Challenges in the Migration to 4G Mobile Systems

Suk Yu Hui and Kai Hau Yeung, City University of Hong Kong

ABSTRACT

With the rapid development of wireless communication networks, it is expected that fourth-generation mobile systems will be launched within decades. 4G mobile systems focus on seamlessly integrating the existing wireless technologies including GSM, wireless LAN, and Bluetooth. This contrasts with 3G, which merely focuses on developing new standards and hardware. 4G systems will support comprehensive and personalized services, providing stable system performance and quality service. However, migrating current systems to 4G presents enormous challenges. In this article these challenges are discussed under the headings of mobile station, system, and service aspects. Proposed solutions to the research problems in each aspect will also be examined.

INTRODUCTION

Second-generation (2G) mobile systems were very successful in the previous decade. Their success prompted the development of third-generation (3G) mobile systems. While 2G systems such as GSM, IS-95, and cdmaOne were designed to carry speech and low-bit-rate data, 3G systems were designed to provide higher-data-rate services. During the evolution from 2G to 3G, a range of wireless systems, including GPRS, IMT-2000, Bluetooth, WLAN, and HiperLAN, have been developed. All these systems were designed independently, targeting different service types, data rates, and users. As these systems all have their own merits and shortcomings, there is no single system that is good enough to replace all the other technologies. Instead of putting efforts into developing new radio interfaces and technologies for 4G systems, which some researchers are doing, we believe establishing 4G systems that integrate existing and newly developed wireless systems is a more feasible option.

Researchers are currently developing frameworks for future 4G networks. Different research programs, such as Mobile VCE, MIRAI, and DoCoMo, have their own visions on 4G features and implementations. Some key features (mainly from the users' point of view) of 4G networks are stated as follows:

- High usability: anytime, anywhere, and with any technology
- Support for multimedia services at low transmission cost
- Personalization
- Integrated services

First, 4G networks are all-IP based heterogeneous networks that allow users to use any system at any time and anywhere. Users carrying an integrated terminal can use a wide range of applications provided by multiple wireless networks. Second, 4G systems provide not only telecommunications services, but also data and multimedia services. To support multimedia services, high-data-rate services with good system reliability will be provided. At the same time, a low per-bit transmission cost will be maintained. Third, personalized service will be provided by this new-generation network. It is expected that when 4G services are launched, users in widely different locations, occupations, and economic classes will use the services. In order to meet the demands of these diverse users, service providers should design personal and customized services for them. Finally, 4G systems also provide facilities for integrated services. Users can use multiple services from any service provider at the same time. Just imagine a 4G mobile user, Mary, who is looking for information on movies shown in nearby cinemas. Her mobile may simultaneously connect to different wireless systems. These wireless systems may include a Global Positioning System (GPS) (for tracking her current location), a wireless LAN (for receiving previews of the movies in nearby cinemas), and a code-division multiple access (CDMA) (for making a telephone call to one of the cinemas). In this example, Mary is actually using multiple wireless services that differ in quality of service (QoS) levels, security policies, device settings, charging methods, and applications. It will be a significant revolution if such highly integrated services are made possible in 4G mobile applications.

To migrate current systems to 4G with the features mentioned above, we have to face a number of challenges. In this article these challenges are highlighted and grouped into various

 0163-6804/03/$17.00

	Key challenges	Proposed solutions
Mobile station		
Multimode user terminals	To design a single user terminal that can operate in different wireless networks, and overcome the design problems such as limitations in device size, cost, power consumption, and backward compatibilities to systems.	A software radio approach can be used: the user terminal adapts itself to the wireless interfaces of the networks [1].
Wireless system discovery	To discover available wireless systems by processing the signals sent from different wireless systems (with different access protocols and incompatible with each other).	User- or system-initiated discoveries, with automatic download of software modules for different wireless systems [2].
Wireless system selection	Every wireless system has its unique characteristic and role. The proliferation of wireless technologies complicates the selection of the most suitable technology for a particular service at a particular time and place.	The wireless system can be selected according to the best possible fit of user QoS requirements, available network resources, or user preferences [3, 4].
System		
Terminal mobility	To locate and update the locations of the terminals in various systems. Also, to perform horizontal and vertical handoff as required with minimum handover latency and packet loss.	Signaling schemes and fast handoff mechanisms are proposed in [5].
Network infrastructure and QoS support	To integrate the existing non-IP-based and IP-based systems, and to provide QoS guarantee for end-to-end services that involves different systems.	A clear and comprehensive QoS scheme for UMTS system has been proposed [6]. This scheme also supports interworking with other common QoS technologies.
Security	The heterogeneity of wireless networks complicates the security issue. Dynamic reconfigurable, adaptive, and lightweight security mechanisms should be developed.	Modifications in existing security schemes may be applicable to heterogeneous systems. Security handoff support for application sessions is also proposed [4].
Fault tolerance and survivability	To minimize the failures and their potential impacts in any level of tree-like topology in wireless networks.	Fault-tolerant architectures for heterogeneous networks and failure recovery protocols are proposed in [7].
Service		
Multi-operators and billing system	To collect, manage, and store the customers' accounting information from multiple service providers. Also, to bill the customers with simple but detailed information.	Various billing and accounting frameworks are proposed in [8, 9].
Personal mobility	To provide seamless personal mobility to users without modifying the existing servers in heterogeneous systems.	Personal mobility frameworks are proposed. Most of them use mobile agents, but some do not [10, 11].

Table 1. *A summary of key challenges and their proposed solutions.*

research areas. An overview of the challenges in future heterogeneous systems will be provided. Each area of challenges will be examined in detail. The article is then concluded.

An Overview of the Challenges in Integrating Heterogeneous Systems

It is convenient to discuss the challenges (and their proposed solutions) by grouping them into three different aspects: mobile station, system, and service. Each of these aspects leads to several important research areas.

Research Challenges

In this section each of the key research areas listed in Table 1 will be described in detail.

Mobile Station

Multimode User Terminals — In order to use the large variety of services and wireless networks in 4G systems, multimode user terminals are essential as they can adapt to different wireless networks by reconfiguring themselves. This eliminates the need to use multiple terminals (or multiple hardware components in a terminal). The most promising way of implementing multimode user terminals is to adopt the software radio approach [1]. Figure 1 shows the design of an ideal software radio. The analog part of the receiver consists of an antenna, a bandpass filter (BPF), and a low noise amplifier (LNA). The received analog signal is digitized by the analog/digital converter (ADC) immediately after the analog processing. The processing in the next stage (usually still analog processing in conventional terminals) is then performed by a reprogrammable baseband digital signal processor (DSP). The DSP will process the digitized signal in accordance with the wireless environment. Unfortunately, the current software radio technology is not completely feasible for all the different wireless networks due to the following technological problems. First, it is impossible to have just one antenna and one LNA to serve the wide range of frequency bands (i.e., to cover the bands of all 4G wireless networks). The only solution is to use

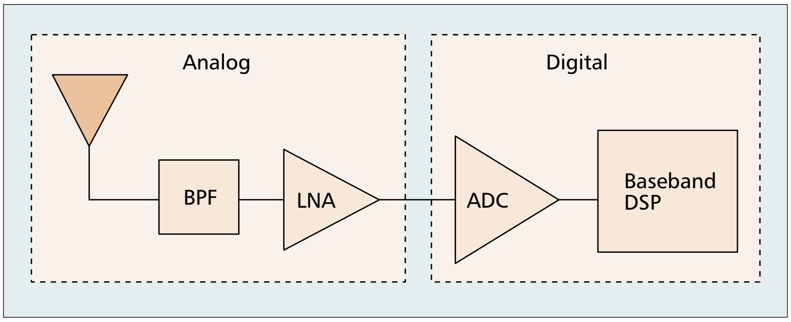

Figure 1. *An ideal software radio receiver.*

multiple analog parts to work in different frequency bands. This certainly increases the design complexity and physical size of a terminal. The second challenge is that existing ADCs are not fast enough. For example, the GSM and Universal Mobile Telecommunications Service (UMTS) waveforms require at least 17 bits resolution with very high sampling rates (over 100 Msamples/s). To provide such bit resolution, the speed of the fastest current ADC is still two to three orders of magnitude slower than required. Finally, in order to allow real-time execution of software implemented radio interface functions such as frequency conversion, digital filtering, spreading, and despreading, parallel DSPs have to be used. This also creates problems such as high circuit complexity, and high power consumption and dissipation.

Wireless System Discovery — To use 4G services, multimode user terminals should be able to select the target wireless systems. In current GSM systems, base stations periodically broadcast signaling messages for service subscription to mobile stations. However, this process becomes complicated in 4G heterogeneous systems because of the differences in wireless technologies and access protocols. One of the proposed solutions is to use software radio devices that can scan the available networks. After scanning, they will load the required software and reconfigure themselves for the selected network. There are a number of ways to facilitate the downloading of software modules [2]. Figure 2 shows an example of how a multimode terminal attached to a WLAN is scanning the available wireless networks. Once the terminal discovers the available systems, it can download the suitable software to reconfigure the software radio. As shown, the software can be downloaded from media such as a PC server, smart card, or memory card, or over the air (OTA). Each downloading method has its own advantages and disadvantages with respect to speed, accuracy, resource usage, and convenience. OTA is the most challenging way to achieve wireless system discovery, but its availability frees users from the tedium of downloading. Operators will also enjoy simplified network management. Le and Aghvami [2] proposed an OTA downloading approach in which multimode user terminals constantly monitor a predefined broadcasting channel (global pilot and download channel, GPDCH) to check for available networks. Once they detect a new available network, they can decide whether or not a change should be made. As pointed out in [2], we still need to solve problems such as the long downloading time and slow speed of the GPDCH.

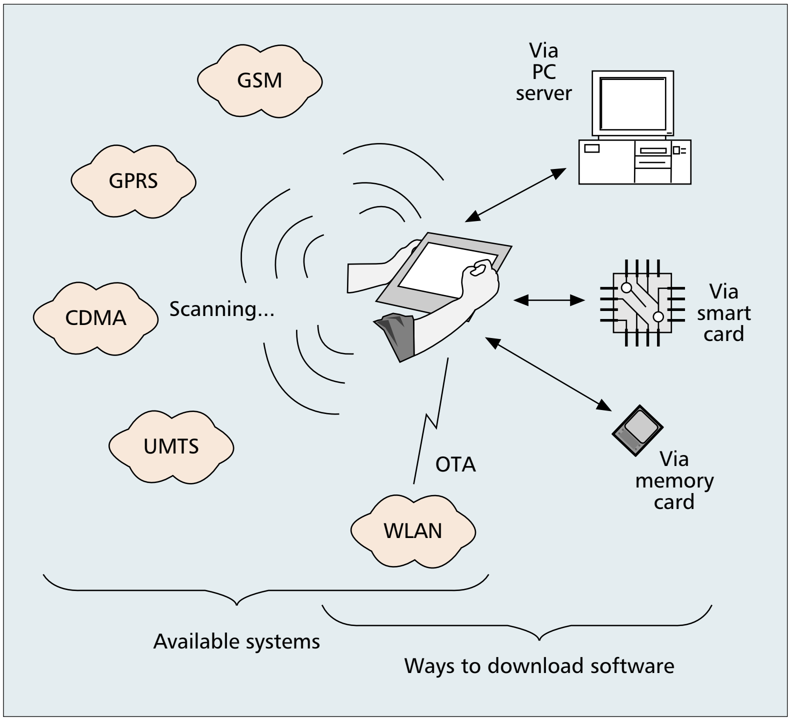

Figure 2. *A multimode terminal attaches to the WLAN and scans the available systems. It can download suitable software manually or automatically.*

Wireless System Selection — With the support of 4G user terminals, we can choose any available wireless network for each particular communication session. As every network has unique features, using a suitable network for a specific service may optimize system performance and resource usage. Furthermore, the right network selection can ensure the QoS required by each session. However, it is complicated to select a suitable network for each communication session since network availability changes from time to time. Moreover, adequate knowledge of each network is required before a selection is made. This includes precise understanding of the supported service types, system data rates, QoS requirements, communication costs, and user preferences. Eguchi *et al.* [3] proposed a selection scheme in which Session Initiation Protocol (SIP) messages, location information of the source mobile node, available networks of both mobile nodes, and user preferences are all taken into account in the selection when a mobile node makes a call to another mobile node. Other researchers also suggest that network resources and minimum QoS requirements should be considered in network selection [5]. Despite these research efforts, we believe that there are many issues to be resolved in selecting the appropriate wireless system.

SYSTEM

Terminal Mobility — In order to provide wireless services at any time and anywhere, terminal mobility is a must in 4G infrastructure. Terminal mobility allows mobile clients to roam across geographic boundaries of wireless networks. There are two main issues in terminal mobility: location management and handoff management. With location management, the system tracks and locates a mobile terminal for possible connection. Location management involves handling all the information about the roaming terminals, such as original and current located cells, authentication information, and QoS capabilities. On the other hand, handoff management maintains ongoing communications when the terminal roams. Mobile IPv6 (MIPv6) is a standardized IP-based mobility protocol for IPv6 wireless systems. In this design, each terminal has an IPv6 home address. Whenever the terminal moves outside the local network, the home address becomes invalid, and the terminal obtains a new IPv6 address (called a care-of address) in the visited network. A binding between the terminal's home address and care-of address is updated to its home agent in order to support continuous communications. However, this handoff process causes an increase in system load, high handover latency, and packet losses. Although some enhanced Mobile IPv6 (MIPv6) schemes have been proposed to solve these problems, more needs to be done to satisfactorily overcome these problems [5]. It is even more difficult to solve these problems in 4G networks. The reason is that besides horizontal handoff, vertical handoff is also needed. Figure 3 shows an example of horizontal and vertical handoff. Horizontal handoff is performed when the terminal moves from one cell to another within the same wireless system. Vertical handoff, however, handles the terminal movement between two different wireless systems (e.g., from WLAN to GSM). Moreover, 4G networks are expected to support real-time multimedia services that are highly time-sensitive. It is unacceptable if the MIPv6 handoff process significantly degrades system performance, especially QoS performance. In addition, it is hard to decide the correct handoff time because measuring handoffs among different wireless systems is very complicated. Finally, the uncertain handoff completion time adds to the complexity in designing good handoff mechanisms. To overcome these problems, researchers are currently investigating new handoff decision policies and new handoff algorithms for heterogeneous networks.

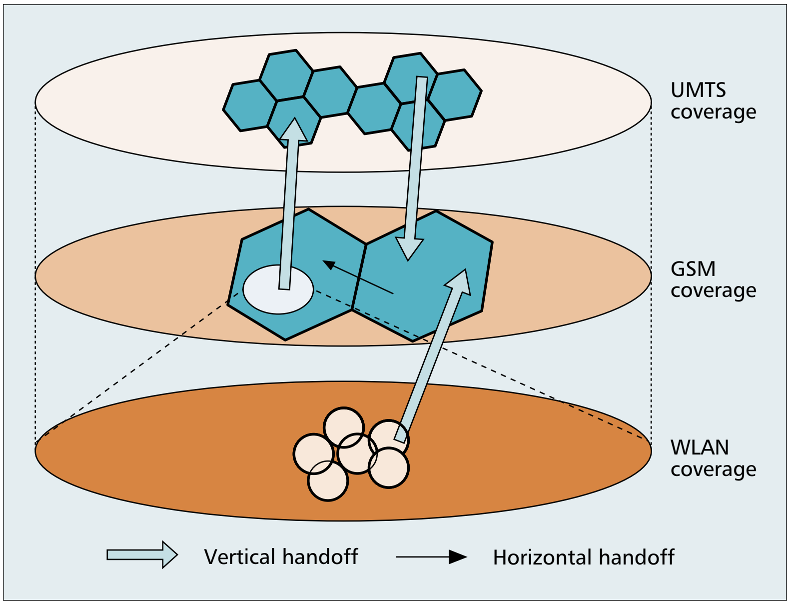

Figure 3. *Vertical handoff and horizontal handoff of a mobile terminal.*

Network Infrastructure and QoS Support — Existing wireless systems can be classified into two types: non-IP-based and IP-based. Many non-IP-based systems are highly optimized for voice delivery (e.g., GSM, cdma2000, and UMTS). On the other hand, IP-based systems are usually optimized for data services (e.g., 802.11 WLAN and HiperLAN). In 4G wireless environments, the problem in integrating these two systems becomes apparent. Research challenges such as QoS guarantee for end-to-end services need to be addressed, although they are by no means easy to tackle, especially when time-sensitive or multimedia applications are considered.

Current QoS designs are usually made with a particular wireless system in mind. For example, the 3G Partnership Project (3GPP) has proposed a comprehensive QoS architecture for UMTS. It realizes QoS in UMTS via the UMTS Bearer Service and its underlying bearer services [6]. There are clear definitions of characteristics and functionalities of each bearer service on a specific layer. These enable the provision of a contracted QoS in all aspects, including control signaling, radio interface transport, and QoS management functionality. Additionally, in order to support various services, the UMTS specification has defined QoS classes and their attributes for dealing with differentiated QoS requirements. However, providing QoS only in UMTS cannot guarantee end-to-end QoS because systems that are non-UMTS are involved. To address this problem, internetworking with most common QoS architectures is studied in 3GPP. We believe that internetworking mechanisms involving layer 3 (or above) operations may be needed.

Security and Privacy — Security requirements of 2G and 3G networks have been widely studied in the literature. Different standards implement their security for their unique security requirements. For example, GSM provides highly secured voice communications among users. However, the existing security schemes for wireless systems are inadequate for 4G networks, as stated in [4]. The key concern in security designs for 4G networks is flexibility. As the existing security schemes are mainly designed for specific services, such as voice service, they may not be applicable to 4G environments that will consist of many heterogeneous systems. Moreover, the key sizes and encryption and decryption algorithms of existing schemes are also fixed. They become inflexible when applied to different technologies and devices (with varied capabilities, processing powers, and security needs). To design flexible security systems, some researchers are starting to consider reconfig-

There are two ways to achieve fault-tolerant architectures to support QoS under failures. The first is to use a hierarchical cellular network system; the second is to use collocated or overlapping heterogeneous wireless networks. However, more work should be done in order to build fault-tolerant 4G systems in both models.

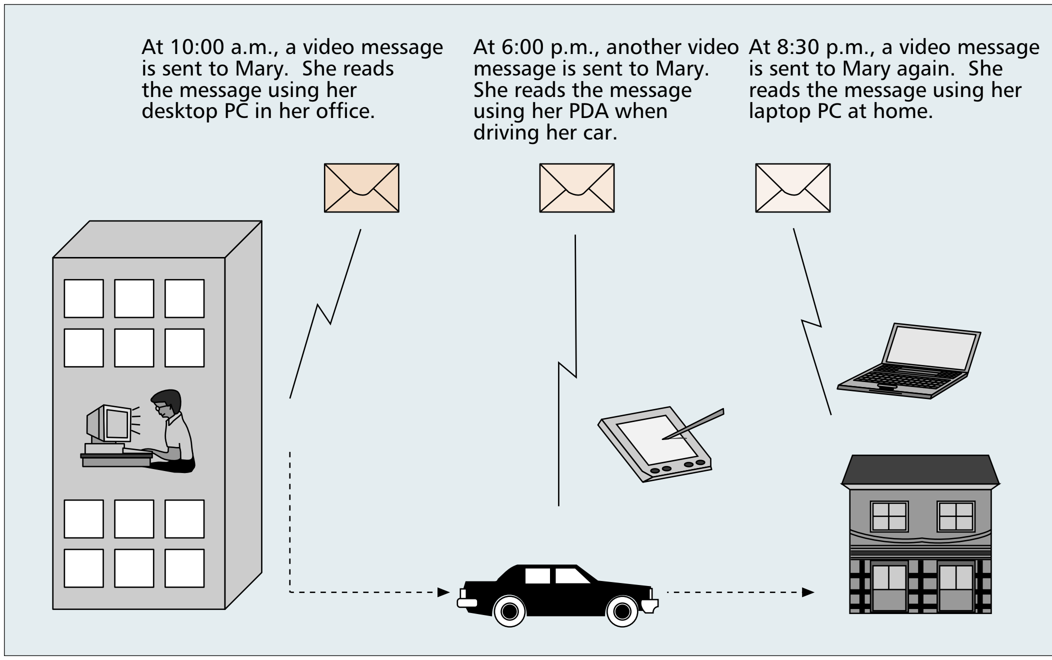

Figure 4. *An example of personal mobility.*

urable security mechanisms. As an example, Tiny SESAME is a lightweight reconfigurable security mechanism that provides security services for multimedia or IP-based applications in 4G networks [6].

Fault Tolerance and Survivability — In the past, extensive work has been done to provide fault tolerance in wired networks and high-speed data networks (e.g., public switched telephone networks and asynchronous transfer mode networks). These attempts have improved the reliability, availability, and survivability of the networks under study. However, there is inadequate study on the survivability of wireless access networks, even though they are more vulnerable than wired networks. A cellular wireless network is typically designed as a tree-like topology that has several levels, including device, cell, switch, and network levels. One major weakness of this topology, though, is that when any level fails (in either hardware or software), all levels below will be affected. For example, damage of a base station in a cell may cause partial or full service loss in that cell. The situation becomes worse when multiple tree topology networks work together in 4G systems. Their fault-tolerant designs should consider power consumption, user mobility, QoS management, security, system capacity, and link error rates of many different wireless networks. To simplify the survivability design, Tipper *et al.* [7] propose three classes of strategies to improve network survivability in different layers: prevention, network design and capacity allocation, and traffic management and restoration. But the work is not for 4G networks, so it remains to be seen whether these strategies are applicable to 4G situations. There are two ways to achieve fault-tolerant architectures to support QoS in failures [6]. The first is to use a hierarchical cellular network system; the second is to use collocated or overlapping heterogeneous wireless networks. However, more work should be done in order to build fault-tolerant 4G systems in both models.

Services

Multiple Operators and Billing System — In today's mobile market, an operator usually charges customers with a simple billing and accounting scheme. A flat rate based on subscribed services, call durations, and transferred data volume is usually enough in many situations. However, with the increase of service varieties in 4G systems, more comprehensive billing and accounting systems are needed. Customers may no longer belong to only one operator, but instead subscribe to many services from a number of service providers at the same time. It may be very inconvenient for a customer to deal with multiple service providers. Instead, a brokering service can be provided. Customers do not have to waste time handling all the financial transactions involved. To achieve this, operators need to design new business architecture, accounting processes, and accounting data maintenance. Moreover, equalization on different charging schemes is also needed. This is because different billing schemes may be used for different types of services (e.g., charging can be based on data, time, or information). It is challenging to formulate one single billing method that covers all the billing schemes involved. Furthermore, 4G networks support multimedia communications, which consists of different media components with possibly different charging units. This adds difficulty to the task of designing a good charging scheme for all customers [8]. Besides, the media components may have different QoS requirements. It is very complicated to decide a good tariff for all the possible components. In order to build a structural billing system for 4G networks, several frameworks have already been studied. The requirements on these frameworks include scalability, flexibility, stability, accuracy, and usability [9].

Personal Mobility — In addition to terminal mobility, personal mobility is a concern in mobility management. Personal mobility concentrates on the movement of users instead of users' terminals, and involves the provision of personal communications and personalized operating environments. Figure 4 demonstrates the concept of personal communications using a personalized video message application. As shown in the figure, when there is a video message addressed to the mobile user, no matter where the user is located or what kind of terminal is being used, the message will be sent to the user correctly. A personalized operating environment, on the other hand, is a service that enables adaptable service presentations (in order to fit the capabilities of the terminals in use regardless of network types). Currently, there are several frameworks on personal mobility found in the literature. Mobile-agent-based infrastructure is one widely studied solution [10, 11]. In this infrastructure, each user is usually assigned a unique identifier and served by some personal mobile agents (or specialized computer programs running on some servers). These agents act as intermediaries between the user and the Internet. A user also belongs to a home network that has servers with the updated user profile (including the current location of the user's agents, user's preferences, and currently used device descriptions). When the user moves from his/her home network to a visiting network, his/her agents will migrate to the new network. Referring to the example shown in Fig. 4, when somebody makes a call request to Mary, the caller's agent first locates Mary's agent by making a location request to her home network. By looking up Mary's profile, her home network sends back the location of Mary's agent to the caller's agent. Once the caller's agent identifies Mary's location, the caller's agent can directly communicate with her agent. Different agents may be used for different services. A mobile-agent-based infrastructure proposed in [10] uses four assistants (user assistant, HTTP assistant, mail assistant, and FTP assistant) to personalize user operating environments. However, there are other personal mobility frameworks that do not rely on mobile agents.

In addition to terminal mobility, personal mobility is another concern in mobility management. Personal mobility concentrates on the movement of users instead of users' terminals, and involves the provision of personal communications and personalized operating environments.

Conclusion

In this article research challenges in the migration to 4G networks are studied and described. The challenges are grouped into three aspects: mobile station, system, and service. Some of the challenges are well studied, such as multimode user terminals, wireless system discovery, terminal mobility, and QoS support. On the other hand, others are less studied. These include wireless system selection, security, failure, and survivability. Moreover, work on the implementations of personal mobility, billing, and accounting systems are also needed in 4G networks. The discussion in this article not only shows that there is much work to be done in the migration to 4G systems, but also highlights that current systems must be implemented with a view to facilitate a seamless integration into 4G infrastructure. Without these infrastructures, 4G services will not be launched easily.

Acknowledgment

The authors are grateful to Professor David Hill, Dr. K. T. Ko and Miss Judy Keung for their helpful comments on the article.

References

[1] E. Buracchini, "The Software Radio Concept," *IEEE Commun. Mag.*, vol. 38, no. 9, 2000, pp. 138–43.

[2] T. H. Le and A. H. Aghvami, "Performance of an Accessing and Allocation Scheme for the Download Channel in Software Radio," *Proc. IEEE Wireless Commun. and Net. Conf.*, vol. 2, 2000, pp. 517–21.

[3] H. Eguchi, M. Nakajima, and G. Wu, "Signaling Schemes over a Dedicated Wireless Signaling System in the Heterogeneous Network," *Proc. IEEE VTC*, Spring 2002, pp. 464–67.

[4] J. Al-Muhtadi, D. Mickunas, and R. Campbell, "A Lightweight Reconfigurable Security Mechanism for 3G/4G Mobile Devices," *IEEE Wireless Commun.*, vol. 9, no. 2, Apr. 2002, pp. 60–65.

[5] N. Montavont and T. Noel, "Handover Management for Mobile Nodes in IPv6 Networks," *IEEE Commun. Mag.*, vol. 40, no. 8, Aug. 2002, pp. 38–43.

[6] 3GPP TS 23.107 v.5.9.0, "Quality of Service (QoS) Concept and Architecture," June 2003.

[7] D. Tipper *et al.*, "Providing Fault Tolerance in Wireless Access Networks," *IEEE Commun. Mag.*, vol. 40, no. 1, Jan. 2002, pp. 58–64.

[8] F. Ghys and A. Vaaraniemi, "Component-based Charging in a Next-generation Multimedia Network," *IEEE Commun. Mag.*, vol. 41, no. 1, Jan. 2003, pp. 99–102.

[9] J. Fleck, "A Distributed Near Real-time Billing Environment," *Telecommun. Info. Net. Architecture*, 1999, pp. 142–48.

[10] A. D. Stefano and C. Santoro, "NetChaser: Agent Support for Personal Mobility," *IEEE Internet Comp.*, vol. 4, no. 2, Mar./Apr. 2000, pp. 74–79.

[11] B. Thai and A. Seneviratne, "IPMoA: Integrated Personal Mobility Architecture," *Comp. and Commun.*, 2001, pp. 485–90.

Biographies

Suk Yu Hui (syhui@ee.cityu.edu.hk) is an M.Phil. candidate in the Department of Electronic Engineering, City University of Hong Kong. Her research interests include mobile communication systems and wireless local area networks. She received a B.Sc. in information technology from City University of Hong Kong.

Kai Hau Yeung [M] (eeayeung@cityu.edu.hk) is an associate professor in the Department of Electronic Engineering, City University of Hong Kong. His research interests include Internet infrastructure security, mobile communication systems, and Internet caching systems. He is also an active industry consultant in the areas of computer networking and communication systems. From 1996 to 1998, he was involved in a project to develop a 900 MHz GSM mobile handset. The project team in City University successfully developed a handset prototype for a listed company in Hong Kong. He is a member of the ACM, and BCS. He is also a Cisco Certified Network Professional and Cisco Certified Academy Instructor.