Received May 31, 2016, accepted June 16, 2016, date of publication August 26, 2016, date of current version October 6, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2597169

Energy-Efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks

KE ZHANG[1], YUMING MAO[1], SUPENG LENG[1], (Member, IEEE), QUANXIN ZHAO[1], (Student Member, IEEE), LONGJIANG LI[1], (Member, IEEE), XIN PENG[2], LI PAN[2], SABITA MAHARJAN[3], (Member, IEEE), AND YAN ZHANG[3], (Senior Member, IEEE)

[1]School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China
[3]Simula Research Laboratory, Oslo 1364, Norway

Corresponding author: Y. Zhang (yanzhang@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61273235, and in part by the Research Council of Norway under Project 240079/F20.

ABSTRACT Mobile edge computing (MEC) is a promising paradigm to provide cloud-computing capabilities in close proximity to mobile devices in fifth-generation (5G) networks. In this paper, we study energy-efficient computation offloading (EECO) mechanisms for MEC in 5G heterogeneous networks. We formulate an optimization problem to minimize the energy consumption of the offloading system, where the energy cost of both task computing and file transmission are taken into consideration. Incorporating the multi-access characteristics of the 5G heterogeneous network, we then design an EECO scheme, which jointly optimizes offloading and radio resource allocation to obtain the minimal energy consumption under the latency constraints. Numerical results demonstrate energy efficiency improvement of our proposed EECO scheme.

INDEX TERMS Energy-efficiency, offloading, mobile edge computing, 5G.

I. INTRODUCTION

As smart mobile devices have seen advanced technology and design, they facilitate us with a pervasive and powerful platform to realize many novel mobile applications [1], [2]. Mobile applications, such as the interactive gaming, virtual reality and natural language processing, typically require intensive computation and result in high energy consumption [3]–[5]. However smart mobile devices have limited computation capabilities and battery power. This conflict between the resource hungry applications and the limited capability of the smart mobile devices brings in unprecedented challenges to implement the novel mobile applications in an energy efficient manner.

A new architecture and technology known as Mobile Cloud Computing (MCC) has the potential to address the aforementioned challenges. By migrating computational tasks from the mobile devices to the infrastructure-based cloud servers, MCC can improve the performance of mobile applications and reduce the energy consumption of mobile devices [6]. However, the infrastructure-based cloud servers are always located centrally in the core network and far away from the mobile devices. The long transmission from the mobile devices to the cloud servers may cause delay fluctuation and invoke extra transmission energy cost [7]. Thus, the computation offloading efficiency can severely degrade.

Mobile Edge Computing (MEC) is envisioned as a promising approach to improve the offloading efficiency. In the MEC framework, cloud computing capabilities are provided within the radio access network in close proximity to these mobile devices [8]. In other words, with the aid of MEC, mobile devices are enabled to offload their tasks to the MEC servers on the edge of the network, rather than utilizing the servers in the core network. This MEC paradigm can provide low latency, high bandwidth and computing agility in the computation offloading process.

With the ever-growing energy consumption for information and communication technology, the communication devices and infrastructure play an important role in global greenhouse gas emissions [9]. Therefore, the development of green 5G networks has become an important topic for the design and implementation of future wireless communications [10]. As MEC is a key component of 5G networks, the energy efficiency has become a mainstream concern for the design of the MEC mechanism.

In this paper, we focus on the design of an energy-efficient computation offloading mechanism for MEC in 5G heterogeneous networks. With the MEC computation offloading, the energy consumption for accomplishing the computation tasks includes two parts. The first part is the energy spent on transmitting the computation files to the MEC servers. Due to the variable wireless channel states and the different sizes of the computation files, the energy consumption for transmission may vary among the mobile devices. Furthermore, in the case where the mobile devices share the radio resources with each other, they may cause severe interference to each others [11]–[13]. The interference will decrease the transmission rates for the files, and hence reduce the energy efficiency of the MEC offloading. As a result, an efficient transmission control scheme is needed in the MEC offloading. The second part is the energy spent on the computing. This amount mainly depends on the computation capabilities of the mobile devices and the MEC servers.

Each mobile device can decide whether to offload its task to the MEC servers for remote computing or to accomplish the task locally on its own device. This decision is made by comparing the energy costs. However, due to the limited communication resources, the interactivity between these mobile devices in the transmission process may affect the transmission energy cost of each device. This effect makes the MEC offloading couple with the wireless resource allocation. Furthermore, considering different QoS constraints required for the computation tasks and the variable computation capabilities of these devices, archiving an energy-efficient offloading by coordinating wireless transmission and task implementation among the mobile devices and the MEC servers is a challenging task.

In this paper, we design an energy-efficient MEC offloading mechanism for mobile devices in 5G heterogeneous networks. This mechanism minimizes the system energy consumption and concurrently ensures the latency constraints of the computation tasks. The main contributions of this paper are as follows:

- We present a multi-device computation offloading framework for mobile edge cloud computing in 5G heterogeneous networks.
- To cope with the multi-access characteristics of 5G heterogeneous networks, we formulate an energy-efficient optimization problem that minimizes the system energy consumption while satisfying the latency constraints.
- In order to overcome the complexity of solving the optimization problem, we design a three-stage energy-efficient computation offloading scheme. In this scheme, through type classification and priority assignment for the mobile devices, the optimization problem can be solved in polynomial complexity.

The rest of the paper is organized as follows. In Section II, we review related work. In Section III, we present the framework of multi-device MEC offloading in a 5G heterogeneous network. The energy-efficient optimization problem is formulated in Section IV. The energy-efficient MEC computation offloading schemes are described in Section V. Performance evaluation is presented in Section VI. Finally, we conclude our work in Section VII.

II. RELATED WORK

The development of cloud computing and virtualization techniques provides an efficient way to decouple the application environment from the underlying hardware resources, and thus greatly improves the utilization of available computing resources [14]–[18]. MCC, which has evolved from cloud computing, is designed to address the computation requirements of new smart mobile phone based applications [19].

In recent years, several studies have addressed the mobile computation offloading in the MCC scenario. In [20], the authors formulated the computation offloading decision of mobile users as a decentralized game, and proposed a game theoretic approach to achieve the efficient computation offloading. Considering the local loads of mobile users and the availabilities of cloudlets, the authors in [21] proposed a Markov decision process based dynamic offloading scheme for mobile users in an intermittently connected cloudlet system. In [22], the authors studied the multi-user computation partitioning problem in a large scale mobile cloud application scenario, and designed an offline heuristic algorithm to minimize the average completion time for all users. In order to form an elastic mobile computing grid, the authors in [23] proposed a resource provisioning framework for organizing the heterogeneous devices in the vicinity. In [24], the authors investigated the impacts of the geographical distribution of cloud resources on the cloud-based mobile augmentation performance.

The cloud servers of a MCC are located in the core network, which leads to high energy consumption by the mobile devices for computation file transmission. Furthermore, the latency caused by the transmission through wide area networks may seriously hamper the interactivity of the real-time mobile applications. MEC is widely considered as a promising technique to tackle these challenges. In MEC, services are hosted on the devices directly attached to radio access network [25]. The proximity of the MEC servers results in the access to cloud functionalities with low transmission energy and latency.

There are a few studies on efficient computation offloading mechanism of MEC. For instance, in [26], the authors investigated the tradeoff between offloading computation tasks to infrastructure clouds and retaining them in mobile edge clouds. In [27], the authors studied the dynamic service migration problem in mobile edge clouds, and proposed a Markov decision process based sequential offloading decision framework. The authors in [28] proposed a low complexity small cell cluster formation and load balancing scheme for the edge cloud in dense deployment 5G network. The authors in [14] formulated the task offloading problem as a joint optimization of the radio resources together with the computational resources, and proposed an iterative algorithm to solve the problem. In [29], the authors presented an analytical model for MEC cloud, and studied the performance of MEC with the presence of user mobility. The authors in [30] made a case study of the formation of a realtime context aware ad-hoc collaboration system through combining 5G and MEC technologies. In [25], the authors studied the virtual network embedding problems, and proposed network virtualization in the context of MEC networks.

Due to the sharp increase of the energy consumption and carbon emission of communication systems, the energy-efficient performance metric has been a critical goal in the design of the offloading mechanism in cloud-based networks [31]. Few studies have addressed the energy-efficient computation offloading problem. For example, the authors in [32] studied the energy-efficient offloading policies for transcoding tasks in a mobile cloud system, and proposed an online offloading algorithm with the objective to minimize the energy consumption while achieving low latency. To reduce energy cost, the authors in [33] presented a framework of the joint optimization of the radio and computation resource for the energy-limited mobile terminals in a femto cell network. Based on the estimated execution time and energy consumption in the computation offloading, the authors in [34] proposed an offloading framework, which reduces the energy consumption and shortens the response time. In MEC, as the computation tasks can be offloaded to the cloud servers located close to the current position of the mobile devices, the energy consumption for data transmission is mainly spent in the radio access network. To improve the energy efficiency in MEC, the authors in [3] studied the multi-user computation offloading problem in a multi-channel wireless network, and designed a distributed offloading algorithm through a game theoretic approach. Nevertheless, few work has taken into account the heterogeneous radio access network in the MEC, and incorporated the multi-transmission mode selection into the energy-efficient offloading schemes.

Different from these studies, in this paper, we concentrate on the computation offloading for MEC in 5G heterogeneous networks and propose the optimal offloading schemes to improve the energy efficiency of the cloud computing system while guaranteeing the delay constraints of the computation tasks.

III. SYSTEM MODEL

Fig. 1 shows the mobile devices offloading their computation tasks to the MEC server through a 5G heterogeneous network. In the system, we consider a set of mobile devices, which is denoted as $\mathcal{N} = \{1, 2, ..., N\}$. Each device has a computation task to be completed within a certain delay constraint. The tasks include interactive gaming, natural language processing, image location and etc [35]. Each computation task can be described in three terms as $T_i = \{d_i, c_i, t_i^{max}\}$, $i \in \mathcal{N}$. For task T_i, d_i is the size of the input data for the computation, which may include program codes, input files etc. c_i denotes the computing ability required for accomplishing this task, which is quantized by the number of CPU cycles. t_i^{max} is the maximum latency required by the computation task.

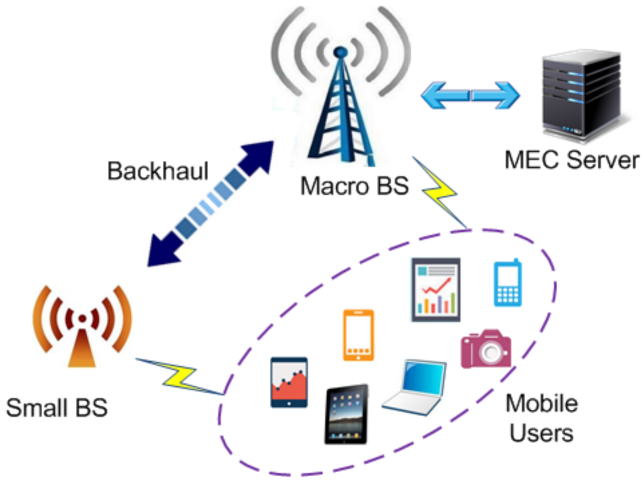


FIGURE 1. The mobile edge computing offloading in a 5G heterogeneous network.

For each mobile device i ($i \in \mathcal{N}$), the task T_i can be either executed locally on itself or on the MEC server via computation offloading. In our model, each task is atomic and can not be further divided. Let f_i^L and δ_i^L denote the local computing ability and the energy consumption for one CPU cycle of device i, respectively. Thus, we can get the time duration of the local execution of task T_i as

$$t_i^L = c_i/f_i^L. \tag{1}$$

The energy consumption of this local execution can be calculated as

$$e_i^L = c_i\delta_i^L. \tag{2}$$

In the 5G heterogeneous network, there is a Macro Base Station (MBS) equipped with an MEC server. The MEC server has the ability to run multiple computation tasks simultaneously. Besides the MBS, there is a Small Base Station (SBS), whose service area is overlaid by that of the MBS. To reuse spectrum efficiently, both the MBS and the SBS operate in the same frequency band. The spectrum is divided into K channels, which are denoted as $\mathcal{K} = \{1, 2, ..., K\}$. The bandwidth of each channel is identical, which is denoted as W. In this paper, we focus on a multi-user OFDMA system in 5G networks, where each channel in the system is orthogonal to the others.

Between the SBS and the MBS, there is a backhaul. This backhaul relays the transmission from the SBS to the MBS. We consider this backhaul is shared with other communication infrastructures. Thus, we ignore the power consumption of this backhaul. The transmission bandwidth of the backhaul is limited. The transmission delay of the backhaul is proportional to the length of the data with the scaling factor φ.

When a mobile device chooses computing its task by the MEC server, the input data can be transmitted to the MEC server through the MBS or the SBS. In the case that mobile device i accesses the MBS on channel k, the obtained uplink data transmission rate can be shown as

$$r_{i,k}^M = W\log_2(1+\frac{p_i^M g_i^M}{I_{i,k}^S+\sigma^2}), \tag{3}$$

where p_i^M is the power of mobile device i transmitting data to MBS in a unit channel. The transmission power can be determined by the MBS through some power control mechanisms [36]. g_i^M is the channel gain between mobile user i and the MBS. $I_{i,k}^S$ denotes the interference at the MBS on channel k, which is caused by the other devices' uplink transmission to the SBS on the same channel. σ^2 is the background noise power.

Similarly, we can give the uplink transmission rate of device i accessing to the SBS on the channel k as

$$r_{i,k}^S = W\log_2(1+\frac{p_i^S g_i^S}{I_{i,k}^M+\sigma^2}). \tag{4}$$

IV. PROBLEM FORMULATION

In this paper, we focus on the energy efficiency of the computation offloading network, and aim to minimize the system energy consumption under the latency constraints of the computation tasks. The energy consumption includes both the computational energy and the communication energy.

Due to the variable computing and communication abilities of these mobile devices, for each device, the energy consumption of computing the task locally may be higher or lower than when the task is executed on the MEC server. Thus, in order to reduce the energy cost, each mobile device should decide whether to offload the task in an energy-efficient manner.

Considering device i is able to offload T_i in two ways, namely through the MBS and the SBS, we denote $a_{i,j,k}$ as the task offloading decisions of the device i, where $a_{i,j,k} = \{0, 1\}$. $a_{i,j,k} = 1$ means mobile device i chooses mode j to accomplish the task T_i, and the computation data is transmitted through channel k. $a_{i,j,k} = 0$, otherwise. Here, $i \in \mathcal{N}$, and $k \in \mathcal{K}$. Let $j = \{1, 2, 3\}$ denote the chosen modes. They are computing locally, transmitting through the MBS and transmitting through the SBS, respectively. It should be noted that as there is no channel in the local computing mode, the item k is meaningless when $j = 1$. Thus we take $a_{i,1,1} = 1$ as the indicator that device i selects local computation.

For the computation task offloading to the MEC server, some extra energy and time cost are incurred by the wireless uplink transmission. In the case that device i offloads its task through the MBS to the MEC server, the total time duration can be calculated as

$$t_i^M = d_i/r_i^M + c_i/f_0^R, \tag{5}$$

where f_0^R is the computing ability of the MEC server. To concentrate our studies on the effects of the 5G heterogeneous network on the computation offloading, we consider f_0^R is a constant for each offloading task. r_i^M is the total uplink rate of the data transmitting from device i to the MBS. r_i^M can be given as

$$\begin{aligned} r_i^M &= \sum\nolimits_{k=1}^K a_{i,2,k} r_{i,k}^M \\ &= \sum_{k=1}^K a_{i,2,k} W\log_2(1+\frac{p_i^M g_i^M}{\sum\limits_{l=1,l\neq i}^N a_{l,3,k} p_l^S g_l^M+\sigma^2}). \end{aligned} \tag{6}$$

The total energy consumption in this case can be shown as

$$e_i^M = b_i^M p_i^M d_i/r_i^M + c_i\delta^R, \tag{7}$$

where δ^R is the energy cost of the MEC server for implementing a unit CPU cycle. As the MEC server always has higher computation energy efficiency than the mobile devices, we consider that $\delta^R < \delta^L$. b_i^M is the number of the channels utilized by device i for transmitting data to the MBS, and $b_i^M = \sum_{k=1}^K a_{i,2,k}$.

Similarly, the time cost for the case where device i chooses offloading the task through the SBS can be given by

$$t_i^S = d_i/r_i^S + d_i\varphi + c_i/f_0^R, \tag{8}$$

where φ is a coefficient representing backhaul transmission time delay for a unit data. The uplink transmitting rate r_i^S in (8) is shown as

$$\begin{aligned} r_i^S &= \sum\nolimits_{k=1}^K a_{i,3,k} r_{i,k}^S \\ &= \sum_{k=1}^K a_{i,3,k} W\log_2(1+\frac{p_i^S g_i^S}{\sum\limits_{l=1,l\neq i}^N a_{l,2,k} p_l^M g_l^S+\sigma^2}). \end{aligned} \tag{9}$$

The total energy cost for offloading the task through the SBS is

$$e_i^S = b_i^S p_i^S d_i/r_i^S + c_i\delta^R, \tag{10}$$

where b_i^S is number of the channels occupied by device i, and $b_i^S = \sum_{k=1}^K a_{i,3,k}$.

To minimize the total energy consumption of the system, the optimization problem is mathematically modeled as

$$\begin{aligned} \min_{\{a_{i,j,k}\}} &\sum_{i=1}^N (a_{i,1,1} e_i^L + s_{i,2}(p_i^M \frac{d_i}{r_i^M}\sum_{k=1}^K a_{i,2,k} + c_i\delta^R) \\ &+ s_{i,3}(p_i^S \frac{d_i}{r_i^S}\sum_{k=1}^K a_{i,3,k} + c_i\delta^R)) \\ s.t.\ & C1: a_{i,1,1}\cdot t_i^L \le t_i^{\max}, \quad i \in \mathcal{N} \\ & C2: r_i^M \ge \frac{d_i}{t_i^{\max} - c_i/f_0^R}, \quad i \in \mathcal{N} \\ & C3: r_i^S \ge \frac{d_i}{t_i^{\max} - c_i/f_0^R - d_i\varphi}, \quad i \in \mathcal{N} \\ & C4: \sum_{k=1}^K a_{i,1,k}\cdot\sum_{k=1}^K a_{i,2,k} = 0, \quad i \in \mathcal{N} \\ & C5: \sum_{k=1}^K a_{i,2,k}\cdot\sum_{k=1}^K a_{i,3,k} = 0, \quad i \in \mathcal{N} \end{aligned}$$

$$
\begin{aligned}
&C6: \sum_{k=1}^{K} a_{i,1,k} \cdot \sum_{k=1}^{K} a_{i,3,k} = 0, \quad i \in \mathcal{N} \\
&C7: \sum_{i=1}^{N} a_{i,j,k} = 1, \quad i \in \mathcal{N},\ j \in \{2,3\} \\
&C8: \sum_{i=1}^{N} \sum_{k=1}^{M} a_{i,j,k} \le M, \quad i \in \mathcal{N},\ j \in \{2,3\} \\
&C9: a_{i,j,k} = \{0,1\}, \quad i \in \mathcal{N},\ j \in \{1,2,3\},\ k \in \mathcal{K}.
\end{aligned} \tag{11}
$$

In (11), $s_{i,j} = 1(\sum_{k=1}^{K} a_{i,j,k} > 0)$, where $j = \{2, 3\}$. Here, $1(x)$ is an indicator function which equals 1 if x is true and 0 otherwise. Constraints C1$\sim$C3 ensure the latency requirements of the three possible implementation ways of task i, respectively. Constraints C4$\sim$C6 state that only one implementation way can be selected for each device i. According to constraint C7, one channel can only be allocated to at most one device. Constraint C8 indicates that the total channels occupied by the mobile devices are less than the maximum number of channels possessed by the MBS or the SBS.

Lemma 1: The optimization problem (11) *can be equally transformed to the problem shown as follows.*

$$
\begin{aligned}
&\min_{\{a_{i,j,k}\}} \sum_{i=1}^{N} (a_{i,1,1} e_i^{L'} + p_i^M \frac{d_i}{r_i^M} \sum_{k=1}^{K} a_{i,2,k} + p_i^S \frac{d_i}{r_i^S} \sum_{k=1}^{K} a_{i,3,k}) \\
&s.t.\ C1 \sim C9,
\end{aligned} \tag{12}
$$

where $e_i^{L'} = e_i^L - c_i \delta^R$.

Proof: The objective function of (11) can be rewritten as $\min_{\{a_{i,j,k}\}} \sum_{i=1}^{N} (a_{i,1,1} + s_{i,2} + s_{i,3}) c_i \delta^R + (a_{i,1,1} e_i^{L'} + s_{i,2} p_i^M \frac{d_i}{r_i^M} \sum_{k=1}^{K} a_{i,2,k} + s_{i,3} p_i^S \frac{d_i}{r_i^S} \sum_{k=1}^{K} a_{i,3,k})$. According to the definition of $s_{i,j}$ and constraint C7 in (11), we can get $a_{i,1,1} + s_{i,2} + s_{i,3} = 1$. Thus, $(a_{i,1,1} + s_{i,2} + s_{i,3}) c_i \delta^R$ can be omitted in the objective function, as it is a constant. Furthermore, according to constraints C4 $\sim$ C6 in (11), we can see that at a given time, for each device i, one and only one of these three items of the objective function is a positive value, and the other two are zeros. Then, we can remove $s_{i,j}$ ($j = \{2, 3\}$) from the function. □

V. ENERGY-EFFICIENT COMPUTATION OFFLOADING SCHEMES

In this section, we focus on solving the optimization problem (12). We design energy-efficient MEC cloud offloading and radio resource allocation schemes in the 5G heterogeneous networks.

In the computation offloading process, the mobile devices choose their tasks' implementation modes through the binary strategies $\{a_{i,j,k}\}$. The decisions of the $\{a_{i,j,k}\}$ not only depend on the tasks' delay constraints, but also on the transmission interference between the mobile devices and the limited resources of the radio access networks. Thus, the problem (12) can be taken as a special maximum cardinality bin packing problem, and is proved NP-hard [3], [37].

To obtain the sub-optimal solution of (12), we propose a scheme named Energy-Efficient Computation Offloading (EECO) to decide the tasks' implementation modes and allocate radio resources to the offloading devices. In order to solve (12) efficiently, the EECO scheme is divided into three stages, which are stated as follows.

- *Stage 1: Mobile device classification.* The mobile devices are classified into three types according to their time and energy cost features of the task computing process.
- *Stage 2: Priority determination.* We derive the priorities of the devices, which choose offloading their tasks to the MEC server. The priorities are used for radio resource allocation, and determined by the wireless communication states and the task requirements.
- *Stage 3: Radio resource allocation.* In this stage, the channels of MBS and SBS are allocated to the mobile devices according to the priorities determined in stage 2.

In the above stages, stage 2 and 3 operate iteratively, until a convergence criterion is satisfied. In the following of this section, we present the EECO scheme in detail.

A. MOBILE DEVICE CLASSIFICATION

Based on the tasks' latency constraints and the comparison of the energy costs between different task implementation modes, we classify the mobile devices into three types.

The first type of devices is defined as the devices, which should compute their tasks on the MEC server. We denote the set of the devices of this type as $\mathcal{G}_R$. For a device with limited computation resource, which cannot satisfy the latency constraint of the task, the device needs to choose offloading the task to the MEC server. Thus, we can get that if $t_i^L > t_i^{max}$, device i belongs to $\mathcal{G}_R$, $i \in \mathcal{N}$.

The second type is defined as the devices should compute the task on their local equipments. We denote the device set of this type as $\mathcal{G}_L$. The condition used to determine the devices belonging to this type is given as follows.

Theorem 1: If $t_i^L \le t_i^{max}$ *and* $e_i^L < \min\{p_i^M \lceil n_0^M \rceil, p_i^S \lceil n_0^S \rceil\} + c_i \delta^R$, *device i belongs to* $\mathcal{G}_L$ *(*$i \in \mathcal{N}$*), where*

$$
n_0^M = \frac{d_i}{(t_i^{\max} - c_i/f_0^R) w log_2(1 + p_i^M g_i^M/\sigma^2)}, \tag{13}
$$

and

$$
n_0^S = \frac{d_i}{(t_i^{\max} - c_i/f_0^R - d_i\varphi) w log_2(1 + p_i^S g_i^S/\sigma^2)}, \tag{14}
$$

where $\lceil \cdot \rceil$ *is the ceil function.*

Proof: If device i chooses offloading its task to the MEC server through the MBS, the highest transmission rate per channel it can obtain is $r_{i,max}^M = w log_2(1 + p_i^M g_i^M/\sigma^2)$. To satisfy the latency constraint t_i^{max}, the least number of channels needed by device i is $n_0^M = d_i/((t_i^{\max} - c_i/f_0^R) w log_2(1 + p_i^M g_i^M/\sigma^2))$. Thus, we can get the least energy for device i to offload task T_i to the MEC server via the MBS is $p_i^M \lceil n_0^M \rceil + c_i \delta^R$. Similarly, we can get the least energy

Algorithm 1 The Algorithm for Classifying the Mobile Devices

Initialization:

Mobile device set: $\mathcal{N} = \{1, 2, \cdots, N\}$;
Transmission power to MBS: $\{p_i^M\}, i \in \mathcal{N}$;
Transmission power to SBS: $\{p_i^S\}, i \in \mathcal{N}$;
Computation tasks: $T_i = \{d_i, c_i, t_i^{max}\}, i \in \mathcal{N}$;
Categorized device sets: $\mathcal{G}_L = \mathcal{G}_O = \mathcal{G}_R = \emptyset$.

```
1:  For Each device i ∈ N Do
2:     Calculate t_i^L and e_i^L according to (1) and (2), respectively;
3:     if (t_i^L > t_i^max) then
4:        i ⇒ G_R;
5:     else if (t_i^L ≤ t_i^max) then
6:        if (e_i^L < min{p_i^M ⌈n_0^M⌉, p_i^S ⌈n_0^S⌉} + c_i δ^R) then
7:           i ⇒ G_L;
8:        end if
9:     else
10:       i ⇒ G_O;
11:    end if
12: End For
```

Output: $\mathcal{G}_L$, $\mathcal{G}_R$ and $\mathcal{G}_O$.

consumed to offload task T_i to the MEC server via the SBS as $p_i^S \lceil n_0^S \rceil + c_i\delta^R$. Thus, we get the conclusion that given the condition $t_i^L \le t_i^{max}$, if the energy spent on local computing is less than the minimal value of the energy consumed for offloading, device i chooses to complete its task locally. □

Besides the two types mentioned before, the third type of the mobile devices is denoted as $\mathcal{G}_O$. The devices belong to $\mathcal{G}_O$ can either decide to implement their tasks locally or to offload the tasks to the MEC server. The decisions mainly depend on the wireless communication states.

The complete device classification process is illustrated in Algorithm 1.

B. MOBILE DEVICE PRIORITY DETERMINATION

Considering the limited capacity of the radio resources and the transmission interference between the mobile devices, we set different priorities for the devices in the radio resource allocation process.

For the devices belonging to $\mathcal{G}_R$, due to their inadequate computation capabilities, the computation tasks need to be offloaded to the MEC server. The radio resource allocation of $\mathcal{G}_R$ should have the highest priority. However, for the devices belonging to $\mathcal{G}_O$, their tasks can be either offloaded to the MEC server or the tasks can be executed locally. Thus, to reduce the energy consumption of the offloading system and to utilize the radio resource more efficiently, the devices in $\mathcal{G}_O$ should be assigned different priorities. Before introducing the priority determination algorithm, we first present two key definitions as follows.

Definition 1: The qualified channels for device i is the unoccupied channels of the MBS or the SBS, from which device i can gain Signal to Interference plus Noise Ratios (SINRs) above a given threshold θ_i.

In Definition 1, the threshold θ_i can be calculated as

$$\theta_i = 2^{\bar{r}_i W} - 1, \tag{15}$$

where $\bar{r}_i$ is device i's additional required data rate to transmit the task file within delay constraint, based on the obtained channels in the last stage iteration. We can get $\bar{r}_i = d_i/t_i^{\max} - r_{i,a}$. Here, $r_{i,a}$ is the total transmission rate on the channels which have been already allocated to device i.

Let $\mathcal{S}$ denote the wireless communication state of the current priorities determination stage. As in the EECO scheme, the stage 2 and 3 operate iteratively, $\mathcal{S}$ is updated in each iteration. Given a state $\mathcal{S}$, the number of qualified channels for device i accessing the MBS can be expressed as

$$h_i^M = \sum_{k \in \mathcal{K}'_M} 1(\gamma_{i,k}^M \ge \theta_i), \tag{16}$$

where $\mathcal{K}'_M$ is the set of the available channels of MBS. $\gamma_{i,k}^M$ is the SINR of device i transmitting in channel k of the MBS, which can be shown as $\gamma_{i,k}^M = p_i^M g_i^M/(I_{i,k}^S + \sigma^2)$.

Similarly, we can get the number of qualified channels for device i accessing the SBS as

$$h_i^S = \sum_{k \in \mathcal{K}'_S} 1(\gamma_{i,k}^S \ge \theta_i), \tag{17}$$

where $\mathcal{K}'_S$ is the set of the unoccupied channels of SBS, and $\gamma_{i,k}^S = p_i^S g_i^S/(I_{i,k}^M + \sigma^2)$.

Definition 2: The priority of device i in the process of the radio resource allocation is defined as

$$p_i = \alpha_1(\frac{t_i^{max}}{\bar{T}_{max}}) + \alpha_2(\frac{h_i}{\bar{H}}) + \alpha_3(\frac{\bar{E}}{e_i^L - c_i\delta^R}), \tag{18}$$

where $\bar{T}_{max} = \sum_{i\in\mathcal{G}_O} t_i^{max}$, $\bar{H} = \sum_{i\in\mathcal{G}_O} h_i$, and $\bar{E} = \sum_{i\in\mathcal{G}_O}(e_i^L - c_i\delta^R)$. α_j $(j = \{1, 2, 3\})$ is a coefficient, $0 \le \alpha_j \le 1$ and $\sum_{j=1}^{3}\alpha_j = 1$. The device with the less p_i value has the higher priority.

The priority definition jointly takes the delay constraint, radio resources and the offloading energy gain into consideration. In (18), the first item indicates the effect of the delay constraint on the priority. The device with more critical delay constraint should have the higher priority. The second item in (18) states the effect of the radio resource availability on the priority. The device that has less qualified channels should be allocated radio resources preferentially. Otherwise, the device may fail to transmit the task file to the MEC server within the delay constraint, due to the insufficient radio resources. The third item in (18) means that the device with higher computing energy difference between the local computing and MEC server computing should have the higher priority.

It is worth noting that as the devices of $\mathcal{G}_O$ can either choose the MBS or the SBS to transmit the task file, the h_i in (18) should be either h_i^M or h_i^S, when device i has decided to access the MBS or the SBS, respectively. However, before device i

Algorithm 2 The Algorithm for the Device Priority Determination

Initialization:
Categorized device set: $\mathcal{G}_O$;
Computation tasks: $T_i = \{d_i, c_i, t_i^{max}\}$, $i \in \mathcal{G}_O$;
Allocated channels for devices: $\mathcal{A}_i$, $i \in \mathcal{G}_O$;
MBS channel set: $\mathcal{K}_M = \{1, 2, ..., K\}$;
SBS channel set: $\mathcal{K}_S = \{1, 2, ..., K\}$;
Wireless communication state: $\mathcal{S}$;
Priority set: $\mathcal{P} = \emptyset$.

For Each device $i \in \mathcal{G}_O$ **Do**
  if $\mathcal{A}_i = \emptyset$ **then**
    Calculate h_i^M and h_i^S according to (16) and (17), respectively;
    Set $h_i = \max\{h_i^M, h_i^S\}$;
  else if $\mathcal{A}_i \in \mathcal{K}_M$ **then**
    Get $h_i = h_i^M$ according to (16);
  else if $\mathcal{A}_i \in \mathcal{K}_S$ **then**
    Get $h_i = h_i^S$ according to (17);
  end if
  Calculate p_i according to (18);
End For

Output: The priority set for the devices $\mathcal{P} = \{p_i\}$, $i \in \mathcal{G}_O$.

has decided accessing a base station, there is no restriction for its access selection. Thus, it chooses the base station with the more qualified channels, i.e., $h_i = \max\{h_i^M, h_i^S\}$.

We present this priority determination process in Algorithm 2.

C. RADIO RESOURCE ALLOCATION

In the third stage, the channels of both the MBS and the SBS are allocated to the devices based on the determined priorities. To ensure the fairness between these devices, each device can get at most one channel in a stage iteration.

In the allocation process, if a device has not decided to access any base station, the device can make the decision based on the comparison of the energy savings between offloading the task through the MBS and the SBS.

For the device that has decided access the MBS, it can only choose a channel from the ones belonging to the MBS. Given device i accesses the MBS, as the transmitting power p_i^M is identical for each channel of the MBS, device i should choose the channel that has the highest SINR. The reason is that higher SINR leads to shorter transmission time and less transmission energy consumption. The new selected channel improves the total transmission rate of device i with the increase of the transmission power on this channel. However, the improving transmission rate reduces the transmission time, which may decrease the total transmission energy. Thus, we should compare the energy cost between the case with the new selected channel and the case without it. If the new selected channel brings higher energy cost, it should not be allocated to device i.

Algorithm 3 The Algorithm of Resource Allocation

Initialization:
Categorized device set: $\mathcal{G}_O$;
Computation tasks: $T_i = \{d_i, c_i, t_i^{max}\}$, $i \in \mathcal{G}_O$;
Allocated channels for devices: $\{\mathcal{A}_i\}$, $i \in \mathcal{G}_O$;
MBS channel set: $\mathcal{K}_M = \{1, 2, ..., K\}$;
SBS channel set: $\mathcal{K}_S = \{1, 2, ..., K\}$;
MBS's unoccupied channel set: $\mathcal{K}'_M$;
SBS's unoccupied channel set: $\mathcal{K}'_S$;
Wireless communication state: $\mathcal{S}$.

Set the temporary set $\mathcal{G}'_O = \mathcal{G}_O$;
while $\mathcal{G}'_O \neq \emptyset$ **Do**
  Select the device i, where $i = \arg\min_i\{p_i\}$, $i \in \mathcal{G}_O$;
  if $\mathcal{A}_i = \emptyset$ **then**
    Under the given state $\mathcal{S}$, choose the channels with the highest SINR from $\mathcal{K}'_M$ and $\mathcal{K}'_S$. The selected channels are denoted as k_0^M and k_0^S, respectively.;
    Compute the transmission rate obtained from k_0^M and k_0^S. The rates are denoted as $r_{i,0}^M$ and $r_{i,0}^S$, respectively.;
    if $p_i^M/r_{i,0}^M \geq p_i^S/r_{i,0}^S$ && $e_i^M < e_i^L$ **then**
      $k_0^M \Rightarrow \mathcal{A}_i$ and $\mathcal{K}'_M = \mathcal{K}'_M \backslash k_0^M$;
    else if $p_i^M/r_{i,0}^M < p_i^S/r_{i,0}^S$ && $e_i^S < e_i^L$ **then**
      $k_0^S \Rightarrow \mathcal{A}_i$ and $\mathcal{K}'_S = \mathcal{K}'_S \backslash k_0^S$;
    end if
  else if $\mathcal{A}_i \in \mathcal{K}_M$ **then**
    Choose the channel k with the highest SNIR in $\mathcal{K}'_M$;
    Compute the energy costs e_i^M with the channels $\mathcal{A}_i$ and $e_i^{M'}$ with the channels $\{\mathcal{A}_i \cup k\}$;
    if $e_i^{M'} < e_i^M < e_i^L$ **then**
      $k \Rightarrow \mathcal{A}_i$ and $\mathcal{K}'_M = \mathcal{K}'_M \backslash k$;
    end if
  else if $\mathcal{A}_i \in \mathcal{K}_S$ **then**
    Choose the channel k with the highest SNIR in $\mathcal{K}'_S$;
    Compute the energy costs e_i^S with the channels $\mathcal{A}_i$ and $e_i^{S'}$ with the channels $\{\mathcal{A}_i \cup k\}$;
    if $e_i^{S'} < e_i^S < e_i^L$ **then**
      $k \Rightarrow \mathcal{A}_i$ and $\mathcal{K}'_S = \mathcal{K}'_S \backslash k$;
    end if
  end if
  $\mathcal{G}'_O = \mathcal{G}'_O \backslash i$;
end while

Output: The updated wireless communication state $\mathcal{S}$.

For the devices that have chosen to access the SBS, the channel allocation is in a similar way. We show the complete radio resource allocation process in Algorithm 3.

D. MAIN PROCESS OF EECO

The detail of the proposed EECO scheme is illustrated in Algorithm 4. At first, we classify the mobile devices into three types. For the devices belonging to $\mathcal{G}_R$, as they cannot accomplish the tasks on their own devices under the latency

Algorithm 4 The Main Process of the Energy-Efficient Computing Offloading Scheme

Initialization:

Mobile device set: $\mathcal{N} = \{1, 2, \cdots, N\}$;
Computation tasks: $T_i = \{d_i, c_i, t_i^{max}\}, i \in \mathcal{N}$;
Allocated channels for devices: $\{\mathcal{A}_i\} = \emptyset, i \in \mathcal{N}$;
MBS channel set: $\mathcal{K}_M = \{1, 2, \cdots, K\}$;
SBS channel set: $\mathcal{K}_S = \{1, 2, \cdots, K\}$;
MBS's unoccupied channel set: $\mathcal{K}'_M = \mathcal{K}_M$;
SBS's unoccupied channel set: $\mathcal{K}'_S = \mathcal{K}_S$;
Wireless communication state: $\mathcal{S}$;
Final determined device sets: $\mathcal{G}_L^{final} = \mathcal{G}_{R'}^{final} = \emptyset$.

1: Clarify the mobile devices into sets $\mathcal{G}_L$, $\mathcal{G}_R$ and $\mathcal{G}_O$ according to Algorithm 1;
2: Allocate radio resources to the device set $\mathcal{G}_R$, and offload their tasks to the MEC server;
3: Update $\mathcal{K}'_M$ and $\mathcal{K}'_S$;
4: **while** $\mathcal{G}_O \neq \emptyset \,||\, \{\mathcal{K}'_M \neq \emptyset \,\&\&\, \mathcal{K}'_S \neq \emptyset\}$ **Do**
5: Get priority set $\mathcal{P}$ according to Algorithm 2;
6: Based on priority set $\mathcal{P}$, allocate radio resource to the device set $\mathcal{G}_O$ according to Algorithm 3, and get the allocated channels $\mathcal{A}'_i$ for device i, $i \in \mathcal{G}_O$;
7: **For** Each device $i \in \mathcal{G}_O$ **Do**
8: **if** $\mathcal{A}'_i \neq \mathcal{A}_i \,\&\&\, t_i^M \leq t_i^{max} (t_i^S \leq t_i^{max})$ **then**
9: $\mathcal{A}_i = \mathcal{A}'_i$;
10: **else if** $\mathcal{A}'_i == \mathcal{A}_i \,\&\&\, t_i^M \leq t_i^{max} (t_i^S \leq t_i^{max})$ **then**
11: $\mathcal{G}_O = \mathcal{G}_O \backslash i$ and $i \Rightarrow \mathcal{G}_{R'}^{final}$;
12: **else if** $\mathcal{A}'_i == \mathcal{A}_i \,\&\&\, t_i^M > t_i^{max} (t_i^S > t_i^{max})$ **then**
13: $\mathcal{G}_O = \mathcal{G}_O \backslash i$, $i \Rightarrow \mathcal{G}_L^{final}$, and put the channels in $\mathcal{A}_i$ to the unoccupied channel set;
14: **end if**
15: Update the channel sets $\mathcal{K}'_M$ and $\mathcal{K}'_S$;
16: **End For**
17: **end while**
18: For device set $\mathcal{G}_L^{final}$, implement their computing tasks locally;
19: For device set $\mathcal{G}_{R'}^{final}$, offload their computing tasks to the MEC server.

constraint, we allocate the channels to the devices with the highest priority. Then, we update the channel allocation state, and the scheme goes into the iterative process. In each iteration, the devices of $\mathcal{G}_O$ are assigned priorities according to Algorithm 2 and allocated channels according to Algorithm 3. At the end of each iteration, each device of $\mathcal{G}_O$ is checked in terms of its channel resource assignment. In every iteration of EECO, each device chooses the channel that brings the highest reduction in the energy consumption. In a given iteration, if a device cannot get a channel that decreases its energy cost, it is less likely that the device will obtain a suitable channel in the next iteration. Thus, we remove the devices from $\mathcal{G}_O$ whose channel assignments are identical in two consecutive iterations. The removed devices are classified into $\mathcal{G}_L$ or $\mathcal{G}_R$ according to the offloading time delay based on their assigned channel resources. The iteration process is terminated when there is no device in $\mathcal{G}_O$ or no available radio resource.

Theorem 2: The computation of EECO scheme has a polynomial complexity.

Proof: Firstly, the mobile device classification in Line 1 of Algorithm 4 has N iterations by categorizing each device in mobile device set $\mathcal{N}$. Secondly, the priority determination in Line 5 has $|\mathcal{G}_O|$ iterations for assigning priorities for devices in $\mathcal{G}_O$. After that, radio resource allocation is applied in Line 6 with $|\mathcal{G}_O|$ iterations in total. The number of iterations of the while-loop is limited by the process of removing devices in $\mathcal{G}_O$ and the eliminating channels in $\mathcal{K}'_M$ and $\mathcal{K}'_S$. The analysis of the iterative process is given as follows. On one hand, a channel is allocated to a device and removed from $\mathcal{K}'_M$ or $\mathcal{K}'_S$ in Line 8, 10, 16 and 22 of Algorithm 3, taking at most $k+K$ iterations of the while-loop in Algorithm 4 in terms of removing channels. On the other hand, each device is removed in Algorithm 4 Line 11 and 13, giving $|\mathcal{G}_O|$ iterations of the while-loop in terms of deleting devices. To sum, the iteration of the while-loop is $(|\mathcal{G}_O| + |\mathcal{G}_O| + |\mathcal{G}_O|)max(|\mathcal{G}_O|, K + K)$. Then, the computational complexity of EECO can be given

$$\begin{aligned} &O(N + (|\mathcal{G}_O| + |\mathcal{G}_O| + |\mathcal{G}_O|)max(|\mathcal{G}_O|, K + K)) \\ &\quad = O(max(|\mathcal{G}_O|^2 + N, |\mathcal{G}_O|K + N)). \end{aligned} \tag{19}$$

□

VI. NUMERICAL RESULTS

In this section, we evaluate the performance of our proposed EECO scheme through the extensive simulations. We consider there are an MBS and an SBS that cover a 1000m×1000m area. Each base station has $K = 50$ channels and the channels belonging to one base station are orthogonal. There is an MEC server located in the MBS, whose computation capability is 4GHz/sec. The energy consumption for the MEC is given as δ^R=1W/GHz [38]. The backhaul time delay coefficient φ is set as 0.0001 sec/KB. $N = 50$ mobile devices are randomly scattered over the area. The number of CPU cycles required by the computation tasks of the devices are randomly distributed between 0.1 and 1 GHz. The corresponding computation file size is randomly distributed between 300 and 800KB. The mobile devices' latency requirements are randomly distributed between 0.5 and 1 second.

Fig. 2 indicates the energy consumption performance of the system with three different task implementation schemes in terms of the number of the mobile devices. The energy consumption includes both the energy cost spent on computation and file transmission. In this figure, the energy consumption of all three schemes increase as the number of devices grows. The energy consumption of the scheme without offloading is higher than that of the other two schemes in all cases, especially with large number of devices. It is worth noting that some tasks' delay constraints may not be satisfied when all the tasks are implemented locally on the devices. Here we use the performance of this scheme as a benchmark to

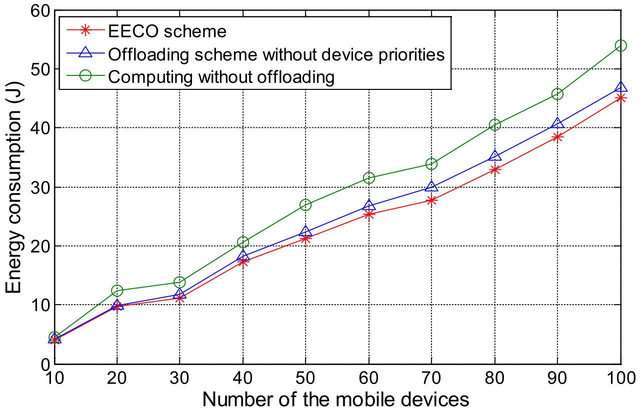


FIGURE 2. The energy consumption of the offloading system with different schemes.

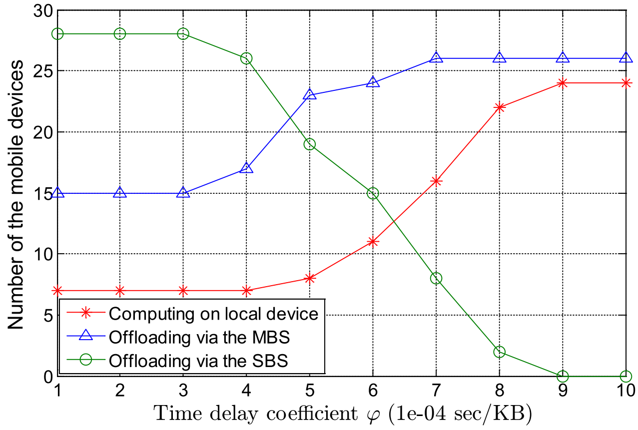


FIGURE 3. The number of the device sets $\mathcal{G}_L$, $\mathcal{G}_O$ and $\mathcal{G}_R$ with different backhaul time delay coefficient.

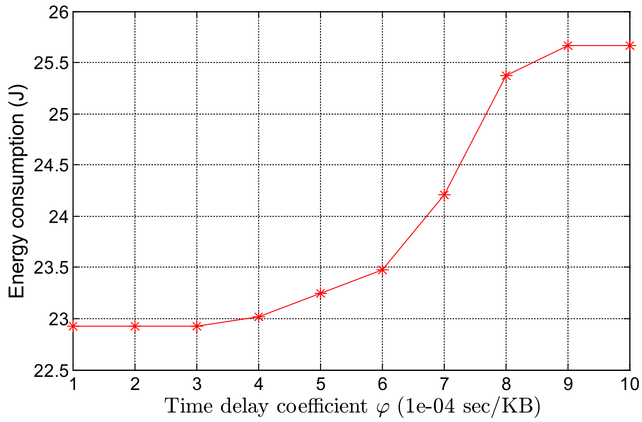


FIGURE 4. The energy consumption of the offloading system with different backhaul time delay coefficient.

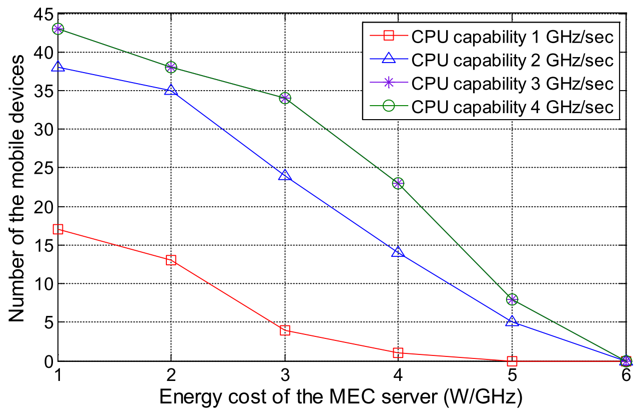


FIGURE 5. The number of the offloading devices with different MEC server CPU capabilities.

measure the energy cost reduction gained by the other two schemes.

Compared to the scheme without offloading, Fig. 2 shows that the average rate of the decrease in energy consumption in our proposed EECO scheme is 18%. From this figure, we can see that the energy consumption difference between the scheme without offloading and the schemes with offloading doesn't continue to increase with the growth of the number of the mobile devices. This is because that the energy consumption for file transmission may vary with the increase of the device number, and there is no linear relationship between them. Furthermore, we can see from the figure that the energy consumption of EECO and that of the offloading scheme without device priorities are almost the same when the number of device is below 30. However, the difference becomes visible as the number increases. The reason is that, when there are few devices in the network, the radio resource is sufficient for the offloading devices. With the increase in the number of device, the channel contention occurs. The offloading scheme without device priorities allocates the channels to the offloading devices randomly. On the contrary, in EECO scheme, the channels are preferentially allocated to the devices, which may result in higher energy cost reduction.

Fig. 3 shows the number of the mobile devices of the three types with different time delay coefficient φ. In Fig. 3, when $\varphi \leq 0.0003$ sec/KB, there is no change in the number of devices. This is because that the low transmission delay of the backhaul may not cause the total time cost of the task accomplishment breaks the delay constraints of the tasks. With the increase of φ, higher transmission delay is posed on the backhaul. Thus, the offloading devices that can not use the SBS for transmitting files under the delay constraint choose to transmit through the MBS. However, due to the radio resource limitation of the MBS, the number of devices offloading the tasks via MBS reaches the maximum value when $\varphi \geq 0.0007$. The left devices that can neither offload tasks via MBS nor via SBS should implement their tasks locally.

Fig. 4 indicates the energy consumption of the system with different time delay coefficient φ. The energy consumption of the system increases with higher φ. This can be explained as follows. First, when φ increases, the devices located near the SBS have to transmit through the MBS due to the time delay constraint. For these devices, transmitting via the MBS causes higher transmission energy cost compared to transmitting through the SBS. Second, the number of devices that implement tasks locally increases. Considering that the computation energy efficiency of the MEC server is higher than that of the devices, some extra energy should be consumed.

Fig. 5 compares the impacts of the MEC server CPU capability on the number of the offloading devices with different MEC server energy cost. The number of devices that choose to offload their tasks to the MEC server decreases with the increase in the server computation energy cost. When energy cost reaches 6 W/GHz, since more computation energy is consumed for the task implemented on the MEC server than on the local devices, no device offloads its task to the MEC server. As the CPU capability gets lower, the number of offloading devices reduces. It is worth noting that the offloading device numbers with CPU capabilities of 4 GHz/sec and 3 GHz/sec are the same in this figure. The reason is that both these CPU capabilities can satisfy the delay constraints of the offloaded tasks. To this end, we can draw a conclusion that the energy cost of the MEC server directly affects the offloading devices. However, the offloading choices of the devices can only be influenced when the MEC server CPU capability is below a certain threshold.

VII. CONCLUSION

In this paper, we investigated the MEC offloading mechanisms in 5G heterogeneous networks. In order to improve the energy efficiency of the offloading system, we formulated a problem to minimize the energy consumption of the computation task implementation together with that of the communication process. To solve the problem more efficiently, we proposed an EECO scheme, which jointly optimizes the computation offloading decisions and the radio resource allocation strategies to minimize the system energy cost under the delay constraints. In addition, we conducted a simulation study, which clearly displays the energy efficiency enhancement in our proposed EECO scheme.

REFERENCES

[1] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 9, pp. 2546–2559, Sep. 2016.

[2] B. Fan, S. Leng, and K. Yang, "A dynamic bandwidth allocation algorithm in mobile networks with big data of users and networks," *IEEE Netw.*, vol. 30, no. 1, pp. 6–10, Jan./Feb. 2016.

[3] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. PP, no. 99, Oct. 2015.

[4] J. Shen, H. Tan, J. Wang, J. Wang, and S. Lee, "A novel routing protocol providing good transmission reliability in underwater sensor networks," *J. Internet Technol.*, vol. 16, no. 1, pp. 171–178, Jan. 2015.

[5] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.

[6] E. Meskar, T. D. Todd, D. Zhao, and G. Karakostas, "Energy efficient offloading for competing users on a shared communication channel," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 3192–3197.

[7] B. Gu and V. S. Sheng, "A Robust regularization path algorithm for υ-support vector classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, no. 99, Feb. 2016.

[8] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, "Mobile-edge computing," Mobile-Edge Comput.-Introductory, Tech. Rep., Sep. 2014.

[9] K. Yang, Q. Yu, S. Leng, B. Fan, and F. Wu, "Data and energy integrated communication networks for wireless big data," *IEEE Access*, vol. 4, pp. 713–723, 2016.

[10] K. M. S. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, and R. L. Aguiar, "Green HetNet CoMP: Energy efficiency analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4670–4683, Oct. 2015.

[11] P. Guo, J. Wang, X. H. Geng, C. S. Kim, and J.-U. Kim, "A variable threshold-value authentication architecture for wireless mesh networks," *J. Internet Technol.*, vol. 15, no. 6, pp. 929–936, Nov. 2014.

[12] C. Shao, S. Leng, Y. Zhang, A. Vinel, and M. Jonsson, "Performance analysis of connectivity probability and connectivity-aware MAC protocol design for platoon-based VANETs," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5596–5609, Dec. 2015.

[13] Q. Wang, S. Leng, H. Fu, and Y. Zhang, "An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 449–458, Jun. 2012.

[14] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.

[15] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 340–352, Feb. 2016.

[16] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.

[17] Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual verifiable provable data auditing in public cloud storage," *J. Internet Technol.*, vol. 16, no. 2, pp. 317–323, Mar. 2015.

[18] Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Trans. Commun.*, vol. E98-B, no. 1, pp. 190–200, Jan. 2015.

[19] A. ur Rehman Khan, M. Othman, F. Xia, and A. N. Khan, "Context-aware mobile cloud computing and its challenges," *IEEE Cloud Comput.*, vol. 2, no. 3, pp. 42–49, May/Jun. 2015.

[20] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 974–983, Apr. 2015.

[21] Y. Zhang, D. Niyato, and P. Wang, "Offloading in mobile cloudlet systems with intermittent connectivity," *IEEE Trans. Mobile Comput.*, vol. 14, no. 12, pp. 2516–2529, Dec. 2015.

[22] L. Yang, J. Cao, H. Cheng, and Y. Ji, "Multi-user computation partitioning for latency sensitive mobile cloud applications," *IEEE Trans. Comput.*, vol. 64, no. 8, pp. 2253–2266, Aug. 2015.

[23] H. Viswanathan, E. K. Lee, I. Rodero, and D. Pompili, "Uncertainty-aware autonomic resource provisioning for mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 8, pp. 2363–2372, Aug. 2015.

[24] S. Abolfazli, Z. Sanaei, M. Alizadeh, A. Gani, and F. Xia, "An experimental analysis on cloud-based mobile augmentation in mobile cloud computing," *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 146–154, Feb. 2014.

[25] M. T. Beck and M. Maier, "Mobile edge computing: Challenges for future virtual network embedding algorithms," in *Proc. Adv. Eng. Comput. Appl. Sci.*, Aug. 2014, pp. 65–70.

[26] U. Drolia *et al.*, "The case for mobile edge-clouds," in *Proc. IEEE 10th Int. Conf. Ubiquitous Intell. Comput., 10th Int. Conf. Auto. Trusted Comput. (UIC/ATC)*, Dec. 2013, pp. 209–215.

[27] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. Chan, and K. K. Leung, "Dynamic service migration in mobile edge-clouds," in *Proc. IFIP Netw. Conf.*, May 2015, pp. 1–9.

[28] J. Oueis, E. C. Strinati, and S. Barbarossa, "The fog balancing: Load distribution for small cell cloud computing," in *Proc. Veh. Technol. Conf.*, May 2015, pp. 1–6.

[29] T. Taleb and A. Ksentini, "An analytical model for follow me cloud," in *Proc. IEEE Global Commun. Conf.*, Dec. 2013, pp. 1291–1296.

[30] S. Nunna *et al.*, "Enabling real-time context-aware collaboration through 5G and mobile edge computing," in *Proc. 12th Int. Conf. Inf. Technol.-New Generat.*, Apr. 2015, pp. 601–605.

[31] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5275–5287, Nov. 2015.

[32] W. Zhang, Y. Wen, and H.-H. Chen, "Toward transcoding as a service: Energy-efficient offloading policy for green mobile cloud," *IEEE Netw.*, vol. 28, no. 6, pp. 67–73, Nov./Dec. 2014.

[33] O. Muñoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4738–4755, Oct. 2015.

[34] Y.-D. Lin, E. T.-H. Chu, Y.-C. Lai, and T.-J. Huang, "Time-and-energy-aware computation offloading in handheld devices to coprocessors and clouds," *IEEE Syst. J.*, vol. 9, no. 2, pp. 393–405, Jun. 2015.

[35] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 507–518, Mar. 2015.

[36] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, no. 2, pp. 210–221, Apr. 2003.

[37] K.-H. Loh, B. Golden, and E. Wasil, "Solving the maximum cardinality bin packing problem with a weight annealing-based algorithm," in *Operations Research and Cyber-Infrastructure*, vol. 47. USA: Springer, 2009, pp. 147–164.

[38] M. Wittmann, G. Hager, T. Zeiser, J. Treibig, and G. Wellein, "Chip-level and multi-node analysis of energy-optimized lattice Boltzmann CFD simulations," *Concurrency Comput., Pract. Exper.*, vol. 28, pp. 2295–2315, May 2016.

KE ZHANG is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China. He is a Lecturer with the University of Electronic Science and Technology of China. He is currently a Visiting Researcher with the Simula Research Laboratory, Fornebu, Norway. His research interests include energy management and communications in the smart grid, design and optimization of next-generation wireless networks, and Internet of Things.

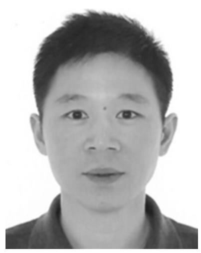
YUMING MAO is a Professor with the University of Electronic Science and Technology of China. He is the Chairman of the Department of Network Engineering. His main research area includes broadband communication network, network organization and protocol analysis, TCP/IP technology, network management and protocol, routing protocol, and network engineering. He was recipient of several awards, including the first grade, second grade, and third grade awards of the Ministry of Electronic Industry for science and technology progress, the second grade National Award for science and technology progress.

SUPENG LENG (M'06) received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He is a Professor with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China. He has been a Research Fellow with the Network Technology Research Center, NTU. He has published over 100 research papers. His research interests include resource, spectrum, energy, routing, and networking in broadband wireless access networks, vehicular networks, Internet of Things, next-generation mobile networks, and smart grids. He serves as an Organizing Committee Chair and a Technical Program Committee Member for many international conferences, and a Reviewer for over ten international research journals.

QUANXIN ZHAO (S'–) is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China. His research interests include resource, spectrum, energy, and networking in broadband wireless access networks, next generation mobile networks, and software defined wireless networks.

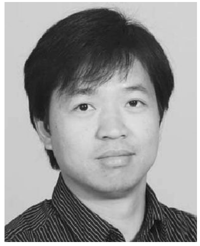
LONGJIANG LI (M'12) received the B.S. degree in computer science and engineering from Xidian University, Xi'an, in 2001, the M.S. degree in computer science and engineering from the Xi'an Telecommunication Institute, in 1998, and the Ph.D. degree in computer science and technology from Shanghai Jiao Tong University, China, in 2007. He is an Associate Professor with the Department of Communication and Information Engineering, University of Electronic Science and Technology of China. His research interest is in the domain of communication and information engineering, especially with regard to mobile ad hoc networks (including VANETs/WSNs), unmanned aerial vehicles, and distributed virtual reality.

XIN PENG received the bachelor's degree in communication engineering, the master's degree in communication and information system, and the Ph.D. degree in computer science from Hunan University, Changsha, China, in 2003, 2008, and 2011, respectively. He is an Associate Professor with the College of Information and Communication Engineering, Hunan Institute of Science and Technology. His research interests include wireless sensor networks and cyber physical systems.



LI PAN received the M.S. degree in computer software and theory from Sun Yat-Sen University, Guangzhou, China, in 2004, and the Ph.D. degree in computer applied technology from Tongji University, Shanghai, China, in 2009. He is currently an Associate Professor with the Department of Information and Communication Engineering, Hunan Institute of Science and Technology, China, and also with the Key Laboratory of Optimization and Control for Complex Systems, College of Hunan Province, Yueyang, China. He has authored over 20 papers. His current research interests include Petri nets, formal method, and computational intelligence.

SABITA MAHARJAN (S'09–M'13) received the M.Eng. degree in wireless communication from the Antenna and Propagation Laboratory, Tokyo Institute of Technology, Tokyo, Japan, in 2008, and the Ph.D. degree in network and distributed systems from the University of Oslo, Oslo, Norway, and the Simula Research Laboratory, Fornebu, Norway, in 2013. She is currently a Post-Doctoral Fellow with the Simula Research Laboratory. Her current research interests include wireless networks, network optimization, security, game theory, smart grid communications, and cyber physical systems.



YAN ZHANG (M'05–SM'10) received the Ph.D. degree from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. He is currently the Head of the Department with the Department of Networks, Simula Research Laboratory, Fornebu, Norway, and an Adjunct Associate Professor with the Department of Informatics, University of Oslo, Oslo, Norway. His current research interests include wireless networks and reliable and secure cyber physical systems (e.g., healthcare, transport, and smart grids). He is an Associate Editor being on the Editorial Boards of a number of well-established scientific international journals, e.g., *Wiley Wireless Communications and Mobile Computing*. He also serves as a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the *IEEE Communications Magazine*, IEEE WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING. He serves as a Chair or a Technical Program Committee Member of numerous international conferences. He has received seven Best Paper Awards. He is a Senior Member of the IEEE Communications and Vehicular Technology Societies, and a fellow of the Institution of Engineering and Technology.