Simulation Modelling Practice and Theory 50 (2015) 83–95

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

Matrix based proactive resource provisioning in mobile cloud environment

Sandeep K. Sood *, Rajinder Sandhu

Computer Science and Engineering, Guru Nanak Dev University, Regional Campus, Gurdaspur 143521, India

ARTICLE INFO

Article history:
Available online 4 July 2014

Keywords:
Cloud computing
Mobile cloud
Resource provisioning
Resource provisioning matrix
Back propagation neural networks
Resource bill calculation

ABSTRACT

Mobile cloud computing is a dynamic, virtually scalable and network based computing environment where mobile device acts as a thin client and applications run on remote cloud servers. Mobile cloud computing resources required by different users depend on their respective personalized applications. Therefore, efficient resource provisioning in mobile clouds is an important aspect that needs special attention in order to make the mobile cloud computing a highly optimized entity. This paper proposes an adaptive model for efficient resource provisioning in mobile clouds by predicting and storing resource usages in a two dimensional matrix termed as resource provisioning matrix. These resource provisioning matrices are further used by an independent authority to predict future required resources using artificial neural network. Independent authority also checks and verifies resource usage bill computed by cloud service provider using resource provisioning matrices. It provides cost computation reliability for mobile customers in mobile cloud environment. Proposed model is implemented on Hadoop using three different applications. Results indicate that proposed model provides better mobile cloud resources utilization as well as maintains quality of service for mobile customer. Proposed model increases battery life of mobile device and decreases data usage cost for mobile customer.

1. Introduction

Cloud user can be an individual or an organization that takes services from the Cloud Service Provider (CSP). CSP provides storage management, universal data access, avoids capital and maintenance expenditure on hardware as well as on software. It is an emerging business model that delivers information technology services over the internet in an elastic, self-provisioned, dynamic and cost-effective manner with guaranteed Quality of Service (QoS) [1]. Cloud computing follows three main service models viz. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The requirement of cloud service model depends upon the needs of end user. CSP provides dynamic memory allocation, processing power and network bandwidth for the clients programs in order to execute them smoothly [2]. As the load of the user programs executing at the cloud increases, more resources are provisioned and as the load decreases, resources are revoked from the cloud user and made available where ever these are needed. CSP can also provides the software development environment to cloud users which will assist in software development process by removing the complexities involved with

* Corresponding author. Tel.: +91 9465204534.
E-mail addresses: san1198@gmail.com (S.K. Sood), rajsandhu1989@gmail.com (R. Sandhu).

http://dx.doi.org/10.1016/j.simpat.2014.06.004
1569-190X/

licensing, number of users and software counterfeit. Cloud users also have not to worry about the updates of software and resources. Cloud provides applications such as gmail, dropbox and google drive over the internet to perform specific tasks. Cloud user can use different applications without worrying about all backend activities [3,4].

With increase in number of smart phone applications and availability of network bandwidth, user can perform many tasks from mobiles device itself irrespective of its location or mobility. Smart phones also have their constraints such as low battery life and less computation power which make these devices unable to run high computational applications for longer period of time. Although, these applications can be offloaded to remote servers where major execution can take place and results are sent back to mobile devices. Cloud computing provides virtually infinite remote servers which are available on demand and with pay as you use pricing model. So, mobile cloud computing can be stated as the concept where cloud resources are used for mobile application execution and only desired results are sent back to user's mobile device using the internet [5]. Mobile devices can also be used to create a peer to peer mobile cloud environment for resource sharing which can be used for different applications such as image processing, natural language processing and disaster recovery [6]. However, along with many benefits of mobile cloud computing, it also inherits many issues such as limited battery life, disconnection due to low reception in remote areas, underutilized cloud resources, increased cost and security of contents hosted on third party cloud. Different users have different resource needs depending upon diverse set of application installed and time of use of these applications. Some users just use mobile devices for mail and messaging services whereas some user watches youtube videos or play online games. Time of use of mobile applications also varies for different users. Most of mobile devices do not use much CSP resources at night. Allocation of static mobile cloud resources to mobile customer will result in either under or over utilization of resources. It will also decrease battery life of mobile device as well as increase network data usage cost for mobile customer because resources are always allocated to mobile customer even when they are not required. Therefore, resource allocation in mobile cloud environment should be dynamic in nature. There are two types of dynamic resource allocation viz. proactive and reactive resource allocation. In proactive resource allocation, resources are predicted before allocating to mobile customer. In this way, resources can be provisioned before they are required. In reactive resource allocation, an action is taken in the response to some situation i.e. resources are provisioned to mobile customer after need is raised. No predictive analysis is done in case of reactive resource allocation strategy. Reactive resource allocation cause many issues such as delay in response, degraded performance and complex allocation process. Proactive resource allocation is cost effective by preventing under and over utilization of cloud resources, increase performance of application and helps CSP to easily manage the cloud infrastructure. Proactive resource allocation is best suited for mobile cloud environment for its optimum and efficient resource management.

Main objective of proposed model involving three entities viz. Cloud Service Provider (CSP), Mobile Customer (MC) and Independent Authority (IA) are:

- Involvement of independent third party cloud to ensure better mobile cloud resource allocation and quality of service for MC.
- Representation of past, current and future allocated or predicted resources in a two dimensional matrix termed as Resource Provisioning (RP) matrix. RP matrix assists in accurate resource prediction and resource bill calculation process.
- Prediction of mobile cloud resources for initial setup of mobile device depending upon the types of applications MC has installed. It will help to allocate required cloud resources accurately even for the first time hence preventing cost wastage. Future prediction is also more accurate due to correct initial allocation of resources.
- Continuous prediction of required mobile cloud resources using back propagation neural network. Therefore, it prevents under and over utilization of mobile cloud resources.
- IA checks, compares and verifies resource usage bill computed by CSP using RP matrices stored over time. That results in cost computation reliability for MC on mobile cloud environment.

The rest of this paper is organized as follows. Section 2 provides the motivation for proposed model. Section 3 summarizes the related works for optimized resource provisioning in a cloud environment and mobile cloud environment. In Section 4, the model is proposed which is designed to solve the resource provisioning issues in the mobile cloud. Section 5 presents the experimental results and performance analysis of the proposed approach. Finally, Section 6 concludes the paper.

2. Motivation

Motivation for the proposed model originated from many challenges of mobile cloud computing. This section discusses various challenges related to mobile cloud computing and also explains how to overcome these challenges.

2.1. Cost of cloud services

CSP should provide an effective and efficient mechanism to run any mobile application over the cloud. Pay as use model is employed by CSP for calculating cloud resource usage cost. Cost of cloud resource usage can be reduced if occupied resources by MC are released when MC is ideal for prolonged duration of time. For example, if MC does not use cloud applications for two (eight hours) out of six (twenty-four hours) quadrants of a day. Then, cloud resource usage cost for eight hours per day