Simulation Modelling Practice and Theory 50 (2015) 72–82

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A novel server selection approach for mobile cloud streaming service

Ran Liu [a,b], Xiaoqun Yuan [c,*], Jie Xu [c], Jianguo chen [a,b], Yuanxiang Zeng [c], Meng Cao [c], Jinhong Liu [c], Lifang Xu [c], Qing Fang [c]

[a] Institute of Public Safety Research, Tsinghua University, China
[b] Department of Engineering Physics, Tsinghua University, China
[c] School of Information Management, Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Mobile cloud streaming
Server selection
Mobility aware
Quality grade

ABSTRACT

Mobile cloud streaming has become very popular recently. However, the centralized datacenter structure may lead to high service delay, especially for real time and high bandwidth streaming services. Deploying cloud edge servers may improve the quality of mobile streaming services in theory, but it cannot well adapt to the user mobility. To deal with this problem, this paper puts forward our *MACSS*, a mobility-aware framework for mobile cloud streaming services, which provides dynamic and optimized server selection functions to support user mobility comprehensively. Then we further propose *CQS3*, a novel quality-aware scheme that frequently redirects user requests to the "best" server according to comprehensive service quality grade. Simulation results show that *CQS3* scales well and guarantees comparable delay performance. In addition, *CQS3* can significantly improve PSNR comparing with CSS and is well adapted to user mobility and channel variation.

1. Introduction

Recently, mobile cloud computing has received large interests form both industry and academic. It has been built to bridge the gap between increasing complex applications and insufficiency of mobile device capacities. With the development of the mobile devices capabilities, such as CPU power, network connectivity and sensors, mobile devices are becoming increasingly popular, and they are used to get access to Internet applications anywhere, anytime. The Internet report of CNNIC points out that the number of mobile Internet users in China has exceeded 460 million, accounted for 78.5% of Internet end users [1]. Huge numbers of users provide unprecedented opportunities for the development of mobile Internet, thus more mobile applications such as online gaming, navigation, streaming and speech recognition are emerging. As a result, more considerable computational power and energy are required. Unfortunately, due to the restrictions on weight, size, battery life, etc., mobile devices are not competent to execute the complex applications, especially for real time and high bandwidth consumption applications like streaming.

To overcome this problem, mobile cloud computing is introduced to offload these computationally expensive tasks to local cloud, such as Amazon's EC2 cloud and Apple's iCloud. Cloud computing provides all kinds of Internet service to end users dynamically with Virtual Machine (VM) technologies [2,19], which makes the service elastic and extendable easily. Unfortunately, these clouds generally adopt the centralized structure, and their datacenters are far from mobile users, leading to high

* Corresponding author. Tel./fax: +86 13343400223.
E-mail address: yuan20030308@163.com (X. Yuan).

http://dx.doi.org/10.1016/j.simpat.2014.06.014
1569-190X/

service delay and low quality of experience(QoE) [3,4]. To reduce service delay and improve QoE, Satyanarayanan [5] presented cloudlets: trusted, resource rich computers in the near vicinity of the mobile user. Based on this work, Verbelen et al. [6] proposed dynamic cloudlets, where all devices in the LAN network cooperated with each other. Both of these two methods used distributed technology to improve mobile service quality. That is to say, by using the "optimal" server near the mobile user to provide service, network congestion is eased and service response delay is reduced. Owing to the resource restrictions and the mobility, mobile devices cannot provide stable resource to other devices, so the latter cannot yield to the expected results. As to the former, guaranteed quality service can be provided on the condition that the appropriate servers are selected.

In fact, server selection, which redirects user requests to the "best" servers with pre-setting strategies, is one of the key technologies in distributed systems, and many researchers have done in-depth researches of it [7,8]. However, these solutions work well just in wired networks, not in mobile networks. Firstly, these solutions just help clients to select the "best" server at the initial stage and do not consider server re-selections. Users may have to tolerate the deteriorated QoE caused by their move. Secondly, they are all driven by application servers, not well considering the user perceived quality. Last but not least, they are based on IP address. Nevertheless, this address may not always truly reflect the position of mobile users. One recent study [13] exposed that in cellular networks, IP addresses can be shared across geographical disjoint regions within a short time. Obviously, IP addresses in mobile networks may not contain enough location information, which prevents these solutions from working well. So researchers began to study the server selection scheme in mobile distributed systems, and presented some new solutions [11–13]. However, the existed works only partially solve the problem caused by user mobility. Thus, it is necessary to propose a server selection strategy for mobile clouds similar to cloudlets.

In this paper, we aim to design a appropriate server selection strategy for mobile clouds to reduce their service delay and improve their quality. To this end, we firstly present a *Mobility-Aware Cloud Streaming System (MACSS)*, a framework to provide high quality services for mobile clouds by selecting appropriate servers dynamically, and the *Comprehensive service Quality Grade (CQG)*, a server selection metric. Then, we further propose a *Client-driven QoS-oriented Server Selection Scheme (CQS3)*, which takes both server-side preference and client-side requirement into account. At last, we design our experiment to verify the scalability and efficiency of the proposed scheme. The results show that, comparing with the Closest Surrogate Scheme (CSS) [14], our strategy significantly improves the Peak Signal-to-Noise Ratio (PSNR), adapting well to user mobility and channel variations while guaranteeing comparable delay performance. The major contributions of this paper are summarized as follows:

(1) We propose a *MACSS* framework, which provides dynamic server selection functions to support user mobility and maximize user perceived service quality.
(2) We present *CQS3* for mobile clouds to select the appropriate server near mobile user. Our server selection scheme takes quality of service (QoS) and QoE guarantee into consideration in a poor wireless network environment and makes the best tradeoff between server-side preference and client-side requirement. This makes our server selection scheme suitable for mobile environment.

The remainder of this paper is organized as follows: Section 2 reviews the related works about mobile Cloud and server selection. In Section 3, we propose our *MACSS* architecture and *CQG* in detail. Section 4 elaborates the working principle of *CQS3*. Section 5 describes the construction of a simulation experiment and the experimental result analysis. Section 6 concludes this paper and proposes future research goals.

2. Related works

This section describes the preliminaries and related works on mobile cloud computing and the server selection for distributed systems respectively.

2.1. Mobile cloud computing

Mobile cloud computing has been introduced to be a potential technology for mobile services based on the explosive growth of the mobile applications and technical maturity of cloud computing recently. It integrates cloud computing into mobile environment and overcomes obstacles of mobile computing such as storage and bandwidth. Due to the complementary property of mobile computing and cloud computing, mobile cloud computing has received large interest from both industry and academic.

In 2010, The Mobile Cloud Computing Forum defined Mobile Cloud Computing (MCC) as an infrastructure where both the data storage and the data processing happen in cloud computing systems outside mobile devices [16]. Then Aepona described MCC as a new paradigm that the data processing and storage had been moved from mobile devices to centralized cloud computing platforms [17]. Besides these mobile cloud computing definitions, Liu et al. [18] and Buyya et al. [19] defined MCC as a combination of mobile web and cloud computing to make mobile users access applications and services on the Internet conveniently. Besides, Pal and Henderson proposed a "Device to Device to Cloud" mobile architecture, which enabled tourists access to cloud computing resources via local users' network connections [26]. Except for offloading mobile device computing resource to clouds, mobile clouds also inherited some advantages from cloud computing, such as extending battery lifetime [20,21], improving data storage capacity and reliability [24] of mobile devices, and processing power [22,23].