Offloading as a Service Middleware for Mobile Cloud Apps

Hamid A Jadad, SQU, Muscat, Oman
Abderezak Touzene, Sultan Qaboos University, Muscat, Oman
Khaled Day, Sultan Qaboos University, Muscat, Oman

ABSTRACT

Recently, much research has focused on the improvement of mobile app performance and their power optimization, by offloading computation from mobile devices to public cloud computing platforms. However, the scalability of these offloading services on a large scale is still a challenge. This article describes a solution to this scalability problem by proposing a middleware that provides offloading as a service (OAS) to large-scale implementation of mobile users and apps. The proposed middleware OAS uses adaptive VM allocation and deallocation algorithms based on a CPU rate prediction model. Furthermore, it dynamically schedules the requests using a load-balancing algorithm to ensure meeting QoS requirements at a lower cost. The authors have tested the proposed algorithm by conducting multiple simulations and compared our results with state-of-the-art algorithms based on various performance metrics under multiple load conditions. The results show that OAS achieves better response time with a minimum number of VMs and reduces 50% of the cost compared to existing approaches.

KEYWORDS
Apps, Load Balancing, Mobile Cloud, Offloading, Scheduling

1. INTRODUCTION

During the last decade, we have witnessed unprecedented growth in the mobile phone industry due to the popularity of smart devices owing to their affordability and advancement in related technologies. One reason for this popularity of smartphones is the rapid demand for mobile applications (GSMA corporate, 2017). For example, as of May 2019, the Google Play store contained around 2.1 million mobile apps, whereas the Apple Store offered 1.8 million apps (Statista, 2019). These apps can be classified into various categories such as entertainment, business, health care, and education. In the recent past, most of these apps required manageable computation, memory, and power. However, many advanced apps like multimedia processing, video gaming, speech recognition, and natural language processing which require a high level of computing power, memory, and energy. This scenario is quickly making it impractical to run such demanding apps locally on the devices. Hence, the cloud computing paradigm has come up with a solution.

DOI: 10.4018/IJCAC.2020040103

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Recently, cloud computing has changed the way of delivering computing services. Cloud providers can offer Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) to the end-user at low cost. Cloud characteristics such as flexibility, reliability, and cost-effectiveness are the major benefits of moving business to the cloud. Many recent cloud computing issues are discussed in (Botta, de Donato, Persico, & Pescapé, 2016), (Skourletopoulos et al., 2017a), and (Singh, Jeong, & Park, 2016). The work in (Botta et al., 2016) discusses the challenges to integrate Internet of Things (IoT) applications and cloud computing. The authors in (Al Ridhawi, Aloqaily, Kantarci, Jararweh, & Mouftah, 2018) introduce a service provision scheme to provide continuous availability of diversified cloud services targeting vehicular cloud users. The study in (Skourletopoulos et al., 2017a) shows current big data techniques and models that exploit cloud technologies. Singh et al. present a comprehensive (Singh et al., 2016) present a comprehensive review of the security challenges in cloud computing.

In order to minimize latency and enhance the performance of the offered services, public cloud providers have expanded their global infrastructure. They group their cloud data centers within a given geographic area called a region. Each region may contain multiple cloud data centers from different cloud providers. For example, IBM deploys 60 cloud data centers within 6 regions (IBM, n.d.). Each data center service has a different price because of different business expenses such as energy, carbon penalties, real estate taxes and operating costs in each data center location. For example, an instance of m4.xlarge on-demand Windows EC2 in the AWS U.S. East region costs $0.404 per hour. The same instance in the Asia Pacific (Singapore) region costs $0.455 (Computing, 2017). This cost difference becomes more significant for a large number of instances over longer periods. Therefore, the cost estimation of utilization of these large-scale systems is not easy to ascertain.

Mobile cloud computing is a new paradigm that appeared from merging cloud computing and mobile computing (Skourletopoulos et al., 2017b). It allows mobile users to utilize cloud services on demand. It is envisioned that this paradigm will help to overcome the limitations of the mobile device’s hardware. In (Lewis & Lago, 2015), the authors have proposed a taxonomy of mobile cloud computing based on the key issues and how they have been tackled in research. One of the key issues is job offloading which consists of migration of jobs (data or code) from resource-constrained mobile devices to the resource-rich cloud.

Most of existing work (Chun, Ihm, & Maniatis, 2011), (Yang et al., 2014), (Guo, Xiao, Yang, & Yang, 2016) use an offloading model where the application has two similar versions; one runs on the mobile device and the other runs on the server-side. The developer predetermines the computation methods to be offloaded from the mobile device to the server for execution. Both mobile and server must have the same runtime environment. The system dynamically decides to either offload computation methods or execute them locally on the mobile device based on different criteria. When a computation method is offloaded to the server for execution, its state (such as registers, stack, etc.) needs to be packaged with user input data and transferred to the server. This model of computation is clearly flawed due to the following reasons: Firstly, it produces an overhead to the application developer to predetermine the offloadable computation methods. Secondly, transferring a huge computation method state from a mobile device to a remote cloud server is costly in terms of Internet data bundle and transmission delay. Thirdly, the mobile device still needs high computing hardware such as processor and memory because the application may run locally if the choice is not to offload the computation to the server. This restricts the mobile device from running more advanced applications compared to desktops. Lastly, it does not make good use of cloud computing benefits since the mobile device still pays the cost of the offloading decision process in terms of battery and computing power.

In this paper, we adopt a novel but simple offloading model where all application computations are performed in the cloud. Only an application interface is installed in the user mobile device, and the user’s input data is transferred to the cloud. This model overcomes the challenges of existing offloading approaches. It is also envisioned for high-speed mobile networks such as 5G.
Related Content

Service Selection Based on Customer Preferences of Non-Functional Attributes
[www.igi-global.com/chapter/service-selection-based-customer-preferences/60890?camid=4v1a](www.igi-global.com/chapter/service-selection-based-customer-preferences/60890?camid=4v1a)

Commercial and Distributed Storage Systems
[www.igi-global.com/chapter/commercial-distributed-storage-systems/77427?camid=4v1a](www.igi-global.com/chapter/commercial-distributed-storage-systems/77427?camid=4v1a)
Identifying the Direct Effect of Experience and the Moderating Effect of Satisfaction in the Greek Online Market
www.igi-global.com/article/identifying-direct-effect-experience-moderating/53533?camid=4v1a

A Consumer Perception Research on the Subject of a New Technology in a Developing Dynamic Market: 3G Technology in Turkey
www.igi-global.com/chapter/consumer-perception-research-subject-new/72543?camid=4v1a