(3) IP Access Network Technologies for 4G Mobile Communications

Daichi Funato, Xiaoning He,
Guangrui Fu and Moo Ryong Jeong

Building an Internet friendly access network is an essential issue for the next-generation 4G mobile communication system that supports mobile multimedia services and reduces network deployment costs. This article introduces the work being conducted at DoCoMo USA Labs on access network technologies including the recent developments in standardization activities.

1. Introduction

Recently, Internet technologies have been widely accepted as basic components for the next-generation 4G mobile communication system. Especially, the Internet Protocol (IP) is generally considered the layer 3 technology that will be widely used in future networks, and is therefore being enhanced with advanced mobility, Quality of Services (QoS), and security functions.

However, incorporating all of these rich functions in the IP layer may complicate the deployment and operation of 4G hosts and networks. The IP layer was originally designed as a simple network layer for end-to-end packet delivery. In other words, the IP layer may not be the most appropriate place to implement all of the advanced capabilities needed for 4G mobile communications. Therefore, we have designed the 4G mobile communication architecture considering the interworking between layers and the desirable location to implement advanced functions [1].

This article provides an overview of our research activities relating to the IP access network and Wireless Local Area Network (WLAN)s, with emphasis on our efforts to implement advanced functions below the IP layer. Chapter 2 presents recent research trends on micro-mobility in the access network, Chapter 3 describes split Medium Access Control (MAC) technology, the centralized management scheme for wireless Access Point (AP)s, Chapter 4 introduces Mobile Firewall for access network security, and Chapter 5 presents a fast WLAN scanning method.

2. Micro-Mobility

2.1 Problems in IP Mobility Management

The Internet Engineering Task Force (IETF) is now in the process of developing a standardized Mobile Internet Protocol (MIP) that will enable the IP to support End-HOST (EHOST) mobility [2]. In MIP, when a Mobile Node (MN) travels from one subnet to another subnet, a Binding Update (BU) message will be sent to both the Home Agent (HA) and to the Correspondent Node (CN) to update the location of the MN.

While MIP can support user mobility at the IP level, it does not fully satisfy all the requirements of the next-generation 4G mobile network. First, MIP does not support seamless handover, because it was designed to only accommodate occasional roaming. This led us to develop the Fast MIP (FMIP) protocol that does effectively handle this seamless handover issue [1]. Second, MIP does not adequately address the problem of scalability. For example, if an MN frequently moves between different subnets, then a large volume of BU messages will be generated and sent to both the HA and the CN. This means that, if there is a large number of MNs, then considerable wireless link and Internet backbone resources (bandwidth, etc.) might be consumed to support all the signaling traffic. To address this scalability issue, efforts have been focused on developing a class of protocols called micro-mobility protocols.

2.2 Micro-Mobility Protocols

A number of micro-mobility protocols have been proposed over the past few years [3]-[5]. Most of these proposed schemes share similar characteristics, and can significantly reduce the number of signaling messages and packet losses by:

1) Dividing the network into multiple micro-mobility domains.
2) Assigning two Care of Addresses (CoA) to each MN. One CoA identifies the micro-mobility domain where the MN is located and the other CoA specifies the MN's current location within the domain.
3) When an MN moves within a micro-mobility domain, BU messages are only sent to a router called a gateway to update the location information within the domain. When an MN moves between micro-mobility domains, the gateway router forwards BU messages to the HA and CN so that HA

and CN can update MN's domain information.

Depending on the layers at which these micro-mobility protocols are used, these protocols can be classified as either IP-based micro-mobility protocols or Multi-Protocol Label Switching (MPLS)-based micro-mobility protocols. The IP-based protocols are based on the MIP. Quite a number of IP-based micro-mobility protocols have been proposed, and a detailed comparison of these schemes can be found in [3].

2.3 MPLS-based Micro-Mobility Protocols

MPLS is a protocol that was standardized by the IETF [6]. MPLS is essentially a sub-IP layer (layer 2.5) protocol that provides a way to map layer 3 traffic to connection-oriented layer 2 transports similar to ATM and Frame Relay. With its ability to control layer 2 resources, MPLS is considered to be suitable for controlling traffic in the resource-limited access network to reduce the effects of delay jitter for the Voice over Internet Protocol (VoIP) application and other services due to the bandwidth variations.

The existing MPLS-based micro-mobility protocol is based on a hierarchical MIP model [4]-[5]. Although the existing protocol provides good mobility support and reduces tunnel header overhead, it does not adequately address QoS management or traffic control issues.

To address both of these issues at the same time, we have been focusing research on a new MPLS-based micro-mobility protocol and domain construction method [7]. **Figure 1**, for example, shows that when using the existing micro-mobility protocol, if bandwidth on R1-R2 and R2-R4 links is strained and the MN moves from the area covered by Access Router (AR) 1 to the area covered by AR2, the routing algorithm will use a path from AR1 to AR2 to establish a tunnel between AR1 and AR2. The important point to note is that this path may not satisfy the QoS requirements of the MN. We can address this problem using an extended MPLS signaling protocol to set up a data path that, while not necessarily the shortest path to the gateway router, ensures the QoS demanded by the MN.

3. Split MAC Technology

3.1 Split MAC Architecture

The WLAN has been considered a cost-effective solution for providing broadband wireless service that is complementary to cellular networks, but this approach does not scale well in its deployment. As the enterprise or campus WLAN network continues to grow, more and more APs must be installed because the transmission range is fairly limited. Each AP must be properly configured to deal with security threats. Typically this setup work is a manual task and thus involves a large amount of time and cost. The problem is common to all systems that have

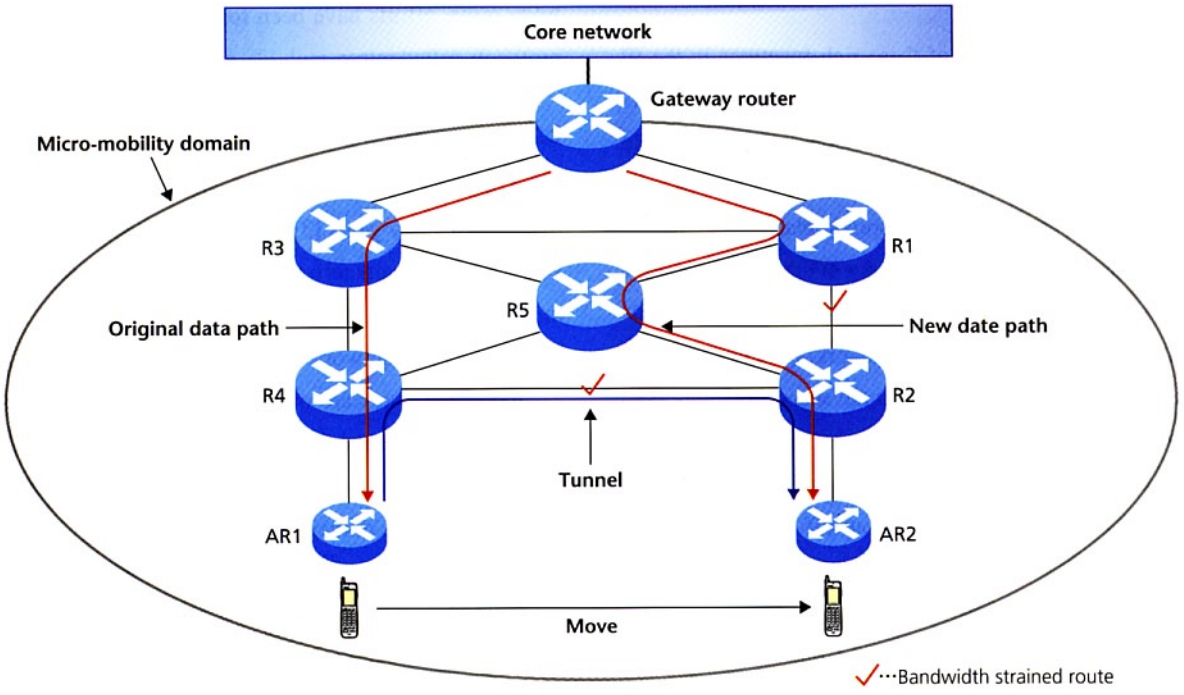


Figure 1 Micro-mobility routing

many APs. The challenge, therefore, is to devise a general approach that can be applied to a broad range of next-generation mobile systems.

Split MAC technology was designed to solve this problem. In this technology, the MAC layer in the wireless part splits the AP and the AR, and assigns MAC frame processing functions to the AR and wireless control functions to the AP. DoCoMo USA Labs and a number of other companies have incorporated this technology in a new protocol called the Light Weight Access Point Protocol (LWAPP) [8] that we have now proposed for standardization to the IETF.

Figure 2 shows the LWAPP protocol stack. In the LWAPP, the AP first automatically discovers the AR and acquires the proper setup parameters from the AR. The AP then executes lower layer IEEE 802.11 radio processing and sends IEEE 802.11 frames received from the mobile to the AR for processing via a tunnel. This allows one AR to function as a control center for multiple APs. With the LWAPP protocol, APs are designated as light, because all the encryption and other complex MAC processing and upper-layer protocol functions needed for authentication have been removed from the APs. Thus it leaves the bare minimum processing capabilities that are required as an AP.

The split MAC technology has been designed to minimize the weight of numerous APs and to automate management of the system. Network administrators are also able to add new functions and capabilities above the MAC layer by simply updating the AR software, which enables services to be upgraded and security settings to be changed quickly and easily.

3.2 Functions of the LWAPP

The main functions implemented by the LWAPP are summarized as follows.

1) AP Initialization and Updating

An AP launches AR discovery process at its initialization stage. Once the AP finds an AR, it sends a configuration request to the AR. The configuration response from the AR includes various parameters needed for the AP to establish a wireless connection including the radio channel frequency, the beacon interval, radio statistic broadcast interval, and so on. Note that these parameters can always be undated by having the AR send an update request. Configuration messages also have the ability to download and install firmware on the APs and upgrade AP functions.

2) MAC Frame Encapsulation

After receiving an 802.11 frame over wireless, the AP forwards the frame to the AR in an 802.3 (Ethernet) frame encapsulation with Received Signal Strength Indication (RSSI) and Signal to Noise Ratio (SNR). The RSSI and SNR values can be used as handoff timing triggers in the mobility management process [9]. The 802.11 frames sent from the AR to the AP are also encapsulated with 802.3.

3) Collection of Radio Statistics

Based on the setting parameters, the AP periodically sends statistical data relating to the radio interface to the AR. By collecting statistical data from a number of APs, the AR is able to comprehensively monitor APs and detect rogue APs.

4) Mobile Station Control

An AR can notify an AP that it is okay to exchange traffic with a particular mobile under specific conditions. This capabil-

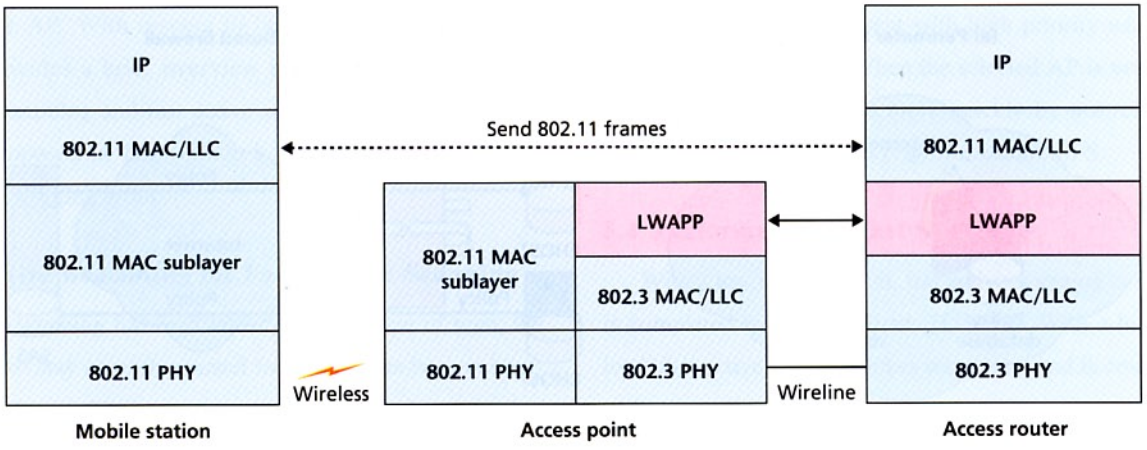


Figure 2 LWAPP protocol stack

ity can be used to implement advanced functions such as access control and QoS management.

4. Mobile Firewall

Our Mobile Firewall technology provides packet filtering protection anywhere mobile users go, and offers a flexible security policy control to both end users and service providers alike.

4.1 Existing Firewall Technologies

Firewalls are extensively used to protect internal networks. The typical firewall separates the network into internal and external parts, and then filters traffic flowing across this separation according to pre-defined security policies.

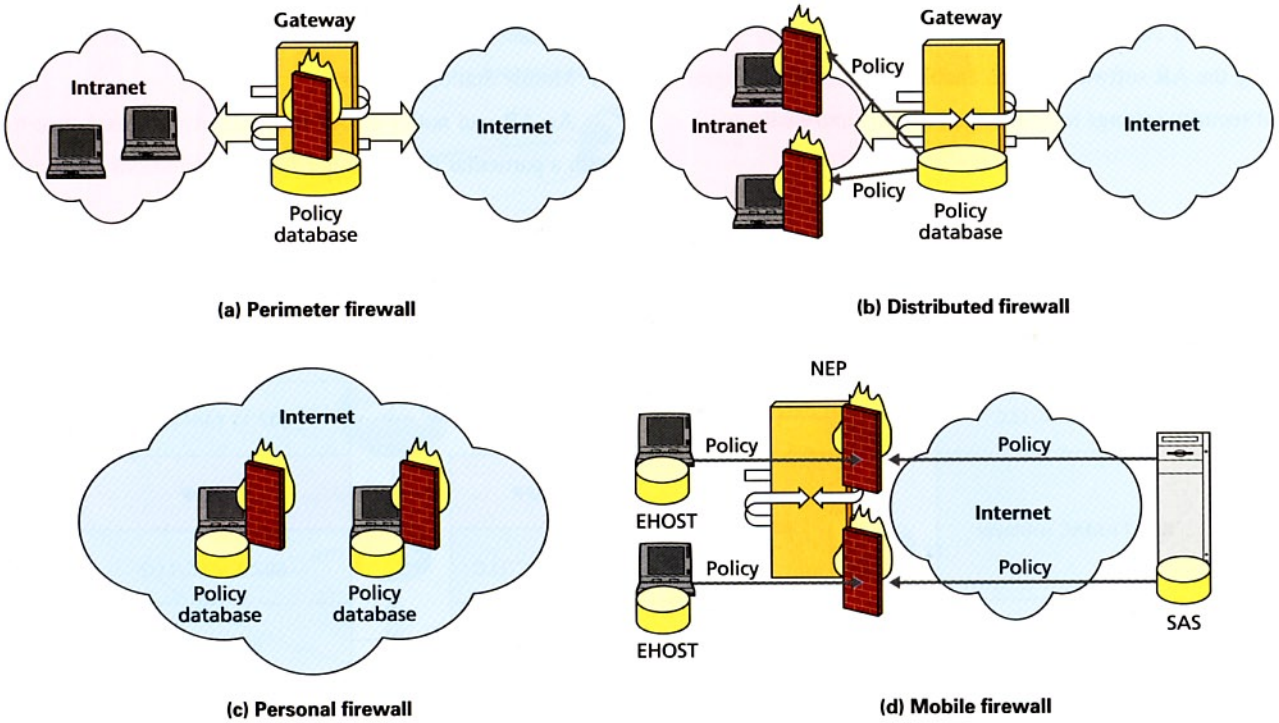
The most common type of firewall is the perimeter firewall such as shown in **Figure 3**(a), in which packets are filtered at network APs such as the network gateways, and usually only network administrators can define the firewall policy. Other existing firewall architectures are distributed firewalls shown in Fig.3(b), and personal firewalls shown in Fig.3(c). In distributed firewalls, packets are filtered at every terminal, but again the policy is defined centrally by network administrators. In personal firewalls, packets are filtered out at each individual terminal and each end user has full control to define the security policy.

These existing firewall technologies have a number of shortcomings when applied to public wireless environments such as hot spots. For example, perimeter firewalls cannot prevent attacks in situation where the attacker and the party being attacked are both located within the same network. Although personal and distributed firewalls are able to protect end users against same-network attacks, these EHOST-based filtering schemes are problematic in that they consume scarce wireless access network bandwidth. These schemes also consume limited battery power and computing resources of EHOSTs, which makes resource-constrained EHOSTs such as mobile terminals and PDAs more vulnerable to Denial of Service (DoS) attacks. Finally, perimeter and distributed firewalls make no accommodation for user mobility.

The shortcomings of the existing firewalls led us to propose the Mobile Firewall shown in Fig.3(d) [13]. This approach conserves wireless access network bandwidth and protects users from attacks within or outside the subnet even as users move around. In additional to conventional packet filtering functions, the Mobile Firewall enables end users, network operators, and third party service providers to set up personalized firewall polices on a service-specific basis.

4.2 Mobile Firewall Architecture

The Mobile Firewall consists of three main network entities: the EHOST which includes mobile terminals, the Network Edge Point (NEP), and Service Administrative Server (SAS). All



(a) Perimeter firewall

(b) Distributed firewall

(c) Personal firewall

(d) Mobile firewall

Figure 3 Firewall architectures

three entities can define firewall policies using eXtensible Markup Language (XML) based description language on their own preferences and security needs.

The NEP is the point where policies are enforced, and similar to the functionality of the AR in the LWAPP, the NEP is logically the first network entity making point-to-point connections. For each legitimate EHOST, the NEP merges all rules defined by the EHOST, the SAS, or the NEP itself into a comprehensive hierarchical rule table, and filters packets for the EHOST based on this table. Using the Mobile Firewall filter transfer protocol, the EHOST and SAS can upload their rules to the NEP, and the previous NEP can transfer these rules to the next NEP as the mobile EHOST moves into a new network. The NEP can also notify the EHOST or SAS if incoming packets match with certain pre-defined traffic patterns.

It is apparent that the Mobile Firewall not only serves the objectives of traditional packet filtering but also supports personal network management and packet filtering tailored to the needs of specific services.

5. WLAN Fast Scanning

5.1 Background

The scanning process-when a mobile searches for an AP to establish a connection –is one of the most time-consuming processes in handoff [14]. The IEEE 802.11 WLAN standard provides two ways of scanning: passive and active. Passive scanning listens for beacon frames from APs. Active scanning involves a transmission of probe request frames for soliciting a probe response frame from nearby APs. When it receives beacon frames or probe response frames from an AP, the mobile gathers information about the reachability and the properties (such as capabilities, supported bit rates, and timing information) of the AP. With respect to fast channel scanning, this chapter provides a brief overview of two new technologies, adaptive beaconing and fast active scanning, which have been recently proposed in Task Group k of the IEEE 802.11 standardization working group.

5.2 Adaptive Beaconing for Fast Passive Scanning

Passive scanning has high latency. In this type of scan, the mobiles must stay on each channel for at least one beacon interval. The value of this interval is usually set to a large number (order of 100 ms) to reduce the beacon transmission overhead and the power consumption of mobiles in power-save mode.

In adaptive beaconing [15], adaptive beacons are transmitted with the frequency based on the network load. Adaptive beacons contain the same fields as those in a beacon frame but do not contain Traffic Indication Map (TIM) indicating traffic buffered for specific mobile stations in power-save mode. Mobiles doing passive scanning quickly gather information about the reachability and the properties of the AP by receiving either regular beacons or adaptive beacons. Mobiles in power save mode save power by waking up only during regular beacon transmissions.

5.3 Fast Active Scanning

Active scanning also has high latency since the mobile must stay on each channel long enough (up to 50 ms [16]) to receive probe responses from as many APs as possible (**Figure 4**(a)). Probe requests are broadcast using the Distributed Coordination Function (DCF), so there is contention among the probe responses from APs and data frames from mobiles. The contention is resolved using random backoff after a DCF InterFrame Space (DIFS).

In fast active scanning [16]-[17], mobiles are allowed to send directed probe requests to APs. These APs are selected using site reports from current AP with neighbor AP information [18]. When it receives a directed probe request, the neighbor AP acknowledges the request and then sends a probe response frame after a DIFS or Point coordination function InterFrame Space (PIFS) interval (Fig.4(b) and (c)), or if possible, sends a probe response within an Short InterFace Space (SIFS) (Fig.4(d)).

When the selected AP is reachable, the mobile receives the probe response more quickly, because unnecessary probe responses from other APs are eliminated, and the desired probe response transmission is sent with high priority using SIFS or PIFS (Fig.4(c) and (d)). When the selected AP is not reachable, the mobile learns this fact more quickly by not receiving any acknowledgement or probe response within SIFS.

5.4 Performance of Fast Scanning

With a low network load, fast active scanning is flexible and is completed in less than 1 ms [17]-[18]. With a high network load, fast active scanning takes more time and is costly in terms of bandwidth consumption, as in conventional active scanning. This is more bandwidth consuming because each mobile station performs scanning with separate exchanges for probe request

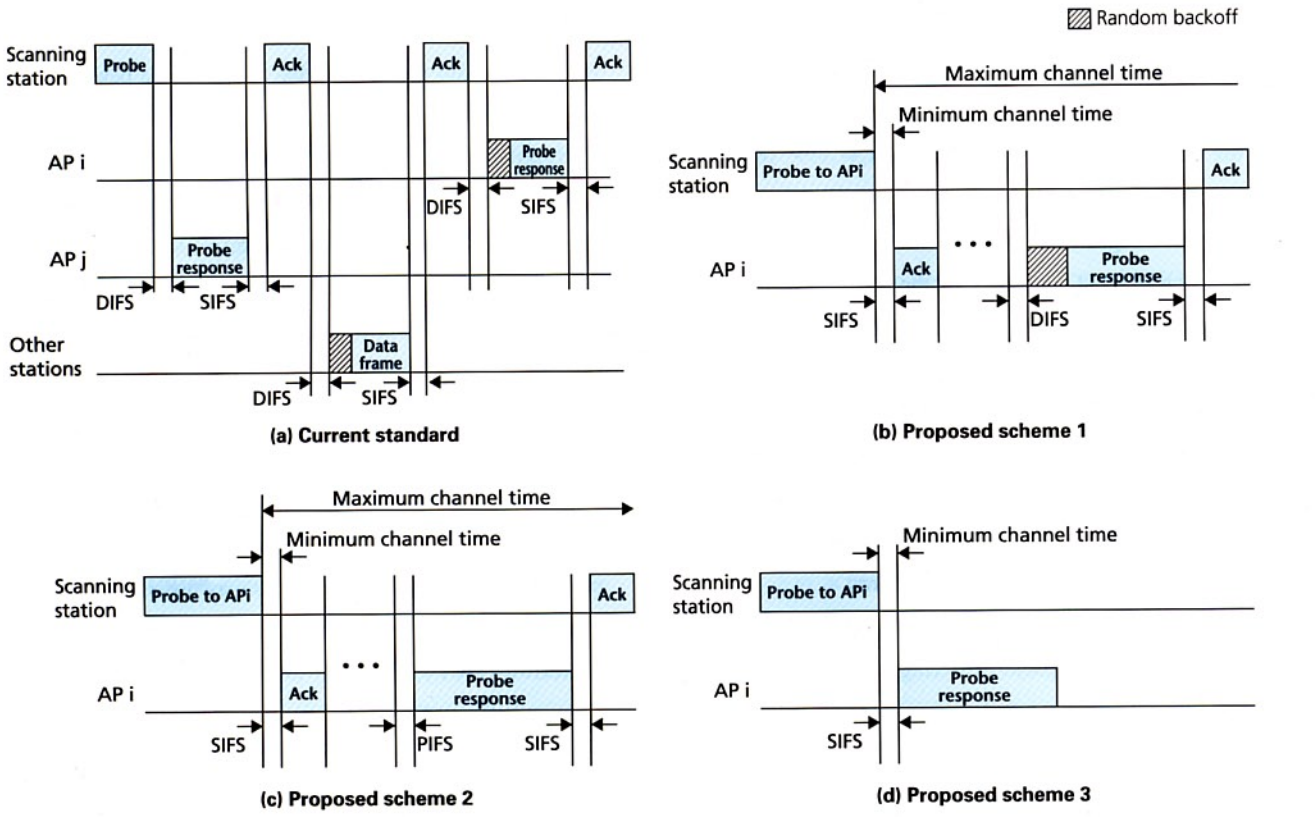


Figure 4 Improved active scanning

and probe response frames.

Adaptive beaconing has a longer scanning time, but consumes less bandwidth by trading off between the scanning time and bandwidth consumption, depending on the network load. An appropriate combination of adaptive beaconing and fast active scanning is required for further study.

6. Conclusion

This article reviewed recent research activities at DoCoMo USA Labs on IP friendly access networks, an essential part of next-generation 4G mobile networks. Specific developments covered in the article included an MPLS-based micro-mobility scheme that supports mobility and QoS control at the same time, and a Split MAC technology permitting centralized automatic setup and management of multiple APs. In addition, a Mobile Firewall was also presented that provides mobile user with a safe and secure operating environment with a full range of flexible packet filtering capabilities. Finally, a fast scanning scheme was proposed that enables seamless handoff in WLANs. In the future, we plan to build a testbed integrating the various technologies highlighted in this article to assess how they interwork in a system.

References

[1] R. Jain, et al: "(2) All-IP Network Architecture for 4G Mobile Communicatons," This issue, Vol. 5, No. 4, pp. 11–16, Mar. 2004.

[2] D. Johnson, C. Perkins and J. Arkko: "Mobility Support in IPv6," IETF work in progress, Jun. 2003.

[3] A. T. Campbell, et al: "Comparison of IP Micro-Mobility Protocols," IEEE Wireless Communications Magazine, Vol.9, No.1, Feb. 2002.

[4] J. Grimminger and H. P. Huth: "Mobile MPLS-an MPLS-based Micro Mobility Concept," Wireless World Research Forum, Meeting 3, Stockholm, Sep. 2001.

[5] Recommendation Y.MIPoMPLS: "Mobile IP Services over MPLS," ITU-TSG13, May 2003.

[6] E. Rosen, et al: "Multiprotocol Label Switching Architecture," IETF RFC 3031, Jan. 2001.

[7] X. He, D. Funato and T. Kawahara: "A Dynamic Micro-Mobility Domain Construction Scheme," The 14th IEEE PIMRC, Sep. 2003.

[8] P. Calhoun, B. O'Hara, S. Kelly, R. Suri, D. Funato and M. Vakulenko: "Light Weight Access Point Protocol," IETF work in progress, June 2003.

[9] A. Yegin, D. Funato, K. Malki, Y. Gwon, J. Kempf, M. Pettersson, P. Roberts, H. Soliman and A. Takeshita: "Supporting Optimized Handover for IP Mobility-Requirements for Underlying Systems," IETF work in progress, Jun. 2002.

[10] W. R. Cheswisk and S. M. Bellovin: "Firewall and Internet Security: Repelling the Wily Hacker," Addison-Wesley, 1994.

[11] S. M. Bellovin: "Distributed Firewalls," Login Magazine, special issue on security, Nov. 1999.

[12] S. Ioannidis, A. D. Keromytis, S. M. Bellovin and J. M. Smith:

"Implementing a Distributed Firewall," ACM Conference on Computer and Communications Security, Nov. 2000.

[13] G. Fu, D. Funato, J. Wood and T. Kawahara: "Mobile Firewall," The Fifth IFIP International Conference on Mobile and Wireless Communication Networks, Oct. 2003.

[14] A. Mishra, M. Shin and W. Arbaugh: "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," ACM Computer Communications Review, 2002.

[15] P. Orava, H. Haverinen and S. Black: "Adaptive Beaconing," IEEE 802.11-03/610, Jul. 2003.

[16] M. Jeong, F. Watanabe and T. Kawahara: "Fast Active Scan for Measurement and Handoff," IEEE802.11-03/416, May 2003.

[17] M. Jeong, F. Watanabe, T. Kawahara and Zhun Zhong: "Fast Active Scan Proposals," IEEE802.11-03/623, Jul. 2003.

[18] IEEE Std 802.11k/D0.6, Specification for Radio Resource Measurement (Draft Supplement to IEEE Std 802.11, 1999 Edition), Mar. 2003.

ABBREVIATIONS

Ack: Acknowledge
AP: Access Point
AR: Access Router
ATM: Asynchronous Transfer Mode
BU: Binding Update
CN: Correspondent Node
CoA: Care of Addresses
DCF: Distributed Coordination Function
DIFS: Distributed coordination function InterFrame Space
DoS: Denial of Service
EHOST: End HOST
FMIP: Fast Mobile Internet Protocol
HA: Home Agent
IEEE: Institute of Electrical and Electronics Engineers
IETF: Internet Engineering Task Force
IP: Internet Protocol
LAN: Local Area Network
LLC: Logical Link Control
LWAPP: Light Weight Access Point Protocol
MAC: Medium Access Control
MIP: Mobile Internet Protocol
MPLS: Multi-Protocol Label Switching
NEP: Network Edge Point
PDA: Personal Digital Assistant
PHY: PHYsical layer
PIFS: Point coordination function InterFrame Space
QoS: Quality of Service
RSSI: Receive Signal Strength Indication
SAS: multimedia Service Agent for Service control
SIFS: Short InterFrame Space
SNR: Signal to Noise Ratio
TGk: Task Group k
TIM: Traffic Indication Map
VoIP: Voice over Internet Protocol
WG: Working Group
XML: Extensible Markup Language