A Priority Based Resource Allocation Algorithm for Transmission Delay Optimization in TD-LTE-A System

Bingchan Fan[1, a] Wenbin Liu[1,b]

[1]Wuhan Technology and Business University, Wuhan, 430065,P.R. China

[a]fanbc1989@126.com [b]cq2200@126.com

Keywords: TD-LTE-A; Transmission Delay Optimization; Scheduling; Resource Allocation; QoS

Abstract. Traditional resource allocation algorithms mainly focus on the maximization of system capacity. However, transmission delay is a crucial performance index for urgent communications in TD-LTE-A System. In this paper, a priority based resource allocation algorithm aiming at optimizing transmission delay of users was proposed. The scheduling priority was analyzed by considering the channel quality characteristics and packet delay of users. And a resource allocation scheme was proposed in which the resources were assigned to users according to the order of users' scheduling priority. By comparing with the conventional resource allocation algorithms, the simulation results verified that the proposed algorithm can achieve better delay performance, while improving the system throughput.

Introduction

With the rapid development of world wireless communication networks, the fourth generation wireless communication is expected to support a variety of high-speed communication services with diverse QoS (Quality of Service) requirements. In this situation, TD-LTE-Advanced (TD-LTE-A) system is proposed as the 4th generation wireless communication system and it can achieve 1Gbps downlink peak date rate by adopting Enhanced Multiple-Input Multiple-Output (MIMO), Carrier Aggregation (CA) and other techniques. However, due to the limited radio resource, time-varying wireless channel and resource competition among users, the QoS of traffics among users is hard to satisfy. To cope with this challenge, the resource allocation scheme is adopted to meet the QoS requirement of multiple users [1].

The orthogonal frequency-division multiplexing (OFDM) has been identified as wireless interface transmission scheme for the TD-LTE-A systems. OFDM can cater the inter-symbol interference and frequency selective fading by dividing the total bandwidth into several narrow subcarriers. The multiple access scheme adopted in TD-LTE-A system is the orthogonal frequency division multiplexing access (OFDMA) where the subcarriers can be assigned to different user. At a given time, a subcarrier that is in deep fade for a user may be in a good channel condition for other users at the same time. So users with better channel condition can be assigned that subcarrier, which eventually result in multi-user diversity. The problem of resource allocation in OFDMA based systems have been recently investigated exclusively. Most traditional resource allocation algorithms mainly focus on the maximization of system capacity. However, in urgent communications system like Ambulance, fire departments and transportation, the transmission delay is a crucial performance index.

In this paper, aiming at decreasing the packet delay of users in TD-LTE-A system, an efficient combined scheduling and resource allocation algorithm is proposed. The scheduling priority of each user is defined based on channel quality characteristics and packet delay of users. Then a resource allocation scheme is presented, where the resources are assigned to users with the highest priority until the instantaneous rate of users are larger than the minimum rates. The simulation results show that the proposed algorithm can achieve better delay performance, as well as improving the system throughput by comparing with the conventional resource allocation algorithms.

System Model

Figure 1 illustrates the resource allocation structure for TD-LTE-A system. In this paper we focus on the downlink of a general single cell of OFDMA system, where base station can transmit multi-user traffic through downlink time varying fading channels. At MAC layer, the randomly incoming packets from upper layers are buffered in a FIFO (first-in-first-out) queue with infinite space and then wait to be transmitted. The resource allocation controller can utilize channel state information (CSI) and queue state to assign suitable resource block to users. After assigning all resources to users, the scheduled packets are modulated and coded adaptively according to CSI. After IFFF transformation and guard interval insertion, the OFDM symbols are transmitted to all users.

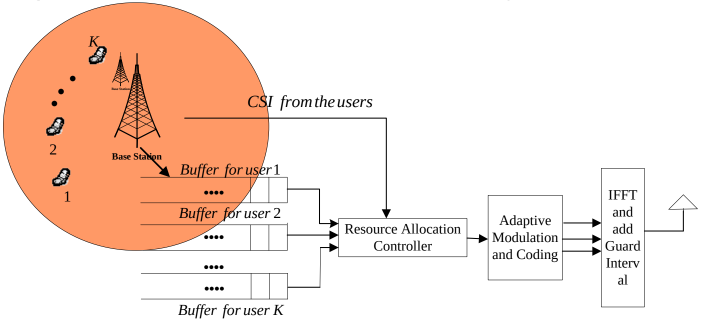

Fig. 1 Resource Allocation Model for TD-LTE-A Systems

According to the 3GPP standard, the radio resource considered in this paper is a basic time-frequency unit called Resource Block (RB). It comprises, in the frequency domain, 12 consecutive sub carriers of 180 kHz bandwidth, and in the time domain it occupies a 0.5 ms time slot. Resource allocation is performed at intervals of 1 ms, i.e. Transmission Time Interval (TTI). Suppose there are K users and N RBs in the current system. For simplicity, we assume power allocated to each RB is the same. At physical layer, we assume the perfect CSI can be sent to Base Station through uplink feedback channel. After allocating resources to each user, the base station can implement the adaptive modulation and coding (AMC) to maximize the data rate on each resource block with CSI. The rate achieved by using AMC is decided by the current channel SNR on each RB and a required bit error rate. According to [2], the achievable instantaneous data rate of kth user on the nth RB at time instant t can be expressed as

$$r_{k,n}(t) = \log_2(1 + \frac{-1.5}{In(5P_b)}\gamma_{k,n}(t)) \qquad (1)$$

Where $\gamma_{k,n}(t)$ is the received SNR on the nth resource block of the kth user at time instant t, and P_b is the required BER for this transmission. According to (1), the total data rate of user k can be expressed as

$$R_k(t) = \sum_{n=1}^{N} w_{n,k} r_{k,n}(t) \qquad (2)$$

Where $w_{n,k}$ is an indicator function. It denotes that the RB k is allocated to user n when $w_{n,k}$ equals to 1, otherwise the RB k is allocated to other users.

Proposed Scheduling and Resource Allocation Algorithm

In this section, we present the combined scheduling and resource allocation algorithm in details. This scheme composes of two steps. The first step is to determine the priority of user on each RB according to channel quality, channel characteristics and delay of users. The second step is to allocate RBs to users according to the computed priority. After data rates of all users are larger than the minimum data rate, the remained RBs are assigned to users with large delay so as to decrease transmission delay of system. More details are as follows.

A. Determination of Scheduling Priority

Each user will be assigned a buffer queue at the eNodeB. Let D_k denotes the packet delay tolerance of user k ($k \in \{1,\cdots,K\}$). When the packet delay of user is larger than D_k, the packets within the queue will be dropped. Since the real traffic in urgent communications has strict delay requirements, we introduce a guard interval G_k to reduce the average packet delay, and it means that the urgent packets of user k should be given a higher priority within the last G_k duration before they timeout.

The delay weight of the user can be defined as

$$C_k = \begin{cases} W_k & if\, W_k \le D_k - G_k \\ \log(W_k)W_k & if\, W_k > D_k - G_k \end{cases} \tag{3}$$

Where W_k is the HOL packet delay of user k at current scheduling instant t. Let $t_c(\mathrm{k})$ denotes the time at which a packet of user k enter the queue in the eNodeB. So the packet delay of user k can be expressed as

$$W_k = t - t_c(k) \tag{4}$$

For every user, we define a variable reflecting current channel state on each resource block. The variable is defined as

$$V_n^k(t) = \frac{r_{\mathrm{k,n}}(t)}{Kr_{i,\max}(t)} \max\{\sum_k | r_{k,n}(t) - r_{\mathrm{k},avg}(t) |, 1\} \tag{5}$$

Where $V_n^k(t)$ denotes the channel state on resource block k for user i, $|\bullet|$ represents the absolute value, and $\max\{a,b\}$ denotes the maximum value between a and b. $r_{\mathrm{k,max}}(t)$ and $r_{\mathrm{k},avg}(t)$ denote the maximum rate and average rate of user k on all resource block alternatively. When $\max\{\sum_k | r_{i,k}(t) - r_{i,avg}(t) |, 1\}$ is larger, the deviation of channel quality is larger, which means some resource block are of low quality and others are of high quality for user k. To decrease transmission delay, the resource block with good channel quality should be given high scheduling priority and assigned to user k.

Based on the analysis above, the priority of users on each resource block is defined as follows

$$P_n^k(t) = C_k V_n^k(t) \tag{6}$$

The scheduling priority described above can assign more RB to users with large delay and good channel quality. After calculating the data rate on each resource block and packer delay of users, we can get the priority of users on different resource block.

B. Resource Allocation

After obtaining the priority of users on each resource block, the number of resource block allocated to selected user can be determined. The proposed resource allocation process is as follows.

After calculating the priority of users, we sort the $P_n^k(t)$ in descending order and get the allocation sequence of users on each resource block. On the selected resource block k, the users sequence is as follows.

$$USR_n = \{i_1, i_2 \cdots, i_K : P_n^{i_1}(t) \ge P_n^{i_2}(t) \ge \cdots \ge P_n^{i_K}(\mathrm{t})\} \tag{7}$$

The resource block is allocated from user i_1, and followed by user i_2, i_3 to i_K. After every resource allocation, we will compute the achieved data rate of each user $r_i^+(\mathrm{t})$. Once the data rate is not less than the minimum service rate $r_{i,\min}(t)$, the resource block will be assigned to other users with lower priority. The achieved rate of user is computed based on the history information of resource block and expressed as follows.

$$r_i^+(t+1) = (1-\mu)r_i^+(t) + \mu r_{i,j}(t+1) \tag{8}$$

Where μ is the window size, $r_i^+(t)$ is the achieved data rate of user i at time t and $r_{i,j}(t+1)$ is the data rate on jth resource block assigned to ith user at time t+1. When the queue of user is empty, this user will not be served so as not to waste resources. Moreover, when the achieved data rate of all users is not less than the minimum rate, the resource allocation sequence of users is sorted in descending order according to delay and the sequence is presented as follows.

$$USR_n' = \{i_1, i_2 \cdots, i_K : \mathrm{W}_{i_1} \geq \mathrm{W}_{i_2} \geq \cdots \geq \mathrm{W}_{i_K}\} \quad (9)$$

The remained resource blocks will be assigned to user according to the order of USR_n' until all resource blocks are all assigned to users. In this way, the system transmission delay can be decreased to some extent. The resource allocation process is described as a pseudo-code in Algorithm 1.

Algorithm 1 The Proposed Algorithm

```
Initialization:
Initialization: Set $r_k(t) = 0$, $r_k^+(t)$=0, Let $\Phi_k = \{\phi\}, \mathrm{k} = \{1,\cdots,K\}$ denotes the assigned RB set for user k; Let $U = \{1,\cdots,K\}$ denotes  user set where users' rate are larger than the minimum data rate; $Q_k$ is the data amount in queue of user k; T is the resource allocation interval.
Iteration:
1: Calculate the instantaneous data rate of each user on each resource block $r_{k,n}(t)$ according to(1);
2: Calculate $W_k$ for all users i, according to (4);
3: Calculate the scheduling priority on each resource block for all users according to (6);
4: Sort $P_n^k(k \in \{1,\cdots K\})$ in descending order and get the user sequence $USR_n$ on each RB according to (7);
5: for each RB $n \in [1, N]$ do
6:    for each user $k \in USR_n$ do
7:       if $r_k^+(t) < r_{k,\min}(t)$ and $Q_k \geq 0$ then
8:           Update $\Phi_k = \Phi_k + \{\mathrm{n}\}$;
9:           Update $r_k^+(t)$ according to (8);
10          Update $Q_k = Q_k - r_{k,n}(t)T$;
11:          break;
12:      else if $r_k^+(t) \geq r_{k,\min}(t)$ then
13:          Update $U = U - \{k\}$;
14:      end if
15:   end for
16:   if U is a empty then
17:      break;
18:   end if
19: end for
20: Sort users in descending order according to users' packet delay  and get the user sequence $USR_n'$ on each RB according to euqation (12).
21: Allocate the remained RBs to users according to the computed sequence $USR_n'$ until the remained RBs are all assigned to users.
22: Transmit each user's packets on the assigned RBs with the corresponding rate.
```

Simulation Results

In this section, system level simulation is performed to evaluate the performance of the proposed scheduling and resource allocation algorithm in terms of system throughout and system average delay. A scenario of a Base Station with 3 cells is considered. Users are randomly distributed within coverage of the base station. Table I shows simulation parameters and corresponding values.

Table I Simulation Parameters

Carrier frequency	2GHz
Carrier Bandwidth	10MHz
Resource Block Number	50
Resource Block BandWidth	180KHZ
eNB Tx power	46dBm
nTx x nRx antennas	2x2
Inter eNB distance	500m
Macrocell pathloss	$128.1+37.6log_{10}R$, R in km
Shadow fading	Lognormal, space-correlated,
Macrocell pading	PedB uncorrelated
Antenna pattern (horizontal)	$A(\theta)=-\min\left[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m\right]$
Thermal noise density	-174dBm/Hz
UE speed	5Km/h

There are 4 different constellations available for AMC at physical layer, which is QPSK, 4QAM, 16QAM, and 64QAM. MCS selection according to SINR thresholds is according to [3]. At MAC layer, we assume each user has video traffic. The video traffic is generated according to the NRTV model in [4]. The duration of each video frame is 100ms. The total number slices of each video frame are deterministic and each slice corresponds to one packet; in the scenario we consider the number is 8. The size of each slice and inter-arrival time between packets follows truncated exponential distribution, the mean slice size is 100 bytes, and max slice size is 150bytes. The packet delay tolerance for video traffic is 100ms. And, the guard interval of video traffic is 60ms.

Using this simulation model, we compare the performance of the proposed scheduling and resource allocation algorithm with respect to three algorithms, namely M-LWDF algorithm [5], DAPS algorithm [6] and Multicarrier Proportional Fair (MPF) algorithm [7]. The simulation results are achieved by comparing the proposed algorithm with the three algorithms described above.

Fig. 2 demonstrates the impact of the number of users on the average Video traffic delay of different scheduling and resource allocation algorithm. As can be seen, the proposed algorithm achieves a lower delay than these algorithms with a wide range of the number of users. This can be explained as follows. Firstly, the proposed algorithm will give higher scheduling priority to users whose delay is closer to the delay bound so as to decrease delay of users. Secondly, when the rate of all users is not less than the minimum rate, the remained resource will be assigned to users with high delay. Fig. 3 shows the average system throughout of Video traffic comparison of the proposed resource allocation algorithm with MPF, DAPS and M-LWDF. It is significant that the proposed resource allocation algorithm achieve the highest average system throughout among all algorithms.

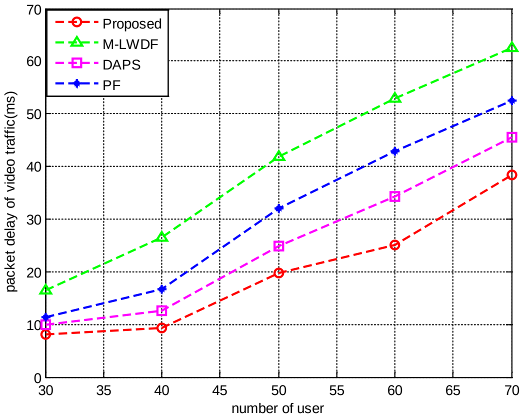

Fig. 2. Packet delay versus number of users with Video traffic

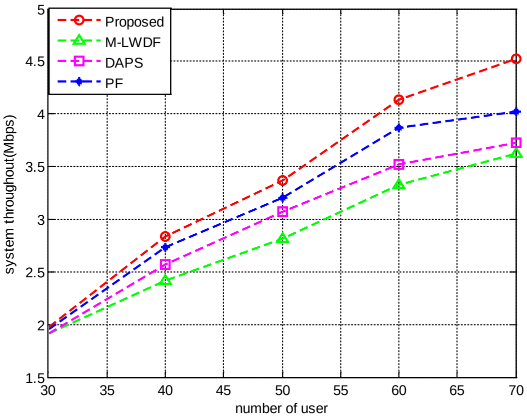

Fig. 3. System Throughout versus number of users with Video traffic

Fig. 4 shows the impact of data rate of video traffic on the average packet delay of users. It is significant that the proposed resource allocation algorithm achieve the best delay performance among all algorithms. When data rate of video traffic is 450kbps, the packet delay in the proposed algorithm is about 140ms.

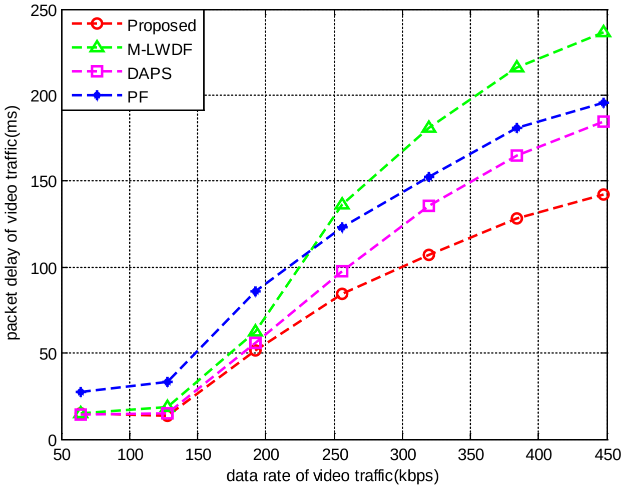

Fig. 4. packet delay versus data rate of video traffic

Summary

In this paper, we proposed an efficient combined scheduling and resource allocation algorithm for transmission delay optimization in TD-LTE-A system. The proposed algorithm provides higher scheduling priority to users with larger delay and better channel quality. The proposed algorithm provides significant performance advantages over other algorithms such as M-LWDF, MPF and DAPS in terms of average packet transmission delay and system throughput. The performance can meet the QoS requirement of Urgent Communication.

References

[1]3GPP TS 36.300 V11.7.0, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network,” www.3gpp.org, September, 2013.

[2] X. Qiu and K. Chawla, ”On the performance of adaptive modulation in cellular systems”,IEEE Trans. Commun., vol. 47, no. 6, pp. 884-895, June 1999.

[3]3GPP TS 36.213 V11.6.0, “Evolved Universal Terrestrial Radio Access (E-UTRA)and Physical layer procedures,” www.3gpp.org, pp. 83, March, 2014.

[4]WG5 Evaluation AHG, 1xEV-DV Evaluation Methodology- Addendum (V6), July 2001.

[5] Kanghee Kim, Insoo Koo, Seokjin Sung and Kiseon Kim, “Multiple QoS Support using M-LWDF in OFDMA Adaptive Resource Allocation,” in Local and Metropolitan Area Networks, pp. 217-222, 2004.

[6]Sueng Jae Bae, Bum-Gon Choi, and Min Young Chung, “Delay-Aware Packet Scheduling Algorithm for Multiple Traffic Classes in 3GPP LTE System,” 17th Asia-Pacific Conference on Communications (APCC), vol.1, pp.33-37, 2011.

[7]H. Kim, K. Kim, Y. Han, and S. Yun, “A proportional fair scheduling for multicarrier transmission systems,” Proc. VTC, vol. 1, pp. 409-413, Los Angeles, CA, September 2004.