DOI: 10.3969/j.issn.1673-5188.2018.02.003

Multi-QoS Guaranteed Resource Allocation for Multi-Services Based on Opportunity Costs

JIN Yaqi, XU Xiaodong, and TAO Xiaofeng
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract

To meet the booming development of diversified services and new applications in the future, the fifth-generation mobile communication system (5G) has arisen. Resources are increasingly scarce in the dynamic time-varying of 5G networks. Allocating resources effectively and ensuring quality of service (QoS) requirements of multi-services come to be a research focus. In this paper, we utilize effective capacity to build a utility function with multi-QoS metrics, including rate, delay bound and packet loss ratio. Taking advantage of opportunity cost (OC), we also propose a multi-QoS guaranteed resource allocation algorithm for multi-services to consider the future condition of system. In the algorithm, according to different business characteristics and the theory of OC, we propose different selection conditions for QoS users and best effort (BE) users to choose more reasonable resources. Finally, simulation results show that our proposed algorithm achieves superior system utility and relatively better fairness in multi-service scenarios.

Keywords

utility function; effective capacity; opportunity cost; QoS guaranteed; resource allocation

1 Introduction

The rapid development of the mobile Internet has driven the demand of users for higher speed, larger data traffic and more intensive network coverage. Therefore, the fifth-generation of mobile communication system (5G) has emerged. With the explosive growth of business demands in 5G, more services which have diverse quality of service (QoS) requirements appear. Due to the scarcity of resources in mobile communication systems, the importance of resource allocation in system performance is decisive. The conflict between limited wireless resources and the ever-increasing QoS needs has become increasingly acute. Therefore, we call for a resource allocation strategy that can improve overall network performance as well as support high quality multi-services.

Researches on resource allocation for QoS guaranteed under multi-service hybrid scenarios have increasingly become a research hotspot. The authors of [1] solved the distributed-resource allocation problems in 5G cellular systems. They utilized the concepts of stable matching, factor-graph-based message passing, and distributed auctions. The authors of [2] investigated the problem of power allocation and sub-channel assignment in heterogeneous small cell network. They considered cross-tier interference mitigation, energy harvesting and incomplete channel state information. In [3], the optimal objective function proposed in the wireless resource allocation algorithm is introduced. The time delay is taken into as a constraint to guarantee QoS. The size of the delay constraint value can be set according to the priority and quality of service requirements. In [4], the researchers proposed a heterogeneous QoS-driven resource allocation scheme by the multiple input multiple output-orthogonal frequency-division multiple access (MIMO-OFDMA) based relaying scheme. Given the heterogeneous statistical QoS constraints, the authors derived the effective capacity under developed optimal power-allocation policies. The authors of [5] proposed a QoS scheduling strategy for multi-users and multi-services. They considered the service type, channel quality, buffer size and fairness based on carrier aggregation.

In this paper, we consider multi-services whose QoS metrics include bandwidth requirements, as well as delay, packet loss ratio, etc. [6]. Therefore, we firstly analyze different typical services' business characteristics. To consider the variety of QoS metrics, we combine the theory of effective capacity (EC) with unified utility function to well characterize multiple QoS constraints [7]–[8]. EC is the maximum constant data rate that a given service rate can support subject to a QoS exponent [9].

This paper is supported by the National Science and Technology Major Project under Grant No. 2016ZX03001009-003 and the Nature and Science Foundation of China under Grants Nos. 61471068 and 111 Project of China B16006.

Besides, in order to consider the dynamic time-varying conditions of 5G ultra-dense networks, the paper introduces the opportunity cost (OC) model. In economics, the original concept of OC is the maximum benefit that can be gained from other uses after the resource is put into a particular use [10]–[11]. As the theory of OC evolves and expands, the opportunity cost is used in not only economics, but also multi-service resources allocation scenarios. According to OC, we propose a multi-service QoS guaranteed resource allocation algorithm, which can improve the system utility while guaranteeing the QoS requirement of multi-users within a long-time frame.

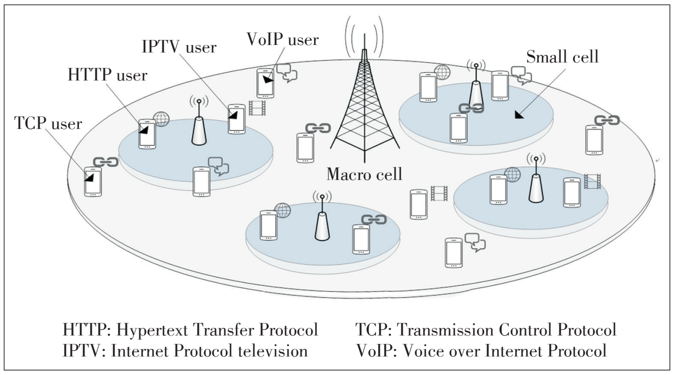
The contributions of this paper can be summarized as follows:

1) Our scheme invites effective capacity based utility function to establish a multi-QoS optimization strategy that put the system utility as the objective function.
2) We introduce opportunity cost to take future network condition into account, which can lead multi-services to make rational choices to conduct resource allocation within the tolerance of delay requirement.
3) The schemes and simulation results show that our algorithm can achieve superior overall user satisfaction and fairness. The correctness of the proposed utility is also evaluated, which could be used as references for related studies.

The rest of this paper is organized as follows. Section 2 provides the system model, EC-based utility function and opportunity cost for multi-service scenarios. Section 3 formulates the multi-service QoS guaranteed resource allocation problem. By comparing opportunity cost and current data rate of QoS users and BE users respectively, we come to the conditions users choose more suitable resources to get better system performance. Simulation results are given and discussed in section 4. Finally, section 5 draws conclusions.

2 System Model

We consider the downlink scenario of multi-service cellular networks as shown in **Fig. 1**. The scenario depicts N base stations (BSs), the power of which is limited by P_n respectively. There are multiple users, which are categorized into four classes: conversational class (VoIP), streaming class (IPTV), interactive class (HTTP) and background class (TCP). The number of VoIP, IPTV, HTTP and TCP users are K_1, K_2, K_3, and K_4 respectively, and the sum of the four kind of users is equal to K. The resource pool has M resource blocks (RBs), which we allow each BS to use. We assume that in each scheduling cycle, a RB m can only be assigned to one user k at most. $\beta \in \{0,1\}^{K \times M \times N}$ is the RB assignment matrix, where $\beta_{k,m,n}=1$ indicates the assignment RB and $\beta_{k,m,n}=0$, otherwise.



▲**Figure 1. The system model of wireless network for multi-QoS mobile services.**

The Signal to Noise Ratio (SNR) between k-th user and n-th BS on RB m is formulated as follows:

$$SNR_{k,m,n} = P_{k,m,n} \cdot \left|H_{k,m,n}\right|^2/(N_0 B), \tag{1}$$

where $P_{k,m,n}$ denotes the transmit power. $\left|H_{k,m,n}\right|^2$ denotes the channel gain. N_0 is the power spectral density of noise and B is the bandwidth of each RB.

Based on Shannon's capacity formula, the rate between the k-th user and the n-th BS on RB m is given by (2):

$$R_{k,m,n} = B\log(1 + P_{k,m,n} \cdot C_{k,m,n}), \tag{2}$$

where $C_{k,m,n} = \left|H_{k,m,n}\right|^2/(N_0 B)$. The total data rate that user k gets from BS n is expressed as:

$$R_{k,n} = \sum_{m=1}^{M} \beta_{k,m,n} \cdot R_{k,m,n}. \tag{3}$$

2.1 Effective Capacity Based Utility Function

To measure user satisfaction, utility function is proposed. The higher the value of $U(r)$, the higher the user satisfaction. To analyze different service characteristics, the utility curves of four typical classes are introduced in **Fig. 2** [12].

From the research of [13], we use the effective capacity based utility function to consider multi-QoS requirements. The function combines the effective capacity with a uniform utility function, which is expresses as

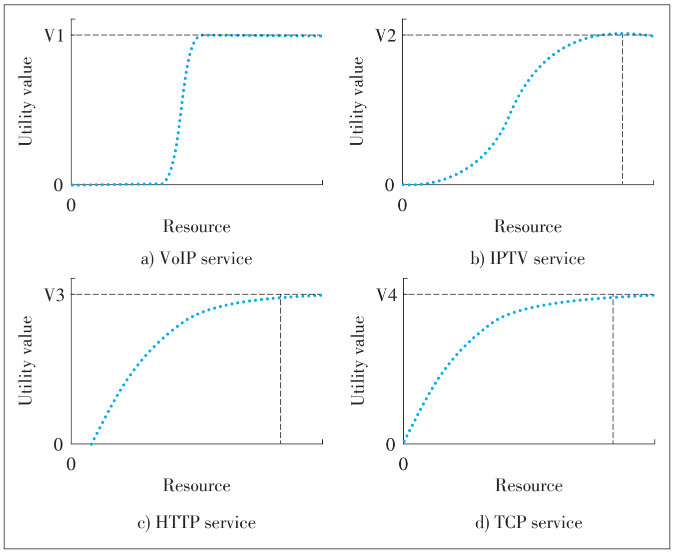
$$U_k(E_{C_k}(\theta_{k,n})) = U_k(P_{k,m,n}, \beta_{k,m,n}, \theta_{k,n}), \tag{4}$$

where $E_{C_k}(\theta_{k,n})$ is the effective capacity with the parameter of QoS exponent θ.

The utility function $U(r)$ can be set by different parameters to form different curves adapted to different typical classes [8].

$$U(r) = \frac{1}{A + Be^{-C(r-d)}} + D, \tag{5}$$

where r is the resource parameter. Parameters A, B and D mainly affect the range of the utility function. C affects the



a) VoIP service b) IPTV service

c) HTTP service d) TCP service

▲**Figure 2. The utility function trend of four classes.**

slope of the curve. And d is the inflexion of the function.

The effective capacity is expressed as following:

$$E_C(\theta) = -\frac{1}{\theta T}\log(E\left[e^{-\theta TR[i]}\right]), \tag{6}$$

where $R[i]$ represents the instantaneous channel capacity during i-th frame. T represents the frame duration.

Moreover, the probability that the delay bound $D_{\max}$ must be below a certain packet loss ratio ε is expressed as (7):

$$\Pr\{D(\infty) > D_{\max}\} \approx e^{-\theta(\lambda)\cdot D_{\max}} \leqslant \varepsilon, \tag{7}$$

where λ represents the service arriving rate.

2.2 Opportunity Cost

Multi-service resource allocation provides corresponding service quality assurance for different user terminals under the condition of limited resources. By allocating available resources such as subcarriers and power properly, we use resource allocation to maximize user satisfaction, system throughput and the efficiency of resource utilization. If a user selects a base station´s resources, it will lose the opportunity to access other base stations. If a base station chooses a user to provide service, it will lose opportunity to allocate resources to other users. In summary, the scenario of multi-service resource allocation meets the three preconditions of opportunity cost: scarcity, diversity and decision rationality of resources.

In this paper, we divide users into QoS users and BE users to research multi-service resource allocation scheme. We assume that choices are made rationally by users aiming to maximize system utility.

For QoS users, when the EC of a QoS user is greater than the expected effective bandwidth, increasing the rate will not greatly improves the utility. However, for BE users, as user data rate continues to increase, their utility will continue to improve. Therefore, if QoS users select the current resource, they may lose the better suited resource when the delay is tolerable. It is also possible that the waiting time exceeds the delay limit so that a serious decline in channel quality causes calls drop.

For BE users, if they choose the current resource, they may lose the larger expected transmission rate that can be obtained when its latency is tolerable. If BE users drop the connection, it is also possible to drop the call. Therefore, from the above resource allocation, the user needs to make a choice to select or abandon a certain resource.

We come to the definition of opportunity cost for resource allocation as follows: the opportunity cost of a user is defined as the expected value of the service transmission rate that the user chooses to wait within the delay allowed by the service.

By introducing the concept of opportunity cost, we can not only allocate according to the current resource situation, but also consider the future changes of the system and maximize the utility value of the system through the rational choice of users.

3 Multi-Service QoS Guaranteed Resource Allocation Based on Opportunity Cost

In this section, we study the multi-service QoS guaranteed resource allocation algorithm based on opportunity cost aiming to improve system utility.

Taking advantage of opportunity cost, taking the opportunity cost into consideration can consider the advantages of the future situation of the system and put forward a more reasonable resource allocation algorithm.

3.1 Problem Formulation

To optimize multi-service resource allocation algorithm, we aim to maximize system utility, that is, the sum utility function of each user. The objective function is shown in the following formula:

$$\max \sum_k U_k(P_{k.m.n}, \beta_{k.m.n}, \theta_{k,n}). \tag{8}$$

In this paper, we suppose users are sorted by the delay bound $D_{\max,k}$ from the smallest to the largest. It is easy to find that user sequence meets $\{k \in K_1, k \in K_2, k \in K_3, k \in K_4\}$. Meanwhile, the channel transmission capabilities of BSs are sorted from the largest to the smallest. And the corresponding channel transmission capabilities of the base stations are $\{E_{C_n} | n = 1, 2, \ldots, N\}$. Without loss of generality, we suppose $E_{C_1} \geqslant E_{C_2} \geqslant \ldots \geqslant E_{C_N}$.

Assume the user’s service duration has an exponential distribution with parameter τ [14]. The distribution function of the service duration T of each user in the base station is as following,

$$F(t) = P(T \leqslant t) = \begin{cases} 1 - e^{-\frac{t}{\tau}}, t > 0 \\ 0, \quad \text{other} \end{cases}. \tag{9}$$

For a full-loaded BS, the probability that the BS will have free resources, that is, at least one user ends the service leaving within the duration t, can be expressed as

$$P(t)=1-P(T_1>t,T_2>t,...,T_K>t). \tag{10}$$

Supposing the service duration of user is mutual independent of each other, the probability is written as

$$P(t)=1-(e^{-t/\tau})^K=1-e^{-Kt/\tau}. \tag{11}$$

From the above analysis, the probability that the $BS-i=1,2,...,N$ will have free resources is $1-e^{-K_i t/\tau}$ respectively, where K_i represents the number of users served by the i-th BS and $K_0=0$.

Assuming that the resource is pre-assigned without considering the future condition of the system, the power pre-allocated to the user is $P'_{k,m,n}$. And if the user occupies the RB m' on the BS n', it is denoted as $\beta'_{k,m,n}=1$. For QoS users, we suppose the allocated data rates all satisfy $E_c(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n})\geqslant E_b^{\,k},k\in K_1,K_2$. Effective bandwidth E_b is defined as the minimum service rate to meet the QoS requirements [15] According to the utility characteristic of QoS users from Fig. 2, when the effective capacity of QoS users is larger than their effective bandwidth, increasing the rate does not greatly improve their utility. Therefore, it is better for QoS user k to obtain desired data rates smaller but meets its QoS requirement during its waiting time t which is less than the delay tolerance range $D_{\max,k}$. Therefore, the resources can be assigned to other users, so as to enhance the overall effectiveness of the system and the system fairness. The condition QoS users choose to wait as shown in (12):

$$E_b^{\,k}\leqslant E_{QoS}[E_C(\theta)]\leqslant E_C(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}). \tag{12}$$

Therefore, QoS users only consider the base station whose transmission capacity is lower than n' to save more resources. We can conclude the expected transmission rate that can be obtained in the waiting time t for QoS users in (13):

$$E_{QoS}[E_C(\theta)]=\sum_{i=n'+1}^{N}(1-e^{-\frac{K_i t}{\tau}})e^{-\sum_{s=i+1}^{N}\frac{K_s t}{\tau}}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}). \tag{13}$$

What is different from QoS users, when BE user is provided more resources, is that BE user satisfaction can still be improved. Therefore, if the opportunity cost of BE user k after the waiting time t is greater than the pre-allocated resources, BE users may choose to wait to access a higher transfer rate, as shown in the following equation:

$$E_{BE}[E_c(\theta)]\geqslant E_C(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}). \tag{14}$$

Unlike QoS users, BE users only consider base stations with transmission capability higher than n' in order to obtain more data rates to improve their utility. The expected transmission rate that can be obtained in the waiting time t is as follows:

$$E_{BE}[E_C(\theta)]=\sum_{i=1}^{n'-1}(1-e^{-\frac{K_i t}{\tau}})e^{-\sum_{s=1}^{i}\frac{K_{s-1} t}{\tau}}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}). \tag{15}$$

If there are multiple users waiting in line for a certain BS service, the waiting queue length is set as j_i (the user is ranked j_i) and $j_0=0$. The probability that the user can get service within the tolerable waiting time range is equal to the probability that at least j_i users leaving within the time frame. It can be seen that the time interval for the user to leave the system obeys the negative exponential distribution of parameter $\lambda_i=K_i/\tau$.

Therefore, the number of users leaving the system in time obeys Poisson distribution. The probability of the user getting the service is converted to (16):

$$P(t_1+t_2+...+t_j<t)=P(k\geqslant j)=\\
1-\sum_{k=0}^{j}\frac{\lambda^k}{k!}e^{-\lambda}. \tag{16}$$

Thus, the expected transmission rate that QoS users can wait for access is:

$$E_{QoS}[E_C(\theta)]=\sum_{i=n'+1}^{N}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=i+1}^{N}(\sum_{k=0}^{j_s}\frac{\lambda_s^k}{k!}e^{-\lambda_s})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}). \tag{17}$$

The condition QoS users choose to wait is shown as following:

$$E_b^k\leqslant\sum_{i=n'+1}^{N}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=i+1}^{N}(\sum_{k=0}^{j_s}\frac{\lambda_s^k}{k!}e^{-\lambda_s})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i})\leqslant\\
E_c(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}). \tag{18}$$

The expected transmission rate that a BE user can wait for access is:

$$E_{BE}[E_C(\theta)]=\sum_{i=1}^{n'-1}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=1}^{i}(\sum_{k=0}^{j_{s-1}}\frac{\lambda_{s-1}^k}{k!}e^{-\lambda_{s-1}})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}). \tag{19}$$

The condition BE users choose to wait is shown as following:

$$\sum_{i=1}^{n'-1}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=1}^{i}(\sum_{k=0}^{j_{s-1}}\frac{\lambda_{s-1}^k}{k!}e^{-\lambda_{s-1}})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i})\geqslant\\
E_C(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}). \tag{20}$$

In addition, users need to tolerate delay during the waiting of the queuing. And each user has the maximum delay limit $D_{\max,k}$, beyond which the user's service quality will be affected. Therefore, the duration a user chooses to wait should be less than its maximum delay limit. When user k queues for the service of a selected base station n, we assume the length of waiting queue is j_n. The probability distribution of the user's service duration obeys the exponential distribution with parameter τ. According to the characteristic of exponential distribution, the expected duration is τ. So the user's queue waiting delay is $j_n\tau$. Therefore, j_n needs to satisfy the following formula:

$$j_n \leqslant [\frac{D_{\max,k}}{\tau}], \tag{21}$$

where $[x]$ represents an integer not greater than x.

3.2 Problem Solution

Based on the above analysis, a multi-service QoS guaranteed resource allocation optimization model based on opportunity cost is formed as follows:

$$\begin{aligned}
&\max \sum_k U_k(P_{k.m.n},\beta_{k.m.n},\theta_{k,n}),\\
&\text{s.t. 1) } E_b^k \leqslant \sum_{i=n'+1}^{N}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=i+1}^{N}(\sum_{k=0}^{j_s}\frac{\lambda_s^k}{k!}e^{-\lambda_s})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}) \leqslant\\
&\qquad E_c(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}), k\in K_1,K_2,\\
&\text{2) } \sum_{i=1}^{n'-1}(1-\sum_{k=0}^{j_i}\frac{\lambda_i^k}{k!}e^{-\lambda_i})\{\prod_{s=1}^{i}(\sum_{k=0}^{j_{s-1}}\frac{\lambda_{s-1}^k}{k!}e^{-\lambda_{s-1}})\}\cdot E_c(P_{k,m,i},\beta_{k,m,i},\theta_{k,i}) \geqslant\\
&\qquad E_C(P'_{k,m,n},\beta'_{k,m,n},\theta_{k,n}), k\in K_3,K_4,\\
&\text{3) } j_n \leqslant [\frac{D_{\max,k}}{\tau}],\ \ n=1,\ldots,N,\\
&\text{4) } \sum_k\sum_n P_{k,m,n} \leqslant P_n,\\
&\text{5) } \sum_k\sum_n \beta_{k,m,n} \leqslant 1, \beta_{k,m,n}\in\{0,1\}.
\end{aligned} \tag{22}$$

The algorithm flow is designed as follows:

1) First, we calculate the channel transmission capacity of each base station and sort it from high to low, numbered $1,\ldots,n,\ldots N$. The users are sorted according to their maximum delay limit from low to high, numbered as $1,\ldots,k,\ldots K$. Then when $k=1,\ldots,K_1+K_2$, step 2 is performed for a QoS user; when $k=K_1+K_2+1,\ldots,K$, step 3 is performed for a BE user.

2) When $k=1,\ldots,K_1+K_2$, we calculate the opportunity cost of QoS users. The algorithm finds the value of n that satisfies the limit conditions 1) and 3) in (22). To make the system utility higher, set $n^*=\arg\max\{n\}$, $\beta^*_{k,m,n}=1$, $j_{n^*}+1$. If not, then make $k+1$.

3) When $k=K_1+K_2+1,\ldots,K$, we calculate the opportunity cost of BE users. And then the algorithm finds the value of n that satisfies the limit conditions 2) and 3) in (22) To make the system utility higher, set $n^*=\arg\max\{n\}$, $\beta^*_{k,m,n}=1$, $j_{n^*}+1$. If not, then make $k+1$.

4) After traversing all the users, the resource allocation is completed. And the power matrix and the allocation matrix are updated when each n' is taken.

4 Simulation Results and Analysis

4.1 Simulation Parameters

In this section, simulation results about the utility-based resource allocation for multi-QoS services with EC are given. There are multiple BSs deployed in a coverage area of 2 km×3 km. All BSs connect with the controller. There are total 50 files and the size of each content is a normal random variable with the mean of 30 Mbits. The requests of users follow a Zipf distribution with Zipf parameter g=0.5. Generally, g indicates the degree of skewness of popularity distribution, a larger g means the content requests are more centralized into popular files. Besides, we let the users of the four classes have the same proportion. More details of simulation environment settings and QoS settings are listed in **Tables 1** and **2** respectively. The QoS parameters are set according to 3GPP TS 23.203 [6]. The four classes have different QoS parameters, such as delay bound and packet loss ratio. Besides, VoIP users and IPTV users have effective bandwidth bound, because they are QoS users who have minimum bandwidth requirement.

4.2 Results and Analysis

According to the proposed resource allocation algorithm, QoS users are expected to seek more suitable resources instead of blindly seeking for better resources by comparing opportunity cost with current resources favorably. In this paper, we compare sum utility, system throughput and fairness between the proposed OC algorithm, the max utility algorithm without OC

▼Table 1. Simulation setting: system parameters

System parameters	
Number of BSs	7
Number of Subchannels	50
Maximum power of BSs	46 dBm
Carrier Frequency	2 GHz
Bandwidth	10 MHz
Cell average radius	500 m
Pathloss model	$PL=128.1+37.6\log_{10}d, d(km)$
Shadowing standard deviation	8 dB
Shadowing correlation distance	50 m
Fast fading	Rayleigh fading
Noise density	-174 dBm/Hz
Average arriving rate	150 kbit/s

BS: base station

▼Table 2. Simulation setting: QoS parameters

Traffic Type	Effective bandwidth (kbit/s)	Delay bound (ms)	Packet loss ratio
VoIP	150	[20, 50]	10^{-2}
IPTV	200	[50, 100]	10^{-3}
HTTP	--	[100, 200]	10^{-6}
TCP	--	[300, 500]	10^{-6}

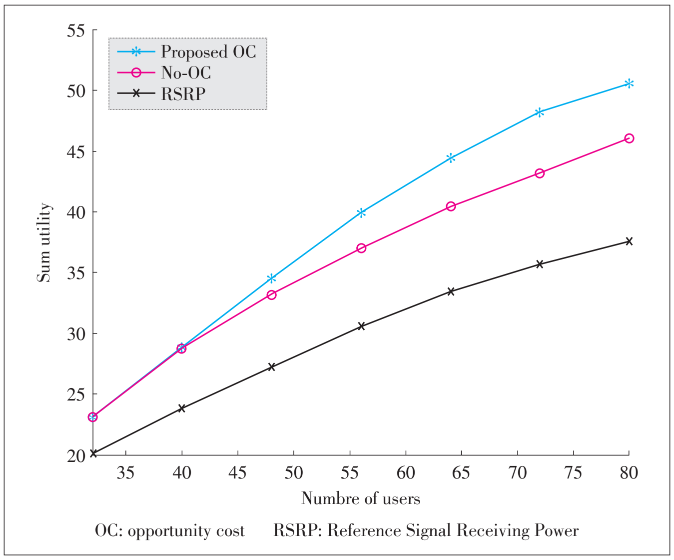
HTTP: Hypertext Transfer Protocol
IPTV: Internet Protocol television
QoS: quality of service
TCP: Transmission Control Protocol
VoIP: Voice over Internet Protocol

(No-OC) and the Reference Signal Receiving Power (RSRP) algorithm.

Fig. 3 depicts the sum utility versus number of users. Since the sum utility is the cumulative result of users´ utility, all three algorithms are monotonically increasing. The OC algorithm and the No-OC algorithm always give higher utility values than the RSRP algorithm, because they guarantee the QoS requirements of the more efficient QoS users. When the number of users is small, the OC algorithm has almost the same utility value as the No-OC algorithm. As the number of users increase, the utility value of the OC algorithm is significantly higher than that of the No-OC algorithm. This is because, when the number of users is small, the resources are sufficient, and both strategies can meet the needs of almost all users. However, as the number of users increases, BE users in the No-OC algorithm cannot be met, while the QoS users in the OC algorithm can obtain more reasonable resources based on the opportunity cost to save more resources for the BE users to implement higher system utility.

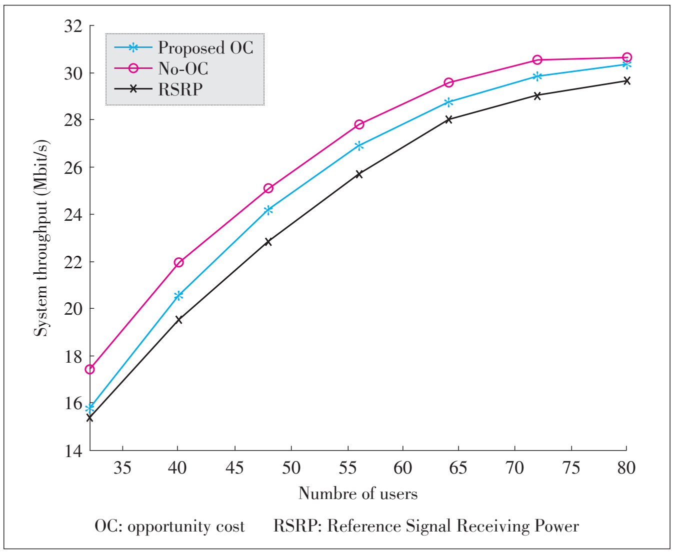
Fig. 4 shows the system throughput versus the number of users. The throughput of the three algorithms depicts an increasing trend with the increase of the number of users. The proposed OC algorithm is between No-OC and RSRP algorithm. This is due to the OC algorithm sacrificing some of the QoS user data rates. When the rate of QoS users meets their requirements, the utility value will not increase any more. Therefore, the resources allocated by QoS are allocated to BE users by OC algorithm to obtain higher system utility. Therefore, the total system throughput is somewhat lower than the No-OC algorithm.

The fairness versus the number of users is shown in **Fig. 5**. The fairness factor is defined as the user-derived data rate normalized variance. The lower the fairness factor indicates, the greater the variance and the more unfair the algorithm are. We observe that the fairness of OC algorithm is better No-OC algorithm. Because the No-OC algorithm always guarantees the QoS requirements of the QoS users and does not consider the BE users well. The OC algorithm takes advantage of the opportunity cost so that the BE users and the QoS users can make constant calculations and thus have better fairness. For RSRP algorithm, all users are equal without distinguishing different users with different QoS requirements. Therefore, its fairness is better than OC algorithm and No-OC algorithm.



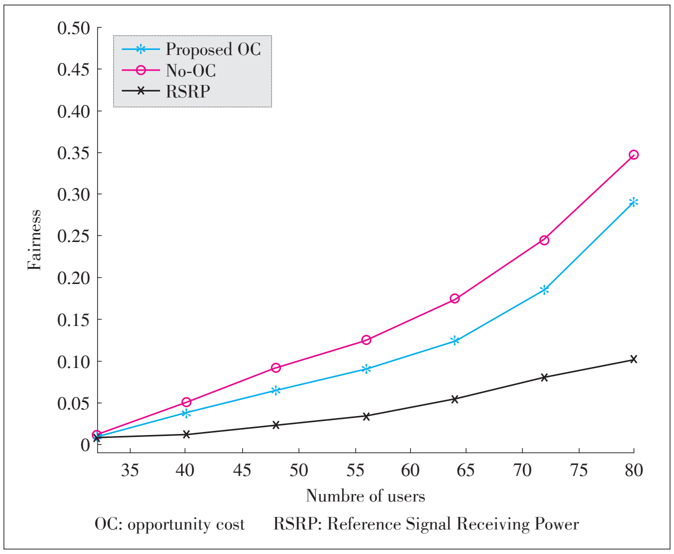
OC: opportunity cost RSRP: Reference Signal Receiving Power

▲Figure 3. Sum utility versus the number of users.



OC: opportunity cost RSRP: Reference Signal Receiving Power

▲Figure 4. System throughput versus the number of users.



OC: opportunity cost RSRP: Reference Signal Receiving Power

▲Figure 5. Fairness versus number of users.

5 Conclusions

In conclusion, the proposed multi-QoS guaranteed resource allocation algorithm for multi-services based on opportunity cost can achieve a well-done balance between user satisfaction and system fairness. We first formulate a unified utility function with effective capacity, which is able to represent the QoS requirements of delay and packet loss rate, to describe the multi-QoS metrics of different services. Then we invite the theory of opportunity cost in economy to form the concept of opportunity cost applied into the multi-service resource allocation scenario by analyzing the utility characteristics of QoS users and BE users respectively. In the multi-services resource allocation scheme, QoS users and BE users make different preferences rationally to maximize system utility. From business characteristic of different users, if boosting the service rate of QoS users when their lowest data rate is met, system utility will not be effectively increased. Meanwhile, BE users always look for a higher data rate to enhance system utility. Therefore, by calculating opportunity cost, QoS users tend to choose resources that have smaller data rates but meet their QoS requirements within delay limits. And BE users prefer resources that provide higher transmission capacity within delay limits. Finally, the simulation results show that our algorithm can achieve superior overall user satisfaction and the algorithm also balances fairness and throughput in a good way.

References

[1] R. Vannithamby and S. Talwar, "Distributed resource allocation in 5G cellular networks," in *Towards 5G:Applications, Requirements and Candidate Technologies*. New York, USA: Wiley, 2017. doi: 10.1002/9781118979846.ch8.
[2] H. Zhang, J. Du, J. Cheng, et al., "Incomplete CSI based resource optimization in SWIPT enabled heterogeneous networks: a non-cooperative game theoretic approach," *IEEE Transactions on Wireless Communications*, pp. 1882–1892, Mar. 2018. doi: 10.1109/TWC.2017.2786255.
[3] H. Y. Liu, S. Y. Xu, K. S. Kwak, et al., "Geometric programming based distributed resource allocation in ultra dense hetnets," in *IEEE 83rd Vehicular Technology Conference*, Nanjing, China, 2016, pp. 1–5. doi: 10.1109/VTCSpring.2016.7504261.
[4] X. Zhang and J. Q. Wang, "Heterogeneous QoS-driven resource allocation over MIMO-OFDMA based 5G cognitive radio networks," in *IEEE Wireless Communications and Networking Conference*, San Francisco, USA, 2017, pp. 1–6. doi: 10.1109/WCNC.2017.7925876.
[5] Q. L. Wang, Q. X. Zhang, Y. H. Sun, et al., "A QoS-guaranteed radio resource scheduling in multi-user multi-service LTE-A systems with carrier aggregation," in *IEEE 2nd International Conference on Computer and Communications*, Chengdu, China, 2016. pp. 2927–2932. doi: 10.1109/CompComm.2016.7925233.
[6] *Technical Specification Group Services and System Aspects, QoS Concept and Architecture (Release 1999)*, 3GPP TS 23.107 v3.7.0, 2002.
[7] D. P. Wu and R. Negi, "Effective capacity: a wireless channel model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, Jul. 2003. doi: 10.1109/TWC.2003.814353.
[8] L. Chen, B. Wang, X. H. Chen, X. Zhang, and D. C. Yang, "Utility-based resource allocation for mixed traffic in wireless networks," in *IEEE Conference on Computer Communications Workshops*, Shanghai, China, 2011, pp. 91–96. doi: 10.1109/INFCOMW.2011.5928944.
[9] S. Ahn, H. Wang, S. Han, and D. Hong, "The effect of multiplexing users in QoS provisioning scheduling," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2575–2581, 2010.
[10] F. M. Xue and W. F. Zhang, "Analysis of opportunity costs," *Market Modernization*, vol. 566, no. 2, pp. 73–73, 2009.
[11] F. Chi, "Research on the meaning, expression and use of opportunity cost," *Journal of Changchun University of Science and Technology*, vol. 6, no. 11, pp. 27–28, 2011.
[12] C. B. Liu, L. Shi, and B. Liu, "Utility-based bandwidth allocation for triple-play services," in *Proc. 4th European Conference on Universal Multiservice Networks*, Toulouse, France, 2007, pp. 327–336. doi: 10.1109/ECUMN.2007.58.
[13] Y. Q. Jin, X. D. Xu, Y. T. Wang, et al., "Multi-QoS mobile services guaranteed resource allocation with effective capacity," in *IEEE 3rd International Conference on Computer and Communications*, Chengdu, China, 2017.
[14] Z. J. Hao, X. D. Xu, and L. J. Li, "System utility based: resource allocation for multi-cell OFDM system," *Journal of China Universities of Posts and Telecommunications*, vol. 17, no. 2, pp. 14–19, 2010.
[15] J. Y. Cao and L. Qiu, "An effective capacity-based hybrid service resource allocation algorithm," *Journal of University of Chinese Academy of Sciences*, vol. 31, no. 11, pp. 685–690, 2014.

Manuscript received: 2018-02-28

Biographies

JIN Yaqi (yqjin@bupt.edu.cn) received the B.S. degree in communication engineering from Harbin Institute of Technology (HIT), China in 2015. She is currently pursuing the M.S. degree in information and communication engineering with Beijing University of Posts and Telecommunications (BUPT), China. Her research interests cover wireless communication, QoS guaranteed techniques, resource management including resource allocation and load balancing. She has published several papers at international conferences.

XU Xiaodong (xuxiaodong@bupt.edu.cn) received his B.S. degree in information and communication engineering and master's degree in communication and information system from Shandong University, China in 2001 and 2004 respectively. He received his Ph.D. degree of circuit and system from Beijing University of Posts and Telecommunications (BUPT), China in 2007. He is currently a professor of BUPT. He has coauthored seven books and more than 120 journal and conference papers. He is also the inventor or co-inventor of 37 granted patents. His research interests cover network architecture, moving network, coordinated multi-point and mobile network virtualization. His reasearch is supported by Beijing Nova Programme on Mobile Networking.

TAO Xiaofeng (taoxf@bupt.edu.cn) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, China in 1993, and the M.S.E.E. and Ph.D. degrees in telecommunication engineering from Beijing University of Posts and Telecommunications (BUPT), China in 1999 and 2002, respectively. He was a visiting professor with Stanford University, USA from 2010 to 2011, was the Chief Architect with the Chinese National FUTURE Fourth-Generation (4G) TDD Working Group from 2003 to 2006, and established the 4G TDD CoMP Trial Network in 2006. He is currently a professor with BUPT and a fellow of the Institution of Engineering and Technology. He is the inventor or co-inventor of 50 patents and the author or co-author of 120 papers in 4G and beyond 4G. He is currently involved in fifth-generation networking technology and mobile network security.