Received July 22, 2017, accepted August 6, 2017, date of publication August 14, 2017, date of current version September 6, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2739343

Mobile Edge Computing Enhanced Adaptive Bitrate Video Delivery With Joint Cache and Radio Resource Allocation

XIAODONG XU, (Member, IEEE), JIAXIANG LIU, AND XIAOFENG TAO, (Senior Member, IEEE)
National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Jiaxiang Liu (liujx@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471068, Grant 61421061, and Grant 61325006, in part by the National Major Project under Grant 2016ZX03001009-003, in part by the 111 Project of China under Grant B16006, and in part by the Shenzhen Science and Technology Project under Grant JSGG20150512153045135.

ABSTRACT Accompanied by the rapid development of mobile video service requirements, the dramatic increase in video streaming traffic causes a heavy burden for mobile networks. Mobile edge computing (MEC) has become a promising paradigm to enhance the mobile networks by providing cloud-computing capabilities within the radio access network (RAN). With the ability of content caching and context awareness, MEC could provide low-latency and adaptive-bitrate video streaming to improve service providing ability of the RAN. In this paper, we propose an MEC enhanced adaptive bitrate (ABR) video delivery scheme, which combines content caching and ABR streaming technology together. The MEC server acts as a controlling component to implement the video caching strategy and adjust the transmitted bitrate version of videos flexibly. A Stackelberg game is formulated to deal with the storage resources occupied by each base station. The joint cache and radio resource allocation (JCRA) is tackled into a matching problem. We propose the JCRA algorithm to solve the matching problem in order to make the cooperation between cache and radio resources. Simulation results reveal that the proposed scheme could improve both the cache hit ratio and the system throughput over other schemes.

INDEX TERMS MEC, content caching, adaptive bitrate streaming, Stackelberg game, matching theory.

I. INTRODUCTION

With the rapid development of mobile communication, the wireless data traffic has experienced an explosive growth for recent years. As one of the most popular services among mobile users, mobile video service is believed to account for more than 70% of the total mobile data traffic in 2020 [1]. Besides the high capacity requirements of mobile video service, mobile video streaming needs low latency to meet its real time characteristic. Nowadays, the demand for high-definition and low-latency mobile video is increasing significantly [2], which poses stringent challenges on the capacity of mobile networks. On the one hand, the backhaul links between base stations (BSs) and the core network are constrained on the transmission bandwidth, which could cause the congestion when fetching the required video from servers of the core network. On the other hand, the wireless channels of the radio access network (RAN) are time-varying, the actual video transmission rate varies according to the wireless channel states. To address above challenges, Mobile Edge Computing (MEC) is considered as a promising approach by providing computing and storage resources within the RAN.

MEC is proposed by the standards organization European Telecommunications Standards Institute [3]. Its main idea is to provide cloud-computing capabilities within the RAN in close proximity to mobile subscribers [4]. It overcomes the drawbacks of mobile cloud computing, which could cause long latency and backhaul bandwidth congestion, and provides the personalizing services according to the local context without the need of centralized coordination [5], [6]. The benefits of MEC consist of low latency, proximity, high bandwidth, real time radio network information and location awareness [7]. In the MEC framework, there are three parts in mobile edge host level which are mobile edge platform, mobile edge applications and virtualization infrastructure [8]. The mobile edge platform could provide the environment for the implementation of the mobile edge applications on the top of virtualization infrastructure. A number of novel services would benefit from the introduction of MEC into the RAN, such as assistance for intensive computation, content caching and intelligent video acceleration service [9].

By deploying the storage resources in close proximity to mobile users, MEC servers could provide the content caching capacity to reduce the reduplicate data traffic of mobile video service. According to the research of [10], the duplicate downloads of a small number of popular contents (e.g. popular videos) consume the most parts of backhaul links and radio resources in the RAN. To handle this challenge, the researchers have proposed the content caching technology in mobile networks to reduce the redundant content transmissions [11]. According to the research of [12], it advocates caching videos at the devices and exchanging data with device-to-device communications. As mentioned in [13], the authors enhance the RAN by the deployment of content caching devices, which are referred as helpers, to cache the video files in advance. The optimal assignment of video files is proposed in order to minimize the expected downloading time. In the investigation of [14], BSs are equipped with different cache capacities and the content placement strategy is made on the purpose of improving the cache hit ratio.

Since the storage resources at devices or BSs are limited, the effect of content caching is constrained by the number of video files that could be cached. With the help of the MEC server at the edge of the RAN, the caching capacity would be significantly improved by the expanded deployment of storage resources. When the requested videos are available in the cache of the MEC server, the requests could be responded with low latency. The experience of mobile users for mobile video service would be improved at the same time. However, it is common for mobile users to request different bitrate versions of the same video according to user preferences and network states. Thus, the content caching problem is more challenging: the caching strategy should consider not only the popularity of videos but also the proper bitrate versions to meet the demand. One traditional method to tackle this problem is to cache all bitrate versions of the video, which is inefficient for cache storages leading to low content cache hit ratio. Taking advantages of cloud computing capacity in the MEC server, the authors of [15] propose the real-time transcoding scheme for videos. The MEC server is competent enough for this computing intensive processing with its cloud computing power. Only one bitrate version of the video needs to be cached while the lower bitrate versions could be transformed at the MEC server. Consequently, MEC could explore the potential of the cooperation between content caching and cloud computing. Meanwhile, the concept of video caching is extended from video level to bitrate version level.

Besides caching videos with the proper bitrate versions, the delivery of videos is also playing an important role in mobile video services. Adaptive Bit Rate (ABR) streaming has become a popular video delivery technique which could adjust the transmission bitrate to adapt the varying network conditions. MEC could utilize its context-aware ability and support the ABR technique more effectively. In [16], the authors propose a video bitrate adaptation algorithm with the RAN caching responding to the varying wireless channel conditions. In [17], a MEC based architecture is proposed to improve the performance of adaptive video streaming. The MEC server could modify the bitrate version of the video on the fly to match current network conditions and devices capacities.

Motivated by the advantages of MEC which could introduce storage and computing resources into the RAN, we propose the MEC enhanced adaptive bitrate (MEC-ABR) video delivery scheme. In our scheme, the MEC server acts as a controlling component in the video delivery chain. Firstly, each BS competes with each other for storage resources of the MEC server to cache popular videos of its own serving users. The traffic load of BSs would be taken into consideration for storage resource allocation of the MEC server. Noting that game theory has been successfully applied to wireless communications for solving competition problems on network resources [18], [19], we model this problem as a Stackelberg game. Secondly, the radio resource allocation could be optimized by storage and computing capacity of the MEC. Finally, the caching strategy and video delivery would be adjusted flexibly to matching the varying wireless channels. The main contributions of this paper could be summarized as follows.

1) We propose a storage resource allocation scheme of the MEC server taking each BS traffic load into consideration. We formulate the problem as a Stackelberg game model in which the MEC server acts as the leader and BSs act as followers. The close-form expressions of storage amount and price for caching are derived at the equilibrium of the Stackelberg game, which reveals the impact of traffic load of BSs.

2) The caching strategy in our scheme is based on not only the popularity of the videos but also the wireless channel conditions of the RAN side. Compared the most existing work such as [12] and [13], the proposed strategy could improve the cache hit ratio and throughput from caching at the same time. The proper bitrate versions of videos are considered in our caching strategy in order to achieve high efficiency of storage resources at the MEC server.

3) The MEC server has the context-aware ability and could adjust the actual transmission bitrate version flexibly based on the QoS supporting ability of time-varying wireless channels. The adaptive bitrate video delivery could make it more effective to utilize network resources.

The rest of this paper is organized as follows. The system model is presented in the aspects of network, radio link and cache in Section II. In Section III, the utility functions and the optimization problems are formulated. In Section IV, the MEC-ABR video delivery scheme is proposed to improve the system performance. The storage allocation for caching is formulated as a Stackelberg game. The joint cache and radio resource allocation is tackled into a matching problem. Simulation results are provided to testify the advantages of our scheme in Section V. Finally, conclusions are drawn in Section VI.

II. SYSTEM MODEL

In this paper, we investigate the video delivery and network resource allocation scheme with MEC. The system model is composed of network model, radio link model and cache model.

A. NETWORK MODEL

We consider the video downlink transmission scenario in mobile networks with MEC, which is illustrated as Fig. 1. The MEC server is located in proximity of BSs to provide the low-latency mobile video service. BSs are connected with the core network and the MEC server simultaneously. The MEC server can either handle a user request and respond directly to the UE or forward the request to the core network [7]. There are N BSs deployed in the network and K users to be served. The total number of resource blocks (RBs) is M, which could be shared by all BSs.

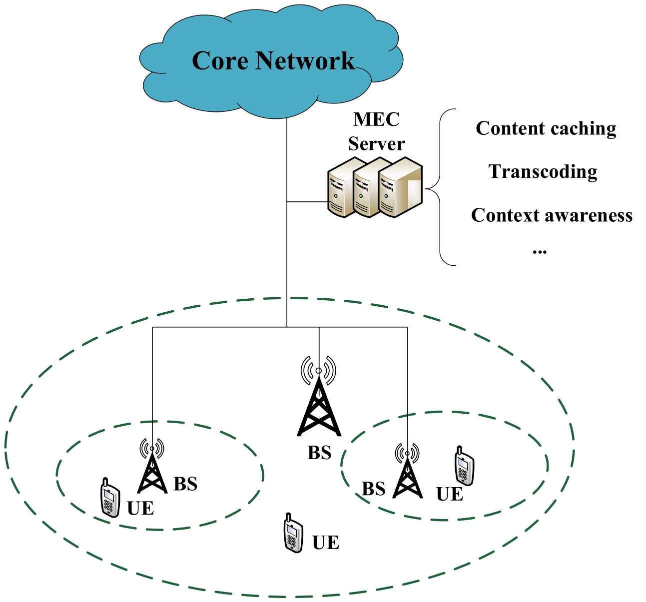


FIGURE 1. Network model for video delivery with MEC.

If the requested video is cached in the storage of the MEC server, which we call content cache hit, the MEC server can respond to the request immediately and user waiting delay is reduced significantly. Otherwise, the requirement will be sent to target servers in the core network, and then the desired videos are sent back to the user. BSs could measure the wireless channel conditions and report them to the MEC server. Thus, the MEC server is context-aware to real time radio network information. The video delivery scheme can be adjusted flexibly according to the wireless link conditions. When wireless channels are capable for transmission, the MEC server will provide the desired bitrate version and inform BS to support the service. However, if the wireless channels are experiencing severe fading, the playback stalling is the most important factor to be considered. The MEC server will choose the lower bitrate version to avoid playback stalling. In this paper, we assume the MEC server has adequate computing resources for transcoding. The impact of computing resources of the MEC server would be investigated in the future. To improve the effectiveness of content caching, the MEC server could determine the proportion of storage resources for each BS considering the wireless traffic load.

B. RADIO LINK MODEL

The wireless channel gain between BS n and user k on RB m is denoted as $|H_{k,m,n}|^2$. The bandwidth of each RB is B. For simplicity, we assume that all BSs allocate the same power P on each RB. $\beta_{k,m,n} \in \{0,1\}$ is the variable to denote whether user k is allocated RB m through BS n. Since RBs are shared by all BSs, the interference occurs when a UE uses the same RBs which are being used by the UE belongs to a neighbor BS. The noise at a user follows Gaussian distribution with a variance σ^2. Therefore, the signal-to-interference-plus-noise ratio (SINR) between user k and BS n on RB m is expressed as follow:

$$\gamma_{k,m,\mathrm{n}} = \frac{\beta_{k,m,n} P |H_{k,m,n}|^2}{\sigma^2 + \sum_{i \neq k} \sum_{j \neq n} \beta_{i,m,j} P |H_{k,m,j}|^2}. \tag{1}$$

According to Shannon capacity formula, the transmission rate achieved for the kth user on the mth RB associating nth BS can be expressed as follow:

$$r_{k,m,n} = B\log_2(1+\gamma_{k,m,n}). \tag{2}$$

Thus, the total transmission rate user k achieved could be calculated as follow:

$$r_k = \sum_m \sum_n r_{k,m,n}. \tag{3}$$

The total traffic load of BS n is calculated as follow:

$$r_n = \sum_k \sum_m r_{k,m,n}. \tag{4}$$

C. CACHE MODEL

The scope of videos requested by mobile users is defined as a video set $V = \{1, 2, \ldots, V\}$. The vth video has two bitrate versions, which are low bitrate version, with standard definition (SD), and high bitrate version, with high definition (HD). We use variable l to indicate the video version. $l = 1$ represents low bitrate version while $l = 2$ represents high bitrate version. The storage size of the vth video with the lth version is defined as s_v^l. The caching capacity of the MEC server is defined as S. The amount of storage resources allocated to BS n is S_n. Let $x_{n,v}^l \in \{0, 1\}$ be the content caching variable which indicates the decision of whether to cache the video v of version l by BS n. We have $x_{n,v}^l = 1$ when the video v of version l is cached by BS n, $x_{n,v}^l = 0$ otherwise. The videos cached for each BS could not exceed S_n, which is expressed by the following formula:

$$\sum_{v}\sum_{l} x_{n,v}^{l} s_{v}^{l} \leq S_n. \tag{5}$$

Both the popularity and quality preference of videos are taken into considerations in our scenario. Without loss of generality, video set V is sorted in descending order of popularity. The probability of users requesting video v is defined as z_v, which follows the Zipf distribution [20] as follow:

$$z_v = \frac{v^{-\alpha}}{\sum\limits_{v=1}^{V} v^{-\alpha}}, \tag{6}$$

where $\alpha > 0$ is the Zipf parameter which indicates the skewness degree of popularity distribution. A higher α corresponds to more video requests concentrated in the popular videos.

The user k's preference of video v with bitrate version l is denoted as $g_k^l(v)$ which is related to different video types. For example, users prefer to select low bitrate version to ensure the fluency of playback if the video is a sport event video, and when the video corresponds to films, users prefer high bitrate version to ensure the watching experience. Thus, the probability of user k requesting for video v with version l is expressed as following formula:

$$f_{k,v}^{l} = \frac{v^{-\alpha}}{\sum\limits_{v=1}^{V} v^{-\alpha}} g_k^l(v). \tag{7}$$

III. PROBLEM FORMULATION

In this section, the utility functions of the MEC server, BSs and the RAN are provided respectively. Based on the utility functions, the optimization problems are formulated. For the MEC server the revenue is defined as the total charge from BSs for caching videos at the MEC server, which are constantly requested by its serving users. The MEC server determines that the price for caching is y per unit storage size. The amount of storage resources purchased by BS n is denoted as S_n. Thus, the utility function of the MEC server can be formulated as follow:

$$U_{MEC} = y \sum_{n=1}^{N} S_n. \tag{8}$$

Given the unit storage price y charged by the MEC server, each BS responds to purchase an optimal amount S_n for caching. According to [21], we define the utility function of BS n as follow:

$$U_{BS}^{n} = \log(1 + \frac{S_n}{\frac{1}{r_n} J_n}) - y S_n, \tag{9}$$

where $J_n = \sum\limits_{i \neq n} S_i$ is the total storage resources occupied by other BSs except BS n. The traffic load r_n has an effect on the revenue of BS n for the reason that more traffic load means more users could utilize the cached videos. The revenue of BS n is also increasing with the respect of S_n. Indeed, the more videos are cached at the MEC server for BS n, the higher is the probability that requests from its serving users could be corresponded locally. The cost of BS n is the charge paid to the MEC server for its desired amount of storage resources.

Based on the storage resources allocated for caching, BSs could utilize them to cache the popular videos with suitable bitrate version among their serving users. The utility of the RAN is related to caching strategy and radio resource allocation. We focus on the transmission rate of videos which are available in the cache of the MEC server. The utility function of the RAN could be expressed as follow:

$$U_{RAN} = \sum_{n}\sum_{k}\sum_{v}\sum_{l} x_{n,v}^{l} f_{k,v}^{l} r_k. \tag{10}$$

Based on the aforementioned utility functions, the optimization problems could be formulated as follows.

Ploblem 1: For the MEC server, the optimization problem could be formulated as follow:

$$\begin{aligned} &\max \ U_{MEC}(y|S_n^*), \\ &s.t. \ \sum_{n} S_n^* \leq S, \end{aligned} \tag{11}$$

where S_n^* represents the optimal amount of storage resources purchased by BS n, the constraint reveals the total amount of storage resources for caching could not exceed caching capacity of the MEC server.

Ploblem 2: For BSs, the optimization problem could be expressed as follow:

$$\begin{aligned} &\max \ U_{BS}^{n}(S_n|y^*), \\ &s.t. \ S_n \geq 0, \\ &\qquad n = 1, 2, \ldots, N, \end{aligned} \tag{12}$$

where y^* is the optimal price set by the MEC server and BS n corresponds to the amount of storage resources that can maximize its own utility function.

Ploblem 3: For utility of the RAN, the optimization of content caching strategy and RB allocation is formulated as follow:

$$\begin{aligned} &\max U_{RAN}(x_{n,v}^{l}, \beta_{k,m,n}|S_n^*), \\ &s.t. \ C1{:} \sum_{k} \beta_{k,m,n} \leq 1, \\ &\quad C2{:} \sum_{v}\sum_{l} x_{n,v}^{l} s_{v}^{l} \leq S_n, \\ &\quad C3{:} l' \leq l, \\ &\quad C4{:} r_k \geq r_{k,v}^{l'}, \\ &\quad C5{:} x_v^l \in \{0, 1\}, \\ &\quad C6{:} \beta_{k,m,n} \in \{0, 1\}. \end{aligned} \tag{13}$$

Constraints C1, C2 enforce the radio and storage resource limits. Constraint C1 states that RB m could be at most allocated to one user within the same BS in order to avoid interference. Constraint C2 implies that the total video files cached for BS n should not exceed the amount of storage resources allocated to BS n. Constraints C3, C4 enforce the video version constraints. Constraint C3 means that the MEC server could adjust the original video request version l to a lower level l' due to the deterioration of wireless channel conditions. Constraint C4 states that the actual wireless transmission rate of user k should be capable to support the video bitrate version l'. Finally, Constraints C5 and C6 ensure the discrete and binary nature of optimization variables.

IV. MEC ENHANCED ADAPTIVE BITRATE VIDEO DELIVERY SCHEME

In this section, we propose the MEC-ABR video delivery scheme to improve system performances. The scheme could tackle the three optimal problems into a sequential decision making process. At first, the MEC server predicts the amount of storage resources purchased by each BS and set the price for unit storage. Furthermore, BSs compete with each other and make decision of the amount purchased to maximize its own utility function. Finally, the many-to-many matching between BSs and users is conducted to make the joint cache and radio resource allocation.

A. STACKELBERG GAME FOR STORAGE RESOURCE ALLOCATION

We tackle the storage resource allocation problem into a Stackelberg game which is suitable to model the interaction between leaders and followers. In our scheme, the Stackelberg game is played between the MEC server and N BSs. The MEC server acts as the leader that provides the caching price y to BSs. BSs act as followers to determine the amount of storage resources S_n purchased from the MEC server according to the price set by the leader. On the one hand, every BS wishes to cache the popular videos of its serving users as many as possible to improve the cache hit ratio. On the other hand, the price set by the the MEC server cause the cost of caching.

Definition 1: Let $S_{-n} = [S_1, \ldots, S_{n-1}, S_{n+1}, \ldots, S_N]$ *be the strategy of BSs except BS n.* (y^*, S_n^*) *is a Stackelberg equilibrium point if for any* (y, S_n) *with* $y \geq 0$, $S_n \geq 0$ *the following conditions satisfied.*

$$U_{MEC}(y^*, S_n^*) \geq U_{MEC}(y, S_n^*),$$
$$U_{BS}^n(y^*, S_n^*, S_{-n}^*) \geq U_{BS}^n(y^*, S_n, S_{-n}^*). \tag{14}$$

Usually, the Stackelberg equilibrium could be obtained by finding its perfect Nash Equilibrium [22]. For a non-cooperative game, the Nash Equilibrium is defined as none of players could improve its utility function by changing its strategy unilaterally. Since the price set by the MEC server is based on the best respond of BSs, we use a backward induce method to find the solution. Thus, we first solve the **Problem 2** with a fixed price y to find the best respond of BSs. Then we substitute the respond function into **Problem 1** and find out the optimal price. At BS side, the purpose of each BS is to maximize its own utility function. The first derivative of U_{BS}^n with respect to S_n is calculated as follow:

$$\frac{\partial U_{BS}^n}{\partial S_n} = \frac{r_n}{r_n + r_n S_n + J_n} - y. \tag{15}$$

Solving $\frac{\partial U_{BS}^n}{\partial S_n} = 0$ we obtain the following solution

$$S_n = \frac{1}{y} - 1 - \frac{J_n}{r_n}. \tag{16}$$

The second derivative of U_{BS}^n with respect to S_n is calculated as follow:

$$\frac{\partial^2 U_{BS}^n}{\partial {S_n}^2} = -(\frac{r_n}{r_n + r_n S_n + J_n})^2 < 0, \tag{17}$$

which guarantees the solution of (16) could achieve the maximum value of BS utility function. Thus, the best respond of BS n could be expressed as follow:

$$S_n^{BR} = (\frac{1}{y} - 1 - \frac{J_n}{r_n})^+, \tag{18}$$

where $()^+$ is a function to ensure the non-negative value of S_n^{BR}.

Theorem 1: Based on the price y set by the MEC server, the equilibrium of each BS is achieved by the following expression:

$$S_n^* = (\frac{1}{y} - 1)\frac{a_n b_n}{R}, \tag{19}$$

where a_n *is*

$$a_n = \begin{cases} 1 - \frac{1}{r_n} - \frac{1 - r_1}{r_1} \sum_{l=2, l \neq n}^{N} \frac{r_n - r_l}{(r_l - 1) r_n} & n \neq 1 \\ 1 - \frac{1}{r_1} - \frac{1 - r_N}{r_N} \sum_{l=2}^{N} \frac{r_1 - r_l}{(r_l - 1) r_1} & n = 1 \end{cases} \tag{20}$$

and b_n *is*

$$b_n = \begin{cases} \prod_{l=2, l \neq n}^{N} (1 - \frac{1}{r_l}) & n \neq 1 \\ \prod_{l=2}^{N-1} (1 - \frac{1}{r_l}) & n = 1 \end{cases} \tag{21}$$

and R is

$$R = (1 - \frac{1 - r_N}{r_N} \sum_{l=1}^{N-1} \frac{1}{r_l - 1}) \prod_{l=1}^{N-1} (1 - \frac{1}{r_l}). \tag{22}$$

Proof: Please refer to Appendix ☐

After predicting each BS respond of price y, the MEC server will solve **Problem 1** to find out the optimal price y^*. By substituting (19) to (8) the utility function of the MEC server could be rewritten as follow:

$$U_{MEC} = (1 - y) \sum_{n=1}^{N} \frac{a_n b_n}{R}. \tag{23}$$

The first derivation of U_{MEC} with respect to y is

$$\frac{\partial U_{MEC}}{\partial y} = -\sum_{n=1}^{N} \frac{a_n b_n}{R} < 0. \tag{24}$$

That means the utility of the MEC server is monotonous decreasing with respect to price y. Furthermore, the caching capacity of the MEC server is finite.

$$\sum_{n} S_n^* = (\frac{1}{y} - 1)\sum_{n} \frac{a_n b_n}{R} \leq S. \tag{25}$$

Solving (25) we could derive the price y has low bound as follow:

$$y \geq \frac{\sum_{n} a_n b_n}{RS + \sum_{n} a_n b_n}. \tag{26}$$

Thus, the optimal price set by the MEC server is expressed as follow:

$$y^* = \frac{\sum_{n} a_n b_n}{RS + \sum_{n} a_n b_n}. \tag{27}$$

Theorem 2: In the proposed Stackelberg game, a unique Stackelberg equilibrium exists.

Proof: The proposed Stackelberg game is composed of two aspects. The MEC server and BSs optimize their utility functions and achieve the Nash equilibrium respectively: the price set by the MEC server is obtained in (27), the amount of storage resources purchased by BSs is derived in (19). Since each aspect exists a Nash Equilibrium, our proposed Stackelberg game exists the Stackelberg equilibrium. Furthermore, the Nash Equilibrium at each aspect is unique. Therefore, the Stackelberg equilibrium of our proposed Stackelberg game is unique. □

B. JOINT CACHE AND RADIO RESOURCE ALLOCATION ALGORITHM

Given the BS n has gotten S_n^* storage resources from the MEC server, it could cache the popular videos among its serving users to reduce duplicate downloads from the core network. Under the perfect knowledge of user download history, which is easy to conduct in MEC scenario, BSs could evaluate the popular profile of video v with version l in its own domain.

The connection between BSs and users through RBs is the critical part of Problem 3. Inspired by [25], we formulate a many-to-many matching problem. Predicting the popularity of users requests, BS n maintains a preference list L_n, which illustrates the priority of user k associates with BS n. The value of each item in L_n is $f_{k,v}^l$ in descending order. As to user aspect, it would like to choose the BS and RB which could provide a high transmission rate. The preference of user k is defined as L_k, which is $r_{k,m,n}$ in descending order. Based on the preference lists of BSs and users, we propose the Joint Cache and Radio resource Allocation (JCRA) algorithm in Algorithm 1.

As shown in Algorithm 1, we initiate a pointer for each user which is pointing to the first item in its preference list. For each round, users propose the requests to BSs for RBs which could provide the suitable video transmission rate. According to the requests from users, each BS conducts a temporary set T_n to record the potential users that may be allocated resources from the BS. In the aspect of BSs, they prefer to corresponds to the requests for the most popular videos to improve the reuse of content caching. Thus, they follow their popularity based preference list to make the cache strategy and radio resource allocation. If the user has achieved its required transmission rate, the pointer of it would not change any more. Otherwise, it would participate the next round of

Algorithm 1 JCRA Algorithm

```
Input:
    BS preference list, $L_n$;
    User preference list, $L_k$;
    Storage resources for caching, $S_n^*$;
Output:
    Caching decision, $x_{n,v}^l$;
    RB allocation decision, $\beta_{k,m,n}$;
    Actual transmission version, $l'$
Initialize $x_{n,v}^l = 0$ for $\forall v \in V, \forall l \in L$, $\beta_{k,m,n} = 0$ for $\forall k \in K, \forall m \in M, \forall n \in N$;
while all unsatisfied users have not scanned all the BSs in their preference lists do
    for user $k$ do
        if $r_k < r_{k,v}^l$ then
            The pointer of $L_k$ moves to the next position and make the request for RB to BSs;
        else
            The pointer keeps current position in the list $L_k$;
        end if
    end for
    if BS $n$ have enough RB resources and caching resources then
        Construct the temporary set $T_n$ of users who propose request to BS $n$;
        BS $n$ choose the most preferred users in $T_n$ based on $L_n$, and set $x_{n,v}^l = 1$, $\beta_{k,m,n} = 1$ according to their requests;
    end if
end while
The MEC server adjusts the video bitrate version
for user $k$ do
    if $r_k \geq r_{k,v}^l$ then
        $l' = l$;
    else
        The MEC server transcodes the video into a lower quality version $l$ which requests a lower transmission bitrate;
        while $r_k < r_{k,v}^{l'}$ do
            $l' - -$;
        end while
    end if
end for
return $x_{n,v}^l$, $\beta_{k,m,n}$, $l'$;
```

matching and the pointer moves to the next position. The process of matching would not terminate until all unsatisfied users have scanned all the BSs in their preference lists. For the users who have not achieved the suitable transmission rate for the desired quality of videos, the MEC server would adjust a lower bitrate version for transmission. That guarantees the wireless link could meet the required transmission rate of the video bitrate version, which could improve the resource efficiency and user experience at the same time.

In summary, the MEC-ABR video delivery scheme which we propose takes full consideration of the radio and storage resources. Based on the traffic loads of BSs, the MEC server decides the storage resources occupied by each BS. The caching strategy in our scheme consider not only the popularity and proper bitrate version of videos but also the wireless channel conditions. Thus, the throughput from caching could be improved significantly. Furthermore, the MEC server could adjust the actual transmission video version flexibly based on the QoS supporting ability of wireless channels with the help of its transcoding abilities.

V. PERFORMANCE EVALUATIONS

In this section, simulation results are presented to evaluate the performance of our MEC-ABR video delivery scheme. There are 50 RBs in the system which are shared by all BSs. The scope of videos is a library of 1000 files with the popularity following a Zipf distribution with parameter $\alpha = 0.95$ as [23]. The bitrate version preference is $g_k^1(v) = \frac{v-1}{V-1}$, $g_k^2(v) = 1 - \frac{v-1}{V-1}$ following [2]. The bitrate of HD quality videos is set as 2Mbps while the bitrate of SD quality videos is 0.45 of HD video bitrate [16]. More details of simulation settings are listed in Table 1.

TABLE 1. Simulation settings.

System Parameters	
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
Fast fading	Rayleigh fading
Noise Density	-174 dBm/Hz
Video Parameters	
Number of videos	1000
Zipf parameter	0.95
HD video size	20 MB
SD video size	9 MB
HD video bitrate	2 Mbps
SD video bitrate	0.9 Mbps

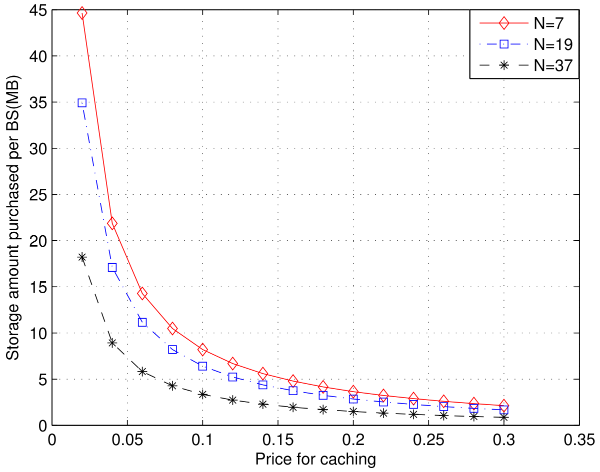


FIGURE 2. Storage amount purchased per BS versus price for caching.

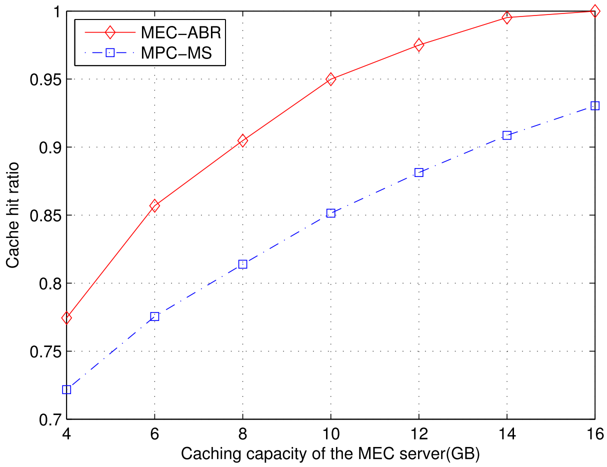


FIGURE 3. Cache hit ratio versus caching capacity of the MEC server.

We first investigate the relationship between the price for caching and the amount of storage resources purchased per BS. The user number K is set as 500. As shown in Fig. 2, the amount of storage resources purchased per BS decreases as the price for caching increases, which is coincident with the analysis of the Stackelberg game. Besides the influence of price for caching, the total number of BSs N would also have an impact on the result of the game. We simulate three cases, as $N = 7, 19, 37$, to illustrate the influence of BS number N in Fig. 2. As more BSs take part into the competition for storage resources of the MEC server, the amount occupied by one BS decrease. That could be explained in two aspects: when the number of BSs is small the storage resources of the MEC server are adequate relatively, one BS could be allocated more storage resources for caching. Another reason is that, compared with the large number of BSs, the traffic load is higher when the BS number is small. Our algorithm would allocate more storage resources of the MEC server correspondingly since that the demand for caching is increasing.

With the storage resources purchased from the MEC server, each BS utilizes them to cache the popular videos of its own serving mobile users. The cache hit ratio is investigated in Fig. 3 revealing the performance of our proposed MEC-ABR video delivery scheme. In this paper, the case of the adjustment of video bitrate version is also calculated into cache hit ratio. The number of mobile users is set as 800 and the BS number is 19 in this simulation. The result demonstrates that the cache hit ratio increases with the improvement of caching capacity. The reason is that more videos could be cached at the MEC server for mobile users. We use the MPC-MS scheme as the comparison scheme, which is proposed in [24]. Its main idea is to cache the most popular contents until the cache is full and users are associated with the BS with the maximum-SINR. As shown in Fig. 3, MEC-ABR video delivery scheme achieves about 10% gain to the MPC-MS scheme in term of cache hit ratio. Although the MPC-MS scheme cache the most popular videos of the network, they may not include the exact requirement of scheduled users. Briefly, the MPC-MS scheme neglects the relationship between caching and radio resource allocation. In our proposed MEC-ABR video delivery scheme, the MEC server could get the real-time information of the wireless channels which could have an impact of video caching strategy. The joint cache and radio resource allocation improves the system performance significantly. Furthermore, MEC-ABR video delivery scheme could select the proper video bitrate version for caching based on the user preference, which is more efficient for storage resources.

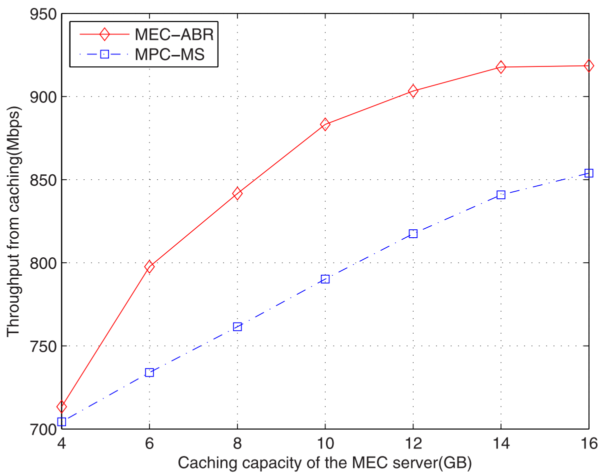


FIGURE 4. **Throughput from caching versus caching capacity of the MEC server.**

Since the cache hit ratio is improved significantly in our proposed MEC-ABR video delivery scheme, it is a natural idea to investigate the throughput from caching further. The simulation is operated in the environment of 19 BSs with 800 mobile users. The numerical results in Fig. 4 demonstrate that the proposed MEC-ABR video delivery scheme also outperforms the MPC-MS scheme in the aspect of throughput from caching. Several important insights could be revealed through the simulation. Firstly, the caching capacity of the the MEC server affects the throughput from caching significantly. When the caching capacity of the MEC server is small, the proposed MEC-ABR video delivery scheme has as similar performance as that achieved by the MPC-MS scheme. Secondly, when the caching capacity of the MEC server is larger, the advantage of our proposed scheme is revealed apparently. This reveals that the joint optimization of cache and radio resource allocation could explore more potential of mobile networks. Finally, when caching capacity of the MEC server is large enough, radio resources become the main influence factor of the system, so the system throughput from caching does not improve any more.

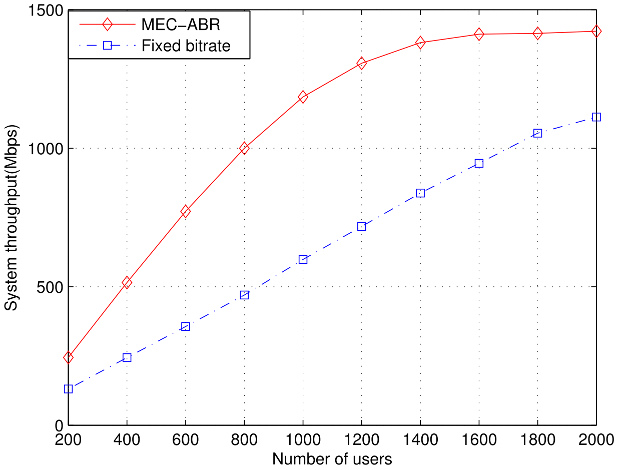


FIGURE 5. **System throughput versus number of users.**

Finally, we investigate the influence of adjusting the transmission bitrate of videos according to the wireless channel conditions. The caching capacity of the MEC server is set as 10GB and the number of BSs is 19. As shown in Fig. 5, the system throughput improves firstly and then tends to saturation with the number of users rising. MEC-ABR video delivery scheme performs better than the one without bitrate adjustment. In our proposed scheme, the MEC server could control the actual transmission bitrate version of videos to adapt the wireless channel conditions. Thus, the potential of wireless channels could be explored by the assistant of the MEC servers. The real-time information of wireless conditions achieved by the MEC server could be utilized to make more intelligent decisions in the aspects of radio resource allocation and video delivery scheme. When the channel state deteriorates, the MEC server could change the lower bitrate version flexibly to avoid the stalling of playback and improve the system throughput at the same time.

VI. CONCLUSION

In this paper, we propose the MEC-ABR video delivery scheme to enhance the mobile video service providing ability of the RAN. As to the storage resource allocation of the MEC server, a Stackelberg game is formulated where the MEC server acts as the leader and BSs act as followers. The MEC server predicts the reaction of BSs and charge the price for caching. The close-form expressions of storage amount and price for caching are derived at the equilibrium of the game. The joint cache and radio resource allocation is tackled into a matching problem. BSs and users maintain prefer lists according to popularity of videos and wireless channel conditions respectively. The JCRA algorithm is proposed to solve the problem and the video bitrate version adjustment is also considered. Simulation results demonstrate that performance gains of our proposed MEC-ABR video delivery scheme in the aspects of cache hit ratio and system throughput.

Therefore, it can be concluded that the proposed scheme gives a promising method to utilize MEC for mobile video service, and provides several important insights for interaction among radio, storage and computing resources.

APPENDIX

In order to find the equilibrium of the Stackelberg game, we could solve N best responds of BSs according to (18) jointly. The matrix function is denoted as follow:

$$\begin{bmatrix} 1 & \frac{1}{r_1} & \cdots & \frac{1}{r_1} \\ \frac{1}{r_2} & 1 & \cdots & \frac{1}{r_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{r_N} & \frac{1}{r_N} & \cdots & 1 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{bmatrix} = \begin{bmatrix} \frac{1}{y} - 1 \\ \frac{1}{y} - 1 \\ \vdots \\ \frac{1}{y} - 1 \end{bmatrix}, \quad (28)$$

we define that

$$R = \begin{bmatrix} 1 & \frac{1}{r_1} & \cdots & \frac{1}{r_1} \\ \frac{1}{r_2} & 1 & \cdots & \frac{1}{r_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{r_N} & \frac{1}{r_N} & \cdots & 1 \end{bmatrix}, \quad (29)$$

$$Y = \begin{bmatrix} \frac{1}{y} - 1 \\ \frac{1}{y} - 1 \\ \vdots \\ \frac{1}{y} - 1 \end{bmatrix}. \quad (30)$$

The amount of storage resources purchased by BS n is solved by the Cramer's rule.

$$S_n = \det(R_n)/\det(R), \quad (31)$$

where R_n is formed by replacing the n-th column of R by the column vector Y. det () is the function to get the determinant of the matrix.

REFERENCES

[1] "Cisco visual networking index: Forecast and methodology," Ciscio, San Jose, CA, USA, White Paper 2014–2019," May 2015.

[2] L. Wu and W. Zhang, "Caching-based scalable video transmission over cellular networks," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1156–1159, Jun. 2016.

[3] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," Eur. Telecommun. Standards Inst., Sophia Antipolis, France, White Paper 11, 2015.

[4] M. Patel *et al.*, "Mobile-edge computing," Eur. Telecommun. Standards Inst., Sophia Antipolis, France, White Paper, 2014.

[5] O. Makinen, "Streaming at the Edge: Local service concepts utilizing mobile edge computing," in *Proc. 9th Int. Conf. Next Generat. Mobile Appl. Services Technol.*, Cambridge, MA, USA, 2015, pp. 1–6.

[6] E. Cau *et al.*, "Efficient exploitation of mobile edge computing for virtualized 5G in EPC architectures," in *Proc. 4th IEEE Int. Conf. Mobile Cloud Comput. Serv. Eng. (MobileCloud)*, Oxford, U.K., Apr. 2016, pp. 100–109.

[7] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[8] ETSI. (2016). *Mobile Edge Computing (MEC); Framework and Reference Architecture*. [Online]. Available: http://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.

[9] ETSI. (2015). *Mobile-Edge Computing (MEC); Service Scenarios*. [Online]. Available: http://www.etsi.org/deliver/etsi_gs/MEC-IEG/001_099/004/01.01.01_60/gs_MEC-IEG004v010101p.pdf

[10] B. A. Ramanan, L. M. Drabeck, M. Haner, N. Nithi, T. E. Klein, and C. Sawkar, "Cacheability analysis of HTTP traffic in an operational LTE network," in *Proc. Wireless Telecommun. Symp. (WTS)*, Phoenix, AZ, USA, 2013, pp. 1–8

[11] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[12] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[13] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[14] B. Serbetci and J. Goseling, "On optimal geographical caching in heterogeneous cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Sep. 2017, pp. 1–6.

[15] S. Dutta, T. Taleb, P. A. Frangoudis, and A. Ksentini, "On-the-fly QoE-aware transcoding in the mobile edge," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Apr. 2016, pp. 1–6.

[16] H. A. Pedersen and S. Dey, "Enhancing mobile video capacity and quality using rate adaptation, RAN caching and processing," *EEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 996–1010, Apr. 2016.

[17] Y. Li, P. A. Frangoudis, Y. Hadjadj-Aoul, and P. Bertin, "A mobile edge computing-based architecture for improved adaptive HTTP video delivery," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Berlin, Germany, Oct. 2016, pp. 1–6

[18] J. Li, J. Sun, Y. Qian, F. Shu, M. Xiao, and W. Xiang, "A commercial video-caching system for small-cell cellular networks using game theory," *IEEE Access*, vol. 4, pp. 7519–7531, 2016.

[19] Z. Yin, F. R. Yu, S. Bu, and Z. Han, "Joint cloud and wireless networks operations in mobile cloud computing environments with telecom operator cloud," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 4020–4033, Jul. 2015.

[20] M. Cha *et al.*, "I Tube, You Tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, New York, NY, USA, 2007, pp. 1–14.

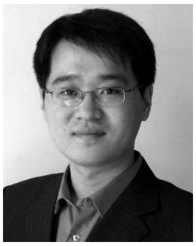
[21] F. Shen, K. Hamidouche, E. Bastug, and M. Debbah, "A Stackelberg game for incentive proactive caching mechanisms in wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Apr. 2016, pp. 1–6.

[22] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2115–2129, Aug. 2016.

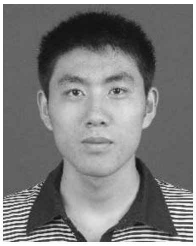
[23] R. Yu *et al.*, "Enhancing software-defined RAN with collaborative caching and scalable video coding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, Sep. 2016, pp. 1–6.

[24] Y. Wang, X. Tao, X. Zhang, and G. Mao, "Joint caching placement and user association for minimizing user download delay," *IEEE Access*, vol. 4, pp. 8625–8633, 2016.

[25] H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, and Z. Han, "Computing resource allocation in three-tier IoT fog networks: A joint optimization approach combining Stackelberg game and matching," *IEEE Internet Things J.*, to be published.



XIAODONG XU (S'06–M'07) received the B.S. degree in information and communication engineering and the master's degree in communication and information system from Shandong University in 2001 and 2004, respectively, and the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications (BUPT) in 2007. He is currently a Professor with BUPT. He has co-authored seven books and over 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, and coordinated multi-point and mobile network virtualization. He was supported by the Beijing Nova Programme on mobile networking.



JIAXIANG LIU received the B.S. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2015, where he is currently pursuing the M.S. degree in information and communication engineering. He has authored or co-authored several papers in journals and international conferences. His research interests cover wireless communication, mobile edge computing, and caching network.

XIAOFENG TAO (S'99–A'02–M'03–SM'13) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1993, and the M.S.E.E. and Ph.D. degrees in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2002, respectively. He was a Visiting Professor with Stanford University, Stanford, CA, USA, from 2010 to 2011; 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.

• • •