Abstract—Cloud computing is an emerging concept combining many fields of computing. The foundation of cloud computing is the delivery of services, software and processing capacity over the Internet, reducing cost, increasing storage, automating systems, decoupling of service delivery from underlying technology, and providing flexibility and mobility of information. However, the actual realization of these benefits is far from being achieved for mobile applications and open many new research questions. In order to better understand how to facilitate the building of mobile cloud-based applications, we have surveyed existing work in mobile computing through the prism of cloud computing principles. We give a definition of mobile cloud computing and provide an overview of the results from this review, in particular, models of mobile cloud applications. We also highlight research challenges in the area of mobile cloud computing. We conclude with recommendations for how this better understanding of mobile cloud computing can help building more powerful mobile applications.

Index Terms—cloud computing; mobile computing; remote execution; distributed systems; automatic offloading.

1. INTRODUCTION

Mobile devices allow users to run powerful applications that take advantage of the growing availability of built-in sensing and better data exchange capabilities of mobile devices. As a result, mobile applications seamlessly integrate with real time data streams and Web 2.0 applications, such as open collaboration, social networking and mobile commerce [1], [2]. The mobile execution platform is being used for more and more tasks, e.g., for playing games; capturing, editing, annotating and uploading video; handling finances; managing personal health, micro payments, ticket purchase, interacting with ubiquitous computing infrastructures. Even mobile device hardware and mobile networks continue to evolve and to improve, mobile devices will always be resource-poor, less secure, with unstable connectivity, and with less energy since they are powered by battery. Resource poverty is major obstacle for many applications [3]. The cloud computing paradigm is often confused about its capabilities, described as general term that includes almost any kind of outsourcing of hosting and computing resources. According to NIST [5] cloud computing is a model for enabling convenient, on-demand network access to computing resources that can be rapidly provisioned.
and released with minimal management effort. The combination of cloud computing, wireless communication infrastructure, portable computing devices, location-based services, mobile Web, etc., has laid the foundation for a novel computing model, called mobile cloud computing, which allows users an online access to unlimited computing power and storage space. Taking the cloud computing features in the mobile domain, we define: "Mobile cloud computing is a model for transparent elastic augmentation of mobile device capabilities via ubiquitous wireless access to cloud storage and computing resources, with context-aware dynamic adjusting of offloading in respect to change in operating conditions, while preserving available sensing and interactivity capabilities of mobile devices." In our survey we consider smart mobile devices that include devices continuously connected to the Internet, with handheld form and a large, high quality graphics display and significant but limited computing power. To make this vision a reality beyond simple services, mobile cloud computing has many hurdles to overcome. Existing cloud computing tools tackle only specific problems such as parallelized processing on massive data volumes [6], flexible virtual machine (VM) management [7] or large data storage [8]. However, these tools provide little support for mobile clouds. The full potential of mobile cloud applications can only be unleashed, if computation and storage is offloaded into the cloud, but without hurting user interactivity, introducing latency or limiting application possibilities. The applications should benefit from the rich built-in sensors which open new doorways to more smart mobile application. The research challenges include how to abstract the complex heterogeneous underlying technology, how to model all the different parameters that influence the performance and interactivity of the application, how to achieve optimal adaptation under different constraints, how to integrate computation and storage with the cloud while preserving privacy and security. In this paper, we provide a systematic and comparative description of mobile application models that go along with the cloud computing ideas. The survey results show that current related projects cover only different subsets of the desired mobile cloud characteristics.

2. CURRENT STATUS IN MOBILE APPLICATIONS

Several researchers, [9]–[10], have identified the fundamental challenges in mobile computing. Mobile computing environments are characterized by severe resources constraints and frequent changes in operating conditions. Mobile devices inherently have and will continue to have limited resources as processing power, memory capacity, display size, and input forms. These have been the forming factors of existing mobile application approaches.

A. Offline Applications:
Most of the applications available for modern mobile devices fall into this category. They act as fat client that processes the presentation and business logic layer locally on mobile devices with data downloaded from backend systems. A fat client is a networked application with most resources available locally, rather than distributed over a network as is the case with a thin client. Offline applications, also often called native applications, offer:

- Good integration with device functionality and access to its features
- Performance optimized for specific hardware and multitasking
On the other hand, the native applications have many disadvantages:
- No portability to other platforms
- Complex code
- Increased time to market.

B. Online Applications
An online application assumes that the connection between mobile devices and backend systems is available most of the time. Mobile has the potential to overcome some of the disadvantages of offline applications because they are:
- multi-platform
- directly accessible from anywhere

However, mobile Web applications have disadvantages:
- no access to device’s features such as camera or motion
- Keeping communication session a over longer period of time.

3. NOVEL APPLICATION MODELS FOR MOBILE CLOUD COMPUTING

Mobile cloud computing could be described as the availability of cloud computing services in a mobile ecosystem. To leverage the full potential of mobile cloud computing we need to consider the capabilities and constraints of existing architectures.

A. Augmented Execution
Augmented execution refers to a technique used to overcome the limitations of Smartphone’s in terms of computation, memory and battery. Chun and Maniatis [12] propose architecture that addresses these challenges via seamlessly offloading execution from the phone to computational infrastructure (cloud) where cloned replica of the smart phone’s software is running.

Fig. 1. Clone Cloud categories for augmented execution (adapted from [12]).

Fig. 1 shows categorization of possible augmented execution for mobile phones: (1) primary functionality outsourcing – more like a client-server application, (2) background augmentation-Good for independent separate process that can run in background like a virus scanning, (3) mainline-in-between primary and background augmentation, (4) hardware - the replica runs on more powerful emulated VM, and (5) multiplicity – helpful for parallel executions. In this architecture, a mobile user exploits VMs to rapidly instantiate customized service software on a nearby cloudlet and uses the service over WLAN. The mobile client acts as thin client, with all significant computation occurring in a nearby cloudlet. This approach relies on technique called dynamic VM synthesis (cf. Fig. 2). A mobile device delivers small VMs overlay to the cloudlet infrastructure that already owns the base VM from which this overlay was derived. The infrastructure applies the overlay to the base to derive the VM which starts executing in the precise state in which it was suspended.

Fig. 2. Dynamic virtual machine synthesis timeline (adapted from [3]).
Figure 3 shows the main concept. After the connection is established, the client requests an application. Then the optimal deployment for the application is computed. Based on that decision, an application description and a list of services to be fetched are sent to the client’s Renderer. The Renderer generates corresponding UI according to the description. Furthermore, for the services that are decided to run on the client side the corresponding service bundles are fetched (on Fig. 3 service S1). Zhang et al. [14], [15] develop a reference framework for partitioning a single application into elastic components with dynamic configuration of execution. The components, called web lets, are platform independent and can be executed transparently on different computing infrastructures including mobile devices or IaaS (Infrastructure as a Service) cloud providers such as Amazon EC2 and S3 [7].

C. Ad-hoc Mobile Clouds

An ad-hoc computing cloud represents a group of mobile devices that serve as a cloud computing provider by exposing their computing resources to other mobile devices. This type of mobile cloud computing becomes more interesting in situations with no or weak connections to the Internet and large cloud providers. Offloading to nearby mobile devices save monetary cost, because data charging is avoided, especially favored in roaming situations. Moreover, it allows creating computing communities in which users can collaboratively execute shared tasks.

4. TOPICS FOR EXPLORATION

To enable the new mobile cloud application model, many challenges exist in different areas, including data replication, consistency, transaction management, cache management, optimal cost-effective execution in heterogeneous computing environments, elastic module lifecycle management and their communication and state synchronization.

A. Programming Abstraction

Development on mobile clouds should be simple and intuitive; however, at the same time the developer should be able to control behavior and location of his application. The developed software modules should be optimized for running on different mobile device hardware. The programming tools should allow a holistic application development for the mobile client, middleware, and cloud, with dynamic shifting of the computation and the storage between them. For example, Zhang et al. [14] have implemented SDK, which is used to develop the basic interfaces of application modules and manage their lifecycle. Different types of applications can be build or adapt without architectural changes.

B. Cost Model

The cost model takes inputs from both device and cloud, and runs optimization algorithms to decide execution configuration of applications (cf. Fig. 5). Zhang et al. [14] use Naïve Bayesian
Learning classifiers to find the optimal execution configuration from all possible configurations with given CPU, memory and network consumption, user preferences, and log data from the application. Every bundle or module composing the application has memory consumption, generated input and output traffic, and code size. Application's distribution between the server and phone is then optimized. The server is assumed to have infinite resources and the client has several resource constraints. The objective function tries to minimize the interactions between the phone and the server, while taking into account the overhead of acquiring and installing the necessary bundles.

C. Adaptation
Adaptation is a key to mobility. Mobile cloud applications, running on relatively limited resources of mobile devices with unreliable networks and changing circumstances, must react on and Fig. 5. Cost model of elastic mobile cloud applications (extended from [14])

dynamically reassign the responsibilities of mobile client and cloud, i.e. must be adaptive. In [9], the range of strategies for application and system adaptation is identified.

D. Cloud Integration
Cloud storage is the most obvious use of cloud computing in mobile applications. Data availability is important for completing tasks in a currently running process. Caching can be used, but the use of cache on distributed databases requires additional efforts such as cache validation coherency. However, many issues exist. First, cloud processing is more complex to implement than cloud storage because it involves both data management and synchronization. Second, to achieve truly computation augmentation of the limited mobile device, the performance of application execution needs to be monitored.

E. Trust, Security and Privacy
A never ending issue will always be security in cloud computing related to multi-tenancy, concurrency, scale and distribution. First, direct concerns arise from aspects such as lacking control over data and code distribution in distributed infrastructures, potential data loss.

5. CONCLUSION
In this paper, we have covered several representative mobile cloud approaches. Much other related work exist, but the purpose of this paper is to give an overview of the wide spectrum of mobile cloud computing possibilities. None of the existing approaches meets completely the requirements of mobile clouds. The former type is using capabilities of mobile devices, but the integration with the cloud is poor. Therefore, we believe that the full potential of mobile cloud applications lies in between these two extremes, while dynamically shifting the responsibilities between mobile device and cloud. Several researchers have shown how to achieve that by, e.g., replicating whole device software image or offloading parts of the application.

To simplify the development a convenient, but effective, programming abstraction is required. Mobile cloud computing will be a source of challenging research problems in information and communication technology for many years to come.
REFERENCES:


