ABSTRACT

Hosting services on mobile devices has been considered as the key solution for domains that have special requirement on portability, timeliness, and flexibility on service deployment. Typical examples include, among many others, military, music, healthcare, gaming, and data sharing. Despite the recent boom of mobile computing makes service deployment in mobile environment possible, significant challenges arise due to the limitations in existing mobile hardware/software capable of managing resource intensive applications. The situation gets worse when managing complex services that allow concurrent clients and requests. This paper addresses the issue related specifically to concurrency control improvement in mobile web servers to support the mobile deployment of complex services. The authors identify key factors that affect a system to respond a request, including request related factors, system resource related factors, and context. Based on this, the authors propose a dynamic heavy request classification model (DHRC) to estimate the heaviness of an incoming request using machine-learning methods. The heavy request will be detected, which requires relatively heavy system resources of the mobile server to generate a response. The authors design a dynamic request management strategy (DRMS), which reduces the number of discarded requests by adding heavy requests to a queue and processing them asynchronously. The proposed solution is implemented on Android-based mobile devices as an extension of I-Jetty web server. Experimental studies are conducted and the result indicates the effectiveness of our solution.

INTRODUCTION

Driven by the significant benefits brought by service computing, a large number of applications are accessed on the Web through web services. Web services have enjoyed a wide spectrum of application fields (Yu, 2008). This leads to the variety and the complexity of online services, which keep increasing over the time. As a result, the traditional platforms for deploying and hosting web services start facing their limitations and become insuf-
cient for those complex services that have special requirement on portability, timeliness, and flexibility. Enabled by the current mobile computing technologies, such services can benefit from deployment on a mobile device delivered in a mobile environment. For example, mass transit systems can use a mobile server to host a geo-location service that provides real-time location as well as scheduling and route information to the commuters. Services hosted on a mobile device also benefit a supply chain management system, where shipped products can be accurately tracked across the world. In military, soldiers in the battlefield can use smartphones with web services hosted on them to directly share information, communicate, and gather intelligence from other soldiers in vicinity, without depending on a third-party service, which introduces extra latency. Skilled people like doctors and nurses can also use a mobile web service in remote areas, in emergent and disastrous situations, to obtain and share patient information in a more proactive and timely way, compared to the communications through a third-party service.

Meanwhile, the spread of mobile devices around the world and the availability of dynamic mobile content have significantly pushed the transition of web service deployment platforms and made hosting complex services on mobile devices possible (Miller, 2008, McFaddin, 2003, Srirama, 2006, Pham, 2005). Recent innovations in wireless broadband technologies such as WiFi, WiMax, 3G, and 4G allow consumers to access and share increasing amount of web services and applications right on their mobile devices. Today, consumers are rapidly transitioning towards a post-PC world where a mobile device is capable of performing essential functions of a tradition personal computer (Mal-ladi, 2002). First developed and introduced by Nokia in 2007\(^1\), the concept of using mobile devices as web servers is still under research (Srirama, 2006). A mobile web server consists of a mobile device (e.g., an Android phone) with an application which functions as a web server such as Apache HTTP Server or Apache Jetty. The application then provides the mobile device the capability to host web sites, web services, and web applications.

Hosting web services on a mobile server will bring flexibility, convenience, and potential to the consumers. This will revolutionize the way we communicate, share data, and gather information. Mobile web servers face unlimited possibilities in the areas of enterprise mobile data access and sharing for lucrative markets such as military, music and health care, to name only a few. Engineering mobile devices around the world as mobile web servers would change the Internet landscape. It would fundamentally alter the way we access web services such as maps with geo-location information, documents, medical records, news, and so on. However, it also faces great performance challenges. Mobile devices are still undergoing major changes and thus, still lack resource capabilities. In addition, battery life is limited in a mobile server. Besides less processing power, less memory, and limited battery life, the concurrency control management on a mobile server further complicates the challenges. For example, because of its limited thread pool and memory capacity, if the number of concurrent clients accessing a web service hosted on a mobile server increases, the device is unable to handle these multiple requests. As a result, the mobile web server discards the client requests. With better concurrency control, it is possible to decrease the number of discarded requests and thus increase the reliability and availability of mobile servers.

Existing research related to complex services in a mobile environment focuses on the enabling framework and performance improvement (Mizouni, 2011, Hassan, 2010, Gehlen, 2005). Mizouni et al proposed an architecture that deploys web services on mobile platform (Mizouni, 2011). They also identified and discussed several factors that might affect the performance of services, such as response time, availability, throughput, and scalability. Hassan et al. proposed a partitioning framework that provides a high level design about how to
Modern Diffusion of Products with Complex Network Models
www.igi-global.com/chapter/modern-diffusion-products-complex-network/53279?camid=4v1a