A SEAMLESS MOBILE VPN DATA SOLUTION FOR UMTS AND WLAN USERS

P M Feder[†], N Y Lee[†], S Martin-Leon[‡]

Bell Laboratories - Mobility Solutions, Lucent Technologies Inc., [†]USA, [‡]UK

Abstract – *Mobile virtual private networks (MVPNs) can provide remote users with easy, secure high-speed access to their enterprise network resources. There is a tremendous market opportunity for operators who can meet the needs of these users. Third-generation (3G) systems, such as Universal Mobile Telecommunications System (UMTS), and IEEE 802.11b wireless local area network (WLAN) systems have complementary strengths. For end users, integrating these systems provides ubiquitous high-speed data coverage, continuous selection of the highest possible data rates and seamlessly maintained user VPN sessions when moving between air interface technologies. For operators, they enable integrated billing and offload UMTS network capacity for higher revenue voice subscribers. In this article we discuss interworking architecture for providing integrated service capability across UMTS and 802.11b based networks. We present tight and loose coupling choices for integration, recommend the loosely-coupled interworking approach, and show how it offers a comprehensive and secure transport layer solution across UMTS and WLAN air interface technologies.*

I. INTRODUCTION

The deployment of third-generation UMTS systems is expected to provide high-speed wireless data service. Early adopters will include enterprises with a highly mobile workforce that values real-time access to secure data and applications residing in the enterprise (corporate intranet) network. Examples of corporate users that may realize significant productivity gains include mobile executives, sales agents, and field technical/operations employees. Such users are seeking high-speed secure connections, good coverage, mobility, and ease of use. Furthermore, enterprises expect these wireless systems to provide intranet accessibility without compromising security, even when access is through a public wireless infrastructure and Internet protocol (IP) networks.

The technique of choice for secure high-speed intranet connectivity today is virtual private networking. Virtual private networks (VPNs) typically use wireline access technologies such as Digital Subscriber Lines (DSLs) and are carried over a public network infrastructure.

Mobile VPNs (MVPNs) extend this concept to environments with a wireless access infrastructure, thus making an enterprise's intranet accessible to mobile users. Mobile users include those who are stationary while connected but portable (connected from different locations), as well as those who are moving. Enterprises also prefer to deal with a minimum number of network operators.

MVPNs are expected to become a major enabling technology for mobile operators to target the very profitable segment of on-the-move business data users. MVPNs offer data users potential connectivity to their enterprise from anywhere wireless access is provided. MVPN users will expect the same intranet access from UMTS networks, public wireless local area networks (WLANs), and their enterprise's private WLANs. UMTS and WLAN access technologies have different mobility characteristics. This is an enormous market opportunity for those wireless carriers who can meet the demands of roaming data users. An integrated VPN service architecture is needed to provide a common service offering across the various wireless technologies.

This paper examines and recommends a solution for how two classes of wireless access networks, WLANs such as IEEE 802.11b and UMTS, can be integrated to provide MVPN service to these data users. IEEE 802.11b and UMTS systems can be integrated to varying degrees. Integration is needed to meet the customer goal of easy to use, secure data connectivity anywhere one of the two access technologies is available. When it come to offering bundled services, existing mobile operators have the advantage over new entrants of being able to leverage their existing infrastructure, billing systems, and customer base. In order to do that, they require a solution that fits easily with their existing back-office.

We begin with the motivation for integrating WLAN with UMTS systems and then compare two approaches to integration, tightly-coupled and loosely-coupled interworking and recommend the latter. Next, we describe the different attributes of Simple-IP and Mobile-IP services and define MVPNs and the various models for providing such services to enterprise users.

II. MOTIVATION FOR INTERWORKING OF WIRELESS ACCESS TECHNOLOGIES

Each wireless access technology has its strengths and weaknesses. Successful integration of technologies with complementary strengths benefits both the mobile operator and the enterprise customer.

IEEE 802.11b (IEEE (1)) was initially designed to extend enterprise networks without the burdensome cabling required for fixed LANs. With its use of unlicensed spectrum, relatively simple deployment, and very low equipment cost, some regard IEEE 802.11b as a direct competitor to third generation (3G) data service. In fact, a new breed of operators known as wireless Internet service providers (WISPs) is providing broadband access using IEEE 802.11b in hot spots, i.e., public areas such as airports, hotels, conference centres and cafés. However, IEEE 802.11b's limited range and lack of mobility management make it an impractical technical and economic alternative to 3G for ubiquitous wide area coverage. Furthermore, mobile operators are better positioned than WISPs to target mobile data users because of their well-established marketing and customer care infrastructures.

IEEE 802.11b networks can best serve as a highly effective complement to UMTS systems, providing service in specific locations where a high concentration of MVPN users is expected. This enables operators to plan their UMTS networks more efficiently, without having to cater to areas of high user concentration. Offloading packet data traffic to WLAN networks also increases the voice capacity available in UMTS systems where spectrum is limited. Furthermore, operators can capitalize on the fact that many enterprise users already have IEEE 802.11b-enabled laptops and personal digital assistants (PDAs) to pull through initial demand for UMTS MVPN services.

III. ARCHITECTURAL CHOICES

Buddhikot et al (2) discusses two ways to integrate WLAN Access with UMTS systems: tightly-coupled interworking and loosely-coupled interworking.

a. Tightly-Coupled Interworking

In a tightly-coupled approach the WLAN network is connected to the UMTS core network in the same manner as the UMTS radio access elements (Figure 1, center). The WLAN network emulates functions natively available in UMTS access networks. In this architecture the WLAN inter working gateway element appears to the UMTS core network as an SGSN. The WLAN gateway hides the details of the 802.11 network from the UMTS core network and implements the required UMTS protocols. Mobile Nodes (MNs) must run the corresponding UMTS protocol stack on top of their standard 802.11 network interface cards and switch from one physical layer to the other as needed. All 802.11 client traffic is injected using UMTS protocols into the core network. The networks share the same authentication, signalling, transport and billing infrastructures.

This approach presents several disadvantages. First, exposing the UMTS interface can imply security threats. Second, injecting 802.11 traffic directly into the UMTS core upsets the carefully engineered, UMTS-optimised capacity and configuration of the network. The entire network setup – including SGSN, router and GGSN – must be re-engineered to sustain the increased load and differing traffic characteristics. Client device configuration also presents issues. First, there is the need to implement the UMTS protocol stack. Second, Universal Subscriber Identity Module (USIM) authentication mechanism ETSI (3) must be used for authentication, implying the use of 802.11 cards with built-in USIM slots or external cards separately plugged into the subscriber devices. The authentication mechanism also forces WLAN operators to interconnect to the UMTS operator's SS7 network.

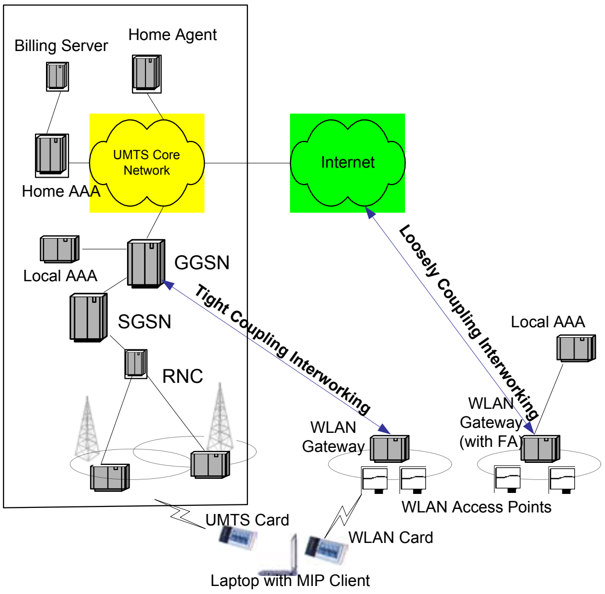

Figure 1. WLAN and UMTS integration: Tightly Coupled (center) vs. Loosely Coupled (right)

b. Loosely-Coupled Interworking

The loosely-coupled approach (Figure 1, right) also calls for the introduction of a WLAN gateway. However, in this design the gateway, which could be integrated in the edge router, connects to the Internet and does not have any direct link to the UMTS core network. The 802.11 network may serve users visiting from other networks as well as local subscribers. This approach completely separates the data paths in the 802.11 and UMTS networks: the 802.11 data traffic is never injected into the UMTS core network, yet the end user still receives seamless access.

In the approach, different mechanisms and protocols can handle authentication, billing and mobility management in the UMTS and 802.11 portions of the network. However, for seamless operation with UMTS, this requires that the 802.11 edge router, or gateway, provides mobility management for access across both networks, as well as authentication, authorization, accounting (AAA) services to interwork with the UMTS networks. This enables the UMTS operator to collect the 802.11 accounting records and generate a unified billing statement indicating usage and various price schemes for both networks. At the same time, the use of compatible AAA services on the two networks would allow the 802.11 gateway to dynamically obtain per-user service policies from their home AAA servers and to enforce and adopt such policies to the 802.11 network.

The evolving UMTS standards (3), ETSI (4) are slowly embracing the IETF based AAA and Mobile-IP protocols Perkins (5), Rigney et al (6), Calhoun et al (7). As these protocols gain acceptance and are integrated into the UMTS core network, adaptation of other access technology such as WLAN will be easily achieved. Integrating Mobile IP (MIP) functionality in the GGSN as indicated in (4) will enable seamless mobility between 802.11 and UMTS. Until this is achieved, co-located Care-of-Address scheme can be used requiring a locally obtained address to serve as the Care-of-Address entity at the UMTS network, limiting the transparency of the address acquisition while roaming across SGSN/GGSN boundaries. Common subscriber database would need to interface to Home Location Register (HLR) for authentication and billing on the UMTS side of the network and to AAA servers for the same operations to be performed while clients roam to the 802.11 networks.

There are several advantages to the loosely-coupled integration approach. It allows the independent deployment and traffic engineering of 802.11 and UMTS networks. UMTS operators can benefit from other operators' 802.11 deployments without extensive capital investments. At the same time, they can continue to deploy 3G networks using well-established engineering techniques and tools. Furthermore, while roaming agreements with many partners can result in widespread coverage, including key hot-spot areas, users benefit from having just one subscription for all network access. They no longer need to establish separate accounts with mobile operators in different regions or covering different access technologies. Unlike the tightly-coupled approach, this architecture allows a mobile operator to provide its own public 802.11 hot-spots, interoperate through roaming agreements with public 802.11 and UMTS operators, and/or manage privately installed enterprise WLANs.

The loosely-coupled approach clearly offers several architectural advantages over the tightly-coupled approach with virtually no drawbacks. Therefore, we advocate the loosely-coupled approach as the preferred architecture for the integration of 802.11 and UMTS networks.

The loosely-coupled approach can also support integrated authentication and unified billing, enabling mobile operators to bundle IEEE 802.11b with UMTS services, retaining ownership of roaming customers and reducing churn. For more discussion on billing and authentication please refer to Benenati (8).

IV. TWO TYPES OF DATA SERVICES

The minimum elements required for UMTS and WLAN integration are a WLAN gateway and AAA infrastructure. For seamless handover, Mobile IP elements are required. This section describes the basic Simple-IP service and the more advanced Mobile-IP service available in UMTS and WLAN networks.

a. Simple-IP Services

WLAN. We recommend that the WLAN network and its gateway support the IEEE 802.1x standard (IEEE (9)) and that the Mobile Node client performs the required 802.1x layer-2 authentication prior to requesting an IP address. The address may be private provided that the WLAN gateway provides Network-Address-Translation (NAT). For non-802.1x deployment, web-based authentication is required. If the user moves to a new subnet location, a new IP address is required and the IP sessions are lost. Therefore, simple-IP service is appropriate when portability but not mobility is desired. The principal advantage of Simple-IP service is that it does not require any specialized client software.

UMTS. Access to the data network is achieved through PDP context activation where either PPP or IP service is established upon device authentication. When AAA infrastructure is added to the UMTS core, network user authentication (NAI and password) can also be added. For ease of HLR administration, a global APN feature can be supported. This allows the mobile operator to utilize a single APN when the subscriber is accessing enterprise resources. From the mobile operator's point of view it can present single virtual APN to all types of enterprise access and map the virtual APN to a unique real APN for each of the enterprises served by the UMTS operators. User NAI can be composed of the identifier "user@domain", where the domain name – UMTSoperator.enterprise.com – corresponds to a unique real APN and the identifier spells out the AAA authentication hierarchy.

b. Mobile-IP Services

Mobile IP services enable loosely-coupled interworking architectures to support session continuity when users roam among heterogeneous networks of different operators and different access technologies. It provides IP-based layer 3 mobility management between air interface technologies, permitting the air interface and data link layers of the various access technologies to evolve independently. MIP also makes user mobility across IP subnets transparent to applications and provides strong authentication techniques. Thus, the MIP set of protocols (5) is well-suited to provide mobility and MVPN connectivity independent of the underlying access technology without requiring any customized WLAN equipment or changes to the well-engineered UMTS network. That is, MIP is an enabler of loosely-coupled interworking. Use of MIP in UMTS is specified in the 3GPP standards ETSI (10).

MIP defines three network components: the foreign agent (FA), the home agent (HA), and the mobile node (MN), which contains the MIP client software. The FA function belongs in the GGSN for UMTS and in an edge router or WLAN gateway for WLAN. The HA function may be attached to the GGSN for UMTS, or in an edge router for a third-party or the enterprise's data centre.

The MIP client software resides in the MN, typically a laptop or PDA. The client is a key element of the layer 3 integration. The MIP client includes multiple network interfaces and is capable of exchanging network and data link messages with any of the wireless and/or wireline interfaces it supports. The MIP client must keep all upper layer data sessions open.

An MN may be in one of three possible operating environments typified in existing UMTS systems. In each, the MN uses a previously assigned identification: either its own static home IP address or its own network access identifier (NAI), such as user@domain. In the first operating environment, the MN is attached to its home network. While the MN is located at home, the network follows standard routing procedures based on normal source and destination IP addresses.

In the second operating environment, the MN attaches through the access technology to a foreign network that contains an FA and supports MIP. It determines that there is an FA when it receives an MIP agent advertisement. The MN sends a Registration Request message to the FA, which forwards it to the HA and HA-AAA, where it is authenticated and authorized. When the HA is accessed over the public Internet, a security association (SA) between the FA and HA is required for mutual FA and HA authentication, and a secure VPN tunnel is established between the FA and HA. Once the tunnel is established, packets destined for the MN are forwarded to the MN's home address, where they are intercepted by the HA, which tunnels them to the current care-of-address (COA) of the FA to which the MN is attached.

In the third operating environment, the MN attaches to a network that does not have an FA. The MN nevertheless is able to use MIP as long as it can acquire an IP address (through WLAN dynamic host control protocol [DHCP] or PDP context activation). The MN uses the assigned IP address as its MIP co-located COA and acts as its own tunnel endpoint. It registers directly with the HA using this co-located COA and the MIP registration procedure. The MN sends packets out directly to their destination. All returning packets go to the MN's home address and are intercepted by the HA, which tunnels the packets to the MN.

The client also handles user authentication and authorization. The client provides authentication credentials to gain access to the domain and, if it is in a visited domain, to establish an SA with its home network or home AAA server. The domain field in the user NAI allows the visited AAA server to determine the home domain that will be used to authenticate and bill the user.

The MN client can simultaneously monitor wireless signal strengths of multiple access technologies and automatically perform access technology handovers based on a pre-defined criteria matrix that may include available bandwidth, signal quality, access cost and SLA agreements between its mobile operator and the visited network. This matrix may be configured by the mobile operator, user, or both. Service disruption and packet loss are minimised when UMTS and WLAN radio coverage overlap because the MN has both network interfaces active, and the MN can use another link in the background to set up the call prior to handover. Figure 2a depicts such a "make before break" scenario where the client observes the WLAN signal over time. At time t_1, when the signal strength exceeds the threshold *H,* and the client attempts to use the WLAN interface. Similarly at time t_2, when the signal strength drops below the threshold *L,* the client reverts to the UMTS airlink. If the client sets up the call and creates a binding (tunnel) at the HA before losing the signal on the current interface, the delays Δ_1 and Δ_2 are masked and handover appears instantaneous to the client application. Packets in transit continue to arrive at the old interface and are not lost. Outgoing packets are routed to the new airlink immediately after switching, minimising packet loss.

In the absence of overlapped coverage, there will be service interruption and packet loss. This "break before make" handover is illustrated in Figure 2b, where the disruption occurs between t_2 and $t_2 + \Delta_2$. This incurred delay could be large (>2 seconds) when a UMTS call is established in the absence of WLAN coverage, or much shorter (tens to hundreds of milliseconds) if a UMTS call is kept up continuously (a client option) and the switching time is only determined by the performance of network latency between the FA, HA, and AAA network elements.

V. MVPNs

Until recently, most VPN services used layer 2 session establishment protocols such as PPP and layer 2

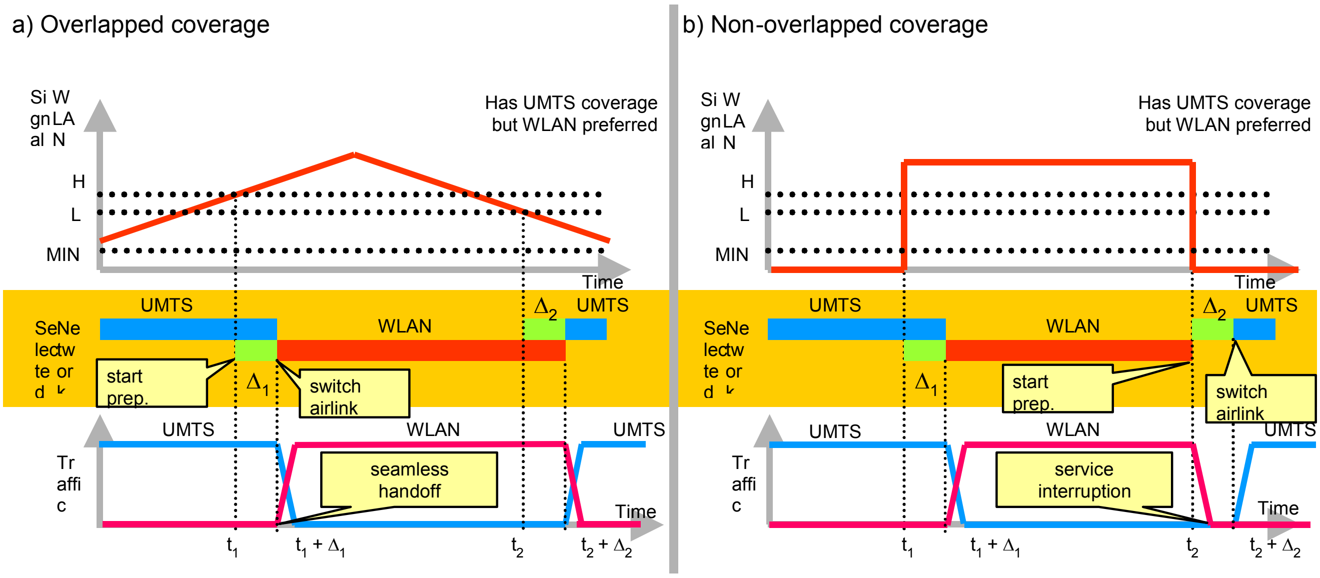

Figure 2. Make Before Break Switching for Overlapped Coverage and Break Before Make Switching for Non-Overlapped Coverage

transport protocol (L2TP). PPP provides methods to authenticate and authorize subscribers to the network as part of session establishment. A PPP link can be extended through an L2TP tunnel to an L2TP network server (LNS) residing at the enterprise location to provide remote access. More recently, VPN service solutions that are based on layer 3 secure tunnels (IPSec) have been gaining popularity and are being deployed, replacing the traditional Layer 2 Tunnelling Protocol (L2TP) network servers.

a. IP Security (IPSec)

IPSec provides layer 3 security by authenticating users and encrypting their data, thus preventing eavesdropping, spoofing the users' identities, and sniffing their traffic. It provides the capability to secure communication across WLAN, wide area networks, and public networks such as the Internet. IPSec components include Internet key exchange (IKE) Harkins and Carrel (11), authentication header (AH) Kent and Atkinson (12), and encapsulation security payload (ESP) mechanisms Kent and Atkinson (13). The latter two are security protocols based on distributed cryptographic keys. The most common algorithms currently used for authentication are message digest MD5 Rivest (14) and secure hash algorithm SHA-1 NIST (15). For encryption, data encryption standard (DES) and triple DES (3DES) are the most common algorithms used.

b. Interoperability of IPSec and MIP

In the protocol stack IPSec can be used on top of MIP to provide secure and seamless MVPN access to enterprise resources across multiple technologies. In such a solution, the user initiates an IPSec session after the MIP session is established and end-to-end tunnels between the client and the enterprise's IPSec gateway are maintained while the user roams across different foreign networks. The IPSec and MIP clients must coexist and interoperate in the MN. The user-initiated IPSec tunnel is transported within an existing FA to HA tunnel. Figure 3 depicts the client components embedded within the Windows OS protocol stack.

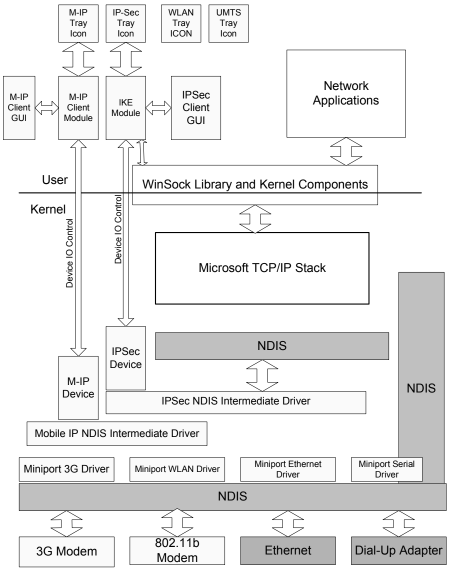

Figure 3. Mobile IP and IPSec drivers and components embedded within the Window stack

VI. END-TO-END SECURED MVPN ENTERPRISE MODEL

End-to-end secured tunnelling is one model for secured

MVPN. As shown in Figure 4, this consists of a user-initiated tunnel from the MN terminating at the enterprise IPSec gateway. End-to-end tunnels provide simple, secure access to a private network. Typically, the IPSec gateway is managed by the enterprise, enabling it to reuse the same VPN infrastructure for any access technology. UDP encapsulation is required to support network address translation if private IP addresses are used, Huttunen (16). Aggregation of traffic from multiple users to a single corporate VPN gateway is not possible, making the solution difficult to scale. Although the operator cannot provide additional value (such as quality of service or egress filtering), this is a very popular solution because of its simplicity, high security, and access network operator independence.

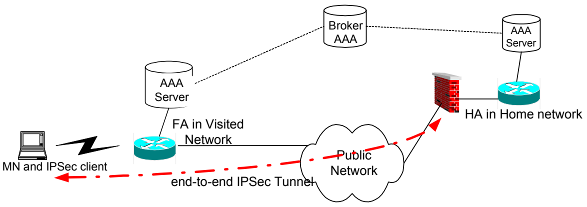

Figure 4. End-to-End IPSec Tunnelling

VII. NETWORK-BASED MVPN ENTERPRISE MODEL

Another model for MVPNs is the network-based VPN model. This model (Figure 5) consists of two parts, a user-initiated tunnel that terminates in the mobile operator's core network and a statically configured tunnel from the core network to the user's enterprise network. This tunnel aggregates traffic from multiple subscribers, making the solution highly scalable.

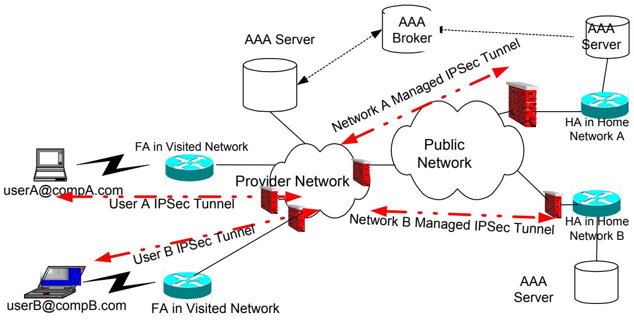

Figure 5. Network-Based IPSec Tunnelling

Alternatively, the user-to-mobile operator tunnel may be omitted if the user considers the radio link spreading and/or encryption to be sufficient, thus eliminating IPSec overhead on the air interface.

In contrast to the end-to-end secured MVPN model, the network-based approach enables operators to provide services such as Web caching, Web proxy, network-based firewall, and off-load functionality wherein certain traffic is routed directly to the Internet or to a server inside the operator's network.

VIII. CONCLUSION

This paper introduced the concept of MVPN service and identified enterprise users as the customers for such services. It presented an overview of the requirements and listed two approaches to interworking. MIP and associated tunnelling protocols solve the challenges of mobility and authentication at the transport layer. IPSec addresses the security concerns with moderate overhead penalty. The result for the end user is ubiquitous, mobile, optimised data coverage in a secure and seamlessly maintained connection. The enhanced access experience of integrated UMTS and IEEE 802.11 systems may stimulate data usage and adoption. However, intelligent MN clients and UMTS standard adaptation is required to shield the user from multiple authentication requests and reduce authentication latency.

With the proliferation of WLANs in public and enterprise environments, interoperability between such networks and UMTS systems can be achieved today through loose coupling, providing all the required mechanisms needed by a mobile operator to attract and retain data users.

It is expected that usage of AAA and MIP within UMTS will soon be adopted (per (10)) as the FA functionality is combined at the GGSN along with the normal GGSN functionality as outlined in (4). In this case, mobility within the UMTS network will be handled with normal SGSN-GGSN procedures, whereas inter-technology handover with 802.11 networks will be handled with MIP procedures.

ACKNOWLEDGMENTS

The authors thank A. Gupta and Y. W. Lee for permission to use their drawings, and R. Atlas, D. Benenati, M. Buddhikot, G. Chandranmenon, S. Han, S. Miller, L. Salgarelli, and R. Shapira for their contributions to the paper.

REFERENCES

1. IEEE 1999, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band", IEEE Standard 802.11b-1999, Part 11

2. Buddhikot M, Chandranmenon G, Han S, Lee Y, Miller S, Salgarelli L, 2003, "Integration of 802.11 and Third-Generation Wireless Data Networks", IEEE Infocom

3. ETSI, 2002, "General Packet Radio Service (GPRS) Service Description, (Stage 2), (Release 99)", 3GPP TS 23.060 v3.14.0 (2002-12) 122 060

4. ETSI, 2003, “Interworking between the Public Land Mobile Networks (PLMN) Supporting Packet Based Services and Packet Data Networks (PDN) (Release 99)”, 3GPP TS 29.061, Version 5.5.0 v3.12.0, release 5

5. Perkins C, 2002, “IP Mobility Support for IPV4”, IETF RFC 3344

6. Rigney C, Willens S, Rubens A, Simpson W, 2000, “Remote Authentication Dial In Users Service (RADIUS)”, IETF RFC 2865

7. Calhoun P, Loughney J, Guttman E, Zorn G, Arkko J, 2002, “Diameter Base Protocol”, IETF Internet Draft, draft-ietf-aaa-diameter-17.txt

8. Benenati D, Feder P, Lee N, Martin-Leon S, Shapira R, 2002, “A Seamless Mobile VPN Data Solution for CDMA2000, UMTS, and WLAN Users”, BLTJ, 7 (2), 143-165

9. Institute of Electrical and Electronics Engineers, 2001, “IEEE Standards for Local and Metropolitan Area Networks: Port Based Network Access Control”, IEEE Standard 802.1x-2001

10. ETSI, 2000, “Combined GSM and Mobile IP Mobility Handling in UMTS IP CN”, TR 23.923, Version 3.0.0, release 3

11. Harkins D and Carrel D, 1998, “The Internet Key Exchange (IKE)”, IETF RFC 2409

12. Kent S and Atkinson R, 1998, “IP Authentication Header”, IETF RFC 2402

13. Kent S and Atkinson R, 1998, “IP Encapsulating Security Payload (ESP)”, IETF RFC 2406

14. Rivest R, 1992, “The MD5 Message-Digest Algorithm”, IETF RFC 1321

15. National Institute of Standards and Technology, 1995, “Secure Hash Standard”, Federal Information Processing Standard Publication 180-1

16. Huttunen A, Swander B, Stenberg M, Volpe V, and DiBurro L, 2003, “UDP Encapsulation of IPSec Packets,” IETF Internet Draft, draft-ietf-ipsec-udp-encaps-06.txt