Int. J. Wireless and Mobile Computing, Vol. 2, No. 4, 2007

Interworking of 3G cellular networks and wireless LANs

Wei Song and Weihua Zhuang*

Centre for Wireless Communications,
Department of Electrical and Computer Engineering,
University of Waterloo,
Waterloo, ON, Canada N2L 3G1
E-mail: wsong@bbcr.uwaterloo.ca
E-mail: wzhuang@bbcr.uwaterloo.ca
*Corresponding author

Aladdin Saleh

Wireless Technology Department,
Bell Canada,
5099 Creekbank Road,
Mississauga, ON,
Canada L4W 5N2
E-mail: aladdin.saleh@bell.ca

Abstract: The Third Generation (3G) cellular networks provide ubiquitous connectivity but low data rates, whereas Wireless Local Area Networks (WLANs) can offer much higher data rates but only cover smaller geographic areas. Their complementary characteristics make the integration of the two networks a promising trend for next-generation wireless networks. With combined strengths, the integrated networks will provide both wide-area coverage and high-rate data services in hotspots. There are many aspects involved in their interworking, such as mobility, security and Quality of Service (QoS) provisioning. In this paper, we present a survey of most recent interworking mechanisms proposed in the literature, and outline some important open issues to achieve seamless integration.

Keywords: 3G/WLAN interworking; tight/loose coupling; mobility management; quality of services; QoS provisioning.

Reference to this paper should be made as follows: Song, W., Zhuang, W. and Saleh, A. (2007) 'Interworking of 3G cellular networks and wireless LANs', *Int. J. Wireless and Mobile Computing*, Vol. 2, No. 4, pp.237–247.

Biographical notes: Wei Song received a BSc in Electrical Engineering from Hebei University, China, in 1998, an MSc in Computer Science from Beijing University of Posts and Telecommunications, China, in 2001 and a PhD in Electrical Engineering from the University of Waterloo, Canada, in 2007. Currently, she is a PostDoctoral Fellow in the Department of Electrical and Computer Engineering, at the University of Waterloo, Canada. Her current research interests include resource allocation and Quality-of-Service (QoS) provisioning for the integrated cellular networks and Wireless Local Area Networks (WLANs). She received the Best Student Paper Award from IEEE WCNC'07.

Weihua Zhuang received a PhD in Electrical Engineering from the University of New Brunswick, Canada. Since October 1993, she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is co-author of the textbook Wireless Communications and Networking (Prentice Hall, 2003). Her current research interests include multimedia wireless communications, wireless networks and radio positioning. She received the Premier's Research Excellence Award in 2001 from the Ontario Government for demonstrated excellence of scientific and academic contributions, the Outstanding Performance Award in 2005 and 2006 from the University of Waterloo, and the Best Paper Awards from IEEE WCNC'07, IEEE ICC'07 and Qshine'07. She is the Editor-in-Chief of *IEEE Transactions on Vehicular Technology* and an Editor of *IEEE Transactions on Wireless Communications*, EURASIP *Journal on Wireless Communications and Networking* and *International Journal of Sensor Networks*.

Aladdin Saleh received a PhD in Electrical Engineering from the University of London, England, in 1984. Since then, he has held several positions both in the academia and the industry. In 1998,

he joined the wireless technology team of Bell Canada, where he designed and implemented several wireless Internet application projects. He has also developed Bell Canada technology strategy for the IEEE 802.11/WiFi, IEEE 802.16/WiMAX, and the integration of 3G cellular networks and WLAN. He is also an Adjunct Professor of Computer and Electrical Engineering at the University of Waterloo, Canada. He is a Fellow of IEE and a Senior Member of IEEE. His current research interests are in the areas of next-generation wireless networks and wireless Internet applications.

1 Introduction

In the last decade, there has been successful deployment and fast evolution of various wireless networks. Different technologies are adopted for different application environments. It is well-recognised that the next-generation wireless networks will integrate heterogeneous technologies to achieve enhanced performance. The attractive and complementary characteristics presented by cellular networks and Wireless Local Area Networks (WLANs) make them promising candidates.

Originally aiming at providing high-quality circuit-switched voice service to mobile users within wide areas, cellular networks have been well deployed around the world and have evolved to the Third Generation (3G). Two major standards for 3G mobile/wireless cellular networks are the Universal Mobile Telecommunication System (UMTS) and cdma2000, which are specified by the 3G partnership projects, that is, 3GPP and 3GPP2, respectively. Both systems are based on Code-Division Multiple Access (CDMA) and augmented with packet-switched data services, such as Multimedia Message Service (MMS) and Wireless Application Protocol (WAP) service. In cdma2000, for example, the nominal 1.25 MHz bandwidth can achieve a data rate up to around 2 Mbps for indoor office environments. However, this is still not enough to satisfy the ever-increasing demands for bandwidth-intensive data applications.

On the other hand, usually operating at unlicensed frequency bands, WLANs provide data services with lower cost. Moreover, the large bandwidth available for WLANs makes it possible to achieve higher data rates. For example, in IEEE 802.11b (one of the most popular WLAN standards), a WLAN can have a bandwidth more than 20 MHz. IEEE 802.11b operates at the licence-exempt Industrial, Scientific and Medical (ISM) frequency band from 2.4 to 2.483 GHz. It extends the physical layer based on Direct Sequence Spread Spectrum (DSSS) specified in the original 802.11 standard and supports a higher data rate up to 11 Mbps. The subsequent revisions such as 802.11a and 802.11g adopt Orthogonal Frequency-Division Multiplexing (OFDM) and offer a maximum rate of 54 Mbps at the unlicensed 5 and 2.4 GHz bands, respectively. However, designed as a wireless extension to the wired ethernet, a WLAN can only cover a small geographic area. For instance, an 802.11b Access Point (AP) can communicate with a Mobile Station (MS) within up to 60 m at 11 Mbps and up to 100 m at 2 Mbps with omnidirectional antennas. Consequently, with lower cost and much higher data rates, WLANs can effectively supplement the 3G networks in hotspot areas, where bandwidth-demanding applications are concentrated. As a result, by effectively combining 3G cellular networks and WLANs into an integrated wireless data access environment, mobile users can be provided with both ubiquitous connectivity and high-rate data services in hotspots.

In the following sections, we first discuss the important challenging issues involved in this integration problem. Then we briefly review some typical interworking solutions proposed in the literature. In the concluding remarks, some open research issues are outlined.

2 Challenges for 3G/WLAN interworking

The heterogeneous technologies employed in 3G cellular networks and WLANs bring many challenges to the interworking. Based on different radio access techniques, cellular networks and WLANs present distinct characteristics in terms of mobility management, security support and Quality of Service (QoS) provisioning. In order to achieve seamless integration, these issues should be carefully addressed while developing the interworking schemes.

After three-generation evolution, relatively mature and complete technologies have been established in cellular networks to address issues such as mobility, security, QoS, etc. With widely deployed infrastructure from radio access networks to core networks, ubiquitous connectivity is provided to mobile uses over wide areas. Different mobility levels are supported from fast vehicles moving on highways to stationary users in an indoor environment.

In contrast, the WLAN specifications only focus on the physical layer and Medium Access Control (MAC) layer. As for the upper layers, it assumes to adopt the same protocols as those in wired networks, for example, the Internet Protocol (IP) suite, with some adaptation for wireless links to avoid performance degradation. A de facto WLAN system is given in Ahmavaara et al. (2003), in which the layer-2 distribution system connects multiple APs, while access routers in turn connect the layer-2 distribution system to an IP backbone network. In some proposed interworking schemes (to be discussed), the access routers offer rich functionalities more than the basic function of IP routing, for example, transferring authentication and charging information between the 3G networks and WLANs. Through border gateways in the IP backbone network, WLAN terminals are provided IP connectivity to external IP networks such as the public Internet or a corporation intranet. Instead of providing continuous coverage over wide areas, WLANs are usually disjointly deployed in public or private hotspots such as cafés, airports and offices. Users in these areas normally have a very low mobility level, as most of these areas are located in indoor environments. Also, cellular coverage is available in these areas. As a result, a non-uniform overlay topology structure has to be considered for 3G/WLAN integration.

2.1 Seamless roaming across 3G cellular networks and WLANs

Taking into account the distinct mobility management mechanisms employed in cellular networks and WLANs, it is a rather challenging task to support seamless roaming across the two networks. Either the cellular networks or WLANs should have inherent mechanisms for location and handoff management to support the layer-2 or link layer mobility. In 3G networks (e.g., UMTS or cdma2000), with the aid of core networks, tunnelling protocols are used to support roaming within the Public Land Mobile Network (PLMN) or across 3G PLMNs of different operators with roaming agreements. For example, in UMTS, the General Packet Radio Service (GPRS) Mobility Management (GMM) is used. However, link layer mobility is not enough to provide network-layer transparency to upper-layer applications. To avoid disruption of upper-layer sessions due to IP address changes when user mobility results in changes of network attachment points, the network-layer IP mobility is needed. In cdma2000, Mobile IP (MIP) (Perkins, 2002) provides IP mobility within the same Packet Data Serving Node (PDSN) and between different PDSNs. However, in current specifications of UMTS, IP mobility is only supported within a UMTS network with the same Gateway GPRS Support Node (GGSN). It is being considered to introduce IP for inter-UMTS or inter-technology IP mobility through a three-state evolution in 3GPP (2000). An overview of the mobility management in UMTS and cdma2000 can be found in Pang et al. (2004).

The mobility management in WLANs is much simpler since they are only oriented to local areas. In IEEE 802.11 WLANs, the distribution system (e.g., an 802.3-type Ethernet) connects multiple Basic Service Sets (BSSs) into an Extended Service Set (ESS). Each BSS is under the control of an AP in the infrastructure mode. In this case, mobility across the BSSs within an ESS is handled by the APs involved. The Inter-Access Point Protocol (IAPP) specified in 802.11 f further facilitates the user roaming between APs of different vendors. When IP connectivity is provided in the WLAN system, IP micromobility protocols can be introduced to further support IP layer mobility.

Keeping in mind the difference between cellular networks and WLANs in mobility management, to achieve seamless roaming across the two networks, either a unified mobility management mechanism is followed, or both networks adopt their individual mobility management mechanisms and maintain proper interoperation.

2.2 Enhanced security level

Network security covers diverse issues such as user authentication, data confidentiality and integrity and key management. The security mechanisms in 3G networks such as UMTS are built upon those used in the second-generation cellular networks, for example, the Global System for Mobile communications (GSM). In particular, user authentication in UMTS adopts the Authentication and Key Agreement (AKA) procedure, which relies on the Universal Subscriber Identity Module (USIM) running in a smart-card at the user terminal. In addition to authenticating the subscriber's identity, cipher and integrity session keys are also generated from the long-term preshared secret key stored in the USIM module and Home Location Register/Authentication Centre (HLR/AuC). A good introduction to the access security in 3G networks is given in Koien (2004); Rose and Koien (2004).

In the original 802.11 standard, rather weak security is provided due to lack of key management and flaws of the Wired Equivalent Privacy (WEP) protocol. As a consequence, it fails the claimed objectives with respect to user identity privacy and data confidentiality. The 802.11i standard aims at improving the security level of WLANs. Port-based access control standard IEEE 802.1X is introduced to enhance authentication and key management using the Extensible Authentication Protocol (EAP) (Aboba et al., 2004). Also, as an interim solution, Temporal Key Integrity Protocol (TKIP) is specified to fix the vulnerabilities of WEP, while the long-term solution is based on Advanced Encryption Standard (AES) in place of the stream cipher RC4 used in WEP. More details on 802.11 security are given in Edney and Arbaugh (2003).

As we know, a system is only as strong as its weakest link. For 3G/WLAN interworking, if comparable security cannot be provided by both networks, adversaries can break into the system through the weakest component in the security chain and in turn defeat all security goals of the entire system. In other words, the two networks must be integrated in such a manner to achieve an enhanced (instead of impaired) security level. Also, appropriate independence between them should be maintained to minimise the security exposure and Domino effect when one of them is broken.

2.3 Consistent end-to-end QoS guarantee

The shared nature of radio link necessitates proper MAC to coordinate multiple connections to access the shared wireless channel. Both UMTS and cdma2000 use CDMA for multiple access. Moreover, by introducing packet-switched mode a higher resource utilisation can be achieved for bursty data traffic by statistical multiplexing. For the multiple access uplink (from MS to base station), a two-phase request-grant access procedure is used in 3G networks. First MSs send transmission requests to the Base Station (BS) through a contention channel. The BS acknowledges those successful requests and reserves resources for data transmission to follow. Then the MSs are notified the resource assignments. This type of centralised control and reservation-based resource allocation, together with proper admission control to limit the traffic load, enables fine QoS provisioning in 3G cellular networks.

On the other hand, the original WLAN specifications only support best-effort data service with contention-based random access protocol, for example, the Distributed Coordination Function (DCF) of 802.11 WLAN based on Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA). In this mode, the AP competes for access with MSs instead of scheduling the resource assignments as a BS in 3G networks. The other centralised control mode, Point Coordination Function (PCF), is based on polling by the AP. It is rarely implemented in reality due to unresolved problems such as uncontrolled transmission time of polled MSs. The distributed and contention-based control leads to weak QoS support capability. To achieve better

QoS, 802.11e develops new approaches by means of MAC enhancements. The above two access modes are improved with service differentiation. Also, admission control and bandwidth reservation (scheduling) are considered to support multimedia applications with stringent QoS requirements. Various QoS mechanisms for 802.11 are explored in Zhu et al. (2004). Nevertheless, WLANs still cannot be expected to support the same level of QoS as 3G networks.

Considering the differences of 3G networks and WLANs in QoS provisioning and the aforementioned overlaying structure, different services can be admitted to either the cellular network or the WLAN according to their traffic characteristics and QoS requirements. For example, real-time services such as voice telephony can be well carried by the cellular network to satisfy the strict delay requirement, whereas the delay-tolerant data traffic can be admitted to the WLAN to enjoy the high throughput. It is also an important issue to maintain consistent or smoothly adapted QoS during vertical handoff (i.e., handoff between the cellular network and the WLAN).

2.4 *Interworking scenarios*

Currently, WLANs may be owned by a 3G operator, a commercial WLAN operator, an authority of a public hotspot (e.g., airports or property management corporations) or a business enterprise for internal use (Salkintzis, 2004). The interworking mechanisms are directly related to the ownership of WLANs. More exposure is possible to integrate the WLANs owned by the cellular operator itself. The objective penetration level between the 3G networks and WLANs and the provisioned services lead to different requirements for the interworking mechanisms. The services supported in future integrated 3G/WLAN networks are envisioned in Axiotis et al. (2004). However, the integration of the two technologies will be a gradual development process. From the perspective of a 3G operator, a step-wise approach is proposed in 3GPP (2003b), which defines six interworking scenarios with each scenario specifying an incremental set of service and operational features. The first scenario only requires common billing and customer care. In the second scenario, a 3G subscriber roaming to the WLAN is authenticated and charged by its 3G home network. Only IP access service via the WLAN is provided to the roaming user. In the third scenario, the 3G packet-switched services are also open to users attached to the WLAN, such as MMS, WAP service, IP multimedia and location-based services. The fourth and fifth scenarios improve the third scenario with higher requirement for service continuity. The sixth scenario allows the access to the 3G circuit-switched services (such as conventional voice calls) via WLANs, and supports seamless mobility across 3G and WLANs. So far, a lot of research in the literature focuses on the four scenarios in the middle.

3 3G/WLAN interworking architectures and mechanisms

The standardisation for 3G/WLAN interworking is well in progress by 3GPP and 3GPP2. The high-level interworking requirements, architecture and procedures (e.g., network selection, authentication, charging, etc.) have been specified in 3GPP (2004a). In ETSI (2001) the integration architecture is classified into two categories according to the interdependence between the two access networks, that is, tight coupling and loose coupling. Different interworking mechanisms are developed to support mobility, security, QoS, charging and billing in the tightly coupled and loosely coupled heterogeneous networks.

3.1 *Tightly coupled and loosely coupled integration*

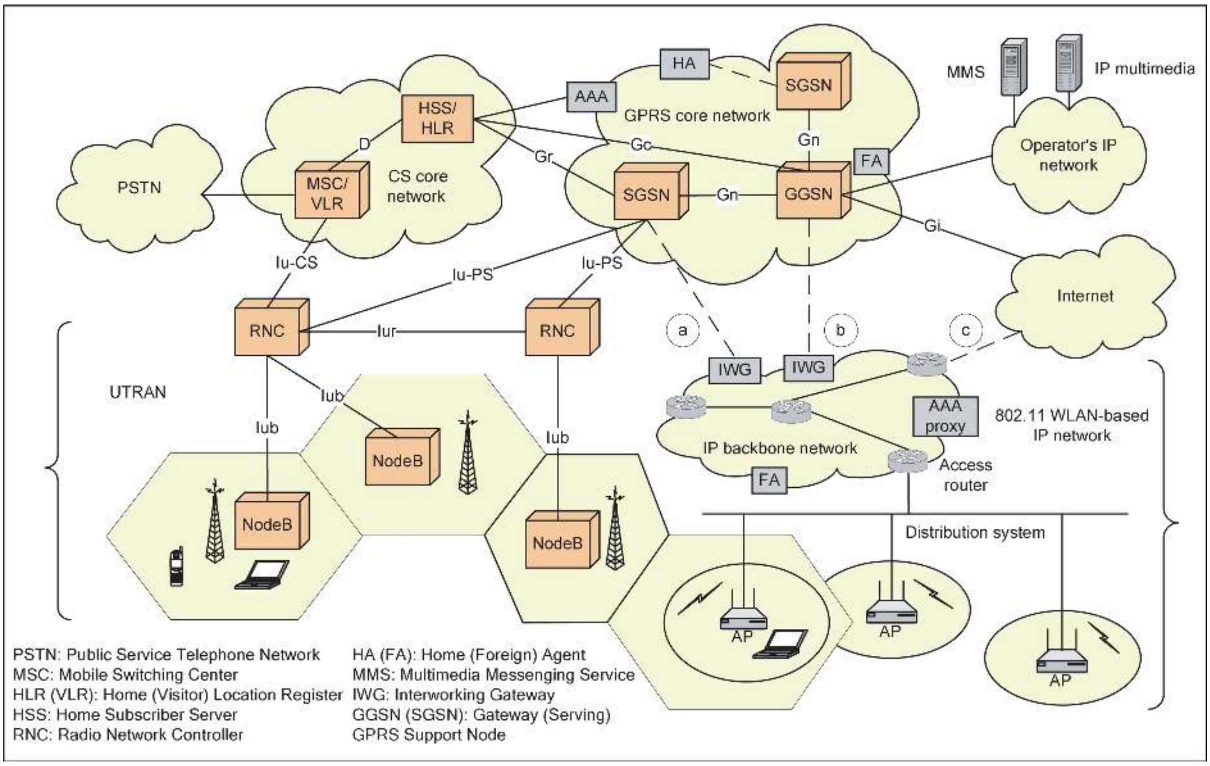
In the tight-coupling architecture, the WLAN is connected to the 3G core network as one 3G radio access network. Figure 1 illustrates a simplified architecture for interworking UMTS/GPRS networks with 802.11 WLANs. The lines with tags '*a*' and '*b*' are examples for tight coupling with different integration points. Typical tight-coupling architectures include the first architecture proposed in Salkintzis et al. (2004) and Buddhikot et al. (2003b).

We can see that, in this type of integration architecture, the cellular radio is simply replaced with WLAN radio providing equivalent functions. As a consequence, the 3G protocols and existing network infrastructures can be reused. For example, the user roaming across the two domains is based on the mobility management protocols of 3G networks, thus enhancing the interdomain mobility management capability. However, the high-layer cellular protocols tailored for highly mobile users in hostile outdoor environments may not operate properly for WLANs (Pichna et al., 2000). The main disadvantages of the tight coupling approach include:

1. an interface in 3G core networks exposed to WLAN is required, which is a challenge as the two domains are likely developed and deployed independently by different operators;
2. a large volume of WLAN traffic will go through the 3G core network, possibly making the latter a network bottleneck;
3. and the WLAN needs to have a protocol stack compatible with that of the 3G networks.

On the contrary, for the loose-coupling approach (shown with the line '*c*' in Figure 1), the WLAN is connected to the cellular core network indirectly through an external IP network such as the Internet. The second architecture proposed in Salkintzis et al. (2004) and Buddhikot et al. (2003b), respectively, and the operator WLAN system in Ala-Laurila et al. (2001) belong to this category. This type of architecture imposes minimal requirements to modify current WLAN standards, and allows for the flexibility and independence of implementing individually different mechanisms within each network. However, the 3G networks may need to be augmented with extra functionalities such as MIP for mobility management and AAA support. Moreover, as the two domains are separated, the mobility signalling may traverse a relatively long path, thus inducing relatively high handoff latency. Nonetheless, there are various enhancement mechanisms for MIP to reduce the handoff latency, such as regional registration and dynamic Home

Figure 1 Interworking architecture for 3GPP UMTS and IEEE 802.11 WLANs: (a) WLAN integrated at SGSN; (b) WLAN integrated at GGSN and (c) WLAN integrated to external IP networks



Agent (HA) assignment. Overall, the loose coupling is the preferred solution for both the 3G and WLAN communities as it allows the gradual deployment of hotspots with no or little modification on the 3G networks (Buddhikot et al., 2003a,b; Koien and Haslestad, 2003).

It is well recognised that the future networks will be *all-IP* networks. Therefore, it is a natural choice to glue 3G networks and WLANs with the pervasive IP technology. In fact, we can see from the later sections that many mechanisms proposed for the interworking follow the de facto standards in the Internet community.

3.2 Authentication and authorisation support

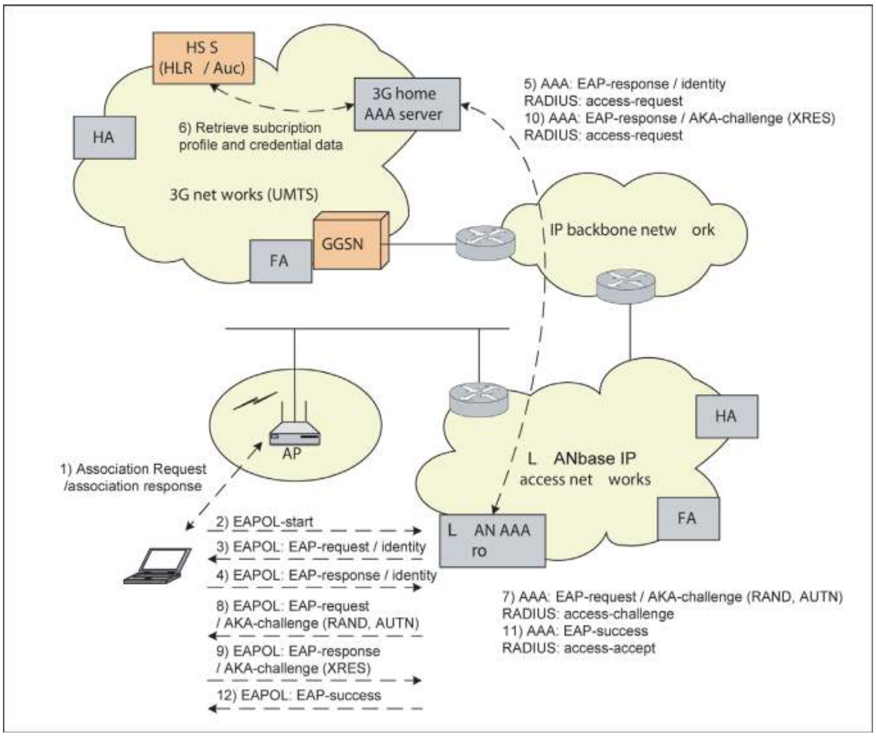
It is expected that in a 3G/WLAN integrated network the wireless terminal will be dual-mode, which means that the terminal will be equipped with network interfaces to both 3G networks and WLANs (Axiotis et al., 2004). However, only one subscription is needed with a 3G operator or a WLAN service provider, who has roaming agreements to support the interworking (Buddhikot et al., 2003a). A lot of research in the literature considers the scenario in which a 3G subscriber is provided WLAN access via independent WLAN systems.

Take the interworking of UMTS and 802.11 WLAN as an example. As mentioned in Section 2, 802.11 WLANs adopt 802.1X for access control, based on the EAP and AAA framework. EAP sets no restriction on specific authentication methods, while the AAA protocols, such as remote authentication dial-in user service (RADIUS) or its enhanced version Diameter, offer customised attributes in the authentication messages. The selection of RADIUS as the AAA protocol is not mandatory in the 802.1X or 802.11i; but it is in Wi-Fi protected access (WPA) for interoperability. Considering the flexibility and interoperability requirement, AAA server is introduced in UMTS (3GPP, 2004b). By means of EAP-AKA, the 3G home AAA server authenticates the UMTS subscribers roaming to a WLAN through the UMTS AKA procedure (Ahmavaara et al., 2003; Buddhikot et al., 2003a; Koien and Haslestad, 2003; Salkintzis, 2004), while EAP-SIM can be used for legacy GSM/GPRS system (Ala-Laurila et al., 2001; Salkintzis et al., 2002). The authentication information and subscription profile can be retrieved from the Home Subscriber Server (HSS) or HLR. Similar arrangements are implemented in the cdma2000/WLAN interworking.

There are three entities defined in 802.11X, that is, supplicant, authenticator and authentication server. It is natural that the wireless terminal and the 3G AAA server will perform the functions of the supplicant and the authentication server, respectively. The authenticator is responsible for relaying EAP frames sent by the supplicant via EAP-Over-LAN (EAPOL), and repacketing them into appropriate AAA messages for onward transmission to the authentication server, and vice versa. In the interworking mechanism, the functions of the authenticator can be carried by the AP (Salkintzis et al., 2004), the WLAN access router, or a separate WLAN AAA proxy (Salkintzis, 2004). A successful authentication procedure is summarised in Figure 2.

In the operator WLAN architecture proposed for legacy GSM/GPRS networks in Ala-Laurila et al. (2001), the access controller actually acts as the access router shown in Figure 1. It provides IP routing for WLANs and also relays authentication messages for the terminal and authentication server. As above, the signalling between the access controller and the authentication server is based on RADIUS, and the SIM-based authentication is executed with the aid of HLR. However, the authentication signalling between the mobile terminal and the access controller is the operator WLAN-specific Network Access Authentication and

Figure 2 Authentication procedure based on EAP-AKA via home AAA server



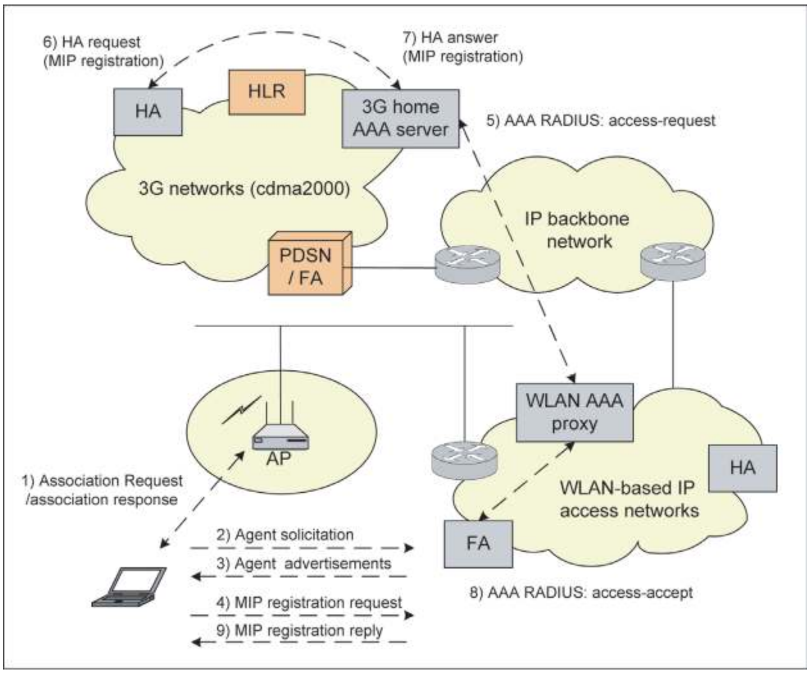
Accounting Protocol (NAAP), which encapsulates GSM authentication messages inside IP packets. This actually limits the applicability of the proposed interworking mechanism and is not compatible with current WLAN or IP standards.

The above approaches actually perform the authentication at layer 2. When MIP is supported in the 3G and WLAN networks, the authentication can also be carried out at the network (IP) layer (Buddhikot et al., 2003a). In the mobile IP mode of cdma2000, PDSN is augmented with Foreign Agent (FA) functionality, while HA is introduced to maintain location information of MSs and forward IP packets to MSs via tunnelling. For UMTS, the technical report 23.923 (not technical specification yet) (3GPP, 2000) outlines the plan to implement MIP in UMTS IP core network through three steps. In the first stage, the FA functionality is added to GGSN, which can support roaming between WLANs and UMTS via MIP. Stage II provides IP mobility for inter-GGSN handoff, while in stage III GGSN and SGSN are further combined to one Intelligent GPRS support node (IGSN) similar to the PDSN in cdma2000. With MIP being implemented, the MS roaming to a WLAN can be authenticated with the MIP registration procedure. An authentication example for cdma2000/WLAN interworking is illustrated in Figure 3. After obtaining a new Care-of Address (CoA) from the `Agent advertisements` of FA, the MS initiates the registration procedure by sending an `MIP registration request` message to the FA, which will contact the 3G home AAA RADIUS server to authenticate the roaming user via mobile-foreign challenge extension and mobile-AAA authentication extension (Perkins et al., 2004). In the mobile IP mode of cdma2000, neither Challenge Handshake Authentication Protocol (CHAP) nor Password Authentication Protocol (PAP) is performed as in the simple IP mode, because an additional AAA traversal will result in longer initial setup time and reestablishment time (3GPP2, 2002). After successful authentication by the RADIUS server, the `MIP registration request` of the MS will be sent to the HA via `HA request` message. After receiving `HA response` indicating the registration is validated by the HA, the AAA RADIUS server acknowledges the PDSN/FA with `RADIUS access accept`. The PDSN/FA will subsequently send an `MIP registration reply` message to the MS terminating the authentication and registration procedure.

In addition, it is possible to perform the authentication at the application layer. In Buddhikot et al. (2003a), it is proposed to authenticate the mobile node via a secured web page for login over a Hypertext Transport Protocol Secured (HTTPS) connection. This type of solution has restricted applications since it requires the support of specific services such as HTTP.

In summary, by introducing the flexible AAA framework, the 3G-specific authentication mechanisms can be reused. At the same time, independent WLAN service providers can implement their own preferred authentication methods, which are generally the popular standards in the Internet community, such as Extensible Authentication Protocol-Message Digest number 5 (EAP-MD5) and Extensible Authentication Protocol- Transport Layer Security (EAP-TLS). On the other hand, by coupling the authentication and authorisation procedure with mobility management, the signalling procedure can be simplified. However, the authentication and authorisation procedure should be efficient enough to minimise its impact on handoff latency. As for higher-layer authentication mechanisms, although more freedom is allowed for underlying technologies, the messaging delay and specific application restriction are their main drawbacks.

Figure 3 IP-level authentication procedure based on MIP registration



3.3 Mobility management across 3G networks and WLANs

Mobility management consists of two aspects, that is, location management and handoff management. Location management continuously tracks the locations of MSs, while handoff management maintains the ongoing connections while switching attachment points. The mobility protocols can operate at different layers such as link, network, transport and application layers (Banerjee et al., 2003). Link-layer mobility support is inherent in each wireless/mobile network, for example, the GMM in UMTS/GPRS networks. In the tightly coupled interworking architecture, the WLANs are integrated as any other 3G radio access networks. As a result, the 3G mobility management mechanism can be reused to support mobility across WLANs and 3G networks (Buddhikot et al., 2003b; Salkintzis et al., 2002).

The most popular network layer mobility protocol is Mobile IP (Perkins, 2002). By introducing mobility agents and IP tunnelling, the upper layer applications are enabled transparency to IP address changes due to user movement. However, the original MIP protocol suffers from the triangle routing problem. The packets have to be first routed to the HA before being tunnelled to the mobile node, while a mobile node sends its packets through a router on the foreign network assuming the routing is independent of source address. When this assumption is not valid, topologically correct reverse tunnels can also be established from the CoA to the HA. As a consequence, an extra delay results from the long route when the visited foreign network is far from the home network. This problem can be solved by route optimisation, in which a direct route is established between the mobile node and its correspondent node. In addition, the location registration to the HA may lead to a heavy signalling load when there are a large number of mobile nodes and also a long signalling delay if the foreign and home networks are far apart. Therefore, MIP is more suited for macromobility with infrequent movement and often between different administrative domains (interdomain). There are many MIP variants to support micromobility, which refers to mobility within one domain (intradomain). The handoff latency can be reduced by means of localising signalling via regional/hierarchical registration (tunneling-based) or host-specific routing (routing-based) (Campbell et al., 2002). For tunnel-based micromobility Protocols, there are Mobile IP regional registration, hierarchical MIP and Intra-Domain Mobility Management Protocol (IDMP). Typical routing-based micromobility protocols include cellular IP and HAWAII (Akyildiz et al. 2004).

In the loosely-coupled interworking architectures proposed in Salkintzis et al. (2004) and Buddhikot et al. (2003a), the MIP approach is used for the mobility across the cellular network and WLANs. Even though the loose coupling allows for the flexibility of implementing different individual mobility management schemes within each network, it has become a trend to apply MIP in 3G networks to pave the way for future all-IP networks. A cross-layer solution is proposed in Akyildiz et al. (2004) to support mobility in all-IP-based wireless networks with heterogeneous access technologies. The architecture can be applied to the 3G/WLAN interworking case. In the scheme, the aforementioned micromobility protocols are used for intradomain mobility, while a cross-layer mobility protocol is proposed for interdomain mobility. With early detection of possible interdomain handoff using link layer information, it is enabled to carry out authentication, authorisation, and MIP registration before the actual handoff. As a result, the interdomain handoff delay is reduced to be comparable to an intradomain handoff delay.

Moreover, transport layer mobility protocols can be used to prevent transport-layer applications such as Transport Control Protocol (TCP) connections from being interrupted by IP address changes due to user mobility. One solution in this type of mobility support is TCP-Migrate (Snoeren and Balakrishnan, 2000). Another transport layer scheme proposed in Ma et al. (2004) supports UMTS/WLAN vertical handoff via Stream Control Transmission Protocol (SCTP). Although performing mobility management at the transport layer provides network-independence, it requires more functions carried by end systems and also many modifications to current end systems.

Application layer mobility protocols can provide mobility support independent of underlying wireless access technologies and network layer protocols (Ma et al., 2004). Hence, they offer another alternative for mobility management in the 3G/WLAN interworked networks. A typical representative of application layer mobility protocols is Session Initiation Protocol (SIP) (Banerjee et al., 2003), which supports pre-call and mid-call mobility with the aid of application-specific redirect server. Although the application layer solution introduces less modification to existing protocol stacks and infrastructures of 3G networks and WLANs, longer handoff latency may be incurred due to the high-layer messaging. In Politis et al. (2004), a hybrid multilayer mobility management scheme is proposed. To handle macromobility, it uses MIP for non-real time services and SIP for real-time services. Existing protocols such as cellular IP, HAWAII and hierarchical MIP are employed to support micromobility. AAA context transfer is further introduced to avoid additional delay induced by the AAA security procedure during handoff.

In summary, for mobility management in 3G/WLAN integrated networks, accurate location information of MSs needs to be maintained, or the MSs can be effectively located by paging, so that the MSs can move freely across the two domains. Moreover, the vertical handoff is required to be fast, smooth and seamless (Gao et al., 2004). Especially for real-time services, the vertical handoff latency needs to satisfy strict delay bounds. Smooth handoff requires minimised packet loss during handoff, while the perceptible service interruption should be minimised to achieve the objective of seamless handoff. In a 3G/WLAN interworking environment involved with multiple domains, when enhanced with micromobility protocols, the Mobile IP approach is a good solution and can well meet the above requirements, especially when a hierarchical layered mobility management architecture is applied. The performance of MIP can be further improved with cross-layer design. Architecturally, the network layer is the right place to handle mobility (Banerjee et al., 2003). The higher-layer approaches such as the transport and application-layer mobility management schemes have the advantage of retaining the network interfaces unchanged. This feature applies well to the heterogeneous nature of 3G/WLAN integrated networks. Nonetheless, compared with the network-layer approaches, these higher-layer solutions are relatively immature and incomplete.

3.4 QoS provisioning architecture

The next-generation wireless networks are required to support multimedia services, which is actually a driving force for the 3G/WLAN interworking. Multimedia applications usually feature more bandwidth-demanding and stringent QoS requirements. We have mentioned in Section 2.3 the differences between 3G networks and WLANs in QoS support at the physical layer and MAC layer. This heterogeneous nature results in technical challenges in end-to-end QoS provisioning for a variety of services in an integrated 3G/WLAN network.

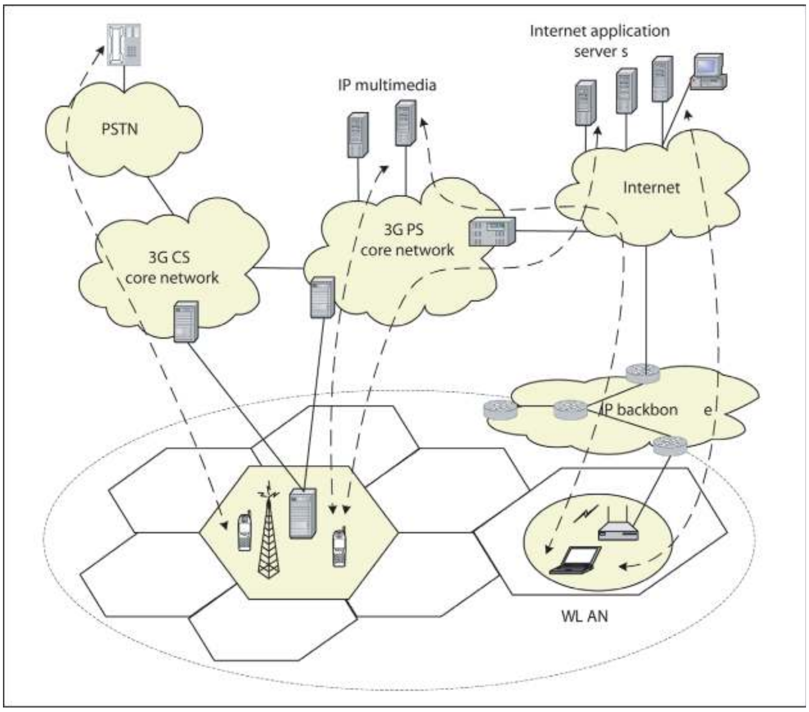
Figure 4 demonstrates the various scenarios requesting end-to-end services. It can be seen that different networks may be involved along the end-to-end path, for example, 3G radio access networks, 3G core networks, WLAN access networks, IP backbone and the wired Internet. In such a case, the following aspects should be well addressed in the QoS provisioning architecture (Fodor et al., 2003):

- *QoS attributes* – The traffic characteristics and end-to-end QoS requirements need to be clearly specified by the QoS attributes. These QoS attributes should be independent of the underlying access technologies, easily interpreted and mapped to service requirements for the networks involved along the end-to-end path.
- *QoS signalling protocols* – The QoS signalling protocols are responsible for QoS negotiation, resource reservation, etc. Although they are not required to be identical across different network domains, it is essential for the QoS attributes to be distributed via signalling along the end-to-end path without modification and being kept consistent.
- *QoS mechanisms* – To satisfy the QoS requirements presented to each network segment along the end-to-end path, proper local QoS mechanisms such as traffic control and scheduling are evoked to provide the requested QoS. Distinct QoS mechanisms are employed in 3G networks and WLANs, which results in different levels of QoS support capability.

A good overview of QoS provisioning in 3G networks is given in Dixit et al. (2001). For UMTS, a layered QoS architecture is specified (3GPP, 2003a). The end-to-end bear required to meet the QoS objectives is decomposed into a concatenation of bear services, which are provided by the different network segments involved. There are four traffic classes supported in UMTS, that is, conversational, streaming, interactive and background, with QoS attributes defined for each class, such as maximum bit rate, transfer delay and traffic handling priority. As for the QoS signalling, the UMTS Packet Data Protocol (PDP) context signalling mechanism is used for the bear services within UMTS. Diverse QoS schemes can be employed in the air interface, radio access networks and core networks to achieve the QoS objectives, for example, power control, bandwidth scheduling and transport technologies.

As mentioned in Section 3, IP technologies offer a common platform to integrate WLANs with 3G networks. In UMTS, the gateway node to external packet networks (i.e., GGSN) is required to provision an IP bearer service manager for interworking with external IP networks. For IP networks, there are two major QoS architectures, that is, Integrated Services (IntServ) and Differentiated Services (DiffServ). In the IntServ architecture, with the aid of Resource Reservation Protocol (RSVP), resources are reserved along the end-to-end path to satisfy the QoS

Figure 4 End-to-end service in 3G/WLAN interworked networks



requirements for each flow. However, the per-flow soft states must be maintained end-to-end and refreshed periodically to hold the reservation. As a result, this model is not well scalable for large networks with a large population.

On the other hand, DiffServ does not need a separate signalling protocol and it only exerts control for aggregate traffic. At the network edge, the traffic flows are classified into one of several classes according to the Service Level Specification (SLS) between the user and the access networks and between different network domains (Grilo et al., 2003). The Type Of Service (TOS) field of IP packets is marked with the Differentiated Service Code Point (DSCP) indicating the service class. Then the packets will receive priority-based treatment according to Per-Hop Behaviour (PHB) defined for each service class. PHBs are actually implemented by means of buffer management and packet scheduling mechanisms (Moon and Aghvami, 2003). In this way, service differentiation can be provided to different classes. In addition, the edge nodes can be augmented with functionalities such as traffic policing and admission control to prevent unacceptable QoS degradation due to saturation. As such, DiffServ only provides relative QoS to aggregate traffic flows instead of absolute QoS guarantee to individual flows as IntServ. Hence, although DiffServ has good scalability, only coarse QoS provisioning is enabled.

Because both IntServ and DiffServ models are proposed for wired IP networks, unique characteristics of wireless links and user mobility should be taken into account for extending them to wireless/mobile networks. The extension mechanisms of RSVP and DiffServ for IP-based wireless mobile networks are surveyed in Moon and Aghvami (2001) and Moon and Aghvami (2003), respectively. Especially, in the case of integrated 3G and WLANs networks, when vertical handoff occurs between the 3G networks and WLANs, the bear services offered by the two access technologies provide different QoS support. These changes should be considered in the QoS context transfer and service reestablishment after handoff.

In Manner et al. (2002), an IP QoS model combining IntServ and DiffServ is proposed for heterogeneous wireless networks. In the wireless access networks, IntServ architecture is applied, since the traffic volume is relatively low and scalability will not be a severe problem. Also, the scarce wireless bandwidth can be efficiently utilised by properly allocating resources to each traffic flow. On the other hand, the core networks adopt the DiffServ architecture for scalability purpose. At the edge of the sender's access networks, the RSVP signalling for end-to-end resource reservation is mapped to DiffServ PHBs, and the RSVP messages are tunnelled through the DiffServ domains till the receiver end. The DiffServ approach for core networks is further justified by the bandwidth over-provisioning in current core networks and the relatively low cost of upgrading transmission capability (Marques et al., 2003). A similar idea is used in the QoS architecture proposed in Grilo et al. (2003), where RSVP is slightly modified. Instead of exchanging the RSVP PATH and RSVP RESV messages end-to-end from the sender to the receiver, the RSVP signalling is localised within the access networks via proxies. This facilitates the reservation path repair after handoff. In addition, it is proposed in Alam et al. (2001) that Multi Protocol Label Switching (MPLS) can be introduced in the DiffServ domain. Due to the intrinsic QoS provisioning capability, MPLS can facilitate scalability by means of flow aggregation and at the same time guarantee individual QoS without the need to maintain per-flow awareness along the path (Alam et al., 2001).

To achieve better QoS support, mobility management and QoS provisioning can be coupled in the architecture design. In Lo et al. (2004), a mobility and QoS architecture is proposed for all-IP based 4 G wireless networks. It takes advantage of the two-layer hierarchical structure of IDMP mobility management to reduce the resource reservation delay due to RSVP reestablishment after handoff. The per-flow RSVP reservation is terminated at a fixed anchor subnet agent, while a forwarding chain is used to track the host mobility within a domain. As such, the reservation messages can be effectively restricted within a local scope.

Another IP-based QoS architecture is proposed in Marques et al. (2003) for 4G networks supporting multiple accesses and multiple service providers. In this architecture, the end-to-end QoS support is integrated with mobility and Authentication, Authorisation, Auditing and Charging (AAAC). In each network domain, an AAAC system is used to handle the network access control, while at least one QoS Broker (QoSB) manages the access network resources. The Common Open Policy Service (COPS) protocol is used for information exchanging between QoSBs, edge routers and AAAC systems, because in this architecture QoSBs and AAAC systems act as Policy Decision Points (PDPs) while edge routers act as Policy Execution Points (PEPs).

In Zhuang et al. (2003), a policy-based multidomain QoS architecture is proposed to provide consistent QoS control over an integrated UMTS and WLAN system. The architecture varies with different interworking scenarios such as interworking with a UMTS operator's WLAN, a WLAN shared by multiple operators or a WLAN of a UMTS operator's customer. This policy-based type of approach views the development of proper QoS architecture for 3G/WLAN interworking from a new perspective. However, there are still some problems that need more careful considerations, such as fastening policy negotiation and securing information exchanging between interconnected policy entities (Zhuang et al., 2003).

4 Concluding remarks

In this paper, we have introduced the complementary characteristics of 3G networks and WLANs and the motivations for their interworking. The differences of the two wireless technologies lead to many challenging issues that need to be addressed in the interworking. To achieve seamless roaming, enhanced security and consistent QoS, there are a variety of interworking mechanisms proposed from different perspectives. In particular, we examine the typical solutions for user authentication, mobility management and QoS provisioning in the 3G/WLAN interworked networks. It is observed that IP technologies play an important role in gluing the two heterogeneous technologies. The demand for ubiquitous Internet access further strengthens this convergence trend. For instance, the well-known AAA framework, MIP protocol, IntServ/RSVP and DiffServ QoS architectures are considered as promising approaches for interworking 3G networks and WLANs.

However, most current research on 3G/WLAN interworking focuses on relatively high-level architecture issues. Although the local QoS mechanisms implemented in the 3G networks and WLANs can provide certain levels of QoS support, proper resource allocation schemes are needed to utilise the integrated resources effectively and efficiently. In Song et al. (2005), we investigate the resource allocation problem in 3G/WLAN integrated networks. In the proposed resource allocation scheme, we try to take advantage of the characteristics of this heterogeneous network with multiple services, so that better QoS assurance and higher utilisation can be achieved. In Luo et al. (2003), a radio resource scheduling scheme is proposed for a 3G/WLAN coupled network. Based on the traffic characteristics, QoS requirements, and network states, a joint scheduler prioritises and splits the traffic over the two networks. A certain amount of resources in each network is allocated to the corresponding traffic substream to achieve the target QoS. A higher utilisation is available due to a larger trunking gain and effective manipulation of the overall resources of the two networks. However, these are only some initial research attempts for the resource allocation problem. It remains to be an open issue that needs more in-depth investigation.

References

Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J. and Levkowetz, H. (Eds) (2004) 'Extensible authentication protocol (EAP)', IETF RFC 3748.

Ahmavaara, K., Haverinen, H. and Pichna, R. (2003) 'Interworking architecture between 3GPP and WLAN systems', *IEEE Communications Magazine*, Vol. 41, No. 11, pp.74–81.

Akyildiz, I.F., Xie, J. and Mohanty, S. (2004) 'A survey of mobility management in next-generation all-IP-based wireless systems', *IEEE Wireless Communications*, Vol. 11, No. 4, pp.16–28.

Ala-Laurila, J., Mikkonen, J. and Rinnemaa, J. (2001) 'Wireless LAN access network architecture for mobile operators', *IEEE Communications Magazine*, Vol. 39, No. 11, pp.82–89.

Alam, M., Prasad, R. and Farserotu, J.R. (2001) 'Quality of service among IP-based heterogeneous networks', *IEEE Personal Communications Magazine*, Vol. 8, No. 6, pp.18–24.

Axiotis, D.I., Al-Gizawi, T., Protonotarios, E.N., Lazarakis, F.I., Papadias, C. and Philippopoulos, P. (2004) 'Services in interworking 3G and WLAN environments', *IEEE Wireless Communications*, Vol. 11, No. 5, pp.14–20.

Banerjee, N., Wu, W. and Das, S.K. (2003) 'Mobility support in wireless Internet', *IEEE Wireless Communications*, Vol. 10, No. 5, pp.54–61.

Buddhikot, M.M., Chandranmenon, G., Han, S., Lee, Y-W., Miller, S. and Salgarelli, L. (2003a) 'Design and implementation of a WLAN/cdma2000 interworking architecture', *IEEE Communications Magazine*, Vol. 41, No. 11, pp.90–100.

Buddhikot, M.M., Chandranmenon, G., Han, S., Lee, Y.W., Miller, S. and Salgarelli, L. (2003b) 'Integration of 802.11 and third-generation wireless data networks', *Proceedings of the IEEE INFOCOM'03*, Vol. 1, pp.503–512.

Campbell, A.T., Gomez, J., Kim, S., Wan, C-Y., Turanyi, Z.R. and Valko, A.G. (2002) 'Comparison of IP micromobility protocols', *IEEE Wireless Communications*, Vol. 9, No. 1, pp.72–82.

Dixit, S., Guo, Y. and Antoniou, Z. (2001) 'Resource management and quality of service in third generation wireless networks', *IEEE Communications Magazine*, Vol. 39, No. 2, pp.125–133.

Edney, J. and Arbaugh, W. (2003) *Real 802.11 Security: Wi-Fi Protected Access and 802.11i*, Addison Wesley Professional.

ETSI (2001) 'Requirements and architectures for internetworking between HIPERLAN/2 and 3rd generation cellular systems', ETSI TR 101 957.

Fodor, G., Eriksson, A. and Tuoriniemi, A. (2003) 'Providing quality of service in always best connected networks', *IEEE Communications Magazine*, Vol. 41, No. 7, pp.154–163.

Gao, X., Wu, G. and Miki, T. (2004) 'End-to-end QoS provisioning in mobile heterogeneous networks', *IEEE Wireless Communications*, Vol. 11, No. 3, pp.24–34.

3GPP (2000) 'Combined GSM and Mobile IP mobility handling in UMTS IP CN', 3GPP TR 23.923 V.3.0.0.

3GPP2 (2002) 'Wireless IP network standard', 3GPP2 P.S0001-B Version 1.0.0.

3GPP (2003a) 'Quality of service (QoS) concept and architecture', 3GPP TS 23.107 V5.12.0.

3GPP (2003b) 'Feasibility study on 3GPP system to wireless local area network (WLAN) interworking', 3GPP TR 22.934 V6.2.0.

3GPP (2004a) '3GPP system to wireless local area network (WLAN) interworking; system description (release 6)', 3GPP TS 23.234 V6.3.0.

3GPP (2004b) 'Wireless local area network (WLAN) interworking security (release 6)', 3GPP TS 33.234 V6.3.0.

Grilo, A., Nunes, M., Sergio, G. and Ciulli, N. (2003) 'Integration of IP mobility and QoS for heterogeneous wireless access in MOICANE', *Proceedings of the IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'03)*, Vol. 1, pp.470–475.

Koien, G.M. (2004) 'An introduction to access security in UMTS', *IEEE Wireless Communications*, Vol. 11, No. 1, pp.8–18.

Koien, G.M. and Haslestad, T. (2003) 'Security aspects of 3G-WLAN interworking', *IEEE Communications Magazine*, Vol. 41, No. 11, pp.82–88.

Lo, S-C., Lee, G., Chen, W-T. and Liuand, J-C. (2004) 'Architecture for mobility and QoS support in all-IP wireless networks', *IEEE Journal of Selected Areas on Communications*, Vol. 22, No. 4, pp.691–705.

Luo, J., Mukerjee, R., Dillinger, M., Mohyeldin, E. and Schulz, E. (2003) 'Investigation of radio resource scheduling in WLANs coupled with 3G cellular network', *IEEE Communications Magazine*, Vol. 41, No. 6, pp.108–115.

Manner, J., Burness, L., Hepworth, E., Lopez, A., and Mitjana, E. (2002) 'Provision of QoS in heterogeneous wireless IP access networks', *Proceedings of the IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, Vol. 2, pp.530–534.

Marques, V., Aguiar, R.L., Garcia, C., Moreno, J.I., Beaujean, C., Melin, E. and Liebsch, M. (2003) 'An IP-based QoS architecture for 4G operator scenarios', *IEEE Wireless Communications*, Vol. 10, No. 3, pp.54–62.

Ma, L., Yu, F., Leung, V.C.M. and Randhawa, T. (2004) 'A new method to support UMTS/WLAN vertical handover using SCTP', *IEEE Wireless Communications*, Vol. 11, No. 4, pp.44–51.

Moon, B and Aghvami, H. (2001) 'RSVP extensions for real-time services in wireless mobile networks', *IEEE Communications Magazine*, Vol. 39, No. 12, pp.52–59.

Moon, B., and Aghvami, H. (2003) 'DiffServ extensions for QoS provisioning in IP mobility environments', *IEEE Wireless Communications*, Vol. 10, No. 5, pp.38–44.

Pang, A-C., Chen, J-C., Chen, Y-K. and Agrawal, P. (2004) 'Mobility and session management: UMTS versus cdma2000', *IEEE Wireless Communications*, Vol. 11, No. 4, pp.30–43.

Perkins, C. (Education) (2002) 'IP mobility support', IETF RFC 3344.

Perkins, C.E., Calhoun, P.R. and Bharatia, J. (2004) 'Mobile IPv4 challenge/response extensions (revised)', Internet Draft, draft-ietf-mip4-rfc3012bis-03.txt.

Pichna, R., Ojanpera, T., Posti, H. and Karppinen, J. (2000) 'Wireless Internet – IMT-2000/wireless LAN interworking', *Journal of Communications and Networks*, Vol. 2, No. 1, pp.46–57.

Politis, C., Chew, K.A., Akhtar, N., Georgiades, M., Tafazolli, R. and Dagiuklas, T. (2004) 'Hybrid multilayer mobility management with AAA context transfer capabilities for all-IP networks', *IEEE Wireless Communications*, Vol. 11, No. 4, pp.76–88.

Rose, G. and Koien, G.M. (2004) 'Access security in CDMA2000, including a comparison with UMTS access security', *IEEE Wireless Communications*, Vol. 11, No. 1, pp.19–25.

Salkintzis, A.K., Fors, C. and Pazhyannur, R. (2002) 'WLAN-GPRS integration for next generation mobile data networks', *IEEE Wireless Communications*, Vol. 9, No. 5, pp.112–124.

Salkintzis, A.K. (2004) 'Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks', *IEEE Wireless Communications*, Vol. 11, No. 3, pp.50–61.

Snoeren, A.C. and Balakrishnan, H. (2000) 'An end-to-end approach to host mobility', *Proceedings of the ACM/IEEE 6th Annual International Conference on Mobile Computing and Networking*, pp.155–166.

Song, W., Jiang, H., Zhuang, W. and Shen, X. (2005) 'Resource management for QoS support in cellular/WLAN interworking', *IEEE Network Magazine*, Vol. 19, No. 5, pp.12–28.

Zhuang, W., Gan, Y-S., Loh, K-J., and Chua, K-C. (2003) 'Policy-based QoS-management architecture in an integrated UMTS and WLAN environment', *IEEE Communications Magazine*, Vol. 41, No. 11, pp.118–125.

Zhu, H., Li, M., Chlamtac, I. and Prabhakaran, B. (2004) 'A survey of quality of service in IEEE 802.11 networks', *IEEE Wireless Communications*, Vol. 11, No. 4, pp.6–14.