Future Generation Computer Systems 29 (2013) 84–106

Contents lists available at SciVerse ScienceDirect

Future Generation Computer Systems

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

Mobile cloud computing: A survey

Niroshinie Fernando *, Seng W. Loke *, Wenny Rahayu

Department of Computer Science and Computer Engineering, La Trobe University, Australia

A R T I C L E I N F O

Article history:
Received 31 August 2011
Received in revised form
22 May 2012
Accepted 30 May 2012
Available online 6 June 2012

Keywords:
Mobile cloud computing
Task offload
Pervasive networks

A B S T R A C T

Despite increasing usage of mobile computing, exploiting its full potential is difficult due to its inherent problems such as resource scarcity, frequent disconnections, and mobility. Mobile cloud computing can address these problems by executing mobile applications on resource providers external to the mobile device. In this paper, we provide an extensive survey of mobile cloud computing research, while highlighting the specific concerns in mobile cloud computing. We present a taxonomy based on the key issues in this area, and discuss the different approaches taken to tackle these issues. We conclude the paper with a critical analysis of challenges that have not yet been fully met, and highlight directions for future work.

1. Introduction

The increasing usage of mobile computing is evident by the study by Juniper Research, which states that the consumer and enterprise market for cloud-based mobile applications is expected to rise to $9.5 billion by 2014 [1]. In recent years, applications targeted at mobile devices have started becoming abundant with applications in various categories such as entertainment, health, games, business, social networking, travel and news. The popularity of these are evident by browsing through mobile app download centers such as Apple's iTunes or Nokia's Ovi suite. The reason for this is that mobile computing is able to provide a tool to the user when and where it is needed irrespective of user movement, hence supporting location independence. Indeed, 'mobility' is one of the characteristics of a pervasive computing environment where the user is able to continue his/her work seamlessly regardless of his/her movement.

However, with mobility comes its inherent problems such as resource scarceness, finite energy and low connectivity as outlined by Satyanarayanan in [2]. These pose the problem of executing many useful programs that could aid the user and create a pervasive environment. According to Tim O'Reilly 'the future belongs to services that respond in real time to information provided either by their users or by nonhuman sensors' [3]. Real time applications are just one type of mobile applications that demand high levels of responsiveness, that in turn, demand intensive computing resources. Some mobile applications, such as location based social networking, process and make use of the phone's various sensor data. However, extensive use of sensors, such as obtaining a GPS reading, is expensive in terms of energy and this limits the mobile phone in providing the user a better service through its embedded sensors. Furthermore, consider applications that require extensive processing – image processing for video games, speech synthesis, natural language processing, augmented reality, wearable computing—all these demand high computational capacities thus restricting the developers in implementing applications for mobile phones. Considering the trends in mobile phone architecture and battery, it is unlikely that these problems will be solved in the future. This is, in fact, not merely a temporary technological deficiency but intrinsic to mobility [4], and a barrier that needs to be overcome in order to realize the full potential of mobile computing.

In recent years, this problem has been addressed by researchers though cloud computing. Cloud computing can be defined as the aggregation of computing as a utility and software as a service [5] where the applications are delivered as services over the Internet and the hardware and systems software in data centers provide those services [6]. Also called 'on demand computing', 'utility computing' or 'pay as you go computing', the concept behind cloud computing is to offload computation to remote resource providers. The key strengths of cloud computing can be described in terms of the services offered by cloud service providers: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) [7]. Extensive surveys on cloud computing such as [6,5,8,7,9,10] can be found in the literature, and here we focus on the potential of, and the challenges faced by mobile cloud computing.

The concept of offloading data and computation in cloud computing, is used to address the inherent problems in mobile

* Corresponding authors.
E-mail addresses: niro_ucsc@yahoo.com, tnfernando@students.latrobe.edu.au (N. Fernando), s.loke@latrobe.edu.au (S.W. Loke), w.rahayu@latrobe.edu.au (W. Rahayu).

0167-739X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2012.05.023

computing by using resource providers other than the mobile device itself to host the execution of mobile applications. Such an infrastructure where data storage and processing could happen outside the mobile device could be termed a 'mobile cloud'. By exploiting the computing and storage capabilities of the mobile cloud, computer intensive applications can be executed on low resource mobile devices.

Some of the key questions needing to be answered are: How does mobile cloud computing differ from cloud computing? What approaches have been made towards mobile cloud computing and how do they differ from each other? How can computation be offloaded and distributed to the cloud efficiently and in which ways does this differ from traditional distributed computing? What incentives can be used to persuade surrounding surrogate devices to participate in sharing the workload? How can context information be used in a beneficial way? How does mobility affect the performance of a mobile cloud?

The goal of this paper is to discuss in detail the current research that addresses these issues. We review the proposed solutions, and explore the upcoming research challenges in mobile cloud computing.

Organization of paper

The remainder of this paper is organized as follows: In Section 2, we present the motivation for mobile cloud computing, and discuss potential applications and scenarios. In Section 3, we briefly introduce cloud computing and mobile cloud computing, and identify three definitions of mobile cloud computing. Next, in Section 4 we propose a taxonomy of mobile cloud computing based on the key issues, and review how each issue has been tackled in related research. We provide a discussion on the current challenges in Section 5. Finally, conclusions and directions for future research are identified in Section 6.

2. Motivation: the need for a mobile cloud

The case for mobile cloud computing can be argued by considering the unique advantages of empowered mobile computing, and a wide range of potential mobile cloud applications have been recognized in the literature. These applications fall into different areas such as image processing, natural language processing, sharing GPS, sharing Internet access, sensor data applications, querying, crowd computing and multimedia search. However, as explained in [11], applications that involve distributed computation do have certain common characteristics, such as having data with easily detectable segment boundaries, and the time to recombine partial results into a complete result must also be small. An example is string matching/manipulation such as grep and word frequency counters. The different applications and scenarios presented in recent literature are described in detail below:

1. Image processing: In [11], the authors have experimented with running GOCR,[1] an optical character recognition (OCR) program on a collection of mobile devices. In a real life scenario, this would be useful in a case of a foreign traveler who takes an image of a street sign, performs OCR to extract the words, and translates words into a known language. A similar scenario is given in [12] where a foreign tourist Peter is visiting a museum in South Korea. He sees an interesting exhibit, but cannot understand the description since it is in Korean. He takes a picture of the text, and starts an OCR app on his phone. Unfortunately his phone lacks the resources to process the whole text. Although he could connect to a remote server via the Internet, that would mean he use roaming data which is too expensive. Instead, his device scans for nearby users/devices who are also interested in reading the description, and requests sharing their mobile resources for the task collaboratively. Those who are interested in this common processing task create an ad hoc network with Peter and together, their mobile cloud is able to extract the text, and then translate it to English. This can be applied to many situations in which a group is involved in an activity together. Another example is a group performing archaeological expeditions in a desert.
2. Natural language processing: As mentioned above, language translation is one possible application, and this is mentioned in [11] as a useful tool for foreign travelers to communicate with locals. Translation is a viable candidate since different sentences and paragraphs can be translated independently, and this is experimentally explored in [11] using Pangloss-Lite [13]. Text-to-speech is also mentioned in [11], where a mobile user may prefer having a file read to them, especially in case of the visually impaired.
3. Crowd computing: Video recordings from multiple mobile devices can be spliced to construct a single video that covers the entire event from different angles, and perspectives [14]. In [15], two scenarios of this nature are described in detail: 'Lost child' and 'Disaster relief'.
   The 'Lost child' scenario takes place at a parade in Manhattan. John, a five year old child who is attending the parade with his parents goes missing among all the people, and his parents only notice he is missing after some time. Fortunately, a police officer sends out an alert via text message to all mobile phones within a two mile radius, requesting them to upload all photographs they have taken in the parade during the past hour, to a server that only the police has access to. With John's parents, the police officer searches through these photographs via an app on his phone. After looking through some pictures, they are able to spot John in one of the images, which they identify to be taken at a nearby location. Soon, the relieved parents are reunited with their child.
   In the 'Disaster relief' scenario, a massive earthquake measuring 9.1 on the Richter scale has occurred in Northern California, resulting in much human loss, and infrastructure and property destruction. Disaster relief teams are facing an uphill task because of limited manpower, lack of transportation, and poor communication. Internet infrastructure has been destroyed. Previous maps on terrain and buildings are suddenly rendered obsolete, contributing to slow disaster relief. Data on Google Earth and Google Maps on this area is now useless since highways, bridges, landmarks and buildings have now all collapsed. To conduct efficient search and rescue operations, new data must be gained and a clear picture of the terrain and buildings state must be constructed. To do this, the relief teams use camera based GigaPan sensing.[2] Local citizens are asked to use their mobile phones to photograph disaster sites, and these are collected at a central server. The collected images are then sewn together to create a whole, panoramic image. The new face of the area emerges, and relief teams can now conduct their work with accurate maps and information on inaccessible areas.
4. Sharing GPS/Internet data: It is more efficient to share data among a group of mobile devices that are near each other, through local-area or peer-to-peer networks. It is not only cheaper, but also faster [14]. Rodriguez et al. [16] present a case study of a hiking party at Padjelanta National Park, which is a deserted land in the Arctic circle lacking power access

1 http://jocr.sourceforge.net/.

2 http://www.gigapan.com/.