

Cross-layer Dynamic Subcarrier Allocation with Adaptive Service Rate Control in SC-FDMA System

Fang Ye¹, Chunxia Su¹, Yibing Li¹, and Xu Zhang¹

¹College of Information and Communication Engineering,
Harbin Engineering University, Harbin, China

[e-mail: yefang0923@126.com, sue212533@126.com, liyibing0920@126.com, zx215347@sina.cn]

*Corresponding author: Yibing Li

*Received May 4, 2017; revised June 22, 2017; accepted July 5, 2017;
published October 31, 2017*

Abstract

In this paper, an improved utility-based cross-layer dynamic subcarrier allocation (DSA) algorithm is proposed for single carrier frequency division multiple access (SC-FDMA) system, which adopts adaptive service rate control (ASRC) to eliminate the service rate waste and improve the spectral efficiency in heterogeneous network including non-real-time traffic and real-time traffic. In this algorithm, furthermore, a first in first out (FIFO) queuing model with finite space is established on the cross-layer scheduling framework. Simulation results indicate that by taking the service rate constraint as the necessary condition for optimality, the ASRC algorithm can effectively eliminate the service rate waste without compromising the scheduling performance. Moreover, the ASRC algorithm is able to further improve the quality of service (QoS) performance and transmission throughput by contributing an attractive performance trade-off between real-time and non-real-time applications.

Keywords: SC-FDMA, heterogeneous network, adaptive service rate control, dynamic subcarrier allocation, quality of service (QoS)

This research was supported by the National Natural Science Foundation of China (Grant No.51509049), the Natural Science Foundation of Heilongjiang Province, China (Grant No.F201345), the Fundamental Research Funds for the Central Universities of China (Grant No.GK2080260140), the National Key Research and Development Program of China (Grant No.2016YFF0102806).

1. Introduction

With the rapid development of multimedia applications and high-speed data transmissions, effective radio resource allocation and management plays an increasingly crucial role in heterogeneous network [1-2]. The dynamic subcarrier allocation (DSA) scheme aiming at efficient wireless spectral resource utilization has drawn an attractive attention [3-7].

The heterogeneous network consists of non-real-time traffic and real-time traffic in terms of the delay sensitivity. For non-real-time applications, the transmission throughput is the main index for performance evaluation because there is no specific QoS requirements. While for real-time applications such as voice and video, the average delay, delay violation probability and loss probability are the QoS metrics for packet scheduling [8-9]. From a network architecture point of view, the current DSA algorithms are mainly based on the strictly layered structure due to its easy implementation in practice. To be specific, the medium access control (MAC) layer based algorithms perform DSA regardless of the time-varying and multipath fading characteristics in wireless channel; while for the physical layer based schemes, such as the greedy algorithm and the adjustable fairness algorithm [10], the traffic burstiness and QoS differentiation are not taken into consideration during allocation. Thus, the existing single-layer DSA frameworks are insufficient in wireless spectral resources utilization and diverse QoS provisioning [11][12]. As a result, it is an urgent need for developing a cross-layer DSA framework to build a bridge between the physical and MAC layers [13]. Under this consideration, some work related to the DSA has been designed for the cross-layer optimization in wireless networks. For instance, a modified largest weighted delay first (M-LWDF) algorithm is presented in [14] by taking the delay of the head-of-line (HOL) into consideration as well as the loss probability. Furthermore, in [15], two utility based cross-layer DSA policies: the alpha-rule and beta-rule are derived for non-real-time and real-time services, respectively. By adjusting the value of the non-negative parameters alpha and beta in utility functions, alpha-rule and beta-rule can provide different degrees of throughput- and delay-based fairness for non-real-time and real-time services, respectively.

However, for all forementioned DSA algorithms, there is no specific scheduling control mechanism to avoid serving the empty queues during allocation, which may cause severe wireless spectral resource waste. In [16], a proportional fair (PF) algorithm is proposed for downlink multiuser OFDMA system with the service rate control that the user with empty queue after service will not participate in the following scheduling procedures. Furthermore, a max-delay-utility (MDU) algorithm is designed for downlink OFDMA network in [17], which fulfills the service rate frugality by means of taking greedy reassignment. While for the above two scheduling control methods, the occurrence of “empty queue after service” may still exist, which means that the service rate waste cannot be thoroughly eliminated. At present, there is still no effective service rate control scheme available for the DSA in SC-FDMA system. On the other hand, the existing cross-layer frameworks for the SC-FDMA system are all based on an infinite queue size assumption, which is obviously inconsistent with the practical queuing condition.

In this paper, we focus on both developing DSA algorithm and formulating service rate control mechanism for uplink SC-FDMA system. In order to meet the realistic wireless transmission scenario, a utility-based cross-layer scheduling framework is established on a finite length queuing model. Based on this architecture, an improved DSA algorithm under the heterogeneous traffic environment is proposed for completely eliminating the service rate

waste so as to effectively improve the spectral efficiency. The remainder of this paper is organized as follows. Section 2 describes the cross-layer DSA model for uplink SC-FDMA system. In Section 3, the proposed cross-layer DSA algorithm with service rate control is developed. Then, simulation results are presented in Section 4 .Finally, Section 5 draws some conclusions.

2. System Model

The cross-layer DSA architecture is illustrated in Fig. 1. Unlike the single-layer based DSA, the cross-layer optimization evaluates both the channel state information (CSI) observed in the physical layer and the queue state information (QSI) obtained at MAC layer at the same time. Thus, it is practically a kind of joint channel- and queue-aware scheduling [18]. The channel states of different access users are almost mutually independent due to the frequency selective nature of the wireless channels, which can provide a high degree of flexibility for DSA. Furthermore, the heterogeneous traffic with various degrees of traffic load will result in different queue states, which is also the important basis for packet scheduling. Thus, by formulating the DSA scheduling as a joint physical and MAC layer optimization problem, the wireless spectral resource can be fully exploited for QoS provisioning.

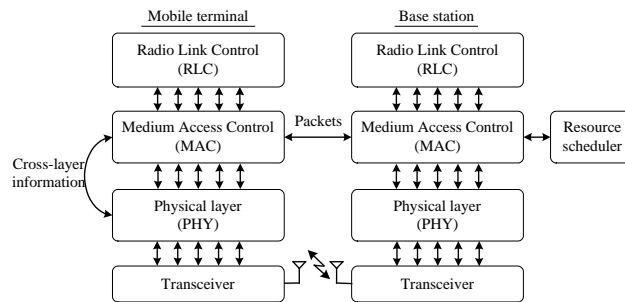


Fig. 1. Cross-layer dynamic subcarrier allocation architecture

Consider time synchronized uplink SC-FDMA transmission with transmission bandwidth B Hz and the total bandwidth is orthogonally subdivided into M subcarriers. A set of sub-carriers constitute a resource block (RB), and one or multiple RBs can be allocated to each user in each transmission time intervals (TTIs) which is a basic unit of time scheduling. The number of subcarriers per RB is M/N , where N denotes the number of RBs. Thus, the number of sub-carriers in a RB is regarded as a minimum resource unit for sub-carrier allocation in the frequency domain [19].

2.1 System Metrics

Fig. 2 shows the block diagram of SC-FDMA system, in which $\mathcal{N} = \{1, 2, \dots, N\}$ and $\mathcal{K} = \{1, 2, \dots, K\}$ represent the RB index set and user index set, respectively [20]. Consider a heterogeneous wireless network with non-real-time traffic user set \mathcal{K}_{NRT} and real-time traffic user set \mathcal{K}_{RT} , and $\mathcal{K} = \mathcal{K}_{NRT} \cup \mathcal{K}_{RT}$. Let $|\mathcal{K}_{NRT}|$ and $|\mathcal{K}_{RT}|$ denote the number of non-real-time traffic user and the number of real-time traffic user, respectively, then $K = |\mathcal{K}_{NRT}| + |\mathcal{K}_{RT}|$.

Assume that the base station has perfect knowledge of the channel gains of all users in the time and frequency domains and the minimum mean square error (MMSE) frequency domain equalizer (FDE) is employed in the receiver for inter-symbol interference (ISI) mitigation. To perform the subcarrier allocation, the resource scheduler located at the base station (BS) broadcasts the subcarrier allocation information to each user via downlink control signals. As indicated in [21], there is no obvious improvement but heavy computational burden to introduce the adaptive power allocation and adaptive modulation and coding to each subcarriers in a RB. Thus, in this paper, we adopt equal-bit-equal-power (EBEP) allocation for each RB to reduce the algorithmic complexity as well as the control signaling overhead.

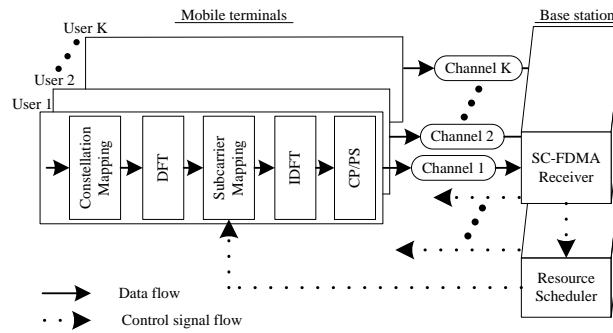


Fig. 2. SC-FDMA system dynamic subcarrier allocation block diagram

In channel-aware network for cross-layer optimization, the channel state is measured in terms of the signal to noise ratio (SNR) by taking channel estimation at BS. Thus, in SC-FDMA system, the SNR of user k at TTI t after MMSE equalization can be given by [22].

$$\gamma_k(t) = \left(\frac{1}{\frac{1}{|I_{sub,k}^{(t)}|} \sum_{i \in I_{sub,k}^{(t)}} \frac{\gamma_{k,i}(t)}{\gamma_{k,i}(t) + 1}} - 1 \right)^{-1} \tag{1}$$

where $\gamma_{k,i}(t)$ is the SNR of subcarrier i for user k at TTI t , $I_{sub,k}^{(t)}$ and $|I_{sub,k}^{(t)}|$ represent the set of subcarriers and the number of subcarriers assigned to user k at TTI t respectively.

According to the Shannon’s formula, the achievable throughput for user k at TTI t can be expressed as

$$C_{k,I_{RB,k}^{(t)}}(t) = \frac{B \cdot |I_{RB,k}^{(t)}|}{N} \cdot \log_2 [1 + \gamma_k(t)] \tag{2}$$

where $I_{RB,k}^{(t)}$ and $|I_{RB,k}^{(t)}|$ represent the set of RBs and the number of RBs assigned to user k at TTI t respectively.

On the other hand, for queue-aware scheduling, we establish a first in first out (FIFO) queuing model with finite space of L . At the beginning of each TTI, all the packets coming from the higher application layer are buffered in the corresponding queues and waiting to be scheduled. Let $Q_k(t)$ presents the amount of bits in the queue of user k at TTI t . The queue length of user k at TTI $(t+1)$ with the finite space limit of L is

$$\begin{aligned} & Q_k(t+1) \\ & = \min\left\{L, Q_k(t) - \min\left(Q_k(t), C_{k, I_{RB,k}^{(t)}}(t) \cdot T\right) + v_k(t)\right\} \end{aligned} \quad (3)$$

where $v_k(t)$ is the amount of arrival bits at TTI t , and T is the length of each TTI.

The number of loss bits due to the buffer overflow for user k at the end of TTI $(t+1)$ is given by

$$\begin{aligned} & D_k(t+1) \\ & = \max\left\{0, \left(Q_k(t) - \min\left(Q_k(t), C_{k, I_{RB,k}^{(t)}}(t) \cdot T\right) + v_k(t)\right) - L\right\} \end{aligned} \quad (4)$$

Thus, the system overall loss probability can be defined as

$$P_{loss}(t) = \frac{\sum_{k \in \mathcal{K}} D_k(t)}{\sum_{k \in \mathcal{K}} Q_k(t)} \quad (5)$$

In practice, the transmission throughput (i.e., service rate) of each user is constrained by its throughput demand corresponding to the amount of bits buffered in the queue, which can be expressed as

$$r_k(t) = \min\left\{C_{k, I_{RB,k}^{(t)}}(t), \frac{Q_k(t)}{T}\right\} \quad (6)$$

Suppose that the real-time user queue length during transmission time is ergodic, so the average queue length of each user can be obtained as

$$Q_k = \lim_{T \rightarrow \infty} \frac{\sum_{t=0}^{T-1} Q_k(t)}{T} \quad (7)$$

According to the Little's law, the average delay of user can be obtained by calculating the average queue length. Thus, the average delay of each user at TTI t can be defined as

$$\begin{aligned} d_k & = \frac{Q_k}{\lambda_k} \\ \lambda_k & = \frac{1}{T} \lim_{t \rightarrow \infty} \frac{v_k(t)}{t} \end{aligned}$$

where λ_k is the average arrival bit rate for user k .

However, the computation complexity involved in (7) is expensive. In practical system, the average queue length can be observed by exploiting an exponentially weighted low-pass window as

$$\bar{Q}_k(t) = (1 - \rho_w) \cdot \bar{Q}_k(t-1) + \rho_w \cdot Q_k(t) \quad (8)$$

where $\rho_w = T/T_w$, $\rho_w \in (0, 1)$ and T_w is the window size.

Finally, the average delay of each user at TTI t can be obtained by

$$d_k(t) = \frac{\bar{Q}_k(t)}{\lambda_k} \quad (9)$$

Let $D_{th,k}$ denote the delay threshold for user k , and δ_k is the corresponding maximum allowable probability. On this basis, the delay violation probability requirement of user k can be defined as

$$\Pr\{d_k > D_{th,k}\} \leq \delta_k \quad (10)$$

2.2 The Utility-based Optimization

Utility theory can be introduced in cross-layer optimization framework to evaluate the degree to which a scheduling algorithm satisfies transmission requirements of users' applications [23]. By formulating utility functions with respect to specific system performance metrics, the benefit of taking certain spectral resource for data transmission can be mapped into the corresponding utility values. To be specific, for non-real-time traffic, the utility functions can be designed regarding the transmission throughput so as to contribute the trade-off between spectral efficiency and fairness. By contrast, the utility functions for real-time applications are generally associated with the average delay and the delay related requirements, thereby effectively meeting the QoS requirements. Thus, by maximizing the aggregate utility over the established utility-pricing model, the cross-layer DSA framework can be optimized. Based on this concept, the utility-based cross-layer optimization problem can be mathematically formulated as

$$\max \sum_{k \in \mathbb{K}} U_k(t) \quad (11)$$

subject to

$$\bigcup_{k_i \in \mathbb{K}} I_{RB,k_i}^{(t)} \subseteq \mathcal{N},$$

$$I_{RB,k_i}^{(t)} \cap I_{RB,k_j}^{(t)} = \emptyset, \quad k_i \neq k_j, \quad \forall k_i, k_j \in \mathbb{K}$$

where $U_k(t)$ represents the utility function for user k at TTI t , and the allocation restricts that each RB cannot be shared among users during each TTI.

3. Cross-layer DSA with Service Rate Control

In this section, we present an improved cross-layer DSA scheme with adaptive service rate control (ASRC) for QoS guarantees in SC-FDMA system, which is designed on a utility-based optimization framework with finite queue length. By introducing service rate control factor, the ASRC algorithm proposed in this paper can completely eliminate service rate waste. Furthermore, the ASRC algorithm can provide a strict grade division of scheduling priorities between real-time and non-real-time traffics by adding adaptive admission threshold, thereby ensuring the QoS demands of real-time applications.

As discussed before, the service rate of each user is determined by two factors: the achievable transmission throughput (i.e., Shannon capacity) and the queue length. Notice that if a user is assigned with the Shannon capacity exceeding its throughput demand, the exceed amount of service rate obviously cannot be utilized for data transmission, and there will be service rate waste during allocation. Based on this concept, the amount of waste service rate caused by the inefficient scheduling control of user k at the end of each TTI can be given by

$$W_k(t) = \max \left\{ 0, C_{k,I_{RB,k}}^{(t)}(t) - \frac{Q_k(t)}{T} \right\} \quad (12)$$

We define the service rate waste probability as the ratio of the total waste service rate over all the TTIs for all the users to the aggregate transmission throughput over these TTIs for these users as follows

$$P_{waste} = \frac{\sum_{k \in \mathcal{K}} W_k}{\sum_{k \in \mathcal{K}} C_{k,I_{RB,k}}^{(t)}} \quad (13)$$

There are usually two approaches to design the service rate control mechanisms. The first one is to develop a kind of additional allocation constraint which is performed independently with the utility-based objective functions. For instance, a commonly used service rate control method presented in [16] restricts the service rate of each user as

$$C_{k,I_{RB,k}}^{(t)}(t) \leq \frac{Q_k(t)}{T} \quad (14)$$

which indicates that the user will be removed from the user index set if its queue becomes empty after the previous service. It is obviously that this “restriction after optimization” approach can only avoid the scheduler serving empty user queues, rather than strictly observing the constraint target indicated in equation (14) during allocation. Thus, it cannot eliminate the service rate waste completely.

The other method for service rate control is to perform resource reassignment according to appropriate scheduling constraints. As presented in [17], a max delay utility (MDU) algorithm is designed by introducing the frugality constraint (FC) to the greedy reassignment process. The algorithm in [17] consists of two steps. In the first step, the scheduler preliminarily allocates all the RBs according to MDU principle. After the preliminary allocation, the value of $r_k^{diff}(t) = \frac{Q_k(t)}{T} - r_k(t)$ for user k can be obtained within each TTI. According to the value of $r_k^{diff}(t)$, we define I_{user}^{non} and I_{user}^{emp} as the set of users with nonempty queues and empty queues after preliminary allocations, respectively. Let I_{RB}^{emp} represent the set of RBs assigned to the user whose queue becomes empty after the previous service. Then the second step implements greedy reassignment (GR) according to FC, that is, the reassignment scheduler in [17] will reallocate RBs in I_{RB}^{emp} to users who can obtain the maximum system utilities, even though the reallocated RBs exceed the practical throughput demand. This means that the FC reassignment is insufficient to eliminate the service rate waste completely.

In order to overcome the above disadvantages, we introduce the service rate constraint into the objective function. Let $S_{k,n}(t)$ be the service rate control factor indicating whether the scheduling satisfies the service rate constraint or not, that is,

$$S_{k,n}(t) = \begin{cases} 1 & C_{k,n}(t) \leq \frac{Q_k(t)}{T} \\ 0 & C_{k,n}(t) > \frac{Q_k(t)}{T} \end{cases} \quad (15)$$

By introducing the service rate constraint as the necessary condition for optimization, it provides an efficient method for service rate waste elimination. However, it is insufficient to guarantee the scheduling superiority of real-time applications, particularly in high-speed transmission scenario with heavy traffic load and outburst arrival data. Furthermore, from a

resource allocation point of view, there is a trade-off between spectral efficiency and QoS support, which means that the improvement in QoS performance is obtained at the expense of transmission throughput deterioration. Most of existing work concerning cross-layer optimization focuses on achieving an effective performance trade-off rather than fully exploiting the available spectral resource for overall performance enhancement.

To address the above problems, the ASRC algorithm is developed by integrating the concept of admission control with cross-layer DSA. The key idea of ASRC is to establish an admission control policy [24] for the active connections, thereby providing a strict grade division of scheduling priorities between non-real-time and real-time applications. The admission control policy is implemented by introducing admission threshold, and the admission threshold sets system admission condition for users. For further enhancing the service provisioning capability, the admission threshold for ASRC algorithm is designed in an adaptive manner, which is aware of the average delay performance and the specific QoS requirements of real-time applications. Based on this architecture, the threshold value at TTI t can be calculated based on the feedback control loop as follows

$$\alpha(t) = \alpha(t-1) - \eta \cdot (\bar{d}(t) - D_{th,k}) \quad (16)$$

where η is the step size that controls the adaptation speed of the threshold value, and the value of η has no obvious effect on the system performance if it is set in a proper range. In this paper, η is set as $1/100000$. $\alpha(t-1)$ denotes the threshold value at TTI $(t-1)$. Besides, $\bar{d}(t)$ represents the average delay for all the real-time applications at TTI t , which is given by

$$\bar{d}(t) = \frac{1}{|\mathbb{K}_{RT}|} \sum_{k \in \mathbb{K}_{RT}} d_k(t) \quad (17)$$

According to equation (16), the ASRC scheduler continuously monitors the system average delay performance and dynamically adapts the admission control threshold to meet the delay requirements of each real-time service during scheduling process. Thus, the ASRC algorithm can offer a great flexibility for connection admission by adjusting the admission threshold to work on a desired operation point. To simplify the algorithm design, we introduce the admission control factor $A_k(t)$ to denote the admission state of each connection, which can only be either 1 or 0, indicating whether the connection is accepted or not. To be specific, when the average delay of the real-time applications exceed the admission threshold, that is, $\bar{d}(t) \geq \alpha(t)$, we have

$$A_k(t) = \begin{cases} 1 & k \in \mathbb{K}_{RT}, \\ 0 & k \in \mathbb{K}_{NRT}, \end{cases} \quad (18)$$

Otherwise, when $\bar{d}(t) < \alpha(t)$, it becomes

$$A_k(t) = \begin{cases} 1 & k \in \mathbb{K}_{NRT}, \\ 0 & k \in \mathbb{K}_{RT}. \end{cases} \quad (19)$$

As can be seen from equation (16), the deterioration in average delay will cause the decrease of admission threshold value, which may create more opportunities for real-time applications to connect to the system, thereby improving the delay performance. Conversely, in the low average delay scenario, the connection chances will switch to the non-real-time applications for transmission throughput enhancement. Under this consideration, the non-real-time traffic in ASRC algorithm can be regarded as the background traffic so as to effectively guarantee the scheduling priorities for real-time services. Furthermore, in order to make the most efficient

use of the available spectral resource, the max-sum-capacity (MSC) scheduling rule is considered for non-real-time traffic in ASRC algorithm, which designs the utility function with respect to transmission throughput as

$$U_{k,n}(t) = r_{k,n}(t) \quad (20)$$

Thus, by considering service rate control and the priority of real-time traffic at the same time, the utility-based optimization framework for ASRC algorithm can be mathematically formulated as

$$\max \sum_{k \in \mathbb{K}} \sum_{n \in \mathbb{N}} A_k(t) \cdot S_{k,n}(t) \cdot U_{k,n}(t) \quad (21)$$

subject to

$$A_k(t) \in \{0, 1\}, \forall k \in \mathbb{K}$$

$$S_{k,n}(t) \in \{0, 1\}, k \in \mathbb{K}, n \in \mathbb{N}$$

$$\bigcup_{k \in \mathbb{K}} I_{RB,k}^{(t)} \subseteq \mathbb{N}$$

$$I_{RB,k}^{(t)} \cap I_{RB,j}^{(t)} = \emptyset, k \neq j, \forall k, j \in \mathbb{K}$$

The objective optimization problem (21) is actually a 0-1 programming problem with multiple complicated constraints, which is difficult to solve by using mathematical tools, so we propose ASRC algorithm. Different from previous service rate control methods, the proposed ASRC algorithm directly introduces the service rate control factor into the optimization objective functions, which can strictly restrict the transmission throughput assigned to each user so as to fully guarantee the service rate constraint. Besides, the proposed ASRC algorithm introduces admission control factor into the optimization objective functions, which can provide the strict grade division of scheduling priorities between non-real-time and real-time applications, thereby adjusting system performance adaptively. We group all the non-real-time users and real-time users waiting to be scheduled in the system into $I_{user}^{NRT}(t)$ and $I_{user}^{RT}(t)$, respectively. The allocation procedure of the ASRC algorithm can be depicted as follows:

The ASRC algorithm

Initialization:

Let $I_{user}(t) \leftarrow \mathbb{K} = \{1, 2, \dots, K\}$;

Let $I_{RB}(t) \leftarrow \mathbb{N} = \{1, 2, \dots, N\}$;

Let $I_{RB,k}(t) \leftarrow \emptyset$ for all $k \in I_{user}(t)$;

Let $A_k(t) \leftarrow 1$ for all $k \in I_{user}(t)$;

Let $S_{k,n}(t) \leftarrow 1$ for all $k \in I_{user}(t)$ and $n \in I_{RB}(t)$;

Calculate $\alpha(t)$ according to (16);

While $I_{RB}(t) \neq \emptyset$

Calculate $\bar{d}(t)$ according to (17), and compare with $\alpha(t)$;

Update $A_k(t)$ for $k \in I_{user}(t)$ according to (18) and (19);

Calculate the objective function $A_k(t) \cdot S_{k,n}(t) \cdot U_{k,n}(t)$ for $k \in I_{user}(t)$;

Find the user $k^* \in I_{user}^{RT}(t)$ and RB $n^* \in I_{RB}(t)$ that

$$(k^*, n^*, t) = \arg \max_{\substack{k \in I_{user}(t) \\ n \in I_{RB,k}(t)}} \{A_k(t) \cdot S_{k,n}(t) \cdot U_{k,n}(t)\};$$

If $C_{k^*,n^*}(t) \leq Q_{k^*}(t)/T$ (the allocation satisfies the service rate constraint)

Allocate RB n^* to user k^* that $I_{RB,k^*}(t) \leftarrow I_{RB,k^*}(t) \cup \{n^*\}$;

Delete RB n^* from $I_{RB}(t)$ that $I_{RB}(t) \leftarrow I_{RB}(t) - \{n^*\}$;

Update $Q_{k^*}(t)$ that $Q_{k^*}(t) \leftarrow Q_{k^*}(t) - C_{k^*,n^*}(t) \cdot T$;

Else if $C_{k^*,n^*}(t) > Q_{k^*}(t)/T$ (the allocation does not satisfy the service rate constraint)

Set $S_{k^*,n^*}(t)$ equal to 0 that $S_{k^*,n^*}(t) \leftarrow 0$;

End if

End while

4. Simulation Results and Analysis

4.1 Simulation Configuration

The simulation parameters for the SC-FDMA system considered in our analysis are presented in **Table 1**. It is assumed that the BS has perfectly acquired the CSI and QSI of all the terminals during each TTI. Furthermore, the channel estimation as well as the scheduling decision transmission is performed instantaneously without taking the DSA feedback delay into account. The wireless channel is modeled as an ITU-R vehicular channel model A with 6 paths, and each path suffers from different Rayleigh fading and transmission delay. We establish a single-cell SC-FDMA cellular with a radius of 1 km. All the mobile terminals are uniformly and independently distributed in the circular cellular, and the BS is located at the center. Moreover, each user is assumed to be stationary or slowly moving, thus, the path loss can be defined as

$$PL_k = 128.1 + 37.6 \log_{10} l_k + \xi_k \quad (dB)$$

where l_k (km) represents the distance between the terminal k and BS, and ξ_k is the log-normal shadowing.

We consider a heterogeneous traffic environment which consists of both non-real-time and real-time traffic. Specifically, the non-real-time traffic is modeled as a Poisson process with dynamic resource rate, such as the data traffic. While for real-time traffic, two typical packetized traffic are considered, i.e., video and voice. In our simulations, the voice and video traffic are both generated on the two-state ON-OFF model with different average periods. In the ON state, the voice packets arrive at a constant rate with fixed-size, and the coming packets for the video traffic are characterized by different packet size as well as variable arrival rate. Whereas in the OFF state for both voice and video traffic, there are no packets generated. The corresponding parameters of the real-time traffic are summarized in **Table 2**.

Table 1. Simulation parameters for SC-FDMA system

System parameters	Values	System parameters	Values
Total available bandwidth	5 MHz	Radio frequency carrier	2 GHz
System sampling rate	200 ns	Cyclic prefix length	4 us
Transmission time interval (T)	1 ms	Total transmit power of each user	1 W
Number of subcarriers	256	Number of chunks	32
Number of users	16	AWGN power spectral density	-160 dBm/Hz
Average window size (T_w)	100 T	Queue size (L)	60000 bits
Slow shadowing standard deviation	8 dB	Bit and power allocation method	EBEP
Modulation method	QPSK	Equalization scheme	MMSE

Table 2. Simulation parameters for real-time traffic

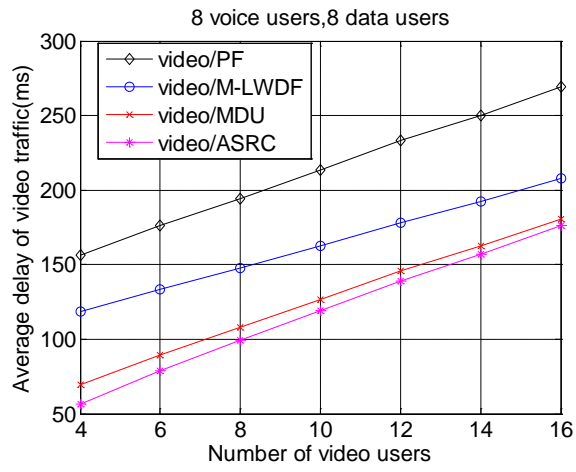
Traffic parameters	Voice traffic	Video traffic
Mean ON period	1.00 s	1.47 s
Mean OFF period	1.35 s	1.92 s
The minimum data rate	32 Kbps	256 Kbps
Average data rate	32 Kbps	512 Kbps
The maximum data rate	32 Kbps	1024 Kbps
Delay limit	50 ms	200 ms
Delay violation probability	5%	5%

4.2 Simulation Results

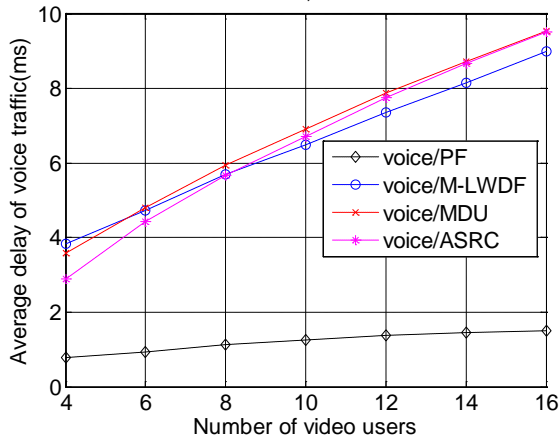
4.2.1 Performance Results with Video Users Increase

Based on the above simulation model, the performance of the proposed ASRC algorithm is compared with the traditional DSA algorithms in terms of five aspects: the average delay, loss probability, transmission throughput, service rate waste probability and delay violation probability. We present comprehensive simulation results considering two kinds of different traffic distribution scenarios. All simulations were run for 500 loops, and each loop includes 500 TTIs, finally, the results were averaged.

In the first traffic distribution scenario, we fix the numbers of data and voice users both to 8, and increase the number of video users from 4 to 16 so as to evaluate the impact of video user number on scheduling performance. Based on this architecture, the performance comparison of all the considered DSA algorithms is shown from [Fig. 3](#) to [Fig. 7](#).



(a) Average delay of video traffic
8 voice users,8 data users



(b) Average delay of voice traffic

Fig. 3. Average delay versus video user number

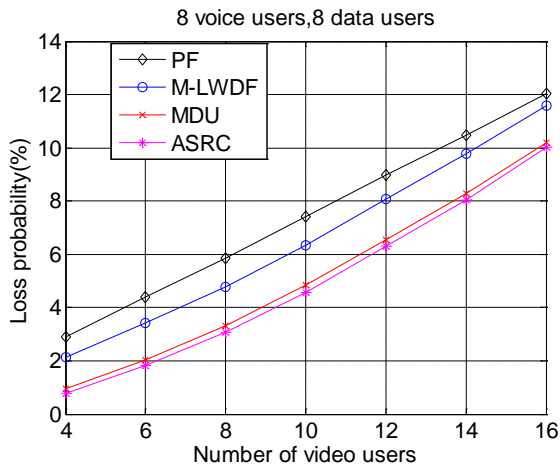


Fig. 4. Loss probability versus video user number

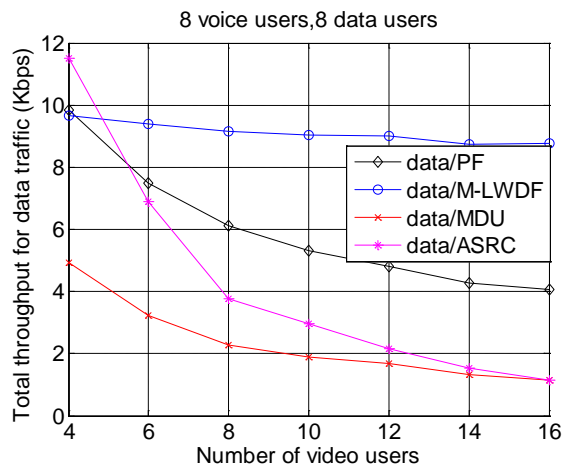


Fig. 5. System throughput versus video user number

As can be seen in [Fig. 3\(a\)](#), the average delay of the video traffic increases as the number of video users increases. This can be explained by the fact that with limited wireless spectral resource, the high traffic density of video applications results in severe network congestion, which will cost long time for video packets waiting to be transmitted in the queues. While in [Fig. 3\(b\)](#), the voice traffic with relatively low packet arriving rate always has good delay performance which shows no obvious increase with video user number increases. Notice that when the packets congestion becomes increasingly severe and exceeds the predetermined finite queue size, there will be packet drop due to the buffer overflow. As illustrated in [Fig. 4](#), the loss probability performance associated with the congestion extent apparently deteriorates with the video user number increases.

To be specific, with the same utility function assumption, the proposed ASRC algorithm can achieve the similar average delay and loss probability performance with the MDU algorithm, which also significantly outperforms the PF algorithm and the M-LWDF algorithms. However, due to the trade-off between spectral efficiency and QoS support, there is an inevitable transmission throughput deterioration in ASRC and MDU algorithm, as shown in [Fig. 5](#). The waste elimination performance is indicated by the service rate waste probability, as depicted in [Fig. 6](#). It can be observed that the service rate waste probability of the ASRC algorithm is much lower than the MDU and M-LWDF algorithms, and the PF algorithm shows the highest waste probability. This means that the additional allocation constraint in [16] as well as the greedy reassignment based FC in [17] are both insufficient to eliminate the service rate waste completely. On the other hand, by directly introducing the service rate control factor into the optimization objective functions, the ASRC algorithm strictly restricts the transmission throughput of each user never exceed the corresponding throughput demand, thus, it can effectively eliminate the service rate waste.

[Fig. 7](#) represents the delay violation probability performance. It can be observed that ASRC algorithm has the lowest delay violation probability, because the adaptive admission control mechanism in ASRC algorithm can provide a rational scheduling priority distribution between real-time and non-real-time traffic, which is implemented by means of continuously monitoring the delay performance and the QoS requirements of each real-time user. Based on this concept, an attractive tradeoff between the QoS support of real-time applications and the transmission throughput of non-real-time applications can be achieved.

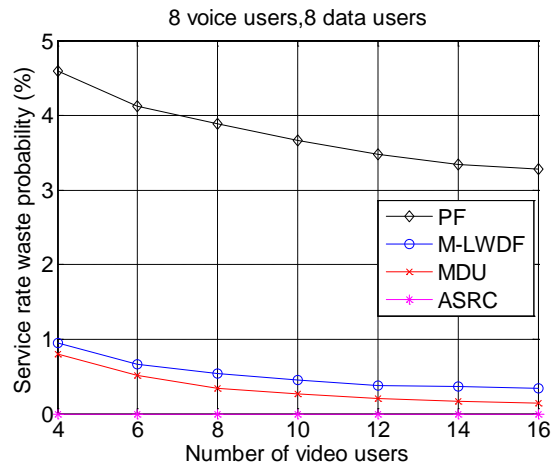


Fig. 6. Service rate waste probability versus video user number

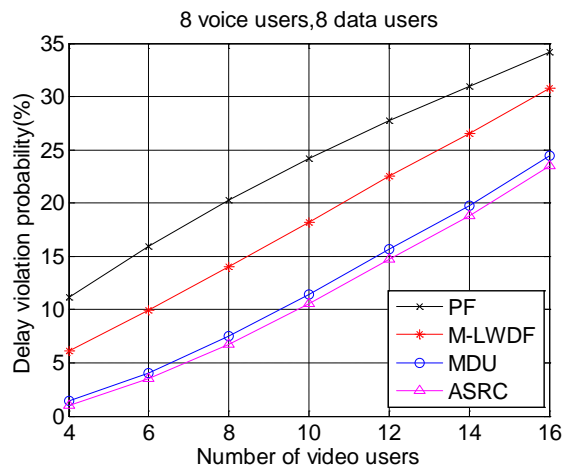


Fig. 7. Delay violation probability versus video user number

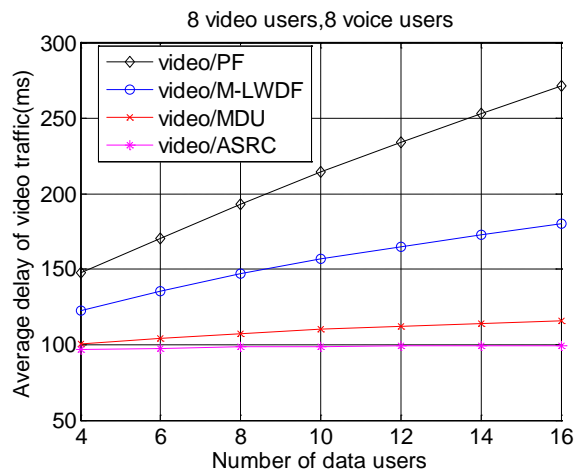
Under this traffic distribution scenario, the full scheduling performance comparison of all the considered DSA algorithms is shown in **Table 3** with 10 video users.

Table 3. Numerical results of simulation with 10 video users

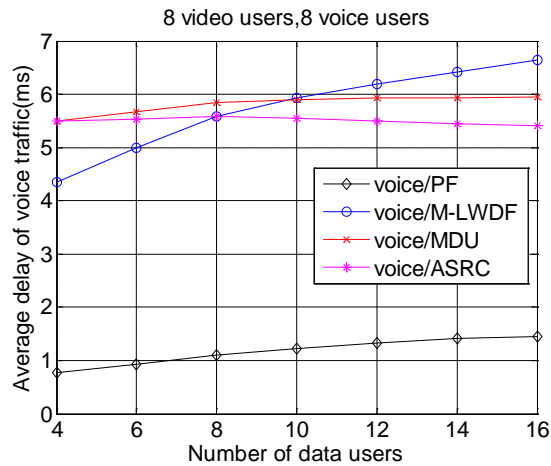
Performance	PF	M-LWDF	MDU	ASRC
Average delay of video user (ms)	213.30	162.67	126.76	118.86
Average delay of voice user (ms)	1.26	6.48	6.92	6.72
Loss probability (%)	7.41	6.34	4.84	4.58
Transmission throughput (Mbps)	5.30	9.03	1.89	2.97
Service rate waste probability (%)	3.66	0.45	0.26	0
Delay violation probability (%)	24.16	18.21	11.42	10.53

4.2.2 Performance Results with Data Users Increase

To demonstrate the performance of the proposed DSA algorithm in various traffic distribution environments, we consider the second traffic scenario in which the numbers of video and voice users are both fixed to 8, and the number of data users is increased from 4 to 16. **Fig. 8** shows that the average delay of the video and voice traffic still remains high-quality performance with the MDU and ASRC algorithm with the increase of data user number, which implies that the increase of non-real-time users in these DSA algorithms has no significant effect on average delay performance of the real-time applications. When observing the loss probability performance in **Fig. 9**, the ASRC algorithm achieves the best loss probability performance.



(a) Average delay of video traffic



(b) Average delay of voice traffic

Fig. 8. Average delay versus data user number

Furthermore, it can be observed from **Fig. 10** that the total transmission throughput of data traffic increases with the increase of data user number. This is because that in this low real-time traffic congested situation, the non-real-time applications can get more opportunities to access to the system, thereby achieving high transmission throughput results from multiuser

diversity. It should be noticed that in this traffic distribution scenario, the ASRC algorithm can contribute an attractive trade-off between the QoS support of real-time traffic and the transmission throughput of non-real-time traffic. Moreover, the ASRC algorithm can effectively eliminate the service rate waste, as shown in Fig. 11. Different from Fig. 7, we can observe from Fig. 12 that the delay violation probability performances of MDU and ASRC remain steady with the increase of data user number, while delay violation probability of PF and M-LWDF increase with the data user number.

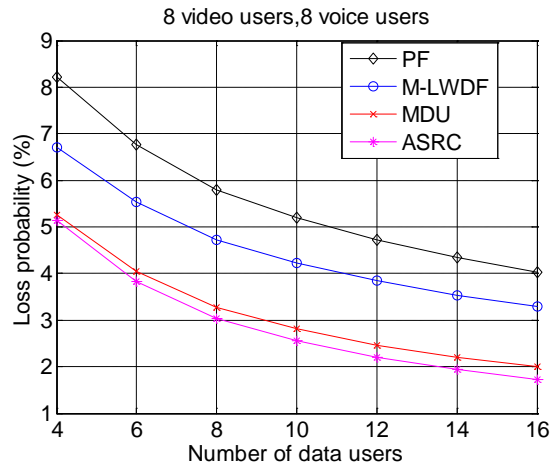


Fig. 9. Loss probability versus data user number

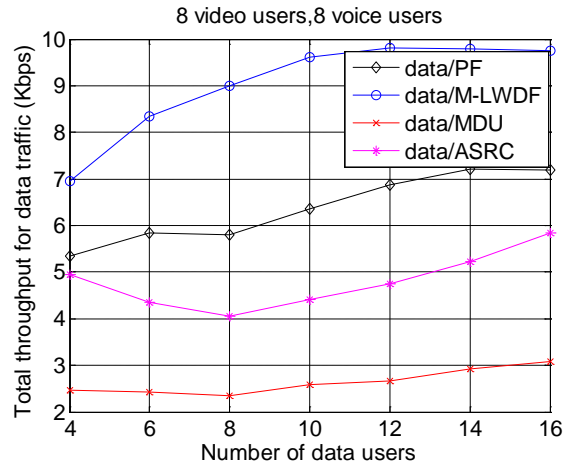


Fig. 10. System throughput versus data user number

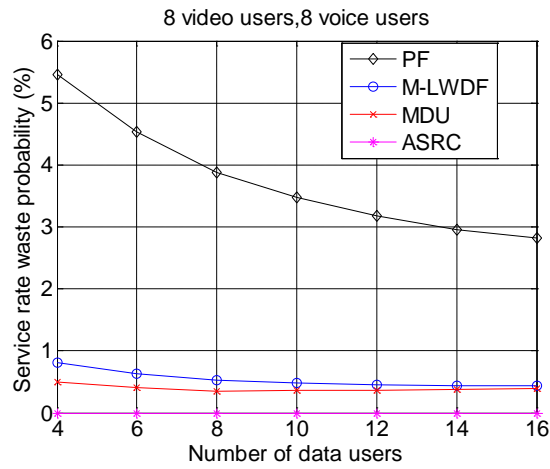


Fig. 11. Service rate waste probability versus data user number

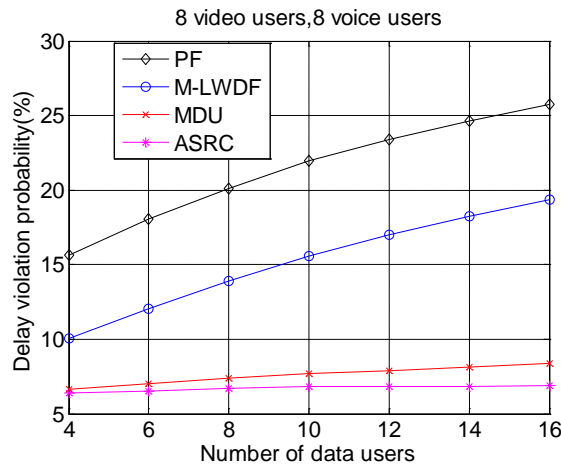


Fig. 12. Delay violation probability versus data user number

The full scheduling performance comparison of all the considered DSA algorithms under the second traffic distribution scenario is shown in Table 4.

Table 4. Numerical results of simulation with 10 data users

Performance	PF	M-LWDF	MDU	ASRC
Average delay of video user (ms)	214.54	156.78	110.08	98.98
Average delay of voice user (ms)	1.23	5.92	5.90	5.55
Loss probability (%)	5.20	4.23	2.81	2.57
Transmission throughput (Mbps)	6.36	9.60	2.58	4.42
Service rate waste probability (%)	3.48	0.48	0.36	0
Delay violation probability (%)	21.96	15.55	7.72	6.80

4.2.3 Computational Complexity

Fig. 13 and Fig. 14 show the average running time of the four comparison algorithms. The simulation was run on a personal computer, and the software and hardware conditions are:

Windows 7 Ultimate Edition, Pentium(R) Dual-Core(2.93-GHz) processor, and 4-GB memory. It is clear that M-LWDF is much more complex than the other three algorithms. The running time of ASRC is similar to MDU and PF, which is tolerable.

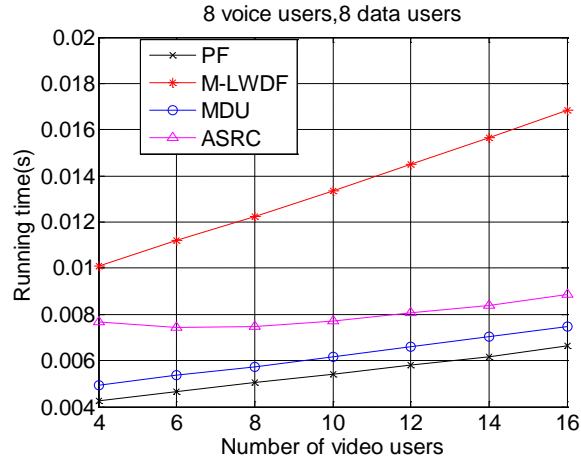


Fig. 13. Running time versus video user number

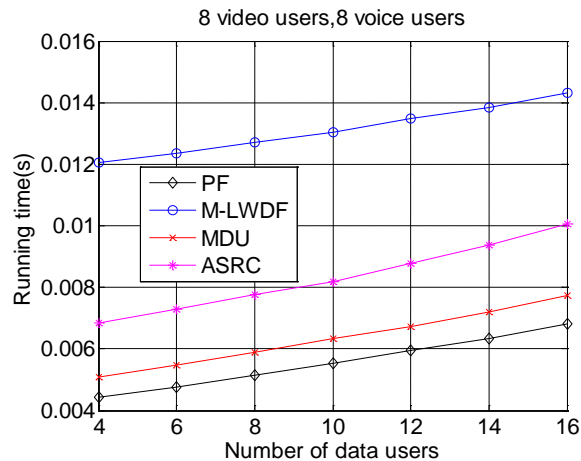


Fig. 14. Running time versus data user number

5. Conclusion

In this paper, we present a utility based cross-layer framework for uplink DSA in SC-FDMA system, which is aware of both the CSI observed in physical layer as well as the QSI obtained at MAC layer. Furthermore, in order to meet the realistic transmission conditions, the proposed scheduling framework is also established on a finite queuing model according to the queuing specifications in data link layer. Based on this architecture, an improved DSA algorithm considered service rate control is designed aiming to eliminate the service rate waste and improve the spectral efficiency for heterogeneous traffic with diverse QoS requirements. Simulation results indicate that by introducing the service rate constraint into the objective functions, the proposed algorithm can achieve excellent service rate waste elimination performance, while maintaining the overall scheduling performance in terms of average delay, loss probability and transmission throughput. Moreover, by establishing an admission control

policy which adapts to the average delay and the QoS requirements of each real-time user, the algorithm provides a reasonable scheduling priority division between non-real-time and real-time applications. As shown in the simulation results, by contributing an attractive performance trade-off between real-time and non-real-time applications, the ASRC algorithm can effectively improve the QoS performance and the transmission throughput, at the same time, the ASRC algorithm can completely eliminate the service rate waste.

References

- [1] Ahmad A., Khan M.T.R. and Kaleem Z., "Uplink optimal power allocation for heterogeneous multiuser SIMO SC-FDMA networks," *Electronics Letters*, vol. 52, no. 24, pp. 1990-1992, November, 2016. [Article \(CrossRef Link\)](#)
- [2] Gu J., Lee T.J. and Chung M.Y., "Water-filling technique-based proportionally fair scheduling algorithm for SC-FDMA," *Transactions on Emerging Telecommunications Technologies*, vol. 27, no. 12, pp. 1664-1671, December, 2016. [Article \(CrossRef Link\)](#)
- [3] Seba Vitomir, Modlic Borivoj and Sisul Gordan, "Resource Allocation Algorithm with Dynamic Subcarrier Assignment in OFDMA-based Wireless Networks," *Automatika*, vol. 55, no. 4, pp. 495-503, October-December, 2014. [Article \(CrossRef Link\)](#)
- [4] Lu X, Ni Q, Li W and Zhang H, "Dynamic User Grouping and Joint Resource Allocation with Multi-Cell Cooperation for Uplink Virtual MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3854-3869, June, 2017. [Article \(CrossRef Link\)](#)
- [5] Li Y, Nie W and Ye F, "A Complex Mixing Matrix Estimation Algorithm Based on Single Source Points," *Circuits Systems and Signal Processing*, vol. 34, no. 11, pp. 3709-3723, November, 2015. [Article \(CrossRef Link\)](#)
- [6] Ahmad A, "Resource allocation and adaptive modulation in uplink SC-FDMA systems," *Wireless personal communications*, vol. 75, no. 4, pp. 2217-2242, April, 2014. [Article \(CrossRef Link\)](#)
- [7] Tsiropoulou E E, Kapoukakis A and Papavassiliou S, "Uplink resource allocation in SC-FDMA wireless networks: A survey and taxonomy," *Computer Networks*, vol. 96, pp. 1-28, February, 2016. [Article \(CrossRef Link\)](#)
- [8] Chen Q and Wu D, "Delay-Rate-Distortion Model for Real-Time Video Communication," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1376-1394, August, 2015. [Article \(CrossRef Link\)](#)
- [9] LI, Y. B., YANG,R. and YE, F, "Improved spectrum sharing algorithm based on feedback control information in cognitive radio networks," *Journal of System Engineering and Electronics*, vol. 24, no. 4, pp. 564-570, August, 2013. [Article \(CrossRef Link\)](#)
- [10] MYUNG, H. G. and GOODMAN, D. J, *Single Carrier FDMA: a New Air Interface for Long Term Evolution*, John Wiley & Sons, 2008.
- [11] Liu Y S and Wu J S, "A Chunk-Based Scheduling Algorithm Considering User Satisfaction for SC-FDMA Transmission," *Wireless Personal Communications*, vol. 92, no. 4, pp. 1387-1405, February, 2017. [Article \(CrossRef Link\)](#)
- [12] Zarakovitis CC, Ni Q, Skordoulis DE and Hadjinicolaou MG, "Power-efficient cross-layer design for OFDMA systems with heterogeneous QoS, imperfect CSI, and outage considerations," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 781-798, February, 2012. [Article \(CrossRef Link\)](#)
- [13] SADR, S., ANPALAGAN, A. and RAAHEMIFAR, K, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 3, pp. 92-106, 2009. [Article \(CrossRef Link\)](#)
- [14] ANDREWS, M., KUMARAN, K., RAMANAN, K., STOLYAR, A., WHITING, P. and VIJAYAKUMAR, R., "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150-154, February, 2001. [Article \(CrossRef Link\)](#)

- [15] Rodrigues, E. B. and Casadevall, F., "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Communications Magazine*, vol. 49, no. 9, pp. 90-98, September, 2011. [Article \(CrossRef Link\)](#)
- [16] KONG, Z., KWOK, Y. K. and WANG, J. Z., "A low-complexity QoS aware proportional fair multicarrier scheduling algorithm for OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 5, pp. 2225-2235, June, 2009. [Article \(CrossRef Link\)](#)
- [17] SONG, G. C., LI, Y., CIMINI, L. J. JR. and ZHENG, H., "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," in *Proc. of IEEE Wireless Communications and Networking Conference*, pp. 1939-1944, March 21-25, 2004. [Article \(CrossRef Link\)](#)
- [18] SONG, G. C. and LI, Y., "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 127-134, December, 2005. [Article \(CrossRef Link\)](#)
- [19] Junsung, L., Myung, H. G., Kyungjin, O. and Goodman, D. J., "Proportional fair scheduling of uplink single-carrier FDMA," in *Proc. of the 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp.1-6, September 11-14, 2006. [Article \(CrossRef Link\)](#)
- [20] Triantafyllopoulou D, Kollias K and Moessner K., "QoS and Energy Efficient Resource Allocation in Uplink SC-FDMA Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3033-3045, June, 2015. [Article \(CrossRef Link\)](#)
- [21] Lee, T. H., and Huang, Y. W., "Resource allocation achieving high system throughput with QoS support in OFDMA-based system," *IEEE Transactions on Communications*, vol. 60, no. 3, pp. 851-861, March, 2012. [Article \(CrossRef Link\)](#)
- [22] Shi T, Zhou S and Yao Y., "Capacity of single carrier systems with frequency-domain equalization," in *Proc. of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, pp. 429-432, May 31-June 02, 2004. [Article \(CrossRef Link\)](#)
- [23] KATOOZIAN, M., NAVAIE, K., and YANIKOMEROGLU, H., "Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 66-71, January, 2009. [Article \(CrossRef Link\)](#)
- [24] Rodrigues, E. B., Cavalcanti, F. R. P. and Wanstedt, S., "QoS-driven adaptive congestion control for voice over IP in multiservice wireless cellular networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 100-107, January, 2008. [Article \(CrossRef Link\)](#)



Fang Ye received the B.S. and Ph.D degrees in Electrical Information Engineering from Harbin Engineering University(HEU) in 2002 and 2006, respectively. She has been a teacher in Harbin Engineering University of China since 2002, and became an associate professor in 2007. During 2007-2008, she stayed in School of Electronics and Computer Science from University of Southampton as a visiting scholar. Now she is an IEEE member, a member of China Institute of Communications and a member of China Computer Federation. Her research interests include LTE technology, adaptive radio resource allocation technology and Cognitive radio.



Chunxia Su received her B.S. degree in communication engineering from Harbin Engineering University in 2014. She is currently working towards the Ph.D. degree in Harbin Engineering University. Her research interests include LTE, SC-FDMA technology and adaptive radio resource allocation technology.



Yibing Li received the B.S., M.S. and Ph.D degrees in Harbin Marine engineering college, Harbin engineering university in 1989, 1997 and 2003, respectively. He has been a teacher in Harbin Engineering University of China since 1989, and became a professor in 2004. During 2007-2008, he stayed in the University of Hong Kong Electronic Engineering lab as a visiting scholar. Now he is an IEEE member, a senior member of China Institute of Communications and a senior member of China Computer Federation. His research interests are in areas of wireless communication technology, LTE technology and information fusion.



Xu Zhang received her B.S. degree in electronic information engineering from Harbin Engineering University in 2009, and received her Ph.D. degree in communication and information system from Harbin Engineering University in 2013. Her research interests include LTE, SC-FDMA technology.