Simulation Modelling Practice and Theory 50 (2015) 151–164

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

Sensitivity analysis of a hierarchical model of mobile cloud computing

Rubens Matos [a,b,d,*], Jean Araujo [a,c], Danilo Oliveira [a], Paulo Maciel [a], Kishor Trivedi [d]

[a] Informatics Center, Federal University of Pernambuco, Brazil
[b] CAPES Foundation, Ministry of Education of Brazil, Brazil
[c] Academic Unit of Garanhuns, Federal Rural University of Pernambuco, Brazil
[d] Dept. of Electrical and Computer Engineering, Duke University, USA

ARTICLE INFO

Article history:
Available online 11 May 2014

Keywords:
Mobile cloud computing
Analytical models
Dependability evaluation
Sensitivity analysis

ABSTRACT

Mobile cloud computing is a new paradigm that uses cloud computing resources to overcome the limitations of mobile computing. Due to its complexity, dependability and performance studies of mobile clouds may require composite modeling techniques, using distinct models for each subsystem and combining state-based and non-state-based formalisms. This paper uses hierarchical modeling and four different sensitivity analysis techniques to determine the parameters that cause the greatest impact on the availability of a mobile cloud. The results show that distinct approaches provide similar results regarding the sensitivity ranking, with specific exceptions. A combined evaluation indicates that system availability may be improved effectively by focusing on a reduced set of factors that produce large variation on the measure of interest. The time needed to replace a fully discharged battery in the mobile device is a parameter with high impact on steady-state availability, as well as the coverage factor for the failures of some cloud servers. This paper also shows that a sensitivity analysis through partial derivatives may not capture the real level of impact for some parameters in a discrete domain, such as the number of active servers. The analysis through percentage differences, or the factorial design of experiments, fulfills such a gap.

1. Introduction

Mobile computing has undergone unprecedented advances, following a huge growth of smartphones and tablet computers market [1]. In addition, increase in the use of wireless networks (WiFi, 3G, and 4G) caused significant changes on the Internet landscape. In 2012, the percentage of traffic created by mobile devices reached about 10% of the total Internet traffic [2], and a growth of 66% in this traffic is estimated by 2017 [3].

Despite the recent advances, mobile computing suffers from resource scarcity, even on the most modern devices. The most common problems are interruption of wireless connectivity, lack of security, hand-off delay, battery discharge and limited computational power [4]. In this context, a new paradigm named Mobile Cloud Computing (MCC) was introduced recently. This aims to utilize cloud computing resources to overcome the limitations of mobile computing, allowing delivery

* Corresponding author at: Informatics Center, Federal University of Pernambuco, Brazil. Tel.: +55 8181309381; fax: +55 81 21268438.
E-mail addresses: rsmj@cin.ufpe.br (R. Matos), jcta@cin.ufpe.br, jean@uag.ufrpe.br (J. Araujo), dmo4@cin.ufpe.br (D. Oliveira), prmm@cin.ufpe.br (P. Maciel), kst@ee.duke.edu (K. Trivedi).

http://dx.doi.org/10.1016/j.simpat.2014.04.003
1569-190X/

of more sophisticated and innovative applications to the user. The mobile cloud computing market is expected to reach 45 billion dollars in revenues by 2016 [5]. Considering this financial impact level, it is essential to provide services that can be justifiably trusted, that is, *dependable* services.

Among the dependability attributes of a system, availability is one which relates directly to the proportion of time that the system is found operational. Achieving high availability is important to avoid revenue loss, and other harmful consequences of service outage (e.g., interruption of a remote patient monitoring system). When a company's workforce needs to move around remote areas to accomplish their duties, the time wasted due to system unavailability may also imply low productivity of its employees.

The availability modeling and analysis of mobile clouds require the investigation of a large number of possible events in client, communication, and server domains. This paper presents sensitivity analysis of mobile cloud availability based on hierarchical analytical models and distinct techniques to assess the impact of each input parameter. This analysis aims to identify the bottlenecks for system improvement in a case study. We also use a combined evaluation of results from three techniques which complement each other to deal with the analysis of this system. The results show that the system availability may be improved effectively by focusing on a reduced set of factors which produce large variation on steady-state availability.

This paper is structured as follows. Section 2 provides important concepts regarding mobile clouds and also presents a background about sensitivity analysis of stochastic models. Section 3 describes a mobile cloud architecture which is analyzed in this paper, whereas Section 4 presents the availability models developed to represent such a mobile cloud system. Section 5 presents the availability results obtained through analytical models and sensitivity analysis performed through distinct techniques. Finally, Section 6 draws the final remarks.

2. Background

This section presents basic concepts related to mobile cloud computing, and sensitivity analysis of stochastic models, which are fundamental to this paper.

2.1. Mobile cloud computing

Scarcity of computational resources is one of the main problems faced by users and developers of mobile applications [6]. Even with the hardware advances in the most modern smartphones—processors with four cores, two gigabytes of main memory, 32 gigabytes of secondary memory—such configurations will always be inferior when compared with desktop computers of the same generation [7]. If we design a heavyweight application suited only for more powerful devices, we will restrain it for a great range of users with more limited smartphones and simpler feature phones. Moreover, because battery technologies have not evolved at the same pace as other hardware components [8], more intensive applications may significantly reduce the time for battery discharge.

Mobile cloud computing (MCC) is a paradigm that has emerged in recent years, with the purpose of surpassing some limitations in mobile computing [9]. As mentioned in [10], the main idea of MCC is to use cloud computing mechanisms to provide computational resources on demand, in order to offload resource intensive tasks, and alleviate resource utilization in mobile devices. This new paradigm can also leverage the provision of smarter mobile applications that extend the user's cognitive capabilities, as for example, speech recognition, natural language processing, computer vision and graphics, machine learning, augmented reality, planning and decision-making [6].

Mobile cloud computing can also be defined as cloud computing when some of the involved devices are mobile [11]. As cloud-based applications have become more popular among users, and the number of smartphone users with Internet access grows, the encounter of those two IT trends is unavoidable. The number of MCC subscribers worldwide is expected to reach almost 1 billion by 2014, and the revenues generated exclusively by 240 million mobile business users may reach 5 billion dollars by the end of 2015 [12].

It is in this context of cloud computing that the concept of Mobile Backend as a Service has emerged [13]. The purpose of MBaaS is to provide an environment to aid the development of mobile applications, reducing its time to market. Some features provided by a MBaaS cloud are: data and processing offloading, users management, push notification, and integration with social networks [14].

2.2. Sensitivity analysis of analytical models

Parametric sensitivity analysis aims to identify the factors for which the smallest variation implies the highest impact in model's output measure [15,16]. The main aim of parametric sensitivity analysis is to predict the effect on outputs (measures) with respect to variations in inputs (parameters), helping to find performance or reliability bottlenecks, and guiding an optimization process [17]. There are many ways of performing parametric sensitivity analysis. Factorial experimental design [18], correlation analysis and regression analysis [19] are some well known techniques. The simplest method is to repeatedly vary one parameter at a time while keeping the others constant. When applying this method, a sensitivity ranking