Energy-Efficient Resource Allocation for Mobile Edge Computing System Supporting Multiple Mobile Devices

Song Jin[1], Qi Gu[2], Xiang Li[3], Xuming An[3], and Rongfei Fan[3](✉)

[1] Institute of Telecommunication Satellite, China Aerospace Science and Technology Corporation, Beijing 100094, China
390923423@qq.com

[2] Beijing Jiaotong University, Beijing 100044, China
15112094@bjtu.edu.cn

[3] Beijing Institute of Technology, Beijing 100081, China
{lawrence,fanrongfei}@bit.edu.cn, 1952590139@qq.com

Abstract. Nowadays, mobile edge computing (MEC) has become a promising technique to provide mobile devices with intensive computation capability for the applications in the Internet of Things and 5G communications. In a MEC system, a mobile device, who has computation tasks to complete, would like to offload part or all the data for computation to a MEC server, due to the limit of local computation capability. In this paper, we consider a MEC system with one MEC server and multiple mobile devices, who access into the MEC server via frequency division multiple access (FDMA). The energy consumption of all the mobile devices is targeted to minimized via optimizing the computation and communication resources, including the amount of data for offloading, the bandwidth for accessing, the energy budget for offloading data, the time budget for offloading, for each mobile device. An optimization problem is formulated, which is non-convex. We decompose it into two levels. In the lower level, a convex optimization problems is formulated. In the upper level, a one-dimensional variable is to be optimized by bisection search method.

Keywords: Edge computing · Data offloading · Multiple users · FDMA

1 Introduction

In the past decade, as smart mobile terminals is obtaining tremendous popularity, mobile users look forward to running real-time applications on their device anywhere [1]. These applications, such as augmented reality (AR), online games and face recognition, has a stringent requirement on latency and computation [2]. However, limited computation capability and battery capacity make it difficult for the mobile devices to fulfill the users' task within the rigid time constraint

B. Li et al. (Eds.): IoTaaS 2019, LNICST 316, pp. 228–234, 2020.
https://doi.org/10.1007/978-3-030-44751-9_19

alone. To solve these problems, mobile-edge computing (MEC) is emerging as a promising technique to provide cloud computing service at the mobile edge network [3]. Rather than transferring all the data for computing to the remote cloud servers, mobile users only need to offload data to the near small-cell base station with MEC server, which avoids the long propagation delay between the mobile device to the remote cloud center and insures the performance for those real-time mobile applications under stringent latency constraint [4].

In a MEC system, the computation task for each mobile users can be offloaded in two ways of scheme, binary and partial offloading. For the unpartitionable computation task, the binary offloading case should be implemented and the whole task should be either offloaded or computed locally entirely. In the partial offloading case, the partitionable task can be divided into two parts and only one of them is offloaded. In the meantime, offloading data to the base station will bring additional communication overhead. Since the battery capacity of mobile devices is limited, it is necessary to balance the tradeoff between the energy cost of local computing and offloading so as to minimize the mobile device's total energy consumption.

There are many works focusing on the energy minimizing problem of offloading scheme in a MEC system. The energy consumption minimization and latency of application execution minimization problem are studied separately under a single-user MEC scenario in [4]. Local CPU frequency, offloading power and data amount are optimized to minimize the energy consumption of the only mobile user. In [5], resource allocation for a multi-user mobile edge computation offloading system based on orthogonal frequency-division multiple access (OFDMA) is studied. Offloading data amount and the allocation of sub-channel are jointly optimized to minimize the sum mobile energy consumption of all mobile users. The problem is a mixed-integer programming problem and a sub-optimal algorithm is obtained. The authors of [6] investigate an multi-user offloading strategy to minimize the total energy consumption under the latency constraints for each user's task. The optimization variable are offloading time and data amount and optimal solution is found by applying KKT conditions. In [7], the authors consider a multi-mobile-users MEC system with unpartitionable computation tasks. To minimize the energy consumption on mobile users, the offloading selection, radio resource allocation, and CPU frequency allocation of the MEC server are optimized coordinately. However, the offloading strategy [7] is binary and is not fit for the partitionable task. The scheme in [4] is hard to apply to the multi-user scenario. [5] doesn't consider the latency constraints for each user's task. Meanwhile, [5,6] and [7] only consider the scenario with a fixed offloading bandwidth for each mobile user and ignore its influence on minimizing the total energy consumption, which leads to a suboptimal offloading scheme.

In this letter, we focus on jointly optimizing communication and computation resources for partial offloading in a MEC system with multiple mobile devices based on FDMA. A multiple mobile users scenario covered by a base station with MEC server is introduced with multiple parallel computation tasks requiring cloud computation resource. Each mobile user can access to the base station through wireless channels with different frequency bands simultaneously. The offloading time, bandwidth and data amount for each mobile user are optimized jointly to minimize overall users energy consumption under the latency constraints for different users' applications. A non-convex optimization problem is formulated. With some transformation, we are able to find the local optimal solution for the proposed optimization problem.

2 System Model and Problem Formulation

In this paper, we consider such a scenario. There are multiple mobile devices and one base station. A MEC server is mounted on the base station. The set of mobile devices are indexed as $\mathcal{K} = \{1, 2, ..., K\}$. Each mobile device, say device k, has a computation task to complete, say $\mathbb{T}_k$, for $k \in \mathcal{K}$. Task k can be described by a three-tuple vector, $\{D_k, C_k, T_k\}$, where D_k denotes the data size for computing (in unit of nat), C_k denotes the required CPU cycles for computing one data nat, and T_k indicates the maximal tolerable computation delay. In this paper, we consider the case that all the T_k are identical, i.e., $T_k = T$ for $k \in \mathcal{K}$. Every task is separable, i.e., every task can be divided into two parts: one part can be computed at the mobile device locally, and the other part can be offloaded to the base station and computed by the MEC server at the base station. For task k, suppose the amount of data nat for local computing is $d_{l,k}$ and the amount of data nat for offloading is $d_{o,k}$. Then there is

$$d_{o,k} + d_{l,k} = D_k, \forall k \in \mathcal{K}. \tag{1}$$

For local computing at mobile device k, the number of CPU cycles required is $C_k d_{l,k}$. The local computing can be completed within a time duration of T_k. Thus the associated energy consumption $E_{l,k}$ can be written as

$$E_{l,k} = c_0 f_k^3 T = \frac{c_0 C_k^3 d_{l,k}^3}{T^2} \tag{2}$$

where c_0 is a coefficient depending on mobile device's CPU structure, which is assumed to be identical over all the mobile devices without loss of generality, and f_k is the computation capability of kth mobile device, which is usually measured by the number of CPU cycles it can process in unit time.

For mobile device k's computing at the base station, the whole available time duration T is divided into three parts. The first part is for data offloading from kth mobile device to the base station, which occupies a time duration of t_o. The second part is for data processing at the base station, which occupies a time duration of t_p. It should be noticed that the data offloading time t_o and data processing time t_p are identical for every mobile user. This setup is for the ease of coordinating among base station and multiple mobile devices. The third part is for data feedback from the base station to every mobile device. Since the amount of data of computational result is generally negligible, we omit the time consumption of this part. Thus there is

$$t_o + t_p = T \tag{3}$$

For data offloading, suppose the normalized channel gain from kth mobile device to the base station is h_k, which is the ratio between the real channel gain and noise spectrum density. h_k for $k \in \mathcal{K}$ are assumed to keep fixed within time T, which is reasonable in slow fading scenario. These K mobile devices access into the base station via FDMA. Denote the bandwidth allocated to the uplink of kth mobile device as w_k, and the system bandwidth as W_T. Then there is

$$\sum_{k=1}^{K} w_k \leq W_T. \tag{4}$$

It should be noticed that the channel gain h_k for $k \in \mathcal{K}$ keeps unchanged within the system bandwidth W_T, which is reasonable in flat-fading scenario. Define the kth mobile device's available energy for offloading data is $E_{o,k}$. Hence the transmit power can be expressed $p_k = \frac{E_{o,k}}{t_o}$. According to Shannon theory, the channel capacity of mobile device k's uplink can be written as

$$\begin{aligned} R_k &= w_k \ln\left(1 + \frac{p_k h_k}{w_k}\right) \\ &= w_k \ln\left(1 + \frac{E_{o,k} h_k}{w_k t_o}\right) \end{aligned} \tag{5}$$

Then the offloaded data $d_{o,k}$ satisfies the following formula

$$\begin{aligned} d_{o,k} &\leq R_k t_o \\ &= w_k t_o \ln\left(1 + \frac{E_{o,k} h_k}{w_k t_o}\right) \end{aligned} \tag{6}$$

For data processing at the base station, the total amount of data processing can be written as $\sum_{k=1}^{K} C_k d_{o,k}$. Suppose the computation capability of base station is f_E. To process offloaded data of all the K mobile devices within time t_p, the following constraint should be satisfied

$$\sum_{k=1}^{K} C_k d_{o,k} \leq t_p f_E. \tag{7}$$

In such a system, the system energy consumption of all the K mobile devices can be written as

$$\begin{aligned} &\sum_{k=1}^{K} \left(E_{o,k} + E_{l,k}\right) \\ = &\sum_{k=1}^{K} \left(E_{o,k} + \frac{c_0 C_k^3 d_{l,k}^3}{T^2}\right) \end{aligned} \tag{8}$$

Our target is minimize the total energy consumption via optimizing t_o, t_p, $E_{o,k}$, $d_{o,k}$, $d_{l,k}$ and w_k for $k \in \mathcal{K}$. By collecting the formulated constraints in (1), (3), (4), (6), (7), and imposing non-negative constraints on all the variables to be optimized, we need to solve the following optimization problem.

Problem 1.

$$\begin{aligned} \min_{\substack{t_o,t_p,\{w_k\},\{E_{o,k}\},\\ \{d_{o,k}\},\{d_{l,k}\}}} \quad & \sum_{k=1}^{K}\left(E_{o,k}+\frac{c_0C_k^3d_{l,k}^3}{T^2}\right) \\ s.t. \quad & \textit{Constraints } (1),(3),(4),(6),(7), \\ & t_o \geq 0, t_p \geq 0, \forall k \in \mathcal{K}, \\ & w_k \geq 0, \forall k \in \mathcal{K}, \\ & E_{o,k} \geq 0, \forall k \in \mathcal{K}, \\ & d_{o,k} \geq 0, d_{l,k} \geq 0, \forall k \in \mathcal{K}. \end{aligned}$$

Substitute t_p with $(T - t_o)$ by following 3, and replace $d_{l,k}$ with $(D_k - d_{o,k})$ in Problem 1, we only need to solve the following optimization problem.

Problem 2.

$$\min_{t_o,\{w_k\},\{E_{o,k}\},\{d_{o,k}\}} \quad \sum_{k=1}^{K}\left(E_{o,k}+\frac{c_0C_k^3(D_k-d_{o,k})^3}{T^2}\right)$$

$$s.t. \quad \sum_{k=1}^{K} w_k \leq W_T, \tag{9a}$$

$$\textstyle\sum_{k=1}^{K} C_k d_{o,k} \leq (T - t_o) f_E, \tag{9b}$$

$$d_{o,k} \leq w_k t_o \ln\left(1+\frac{E_{o,k}h_k}{w_k t_o}\right), \forall k \in \mathcal{K} \tag{9c}$$

$$t_o \geq 0, \forall k \in \mathcal{K}, \tag{9d}$$

$$w_k \geq 0, \forall k \in \mathcal{K}, \tag{9e}$$

$$E_{o,k} \geq 0, \forall k \in \mathcal{K}, \tag{9f}$$

$$d_{o,k} \geq 0, \forall k \in \mathcal{K}. \tag{9g}$$

3 Solution

In this part, we turn to solve Problem 2. It can be checked that Problem 2 is non-convex since the function $w_k t_o \ln\left(1+\frac{E_{o,k}h_k}{w_k t_o}\right)$ in (10c) are non-concave with w_k and t_o. Hence it is difficult to get the optimal solution of Problem 2. Next we turn to develop sub-optimal solution of Problem 2. In the first step, we decompose Problem 2 into two levels. In the lower level, the variable t_o is fixed, and the rest of variables of Problem 2 are optimized. Specifically, for the lower level, by fixing t_o, Problem 2 turns to be the following optimization problem

Problem 3.

$$F(t_o) \triangleq \min_{\{w_k\},\{E_{o,k}\},\{d_{o,k}\}} \quad \sum_{k=1}^{K}\left(E_{o,k}+\frac{c_0C_k^3(D_k-d_{o,k})^3}{T^2}\right)$$

$$s.t. \quad \sum_{k=1}^{K} w_k \leq W_T, \tag{10a}$$

$$\textstyle\sum_{k=1}^{K} C_k d_{o,k} \leq (T - t_o) f_E, \tag{10b}$$

$$d_{o,k} \leq w_k t_o \ln\left(1 + \frac{E_{o,k} h_k}{w_k t_o}\right), \forall k \in \mathcal{K} \quad (10c)$$
$$t_o \geq 0, \forall k \in \mathcal{K}, \quad (10d)$$
$$w_k \geq 0, \forall k \in \mathcal{K}, \quad (10e)$$
$$E_{o,k} \geq 0, \forall k \in \mathcal{K}, \quad (10f)$$
$$d_{o,k} \geq 0, \forall k \in \mathcal{K}. \quad (10g)$$

In the upper level, the variable t_o is to be optimized to minimize the function $F(t_o)$, i.e., we need to solve the following optimization problem.

Problem 4.

$$\min_{t_o} F(t_o)$$
$$s.t.\ \ 0 \leq t_o \leq T$$

It can be checked that solving Problem 4 is equivalent with solving Problem 2.

For Problem 3, there is such a lemma.

Lemma 1. *Problem 3 is a convex optimization problem.*

Proof. In Problem 3, it can be easily checked that the objective function of Problem 3 is convex with $E_{o,k}$ and $d_{o,k}$, it can be also checked that the function $w_k t_o \ln\left(1 + \frac{E_{o,k} h_k}{w_k t_o}\right)$ in (10c) is concave with w_k when t_o is fixed. Hence Problem 3 is a convex optimization problem [8].

This completes the proof.

Remark: Problem 3 is a convex optimization problem. Hence it can be solved by existing numerical methods optimally.

For Problem 4, the optimization of t_o is essentially a one-dimensional search problem. We adopt bisect search method to find the optimal solution. Local optimality can be guaranteed at least.

4 Conclusion

In this paper, a MEC system with one MEC server and multiple mobile devices under FDMA mode is investigated. The energy consumption of all the mobile devices is to be minimized by optimizing various computation and communication resources. The associated optimization problems are formulated, and found to non-convex. To solve the formulated optimization problem, we decompose the formulated problem to two levels, in which a convex optimization problem need to be solved in the lower level and a bisection search need to be run in the upper level. The proposed method offer an solution at least with local optimality for the formulated optimization problem. In the future, it is of importance and interesting to study the security issue [9,10], as each mobile device offload private data to the MEC server.

References

1. Barbarossa, S., Sardellitti, S., Lorenzo, P.D.: Communicating while computing: distributed mobile cloud computing over 5G heterogeneous networks. IEEE Signal Proc. Mag. **31**(6), 45–55 (2014)
2. Mao, Y., You, C., Zhang, J., Huang, K., Letaief, K.B.: A survey on mobile edge computing: the communication perspective. IEEE Commun. Surv. Tut. **19**(4), 2322–2358 (2017)
3. Milan, P., Jerome, J., Valerie, Y., Sadayuki, A.: Mobile-edge computing introductory technical white paper. White Paper (2014)
4. Wang, Y., Sheng, M., Wang, X., Wang, L., Li, J.: Mobile-edge computing: partial computation offloading using dynamic voltage scaling. IEEE Trans. Commun. **64**(12), 4268–4282 (2016)
5. You, C., Huang, K., Chae, H., Qi, H., Kim, B.: Energy-efficient resource allocation for mobile-edge computation offloading. IEEE Trans. Wirel. Commun. **16**(3), 1397–1411 (2017)
6. Tao, X., Ota, K., Dong, M., Qi, H., Li, K.: Performance guaranteed computation offloading for mobile-edge cloud computing. IEEE Wirel. Commun. Lett. **6**(6), 774–777 (2017)
7. Zhao, P., Tian, H., Qin, C., Nie, G.: Energy-saving offloading by jointly allocating radio and computational resources for mobile edge computing. IEEE Access **6**, 11255–11268 (2017)
8. Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
9. Tu, W., Lai, L.: Keyless authentication and authenticated capacity. IEEE Trans. Inf. Theory **64**(5), 3696–3714 (2016)
10. Tu, W., Goldenbaum, M., Lai, L., Poor, H.V.: On simultaneously generating multiple keys in a joint source-channel model. IEEE Trans. Inf. Forensics Secur. **12**(2), 298–308 (2017)