Simulation Modelling Practice and Theory 58 (2015) 140–156

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A modeling and simulation framework for mobile cloud computing

M. Amoretti *, A. Grazioli, F. Zanichelli

Department of Information Engineering, Università degli Studi di Parma, Italy

ARTICLE INFO

Article history:
Available online 6 June 2015

Keywords:
Mobile cloud
Discrete event simulation
Adaptive loops

ABSTRACT

Mobile cloud computing (MCC) is an emerging paradigm for transparent elastic augmentation of mobile devices capabilities, exploiting ubiquitous wireless access to cloud storage and computing resources. MCC aims at increasing the range of resource-intensive tasks supported by mobile devices, while preserving and extending their resources. Its main concerns regard the augmentation of energy efficiency, storage capabilities, processing power and data safety, to improve the experience of mobile users. The design of MCC systems is a challenging task, because both the mobile device and the Cloud have to find energy-time tradeoffs and the choices on one side affect the performance of the other side. The analysis of the MCC literature points out that all existing models focus on mobile devices, considering the Cloud as a system with unlimited resources. Also, to the best of our knowledge, no MCC-specific simulation tool exists. To fill this gap, in this paper, we propose a modeling and simulation framework for the design and analysis of MCC systems, encompassing all their components. The main pillar of the proposed framework is the autonomic strategy consisting of adaptive loops between every mobile devices and the Cloud. The proposed model of the mobile device takes into account online estimations of the actual Cloud performance – not only the nominal values of the performance indicators. At the same time, the model of the Cloud takes into consideration the characteristics of the workload, to adapt its configuration in terms of active virtual machines and task management strategies. Moreover, the developed discrete event simulator is an effective tool for the evaluation of an MCC system as a whole, or single components, considering different classes of parallel jobs.

1. Introduction

Mobile cloud computing (MCC) is an emerging paradigm that leverages ubiquitous access to unified elastic resources provided by varied cloud and network technologies, with the purpose of augmenting mobile devices capabilities in terms of energy efficiency, storage capabilities, processing power and data safety [1]. Mobile users' experience gets enriched by the possibility of executing resource-intensive applications whose performance would be hindered by intrinsic deficiencies of mobile devices. Examples of applications which can benefit from MCC include computing-intensive software, such as

* Corresponding author.
E-mail addresses: michele.amoretti@unipr.it (M. Amoretti), alessandro.grazioli1@studenti.unipr.it (A. Grazioli), francesco.zanichelli@unipr.it (F. Zanichelli).

http://dx.doi.org/10.1016/j.simpat.2015.05.004
1569-190X/

speech recognition and natural language processing, data-intensive programs, such as enterprise applications and communication-intensive applications, such as online video streaming [2].

The design of MCC systems is a challenging task, because both the mobile device and the Cloud have to find energy-time tradeoffs and the decisions of one part affect the performance of the other part. Abolfazli et al. [2], elaborated the entire history of MCC and presented the most interesting models and solutions. Their analysis points out that all MCC models focus on mobile devices, considering the Cloud as a system with unlimited resources [3,6]. Moreover, to the best of our knowledge, no MCC-specific simulation tool exists.

With the purpose to contribute in filling this gap, we propose a modeling and simulation framework for the design and analysis of MCC systems, encompassing both their sides. Such a framework is based on four pillars, that we list in order of importance.

The first pillar is the concept of *offloading probability* for the mobile device. Offloading decisions depend on energy and performance considerations. One approach is to compare, for each job, the energy cost for both scenarios; the one where the job is executed on the mobile device and the one where the job is sent to the Cloud for execution. Meeting the job deadline may be considered as a constraint. Jobs are assumed to be made of independent tasks – *e.g.*, image processing applications, where a single image may be split before processing, or a sequence of images may be batch-processed. The case of dependent tasks will be investigated in a future work.

The second pillar is the *characterization of the Cloud*, in terms of task dispatching and scheduling strategies. In this paper, we focus on only one type cloud-based resources, which are distant giant data centers. Thus, the Cloud model includes a Dispatcher (modeled as a FCFS queue) and a set of VMs. The Dispatcher assigns incoming tasks to VMs. Several dispatching strategies are considered, taking into account for example the length of the VMs' queues, or their current load.

The third pillar is the *discrete event simulator* of the proposed MCC model, which has been developed using the DEUS general-purpose discrete event simulation platform [17,18]. The simulator allows designers to characterize jobs, mobile devices, communication channels, the Cloud, as well as the events and processes that define the dynamics of the whole MCC system – *e.g.*, adaptive loops. The simulator has been validated against the OpenStack-based private Cloud of our Department.

The fourth (and most important) pillar is the *autonomic strategy consisting of adaptive loops* that involve mobile devices and the Cloud, aimed at automatically determining energy-time tradeoffs, and achieve better global performance. With respect to the reference strategy proposed by Kumar and Lu [3], we take into account the variable performance of the Cloud, with a concrete and effective approach. For example, when offloading decisions have to be taken for a given job, the mobile device can obtain up-to-date estimates of the Cloud speedup and available bandwidth. At the same time, the Cloud can periodically update the number of active VMs, based on workload characterization, in order to maintain a given level of service. Such processes affect each other, until a stable global state is reached. As the size and nature of the workload may change, the MCC system moves from a stable state to another.

The paper is organized as follows. In Section 2, we discuss related work in MCC modeling. In Section 3, the MCC model is illustrated in detail, also supported by a practical example. In Section 4, the discrete event simulator of the MCC model is illustrated. In Section 5, an extensive performance evaluation of an MCC system characterized by different classes of parallelizable jobs, carried out by means of the proposed simulator, is presented. Finally, in Section 6, conclusions and future research directions are illustrated.

2. Related work

Although MCC can be realized according to different approaches, which differ in the computation granularity considered by the offloading process (ranging from device cloning to application partitioning and migration) [7,8] and in the degree of involvement of the Cloud, the main question on the mobile platform is always: *to offload or not to offload*?

Barbera et al., considering an architecture where each real device is associated to a software clone in the Cloud, have given a precise evaluation of the offloading feasibility and costs in terms and bandwidth and energy consumption on the real device [9]. The first important observation is that almost 50% of the time users are connected to a WiFi access point. Second, synchronizing back-clones (for backup purposes) requires less network traffic and less energy overhead than synchronizing off-clones (that handle mobile computation offload). In this paper, we focus on the second scenario, which is generally considered more challenging.

Kumar and Lu have proposed a simple but effective formula for energy analysis related to computation offloading [3]. According to that formula, offloading is energetically favorable when the cost of transferring data is compensated by the cloud speedup. Kumar and Lu's formula is one of the fundamental ingredients we consider to define our MCC model, in Section 3.

Wen et al. have proposed and alternative approach, where the offloading decision is based on the solution of two optimization problems [10]. The first one is related to the minimization of the computation energy when mobile execution is chosen, by optimally configuring the clock frequency of the chip, via the dynamic voltage scaling (DVS) technology. The solution of the second problem, instead, is the minimum amount of energy consumed by the mobile device for the cloud execution, under an optimal transmission scheduling. This approach is based on detailed models of the device and communication channel, thus resulting in a more complex formulation with respect to Kumar and Lu's one. However, the