Computers and Electrical Engineering 44 (2015) 218–240

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Context-aware multi-objective resource allocation in mobile cloud ☆

Simin Ghasemi-Falavarjani [a,*], Mohammadali Nematbakhsh [a], Behrouz Shahgholi Ghahfarokhi [b]

[a] Department of Software Engineering, University of Isfahan, Isfahan, Iran
[b] Department of Information Technology Engineering, University of Isfahan, Isfahan, Iran

ARTICLE INFO

Article history:
Received 15 March 2014
Received in revised form 6 February 2015
Accepted 9 February 2015
Available online 16 March 2015

Keywords:
Mobile cloud
Offloading
Context-awareness
Resource allocation
Multi-objective optimization

ABSTRACT

Mobile Cloud Computing (MCC) augments capabilities of mobile devices by offloading applications to cloud. Resource allocation is one of the most challenging issues in MCC which is investigated in this paper considering neighboring mobile devices as service providers. The objective of the resource allocation is to select service providers minimizing the completion time of the offloading along maximizing lifetime of mobile devices satisfying deadline constraint. The paper proposes a two-stage approach to solve the problem: first, Non-dominated Sorting Genetic Algorithm II (NSGA-II) is applied to obtain the Pareto solution set; second, entropy weight and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method are employed to specify the best compromise solution. Furthermore, a context-aware offloading middleware is developed to collect contextual information and handle offloading process. Moreover, to stimulate selfish users, a virtual credit based incentive mechanism is exploited in offloading decision. The experimental results demonstrate the ability of the proposed resource allocation approach to manage the trade-off between time and energy comparing to traditional algorithms.

1. Introduction

In spite of significant improvements of mobile device technologies, mobile devices still encounter resource scarcity (limited processing power, memory, battery life) running intensive computing applications such as image processing, speech recognition and data mining. To overcome this resource restriction, it has been suggested to offload computation from mobile device to the cloud [1].

Doing this, Mobile Cloud Computing (MCC) augments computing capabilities of mobile devices to enable them to run computational intensive applications to provide better user-experiences. On the other hand, mobile devices have sensing abilities to gather contextual information of user environment. Therefore MCC can provide adoptive, efficient, context-aware services for users which is a common view of pervasive computing [1,2].

A three-tier architecture is defined for MCC which consists of remote cloud servers, nearby computing servers known as cloudlets and neighboring mobile devices [1,3,4]. However, as cloud servers are not always available, offloading to remote cloud can be costly and introduces latency [5–7] while offloading to local cloudlets limits mobility of devices [1,6,8]. In this

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr Hamed Vahdat-Nejad.
* Corresponding author.
E-mail addresses: simin.ghasemi@gmail.com (S. Ghasemi-Falavarjani), nematbakhsh@eng.ui.ac.ir (M. Nematbakhsh), shahgholi@eng.ui.ac.ir (B. Shahgholi Ghahfarokhi).

http://dx.doi.org/10.1016/j.compeleceng.2015.02.006
0045-7906/

paper third tier is considered which is defined as 'Mobile Cloud' [9,10] consisting of neighboring mobile devices (mobile phones, PDAs, laptops, etc.) which are encountered to each other opportunistically and belong to different individuals. In the Mobile Cloud, both service requester and service providers are mobile devices.

Given an offloading service request in Mobile Cloud, as shown in Fig. 1, it is a challenging issue to determine whether and where to offload the subtasks among vicinal mobile devices minimizing the completion time of tasks as well as consumed energy of all mobile devices along satisfying deadline constraint. In addition, in order to provide dynamic and accurate resource allocation, it is required to gather contextual information of devices, applications, and environment to be used in decision [11]. Moreover, an integrated system is required to manage the offloading process in mobile cloud.

Most of the previous works on resource allocation in MCC attempt to decide only which portion of application to be offloaded to the cloud in order to optimize execution time or energy [3,4,8,12,13]. Some other studies try to select appropriate resource providers among available mobile devices to allocate tasks [5,9,10,14–17]. For example, Scavenger [14] uses a dual-profiling scheduler and develops a cyber foraging framework to allocate tasks to available surrogates towards minimizing execution time. In [5], a context aware application offloading scheme is proposed for mobile peer-to-peer environments. The authors define three context aware offloading policies that select a surrogate for each application partition in order to minimize the initiator device's energy consumption, or minimize the execution time of the application or maximize the success in application execution. In Serendipity [9], initiator device assigns subtasks to resource providers among intermittently encountered mobile devices. Time-optimizing Serendipity only minimizes completion time while energy-aware Serendipity only minimizes energy consumption of mobile devices. In [10] authors define a Mobile Device Cloud (MDC) environment consisting of highly collaborative mobile devices where the offloading schema is to only maximize the lifetime of the MDC. Authors in [16] introduce HYCCUPS, an offloading framework for Hybrid Contextual Cloud for Ubiquitous Platforms comprising of Smart phones and propose an adaptive contextual search algorithm to schedule mobile applications. In local decision, the initiator device decides whether to execute the task locally or offload it onto another device in order to minimize overall execution time regarding the battery energy, resource availability and mobility of mobile while in remote decision, a remote device determines whether it can execute that task or not.

However, to the best of our knowledge, those methods mostly focus on a single decision metric (usually time or energy) to decide about the offloading target. Specifically, none of the previous literatures in MCC provides a resource allocation approach which optimizes both overall execution time and consumed energy simultaneously while minimizing completion time along the energy consumption is a significant problem since in addition to keeping tasks deadline, minimizing the tasks completion time improves the quality of users' experience from using MCC and minimizing energy consumption simultaneously increases the lifetime of the mobile cloud.

In Our previous work [15], we have investigated the resource allocation problem among neighboring mobile devices aiming to minimize the completion time of offloading while maximizing lifetime of mobile devices. An Optimal Fair Multi-criteria Resource Allocation (OFMRA) algorithm has been proposed to solve the resource allocation problem assuming that the subtasks are identical. However most of applications have subtasks with different computational workloads. In such case, the resource allocation problem is a more complex problem. In addition, an Offloading Mobile Cloud Framework (OMCF) has been developed which collects profile information and handles the offloading process.

In this paper, the resource allocation problem is formulated as a multi-objective optimization that aims to minimize completion time of tasks as well as consumed energy of all participating mobile devices along satisfying deadlines. In contrast to our previous work [15], the subtasks are assumed to have different size which makes the resource allocation problem more complex than previous study. Therefore, to solve the multi-objective optimization in reasonable time, a two-stage approach is employed. First, a Non-dominated Sorting Genetic Algorithm II (NSGA-II) is applied to the problem to obtain the set of Pareto solutions. Second, a Multi-Attribute Decision-Making (MADM) technique is used to specify best compromise solution which is based on entropy weight and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). It is

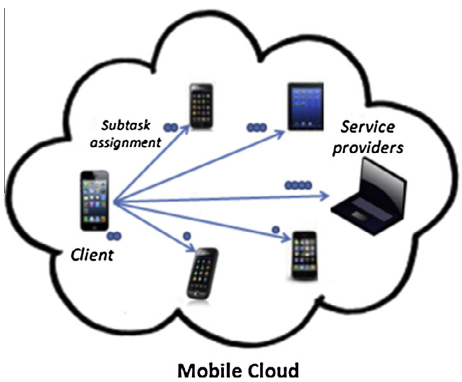

Fig. 1. Resource allocation in Mobile Cloud consisting of neighboring mobile devices.