008.

[4] Jonathan Duplicy, Biljana Badic, Rajarajan Balraj, Rizwan Ghaffar, Peter Horvath, Florian Kaltenberger, Raymond Knopp, IstvanZ Kovacs, HungT Nguyen, Deepaknath Tandur, and Guillaume Vivier. MU-MIMO in LTE systems. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):496763, 2011.

[5] H. Ekstrom. QoS control in the 3GPP evolved packet system. *Communications Magazine, IEEE*, 47(2):76–83, february 2009.

[6] R. Jain, D.-M. chiu, and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer system. Technical Report TR-301, Digital Equipment Corp., 1984.

[7] Guang Ji and Ben Liang. Stochastic rate control for scalable VBR video streaming over wireless networks. In *GLOBECOM 2009. IEEE*, pages 1 –6, 30 2009-dec. 4 2009.

[8] V.V. Kulkarni, J. Biswas, Ren Ping Liu, I.B. Collings, and S.K. Jha. Robust power allocation for MIMO beamforming under time varying channel conditions. In *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pages 1 –5, sept. 2011.

[9] S. Lee, I. Pefkianakis, S. Choudhury, S. Xu, and S. Lu. Exploiting spatial, frequency, and multiuser diversity in 3GPP LTE cellular networks. *Mobile Computing, IEEE Transactions on*, PP(99):1, 2011.

[10] Lifang Li and A.J. Goldsmith. Capacity and optimal resource allocation for fading broadcast channels .i. ergodic capacity. *Information Theory, IEEE Transactions on*, 47(3):1083–1102, mar 2001.

[11] Lingjia Liu, Runhua Chen, S. Geirhofer, K. Sayana, Zhihua Shi, and Yongxing Zhou. Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO. *Communications Magazine, IEEE*, 50(2):140 –147, february 2012.

[12] G. Piro, L.A. Grieco, G. Boggia, and P. Camarda. A two-level scheduling algorithm for QoS support in the downlink of LTE cellular networks. In *Wireless Conference (EW), 2010 European*, pages 246 –253, april 2010.

[13] A. Razi, D.J. Ryan, I.B. Collings, and Jinhong Yuan. Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding. *Wireless Communications, IEEE Transactions on*, 9(1):356 –365, january 2010.

[14] Z. Rosberg, A. Cantoni, and R.P. Liu. Resource allocation for QoS multiuser MIMO with zero forcing and MMSE beamforming. In *Quality of Service (IWQoS), 2010 18th International Workshop on*, pages 1 –6, june 2010.

[15] T. Schierl, T. Stockhammer, and T. Wiegand. Mobile video transmission using scalable video coding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(9):1204–1217, sept. 2007.

[16] Hooman Shirani-Mehr, Giuseppe Caire, and Michael J. Neely. MIMO downlink scheduling with non-perfect channel state knowledge. *Communications, IEEE Transactions on*, 58(7):2055 –2066, july 2010.

[17] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell. Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel. *Communications*, oct 2004.

[18] M. Torabzadeh and W. Ajib. Packet scheduling and fairness for multiuser MIMO systems. *Vehicular Technology, IEEE Transactions on*, 59(3):1330–1340, march 2010.

[19] Cheng Wang and R.D. Murch. Optimal downlink multi-user MIMO cross-layer scheduling using hol packet waiting time. *Wireless Communications, IEEE Transactions on*, 5(10):2856–2862, oct. 2006.

[20] Taesang Yoo and A. Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *Selected Areas in Communications, IEEE Journal on*, 24(3):528–541, march 2006.