WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2005; **5**:805–823
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/wcm.344

Architecture, mobility management, and quality of service for integrated 3G and WLAN networks

Yang Xiao[1]*,†, Kin K. Leung[2], Yi Pan[3] and Xiaojiang Du[4]

[1]*Department of Computer Science, The University of Memphis, 373 Dunn Hall, Memphis, TN 38152, U.S.A.*
[2]*Electrical and Electronic Engineering and Computing Departments, Imperial College, Exhibition Road, London SW7 2BT, United Kingdom*
[3]*Department of Computer Science, Georgia State University, 34 Peachtree Street, Suite 1450 Atlanta, GA 30302-4110, U.S.A.*
[4]*Department of Computer Science, North Dakota State University, Fargo, ND 58105, U.S.A.*

Summary

Integration of 3G and wireless LAN (WLAN) becomes a trend in current and future wireless networks, and brings many benefits to both end users and service providers. In this paper, we provide a comprehensive survey on integration of 3G and WLAN. We discuss issues such as underline network architectures, integrated architectures, mobility management, and quality of service (QoS). We particularly study handoff QoS mapping and guarantee between 3G and WLAN, as well as how seamless voice/multimedia/data handoff becomes possible. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: IEEE 802.11; IEEE 802.11e; CDMA 2000; WCDMA; UMTS; 3G/WLAN integration; mobility management; quality of service; medium access control

1. Introduction

Integration of the IEEE 802.11 wireless LANs (WLANs) [1–4] and 3G networks, such as universal mobile telecommunications service (UMTS) [5] and code-division multiple access 2000 (CDMA 2000) [6] networks, becomes a trend in providing wireless services. The IEEE 802.11 WLANs and the 3G networks have complimentary characteristics. The IEEE 802.11 WLANs have been rapidly gaining popularity to provide high-speed wireless access for indoor networks, enterprise networks, and public hotspots. Particularly, public hotspots include airports, coffee houses, convention centers, hotels, school campuses, and libraries, which have a high demand for wireless data services. The IEEE 802.11 WLANs' success is partially due to low price of WLAN cards and high data rates. The original IEEE 802.11 standard specified in 1997 is for the 2.4 GHz unlicensed band with data rates up to 2 Mbps [1]. The IEEE 802.11b and 802.11a standards specified in 1999 can provided up to 11 and 54 Mbps using the 2.4 and 5 GHz bands, respectively [2,3]. The IEEE 802.11g specified in 2003 can provide up to 54 Mbps in the 2.4 GHz band. The emerging IEEE 802.11n standard will provide much higher data rates for the IEEE 802.11

*Correspondence to: Yang Xiao, Department of Computer Science, The University of Memphis, 373 Dunn Hall, Memphis, TN 38152, U.S.A.
†E-mail: yangxiao@ieee.org

extension [32–34]. However, an IEEE 802.11 basic service set (BSS) covers only a few thousand square meters with no specific mobility support. On the other hand, the 3G standards, the CDMA 2000 and UMTS, provide wide-area coverage of several kilometers in cell radius with high mobility management support, but with relatively low data rate from 64 Kbps to 2 Mbps (a theoretical maximum). The UMTS adopts wideband code-division multiplexing access (WCDMA) technology, providing much better services and higher data rates than 2G or 2.5G cellular networks. NTT DoCoMo launched the world's first WCDMA network in 2001 in Japan, making the UMTS a reality, and commercial WCDMA networks are now operating in Australia, Austria, Italy, Sweden, and the United Kingdom [21]. Although wireless cellular systems provide most of public wireless access, WLAN systems have rapidly become an important broadband public wireless access. Furthermore, the costs of obtaining radio spectrum as well as devices for 3G networks are very expensive, whereas, WLANs use license-free radio spectrum to provide higher speed wireless services. Therefore, 3G and WLAN are complementary technologies, and integration of WLAN and 3G provides benefits to both the end users and service providers with advantages of both technologies.

There are two operation modes in the IEEE 802.11 standards: infrastructure mode and ad hoc mode [1]. In an infrastructure mode network, an access point (AP) is present and terminals can communicate only with an AP at a given time, whereas in an ad hoc mode network, no AP is defined and terminals can communicate with each other as long as the radio link between them can support the communication. Typically, infrastructure mode networks are used for integration of 3G and WLAN. To access an 802.11 WLAN, a station needs to go through authentication and association procedures first. The packet transmissions between the AP and stations can be optionally protected using a symmetric key based RC4 encryption called wired equivalency privacy (WEP) [1]. APs connect distributed system, which connects Internet with TCP/IP protocol suite. A dynamic host configuration protocol (DHCP) server is needed for configuration of the WLAN MS's IP stack. An MS in WLAN is typically a laptop computer or a PDA with a built-in WLAN module or a PCMCIA card [14]. In a BSS of WLAN, access point acts as a bridge for the wired and wireless parts of the network. The 802.11 standard defines only the MAC and physical layers, and hence the authentication procedures, QoS, and mobility management mechanisms, if available, vary from provider to provider. Since the 3G standards also define protocols above the MAC layer, they can handle all these functionalities. Furthermore, WLAN lacks capabilities such as subscription and roaming services, which are provided by the 3G networks. Compatibility issues also rise when extending these characteristics to the WLAN. In addition, providing consistent QoS control over an integrated 3G/WLAN network is necessary [21].

Many researchers have been reported their work on the integration of 3G and the 802.11 WLAN during the last few years [8–11,14,18–28]. Furthermore, the third generation partnership project (3GPP) also develops a cellular–WLAN interworking architecture for the 3GPP cellular system standards [14,15]. 3GPP2 has recently started to examine the issues of 3G/WLAN interworking. In this paper, we provide a comprehensive survey on the integration of 3G and WLAN. We cover issues such as underline network architectures, integrated architectures, mobility management, and QoS. Furthermore, most of the existing work focuses on the protocol design or providing a preliminary evaluation for data sessions [8–11, 14, 18–28]. In this paper, we also consider how resource management works, how the UMTS QoS is mapped into the WLAN QoS, as well as how QoS can be guaranteed in the integrated 3G and the emerging 802.11e WLAN networks.

The rest of the paper is organized as follows. Section 2 introduces networks of UMTS, CDMA 2000, 802.11 WLAN, and Mobile IP. Section 3 discusses integrated architectures and mobility management for integrated 3G/WLAN networks. In Section 4, we discuss QoS for integrated 3G/WLAN networks. We conclude this paper in Section 5, along with some future research directions.

2. Underline Network Architectures

In this section, we briefly discuss two 3G network standards (UMTS and CDMA2000), the IEEE 802.11 WLAN, and Mobile IP.

2.1. UMTS and CDMA 2000

The International Telecommunication Union (ITU) requires the 3G networks to support 144 Kbps as the minimum transmission rate in mobile (outdoor) and 2 Mbps in fixed (indoor) environments [21]. In both the 3GPP and 3GPP2, 3G/WLAN integration is considered. 3GPP is a worldwide specification forum responsible for GSM and UMTS specifications, and 3GPP2 is another worldwide specification forum responsible for CDMA2000 technology specifications. In September 2001, 3GPP decided to study the UMTS/WLAN integration. 3GPP has currently specified an interworking architecture that enables users to access their 2G and 3G data services from WLANs. 3GPP2 has recently started to examine the issues of 3G/WLAN interworking.

The 3GPP also proposes UMTS all-IP architecture to integrate IP and wireless technologies [29]. All-IP architecture for UMTS evolves from global system for mobile communications (GSM), general packet radio service (GPRS), UMTS release 1999 (UMTS R99), and UMTS release 2000 (UMTS R00). UMTS R00 has two releases: Release 4 for the next-generation network architecture for the circuit-switched (CS) domain, and Release 5 for the IP multimedia subsystem on the topic of the packet-switched (PS) domain. In Reference [29], the authors introduce two options for all-IP architectures: Option 1 architecture supports PS domain multimedia and data service; Option 2 architecture extends Option 1 network by accommodating CS domain voice services over a packet-switched core network. Furthermore, 3GPP proposes UMTS traffic classes and QoS parameters in [30]. We will further discuss UMTS QoS in Section 5.

The CDMA2000 is another 3G standard, a solution evolved from the North American Standard IS-95 [21]. The first phase of the CDMA2000 is called CDMA2000 1X with data rate of 144 Kbps, using the same spectral bandwidth as IS-95. The CDMA2000 1xEV (EV for evolution) is the second phase of the 3G network to provide 2 Mbps data rate. SK Telecom (Korea) launched the first CDMA2000 commercial system in October 2000, and now the CDMA2000 1X has been deployed in Asia, North America, South America, and Europe. SK Telecom and KT Freetel launched CDMA2000 1xEV-DO (data only) in 2002 [21]. Verizon also uses the EV-DO network to provide data service in several major U.S. cities.

In a CDMA2000 cellular network shown in Figure 1, multiple base stations (BSs) are connected a radio network controller (RNC) via T1/T3 lines [6]. In turn, each RNC is connected to a packet data serving node (PDSN) via a packet control function (PCF). The purpose of the PCF is to control transmission of packets between BSs and the PDSN. Between a mobile station (MS) and the RNC, the radio link protocol (RLP) defined in the CDMA2000 standard is used to control data transport between an MS and the RNC. On the other hand, the point-to-point protocol (PPP) is employed between the MS and the PDSN. Multiple RLP sessions between MSs and the RNC are

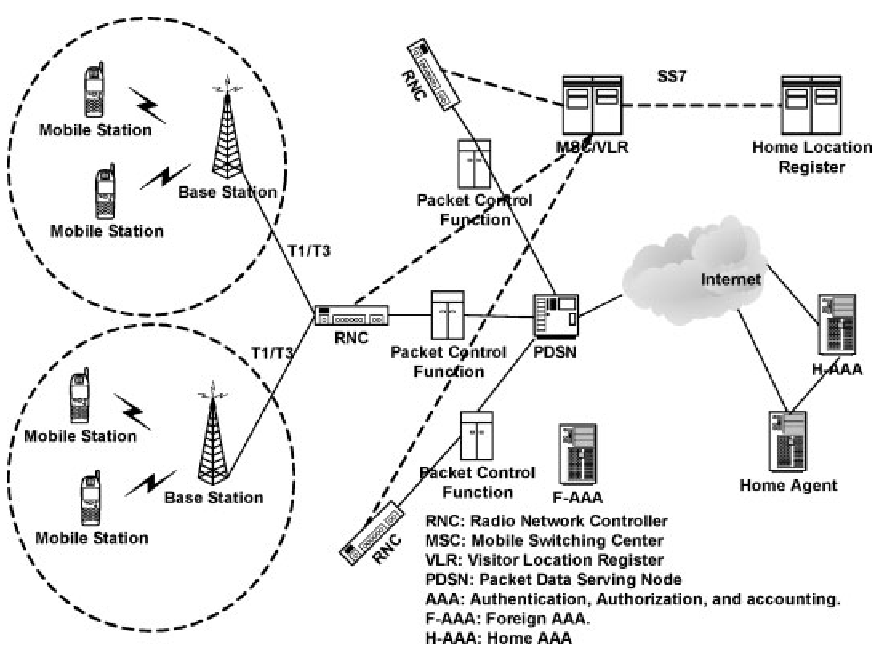

Fig. 1. A code division multiple access (CDMA) 2000 network.

handled by the RNC to share the 144 Kbps carrier throughput in the 3G-1X cellular networks. If an MS moves from one RNC to the other, the corresponding RLP session is disconnected and a new session needs to be established with the new RNC. The PDSN connects Internet. Foreign Agent (FA) function of mobile IP [7] is implemented in the PDSN for inter-PDSN mobility.

2.2. IEEE 802.11 WLAN

The IEEE 802.11 PHY [1] include three different physical layer implementations: frequency hopping spread spectrum (FHSS), direct sequence spread spectrum (DSSS), and infrared (IR). Both the FHSS and the DSSS utilize the 2.4 GHz industrial, scientific, and medical (ISM) band. The FHSS adopts two-level Gaussian frequency shift keying (GFSK), and the DSSS adopts differential binary phase shift keying (DBPSK) and differential quadrature phase shift keying (DQPSK). The IR specification is designed for indoor use only and operates with nondirected transmissions with 16-pulse position modulation (PPM) and 4-PPM modulation. The IEEE 802.11b PHY [2] uses complementary code keying (CCK) and DSSS modulation schemes at 2.4 GHz. The IEEE 802.11a PHY uses a convolutionally coded adaptation of orthogonal frequency division multiplexing (OFDM) for encoding and transmission called coded OFDM-(COFDM), which is a frequency division multiplexed (FDM) multicarrier communications scheme that includes the application of convolutional coding to achieve higher raw data rates. IEEE 802.11g is a superset of the 802.11b PHY, including the 802.11b modulation schemes and the OFDM schemes originally defined for 802.11a PHY at the 5 GHz band.

IEEE 802.11 medium access control (MAC) employs a mandatory contention-based channel access function called distributed coordination function (DCF), and an optional centrally controlled channel access function called point coordination function (PCF) [1]. The DCF adopts a carrier sense multiple access with collision avoidance (CSMA/CA) with binary exponential backoff, and the PCF is based on a polling mechanism.

The DCF and the optional PCF determine when a station, operating within a basic service set (BSS) or independent BSS (IBSS), is permitted to transmit. There are two types of 802.11 networks: infrastructure network, i.e., BSS, in which an access point (AP) is present and ad hoc network, i.e. IBSS, in which an AP is not present. In the long run, time is always divided into repetition intervals called superframes. Each superframe starts with a beacon frame, and the remaining time is further divided into an optional contention-free period (CFP) and a contention period (CP). The DCF works during the CP and the PCF works during the CFP. The DCF defines a basic access mechanism and an optional request-to-send/clear-to-send (RTS/CTS) mechanism. In the DCF, a station with a frame to transmit monitors the channel activities until an idle period equal to a distributed inter-frame space (DIFS) is detected. After sensing an idle DIFS, the station waits for a random backoff interval before transmitting. The backoff time counter is decremented in terms of slot time as long as the channel is sensed idle. The counter is stopped when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for more than a DIFS. The station transmits its frame when the backoff time reaches zero. At each transmission, the backoff time is uniformly chosen in the range [0, $CW - 1$] in terms of timeslots, where CW is the current backoff window size. At the very first transmission attempt, CW equals the initial backoff window size CW_{min}. After each unsuccessful transmission, CW is doubled until a maximum backoff window size value CW_{max} is reached. After the destination station successfully receives the frame, it transmits an acknowledgment frame (ACK) following a short inter-frame space (SIFS) time. If the transmitting station does not receive the ACK within a specified ACK timeout, or it detects the transmission of a different frame on the channel, it reschedules the frame transmission according to the above backoff rules. The above mechanism is called the basic access mechanism. To reduce the hidden station problem, an optional four-way data transmission mechanism called RTS/CTS is also defined in the DCF. In the RTS/CTS mechanism, before transmitting a data frame, a short RTS frame is transmitted. The RTS frame also follows the backoff rules introduced above. If the RTS frame succeeds, the receiver station responds with a short CTS frame. Then a data frame and an ACK frame will follow. All four frames (RTS, CTS, data, ACK) are separated by an SIFS time. In other words, the short RTS and CTS frames reserve the channel for the data frame transmission, which follows.

The PCF is an optional centrally controlled channel access function, which provides contention-free (CF) frame transfer. The PCF is designed for supporting time-bounded services, which can provide limited

QoS. It logically sits on top of the DCF, and performs polling, enabling polled stations to transmit without contending for the channel. It has a higher priority than the DCF by adopting a shorter inter-frame space (IFS) called point inter-frame space (PIFS). Under the PCF, the AP sends a poll frame to a station to ask for transmitting a frame. The poll frame may or may not include data to that station. After receiving the poll frame from the PC, the station with a frame to transmit may choose to transmit a frame after a SIFS time.

2.3. Mobile IP

Mobile IP [7] preserves user sessions when a user roams among heterogeneous networks. It allows a user to maintain the same IP address and maintains connections while roaming between IP networks. In mobile IP, an MS keeps a fixed IP address in the home network called home address (HoA). Two agents, a home agent (HA) in the home network and a foreign agent (FA) in the visited network, are adopted. Both the HA and the FA are routers with some defined functions. An MS in a visited network, registers the local FA's address in its HA as a care-of address (CoA). The HA maintains an association between the MS's home IP address and its CoA, and forwards packets from any correspondent node (CN) to the MS through tunneling encapsulated IP packets to the FA, which forwards packets to the MS. The MS sends packets using its home IP address, even in a visited network.

In mobile IP, all messages between the MS and the HA, protected by a 128-bit symmetric key, and authenticated by a keyed message digest algorithm 5 (MD5) in 'prefix + suffix' mode. Hash-based message authentication code (HMAC-MD5) is also supported. Between the MS and the FA, optional authentication can be used. An identification field (32 bits), used as a timestamp and changed each time, and sequence number are used as replay protection.

AAA (authentication, authorization, accounting) servers, such as RADIUS (remote authentication dial-in user service) [36] and DIAMETER [37], are used for authentication and authorization for Mobile IP [35].

3. WLAN/3G Integration Architectures and Mobility Management

There are several WLAN/3G integration architectures reported in the literature, based on the amount of interdependence between WLANs and 3G networks [11]. Integration architectures include tightly-coupled integration, loosely-coupled integration, peer integration, and hybrid-coupled integration [8–11,25,27]. In the tightly coupled integration, the 802.11 network appears to the 3G core network as another 3G access network, whereas in the loosely-coupled integration, the 802.11 network connects the 3G core network via Internet. In Reference [10], the authors also introduce a peer network approach, in which the 802.11 network acts a peer network. In Reference [27], a hybrid-coupled integration is proposed to differentiate the data paths according to the type of traffic.

3.1. Tightly-coupled Integration

We will discuss tightly-coupled integration for WLAN/CDMA2000 and WLAN/UMTS in the following two subsections.

3.1.1. CDMS 2000 and WLAN

In the tightly-coupled integration, a WLAN emulates functions of a 3G radio access network, and therefore is treated as a 3G access network from the point view of the 3G core network.

Figure 2 illustrates both the tightly-coupled integration and the loosely-coupled integration in CDMA2000. In the tightly-coupled integration, a RNC/packet-control-function emulation unit is needed to connect PDSN and access point. The emulation function unit implements all 3G protocols such as mobility management, authentication, etc., hiding the details of the WLAN. MSs in the tightly-coupled integration implements both 3G and WLAN interfaces in the physical layer. Furthermore, the 3G protocol stack should be implemented on top of WLAN standard in these MSs. WLAN and 3G networks use the same authentication, signaling, and billing functions.

Drawbacks of the tightly-coupled integration include complexity and high cost: (1) both WLAN and 3G networks should be owned by the same operator; (2) both WLAN and 3G devices and configurations should be modified; (3) MSs need both the physical layer and the upper layer modifications; and (4) wireless cards in MSs become expensive since both interfaces are needed. One advantage of the tightly-coupled integration is that it is more easily to control quality of service (QoS) for time-sensitive traffic.

The PDSN in CDMA2000 as shown in Figure 2 implements mobile IP to support inter-PDSN handoff.

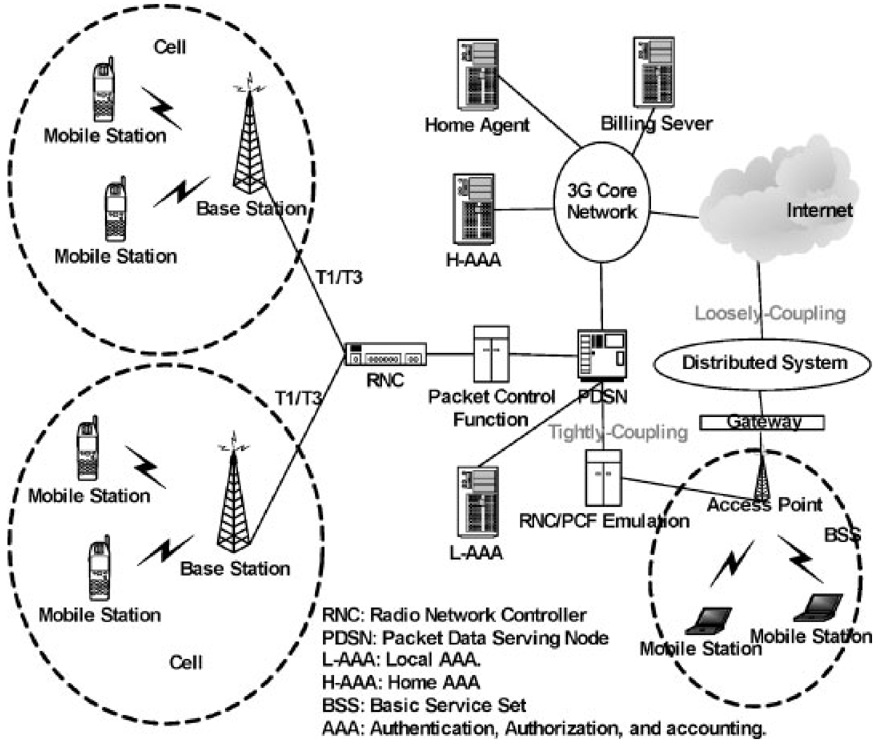

Fig. 2. Tightly-coupled and loosely-coupled in CDMA2000.

The MS conducts handoffs when its signal in one wireless network is weak or when it finds a better wireless signal in another wireless network.

3.1.2. UMTS and WLAN

In this subsection, we discuss integration of UMTS and WLAN, which is proposed in Reference [26]. An architecture integrating the UMTS cellular network and 802.11 WLAN allow an MS to maintain connections to the WLAN and UMTS simultaneously: a packet data service through WLAN and a circuit switched voice service through UMTS. UMTS provides packet-switching (PS) and circuit-switching (CS) services, and GPRS is integrated into UMTS for packet data service. In UMTS, RNC and nodes B's constitute radio access network (RAN) called UMTS RNS. Each node B has a cluster of base stations. Several node B's connect to a RNC. The core network (CN) is comprised of SGSN (serving GPRS support node) and GGSN (gateway GPRS support node). The RNC acts as a mediator for converting the radio frames to IP packets and IP packets to radio frames via SGSN. The IP packets are tunneled between GGSN and SGSN and then between SGSN and RNCs. Several APs are connected to the IP routed network via AR. When an MS moves across APs connected to the same AR, intra-AR handoff takes place and is handled by the AR. During inter-AR handoff, the new AR performs the IP handover. Mobile IP handles any intra-domain mobility. The IP service layer is used to exchange the device specific context information with higher layers and also provides synchronization between the two device drivers. The IP layer protocols implement RSVP (Resource Reservation Setup Protocol) and MMP (mobility management protocol) for QoS signaling and reservation, and mobility management in WLAN network, respectively. The 802.11 device driver implements 802.11 MAC control functions and the UMTS device driver implements GPRS user and control plane protocols.

When an MS is powered in UMTS, it receives beacons from UMTS and thus activates the UMTS interface, and if the MS is powered in 802.11, it can be connected to both UMTS and 802.11 WLAN. In this case, it receives beacons from UMTS and 802.11. However, since UMTS provides basic wireless service, it runs UMTS-GPRS power up procedure through UMTS interface ignoring the beacons from

802.11 interfaces. After UMTS power up procedure, the MS responds to the 802.11 beacons by sending an association request. After associated with the AP, UMTS-WLAN handover procedure handovers the PS connection to the WLAN network. In WLAN, the MS uses the same IP address obtained from GGSN in UMTS. The MS obtains a temporary address which the packets are tunneled via SGSN. In GPRS, the PDP (packet data protocol) context signaling is used to set up the connection and reserve the resources. The RAB (radio access bearers) signaling is used to set up radio channels and reserve radio resources between SGSN and RNC. The MS communicates with the SGSN for PDP context setup, which in turn coordinates with the GGSN and RAB. In WLAN, the RSVP is used for resource reservation. After the MS sends a RSVP PATH message to SGSN, the SGSN negotiates the session setup with the GGSN using PDP context messages and responds to the MS with RSVP RESV message. GPRS mobility context is the MS's mobility context within UMTS and is stored both at the SGSN and the MS. WLAN mobility context is the MS mobility context within WLAN and is stored both at AR and the MS. In WLAN, the MS is connected to the GPRS and maintains both UMTS and WLAN mobility context. The architecture allows an MS to maintain a PS connection through WLAN and CS connection through UMTS simultaneously.

3.2. Loosely-Coupled Integration

In the loosely-coupled approach, shown in Figure 2, AP connects a gateway, which connects the distributed system and then internet. An MS contacts 3G via AP, gateway, distributed system, and internet. The gateway implements mobile IP and AAA service to interwork with the 3G's home AAA servers. The AAA service enables exchanging accounting information and billing information between a 3G network and a WLAN.

The PDSN in CDMA2000 as shown in Figure 2 implements Mobile IP to support inter-PDSN handoff. The WLAN gateway in Figure 2 also needs to implement mobile IP. The MS conducts handoffs when its signal in one wireless network is weak or when it finds a better wireless signal in another wireless network.

Advantages of the loosely-coupled integration include less complexity and low cost: (1) independent ownership, deployment, and traffic engineering of the WLAN and 3G networks; (2) fewer modifications on 3G networks; (3) fewer efforts on billing and accounting issues. One drawback of the loosely-coupled integration is that it is very difficult to provide QoS guarantee for time-sensitive traffic since Internet QoS itself is difficult to guarantee.

3.3. Peer Integration

In Reference [25], a peer integration scheme is proposed, and we summarized it as follows. The mobile IP is implemented, and users may subscriber to either the IEEE 802.11 WLAN or UMTS. UMTS CN (core network) includes the HA as well as functionality of AAA servers. In the 802.11 WLAN, multiple extended service sets (ESSs) are connected to access gateway (AGW) which interfaces to the core IP network. The IEEE 802.11 WLAN includes AAA and the HA to support mobility to other peer networks and it also includes emulator of HLR to support mobility to UMTS.

When an MS is subscribed to UMTS and roams to an IEEE 802.11 WLAN network, the MS associates with an AP first and then performs AAA functions with a local AAA server. After authentication, authorization, and obtaining a CoA, the MS sends a binding update, and the HA sends binding acknowledgement to the MS. A location update message is sent to the HSS (home subscriber server) by the HA, and the HSS cancels location with the previous SGSN, which initiates the deletion of PDP context with the GGSN. Packets arriving from the CN to the GGSN are tunneled to the MS. The MS sends a binding update to the CN so that the subsequent packets are directly sent to the MS instead of tunneling to the MS. The MS establishes radio bearers for UTRAN (universal terrestrial radio access network) when returning back to the UMTS and performs UMTS attach followed by PDP context activation. There is a possibility that the MS may have to go through authentication process. The MS then performs Mobile IPv6 registration by sending the binding update to the HA, which then sends a binding acknowledge message after updating the binding cache. binding update message is also sent to the previously serving router to forward any packets destined to the MS. For the CN to send packets directly to the MS, the MS sends a binding update to the CN.

When the MS subscribed to an IEEE 802.11 network is performing a handoff to UMTS network, the MS establishes UTRAN radio bearers and sends an attach message. The SGSN interacts with the HLR of IEEE 802.11 WLAN to authenticate the MS, which then performs PDP context activation with the UMTS network, and completes mobile IPv6 procedures by forming a new CoA and exchange of binding update/acknowledge with its HA in IEEE 802.11 WLAN. Packets arriving in IEEE 802.11 WLAN are tunneled to the GGSN and further tunneled to the MS. When the MS is returning to the 802.11 WLAN, it first associates with an AP and AAA functions are performed with the home AAA server. The MS then performs mobile IPv6 procedures by exchanging binding update/acknowledge with the HA and also sends binding update to the previous GGSN. The GGSN deletes the PDP contexts and any packets to the MS are tunneled. The subsequent packets are forwarded directly to the MS, and the MS sends binding update to the CN.

Mobility management procedures make the integration of UMTS and 802.11 WLAN more effective. In peer integration, the two infrastructures are independent and just AAA linkage is added and hence is least complex. Mobility management is least complex for peer networking and a bit more complex for tight coupling and the most complex for loose coupling architecture.

3.4. Hybrid-Coupled Integration

The tightly-coupled integration has a drawback that the capacity of UMTS core network nodes is not enough to accommodate the bulky data traffic from WLAN since the core network nodes are designed to handle circuit voice calls or short packets, whereas it is difficult for the loosely-coupled integration to support service continuity to other access network during handover, and both the handover latency and packet loss are large [27]. In Reference [27], a hybrid-coupled integration is proposed, and the scheme differentiates the data paths according to the type of the traffic and can accommodate traffic from WLAN efficiently with guaranteed seamless mobility. The hybrid coupled scheme detours different types of traffic by using different traffic paths: real-time traffic such as voice packets use the path of APGW (access-pointer gateway)-SGSN-GGSN, which is a tightly-coupled integration; and non-real time traffic such as FTP traffic uses the path of APGW-AR-HA, which is a loosely-coupled integration [27]. In the loosely-coupled integration, real-time traffic has longer delay with a higher packet losing rate. The tightly-coupled integration provides low delay and low packet loss, but it cannot handle large data traffic. The hybrid-coupled scheme can transmit large and non-real time traffic through WLAN to internet, solving the problem that non-real time traffic uses the core network of UTMS. The hybrid-coupled scheme combines the advantages of the UMTS and WLAN to provide seamless handover, low dropping probability and packet loss probability. Two FAs, one for WLAN and another for UTMS, are used to provide user's mobile IP service. To support the binding option of mobile IP, the HA implements complementary functionality: (1) the default operation of binding option is implemented; and (2) HA can send packets to the appropriate FA according to the traffic type specified by the IP header of packets. RNC makes the decision to begin the handover and send the relocation required message to SGSN, which forwards this message to suitable APGW. When user equipment (UE) accesses to an AP, APGW set up radio access bearer. Then, GGSN update PDP context with UE. IP mobility is supported by mobile IP and the binding option. Therefore, when UEs in UMTS, they use FA provided by UMTS. When UEs are in WLAN, they use two FAs. There are in UMTS and WLAN respectively. It is no need to register to GGSN, when handover to UMTS. But it is need to register to the FA in WLAN, when handovering to WLAN [27].

3.5. Vertical Handoff

In Reference [20], authors propose a vertical handoff scheme, in which, MSs utilize high-bandwidth WLANs in hotspots and switch to 3G cellular networks when the coverage of WLAN is not available or the network condition in WLAN is not good enough, and a virtual connectivity manager that uses an end-to-end principle to maintain a connection without additional network infrastructure support. In a seamless vertical handoff, the handoff procedure should be transparent to upper-layer applications, and bases on handoff metrics and handoff decision algorithms. Horizontal handoff, defined as handoff between base stations (BSs) or between APs, is much easier than vertical handoff [20], which has the following difficult issues: when an MS moves from 3G to WLAN, the handoff cannot be triggered by signal decay of the current system, as in horizontal handoff, and there is no comparable signal strength available to aid the decision as in horizontal handoff [20]. Therefore, network conditions such as available bandwidth and delay and user preference are used rather than the physical layer parameters such as received signal strength and signal-to-interference ratio. Mobile IP can be used for mobility management after a vertical handoff. In the proposed system [20] integrates a connection manager that intelligently detects the wireless network changes and a virtual connectivity manager that maintains connectivity using the end-to-end principle.

3.6. DHCP-Based Mobility Support

In Reference [24], the authors extend DHCP to address issues, involved in global mobility among heterogeneous access networks using IP protocol, such as subnet detection and terminal configuration, based on the external triggers. DHCP supports host registration and configuration, and is an UDP-based client/server protocol for assigning IP addresses dynamically for a lease time period. The DHCP client requests the DHCP server for network parameters, and the DHCP server responds to the client's requests and may not be on the same link as the client. The DHCP relay acts as a relay for exchanging messages between a client and the server, and it is always located on the same link as the client. When the DHCP client broadcasts the DHCP DISCOVER message, including the lease time and options for network address, on its local subset, the relay agents pass the message to the DHCP servers not on the same subset. The relay agent transmits the DHCP OFFER message from each server including the IP address. After receiving one or more messages, the client chooses one server among them and broadcasts the DHCP REQUEST indicating the chosen server, which responds with a DHCP ACK message to the client containing the configuration parameters. After the lease period, the client can extend the lease following the renew procedure. After configuration in DHCP, clients go into a state of not listening to DHCP messages until the renewal of the lease. This fact makes difficult supporting mobility, and the client should be able to detect subnet or access network changes independently of the lease time associated to the current IP address [24]. In Reference [24], two proposed external triggers allow the client to jump into active DHCP state such as sending a DHCP DISCOVER message and detect the changes in subnet or IP access point to reconfigure. In Layer-2 triggers, the handoff indicators allow the client to jump into an active DHCP state by 'movement detection' mechanisms [24]. In Layer-3 triggers, a router broadcasts ICMP (internet control message protocol) messages periodically to the hosts on the network containing the IP addresses of the routers and their preference level, and the host verifies the IP address to its current address to determine the change of subnet and can start the DHCP procedure for reconfiguration [24].

4. Quality of Service

In this section, we study how seamless voice/multimedia/data handoff becomes possible and how QoS can be mapped and guaranteed in integrated 3G/WLAN networks with the emerging IEEE 802.11e standard. The UMTS/WLAN integration is based on the all-IP architectures of UMTS [29]. However, it can be easily adapted to non-all-IP architectures of UMTS. Furthermore, we adopt the hybrid coordination function's controlled channel access mechanism to provide seamless voice/multimedia/data handoff with QoS guaranteed between a WLAN BSS and a cell in a cellular network. We only consider tightly coupled integration in this section since it is difficult to guarantee QoS in the loosely coupled integration due to the fact that QoS depends on internet.

In Subsections 4.1 and 4.2, we introduce the IEEE 802.11e MAC and UMTS QoS, respectively. 3G/WLAN architecture with all-IP UMTS is presented in Section 4.3. Mobility and Handoff are discussed in Subsection 4.4. Subsection 4.5 studies QoS mapping between UMTS and WLAN. Resource management is discussed in Subsection 4.6.

4.1. IEEE 802.11e MAC

To support MAC-level QoS, the IEEE 802.11 working group is currently working on the IEEE 802.11e standard [4], which is in the final stage for approval. The emerging IEEE 802.11e standard provides QoS features and multimedia support to the existing 802.11b [2] and 802.11a [3] standards, while maintaining full backward compatibility with them. The IEEE 802.11e MAC employs a channel access function, called hybrid coordination function (HCF), which includes a contention-based channel access and a contention-free centrally controlled channel access mechanism. The contention-based channel is also referred to as enhanced distributed channel access (EDCA). The EDCA provides a priority scheme by differentiating the inter-frame space, the initial and the maximum window sizes for backoff procedures. The HCF controlled channel access mechanism (HCCA) is based on a polling mechanism with some enhanced QoS-specific mechanisms and frame subtypes to support data QoS during collision-free periods.

In the IEEE 802.11e, the QoS enhancements are available to QoS enhanced stations (QSTAs) which can be associated with a QoS enhanced access point (QAP) in a QoS basic service set (QBSS), or can be in a QoS independent basic service set (QIBSS) without a QAP. A collision free period (CFP) under the HCCA and a collision period (CP) under the EDCA alternate over time, and time is always divided into repetition intervals called superframes. Each superframe starts with a beacon frame, and the remaining time is further divided into a CFP and a CP. The EDCA works during the CP and the HCCA works during the CFP.

The concept transmission opportunity (TXOP) is introduced in the IEEE 802.11e for both the EDCA and the HCCA. A TXOP is a time period when a station has the right to initiate transmissions onto the wireless medium. It is defined by a starting time and a maximum duration. A station cannot transmit a frame that extends beyond a TXOP.

The HCCA allows for the reservation of TXOPs with a hybrid coordinator (HC), a type of point coordinator handling rules defined by the HCF. The HC, collocated with a QAP, performs bandwidth management including the allocation of TXOPs to QSTAs. The HC can transmit the beacon frame to initiate a CFP when it senses the medium idle for a point inter-frame space (PIFS) interval, and terminate the CFP by transmitting a CF-end frame when it senses the medium idle for a PIFS interval.

A QSTA based on its requirements requests the HC for TXOPs for both its own transmissions and transmissions from the HC to itself. QSTAs may send TXOP requests using the QoS control field in the frame directed to the QAP, with the request duration or queue size indicated to the QAP. The HC, based on an admission control policy either accepts or rejects the request. If the request is accepted, it schedules TXOPs for the QSTA. For transmissions for the station, the HC polls a QSTA based on the parameters supplied by the QSTA at the time of its request. For transmissions to the QSTA, the HC queues the frames and delivers them periodically, again based on the parameters supplied by the QSTA. This mechanism is used for applications such as voice and video, which may need a periodic service from the HC. Readers please refer to Reference [4] for more information about IEEE 802.11e.

4.2. UMTS QoS

Network services in UMTS are considered end-to-end [30]. A UMTS bearer service layered architecture is depicted in Figure 3, in which each bearer service on a specific layer offers it is individual services using services provided by the layers below. A terminal equipment (TE) is connected to the UMTS network by use of a mobile termination (MT). The end-to-end service on the application level uses the bearer services of the underlying network(s). The end-to-end-service used by the TE will be realized using a TE/MT

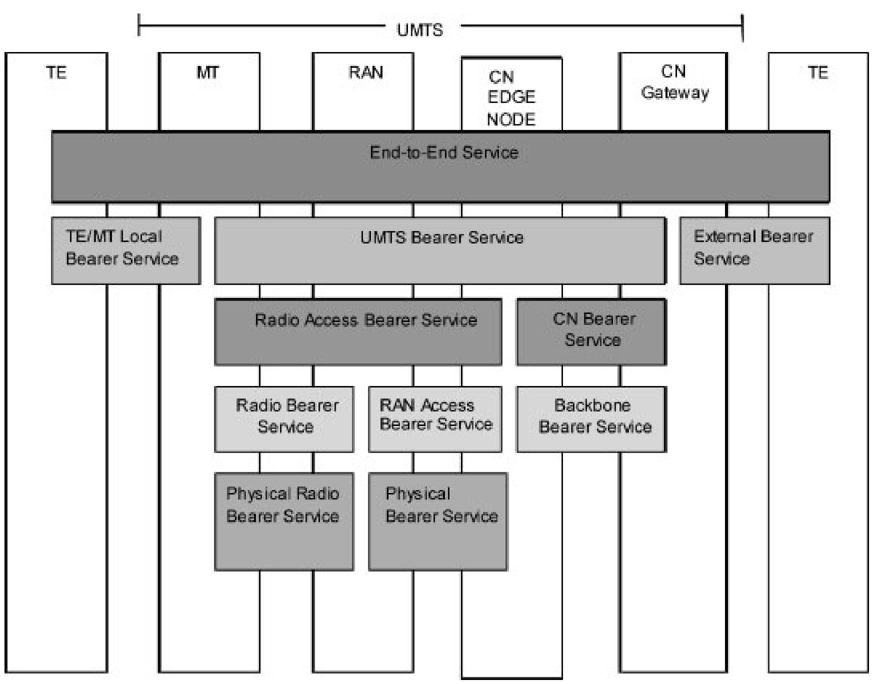

Fig. 3. UMTS QoS architecture.

local bearer service, a UMTS bearer service, and an external bearer service. The UMTS bearer service consists of two parts, the radio access bearer service and the core network bearer service. The radio access bearer service provides confidential transport of signaling and user data between MT and CN edge node with the QoS adequate to the negotiated UMTS bearer service or with the default QoS for signaling. The core network bearer service of the UMTS core network connects the UMTS CN edge node with the CN gateway to the external network. The role of this service is to efficiently control and utilize the backbone network in order to provide the contracted UMTS bearer service. The UMTS packet core network shall support different backbone bearer services for variety of QoS. A radio bearer service and an RAN access bearer service realize the radio access bearer service. The radio bearer service covers all the aspects of the radio interface transport. The RAN access bearer service together with the physical bearer service provides the transport between RAN and CN. RAN Access bearer services for packet traffic shall provide different bearer services for variety of QoS.

4.2.1. UMTS QoS traffic class

There are four different UMTS QoS traffic classes: conversational class, streaming class, interactive class, and background class, shown in Table I [30]. Conversational class is the most delay sensitive traffic class, whereas background class is the most delay insensitive traffic class. Conversational and streaming classes are used to carry real-time traffic flows, whereas interactive and background classes are mainly used for best-effort traffics like WWW, email, Telnet, and FTP. Interactive and background classes provide better error rate by means of channel coding and retransmission than Conversational and streaming classes due to looser delay requirements.

4.2.2. UMTS QoS parameters/attributes

There are many QoS parameters/attributes defined for UMTS: maximum bitrate (kbps), guaranteed bitrate (kbps), delivery order (y/n), maximum SDU (service data unit) size (octets), SDU format information (bits), SDU error ratio, residual bit error ratio, Delivery of erroneous SDUs (y/n/-), transfer delay (ms), traffic handling priority, allocation/retention priority, source statistics descriptor ('speech'/'unknown'), and Signaling indication (yes/no).

Maximum bitrate is the maximum number of bits delivered by UMTS and to UMTS at a SAP within a period of time, divided by the duration of the period. The traffic is conformant with maximum bitrate as long as it follows a token bucket algorithm where token rate equals maximum bitrate and bucket size equals maximum SDU size. The maximum bitrate is the upper limit a user or application can accept or provide. All UMTS bearer service attributes may be fulfilled for traffic up to the maximum bitrate depending on the network conditions. Guaranteed bitrate is the guaranteed number of bits delivered by UMTS at a SAP within a period of time (provided that there is data to deliver), divided by the duration of the period. The traffic is conformant with the guaranteed bitrate as long as it follows a token bucket algorithm where token rate equals Guaranteed bitrate and bucket size equals maximum SDU size. UMTS bearer service attributes, for example, delay and reliability attributes, are guaranteed for traffic up to the guaranteed bitrate. For the traffic exceeding the guaranteed bitrate the UMTS bearer service attributes are not guaranteed. Delivery order (y/n) is to indicate whether the UMTS bearer shall provide in-sequence SDU delivery or not. Maximum SDU size (octets) is the maximum SDU size for which the network shall satisfy the negotiated QoS. SDU format information (bits) is a list of possible exact sizes of SDUs. SDU error ratio is to indicate the fraction of SDUs lost or detected as erroneous. SDU error ratio is defined only for

Table I. UMTS classes.

	Traffic class	Features	Examples
Real-time	Conversational	Preserve time relation (variation) between information entities of the stream; Conversational pattern (stringent and low delay)	Voice
	Streaming	Preserve time relation (variation) between information entities of the stream	Video
Best effort	Interactive	Request response pattern; Preserve payload content	Web browsing
	Background	Destination is not expecting the data within a certain time; Preserve payload content	Background download of emails

Table II. UMTS QoS attributes defined for each class.

Traffic class	CS	SC	IC	BC
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
Maximum SDU size	X	X	X	X
SDU format information	X	X		
SDU error ratio	X	X	X	X
Residual bit error ratio	X	X	X	X
Delivery of erroneous SDUs	X	X	X	X
Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/retention priority	X	X	X	X
Source statistics descriptor	X	X		
Signalling indication			X	

CS, conversational class; SC, streaming class; IC, interactive class; BC, background class.

conforming traffic. Residual bit error ratio is to indicate the undetected bit error ratio in the delivered SDUs. If no error detection is requested, residual bit error ratio indicates the bit error ratio in the delivered SDUs. Delivery of erroneous SDUs (y/n/-) is to indicate whether SDUs detected as erroneous shall be delivered or discarded. Transfer delay (ms) is to indicate maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP. Traffic handling priority is to specify the relative importance for handling of all SDUs belonging to the UMTS bearer compared to the SDUs of other bearers. Allocation/retention priority is to specify the relative importance compared to other UMTS bearers for allocation and retention of the UMTS bearer. The allocation/retention priority attribute is a subscription attribute, which is not negotiated from the mobile terminal. Source statistics descriptor ('speech'/'unknown') is to specify characteristics of the source of submitted SDUs. Signaling indication (yes/no) is to indicate the signaling nature of the submitted SDUs. This attribute is additional to the other QoS attributes and does not over-ride them.

In Table II, the defined UMTS bearer attributes and their relevancy for each bearer traffic class are summarized.

4.3. Integrated 3G/WLAN Network with all-IP UMTS

The UMTS infrastructure includes the CN and the UMTS terrestrial radio access network (UTRAN), as shown in Figure 4(a) and (b) (left). Here, we adopt all-IP UMTS architectures introduced in Reference [29]: (a) *Option 1* for PS domain, and (b) *Option 2* for CS domain. The CN is responsible for switching/routing calls and data connections to the external networks, and the UTRAN handles all radio-related functionalities. The CN consists of two service domains, the PS service domain and the CS service domain. The PS domain provides the access to the IP-based networks, and the CS domain provides the access to the PSTN/ISDN. Figure 4(a) shows integration of WLAN and all-IP UMTS in PS domain, and Figure 4(b) shows integration of WLAN and all-IP UMTS in CS domain. The serving GPRS support node (SGSN) is equivalent to that of the MSC/VLR in the GSM network. The gateway GPRS support node (GGSN) is primarily a router with switching and routing functions. The HSS is the master database containing all 3G user-related subscription information such as IP multimedia user database, a subset of the HLR for the PS domain, and a subset of HLR for the CS domain. The IP multimedia subsystem is located behind the GGSN for functions such as call control of SIP, and voice over IP (VoIP) functions. The application and server network supports flexible services through a service platform. The UTRAN consists of node Bs (base stations) and the RNCs. MSs communicate with node Bs through the radio interface based on the WCDMA technology. Figure 4(b) (left) shows *Option 2* of all-IP UMTS in the CS domain, which is a superset of *Option 1* [29]. Note that details about *Option 1* are omitted in Figure 4(b). Two control elements, the MSC and the GMSC servers, are introduced. UTRAN accesses the core network via a CS-MGW (user plane) separated from the MSC server over Iu interface. There are one or more CS-MGWs in the figure for voice format conversion between PS and CS networks. Please refer to Reference [29] for more details about the all-IP UMTS architectures.

The MSs are dual-mode terminals supporting both UMTS and IEEE 802.11. Figure 4(a) and (b) (right) show an IEEE 802.11 Gateway connecting multiple access points (AP). The IEEE 802.11 Gateway domain is an extended service set (ESS). We assume that the IEEE 802.11e is implemented. The 802.11 gateway may connect either a RNC emulator or a SGSN emulator. In Figure 4, we only show the later case. If the 802.11 gateway connect a RNC emulator, the RNC emulator connects SGSN in the UMTS CN. If the 802.11 gateway connect a SGSN emulator, the SGSN emulator connects GGSN, HSS and application and service networks in the UMTS CN, as shown in Figure 4(a). Note that if the IEEE 802.11 gateway

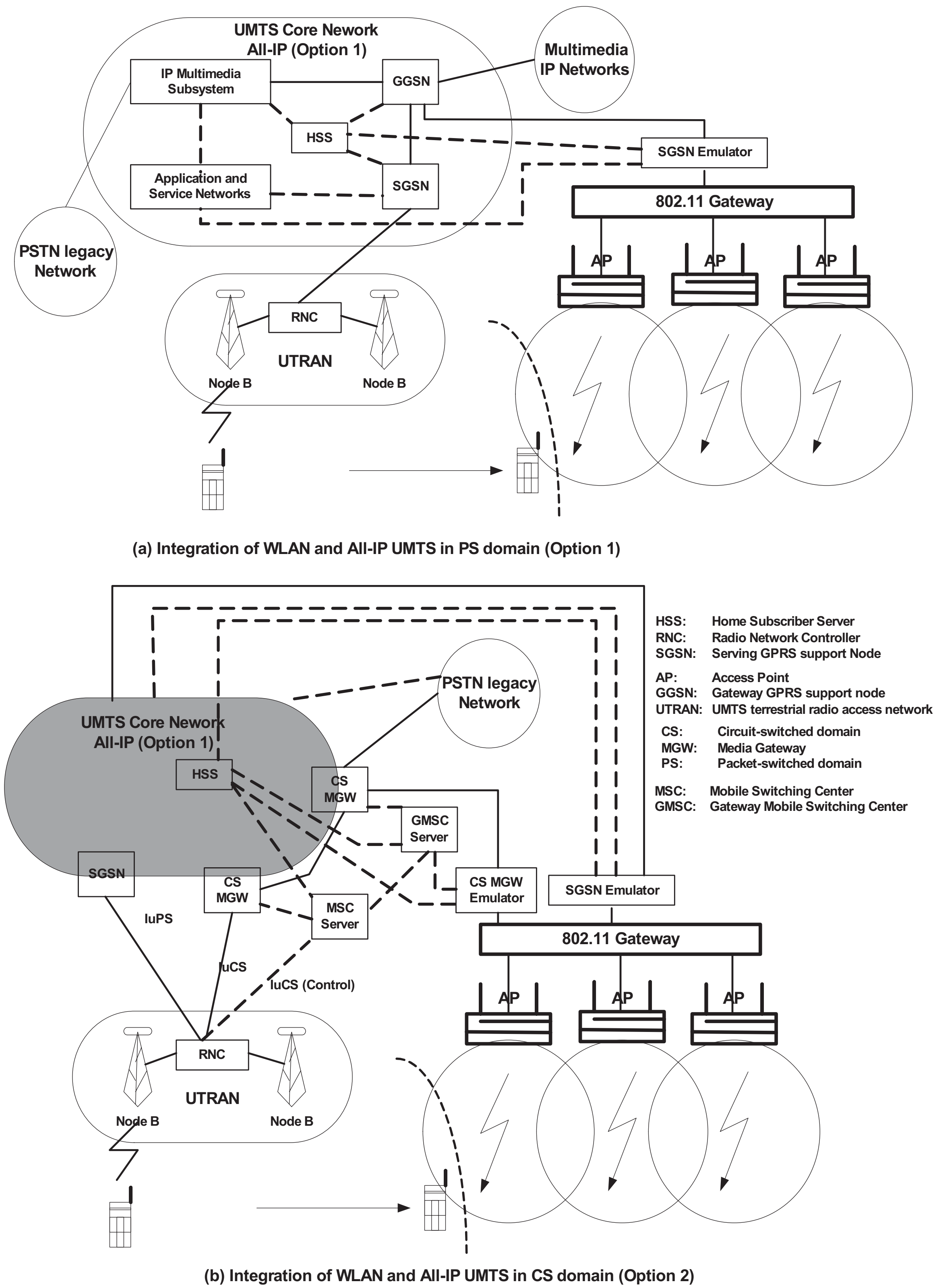

Fig. 4. Tightly coupled integration (simplified) of WLAN and all-IP UMTS: (a) *Option 1* in PS domain; (b) *Option 2* in CS domain.

connects a RNC emulator instead of a SGSN emulator, Figure 4 can be further simplified.

For the CS domain of UMTS, the IEEE 802.11 gateway connects a CS MGW emulator, which connects a CS MGW, and then connects the PSTN legacy network. With the architectures in Figure 4, a seamless voice/multimedia/data handoff is possible in the integrated UMTS/WLAN network. As illustrated in Figure 4, when an MS moves from a UMTS network to a BSS in a wireless LAN, a handoff procedure is needed. Next, we briefly introduce the mobility management and handoff.

4.4. Mobility Management and Handoff

In the UMTS PS domain, the cells are grouped into RAs. The RA of an MS is tracked by the SGSN. The cells in an RA are further grouped into UTRAN registration areas URAs. UMTS utilizes a three-level location management strategy, i.e., an MS is tracked at cell level during packet transmission session, at the URA level during the idle period of an ongoing session, and at the RA level when the MS is not in any communication session [31]. The IEEE 802.11 network can be treated as a special URA in a special RA, within which, the IEEE 802.11 mobility management is adopted. If the IEEE 802.11 gateway connects a SGSN emulator, as shown in Figure 4, the IEEE 802.11 gateway also needs to emulate a RNC function, and the ESS domain is both a URA and a RA. If the IEEE 802.11 Gateway connects a RNC emulator, the IEEE 802.11 network can be treated as a special 'RNC,' or a special URA in a special RA. When an MS moves in the UMTS network, mobility management of UMTS is used, while when an MS moves in the IEEE 802.11 network, the IEEE 802.11 mobility management is used. Mobility from the UMTS network to the IEEE 802.11 network or from the IEEE 802.11 network to the UMTS network causes an inter-RNC URA update and a RA update. Users of the IEEE 802.11 network and users of the UMTS network may also share the same poll of IP addresses assigned by the GGSN, and therefore, mobility across between the UMTS network and the IEEE 802.11 network does not cause a change of the IP address. When an MS moves within the IEEE 802.11 network, the associations between the MS and APs change [1].

If an MS moves from the UMTS network to the IEEE 802.11 network, the MS first performs association with the corresponding AP, and then an inter-RNC URA update and a RA update are performed. If an MS moves from the IEEE 802.11 network to the UMTS network, the MS first performs de-association with the corresponding AP, and then an inter-RNC URA update and a RA update are performed.

If the MS has an on-going call or multimedia/data connection when moving between UMTS and IEEE 802.11, a cell level update is performed (the IEEE 802.11 gateway emulates the effect), and a handoff occurs, which involves resource management as well as QoS mapping to be discussed in later sections.

Therefore, when an MS moves from the UMTS network to the IEEE 802.11 network, one of following cases may happen:

- If the MS has neither an on-going voice call (VoIP) nor a data connection, only mobility management is involved.
- If the MS has an on-going voice call but not a data connection, the voice call still stay in the UMTS network if possible, and only mobility management is involved. Or
- If the MS has an on-going voice call but not a data connection, a voice handoff is needed. Note this can happen since we adopted all-IP UMTS architectures. Otherwise, a voice call's handoff between WLAN and UMTS cannot be performed since UMTS uses circuit switched technology, but WLAN uses packet switched technology.
- If the MS has a data connection but not an on-going voice call, a data handoff is needed.
- If the MS has both an on-going voice call and a data connection, both a voice handoff and a data handoff are needed. It may happen that the IEEE 802.11 network accepts one kind of handoff, but rejects another if there is not enough resource. Or
- If the MS has both an on-going voice call and a data connection, another scenario is also possible: the MS's voice call still use the UMTS network if possible, but a data handoff is needed to enjoy higher bandwidth.

One reason of the necessities of voice handoff call to WLAN is that at hotspots where integration of 3G and WLAN happens, there are potential more cellular customers. Therefore, a voice handoff from UMTS to WLAN can help to reduce the load of UMTS. However, it could be optional.

4.5. Mapping UMTS QoS with WLAN QoS

Mapping UMTS QoS with WLAN QoS is a challenging task since UMTS and WLAN are totally different networks. Exactly mapping all UMTS QoS parameters is not possible. In this article, we will show how to map some major QoS parameters listed in Table II, such as maximum bitrate, delivery order, maximum SDU size, transfer delay, and guaranteed bit rate. Furthermore, we will show how these UMTS QoS parameters can be guaranteed in WLAN in the next section.

Maximum bitrate can be achieved in IEEE 802.11e via a leaky bucket algorithm and a token bucket algorithm [4]. Delivery order can be easily achieved in the HCCA of IEEE 802.11e. Maximum SDU size can also be achieved via the fragmentation threshold in WLAN.

QoS in the HCCA of the IEEE 802.11e can be defined by a set of parameters such as mean data rate, nominal frame size, and maximum service interval or delay bound. Maximum service interval is the maximum interval between the start of two successive QoS CF-Polls, and is a very close to delay bound since the HCCA is an immediately acknowledged system. Mean data rate is equivalent to guaranteed bit rate in UMTS QoS, whereas delay bound in WLAN and transfer delay in UMTS QoS are somewhat different since transfer delay is an end-to-end measurement, whereas delay bound is one-hop delay within a BSS. However, if we can know the delay (referred to as external delay) beyond the WLAN within 3G domain, we can have the following relationship.

$$\text{Delay bound} = \text{Transfer delay} - \text{External delay} \tag{1}$$

To obtain external delay, we have the following example. For a voice call in all-IP UMTS, transfer delay is the delay sum of from the MS to Node B, from Node B to RNC, from RNC to CS MGW, from CS MGW to zero or more CS MGWs, and from CS MGW to PSTN legacy network, shown in Figure 4(b). After the voice call handoffs to WLAN, transfer delay is the delay sum of delay bound, from the AP to CS MGW emulator, from CS MGW emulator to zero or more CS MGWs, and from CS MGW to PSTN legacy network, shown in Figure 4(b). In other words, external delay can be obtained.

Guaranteeing transfer delay and guaranteed bit rate can be a little challenging, and will be discussed in the next section.

4.6. Resource Management

If an MS, originally from the UMTS network, with an on-going call or a multimedia/data connection moves back to the UMTS network from the IEEE 802.11 network, the resource management follows the UMTS resource management for an inter-RNC handoff. The challenging issue is how to perform resource management when an MS with an on-going call or a multimedia/data connection moves from the UMTS network to the IEEE 802.11 network. There are two difficulties. First, the UMTS network is a connection-oriented network in the sense that both a voice call and a multimedia/data session have connections, whereas the IEEE network is connectionless-featured network so that there is no connection concept there. Furthermore, how to guarantee QoS in the IEEE 802.11 MAC layer is still an open issue. Note that without the MAC layer support, QoS guarantee at higher layers is not possible. In the section, with the IEEE 802.11e, we show that QoS guarantee can be achieved. To this end, we need to design an admission control and scheduling algorithm for the HCCA in the IEEE 802.11 network. The HCCA is adopted in the integrated 3G/WLAN architecture since it is relatively deterministic for QoS issues compared to the contention-based EDCA.

In this Subsection, we only consider resource management for those MSs who moves from the UMTS network to the IEEE 802.11 network and moves back from the IEEE 802.11 network to the UMTS network. In other words, MSs originally residing in the UMTS network are considered. In this section, we assume that all the resource under the HCCA is only for MSs from the UMTS networks, whereas the local traffic of the IEEE 802.11 networks will use the contention-based EDCA. The scheme can be easily adapted to the case when the above assumption is removed. For the rest of the section, 'MS' stands for an MS who originally moved or is moving from the UMTS network to the IEEE 802.11 network, and it has both UMTS and 802.11/802.11e enabled.

For the HCCA in IEEE 802.11e, when the HC provides controlled channel access to MSs, it is responsible to grant or deny polling service based on the admitted voice calls and admitted multimedia/data connections. In general, we use 'a request' to refer to as any kind of arrival request, either voice or multimedia/data.

All of these criteria affect the admissibility of a given request. If both maximum service interval and delay bound are specified, uses the maximum service interval for the calculation of the schedule. The schedule for an admitted request is calculated in two steps: (a) Calculation of the scheduled service interval (SI); (b) TXOP duration for a given SI is calculated for the request. First calculates the minimum of all maximum service intervals for all admitted requests. Let this minimum be '*m*.' Second, the scheduler chooses a number lower than '*m*' that it is a submultiple of the beacon interval. This value is the scheduled service interval for all MSs with admitted requests.

When a new request requests for admission, the admission control process is done in three steps. First, calculates the number of frames that arrive at the mean data rate during the scheduled service interval. Second, calculates the TXOP duration that needs to be allocated for the request. Finally, the admission control unit (ACU) determines that the request can be admitted when the following inequality is satisfied: where k is the number of existing connections/flows and $k+1$ is used as index for the newly arriving request. T indicates the beacon interval and T_{CP} is the time used for EDCA traffic.

$$\frac{\mathrm{TXOP}_{k+1}}{\mathrm{SI}} + \sum_{j=1}^{k} \frac{\mathrm{TXOP}_j}{\mathrm{SI}} \leq \frac{T - T_{\mathrm{CP}}}{T} \tag{2}$$

All admitted requests have guaranteed access to the channel. For the calculation of the TXOP duration for an admitted request, the Simple Scheduler uses the following parameters: mean data rate (r) and nominal MSDU size (L) from the negotiated traffic request, the scheduled service interval (SI) calculated above, physical transmission rate (R), size of maximum frame, i.e., 2304 bytes (M) and overhead in time units (O). The physical transmission rate is the minimum PHY rate negotiated in the traffic request. The overhead in time includes interframe spaces, ACKs and CF-Polls. For the calculation of the TXOP duration for an admitted request, the simple scheduler uses the following parameters. First the scheduler calculates the number of frames that arrived at the mean data rate during the SI: N_i; then the scheduler calculates the TXOP duration as the maximum of (1) time to transmit N_i frames at R_i and (2) time to transmit one maximum size MSDU at R_i (plus overheads):

$$N_i = \left\lceil \frac{\mathrm{SI} * \rho_i}{L_i} \right\rceil \tag{3}$$

$$\mathrm{TXOP}_i = \max\left(\frac{N_i * L_i}{R_i} + O, \frac{M}{R_i} + O\right) \tag{4}$$

An example of the scheduling is shown in Figure 5. Request from MS '*i*' is admitted in Figure 5(a). The beacon interval is 100 ms and the maximum service interval for the request is 60 ms. The scheduler calculates a scheduled service interval (SI) equal to 50 ms using the steps explained above. The same process is repeated continuously while the maximum service interval for the admitted request is smaller than current SI. If a new request is admitted with a maximum service interval smaller than the current SI, the scheduler needs to change the current SI to a smaller

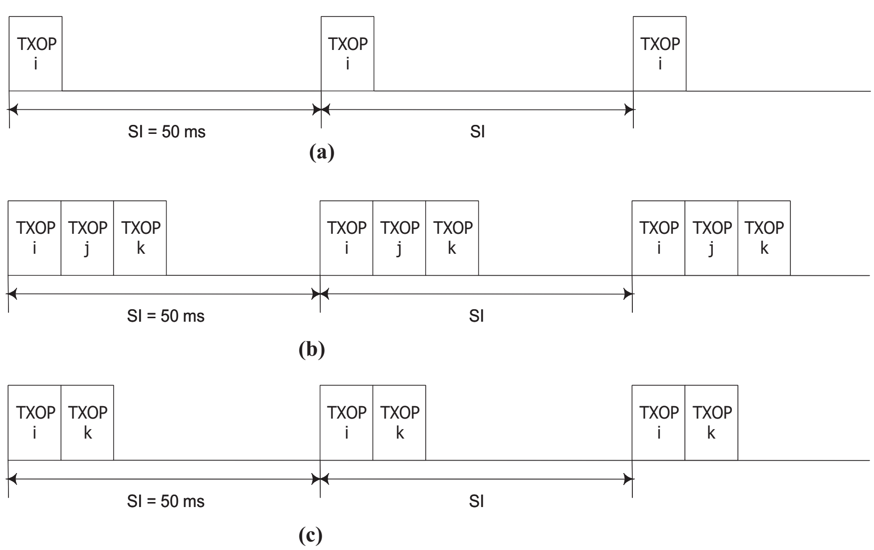

Fig. 5. An example of the scheduler: (a) Schedule for request from MS '*i*'; (b) Schedule for requests from MS '*i*' to '*k*'; (c) Reallocations of TXOPs when a session has finished.

number than the maximum service interval of the newly admitted request. Therefore the TXOP duration for the current admitted requests needs also to be recalculated with the new SI. If the corresponding MS leaves or finishes the voice/multimedia/data session, shown in Figure 5(c), the scheduler has additional available resource.

5. Conclusions and Future Research Directions

Interoperation of 3G and WLAN can support service diversity and optimal connectivity, by improving both mobility and QoS requirements.

In this paper, we have provided a comprehensive survey on integration of 3G and WLAN. We have discussed several issues such as underline network architectures, integrated architectures, mobility management, and QoS. We particular have discussed an integrated 3G/WLAN networks with all-IP UMTS, and a QoS mapping and guarantee mechanism for resource management for seamless voice/multimedia/data handoffs between 3G and all-IP UMTS with the emerging IEEE 802.11e standard.

Further research directions are mostly related to heterogeneous WLAN and 3G networks including:

- Studying different efficient handoff procedures between 3G and WLAN so that the delay of handoff procedure is minimal.
- Studying efficient resource allocation schedules between 3G and WLAN so that QoS guaranteed for real-time traffic is achieved while utilization is maximized.
- Studying fast authentication schemes between 3G and WLAN so that both strong security and fast handoff are achieved.
- Studying practical billing schedules for integrated 3G/WLAN network.

Integration of WLAN and 3G has a promising future since it provides benefits to both the end users and service providers with advantages of both technologies. However, many challenge issues still exist for realization of the technologies. For example, QoS guarantee in integrated 3G/WLAN networks deserves further investigations, and is an extremely challenging issue due to many reasons such as different network architectures, different radio technologies, different upper layer protocols, and different network capacities.

References

1. IEEE 802.11 WG, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification, standard. *IEEE*, August 1999.
2. IEEE 802.11b WG, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification: high-speed physical layer extension in the 2.4 GHz band. *IEEE*, September 1999.
3. IEEE 802.11a WG, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification: high-speed physical layer in the 5 GHz band. *IEEE*, September 1999.
4. IEEE 802.11e WG, Draft Supplement to Part 11: Wireless medium access control (MAC) and physical layer (PHY) specifications: medium access control (MAC) enhancements for quality of service (QoS), Std 802.11e/D3.3.2. *IEEE*, November 2002.
5. 3GPP. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; RRC Protocol Specification for Release 1999. Technical Specification 3G TS 25.331 version 3.5.0 (2000–12), 2000.
6. 3GPP2. TIA/EIA/IS-835B cdma2000 Wireless, IP Network Standard, 2000.
7. Perkins C. IP Mobility Support for IPv4, IETF RFC 3344, August 2002.
8. Salkintzis AK, Fors C, Pazhyannur C. WLAN-GPRS integration for next-generation mobile data networks. *IEEE Wireless Communications* 2002; **9**(5): 112–124.
9. Buddhikot M, Chandranmenon G, Han S, Lee YW, Miller S, Salgarelli L. Integration of 802.11 and third-generation wireless data networks. In *Proceedings of IEEE INFOCOM*, 2003.
10. Varma VK, Ramesh S, Wong KD, Friedhoffer JA. Mobility management in integrated UMTS/WLAN networks. *Proceedings of IEEE ICC*, 2003.
11. Buddhikot MM, Chandranmenon G, Han S, Lee YW, Miller S, Salgarelli L. Design and implementation of a WLAN/CDMA2000 interworking architecture. *IEEE Communications Magazine* 2003; **41**(11): 90–100.
12. Borisov N, Goldberg I, Wagner D. Intercepting mobile communications: The insecurity of 802.11. *Proceedings of MobiCom*, 2001.
13. IEEE Std. 802.11i/D4.0. Draft amendment to standard for telecommunications and information exchange between systems—LAN/MAN specific requirements—Part 11: wireless medium access control (MAC) and physical layer (PHY) specifications: medium access control (MAC) security enhancements, *IEEE*, May 2003.
14. Ahmavaara K, Haverinen H, Pichna R. Interworking architecture between 3GPP and WLAN systems. *IEEE Communications Magazine* 2003; **41**(11): 74–81.
15. 3GPP. Group services and system aspects; 3GPP systems to wireless local area network (WLAN) interworking; system description (release 6), TS 23.234. v. 1.10.0, May 2003.
16. Haverinen H, Salowey J. EAP SIM authentication. IETF draft-haverinen-pppext-eap-sim-10.txt, February 2003.
17. Arkko J, Haverinen H. EAP AKA authentication. IETF draft-arkko-pppext-eap-aka-09.txt, February 2003.
18. Køien GM, Haslestad T. Security aspects of 3G-WLAN interworking. *IEEE Communications Magazine* 2003; **41**(11): 82–88.
19. Zhuang W, Gan YS, Loh KJ, Chua KC. Policy-based QoS management architecture in an integrated UMTS and WLAN environment. *IEEE Communications Magazine* 2003; **41**(11): 118–125.
20. Zhang Q, Guo C, Guo Z, Zhu W. Efficient mobility management for vertical handoff between WWAN and WLAN. *IEEE Communications Magazine* 2003; **41**(11): 102–108.

21. Varma VK, Wong KD, Chua KC, Paint F. Guest Editorial—Integration of 3G wireless and wireless LANs. *IEEE Communications Magazine* 2003; **41**(11): 72–74.
22. Mccann S, Flygare F. Hiperlan/2 public access interworking with 3G cellular systems. *Wireless Networks* 2004; **10**: 43–51.
23. Luo H, Jiang Z, Kim BJ, Shankaranarayanan NK, Henry P. Integrating wireless LAN and cellular data for enterprise. *IEEE Internet Computing* 2003; **March–April**: 25–33.
24. Floris A, Tosetti A, Veltri L. Solutions for mobility support in DHCP-based environments. In *Proceedings of IEEE ICC'03*.
25. Varma VK, Ramesh S, Wong KD, Friedhoffer JA. Mobility management in integrated UMTS/WLAN networks. In *Proceedings of IEEE ICC*, 2003.
26. Jaseemuddin M. An architecture for integrating UMTS and 802.11 WLAN networks. In *Proceedings of IEEE ISCC*, 2003; pp. 716–723.
27. Song JY, Lee SW, Cho DH. Hybrid coupling scheme for UMTS and wireless LAN interworkings. In *Proceedings of IEEE VTC* 2003.
28. Haverinen H, Mikkonen J, Takamäki T. Cellular access control and charging for mobile operator wireless local area networks. *IEEE Wireless Communications* 2002; **9**(6): 52–60.
29. Lin YB, Pang AC, Huang YR, Chlamtac I. An all-IP approach for UMTS third-generation mobile networks. *IEEE Network* 2002; 8–19.
30. 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture, TS 23.107 Version 5.9.0, (2003–6), 2003.
31. Yang SR, Lin YB. Performance evaluation of location management in UMTS. *IEEE Transactions on Vehicle Technology* 2003; **52**(6): 1603–1615.
32. Xiao Y, Rosdahl J. Performance analysis and enhancement for the current and future IEEE 802.11 MAC protocols. *ACM SIGMOBILE Mobile Computing and Communications Review* 2003; **7**(2): 6–19.
33. Xiao Y. Packing Mechanisms for the IEEE 802.11n Wireless LANs. In *Proceedings of IEEE GLOBECOM*, 2004.
34. Xiao Y. Efficient MAC strategies for the IEEE 802.11n wireless LANs. *Wireless Communications and Mobile Computing* (in press).
35. Perkins CE, Calhoun PR. AAA registration keys for mobile IP. Internet Draft, 22 June 2003.
36. Rigney C, Rubens A, Simpson W, Willens S. Remote authentication dial in user service (RADIUS). RFC 2865, Internet Engineering Task Force, June 2000.
37. Calhoun PR, Loughney J, Guttman E, Zorn G, Arkko J. DIAMETER base protocol (work in progress). Internet Draft, internet engineering task force. draft-ietf-aaa-diameter-15.txt, October 2002.

Authors' Biographies

Yang Xiao is assistant professor of Department of Computer Science, The University of Memphis. Dr Yang Xiao is an IEEE senior member. He was a voting member of IEEE 802.11 Working Group from 2001 to 2004. He currently serves as an associate editor or on editorial boards for six refereed journals: (Wiley) *International Journal of Communication Systems*, (Wiley) *Wireless Communications and Mobile Computing* (WCMC), *EURASIP Journal on Wireless Communications and Networking*, *International Journal of Wireless and Mobile Computing*, *International Journal of Signal Processing*, and *International Journal of Information Technology*. He serves a (lead) guest editor for EURASIP Journal on Wireless Communications and Networking, special issue on 'Wireless Network Security' in 2005, a (sole) guest editor for (Elsevier) Computer Communications journal, special Issue on 'Energy-Efficient Scheduling and MAC for Sensor Networks, WPANs, WLANs, and WMANs' in 2005, a (lead) guest editor for (Wiley) *Journal of Wireless Communications and Mobile Computing*, special Issue on 'Mobility, Paging and Quality of Service Management for Future Wireless Networks' in 2004–2005, a (lead) guest editor for *International Journal of Wireless and Mobile Computing*, special issue on 'Medium Access Control for WLANs, WPANs, Ad Hoc Networks, and Sensor Networks' in 2004–2005, and an associate guest editor for *International Journal of High Performance Computing and Networking*, special issue on 'Parallel and Distributed Computing, Applications and Technologies' in 2003. He serves as co-editor for four edited books: Wireless LANs and Bluetooth, Security and Routing in Wireless Networks, Ad Hoc and Sensor Networks, and Design and Analysis of Wireless Networks. He serves as a technical program vice chair for The 2005 International Conference on Wireless Networks (ICWN 2005). He serves as a technical program vice chair on Wireless and Mobile Computing for the 2005 International Conference on High Performance Computing and Communications (HPCC-05). He serves as a symposium co-chair for International Symposium on Wireless Local and Personal Area Networks in WirelessCom 2005. He serves as a symposium co-chair for Symposium on Data Base Management in Wireless Network Environments in IEEE VTC 2003. He serves as a TPC member for many conferences such as IEEE ICDCS, IEEE ICC, IEEE GLOBECOM, IEEE WCNC, IEEE ICCCN, IEEE PIMRC, ACM WMASH, etc. He serves as a referee/reviewer for many journals, conferences, and funding agencies such as Research Grants Council (Hong Kong), Canada Foundation for Innovation, and Louisiana Board of Regents. He has served as a panelist for NSF in 2005. Dr. Xiao's research areas include Wireless LANs, Wireless PANs, Wireless MANs, Wireless WANs (cellular networks), and Ad hoc and Sensor networks. He has published many papers in major journals and refereed conference proceedings related to these research areas, such as *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Vehicular Technology*, *IEEE Communications Letters*, *IEEE Communications Magazine*, *IEEE Wireless Communications*, *ACM/Kluwer Mobile Networks and Applications (MONET)*, etc. His research interests are Security/Reliable Communications, Medium Access Control, Mobility/Location/Paging Managements, Cache Access and Replacement Policies, Quality of Service, Energy Efficiency, and Routing in wireless networks and mobile computing.

Kin K. Leung is received his B.S. degree from the Chinese University of Hong Kong in 1980, and his M.S. degree and his Ph.D. in Computer Science from University of California, Los Angeles, in 1982 and 1985, respectively. He started his career at AT&T Bell Labs in 1986. Following Lucent Technologies spun off from AT&T in 1996, he was with AT&T Labs from 1996 to 2002. In 2002, he re-joined Bell Labs of Lucent Technologies. Since 2004, he has been the Tanaka Chair Professor in Internet Technology at Imperial College. His research interests include radio resource allocation, MAC protocol, TCP/IP protocol, mobility management, network architecture, real-time applications and teletraffic issues for broadband wireless networks. He is also interested in a wide variety of wireless technologies, including 802.11, 802.16, and 3G and future generation wireless networks. e received the Distinguished Member of Technical Staff Award from AT&T Bell Labs in 1994, and was a co-recipient of the 1997 Lanchester Prize Honorable Mention Award. He is an IEEE fellow. He has published widely and acquired patents in many areas of communication networks. He has actively served on conference committees, including as the committee co-chair for the Multiaccess, Mobility and Teletraffic for Wireless Communications (MMT'98) and the committee vice-chair for the IEEE ICC 2002. He was a guest editor for the *IEEE Journal on Selected Areas in Communications (JSAC)* and the *MONET Journal*, and as an editor for the *JSAC: Wireless Series*. Currently, he is an editor for the *IEEE Transactions on Communications and the Transactions on Wireless Communications*.

Yi Pan is the chair and a full professor in the Department of Computer Science at Georgia State University. Dr Pan received his B.Eng. and M.Eng. degrees in Computer Engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. in Computer Science from the University of Pittsburgh, U.S.A., in 1991. Dr Pan's research interests include parallel and distributed computing, optical networks, wireless networks, and bioinformatics. Dr Pan has published more than 80 journal papers with 29 papers published in various IEEE journals. In addition, he has published over 100 papers in refereed conferences (including IPDPS, ICPP, ICDCS, INFOCOM, and GLOBECOM). He has also co-edited over 20 books (including proceedings) and contributed several book chapters. His pioneer work on computing using reconfigurable optical buses has inspired extensive subsequent work by many researchers, and his research results have been cited by more than 100 researchers worldwide in books, theses, journal, and conference papers. He is a co-inventor of three U.S. patents (pending) and five provisional patents, and has received many awards from agencies such as NSF, AFOSR, JSPS, IISF, and Mellon Foundation. His recent research has been supported by NSF, NIH, NSFC, AFOSR, AFRL, JSPS, IISF, and the states of Georgia and Ohio. He has served as a reviewer/panelist for many research foundations/agencies such as the U.S. National Science Foundation, the Natural Sciences and Engineering Research Council of Canada, the Australian Research Council, and the Hong Kong Research Grants Council. Dr. Pan has served as an editor-in-chief or editorial board member for eight journals including three *IEEE Transactions* and a guest editor for seven special issues. He has organized several international conferences and workshops and has also served as a program committee member for several major international conferences such as INFOCOM, GLOBECOM, ICC, IPDPS, and ICPP. Dr Pan has delivered over 50 invited talks, including keynote speeches and colloquium talks, at conferences and universities worldwide. Dr Pan is an IEEE Distinguished Speaker (2000–2002), a Yamacraw Distinguished Speaker (2002), a Shell Oil Colloquium Speaker (2002), and a senior member of IEEE. He is listed in Men of Achievement, Who's Who in Midwest, Who's Who in America, Who's Who in American Education, Who's Who in Computational Science and Engineering, and Who's Who of Asian Americans.

Xiaojiang Du is an assistant professor in Department of Computer Science, North Dakota State University. Dr Du received his B.E. degree in EE from Tsinghua University, Beijing, China in 1996, and his M.S. degree and Ph.D. in EE from University of Maryland, College Park in 2002 and 2003, respectively. Dr Du's research interests are wireless sensor networks, mobile ad hoc networks, wireless networks, computer networks, network security, and network management. He is a technical program committee member for several international conferences (including IEEE ICC 2006, Globecom 2005, BroadNets 2005, WirelessCom 2005, IPCCC 2005, and BroadWise 2004). He is a member of IEEE.