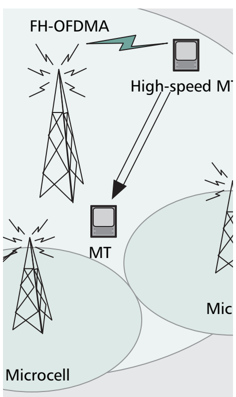
ALL-IP 4G NETWORK ARCHITECTURE FOR EFFICIENT MOBILITY AND RESOURCE MANAGEMENT

YOUNG-JUNE CHOI, UNIVERSITY OF MICHIGAN
KWANG BOK LEE AND SAEWOONG BAHK, SEOUL NATIONAL UNIVERSITY



Third-generation (3G) access networks, like WCDMA (wideband code division multiple access) and cdma2000, have a complicated network structure and define many protocols to cover the system structure. Accordingly, it is expected that fourth-generation (4G) networks will have a simple structure based on all-IP.

ABSTRACT

In this article, we investigate 4G network architecture and consider two underlying layers: PHY and MAC. We compare two models of wireless access network: pure all-IP and subnet based networks. The pure all-IP model is simple and cost-efficient but causes implementation issues of mobility management and resource coordination. In contrast, the subnet-based network enables layer 2 and layer 3 handoffs to be executed independently, deploying several access points under an access router. Further, to handle various cases efficiently according to traffic class and mobility, we present an advanced model of a hierarchical cellular system that combines multiple access techniques of OFDMA and FH-OFDMA with microcells and macrocells. Finally, as an integrated approach to support diverse QoS requirements, we consider an IP-triggered resource allocation strategy (ITRAS) that exploits IntServ and DiffServ of the network layer to interwork with channel allocation and multiple access of MAC and PHY layers, respectively. These cross layer approaches shed light on designing a QoS support model in a 4G network that cannot be handled properly by a single layer based approach.

INTRODUCTION

Third-generation (3G) access networks, like WCDMA (wideband code division multiple access) and cdma2000, have a complicated network structure and define many protocols to cover the system structure. Accordingly, it is expected that fourth-generation (4G) networks will have a simple structure based on *all-IP* [1] where Internet Protocol (IP) packets traverse an access network and a backbone network without any protocol conversion. Since 3G networks basically evolved from a circuit-switched cellular network, they have their own gateways to interpret IP from the backbone network. They also have their own protocols and interfaces for the communication within themselves. To overcome these problems, 4G networks are expected to become an all-IP based packet-switched system, similar to the IP backbone network.

4G networks have two different visions: *revolution* — developing an innovative system and *evolution* — interworking with existing systems. The interworking model takes an approach that integrates cellular networks, wireless metropolitan area networks (WMAN), wireless local area networks (WLAN), and wireless personal area networks (WPAN). This model covers a future scenario of ubiquitous networking where anyone can access a network anytime, anywhere, and anyway. The IEEE 802.11 WLAN achieves system throughput up to 54 Mb/s while the service area is limited to two or three hundred meters. In contrast, a current cellular network such as cdma2000 1x EV-DO (evolution-data only) covers several kilometers, but its cell throughput is at most, 2 Mb/s. Therefore, it is essential to develop an innovative system with high throughput and wide coverage.

The new system is expected to employ novel techniques such as orthogonal frequency division multiplexing (OFDM) and multiple input multiple output (MIMO) antennas [1]. If the system considers various conditions such as high speed mobile or nomadic users, data or voice traffic, and cell center or boundary conditions, it may be required to exploit hybrid multiple access techniques. As a candidate for the 4G network, the IEEE 802.16 standard [2] sets a goal of WMAN/WLAN based on OFDM or orthogonal frequency division multiple access (OFDMA). The IEEE 802.16e standard supplements it for mobility support such as IEEE 802.20.

In Korea, a new service for the 2.3 GHz band based on IEEE 802.16 OFDMA, named WiBro, started in the second quarter of 2006. It is designed to support the maximum mobility of 60 km/h. To support high mobility, the authors in [3] developed a hybrid multiple access scheme combining OFDMA and frequency hopping (FH)-OFDMA, where fast-moving users access the network via FH-OFDMA. Decoupling multiple access techniques for the hierarchical cell structure, we consider a new wireless network, comprised of OFDMA microcells and FH-OFDMA macrocells.

Although innovative 4G systems — medium

1536-1284/07/$20.00 © 2007 IEEE

access control (MAC) layer and physical (PHY) layer — are under development, there exists little work that considers the network architecture. In this article, we design a new 4G network that enhances MAC and PHY performance. Our work expands cross layer techniques, which are dedicated to the cooperation between MAC (L2) and PHY (L1) layers, to cover the network layer (L3) together.

In this article, we consider the following issues. First, we explain two models of all-IP cellular network architecture. We develop a subnet-based network that can support L2 and L3 handoffs separately and compare it with a pure all-IP network. We then design a hierarchical cellular network that consists of microcells and macrocells, each having its own multiple access mechanism. The network chooses cell type and multiple access method according to mobile speed and traffic type. Finally, we present a quality of service (QoS) support methodology that tightly couples all three layers.

Network Architecture

All-IP Cellular Network

Existing cellular networks, based on circuit-switching, consist of base stations (or base transceiver stations), base station controllers, switching centers, gateways, and so on. The base station (BS) plays the role of physical transmission with fast power control and wireless scheduling. The base station controller (BSC) executes the largest part of the radio resource management. Whenever a mobile terminal (MT) moves into another cell, it requires handoff to another base station.

In contrast, the 4G network has a simple structure where each BS must function intelligently to perform radio resource management as well as physical transmission. This makes the BS act the role of an access router (AR). This architecture is shown in Fig. 1. It incurs high overhead, however, especially when an MT configures mobile IP (MIP) addresses for handoff. As we know, it takes several seconds to run the MIP handoff [4], and MIP hinders an MT from performing smooth handoff. In addition, the 4G network is expected to have a small cell radius due to its use of high frequency band, which possibly results in short cell residence time. For this matter, reducing the latency is still a challenging issue. In particular, a *fast handoff scheme* [5] proposes to decrease the address resolution delay by pre-configuration.

To solve the fundamental problem of all-IP cellular networks, we can separate the functionality of an AR from that of an access point (AP) so that each undertakes L3 and L2 protocols, respectively. Figure 2 shows an example of a simple network where an AR manages several AP. This relation is similar to that between BSC and BS in existing cellular networks. A subnet consisting of an AR and several AP can be configured by Gigabit Ethernet. Then, an MT moving within the subnet (i.e., changing AP) performs L2 handoff without changing MIP attachment. When the MT moves into another AR area, it experiences L3 handoff.

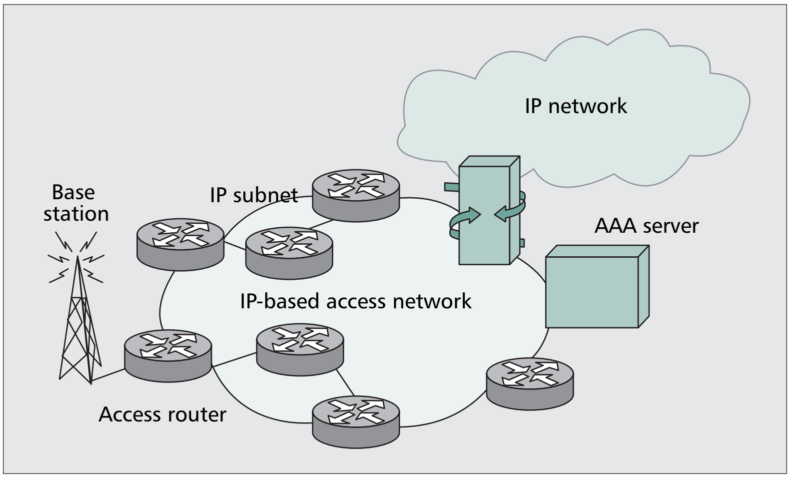


Figure 1. *The pure all-IP 4G network.*

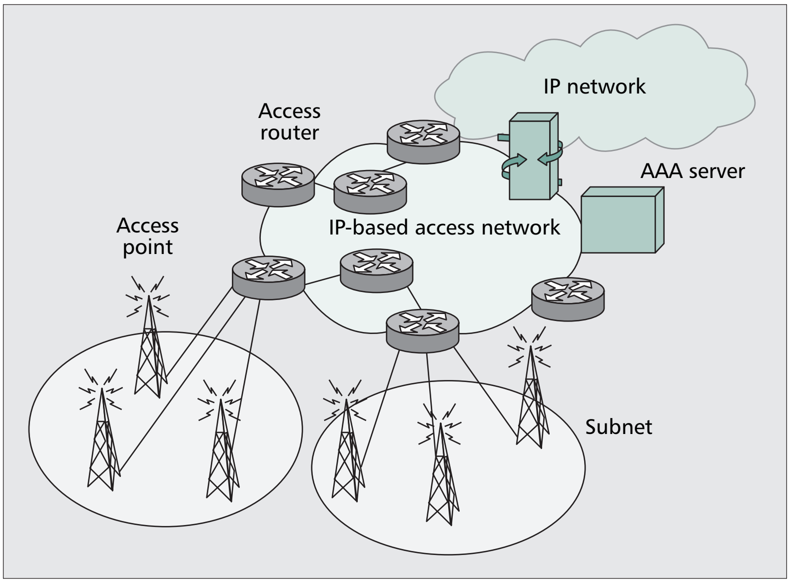


Figure 2. *The subnet-based 4G network.*

Table 1 compares the network architecture of pure and subnet-based all-IP cellular networks. A main difference is that the former is decentralized, and the latter is centralized. Since the pure all-IP network incurs L3 protocol in the end access link, it requires long handoff latency and high signaling overhead. However, the architecture is simple and cost-efficient for implementation. On the other hand, the subnet-based all-IP network implements hierarchical architecture, so it is possible to do efficient resource management in spite of its inflexibility.

Network Architecture for Efficient Multiple Access

We now solve the mobility problem by another approach. Generally, cells are categorized into macrocells, microcells, and picocells depending on size. Macrocells and microcells are usually deployed in rural and urban regions, respectively, and the picocells are deployed in buildings. In some regions such as a hot-spot zone, an MT can access both macrocells and microcells such as in Fig. 3. The authors in [6] designed a service model by mobility, such that macrocells and microcells cover high speed and low speed MT, respectively. This structure is effective because a high speed MT must change cells frequently, when covered by microcells.

We extend the hierarchical cell structure by integrating multiple access techniques. Some systems under development are based on OFDMA that combines OFDM and frequency division multiple access (FDMA) [2]. Since an OFDMA system has several channels in a frequency domain, it has higher allocation granularity than an OFDM system. It also can take advantage of adaptive modulation and coding (AMC), but its application is limited to low mobility. If an MT using OFDMA has high mobility, it cannot perform coherent detection properly due to the long symbol.

An *FH-OFDMA system*, which combines frequency-hopping and OFDMA, has the advantage of exploiting diversity [1]. Though it experiences difficulty in supporting high data rates and AMC, it can overcome channel fading and multi-user interference through an FH pattern. Accordingly, it is a viable combination that microcells for low mobility use OFDMA that has fine granularity, while macrocells for high mobility use FH-OFDMA that is robust to channel fading and interference.

As well, each cell plane can handle traffic classes differently. High rate data services are suitable for OFDMA that has high spectral efficiency and supports various data rates by AMC. In contrast, low rate services, such as voice, are adequate for FH-OFDMA that is easy to use diversely. If an MT can support dual modes, it can switch cells according to mobility and traffic type in a manner of using vertical handoff that offers the additional merit of load balancing.

In summary, the hierarchical network that consists of OFDMA microcells and FH-OFDMA macrocells can support various users with different mobility and traffic types. As the network considered is based on a common OFDMA platform, it is more manageable than other heterogeneous networks. Also, it provides flexibility in network planning. 4G networks most probably will be overlaid with 2G or 3G cellular networks. As existing cellular networks are basically designed for circuit-switched voice service, in some regions, they will continue to undertake voice users and high mobility users such as the FH-OFDMA macrocell system, while 4G networks will focus on data traffic users by using the OFDMA microcell system.

	Pure all-IP network	Subnet-based all-IP network
Access network components	AR	AP + AR
Operation type	Decentralized	Centralized
Coordination among cells	Complex but flexible	Simple but fixed
Handoff overhead	High	Low
Handoff protocol	L3	L2 + L3
Cost	Low	High

Table 1. *Comparison of two network types.*

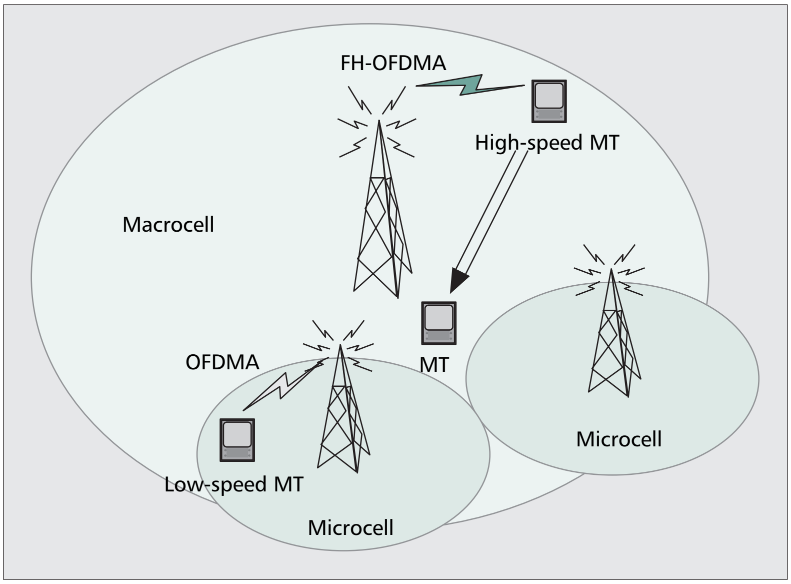


Figure 3. *The model of a hierarchical cellular network.*

IP-Triggered Resource Allocation Strategy (ITRAS)

To support IP QoS, the Internet Engineering Task Force (IETF) recommends integrated services (IntServ) [7] and differentiated services (DiffServ) [8]. These services also are expected to be effective in all-IP-based 4G networks. Since 4G networks will support multimedia traffic, we must visit the issue of providing IP QoS in IP based wireless access networks and propose ITRAS for QoS support in 4G networks, where the decision of radio resource allocation follows IntServ or DiffServ policy.

IP QoS

IntServ [7] uses Resource Reservation Protocol (RSVP) to reserve bandwidth during the session setup. As a first step in RSVP, the source sends a QoS request message of PATH to the receiver through intermediate routers that run an admission and a policy control. If the sender receives RESV returned from the receiver through the reverse route as an indication of QoS guarantee, it initiates the session. If each router along the path receives packets, it classifies and schedules them. IntServ ensures strict QoS, but each router must implement RSVP and maintain per-flow state, which can cause difficulties in a large-scale network.

DiffServ [8], on the other hand, does not require a signaling protocol and cooperation among nodes. As the QoS level of a packet is indicated by the DS field of IP header (TOS (type of service) field in IPv4, Traffic Class field in IPv6), each domain can deal with it independently. After the packet is classified, each router can mark, shape, or drop it according to network status. Since DiffServ is not so rigorous as IntServ, it is scalable in supporting QoS statistically.

QoS of Wireless Access Networks

In general, a wireless access network can manage QoS independently of the IP network because it becomes a bottleneck for providing end-to-end QoS. QoS control is possible by using access and scheduling methods. Recently, the QoS of IEEE 802.11 WLAN system was supplemented by the IEEE 802.11e standard. It defines extended distributed contention access (EDCA), which assigns a small backoff number to high priority traffic and hybrid coordination function (HCF), which improves the conventional polling scheme of point coordination function (PCF). Also, cdma2000 1x EV-DO and WCDMA-HSDPA (high speed downlink packet access) adopted opportunistic scheduling to exploit channel fluctuation. The opportunistic scheduling is a hot issue in designing various scheduling algorithms for QoS [9].

The Third Generation Partnership Projects (3GPP and 3GPP2) define four traffic classes and their related parameters for QoS provisioning. There exist gateways between IP backbone and access networks that perform protocol conversion and QoS mapping between IP and access networks [10]. However, direct translation is difficult since access networks have their own QoS attributes that require strict QoS provisioning within them.

Meanwhile, the importance of unified QoS management grows in 4G networks as QoS management for both access networks and IP networks becomes cumbersome in all-IP networks. If each network has an individual QoS model, it requires a rule that integrates the QoS models to ensure end-to-end QoS. For unified QoS management, we propose ITRAS that considers L1, L2, and L3 together. In ITRAS, L1 and L2 allocate radio resources and logical channels, respectively, according to the QoS indication of L3.

ITRAS

ITRAS concerns the information about IntServ and DiffServ for the resource management of L1 and L2. When IntServ establishes a real-time session, MAC reserves a dedicated channel. In contrast, when DiffServ is used for low mobility users, MAC can exploit either a dedicated or a shared channel. If the shared channel is allocated for DiffServ, the wireless scheduler runs a scheduling algorithm for QoS provisioning. In contrast, the dedicated channel allocation requires an admission control that enables a limited number of users into the network for QoS support. Therefore, IP QoS information helps MAC and PHY manage the following resources in a flexible manner.

- Cell type — microcell or macrocell
- Multiple access — OFDMA or FH-OFDMA
- MAC channel — dedicated or shared
- PHY scheduling — priority or fairness

IntServ is easy to involve in radio resource management because wireless access is usually accompanied by signaling. When an MT requests a real-time service in a 4G network, the corresponding AR can initiate IntServ and allocate a dedicated channel. For a downlink call, the AR can adjust the bandwidth of a dedicated channel with the aid of RSVP. As real-time traffic usually requires a constant data rate, the dedicated channel is recommended to use power control rather than AMC. In this aspect, FH-OFDMA and CDMA may be more suitable than OFDMA for real-time services.

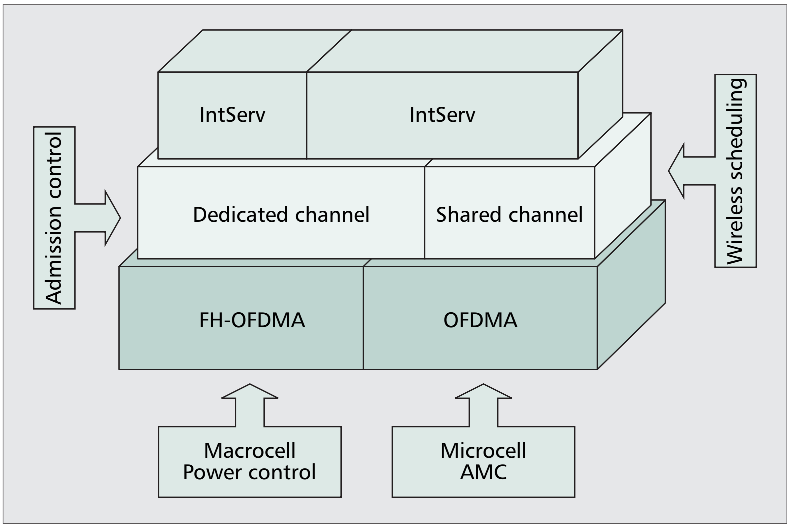


Figure 4 . *The coupled layering for resource management.*

Regarding DiffServ in 4G networks, it is sufficient for an MT to set the DS field properly for uplink packets, because the AP controls radio resources before transferring them to the AR. For downlink traffic, the AR classifies packets according to the DS field and chooses a multiple access method, and accordingly, the AP allocates a dedicated or shared channel. The dedicated channel has the advantage of simple management, while the shared channel works well with DiffServ because both require scheduling. In contrast to scheduling in routers, which must handle several flows, wireless scheduling handles fewer connections, which enables it to use a per-user buffer. Therefore, the wireless scheduler can exploit an algorithm with high granularity of radio resources. Figure 4 and Table 2 summarize tightly coupled resource management among three layers through a unified QoS strategy.

Implementing ITRAS requires further study. Specifically, when the subnet-based all-IP network is deployed, an AR should cooperate with its subordinate AP in performing ITRAS functions. While resource management functions are primarily handled by BSC in the existing cellular networks, more functions will be imposed on AP in 4G networks. Basically AR will be responsible for IP QoS, and AP will play the primary role of resource management. Another challenge is the application of ITRAS to the hierarchical network. In this case, a coordinator is required to decide whether an incoming session is served by a macrocell or a microcell. It also will have the capability of load balancing by triggering vertical handoff.

Our approach of tight layer coupling provides a neat and flexible solution to accommodate various demands of 4G networks that cannot be handled properly by a conventional simple layered approach.

Traffic class	Mobility	IP QoS	Logical channel	Multiple access
Real-time	High	IntServ/DiffServ	Dedicated	FH-OFDMA
	Low	IntServ	Dedicated	FH-OFDMA/ OFDMA
		DiffServ	Dedicated/shared	
Non-real-time	High	DiffServ	Shared	FH-OFDMA
	Low	DiffServ	Shared	OFDMA

Table 2. *An example of ITRAS.*

Conclusion

This article discusses a new approach for designing an architecture and QoS model in 4G networks that uses cross layer techniques to cover L1 through L3. We devise a subnet-based cell structure that consists of AP and AR, each handling L2 and L3 handoffs. Also, we combine the multiple access schemes of OFDMA and FH-OFDMA with cell selection to support various mobile speed and traffic types. This approach easily can be compatible with those of existing cellular networks. Finally, we design a unified QoS strategy, named ITRAS that considers L1, L2, and L3 all together. In this approach, IP QoS such as IntServ and DiffServ can determine resource attributes of the wireless access network, that is, MAC channel and PHY resource as well as multiple access type. Our approach of tight layer coupling provides a neat and flexible solution to accommodate various demands of 4G networks that cannot be handled properly by a conventional simple layered approach.

Acknowledgment

This research was supported partially by the National Research Laboratory (NRL) Project and Seoul City R&BD Project, Korea.

References

[1] Y. Kim *et al.*, "Beyond 3G: Vision, Requirements, and Enabling Technologies," *IEEE Commun. Mag.*, Mar. 2003, pp. 120–24.

[2] IEEE 802.16-REVd/D4-2004, "Part16: Air Interface for Fixed Broadband Wireless Access Systems," Mar. 2004.

[3] W.-I. Lee *et al.*, "An OFDMA-based Next-Generation Wireless Downlink System Design with Hybrid Multiple Access and Frequency Grouping Techniques," *J. Commun. and Networks*, vol. 7, June 2005, pp. 115–25.

[4] H. Yokota *et al.*, "Link Layer Assisted Mobile IP Fast Handoff Method Over Wireless LAN Networks," *Proc. ACM MOBICOM '02*, Atlanta, GA, Sept. 2002.

[5] R. Koodli, "Fast Handovers for Mobile IPv6," draft-ietf-mipshop-fast-mipv6-03.txt, Oct. 2004.

[6] Chih-Lin I, L. J. Greenstein, and R. D. Gitlin, "A Microcell/Macrocell Cellular Architecture for Low- and High-Mobility Wireless Users," *IEEE JSAC*, vol. 11, no. 6, Aug. 1993, pp. 885–91.

[7] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," RFC 1633, June 1994.

[8] S. Blake *et al.*, "An Architecture for Differentiated Service," RFC 2475, Dec. 1998.

[9] M. Andrews *et al.*, "Providing Quality of Service Over a Shared Wireless Link," *IEEE Commun. Mag.*, vol. 39, Feb. 2001, pp. 150–54.

[10] J.-C. Chen and T. Zhang, *IP-Based Next Generation Wireless Networks*, Wiley, 2004, pp. 380–403.

Biographies

Young-June Choi (yjunchoi@eecs.umich.edu) received his B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering and Computer Science, Seoul National University in 2000, 2002, and 2006, respectively. He is currently a postdoctoral researcher in the Electrical Engineering and Computer Science Department at the University of Michigan. His research interests include fourth-generation wireless networks, wireless resource management, and cross-layer design. He received a Gold Prize at the Samsung Humantech Thesis Contest in 2006.

Kwang Bok Lee (klee@snu.ac.kr) received his B.A.Sc. and M.Eng. degrees from the University of Toronto, Ontario, Canada, in 1982 and 1986, respectively, and his Ph.D. degree from McMaster University, Hamilton, Ontario, Canada, in 1990. Currently, he is a professor in the School of Electrical Engineering, Seoul National University, Korea. He was a vice chair of the School of Electrical Engineering from 2000 to 2002. He has been serving as a consultant to a number of wireless industries. His research interests include mobile communications and communication technique covering the physical and upper layers. He was with Motorola Canada from 1982 to 1985 and Motorola USA from 1990 to 1996 as a senior staff engineer. At Motorola, he was involved in the research and development of wireless communication systems. He was with Bell-Northern Research, Canada, from 1989 to 1990. He holds 11 U.S. patents and nine Korean patents, and has a number of patents pending. He was an editor of the *IEEE Journal on Selected Areas in Communications* Wireless Series in 2001 and has been an editor of *IEEE Transactions on Wireless Communications* since 2002. He was a co-chair of the (ICC '05 Wireless Communication Symposium and has been a chair of IEEE Communications Society Seoul Section since 2005. He received the Best Paper Award from CDMA International Conference 2000 (CIC 2000) and the Best Teacher Award in 2003 and 2006, respectively, from the College of Engineering, Seoul National University. He received the Special Award from Samsung Advanced Institute of Technology in 2005.

Saewoong Bahk (sbahk@snu.ac.kr) received his B.S. and M.B. degrees in electrical engineering from Seoul National University in 1984 and 1986, respectively, and his Ph.D. degree from the University of Pennsylvania in 1991. In 1994 he joined the School of Electrical Engineering at Seoul National University and currently serves as a professor. From 1991 through 1994 he was with the Department of Network Operations Systems at AT&T Bell Laboratories as a member of technical staff. His areas of interest include performance analysis of communication networks and network security.