A Survey of Mobile Cloud Computing

Xiaopeng Fan[1], Jiannong Cao[2], and Haixia Mao[3]

(1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences;
2. Department of Computing, Hong Kong Polytechnic University;
3. Department of Land Surveying and Geo–informatics, Hong Kong Polytechnic University)

Abstract: Mobile Cloud Computing (MCC) is emerging as one of the most important branches of cloud computing. In this paper, MCC is defined as cloud computing extended by mobility, and a new ad–hoc infrastructure based on mobile devices. It provides mobile users with data storage and processing services on a cloud computing platform. Because mobile cloud computing is still in its infancy, we aim to clarify confusion that has arisen from different views. Existing works are reviewed, and an overview of recent advances in mobile cloud computing is provided. We investigate representative infrastructures of mobile cloud computing and analyze key components. Moreover, emerging MCC models and services are discussed, and challenging issues are identified that will need to be addressed in future work.

Keywords: mobile cloud computing; cloud computing

1 Introduction

Since the first computer was invented in 1946, people have dreamt of using computing resources as a utility—like water, electricity, gas, or the telephone. There has been a long–held perception that computing will one day be the fifth utility. Cloud computing [1]–[6] might be the most promising way of realizing this dream.

In recent research on cloud computing, one of the most important issues has been how to sweep up data and programs from desktop PCs [1] and install them in the compute cloud. Cloud computing refers to applications delivered as services over the Internet and also the hardware and systems software in the data centers that provide those services [2]. However,

whether it is called cloud computing or on–demand computing, Software as a Service (SaaS) or Internet as platform, the overriding idea is a shift in the geography of computation.

In general, cloud computing is web–based processing [7], whereby shared resources, software, and information are provided on demand to computers, smartphones [8], and other similar devices. Cloud computing is also a new style of computing in which dynamically scalable resources are provided as virtualized services [9]. This allows service providers and users to adjust their computing capacity depending on how much is needed at a given time or for a given task.

Mobile computing [10] means using portable devices to run stand–alone applications and/or accessing remote applications via wireless networks.

Increasing the number of mobile applications demands greater resources and improved interactivity for better experience. Resources in cloud platforms such as Amazon EC2 [11], Microsoft Azure [12] and Google AppEngine [13] can remedy the lack of

resources in mobile devices. But what does mobile cloud computing really mean? Different people hold different views, and there are several existing definitions of mobile cloud computing.

In [14], mobile cloud computing was first referred to as an infrastructure where data storage and processing could happen outside the mobile device, enabling a new class of applications—especially context–aware mobile social networks. Mobile cloud applications move computing power and data storage away from mobile phones and into the cloud. This brings mobile applications and computing not just to smartphone users but to a broad range of mobile subscribers.

In terms of services [7], mobile cloud computing extends processing and storage beyond devices and seamlessly integrates with other services and sensor data to open up new classes of application.

In [9], mobile cloud computing is defined as an extension of cloud computing in which foundation hardware consists at least partly of

mobile devices. This definition recognizes the opportunity to harness collective sensing, storage, and computational capabilities of multiple networked wireless devices to create a distributed infrastructure that supports a wealth of new applications.

Therefore, our goal in this survey is to eliminate confusion by clarifying the terms and concepts, predicting state−of−the−art mobile cloud computing, discussing future trends, and identifying technical and non−technical obstacles and opportunities.

2 Mobile Cloud Computing

Definitions of mobile cloud computing can be divided into two classes. The first refers to carrying out data storage and processing outside mobile devices [15]. Mobile devices are simply terminals in cloud computing, only intended to provide a more convenient way of accessing services in the cloud. The benefit of this is that storage and computing limitations of mobile devices are avoided, and a new level of security is provided by centralizing maintenance of security−critical software.

The second class of definitions refers to computing where data storage and processing are also carried out on mobile devices [9]. The infrastructure of the cloud is different from that of data−center "pay−as−you−use" cloud computing. Each node is owned by a different user and is likely to be mobile. Using mobile hardware for cloud computing has advantages over using traditional hardware. These advantages include computational access to multimedia and sensor data without the need for large network transfers, more efficient access to data stored on other mobile devices, and distributed ownership and maintenance of hardware.

In this paper, mobile cloud computing is defined as cloud computing extended by mobility and a new ad−hoc infrastructure based on mobile devices. It provides mobile users with data storage and processing services on a cloud computing

platform. Using this definition, we clarify two differences between mobile computing and cloud computing.

First, there is an overlap between the two computing paradigms. Cloud computing aims to provide services to users without letting them know where these services are hosted or how they are delivered. Mobile computing aims to support mobility so that users can access resources with wireless technology.

Second, it is possible to build up clouds within mobile devices for data storage and processing. Although mobile devices nowadays are weak in terms of energy, power, storage, and communication, this situation will be improved dramatically with the development of technology. In [9], Hyrax shows that a platform derived from Hadoop [16] can support cloud computing on Android [17] smartphones. It demonstrates the possibility that a cloud composed of mobile devices can provide basic functions and services of cloud computing.

Why do we need mobile cloud computing? There are at least three possible explanations.

First, mobility support is very important in allowing users access to cloud services. Service availability is considered the first obstacle [9] for cloud computing. Mobile cloud computing exploits information about a user's location, context, and requested services and significantly evolves the heterogeneous access management schemes for traditional heterogeneous access scenarios. This is called "intelligent access" [15]. It can noticeably improve user experience.

Second, the capacity of mobile devices has increased dramatically with the growth of mobile users, especially with the prevalence of smartphone users. Each mobile device has storage, computing, sensing, and power resources. By using these resources, applications or services can be easily accessed by users inside and outside the mobile network.

Third, there are still some obstacles to implementing cloud computing, and mobile computing technology can help

overcome these obstacles. For example, long WAN latencies are a fundamental obstacle when a mobile device executes a resource−intensive application on a distant high−performance server or cluster. A solution is to instantiate customized service software on a nearby cloudlet [18] and then to use that service over a wireless LAN.

3 Existing Works on Mobile Cloud Computing

According to our definition of mobile cloud computing, there are two key issues: How to build up a mobile cloud using mobile devices and how to access services provided by clouds in a mobile way. Thus, existing work on mobile cloud computing can be classified as that based on the cloud platform, or that based on access schemes. For the former, we investigate existing work on mobile cloud platforms consisting of mobile devices. The Hyrax [9] platform is derived from Hadoop [16] and supports cloud computing on Android [17] smartphones. To the best of our knowledge, it is the only platform comprised of mobile devices. For the latter, we investigate existing access schemes in mobile cloud computing. Mobile cloud computing is still in its infancy, and we investigate Virtual Machine (VM)−based cloudlets [18] and intelligent access schemes [15].

3.1 Hyrax

The goal of Hyrax is to develop a mobile cloud infrastructure that enables smartphone applications with distributed data and computation. Hyrax allows applications to conveniently use data and execute computing jobs on smartphone networks and heterogeneous networks of phones and servers. Research has been focused primarily on implementation and evaluation of mobile cloud computing infrastructure based on MapReduce [19]. MapReduce, Hadoop, and Android should be introduced before describing Hyrax infrastructure.

MapReduce is considered a programming model for processing and
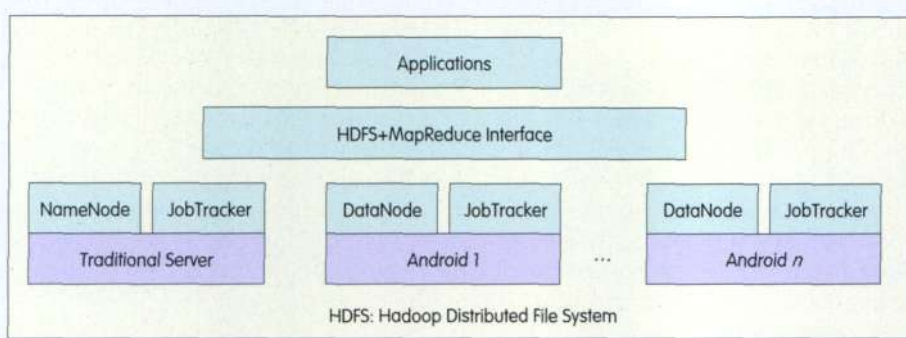
▲Figure 1. Hyrax hardware and software layers.

generating large data sets. A user specifies a map function that processes a key/value pair in order to generate a set of intermediate key/value pairs. It also has a reduce function that merges all intermediate values associated with the same intermediate key. The MapReduce runtime system splits input data, schedules map and reduce tasks, and transfers input and output data to machines running the tasks. There is a master to manage jobs, assign tasks to slave machines, and provide locations of intermediate values to reduce tasks.

Apache Hadoop is an open source implementation of MapReduce that is used by many organizations for large−scale data processing. It is designed to operate data stored in a distributed file system, Hadoop Distributed File System (HDFS). There are four types of processes in Hadoop instances: NameNode, JobTracker, DataNode, and TaskTracker. There is one NameNode and one JobTracker in a Hadoop cluster. NameNode schedules jobs and coordinates sub−tasks among TaskTrackers. A DataNode instance and a TaskTracker instance both run on each worker machine. DataNode stores and provides access to data blocks, and TaskTracker executes tasks assigned to it by JobTracker. Clients access files by first requesting block locations from NameNode and then requesting blocks directly from these locations. The hardware and software layers of Hyrax are described in Fig. 1.

Hyrax ports Hadoop to the Android platform. In Hyrax, each machine runs one instance of NameNode and one instance of JobTracker. DataNode and TaskTracker are run on each phone in separate Android service processes within the same application. Android applications may consist of multiple processes, some of which run as background services. Since DataNode and TaskTracker are run as Android services, they can run in the background of other applications.

A thread is also spawned to record information about the system load—including power level and CPU, memory, network, and disk I/O statistics—into the local file system. Within the application, a server is run to allow external scripts to control data uploading, kill the program, and check the program status. Fig. 2 illustrates the data interactions between all the software components on each phone.

Hyrax is implemented in a real testbed consisting of a cluster of 10 Android G1 (HTC Dream) phones and 5 HTC Magic phones, each running Android 1.5 Cupcake. Since Android does not support peer−to−peer networking yet, the phones communicate with each other on an isolated 802.11g network via a Linksys WRT54G wireless router with no firmware modifications. NameNode and JobTracker processes are run on a desktop machine connected behind this router via Ethernet. The phones are connected via USB to a controller that executes experiment scripts. These scripts are used to install Hyrax, initialize the cluster, run benchmarks, and collect and post process data.

To determine the advantages and drawbacks of Hyrax, an application was developed on it. The Hyrax multimedia search and sharing application, HyraxTube, allows users to browse videos and images stored on a network of phones and search by time, location, and quality. Quality ratings based on sensor data are generated by periodically executing a MapReduce job. Requests are serviced by reading results generated by the MapReduce job from HDFS. The client interface is implemented as a web application so that it can be used on mobile devices and desktop machines.

3.2 VM−Based Cloudlets

Satyanarayanan et al. [18] present a new vision for mobile cloud computing. They foresee a new world in which mobile computing seamlessly augments users' cognitive abilities via
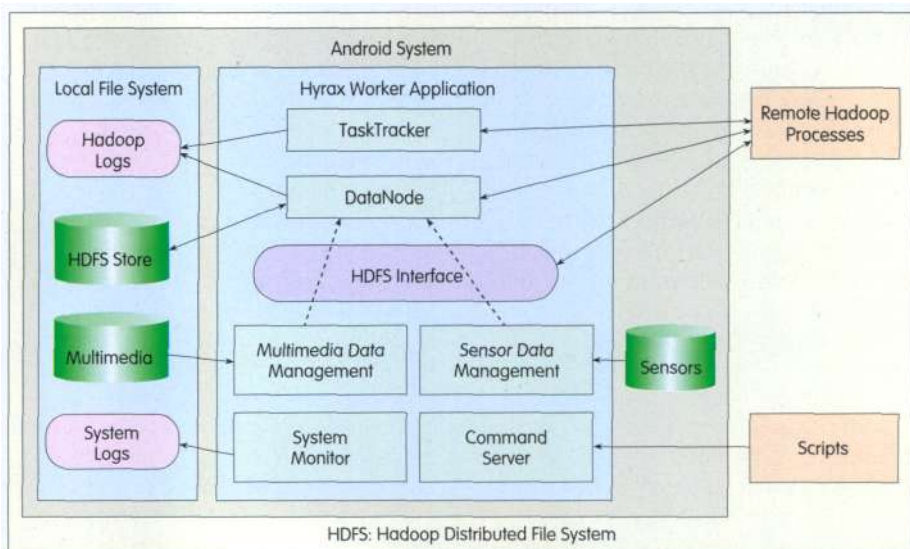
▲Figure 2. Hyrax worker application component interaction diagram.

intensive capabilities such as speech recognition, natural language processing, computer vision and graphics, machine learning, augmented reality, planning, and decision making.

The means of achieving this vision is to let a mobile user exploit VM technology to rapidly instantiate customized service software on a nearby cloudlet. That service is then used over a wireless LAN. A cloudlet is a trusted, resource-rich computer or cluster of computers that is well-connected to the Internet and available for use by nearby mobile devices. Mobile devices act as a thin client in a 3-layer cloud computing architecture.

This solution mainly addresses the resource poverty of mobile devices, and long WAN latency is a fundamental obstacle. The predominant view is that WAN latency can hurt usability by degrading the crispness of system response, and it is unlikely to be improved. Bandwidth-induced delays also hurt user experience. The transmission of large data items should occur within a tight user-machine interaction loop.

Real-time interactive response can be achieved through low-latency, one-hop, high-bandwidth wireless access to the cloudlet. Fig. 3 illustrates cloudlets and how they work. A cloudlet resembles a data center in a box; it is self-managing and requires little more than power, Internet connectivity, and access control for setup. Such simplicity of management is similar to an appliance model of computing resources and makes cloudlets simple to deploy in a business premises. Internally, a cloudlet resembles a cluster of multicore computers, with gigabit internal connectivity and high-bandwidth wireless LAN.

Cloudlet infrastructure is deployed in much the same way as Wireless Fidelity (Wi-Fi) access points. The key challenge is to simplify cloudlet management, and a proposed solution is transient customization of cloudlet infrastructure using hardware VM technology. Pre-use customization and post-use cleanup ensure that cloudlet
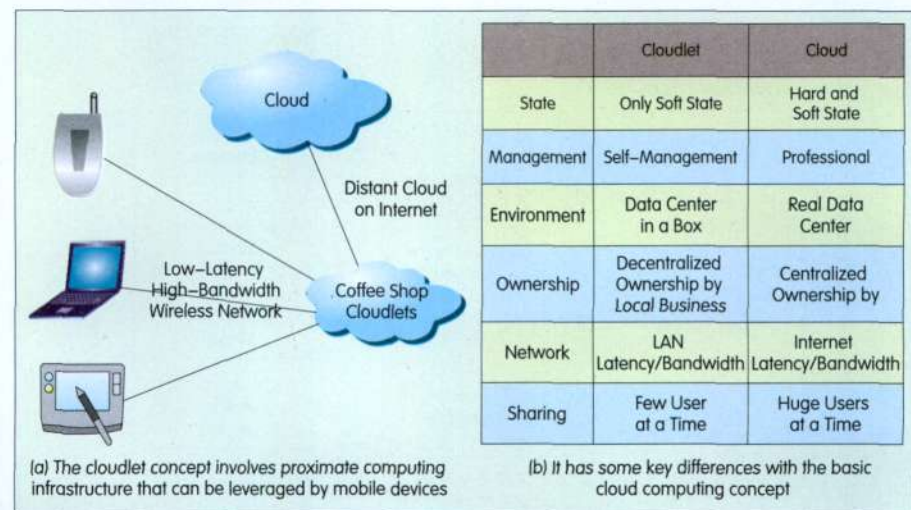
▲Figure. 3. The cloudlet concept.

infrastructure is restored to pristine state after each use, without manual intervention. A VM cleanly encapsulates and separates the transient guest software environment from the cloudlet infrastructure's permanent software environment. Therefore, a VM-based approach is less brittle than alternatives such as process migration or software virtualization. It is also less restrictive and more general than language based virtualization approaches that require applications to be written in a language such as Java or C#.

3.3 Intelligent Access Schemes

Mobile cloud computing relies on permanent connectivity. Intelligent access schemes [15] try to meet this requirement by specific information such as the user's location, context, and request services provided by the so-called mobile cloud controller.

Efficient network access management across different Radio Access Technologies (RATs)—Heterogeneous Access Management (HAM)—is necessary. Radio resources are limited due to physical, technology-specific, and regulatory constraints. Context provided by terminals, network nodes, or sensors deployed in the user's environment are used to significantly reduce wasting scarce radio resources and to efficiently manage wireless access across heterogeneous RATs.

The proposed solution is designed

for a heterogeneous access scenario with a wide range of different radio access technologies and context information. The proposed Intelligent Radio Network Access (IRNA) concept takes into account the characteristics and status of each RAT while simultaneously trying to accommodate users who face different environmental conditions in order to provide the best possible end-to-end performance. IRNA is shown in Fig. 4.

The proposed Context Management Architecture (CMA) acquires, processes, manages, and delivers context information. A Context Quality Enabler (CQE) is incorporated into the architecture. The CQE controls the provision of context information according to the requirements of the Mobile Cloud Controller. A Context-Aware Radio Network Simulator (CORAS) is proposed to model context availability, accuracy, and delay, and enable an evaluation to be made on the impact of context relevance, confidence, and quality on simulation results.

4 Conclusions

Wireless and mobile computing technologies provide more possibilities for accessing services conveniently. Mobile devices will be improved in terms of power, CPU, and storage. Mobile cloud computing has emerged as a new paradigm and extension of
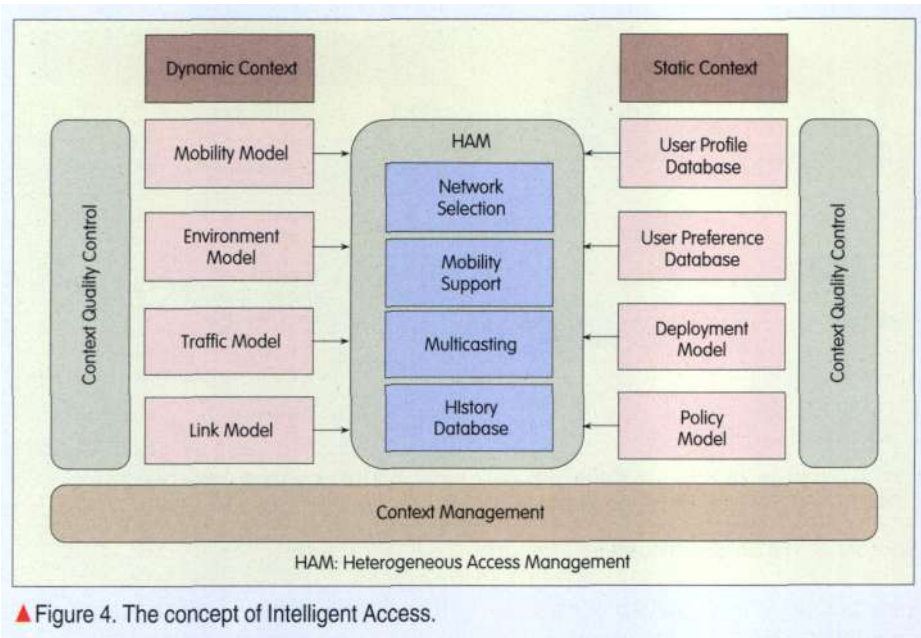
▲ Figure 4. The concept of Intelligent Access.

cloud computing.

In this survey, terminologies and concepts are clarified, and a definition of mobile cloud computing is provided based on an understanding of underlying technologies and applications. Existing work has also been surveyed and two classes of architecture framework are described for mobile cloud computing.

For future research, there are still some challenging issues related to the models of mobile cloud computing and services. They are:

(1) Mobile devices are constrained in storage and processing capacity. How can efficient use be made of the limited resources for cloud computing?

(2) There are several operating systems for mobile devices, especially mobile smartphones. These include Android, Symbian, iOS [20], Chrome, and MeeGo [21]. Is it possible to provide a general access platform for mobile cloud computing on top of these various OS platforms?

(3) From Hyrax and Cloudlet, mobile devices act not only as clients but also as components of clouds. What is the impact of mobility of mobile devices on performance of the cloud? What are the benefits and disadvantages?

(4) Users in mobile clouds are still exposed to security threats both inside and outside the cloud. What is the best solution for such a mobile environment?

Demand and applications of mobile cloud computing will grow rapidly, and researchers and developers should pay more attention to this area of cloud computing research and development.

References
[1] B. Hayes, "Cloud computing," *Communications of the ACM*, vol. 51, no. 7, pp. 9–11, Jul. 2008.
[2] M. Armbrust et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no.4, pp. 50–58, Apr. 2010.
[3] R. Buyya et al., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–606, Jun. 2009.
[4] "Cloud Computing," Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Cloud_computing
[5] Expert Group Report, "The Future of Cloud Computing. Opportunities for European Cloud Computing Beyond 2010," 2010. [Online]. Available: http://cordis.europa.eu/fp7/ict/ssai/docs/cloud–report–final.pdf
[6] M. Armbrust, A. Fox, R. Griffith et al. (2009, Feb.) "Above the clouds: A Berkeley view of cloud computing." EECS Dept. University of California, Berkeley. Tech. Rep. UCB/EECS–2009–28. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS–2009–28.html
[7] D. Kovachev, D. Renzel, R. Klamma, and Y. Cao, "Mobile community cloud computing: emerges and evolves," *Proc. 1st Intl. Workshop on Mobile Cloud Computing (in conjunction with) 11th Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010, pp. 393–395.
[8] "Smartphone," Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Smartphone/
[9] E. Marinelli, "Hyrax: cloud computing on mobile devices using MapReduce," Master thesis, Carnegie Mellon University, 2009.
[10] Mahadev Satyanarayanan, "Mobile computing: The next decade," *Proc. 11th Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010.
[11] "Amazon elastic compute cloud (EC2)." AWS. [Online]. Available: http://www.amazon.com/ec2/
[12] Microsoft azure. [Online]. Available: http://www.microsoft.com/azure/
[13] Google app engine. [Online]. Available: http://appengine.google.com/
[14] "The 1st ACM workshop on mobile cloud computing & services: Social networks and beyond," MCS 2010. [Online]. Available: http://www.cs.colorado.edu/~rhan/MCS2010.html
[15] A. Klein, C. Mannweiler, J. Schneider, and H. Schotten, "Access schemes for mobile cloud computing," *Proc. 11st Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010, pp. 387–392.
[16] Apache Hadoop. [Online]. Available: http://hadoop.apache.org/
[17] Android. [Online]. Available: http://www.android.com/
[18] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, "The case for VM–based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp.14–23, Oct. 2009.
[19] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Proc. 6th Symp. on Operating System Design and Implementation (OSDI'0 04)*, San Fransisco, CA, pp.137–150.
[20] Apple IOS. [Online]. Available: http://www.apple.com/iphone/ios4/
[21] MeeGo. [Online]. Available: http://meego.com/

Biographies

Xiaopeng Fan (xp.fan@siat.ac.cn) received B.E. and M.E. degrees in computer science from Xidian University, China. He received his Ph.D. in computer science in 2010 from Hong Kong Polytechnic University. He is currently an assistant researcher at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include cloud computing, wireless communication, and software engineering. His recent research has been focused on data center and mobile data management. He has reviewed for several international journals and conference proceedings.

Jiannong Cao (csjcao@comp.polyu.edu.hk) received his M.Sc. and Ph.D. degrees in computer science from Washington State University, Pullman. He is a professor and acting head of the Department of Computing at Hong Kong Polytechnic University. He is also the director of the department's Internet and Mobile Computing Lab. Before joining Hong Kong Polytechnic University, he taught in the faculty of computer science at James Cook University, the University of Adelaide, and City University, Hong Kong. Dr. Cao's research interests include computer networks, mobile and pervasive computing, parallel and distributed computing, fault tolerance, and middleware. He has co–authored a book on mobile computing, co–edited 9 other books, and published over 300 papers. Dr. Cao is a senior member of the China Computer Federation, a senior member of the IEEE's Computer Society and Communication Society, and a member of the ACM. He is also a member of the IEEE Technical Committee on Distributed Processing, IEEE Technical Committee on Parallel Processing, and IEEE Technical Committee on Fault–Tolerant Computing.

Haixia Mao (lsjessie@polyu.edu.hk) received B.E. and M.E. degrees in remote sensing and geo–informatics at Wuhan University, China. She received her Ph.D. degree in remote sensing and geo–informatics from Hong Kong Polytechnic University in 2010. She is currently a research associate in Hong Kong Polytechnic University. Her research interests include remote sensed image processing and mass geo–spatial data mining in geo–informatics.