ARTICLE IN PRESS

Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

User mobility aware task assignment for Mobile Edge Computing

Zi Wang [a], Zhiwei Zhao [a,c,*], Geyong Min [b], Xinyuan Huang [a], Qiang Ni [d], Rong Wang [e]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[b] College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK
[c] The Research Institute of Information Technology and National Laboratory for Information Science and Technology (TNList), Tsinghua University, China
[d] School of Computing and Communications, Lancaster University, UK
[e] College of Computer Science, Zhejiang University, China

HIGHLIGHTS

- We formally model the problem of task assignment in MEC as a constraint satisfaction problem, which jointly considers user mobility and resource distributions in the MEC end.
- We propose a lightweight algorithm for assignment, which can increase the resource utilization of the MEC nodes.
- We propose an accurate delay estimation scheme which supports accurate offloading decisions on mobile devices.
- We conduct simulation experiments for performance evaluation. The results show that the proposed work can reduce the task execution delay and increase the resource utilization.

ARTICLE INFO

Article history:
Received 17 August 2017
Received in revised form 26 January 2018
Accepted 8 February 2018
Available online xxxx

Keywords:
Mobile Edge Computing
Task assignment
User mobility
Offloading

ABSTRACT

Mobile Edge Computing (MEC) has emerged as a prospective computing paradigm to provide pervasive computing and storage services for mobile and big data applications. In MEC, many small cell base stations (sBSs) are deployed to establish a mobile edge network (MEN). These sBSs can be usually accessed directly by mobile users. The computational tasks are first offloaded from mobile users to the MEN and then executed in one or several specific sBSs in the MEN. While the offloading decision has been well studied, the task execution delay on the MEN side is overlooked. This paper aims at reducing the task execution delay by task scheduling in MENs. Specifically, we jointly consider the task properties, the user mobility and network constraints. The problem is formalized as a constraint satisfaction problem and a lightweight heuristic solution is proposed for fast scheduling. We conduct simulation experiments to study the performance of the proposed work. The results show that our work is able to significantly reduce the task execution delay in MENs and thus reduces the end-to-end delay for MEC tasks.

1. Introduction

Modern mobile devices are becoming more and more powerful and the mobile applications are becoming increasingly computation-intensive and delay-sensitive, such as real-time online gaming [1], augmented/virtual reality (AR/VR) [2], image/video processing APPs [3] and the vehicle networking systems [4]. These applications can often introduce a large amount of traffic and computational workload, which can potentially cause the battery drain problem on mobile devices [5]. To deal with the resource constraints of mobile devices, Mobile Edge Computing (MEC) has emerged as a promising computing paradigm. In MEC, a number of small cell base stations (sBSs) with computation and storage capability [3] are deployed to construct a network termed as the mobile edge network (MEN). The MENs provide computation and storage services in close proximity to subscribers to meet the high-workload and low-latency requirements [6]. Some projects such as TROPIC [7] and SESAME [8] introduce powerful sBSs, like picocells or femtocells, to share their enhanced high performance computing capabilities for high-performance edge computing. In the MEC paradigm, computational tasks are first offloaded from mobile users to the MENs and then executed in one or several specific sBSs in the network. Compared to traditional cloud computing, MEC is more close to users and incurs much less computational and transmission delay. The deployment of MEC can also provide more flexible resource scheduling for the mobile tasks.

* Corresponding author at: School of Computer Science and Engineering, University of Electronic Science and Technology of China, China.
E-mail addresses: wangzi@mobinets.org (Z. Wang), zzw@uestc.edu.cn (Z. Zhao), g.min@exeter.ac.uk (G. Min), xinyuan@mobinets.org (X. Huang), q.ni@lancaster.ac.uk (Q. Ni), zjajwr@zju.edu.cn (R. Wang).

https://doi.org/10.1016/j.future.2018.02.014
0167-739X/© 2018 Published by Elsevier B.V.

The current research works on MEC often focus on the offloading decision problem on mobile devices [9,10]. For example, X. Chen et al. [9] employed a game-theoretic computation offloading approach for multi-user systems. S. Guo et al. [10] proposed a dynamic offloading and resource scheduling policy to reduce energy consumption and the application completion time. Some other works such as [11] aimed at utilizing the synergy between the computation resources of MEC and the mobile devices. These works can effectively reduce the computational delay on mobile devices. Although the user-end delay is reduced, we notice that much delay is consumed at the MEC end. Different from the traditional cloud computing where execution delay is stable [12], it can be highly diverse in MEC because the tasks may be executed in different sBSs which have different resource situations. There are several research works on optimizing the task execution delay in MENs [13,14]. Among the problems, the task assignment in the MENs plays an essential role in reducing the execution delay, i.e., assigning the tasks to appropriate sBSs to achieve minimized execution delay in MENs.

The user mobility, which is a fundamental characteristic in MEC, have a large impact on the task execution in two ways. First, mobile users lead to time-varying workload distribution. Different sBSs have different numbers of connected mobile users. Second, mobile users often upload tasks and receive results via different sBSs, involving in-MEN communications and computation workload. However, the existing works often overlook the impact of user mobility and assign tasks to the directly connected sBSs. Some works consider mobility based on intuitive models, which may not reflect the exact impact of user mobility and thus yield inefficient task executions [13].

To address the impact of user mobility on task execution within MENs, we consider task assignment based on user trajectory prediction [15,16]. By jointly considering the user mobility, task properties and the resource distribution in the MEN, we formally model the problem as a constraint satisfaction problem. We then proposed a lightweight heuristic approach to the problem. Based on the task assignment/scheduling scheme, we further propose a delay estimation for MEC tasks to support accurate task offloading on mobile devices. Considering MEC is usually deployed in scenarios such as flight terminals, shopping malls, etc., the user mobility can be highly predictable [16]. We conduct extensive simulation experiments and the results show that the proposed work can significantly reduce the execution time of tasks in MEC networks.

The major contributions of this paper are summarized as follows:

1. We formally model the problem of task assignment as a constraint satisfaction problem, which jointly considers the user mobility, task properties and the resource distributions in the MEC network.
2. We propose a lightweight algorithm for the problem, which can effectively increase the resource utilization at the MEC end.
3. Based on the proposed assignment, we propose an accurate delay estimation scheme which can support accurate offloading decisions on mobile devices.
4. We conduct simulation experiments and show that the proposed work outperforms the existing works in terms of task execution delay.

This rest of the paper is organized as follow. Section 2 summarizes the related works. Section 3 presents the proposed model and algorithm. Section 4 presents the evaluation of the proposed work. Section 5 concludes this work and points future directions.

2. Related work

A large number of existing works have studied the optimization problems on mobile edge computing. Most of these works focus on the decision making problem for task offloading at the user end. These works will first estimate the expected delay or energy consumption for target tasks and then offload them to the mobile edge if the delay or energy consumption is reduced. According to the optimization goal, we divide these works into two categories: works on optimizing energy consumption and works on reducing the task delay. Our work falls into the second category and differs from the existing works in that our aim is to reduce the delay at the network end instead of the user end.

Works on optimizing energy consumption. K. Zhang et al. [17] presented a multi-device computation offloading framework for MEC and formulated an optimization problem that minimized the device energy consumption. A three-stage offloading scheme was proposed to obtain the sub-optimal solution, which (1) classified the mobile device, (2) determined the priority and (3) allocated the radio resource. X. Chen et al. [9] further considered the interference and collisions when there were too many users trying to offload tasks to the same sBS, which can significantly increase the energy consumption of mobile devices. The offloading was formulated as a multi-user game, which was proved always admitting a Nash equilibrium. W. Labidi et al. [18] considered the time varying channel state for wireless offloading and proposed a scheduling scheme for task offloading, which tried to make the best use of wireless channels and user buffers to reduce the energy consumption.

Works on reducing the task delay. J. Liu et al. [11] tried to minimize the execution delay for single users with one-dimensional search algorithm. The algorithm outputted a policy for offloading decision according to the application buffer queuing state. Besides, the characteristics of wireless channels were also considered. Y. Mao et al. [19] jointly optimized the task offloading scheduling and transmitted power allocation problem to reduce the offloading delay. Plachy et al. [20] took the spatial diversity of the sBSs into account in the offloading process. The sBS that was responsible for task execution was chosen by the users. Then the results would be returned to the users via the sBS with the highest RSSI of the wireless links. However, the work was designed for single task offloading.

There are also some works that jointly optimize both energy and task delay. Reducing task delay often adds additional energy consumption in MEC, especially for the tasks that execute faster on mobile devices than on mobile edges. Some works set an energy threshold and then minimize the execution delay without exceeding the energy threshold. For example, Y. Mao et al. [14] proposed a dynamic offloading scheme for single user to minimize the execution delay for energy harvesting devices, where the energy harvesting technique added complexity to the offloading algorithms. A lightweight approximation approach was proposed to achieve a good tradeoff between complexity and the delay minimization. J. Yang et al. [21] used a multi-stage sequential game model to meet the energy and delay requirements at the same time.

Different from the aforementioned works, our work emphasizes on minimizing the delay within the MENs. The delay in MENs not only adds to the overall task execution delay but also impacts the decision making process at the user end. We propose a novel in-network task scheduling approach to reduce the task execution delay, where task information, resource information and user mobility are jointly considered. Different from the [13] which used the contact rate to capture user mobility, our work is built on top of the user trajectory prediction [15,16], which can accurately capture user mobility in most MEC scenarios such as airports, shopping malls, etc. Besides, to deal with the impact on offloading decision, we propose a novel delay estimation scheme for mobile devices.

Please cite this article in press as: Z. Wang, et al., User mobility aware task assignment for Mobile Edge Computing, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.02.014.