Technical Study

Secure Mobile Architecture (SMA) Vision and Architecture

Technical Study

Secure Mobile Architecture (SMA) Vision and Architecture

Document Number: E041

Published by The Open Group, February 2004.

Comments relating to the material contained in this document may be submitted to:

The Open Group
Apex Plaza
Forbury Road
Reading
Berkshire, RG1 1AX
United Kingdom

or by electronic mail to:

ogspecs@opengroup.org

Contents

Preface

The Open Group

The Open Group is a vendor-neutral and technology-neutral consortium, whose vision of Boundaryless Information Flow will enable access to integrated information within and between enterprises based on open standards and global interoperability. The Open Group works with customers, suppliers, consortia, and other standards bodies. Its role is to capture, understand, and address current and emerging requirements, establish policies, and share best practices; to facilitate interoperability, develop consensus, and evolve and integrate specifications and Open Source technologies; to offer a comprehensive set of services to enhance the operational efficiency of consortia; and to operate the industry's premier certification service, including UNIX certification.

Further information on The Open Group is available at www.opengroup.org.

The Open Group has over 15 years' experience in developing and operating certification programs and has extensive experience developing and facilitating industry adoption of test suites used to validate conformance to an open standard or specification.

More information is available at www.opengroup.org/testing.

The Open Group publishes a wide range of technical documentation, the main part of which is focused on development of Technical and Product Standards and Guides, but which also includes white papers, technical studies, branding and testing documentation, and business titles. Full details and a catalog are available at www.opengroup.org/pubs.

As with all *live* documents, Technical Standards and Specifications require revision to align with new developments and associated international standards. To distinguish between revised specifications which are fully backwards-compatible and those which are not:

- A new *Version* indicates there is no change to the definitive information contained in the previous publication of that title, but additions/extensions are included. As such, it *replaces* the previous publication.
- A new *Issue* indicates there is substantive change to the definitive information contained in the previous publication of that title, and there may also be additions/extensions. As such, both previous and new documents are maintained as current publications.

Readers should note that updates – in the form of Corrigenda – may apply to any publication. This information is published at www.opengroup.org/corrigenda.

This Document

This document is the Technical Study for Secure Mobile Architecture (SMA) Vision and Architecture. It has been developed and approved by The Open Group. It describes an integration architecture that provides the framework and the building blocks for implementing a secure mobile environment. At the time of publication, some of the specifications that describe the framework are not yet approved standards, and some of the building blocks assume products that are not commercially available.

Intended Audience

This document is intended for any IT information worker who must consider how to deal with mobile workers or the integration of mobile devices into an enterprise or government network. These IT information workers include those involved in networking, security, applications, mobility, workflow, location, Voice Over IP (VOIP), and wireless.

Trademarks

Acknowledgements

The Open Group gratefully acknowledges the contribution of the following people in the development of this document:

- Bill Estrem, University of St. Thomas
- Peter George, Wheatstone Consulting Ltd.
- Chandra Olson, Lockheed-Martin
- Richard Paine, Boeing
- Emil Sturniolo, NetMotion Wireless
- Jim Wojnarowski, Motorola

Referenced Documents

The following documents are referenced in this Technical Study:

- draft-ietf-seamoby-mobility-terminology-03.txt, M. Kojo, J. Manner, published by the IETF, March 2003.
- Secure Mobile Architecture (SMA) Issues and Requirements, S. Coetzee, W. Estrem, P. George, C. Olson, R. Paine, E. Sturniolo, J. Wojnarowski, published by The Open Group, January 2003 [http://www.opengroup.org/tech/mmf/arch/protected/uploads/10_1049811149__Secure_Mobile_Architecture_Requirements_Issues_Version1.pdf].
- Session Management Issues and Requirements, The Open Group Mobile Management Forum (MMF), published by The Open Group, July 2002.

1 Introduction

1.1 Why Develop a Secure Mobile Architecture (SMA)?

The Internet has emerged as the world's primary communications infrastructure to support data applications and services, but it does not inherently support mobility. In contrast, mobile networks in the past decade were built primarily to deliver voice services with limited data capability. However, as users of the Internet have become increasingly more mobile, the need has arisen for a convergence to a mobile Internet environment that has the ability to support data, voice, and video anywhere, anytime, anyplace from a multitude of devices.

The WLAN technology and standards have revolutionized the access speeds for mobility. Mobile users are experiencing the same wireless and mobile speeds that they are familiar with in the office and therefore they want to do the same things mobile that they have been able to do in the office over a wire. Although, those users have become more mobile and demand mobility of their computing devices, those mobile devices are generally more resource constrained due to size and memory constraints of the smaller, handheld devices. The Secure Mobile Architecture (SMA) does not address the resource constraints of the mobile devices, but deals with the means by which they communicate. The protocols and underlying infrastructure to support mobility are the areas of emphasis of the SMA.

With the rise of Voice Over IP (VOIP) and the availability of WLANs, cellular data, and satellite data systems, comes the need to deal with rapid hand-offs and streaming data using Internet protocols. This add-on nature of security and mobility to the Internet has led to fragmentation of the services market through proprietary technology implementation.

Lack of standards has become the biggest obstacle to the further growth of the Internet and its use as a mobility communications infrastructure. Much as the cellular industry met the requirement to do fast hand-offs and shifting the stream from one provider to another, so must the Internet industry meet those same requirements to support VOIP. Early standardization efforts such as MobileIP (MIPv4 and MIPv6) efforts have not met with success. This may be attributed to inherent delays incurred by trying to solve mobility by using care-of-addresses and the long delay times associated with multiple transits across the Internet.

The Open Group Mobile Management Forum (MMF) addressed user mobility requirements using a client-server model to address roaming in the Session Management Issues and Requirements document published in July 2002. In this architecture, the server maintains state as the client roams, as depicted in Figure 1. This approach has been implemented in vendor products such as NetmotionWireless (architecturally depicted in Figure 1) and Brand Communications (Apollo).

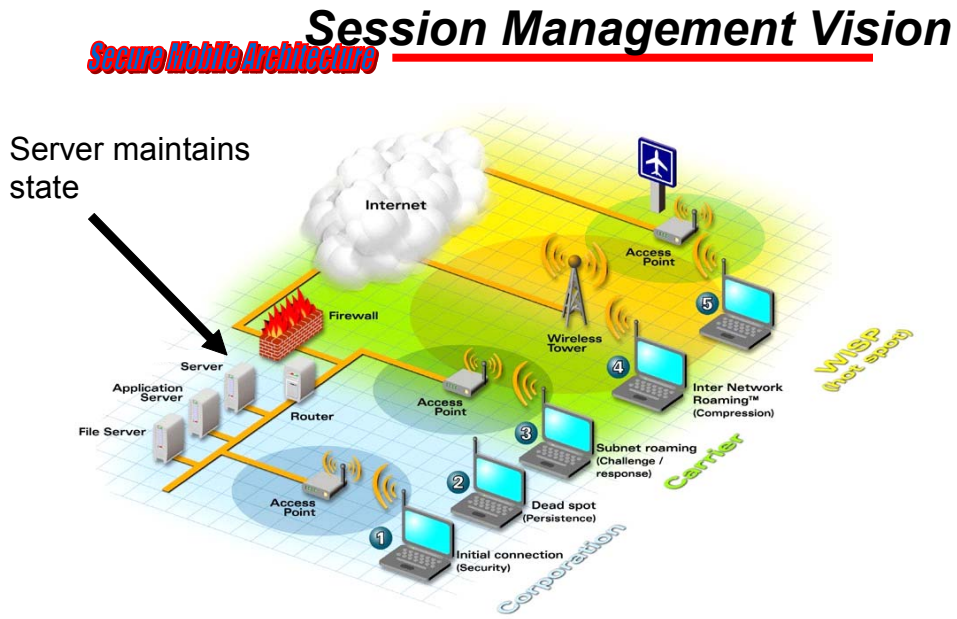

Figure 1: Session Management Vision (c. July 2002)

To take session management to the next step requires addressing mobility using a protocol-based architecture in which the protocols maintain the state. In Figure 2, the server has been removed and the mobile state is maintained by protocols and context transfers as the mobile end user moves. The SMA addresses this by providing a standards-based approach that could be used to create a secure mobile Internet environment.

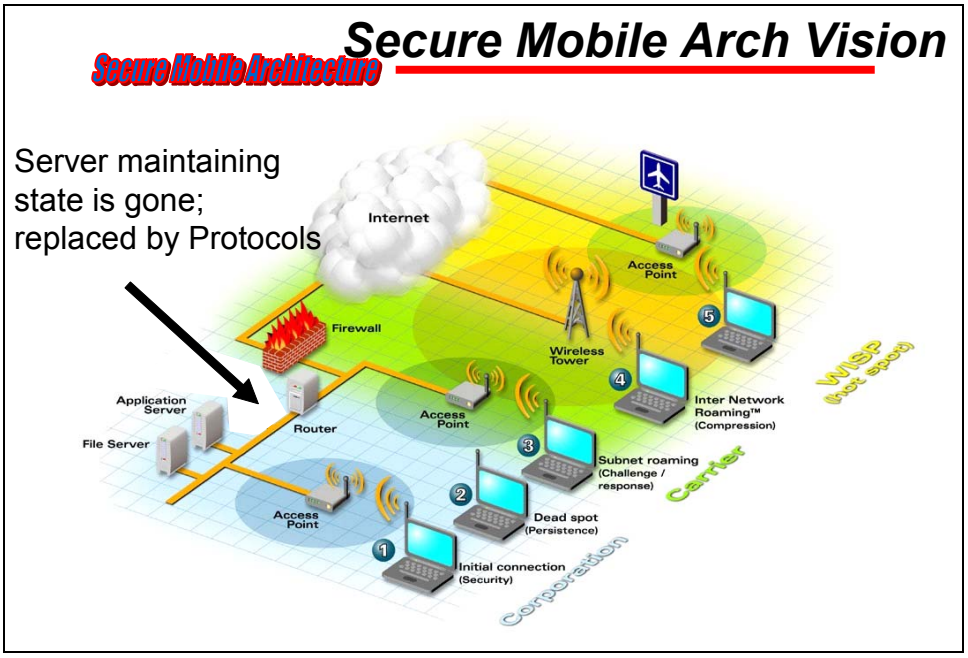

Figure 2: SMA Vision (2003)

In this document, the requirements and issues of an SMA are addressed first to build a foundation. The discussion then builds a mobile Internet architecture by addressing necessary components including the protocol stack, security, and roaming. The document brings all of the concepts together as it concludes with a recommendation on how to achieve the SMA vision.

1.2 SMA Issues and Requirements

The SMA Issues and Requirements document was published by The Open Group in January 2003 (see Referenced Documents). This document states the issues and the need for such an architecture.

1.3 Requirements-Based Principles for the SMA

Rather than state hard requirements for the SMA, there is a need to state the requirements in terms of principles. This allows the architecture, based on the SMA Issues and Requirements document, to have the flexibility to adjust for the future as well as the present, as depicted in Table 1.

Having a foundation of principles will give the SMA a longer-term perspective and persistence in the long run. These principles are outlined below.

Table 1: SMA Principles

Note: We are using the Open Systems Interconnection (OSI) layers to demonstrate where within the protocol stack each principle applies. We are, however, assuming an IP-only network model.

Description	Issue	Principles	OSI Layer Mapping
IP-only	Most data and voice are moving to the Internet	The primary principle of the SMA is that it be IP-only. This principle taken as a requirement allows for flexibility, meeting the needs of all future implementations of the IP infrastructure.	3
Session Security	Securing the session	The security of the mobile user should be based on the session of communication being used. This principle enables a concentration on the security needed for a unicast or multicast session to an individual or a group.	4 and 5

Description	Issue	Principles	OSI Layer Mapping
Standards-based Network Statistics	Differing standards are being employed by different networks	The information about security and communications should be based on standards. This principle enables the architecture to be pointed toward standards. The variation is thus decreased and made more effective. Vendors' equipment that will work together are pursued and deployed by enterprises.	1 thru 7
Session Initiation Protocol (SIP)	SIP is becoming the future technology of choice	The SIP is the foundation protocol for streaming media such as voice and video. This principle enables the architecture to be focused on the next generation of IP technology and standards.	5
Personal Firewall on Every Device	Mobile devices are vulnerable and exposed on the Internet	A personal firewall must be on every mobile device to protect the device from hackers on the Internet. The firewall can take the form of a firmware firewall or NAT and persist as protection over the lifetime of the device. This principle makes the hardware platform secure and minimally protected from the Internet.	2 thru 7
Host Identity Security (moving away from security based on MAC and IP address)	Internet bases security on IP addresses; vulnerable to man-in-the-middle	Host identity security will move us away from the Internet practice of basing security on MAC and IP addresses. This principle enables certificates and biometrics to accompany the communication of information across the mobile network.	3 and 4
Common Information Model/DEN-NG	Lack of information about systems (end systems and networks)	The information model underlying the architecture is standards-based and written in a modeling language like UML.	2 thru7
Policy Engine	Lack of ways to make decisions and enforce them	There will be a policy engine to run simple associations at the network level for affecting port assignment and route determination.	2 thru 7

Description	Issue	Principles	OSI Layer Mapping
Location Used for Zoning (Safety, Security, ITAR, Work-Flow, etc.)	Mobile devices can be anywhere and must be located (E911 and other req)	Location will be a key component of the wireless network and is used for a myriad of industrial uses, including E911, denying service for those not location authorized, and to improve productivity by delivering the correct tools and software based on the location of the individual or device.	2 thru 7
Mobile Context Transfer to enable smooth hand-offs	VOIP requires very fast hand-off as a prerequisite	Moving around (roaming) is only effective when hand-offs and transfers are fast enough for VOIP.	2 thru 4
Network Statistics are Standards-based	Network measurements are either not consistently represented or non-existent	Network statistics must be based on Internet standards. The standards are from the Internet Engineering Task Force (IETF) and IP-only. This principle ensures that the information that is available for enabling security and mobility has a common basis.	1 thru 7
End-to-End Security	Security is not guaranteed between ultimate peers	Security (Authentication, Authorization, and Encryption) must be guaranteed between the end points of communication.	3 thru 5

1.4 Basic Elements of an SMA

The basic elements of an SMA are that the focus of security be moved away from address-based security to a host identity scheme and that mobile devices be enabled to move between disparate network mediums. In particular, the host identity must be used to identify each IP packet generated between the ends of the security association, and that roaming be normalized between LAN and WAN systems.

1.4.1 Security Based on the Host Identity

Moving the security from address-based security to host identity security moves away from a world of man-in-the-middle worries and spoofing of addresses. In the Internet world today, any identity-grabbing teenager or hacker can assume the identity of an individual Internet stream of consciousness because the addresses are the basis of the security. By moving the fundamental underlying security basis to host identity, the fundamental security falls back to the identity of an individual component, which cannot be spoofed or attacked by a man-in-the-middle.

1.4.2 Mobility Across the WPAN, WLAN, Cellular Data, and Satellite Data Environments

The ability to be mobile implies that at some point in time a device can be detached from its initial point of contact with the network and then reattached without disruption of application-level sessions. Thus, the fundamental requirement of an SMA is the ability to transparently (not noticed by the user) migrate across disparate network technologies seamlessly and while maintaining the communication session(s) and security state. The speeds and capability of network access may change across the boundaries, but the seamless movement and transition of sessions is maintained. Current analysis has shown that the harshest current day requirement is maintaining a VOIP session. If a solution can secure and functionally not interfere with an active VOIP session between peers while transitioning network access between cellular data and WLANs, then such an architecture has met the criteria for a framework for the future.

2 Architectural Implications of the SMA Protocol Stack

The Secure Mobile Architecture (SMA) is an integration architecture providing the framework and the building blocks for implementing a secure mobile environment. As an integration architecture, this approach enables individual designs and implementations that address specific design and implementation requirements. For example, the issues of how to provide a host identity protocol and its attendant private keys is a design requirement – you may choose to design in user and machine certificates to provide the host identity. The SMA is meant to be used as a integration framework to build upon. The major elements of the architecture (information model for both stable and real-time data, a policy engine, a role-based access control, location-enabled zones, a Context Transfer Protocol (CTP), a host identity protocol, and transport layer encryption) provide that high-level framework.

The SMA protocol stack is based on IP. However, we will be discussing the impacts of the protocol in terms of the OSI stack. There is a need to discuss session, presentation, and application layers to complete the discussions about the elements of an SMA.

2.1 Generic OSI Stack

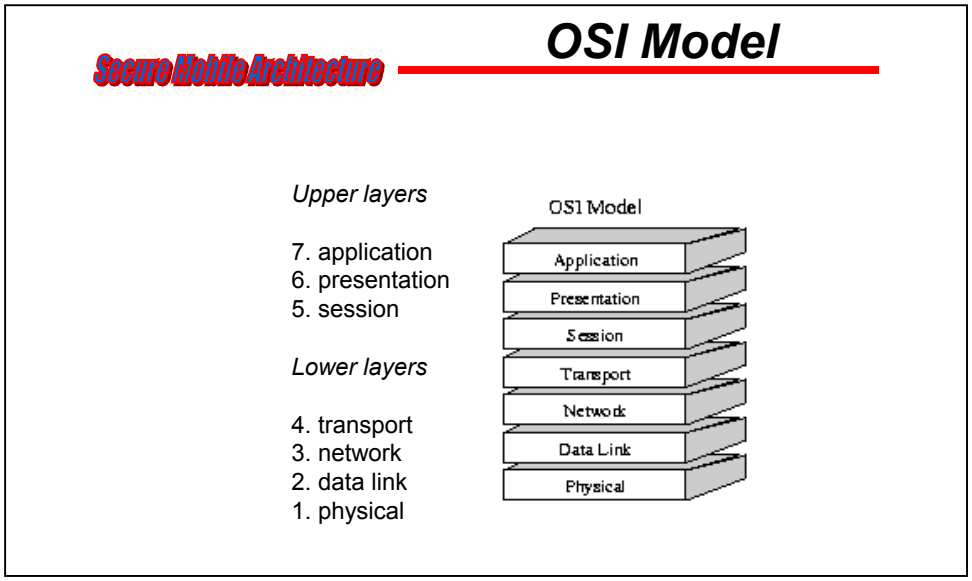

Figure 3: OSI Model Network Stack

2.1.1 Differences with the TCP/IP Stack

The TCP/IP stack places session, presentation, and application layers in the "Applications" layer. In other words, TCP/IP does make inferences about how the protocol stack handles sessions, how it formats the information, or how its applications use the information.

2.2 TCP/IP Stack

The following picture of the TCP/IP stack shows this placement together of the upper layers of the OSI model.

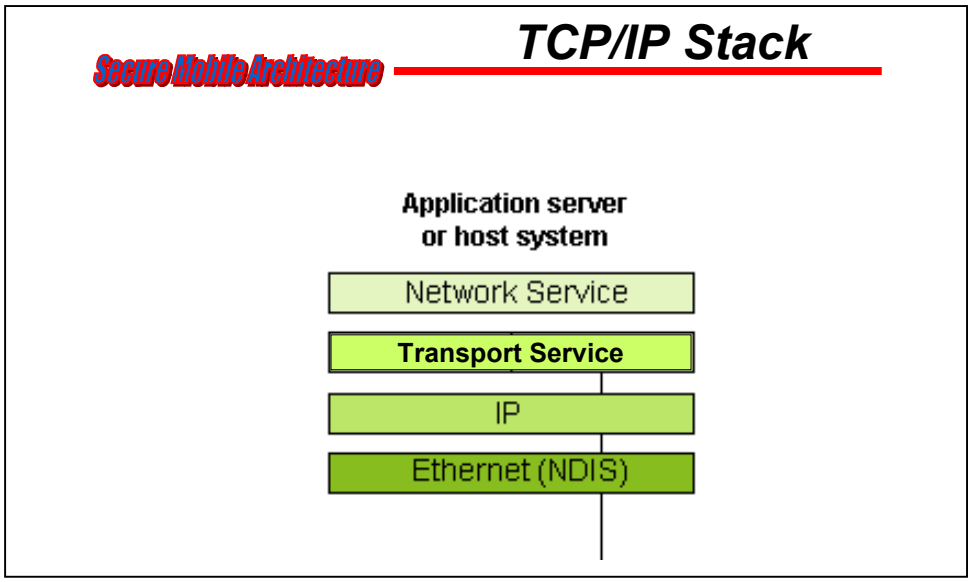

Figure 4: TCP/IP Protocol Stack

2.3 Implications of Using IPv4 versus IPv6

When Version 4 of the Internet Protocol (IP) was initially designed, mobility of a device was not even a twinkle in the inventor's eyes. No one at that point envisioned strapping a room-sized computer on someone's back and moving it from place to place while staying connected to the network. Each system was stationary, most likely bolted to the floor, and was assigned a unique identifier or address from a global authority known as IANA. Once the device was registered, this address allowed interconnected peers to correctly route information between the two across the Internet.

During the early 1990s, the Internet "came of age"; no longer the purview of just research and academic institutions, but corporations, business, and regular consumers. Cost to access the Internet also significantly decreased in this timeframe from thousands or hundreds to just tens of dollars a month. This phenomenon, combined with the rise in popularity of microcomputers in business and home settings, put demands on the assignment of addresses. It was clear that if

something didn't change, there would be a shortage of addresses. This problem was further exacerbated by the fact that current research work in the area of mobility might require not one but two distinct addresses when a device became mobile.

To solve this problem, engineers create a two-pronged approach:

1. To extend the life of the existing addressing structure
2. To create a new addressing structure so the community would not have to confront this same issue for the foreseeable future

2.3.1 IPv4

The heart of the addressing problem surrounding the IPv4 protocol is that it is limited by a 32-bit field that contains both its host and network identity fields. Although it is theoretically possible to generate over four billion addresses in this space, due to the hierarchical nature of the addressing scheme used for the efficient routing of frames, this actual yield of potential address is significantly reduced. To resolve the impending depletion, Version 4 standardized on some unique solutions to extend the life of the available address space while the world waited for the next generation addressing scheme to be completed (IPv6 – see below).

One of the most popular methods employed today uses an address reuse scheme. More commonly know as Network Address Translators (NATs), this functionality allows for more then one device to gain access to the Internet, while only consuming one globally referencable address. This, however, causes other problems, as each node connected to the network is no longer globally identifiable. Routing therefore can become somewhat of a challenge.

Another major advancement in extending the life of Version 4 addressing was the dynamic distribution of addresses using a protocol known as the Dynamic Host Configuration Protocol (DHCP). Through the use DHCP, a specific device "leases" and IP address for a specific period of time. If these two methods are combined, the fact that addresses are no longer unique, and they may be transiently assigned (both of which are the norm today), keeping track of what address relates to what device at any one point it time can be significantly difficult.

Considering that an address isn't permanent, and may not be globally unique, presents many challenges for a secure mobility solution. For example, most currently standard security measures use the IP address as part of the material used for identifying an associated security context. If the IP address changes for whatever reason, the security association is invalidated.

2.3.2 IPv6

To solve the issues surrounding the scarcity of globally unique addressing, the IPv6 standard has expanded the address space from only 32 to 128 bits (more then 300 addresses for every square centimeter on the earth). However, addresses can still be assigned dynamically and security standards continue to reference them to help identify a security association.

2.3.3 MIPv4 versus MIPv6

To address issues surrounding the mobility of devices connected to the Internet,, the Internet Engineering Task Force (IETF) has designed two standards. MIPv4 is an overlay to the existing standard, and MIPv6 was designed in from the beginning. Each, however, only deals with the

correct Layer 3 routing of frames between communicating peers, leaving other issues of security and persistence to the higher layers.

2.3.4 Alternatives to MIPv4 and MIPv6 – Dynamic DNS (DDNS)

Alternatives include the previously defined Open Group Session Management Architecture, or the use of Dynamic DNS (DDNS) to maintain an updated list of the IP address associations. A paper by Pappas, et al[1] presents the viability of such an approach and the desirability of this approach over the MIPvX approach.

2.4 Implications of IP-Only

In an IP-only environment, the addressing, routing, and connectivity imply that these packets can pass over the Internet transparently. They also imply that the transport is constrained by the rules of the IETF and the TCP/IP RFCs.

2.4.1 Implications of Streaming over IP

Streaming of secure voice and video may be the major requirement of the SMA as many new services are being designed with multimedia content in mind. However, some of the real-time requirements of applications such as Voice Over IP (VOIP) place significant demands on the network fabric, especially when transitioning between networks seamlessly.

2.4.2 Signaling over IP

Signaling in streamed and switched voice environments has typically been managed by Signaling System 7, an ITU-developed standard for switched systems. Signaling on the Internet is a relatively new concept and generally has been accommodated by the Session Initiation Protocol (SIP). Basically, signaling on the Internet is a somewhat foreign concept to the TCP/IP stack and therefore is the subject of much discussion.

2.4.3 ICMP

The purpose of these control messages is to provide feedback about problems in the communication environment, not to make IP reliable. There are still no guarantees that a datagram will be delivered or a control message will be returned.

2.4.4 Messaging over IP

Messaging over IP is only critical when there is a real-time and streaming source of the messaging. The nature of the Internet is such that bursty traffic is the norm and streams of real-time traffic, like voice, can only be accommodated when the bursty nature does not interfere with the requirement of interpacket latency of less than 20 msec.

1 Available at: http://www.ee.ucl.ac.uk/lcs/papers2002/LCS072.pdf.

2.4.5 Anything that Runs over UDP

UDP is the protocol that does not require a guaranteed receipt at the other end. VOIP and streaming video can tolerate the loss of several packets without a noticeable interruption in the stream because of the nature of human-perceived (human brain) reception. UDP is the protocol that is used for VOIP and streaming video in order to speed up the process and avoid delays in the checking of every packet.

2.4.6 SIP Definition

The SIP is a session protocol for Internet conferencing, telephony, presence, events notification, and instant messaging. SIP was developed within the IETF MMUSIC (Multiparty Multimedia Session Control) Working Group, with work proceeding since September 1999 in the IETF SIP Working Group.

2.4.7 Implications of SIP in the SMA

The implications of SIP in the SMA are that an architecture with a fundamental element being SIP would function somewhat differently than one that did not. For example, with the SMA using a Host Identity Payload (HIP), a session would have a verifiable identity associated with it and therefore be less susceptible to man-in-the-middle and spoofing attacks. These implications make for an Internet environment that is secure and that can accommodate moving around throughout the Internet seamlessly and providing services like data, voice, and video.

2.5 Session Management Architecture Stack

The Open Group developed a Session Management Issues and Requirements document and presented a session management architecture. The architecture is a client-server model allowing for the server to maintain state of the communication. In this architecture, the device can move out of range of the radio and be disconnected, but it can come back to exactly where it left off when it regains connection.

Figure 5 shows an example protocol stack associated with one of The Open Group session management-compliant products.

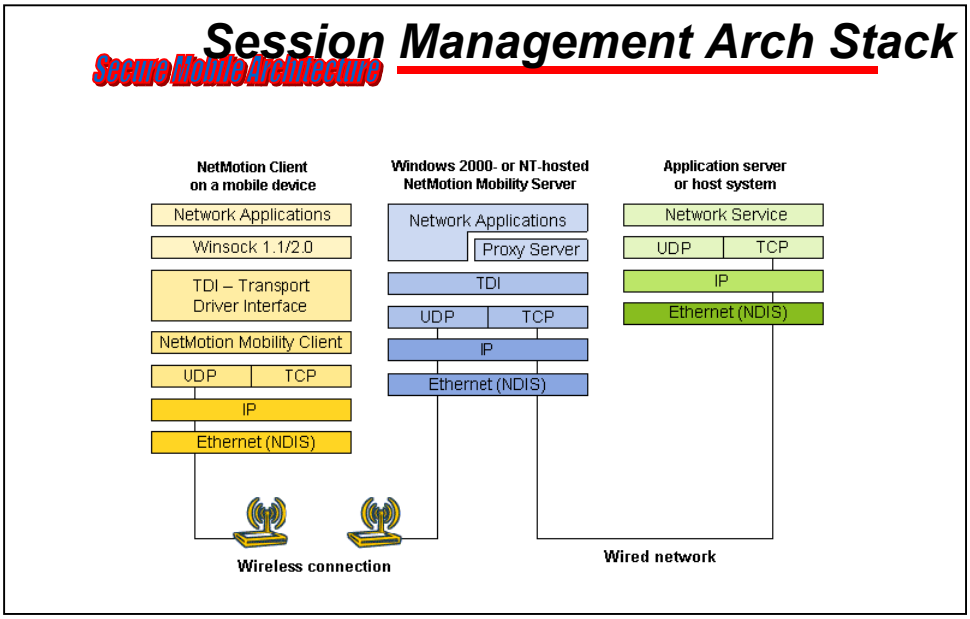

Figure 5: Example Session Management Architecture Protocol Stack

2.6 IP-Only

The SMA addresses IP only. Previous work in The Open Group – called Session Management Vision and Architecture – addressed multiple protocols and their appropriate stacks. We believe that IP is the only protocol stack of consequence in the future. The modifications of TCP/IP will be through a standards process (IETF) that has proven to be successful and world-wide in scope and therefore worthy of dedication of the SMA architectural model. This IP-only limitation inserts some consequences to the SMA framework.

The consequences of IP-only are the following:

- The TCP/IP stack is limited to Layers 1 to 4 of the OSI model
- Two different address spaces, IPv4 and IPv6
- Two mobile approaches, MobileIPv4 (MIPv4) and MobileIPv6 (MIPv6)
- Use of Domain Name Service (DNS) and Dynamic DNS (DDNS)
- Use of UDP for streams
- Use of signaling over IP
- Use of the SIP
- Adherence to IETF and IRTF protocols

3 Security

3.1 Security Concepts

The security concepts for the Secure Mobile Architecture (SMA) include the Host Identity Payload (HIP) and the use of secure Context Transfer Protocols (CTPs)[2] to move security and security context across network boundaries to enable seamlessness. These concepts, then, enable secure transition while mobile without interruption, including for voice and video. In the four levels of wireless, which are:

1. Personal Area Networks (PAN)
2. Local Area Networks (LAN)
3. Metropolitan Area Networks (MAN)
4. Wide Area Networks (Satellite-based WAN)

the one common denominator in the principles stated at the start of this document, is that a fundamental principle is the SMA address IP-only. So, the fundamental environment is that of IP-only, which stretches across the four levels of wireless. In addition to being IP-only, the secure mobile security happens at different levels of the protocol stack. Figure 6 reflects the security that must be taken into consideration as the end-user device moves around and adjusted as the movement occurs.

[2] From IETF's Seamoby CTP Internet Draft:
"This document presents a context transfer protocol that enablesauthorized context transfers. Context transfers allow better support for node-based mobility so that the applications running on MNs can operate with minimal disruption. Key objectives are to reduce latency, packet losses, and avoid re-initiation of signaling to and from the MN."

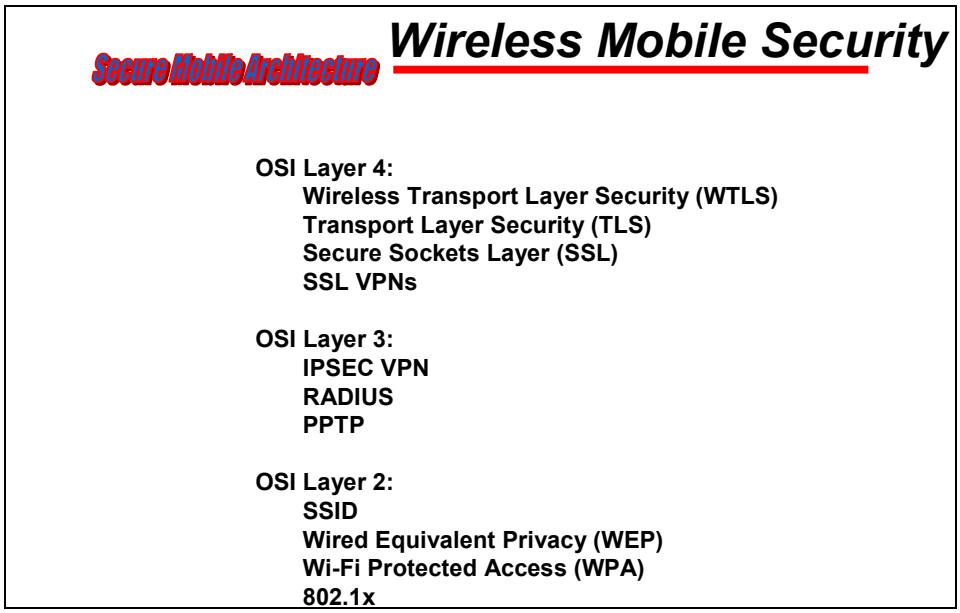

Figure 6: Varieties of Security at the Network Stack Levels

3.2 Internet Security Concepts

3.2.1 AES

The Advanced Encryption Standard (AES) is the US government-approved encryption standard. It is in the process of replacing triple DES as being the most used encryption standard in the world and is being built into the WLAN standards and most of the security standards in the world. The implications are that AES will be the encryption method most widely pursued in the timeframe that the SMA will be functioning.

3.2.2 802.1x

802.1x is the IEEE standard for authenticating to specific ports on networking equipment. 802.1x, as first proposed, attempted to provide a practical solution to a simple scenario. Ports allowing access to a LAN were to be placed in a semi-public space – e.g., a conference room used by visitors to a company – and access to the LAN would be controlled by a dedicated bridge port taking instructions from an authentication server for the company. From the point of view of getting a result – that is, producing a working specification – in a finite time this scenario is effective as it admits very few threats:

1. The human (let's call him or her 'H') who is attempting to connect to the LAN with a laptop (the supplicant) is presumed (if authenticated and authorized) to be well intentioned, moderately competent, and careful of the LAN and attached resources.

2. H can inspect the laptop, provide the physical cable from the laptop to the wall jack, and verify by inspection that no intruder is attached (there is no shared hub half way down the cable).

3. The wall jack is firmly attached to the wall, and intruders are not believed to have sufficient uninterrupted access to complete building works to insert equipment behind the jack.

4. The physical wiring behind the wall jack is believed to be wired into the bridge port controlling access with semi-permanent point-to-point wiring using the cable routing commonly used in office walls and cube layouts, so is not readily susceptible to accidental network additions by the company's own non-IT staff.

5. The company's own staff, who have direct physical access to the secured LAN, are presumed not to be directly interested in helping intruders access resources. To put it another way – if they are of mischievous intent, the company has much more to worry about than controlling open access wall jacks in meeting rooms, since they can physically access all sorts of equipment in the network directly.

The implications for the SMA are that 802.1x will be the predominant security authentication methodology when the SMA is being practiced, especially with the predominance of IEEE 802.11i, which is based primarily on 802.1x. The WLANs have proven and will continue to be the predominant communications technology for access to the Internet over the next twenty or so years. Therefore, the mechanism for authentication is going to be based on 802.1x, and other IP-based technologies, such as cellular data, may also adopt 802.1x as the predominant authentication mechanism for Internet IP-only access.

3.2.3 HIP

In addition to some port security is the concept that the host at the initiating end can prove who they say they are without some convoluted ticketing or certificate-based authentication process. The HIP[3] offers such a mechanism. HIP is the creation of Robert Moskowitz and answers many of the security ills of the Internet. HIP is the equivalent of packets carrying your driver's license in every packet. Such a mechanism, in combination with port security, makes the identity of an individual, as instantiated in the network, verifiable and therefore resistant to man-in-the-middle attacks that are so inherent in the Internet architecture.

The implications for the SMA are that, if the Internet is ever going to be secure in a less complicated way, some kind of host identity mechanism will be required to be implemented Internet-wide. That means such a security mechanism is imperative to a really secure and manageable mobile technology.

3.2.4 IPSEC

IPSEC may also play a role in the security of the SMA, but only when tunneling is possible and viable, which it may not be. The Internet experience of being mobile and needing to tunnel with IPSEC has been less than satisfactory. In the case of the Layer 3 tunneling, IPSEC will not go through most gateways or firewalls because of the inability to read and interpret the headers. What has been more successful is using the Secure Sockets Layer (SSL); mechanisms which do not count on Layer 3 to provide the addressing. So, IPSEC can only be used in controlled environments, such as a corporate setting, that enable Layer 3 tunneling. In the SMA, HIP and

[3] The HIP documentation are available at http://homebase.htt-consult.com/HIP.html.

SSL encryption can enable roaming through NATs and simply replace much of the functionality of IPSEC without loss of header information.

3.2.5 WPAN Security

Bluetooth (or IEEE 802.15.1) is the primary example of WPAN security and is not adequate for corporate security infrastructures. The need for a verifiable end user will lead enterprise users to do the same thing they have done with WLAN security; require strong authentication and tunneling for every wireless connection. There is a need to address security in the IEEE 802.15.3 Wireless PAN (WPAN) standard as devices in a WPAN have no assurances about whom they are communicating with and that their messages have not been eavesdropped upon, or altered, undetected.

An IEEE WPAN security framework proposal was presented in February 2002 in the 802.15.3 MAC to promote further discussion of security models and cipher suites. The paper suggests enabling user established trust in WPANs and that authentication not require the use of digital certificates or a certificate authority.

In October 2002 a Security and Security Architectural Recommendations document was introduced in the 802.15.3 High Rate WPAN Working Group. This document presented security modes for authentication of devices, privacy protection of message traffic between devices, and verification of the public key.

3.2.6 WLAN Security

WLAN security has been the weak link in the movement to wireless. The initial WEP encryption standard implemented in the 1997 802.11 standard made the standard vulnerable to interception. Airsnort was the precipitator of the move to make WLAN security less vulnerable, because it was released as Open Source software to break the WEP key. At that point, all IEEE 802.11 WLANs were vulnerable to determining the WEP key. IEEE 802.11i is the task group within IEEE addressing WLAN security. The Wi-Fi Alliance, which is a WLAN industry group providing interoperability standards for WLANs, has taken the work of 802.11i and specified a minimum standard called Wi-Fi Protected Access (WPA) that is the minimum standard for providing WLAN security without requiring strong authentication and an encrypted tunnel. IEEE 802.11i has two stages, one called Temporal Key Integrity Protocol (TKIP) that WPA is based on, and another using the AES standard. Both require an authentication to the Access Point (AP). These wireless security methods being developed are far superior to any of the previous wireless security proposals, including Bluetooth, any of the cellular methods, or WAP.

3.2.7 WPA

Wi-Fi Protected Access (WPA) is a specification of standards-based, interoperable security enhancements that strongly increase the level of data protection and access control for existing and future WLAN systems. It is designed to run on existing hardware as a software upgrade. WPA is derived from and will be forward-compatible with the upcoming IEEE 802.11i standard. WPA was constructed to provide improved data encryption through Temporal Key Integrity Protocol (TKIP), which was weak in Wired Equivalent Privacy (WEP), and to provide enterprise-level user authentication via 802.1x and the Extensible Authentication Protocol (EAP), which was largely missing in WEP. It should be used in conjunction with an

authentication server such as Remote Authentication Dial-In User Service (RADIUS) to provide centralized access control and management.

3.2.8 Cellular Data Security

Cellular data (IP-only) security has been based on proprietary security methods and may therefore be less vulnerable, but restrictive due to their proprietary nature and they are indeed breakable.

The implications for the SMA are that WLANs will be the primary means of Internet access, and cellular data will be used secondarily if WLANs are not available. The implications are that a secure seamless hand-off between cellular data and the WLAN technology will be a requirement and an immediate need.

3.2.9 Satellite Data Security

The approach to satellite data has been much like the cellular data example and is quite breakable. The implications for the SMA are that WLANs will be the primary means of Internet access and satellite data systems will be used secondarily or in a tertiary capacity if WLANs are not available.

3.3 Implications of Session Security

Session end-to-end tunneling security is the only secure means of connectivity today. In addition, the only way to securely move between communications infrastructures is via session management. The Open Group Session Management Architecture is the standard for this kind of secure seamless mobility and is epitomized by NetMotion Wireless and Brand Communications. There are lots of problems with session security when using IPSEC because the header will not go through many of the gateways. The Open Group Session Management Architecture is a client-server model and therefore the VPN model and state is carried in the server.

The Secure Mobile Architecture (SMA) is the next step in the progression of a stateful seamless mobility. It uses protocols to carry state. This stateful communication is enabled by Host Identity, Context Transfers, and Policy-Based Networking to accomplish its secure seamless nature.

3.3.1 Transport Layer Security (TLS)

TLS is another approach to security for the Transport Layer. This approach moves the security mechanisms up to and including the transport layer and gets around the issue of encrypting the IP header. Wireless TLS (WTLS) and Mutual TLS (MTLS) are variations of the TLS that accommodate fast roaming and other specialized treatment of WLAN security.

3.3.2 Wireless TLS (WTLS)

WTLS is a wireless adaptation of the TLS protocol. At a first glance, TLS and WTLS present very similar services, such as strong encryption, authentication, signing, and hashing. Like TLS, WTLS can be used to enforce strong end-to-end security on an application-to-application basis, which means that encryption is maintained past any corporate firewall or gateway all the way to

the transport endpoint if needed. If end-to-end security is not required, WTLS encryption can be terminated in a border gateway like traditional VPN solutions.

However, there is one important difference between the protocols. Secure session establishment is a process that requires heavy computations and quite a few message exchanges between the two communicating peers. If the communication device has limited processing power, session establishment can take a significant amount of time to perform. As connection instabilities are quite common in wireless communication, it is very important that the secure session is able to survive roaming and periods of network inaccessibility. To handle these issues, WTLS supports a fast re-establishment method for sessions that have been disrupted. The session resume functionality allows for long-lived WTLS sessions that are able to survive network disconnects. If connectivity is lost, a new session can be re-established without significant overhead, and at a fraction of the original setup time. TLS on the other hand does not contain this resume functionality, and a full-scale handshake including key exchange mechanisms and capability negotiations must be performed each time the peers reconnect.

3.3.3 SSL

The SSL is yet another approach above the transport and above the IP layer. This approach enables a tunnel that is not affected by the IP layer and TLS.

3.3.4 True End-to-End Security (Not VPN)

True end-to-end security is that which does not necessarily require a tunnel. Some of the difficulties are when the communications are one-to-many or many-to-many. In these cases, a multi-conferencing unit (MCU) is required to make the communications work. More appropriate, perhaps, is a method by which the host identity, authentication, encryption, tunneling, etc. take place through the use of general-purpose protocols rather than client-server mechanisms.

3.4 Personal Firewalls and their Implications

In the environment where there is native secure mobile communications, the requirement is that each end-user device must have a personal firewall in order to protect the end device from hacking from the Internet. This is one of the fundamental requirements and constructs of the SMA.

3.5 Network Statistics

Also necessary in the SMA are network statistics that give information about the wireless environment and who is accessing whom.

3.5.1 WLAN – Radio Resource Measurement 802.11k

The IEEE 802.11k is the Task Group in 802 investigating the WLAN measurements. The measurements are required for making information about the WLAN environment available to the layers above Layer 2 and the applications needing that kind of information to make decisions about moving and pre-authentication to speed up fast roaming.

3.5.2 WPAN

802.15.1 (Bluetooth), 802.15.3 (High Rate), 802.15.3a (Ultra-Wide Band)

PANs will need to pursue the same kind of measurements that 802.11 has been pursuing in order to provide statistics and information to higher layers about what is being used in the frequencies that 802.15 operates. These frequencies, so far, are 2.4GHz ISM band and the 3.1GHz to 10.6GHz band for UWB (in the noise floor). Such a mechanism may be addressed in the move to do an 802 wireless family (802.11, 802.15, 802.16, 802.20) hand-off mechanism.

3.5.3 Cellular Data

The cellular data providers are the source of 56-200Kbps data communications switched connections. These providers do not address the handing off to other providers, but want to enable Internet access via the existing cellular systems.

3.5.4 GPRS

The interim GSM cellular data network is called the General Packet Radio Service (GPRS). This service delivers 30Kbps connections to the cellular system. GPRS is available in Europe and the US in locations where there are GSM services.

3.5.5 CDMA

Code Division Multiple Access (CDMA) is the technology underlying about a third of the cellular companies in the US (Verizon and Sprint PCS). These cellular companies have based their cellular data offerings on a fundamentally more inherent IP infrastructure and are capable of delivering Voice Over IP (VOIP) over the CDMA data network.

3.5.6 PCCA Standard 201

The Portable Computer and Communications Association has worked with many vendors over the years to promote a common object model for managing wireless devices at all four levels of the SMA. One of these efforts resulted in the specification and ultimate adoption of generic and network-specific management objects that were adopted by Microsoft for their proprietary Network Driver Interface Specification (NDIS).

3.5.7 Satellite Data

Satellite data is inherently affected by the fact that the signal must travel 22,000 nautical miles into space to use the geosynchronous satellite for communications. Such real-time offerings as video and voice are affected by this distance and the delays incurred by the transit time to and from the satellite.

3.5.8 QoS

The QoS pursuit is somewhat misguided from the standpoint that unless it is all QoS-enabled, none of it is QoS-enabled. There is information that can help to determine whether QoS can be pursued, however, and that is through Data Rate and other Layer 2 information about what kinds of bandwidth is available to permit QoS to happen.

3.6 Host Identity

Security is now quite insecure. The Internet is based on MAC and IP addresses that can easily be spoofed. Many methods have been developed to provide security based on addresses. This address-based security has led to the need for tunneling and ticketing to prevent man-in-the-middle attacks. Elaborate means have been developed to get around this vulnerability. The elaborate means have, perhaps needlessly, complicated and made complex the entire Internet world of security. The simple answer is to change the basic nature of the communications. Instead of basing the security on the address, base the security on the identity of the host. Use the method of security used for thousands of years; personal or enterprise identity. Much like the use of a national identity card, or drivers' license, or certificate, or social security number; embed the host identity with each packet.

4 Roaming

4.1 Implications of Roaming

Roaming is the ability to move into different macro-mobility cells and still have connectivity. In the cellular world, the roaming means that a cell phone user can move from cell to cell and maintain a conversation even if the provider is in a different company. In the data world, roaming should mean the same thing.

4.1.1 Internet Service Providers (ISPs)

Wireless IPSs (WISPs) provide Internet access via wireless means. These wireless means include cellular data, WLANs, and Satellite service providers. There are several organizations offering to do wireless roaming for WISPs and deliver billing methods to enable this roaming between WISPs. Examples include PassOne and Boingo.

4.2 Context Transfer Protocol (CTP)

In order to make seamless transfers happen without interruption, there is a requirement to pass context, first of all between micro-mobility vendors, like WLANs, and eventually between WISPs that cross macro-mobility boundaries, like between cellular data and WLAN WISPs. The context carried includes security, frequency, bandwidth, etc. in order to move an application seamlessly between providers.

4.2.1 Examples

CTPs enable specific information to be passed that carry the state or context of a condition between communicating entities on a network. There are several examples of CTPs that enable worthwhile communication between two entities; Seamoby is an IETF effort which is short for "seamless mobility"; IEEE 802.11f is a recommended practice for Inter Access Point Protocol (IAPP); and there are the examples of Layer 2 routing in which hand-offs can occur much faster.

4.2.2 Seamoby

The Seamoby CTP is a Layer 3 CTP. It is not entirely dependent on Layer 3, so it has been called a Layer 2.5 protocol. It is defined in the Internet draft CTP from the Seamoby Working Group and uses Layer 2 "triggers" to start the context transfers.

4.2.3 802.11f

The 802.11f IAPP is a CTP that is specifically for communicating stateful information between Access Points (APs) in a WLAN. There are proprietary means for doing that today, but 802.11f answers the question of how you perform inter-AP communications using standards.

4.2.4 Voice Over IP (VOIP) Issues

There are Voice Over IP (VOIP) issues and CTPs. These are specifically related to how quickly a protocol and a device can perform the transfer of QoS state and security state to enable a break-less telephony call across APs or across networks. The actual break time has been published to be 20ms. Much work can be done in the background using CTPs without dipping into this 20ms. For example, the pre-authentication and pre-authorization can be done using secure CTPs before the actual hand-off is required (802.11f is such a secure CTP).

4.2.5 Roaming via CTPs

As in the VOIP discussion in the previous paragraph, roaming has inherent requirements that include the need to pass state to the next entity without eating into the routing or hand-off issues. CTPs allow for this to occur. Instead of the state being held by a server, it is now passed using a CTP that enables stateful transfer of critical trusted information between entities on the network.

5 Secure Mobile Architecture (SMA) Vision

In the previous work of the Mobile Management Forum, the Session Management Architecture uses a client-server architecture to carry state between networks.

In the Secure Mobile Architecture (SMA) vision, the state is not carried by the server in a client-server architecture, but is carried by protocols between the elements of the network and across networks.

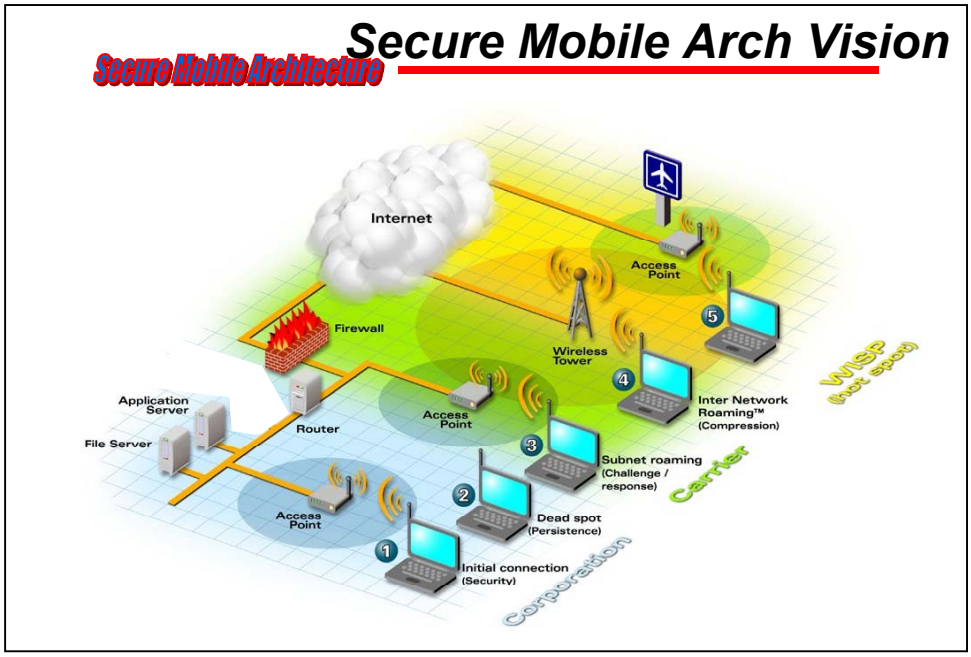

Figure 7: Secure Mobile Architecture (SMA) Vision

5.1 Stateful Protocols to Pass State Across/Between Networks

There are particular elements that have state and are of critical importance to the wireless and mobile network. These include QoS and security specifically and could, less critically, include Access Point (AP) state and STA (end device radio) state.

5.1.1 IEEE 802.11e QoS State

The QoS state includes the reservation requirements of a particular application asking for inclusion in the QoS network. This state might include the available bandwidth over the air, in the case of the wireless network, and the available bandwidth end-to-end over a wire in the case

of the wired network. There are, of course, other important parameters such as jitter and performance. However, of most importance is how much bandwidth the application is going to take and does the network being moved have that much bandwidth to offer?

5.1.2 IEEE 802.11f AP Stateful Protocols

In the case of the IEEE 802.11f Inter-Access Point Protocol (IAPP), the protocol does not define the context blob. In fact, it is left to the imagination of the developer what state is transferred between the APs. In the case of VOIP, there needs to be pre-authentication and pre-authorization information that must be passed in order to securely transfer information about a hand-off (including the IP address, the MAC address, the state of the radio environment, and an estimate of when it will be moving).

5.1.3 IEEE 802.11i Security State

IEEE 802.11i is the task group within 802.11 that is specifying the security standards for WLANs. The security state defined by this group is the state of the authentication and authorization within the Basic Service Set and the Extended Service Set of 802.11. Each AP has the responsibility, under 802.11i, to manage the security of itself and pass the security information to other members of its BSS and ESS. The security state is maintained across multiple APs through the use of its own security context transfer.

5.1.4 IEEE 802.11k Measurement State

The state of the radio environment in a WLAN is defined by 802.11k. This task group is defining how the WLANs obtain information about their radio state and how they are able to communicate (passively listen and actively seek measurements) this state. The goal is to have each radio be able to understand its radio environment and pass that information to higher layers so applications can respond to it.

5.2 Protocols to Carry State Across/Between Networks

There are two primary efforts to carry state across and between networks: the IEEE 802.11f IAPP and the IETF's Seamless Mobility (Seamoby) Working Group's Context Transfer Protocol (CTP).

5.2.1 IEEE 802.11f

IEEE 802.11f is a secure method of handing information to another AP. The information element is a context blob that has no definition within the packet, but is defined by the protocols that use it to pass context between APs.

5.2.2 Seamoby CTP

The Seamoby CTP is a context transfer protocol that enables authorized context transfers based on Layer 2 triggers Context transfers allow better support for node-based mobility so that the applications running on Mobile Nodes (MNs) can operate with minimal disruption. Key objectives are to reducing latency, packet losses, and avoid re-initiation of signaling to and from the MN.

5.3 Security Based on Host Identity (Three Encapsulations of Data)

5.3.1 Host Identity Payload (HIP)

The HIP proposal suggests that a new cryptographically-based name space may solve problems in today's Internet, including routing table growth due to site multi-homing, lightweight IPsec key establishment, and mobility management across multiple IP addressing realms. The fundamental idea is to assign a (statistically) globally unique name for any host with an IP stack. By making this name cryptographically-based (a public key), this host identity can be used to authenticate transactions. A HIP protocol layer is effectively inserted between the IP and transport layers, allowing for decoupling of transport connections from IP addresses, and all packets carry a representation of the host identity, either implicitly or explicitly. The host identity could be stored in DNS or in a public-key infrastructure (PKI), or it could be anonymous, in which case it still can be used to prevent connection hijacking.

5.3.2 IPSEC

The IPSEC protocols enable the end-to-end encryption of a packet exchange. This is done through a Layer 3 encrypted tunnel. There are tunnels and encryption methods at other levels of the protocol stack, but IPSEC is the basic, and any method of secure mobile communications must be able to use an IPSEC tunnel.

5.3.3 CTP Authentication

There is an example of context transfer authentication that is used in the IEEE 802.11f. It uses RADIUS as its source of a shared key for use as a CTP between APs.

5.3.4 WISP Protocols for Account Information

Roaming in the WLAN world requires an exchange of account information to allow the transition from one WISP to another. Pass-One is a company whose charter is to enable the move between WISPs by enabling roaming agreements between them. Much like the early days of cellular telephony, not all cell phone companies had roaming agreements and when you left the area of your local cell site, there was no service.

5.3.5 CTPs for Accounting

Context transfer is envisioned to occur via CTPs; for example, the 802.11f context transfer. If the context transfer is over the wireless connection between APs or 802.11f over the wired connection, it is on the same subnet. The context transfer could also happen using the Seamoby CTP if the transfer occurs at Layer 3 that has been triggered by a Layer 2 mechanism or a Layer 2 measurement.

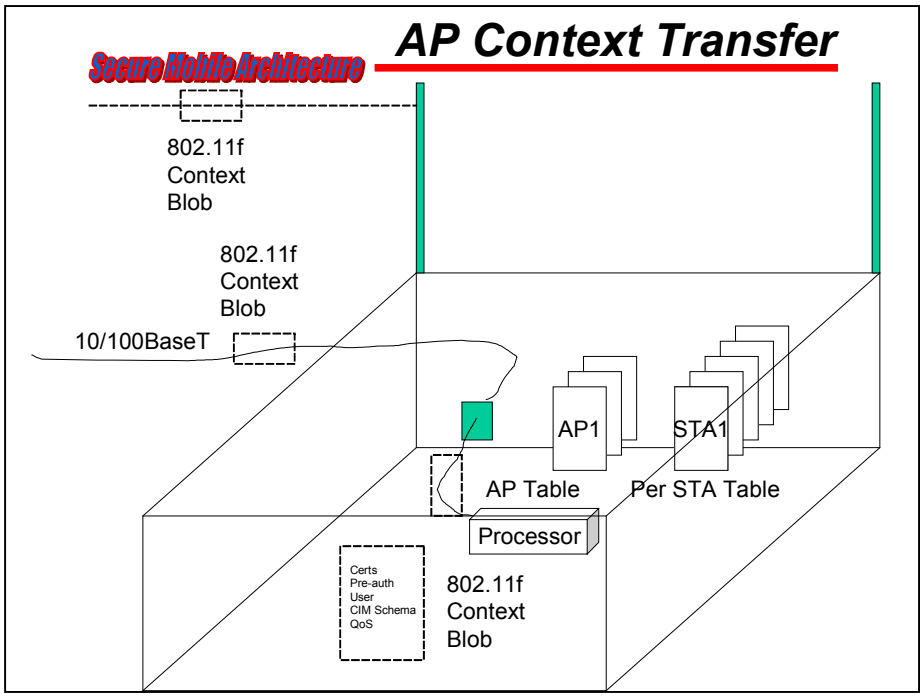

Figure 8: AP Knowledge of the Radio Environment

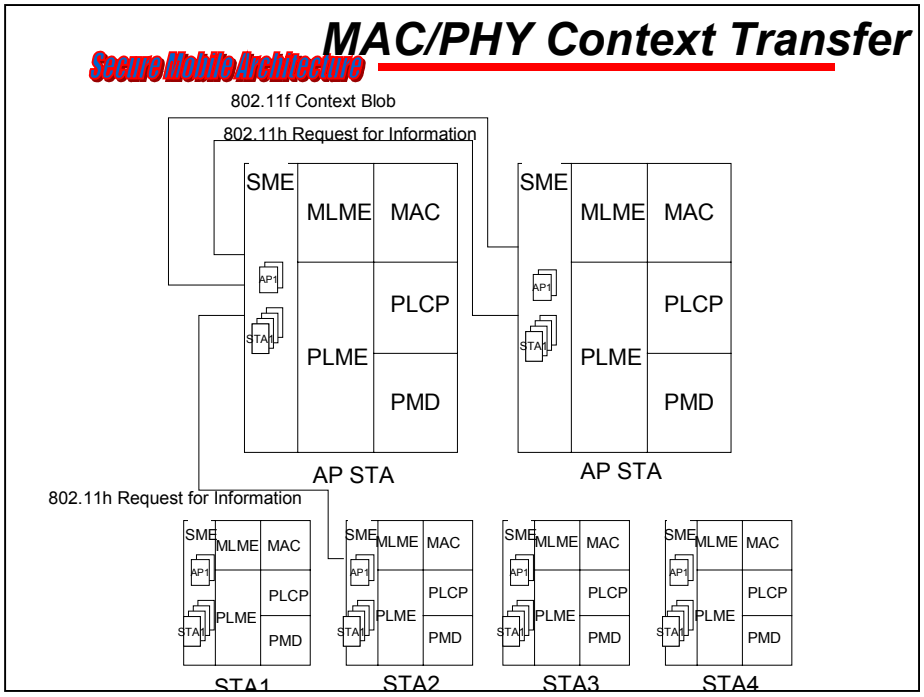

Figure 9: AP or STA Knowledge of the Radio Environment

5.4 Architectural Vision

The Secure Mobile Architecture (SMA) architectural building blocks include the components that make up an environment to enable knowledge about the infrastructure, policies that tell the components when to move, roles to imply access approval, location to enable zones of security, CTPs to facilitate pre-move authentication and authorization, and a host identity that enables identification instead of an address. Figure 10 gives a pictorial representation of the communication and the roam of a device using the SMA. Following Figure 10 is a listing of the components of the architecture and their relevance to the overall process of securely roaming in an IP-only mobile environment.

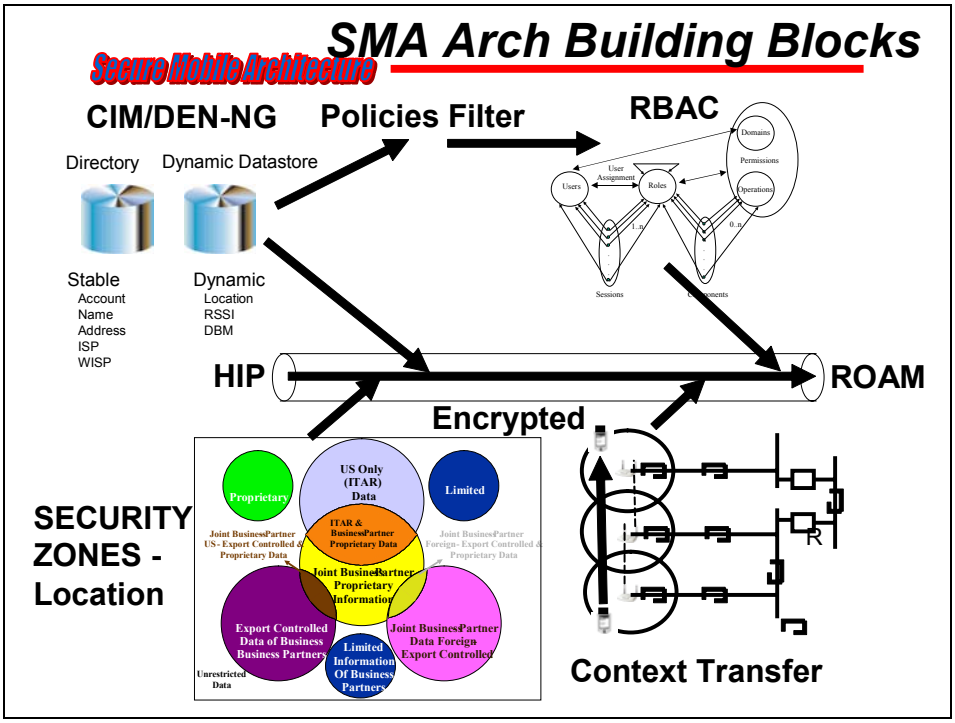

Figure 10: SMA Architectural Building Blocks

5.4.1 Components

Common Information Model (CIM)

A schema that exists in a directory (relatively static information) or in a real-time datastore (dynamic information) to maintain information about the current state of the mobile and static systems and their network.

Policy Engine (runs on the device)

1. Discovery and Binding Service – multi-layer switching algorithm which helps to figure out the services available.
2. Role-based authorization access control.

Identity Management and Authorization

(Of devices and users.) Establishes trusted computing base end-to-end (socket-to-socket). Provides the authenticated host at each end.

Security Zone

Location-specific policy.

Devices

1. IP client on device
2. Policy runs here

Transport Pipe

(IP-only is assumed.)

5.4.2 Protocols

Context Transfer Protocol (CTP)

1. Protocol exchange aids the hand-off process through pre-authentication and pre-authorization
2. IEEE 802.11e
3. IEEE 802.11f
4. IEEE 802.11i

Host Identity Tag (HIT)

1. Abstracts the session stream from the Layer 3 routing information (the HIT goes into each packet)
2. Provides end-to-end verification

Security at the Transport Layer

1. Provides end-to-end encryption, data integrity
2. Available to all applications thus removing this responsibility from the application layer. Any application's communications transfer inherits security.

5.4.3 SMA Example

The "Executive on the Move", as published by The Open Group Mobile Management Forum (MMF), presents the example of an executive moving from her offices to an airport and the

communications transitions that must occur to remain connected. The example scenario starts with a laptop docked at the desk over wire-line Ethernet connection and then migrates to a wireless connection over Wi-Fi within a corporate facility, followed by a second migration to a WAN connection (such as GPRS or other terrestrial IP network).

Thus we describe below a wire-line Ethernet connection to Wi-Fi to GPRS hand-off sequence of the same device and user. This is also moving from a corporate intranet to a public network. The initial assumptions are that each IP network interface has already been provisioned by the attached network, policy has been distributed and a common Dynamic DNS (DDNS) server is globally available.

Phase 1: Moving from Cable Ethernet to Wi-Fi (Intra-Network)

Step 1: A communications session is established with a peer.

1. DNS lookup for the peer's DNS name:
    a. HIT (immutable)
    b. Current IP address (transient)
2. Establish a secure transport-layer session:
    c. Uses the HIT instead of the IP address as the part of its session identifier for the desired transport protocol.
    d. Transport Layer Security (TLS) is the transport that leaves the header information unencrypted and encrypts the payload.
3. Once the session is established it is now secure and application traffic can now flow.

Step 2: When the client detects the IP interface it is using to communicate with, its peer is no longer available.

1. The disconnect triggers the policy engine service.
2. The policy engine then evaluates alternative paths available for communications to the peer.
3. Having selected the alternative route, the client notifies the peer that its IP address has changed for the established session. Its HIT remains unchanged.
4. Once the session is re-established it is now secure and application traffic can now flow again.

Phase 2: Wi-Fi to GPRS (Inter-Network Transfer)

The Wi-Fi signal weakens and the device disassociates from the AP. The policy engine determines that the only alternative is to move to the GPRS network and, knowing the GPRS account information, establishes a secure session with GPRS and the device enables packets to be sent out over the GPRS network rather than the Wi-Fi. The Wi-Fi connection disassociates and the hand-off is accomplished without a user perceptible disconnect.

5.4.4 Exception Handling

Both MNs roam at the same time:

1. If both MNs roam at the same time, then they will lose the ability to inform each other of their respective new addresses and so they must both do a secure update to the common DDNS.

2. Each peer does a DDNS lookup to re-find the other.

5.5 Policy

A policy engine is required to interpret a policy and send that policy to Policy Decision Points (PDPs) and Policy Enforcement Points (PEPs). The policy-based interpretation of such information as location or radio signal and its affect on the network and network service are a necessary element of the SMA. Without such a policy mechanism, there is no way to have relevant changes interpreted and incorporated into the network infrastructure.

5.5.1 Roaming and Security Policy Available to PDPs and PEPs

There may be restrictions on roaming and network availability when the policies are decided upon or enforced. In the location example, if a radio is outside the fence or in the parking lot, there could be restrictions on access to the network based on the location of the device, rather than on radio signal strength.

5.5.2 Policy Enforceable at the Network Level

Policies may be enforced at the network level instead of at the application level. In this case the network equipment, such as switches and routers, are the means by which policies are enforced. In fact, there are a number of policies that can only be enforced at the switches and routers, such as ports and routes.

5.5.3 Policy Decisions and Enforcement at the Application Level

The policies interpretable at the application level include those things, such as Role-Based Access Control (RBAC) that can impact the availability of the network to those authorized to access it. This may include such mechanisms such as the Microsoft group authentication mechanisms.

5.6 Infrastructure

The infrastructure of the SMA includes the physical network elements of the roaming – that is, the wireless networks – that enable roaming.

5.6.1 PAN Infrastructure

PANs do not enable roaming in the sense that they are subject to very short distances (10m) and centered on the end user with a personal space in which there is a network. The PANs are really

only going to be cable replacement technologies and therefore not a part of the SMA roaming network.

5.6.2 Enterprise WLAN Infrastructure

The roaming between cellular and WLAN infrastructure is the core of the SMA infrastructure. The ability to move in and out of hotspots and hotzones and onto the cellular data networks makes a complete story about roaming for an enterprise or organization with a requirement for seamless mobility.

5.6.3 DHCP

The need for a background infrastructure to deliver IP addresses across the WLAN/cellular data is a core part of the SMA roaming infrastructure. These components exist in the separate networks, but must interplay in order for such roaming to work.

5.6.4 DDNS

The need for a DNS is an integral part of the SMA roaming infrastructure. One aspect is a repository for the HIP hash. The other is the integral nature of DNS as the address interpreter of the network. Another aspect is the ability to accept and deliver dynamic changes to the naming; specifically DDNS.

5.6.5 Session Persistence

The sessions must persist in the SMA-based network. They must also persist with state being transferred by protocol rather than held by a server.

5.6.6 Billing

The SMA network must carry the information that enables billing, whether the billing is done by time or by volume, by connection, by flat fee, or by free service.

5.6.7 Hand-Off

In the SMA network, hand-off occurs at Layer 2. CTPs pass information between cells (either WLAN or cellular data cells) to do the passing of information.

5.6.8 Cellular Infrastructure

The integration of the hand-off mechanisms on the cellular side already exists in a controlled manner by the cellular providers. The integration required on the cellular side includes the billing mechanisms, location, and hand-off between the cellular providers and the WLANs. Basically, the mechanisms must be relevant to the TCP/IP routing and switching infrastructure that is becoming prevalent in both cellular data and WLAN.

5.6.9 Satellite Infrastructure

In the SMA, the satellite infrastructure also needs roaming characteristics and features to transition it to the WLAN infrastructure.

5.6.10 Directory-Enabled Network (DEN)

The DEN forms the core of an SMA and contains the static information about movement, location, hand-off, and other relevant roaming information.

5.6.11 Real-Time Databases

A real-time datastore forms the core of an SMA and contains the nomadic information about movement, location, hand-off, and other relevant roaming information.

5.7 Secure Mobile Architecture (SMA)

5.7.1 Security Framework

The security framework, being based on the HIP, changes the basic nature of Internet communications. In fact, HIP can be used as the addressing infrastructure for transport by using the host identity to transport the packets in a session between two hosts. In addition to a simplification of the routing and transport, HIP ends the perennial discussion about the vulnerability of TCP/IP to the man-in-the-middle attacks and the address spoofing that has gone on since the early days of the Internet.

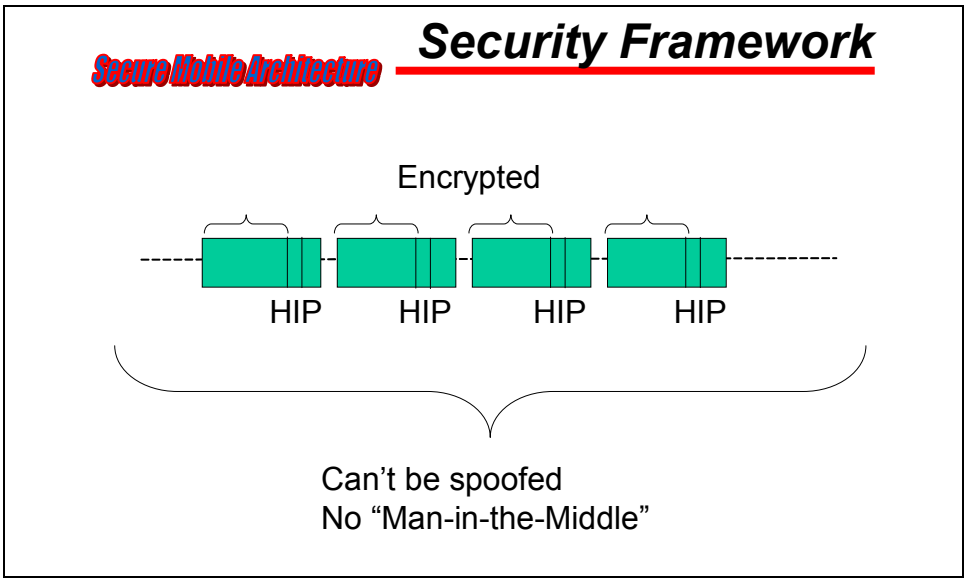

Figure 11: HIP-Enabled Security Framework

5.7.2 Mobility Framework

One of the premises at the start of this effort was that the basic requirement for the SMA is Voice Over IP (VOIP) over the WLAN and being able to transition onto the cellular data network when out of range of WLAN hotspots. Figure 12 gives a view of that mobility framework. The Internet extends across the world. The cellular WWAN has a larger footprint than the WLAN hotspot or hotzone, and the transition must occur when the roamer goes out of range of the WLAN. This transition is an integral process enabled by the hand-off of state between the cellular WWAN and the WLAN hotspot or hotzone. The organizations with the potential to achieve this are IEEE or the IETF, by enabling a CTP that works in a standardized manner between the IEEE's 802 PHY/MAC protocols and the IETF's IP layer protocols. IEEE 802 is working on a seamless CTP hand-off between the 802 wireless families of protocols (802.11, 802.15, 802.16, and 802.20). Coordination between the IEEE working on these protocols and the IETF's Seamoby Working Group may enable this to happen.

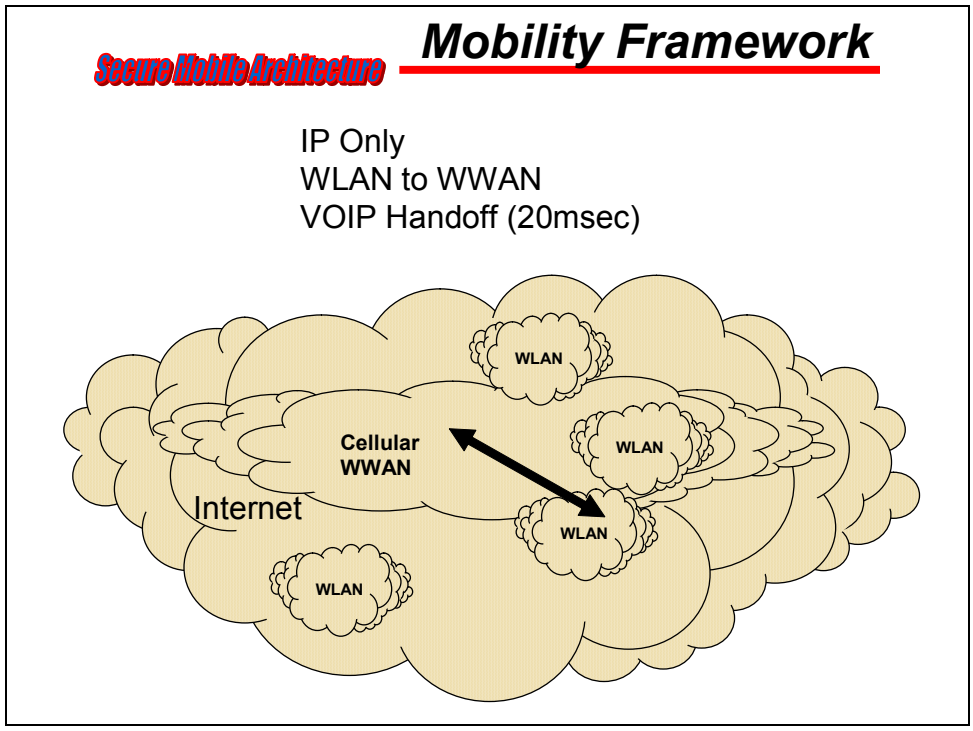

Figure 12: Mobility Framework for Roaming in IP Networks

5.7.3 Implementation Framework

The implementation framework consists of a number of services. The distribution service consists of a secure Distributed Name Service (DNS) that holds the addresses and the host identity.

The Location Enabled Network Service (LENS) makes location available as a tool for security, provisioning, and other workflow mechanisms to contribute to whatever process is being followed to improve productivity.

The addressing service also contains a Dynamic Host Configuration Protocol (DHCP) that issues IP addresses and associates them with MAC addresses.

The transport service may be based on Session Initiation Protocol (SIP) or HIP as a means of doing session-oriented VOIP or collaboration.

The HIP service enables the packets to be identifiable as belonging to a particular host and differentiates each packet by its host identity.

The new security service uses the combination of IPSEC and host identity, plus the location for establishing a relationship of host to process and to security zones to effectively eliminate man-in-the-middle and spoofing attacks.

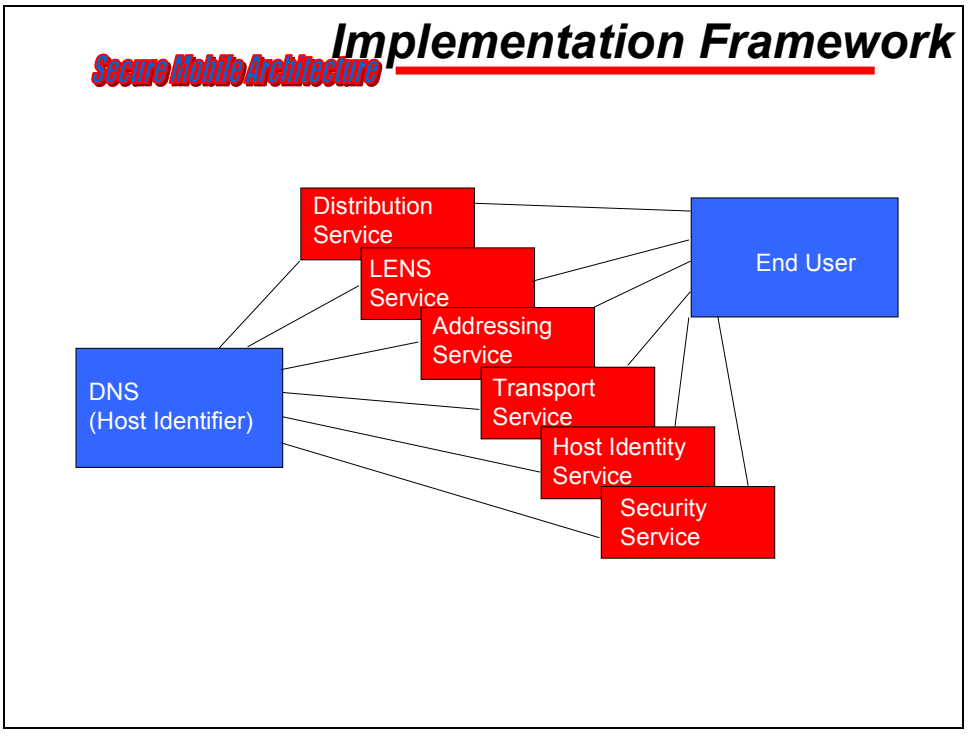

Figure 13: Network Services to Support Mobility and Roaming

5.7.4 Deployment Framework

The deployment framework is an addition to the existing infrastructure and can be deployed by using DNS proxies and DHCP proxies to implement this HIP/SIP environment without disruption to the existing infrastructure. The following is a sample prototype.

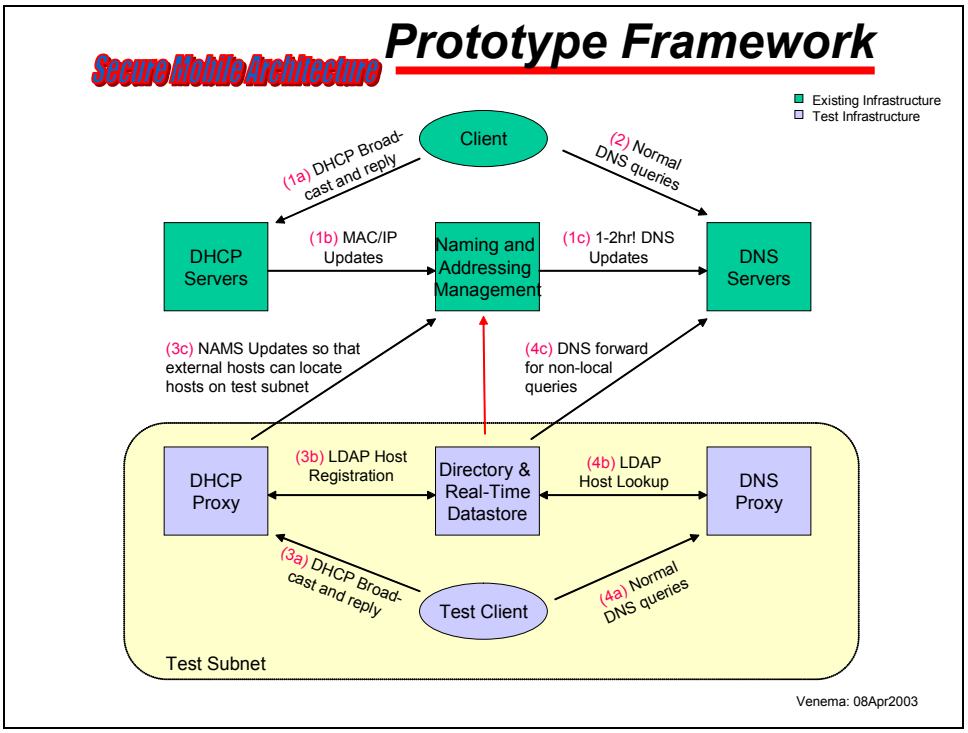

Figure 14: Potential Concept Demonstration Layout for SMA Roaming

In this example, there are a number of specific devices and software systems that are needed to prototype an SMA infrastructure. Figure 15 shows a sample set of the software systems needed to prototype.

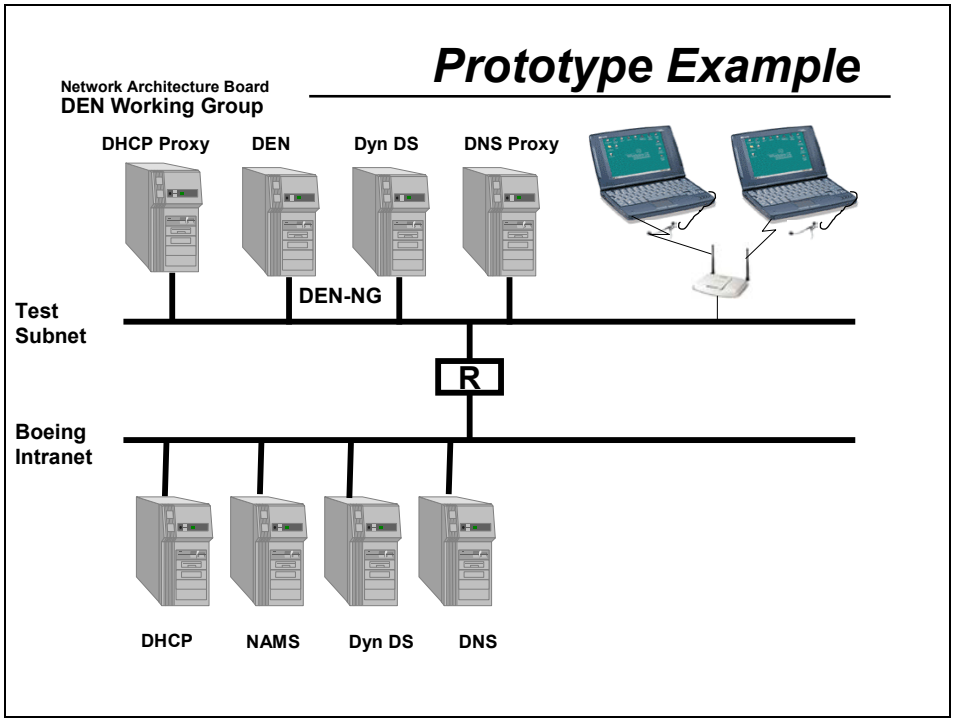

Figure 15: Networked Devices for Proving SMA Concepts for Roaming

6 SMA Recommended Practices

6.1 Recommended Practice #1: Session Management

Until ubiquitous wireless coverage is a reality, session management is a necessary part of doing business in a mobile environment. Plan for and develop a session management client-server solution for mobile and roaming users.

6.2 Recommended Practice #2: Wireless

Meet the need for a WLAN solution first. The WLAN technology is of the most immediate concern. Deal with the most immediate need first and then pursue the interfaces and transition issues around moving to other wireless Internet infrastructures, such as cellular and satellite data services.

6.3 Recommended Practice #3: Security

Develop a security architecture that takes into account all the layers of wireless security (see Figure 6), deals with roaming, and is an integrated approach to end-to-end security. The nature of the SMA approach is to use a Host Identity Payload (HIP) as the core security feature of the network architecture to move away from address-based security. The SMA approach also uses Context Transfer Protocols (CTPs) to pass security from wireless Access Point (AP) to wireless AP.

6.4 Recommended Practice #4: Roaming

Design the mobile network approach to accommodate Voice Over IP (VOIP) over the WLAN. Such an approach means dealing with very fast hand-offs and most likely includes the use of CTPs to move a voice stream from cell to cell.

6.5 Recommended Practice #5: Vision

Create an IT body to deal with wireless and mobility issues. Create an architecture that technically meets the wireless and mobility requirements within your organization. There are enough technology issues around wireless and mobility that the approach must be dealt with in a consistent and persistent manner.

Glossary

General Terms

Asymmetric Link
: A link with transmission characteristics which are different depending upon the relative position or design characteristics of the transmitter and the receiver of data on the link. For instance, the range of one transmitter may be much higher than the range of another transmitter on the same medium.

Bandwidth
: The total capacity of a link to carry information (typically bits) per unit time.

Bandwidth Utilization
: The actual rate of information transfer achieved over a link, expressed as a percentage of the available bandwidth on that link.

Beacon
: A control message broadcast by a node (especially, a base station) informing all the other nodes in its neighborhood of the continuing presence of the broadcasting node, possibly along with additional status or configuration information.

Binding Update (BU)
: A message indicating an MN's current mobility binding, and in particular its care-of address.

Care-of Address (CoA)
: An IP address associated with an MN while visiting a foreign link; the subnet prefix of this IP address is a foreign subnet prefix. Among the multiple care-of addresses that an MN may have at any given time (e.g., with different subnet prefixes), the one registered with the MN's home agent is called its “primary” care-of address.

Channel
: A subdivision of the physical medium allowing possibly shared independent uses of the medium. Channels may be made available by subdividing the medium into distinct time slots, or distinct spectral bands, or decorrelated coding sequences.

Channel Access Protocol
: A protocol for mediating access to, and possibly allocation of, the various channels available within the physical communications medium. Nodes participating in the channel access protocol agree to communicate only when they have uncontested access to one of the channels, so that there will be no interference.

Control Message
: Information passed between two or more network nodes for maintaining protocol state, which may be unrelated to any specific application.

Distance Vector
: A style of routing protocol in which, for each desired destination, a node maintains

information about the distance to that destination, and a vector (next hop) towards that destination.

Fairness A property of channel access protocols whereby a medium is made fairly available to all eligible nodes on the link. Fairness does not strictly imply equality, especially in cases where nodes are given link access according to unequal priority or classification.

Flooding The process of delivering data or control messages to every node within the network under consideration.

Foreign Subnet Prefix
A bit string that consists of some number of initial bits of an IP address which identifies a node's foreign link within the Internet topology.

Forwarding Node
A node which performs the function of forwarding datagrams from one of its neighbors to another.

Home Address
An IP address assigned to an MN, used as the permanent address of the MN. This address is within the MN's home link. Standard IP routing mechanisms will deliver packets destined for an MN's home address to its home link.

Home Subnet Prefix
A bit string that consists of some number of initial bits of an IP address which identifies a node's home link within the Internet topology (i.e., the IP subnet prefix corresponding to the MN's home address).

Interface A node's attachment to a link.

IP Access Address
An IP address (often dynamically allocated) that a node uses to designate its current point of attachment to the local network. The IP access address is typically to be distinguished from the MN’s home address; in fact, while visiting a foreign network the former may be considered unsuitable for use as an end-point address by any but the most short-lived applications. Instead, the IP access address is typically used as the care-of address of the node.

Link A communication facility or physical medium that can sustain data communications between multiple network nodes, such as an Ethernet (simple or bridged). A link is the layer immediately below IP.

Link Establishment
The process of establishing a link between the MN and the local network. This may involve allocating a channel, or other local wireless resources, possibly including a minimum level of service or bandwidth.

Link-layer Trigger (Layer 2 Trigger)
Information from Layer 2 that informs Layer 3 of the detailed events involved in handover sequencing at Layer 2. Layer 2 triggers are not specific to any particular

Layer 2, but rather represent generalizations of Layer 2 information available from a wide variety of Layer 2 protocols.

Link-level Acknowledgement
: A protocol strategy, typically employed over wireless media, requiring neighbors to acknowledge receipt of packets (typically unicast only) from the transmitter. Such strategies aim to avoid packet loss or delay resulting from lack of, or unwanted characteristics of, higher-level protocols. Link-layer acknowledgements are often used as part of ARQ algorithms for increasing link reliability.

Link State
: A style of routing protocol in which every node within the network is expected to maintain information about every link within the network topology.

Local Broadcast
: The delivery of data to every node within range of the transmitter.

Loop-free
: A property of routing protocols whereby the path taken by a data packet from source to destination never transits the same intermediate node twice before arrival at the destination.

Medium-Access Protocol (MAC)
: A protocol for mediating access to, and possibly allocation of, the physical communications medium. Nodes participating in the medium access protocol can communicate only when they have uncontested access to the medium, so that there will be no interference. When the physical medium is a radio channel, the MAC is the same as the Channel Access Protocol.

Mobile Network Prefix
: A bit string that consists of some number of initial bits of an IP address which identifies the entire mobile network within the Internet topology. All nodes in a mobile network necessarily have an address named after this prefix.

Mobility Factor
: The relative frequency of node movement, compared to the frequency of application initiation.

Multipoint Relay (MPR)
: A node which is selected by its one-hop neighbor to re-transmit all broadcast messages that it receives. The message must be new and the time-to-live field of the message must be greater than one. Multipoint relaying is a technique to reduce the number of redundant retransmissions while diffusing a broadcast message in the network.

Neighbor
: A neighbor is any other node to which data may be propagated directly over the communications medium without relying on the assistance of any other forwarding node.

Neighborhood
: All the nodes that can receive data on the same link from one node whenever it transmits data.

Next Hop A neighbor which has been selected to forward packets along the way to a particular destination.

Payload The actual data within a packet, not including network protocol headers that were not inserted by an application. Note that payloads are different between layers: user data is the payload of TCP, which is the payload of IP, which is the payload of link-layer protocols, etc. Thus, it is important to identify the scope when talking about payloads.

Prefix A bit string that consists of some number of initial bits of an address.

Route Activation
The process of putting a route into use after it has been determined.

Route Entry
An entry for a specific destination (unicast or multicast) in the route table.

Route Establishment
The process of determining a route between a source and a destination.

Route Table
The table where forwarding nodes keep information (including next hop) for various destinations.

Routing Proxy
A node that routes packets by overlays; e.g., by tunneling, between communicating partners. The Home Agent and Foreign Agent are examples of routing proxies, in that they receive packets destined for the MN and tunnel them to the current address of the MN.

Signal Strength
The detectable power of the signal carrying the data bits, as seen by the receiver of the signal.

Source Route
A source route from node A to node B is an ordered list of IP addresses, starting with the IP address of node A and ending with the IP address of node B. Between A and B, the source route includes an ordered list of all the intermediate hops between A and B, as well as the interface index of the interface through which the packet should be transmitted to reach the next hop.

Spatial Re-use
Simultaneous use of channels with identical or close physical characteristics, but located spatially far enough apart to avoid interference (i.e., co-channel interference).

System-wide Broadcast
Same as flooding, but used in contrast to local broadcast.

Topology A network can be viewed abstractly as a “graph” whose “topology” at any point in time is defined by set of “points” connected by (possibly directed) “edges”.

Triggered Update
: An unsolicited route update transmitted by a router along a path to a destination.

Mobile Access Networks and Mobile Networks

In order to support host mobility, a set of nodes towards the network edge may need to have specific functions. Such a set of nodes forms a mobile Access Network (AN) that may or may not be part of the global Internet. Figure 1 presents two examples of such AN topologies. The figure depicts a reference architecture that illustrates an IP network with components defined in this section.

We intend to define the concept of the AN which may also support enhanced mobility. It is possible that to support routing and QoS for MNs, existing routing protocols (e.g., OSPF or other standard IGPs) may not be appropriate to maintain forwarding information for these MNs as they change their points of attachment to the AN. These new functions are implemented in routers with additional capability. We can distinguish three types of AN components: Access Routers (AR) which handle the last hop to the mobile, typically over a wireless link; Access Network Gateways (ANG) which form the boundary on the fixed network side and shield the fixed network from the specialized routing protocols; and (optionally) other internal Access Network Routers (ANR) which may also be needed in some cases to support the protocols. The AN consists of the equipment needed to support this specialized routing; i.e., AR or ANG. AR and ANG may be the same physical nodes.

In addition, we present a few basic terms on mobile networks; that is, Mobile Network, Mobile Router (MR), and Mobile Network Modes (MNN). A more thorough discussion on mobile networks can be found in the working group documents of the NEMO Working Group.

Access Link (AL)
: A last-hop link between an MN and an AR; that is, a facility or medium over which an AP and the MN can communicate at the link layer; i.e., the layer immediately below IP.

Access Network (AN)
: An IP network that includes one or more ANRs.

Access Network Gateway (ANG)
: An ANR that separates an AN from other IP networks, much in the same way as an ordinary gateway router. The ANG looks to the other IP networks like a standard IP router.

Access Network Router (ANR)
: An IP router in the AN. An ANR may include AN-specific functionalities; for example, related to mobility and/or QoS. This is to distinguish between ordinary routers and routers that have AN-related special functionality.

Access Point (AP)
: An AP is a Layer 2 device that is connected to one or more ARs and offers the wireless link connection to the MN. APs are sometimes called base stations or AP transceivers. An AP may be a separate entity or co-located with an AR.

Access Router (AR)

An ANR residing on the edge of an AN, and connected to one or more APs. The APs may be of different technology. An AR offers IP connectivity to MNs, acting as a default router to the MNs it is currently serving. The AR may include intelligence beyond a simple forwarding service offered by ordinary IP routers.

Administrative Domain (AD)

A collection of networks under the same administrative control and grouped together for administrative purposes.

Candidate Access Router (CAR)

An AR to which the MN may do a hand-off.

Fixed Node (FN)

A node, either a host or a router, unable to change its point of attachment to the network and its IP address without breaking open sessions.

Mobile Host (MH)

An MN that is an end host and not a router. An MH is capable of sending and receiving packets; that is, being a source or destination of traffic, but not a forwarder of it.

Mobile Network

An entire network, moving as a unit, which dynamically changes its point of attachment to the Internet and thus its reachability in the topology. The mobile network is composed of one or more IP-subnets and is connected to the global Internet via one or more MRs. The internal configuration of the mobile network is assumed to be relatively stable with respect to the MR.

Mobile Network Node (MNN)

Any node (host or router) located within a mobile network, either permanently or temporarily. An MNN may either be an MN or an FN.

Mobile Node (MN)

An IP node capable of changing its point of attachment to the network. An MN may or may not have forwarding functionality.

Mobile Router (MR)

A router capable of changing its point of attachment to the network, moving from one link to another link. The MR is capable of forwarding packets between two or more interfaces, and possibly running a dynamic routing protocol modifying the state by which to do packet forwarding.

The interface of an MR attached to a link inside the mobile network is called the ingress interface. The interface of an MR attached to the home link if the MR is at home, or attached to a foreign link if the MR is in a foreign network, is called the egress interface.

An MR acting as a gateway between an entire mobile network and the rest of the Internet has one or more egress interface(s) and one or more ingress interface(s).

Packets forwarded upstream to the rest of the Internet are transmitted through one of the MR's egress interfaces; packets forwarded downstream to the mobile network are transmitted through one of the MR's ingress interfaces.

New Access Router (NAR)
: The AR that offers connectivity to the MN after a handover.

Old Access Router (OAR)
: An AR that offered connectivity to the MN prior to a handover. This is the SAR that will cease or has ceased to offer connectivity to the MN.

Previous Access Router (PAR)
: An AR that offered connectivity to the MN prior to a handover. This is the SAR that will cease or has ceased to offer connectivity to the MN. Same as OAR.

Radio Cell
: The geographical area within which an AP provides radio coverage; i.e., where radio communication between an MN and the specific AP is possible.

Serving Access Router (SAR)
: The AR currently offering the connectivity to the MH. This is usually the point of departure for the MN as it makes its way towards a NAR (then the SAR takes the role of the OAR). There may be several SARs serving the MN at the same time.

Handover Terminology

These terms refer to different perspectives and approaches to supporting different aspects of mobility. Distinctions can be made according to the scope, range overlap, performance characteristics, diversity characteristics, state transitions, mobility types, and control modes of handover techniques.

Handover
: (Also known as hand-off.) The process by which an active MN changes its point of attachment to the network, or when such a change is attempted. The AN may provide features to minimize the interruption to sessions in progress. There are different types of handover classified according to different aspects involved in the handover.

Roaming
: An operator-based term involving formal agreements between operators that allows a mobile to get connectivity from a foreign network. Roaming (a particular aspect of user mobility) includes, for example, the functionality by which users can communicate their identity to the LAN so that inter-AN agreements can be activated and service and applications in the MN's home network can be made available to the user locally.

Scope of Handover

Note that the definitions of horizontal and vertical handover are different than the ones commonly used today. These definitions try to look at the handover from the IP layer's point of view; the IP layer works with network interfaces, rather than specific technologies used by those interfaces.

Horizontal Handover

A handover in which the MN's network interface does not change (from the IP point of view); the MN communicates with the AR via the same network interface before and after the handover. A horizontal handover is typically also an intra-technology handover, but it can be an inter-technology handover if the MN can do a Layer 2 handover between two different technologies without changing the network interface seen by the IP layer.

Inter-AN Handover

When the MN moves to a new AN, then this handover occurs. This requires some sort of host mobility across ANs, which typically is be provided by the external IP core. Note that this would have to involve the assignment of a new IP access address (e.g., a new care-of address) to the MN.

Inter-technology Handover

A handover between equipment of different technologies.

Intra-AN Handover

When the MN changes ARs inside the same AN, then this handover occurs. Such a handover is not necessarily visible outside the AN. In case the ANG serving the MN changes, this handover is seen outside the AN due to a change in the routing paths. Note that the ANG may change for only some of the MN's data flows.

Intra-AR Handover

A handover which changes the AR's network interface to the mobile. That is, the SAR remains the same but routing changes internal to the AR take place.

Intra-technology Handover

A handover between equipment of the same technology.

Layer 2 Handover

When an MN changes APs (or some other aspect of the radio channel) connected to the same AR's interface, then a Layer 2 handover occurs. This type of handover is transparent to the routing at the IP layer (or it appears simply as a link layer reconfiguration without any mobility implications).

Vertical Handover

In a vertical handover the MN's network interface to the AN changes. A vertical handover is typically an inter-technology handover, but it may also be an intra-technology handover if the MN has several network interfaces of the same type. That is, after the handover, the IP layer communicates with the AN through a different network interface.

Note that the horizontal and vertical handovers are not tied to a change in the link layer technology. They define whether, after a handover, the IP packet flow goes through the same (horizontal handover) or a different (vertical handover) network interface. These two handovers do not define whether the AR changes as a result of a handover.

Handover Control

A handover must be one of the following two types (a):

Mobile-initiated Handover
: The MN is the one that makes the initial decision to initiate the handover.

Network-initiated Handover
: The network makes the initial decision to initiate the handover.

A handover is also one of the following two types (b):

Mobile-controlled Handover (MCHO)
: The MN has the primary control over the handover process.

Network-controlled Handover (NCHO)
: The network has the primary control over the handover process.

A handover is also either of these three types (c):

Mobile-assisted Handover
: Information and measurement from the MN are used by the AR to decide on the execution of a handover.

Network-assisted Handover
: A handover where the AN collects information that can be used by the MN in a handover decision.

Unassisted Handover
: A handover where no assistance is provided by the MN or the AR to each other.

Note that it is possible that the MN and the AR both do measurements and decide on the handover.

A handover is also one of the following two types (d):

Backward Handover
: A handover either initiated by the OAR, or where the MN initiates a handover via the OAR.

Forward Handover
: A handover either initiated by the NAR, or where the MN initiates a handover via the NAR.

The handover is also either proactive or reactive (e):

Planned Handover
: A proactive (expected) handover where some signalling can be done in advance of the MN getting connected to the NAR; e.g., building a temporary tunnel from the OAR to the NAR.

Unplanned Handover
: A reactive (unexpected) handover, where no signalling is done in advance of the MN's move of the OAR to the NAR.

The five handover types (a-e) are mostly independent, and every handover should be classifiable according to each of these types.

Simultaneous Connectivity to Access Routers

Break-before-make (BBM)
: During a BBM handover the MN cannot communicate simultaneously with the old and the NAR.

Make-before-break (MBB)
: During an MBB handover the MN can communicate simultaneously with the old and NAR. This should not be confused with “soft handover” which relies on macro diversity.

Performance and Functional Aspects

Exposed-terminal Problem
: The problem whereby a transmitting node prevents another node from transmitting although it could have safely transmitted to anyone else but that node.

Fast Handover
: A handover that aims primarily to minimize delay, with no explicit interest in packet loss.

Goodput
: The total bandwidth used, less the volume of control messages, protocol overhead from the data packets, and packets dropped due to CRC errors.

Handover Latency
: Handover latency is the time difference between when an MN is last able to send and/or receive an IP packet by way of the OAR, until when the MN is able to send and/or receive an IP packet through the NAR.

Hidden-terminal Problem
: The problem whereby a transmitting node can fail in its attempt to transmit data because of destructive interference which is only detectable at the receiving node, not the transmitting node.

Pathloss
: A reduction in signal strength caused by traversing the physical medium constituting the link.

Seamless Handover
: A handover in which there is no change in service capability, security, or quality. In practice, some degradation in service is to be expected. The definition of a seamless handover in the practical case should be that other protocols, applications, or end users do not detect any change in service capability, security, or quality, which would have a bearing on their (normal) operation.

Smooth Handover
: A handover that aims primarily to minimize packet loss, with no explicit concern for additional delays in packet forwarding.

Throughput
: The amount of data from a source to a destination processed by the protocol for which throughput is to be measured; for instance, IP, TCP, or the MAC protocol. The throughput differs between protocol layers.

Micro Diversity, Macro Diversity, and IP Diversity

Certain air interfaces (e.g., the Universal Mobile Telephone System (UMTS) Terrestrial Radio Access Network (UTRAN) running in Frequency Division Duplex (FDD) mode) require or at least support macro diversity combining. Essentially, this refers to the fact that a single MN is able to send and receive over two independent radio channels (diversity branches) at the same time; the information received over different branches is compared and that from the better branch passed to the upper layers. This can be used both to improve overall performance, and to provide a seamless type of handover at Layer 2, since a new branch can be added before the old is deleted.

It is necessary to differentiate between combining/diversity that occurs at the physical and radio link layers, where the relevant unit of data is the radio frame, and that which occurs at Layer 3, the network layer, where what is considered is the IP packet itself.

In the following definitions, micro- and macro diversity refer to protocol layers below the network layer, and IP diversity refers to the network layer.

IP Diversity
: The splitting and combining of packets at the IP level.

Macro Diversity
: Duplicating or combining actions taking place over multiple APs, possibly attached to different ARs. This may require support from the network layer to move the radio frames between the base stations and a central combining point.

Micro Diversity
: For example, two antennas on the same transmitter send the same signal to a receiver over a slightly different path to overcome fading.

Paging, and MN States and Modes

Mobile systems may employ the use of MN states in order to operate more efficiently without degrading the performance of the system. The term “mode” is also common and means the same as “state”.

An MN is always in one of the following states:

Active State
: When the AN knows the MN's SAR and the MN can send and receive IP packets. The AL may not be active, but the radio layer is able to establish one without assistance from the network layer. The MN has an IP address assigned.

Dormant State

A state in which the mobile restricts its ability to receive normal IP traffic by reducing its monitoring of radio channels. The AN knows the MH's Paging Area, but the MH has no SAR and so packets cannot be delivered to the MH without the AN initiating paging.

Inactive State

The MH is in neither the Active nor Dormant state. The host is no longer listening for any packets, not even periodically, and not sending packets. The host may be in a powered-off state, it may have shut down all interfaces to drastically conserve power, or it may be out of range of a radio AP. The MN does not necessarily have an IP access address from the AN.

Time-slotted Dormant Mode

A dormant mode implementation in which the mobile alternates between periods of not listening for any radio traffic and listening for traffic. Time-slotted dormant mode implementations are typically synchronized with the network so the network can deliver traffic to the mobile during listening periods.

Note that as well as the MN being in one of these states, the AN also stores which state it believes the MN is in. Normally these are consistent; the definitions above assume so.

Here are some additional definitions for paging, taking into account the above state definitions.

Location Updating

A procedure initiated by the MN, by which it informs the AN that it has moved into a new paging area.

Paging A procedure initiated by the AN to move an Idle MN into the Active state. As a result of paging, the MN establishes an SAR and the IP routes are set up.

Paging Area

A part of the AN, typically containing a number of ARs/APs, which corresponds to some geographical area. The AN keeps and updates a list of all the Idle MNs present in the area. If the MN is within the radio coverage of the area it will be able to receive paging messages sent within that Paging Area.

Paging Area Registrations

Signaling from a dormant mode MN to the network, by which it establishes its presence in a new paging area. Paging Area Registrations thus enable the network to maintain a rough idea of where the mobile is located.

Paging Channel

A radio channel dedicated to signaling dormant mode mobiles for paging purposes. By current practice, the protocol used on a paging channel is usually dictated by the radio link protocol, although some paging protocols have provision for carrying arbitrary traffic (and thus could potentially be used to carry IP).

Traffic Channel

The radio channel on which IP traffic to an active mobile is typically sent. This channel is used by a mobile that is actively sending and receiving IP traffic, and is

not continuously active in a dormant mode mobile. For some radio link protocols, this may be the only channel available.

Context Transfer

Context
: The information on the current state of a routing-related service required to re-establish the routing-related service on a new subnet without having to perform the entire protocol exchange with the MH from scratch.

Context Transfer
: The movement of context from one router or other network entity to another as a means of re-establishing routing-related services on a new subnet or collection of subnets.

Feature Context
: The collection of information representing the context for a given feature. The full context associated with a MH is the collection of one or more feature contexts.

Routing-related Service
: A modification to the default routing treatment of packets to and from the MH. Initially establishing routing-related services usually requires a protocol exchange with the MH. An example of a routing-related service is header compression. The service may also be indirectly related to routing; for example, security. Security may not affect the forwarding decision of all intermediate routers, but a packet may be dropped if it fails a security check (can't be encrypted, authentication failed, etc.). Dropping the packet is basically a routing decision.

Candidate Access Router Discovery

Candidate Access Router (CAR)
: An AR to which the MN has a choice of performing IP-level hand-off. This means that the MN has the right radio interface to connect to an AP that is served by this AR, as well as the coverage of this AR overlaps with that of the AR to which the MN is currently attached.

Capability of Access Router
: A characteristic of the service offered by an AR that may be of interest to an MN when the AR is being considered as a hand-off candidate.

Target Access Router (TAR)
: An AR with which the procedures for the MN's IP-level hand-off are initiated. TAR is selected after running a TAR Selection Algorithm that takes into account the capabilities of CARs, preferences of the MN, and any local policies.

User, Personal, and Host Mobility

Different sorts of mobility management may be required of a mobile system. We can differentiate between user, personal, host, and network mobility.

Host Mobility
: Refers to the function of allowing an MH to change its point of attachment to the network, without interrupting IP packet delivery to/from that host. There may be different sub-functions depending on what current level of service is being provided; in particular, support for host mobility usually implies active and idle modes of operation, depending on whether the host has any current sessions or not. AN procedures are required to keep track of the current point of attachment of all the MNs or establish it at will. Accurate location and routing procedures are required in order to maintain the integrity of the communication. Host mobility is often called “terminal mobility”.

Network Mobility
: Network mobility occurs when an entire network changes its point of attachment to the Internet and, thus, its reachability in the topology, which is referred to as a mobile network.

Personal Mobility
: Complements user mobility with the ability to track the user's location and provide the user's current location to allow sessions to be initiated by and towards the user by anyone on any other network. Personal mobility is also concerned with enabling associated security, billing, and service subscription authorization made between ADs.

User Mobility
: Refers to the ability of a user to access services from different physical hosts. This usually means the user has an account on these different hosts or that a host does not restrict users from using the host to access services.

Two subcategories of mobility can be identified within either host mobility or network mobility:

Global Mobility
: Same as Macro Mobility.

Local Mobility
: Same as Micro Mobility.

Local Mobility Management
: Local Mobility Management (LMM) is a generic term for protocols dealing with IP mobility management confined within the AN. LMM messages are not routed outside the AN, although a handover may trigger Mobile IP messages to be sent to correspondent nodes and home agents.

Macro Mobility
: Mobility over a large area. This includes mobility support and associated address registration procedures that are needed when an MH moves between IP domains.

Inter-AN handovers typically involve macro-mobility protocols. Mobile-IP can be seen as a means to provide macro mobility.

Micro Mobility
: Mobility over a small area. Usually this means mobility within an IP domain with an emphasis on support for active mode using handover, although it may include idle mode procedures also. Micro-mobility protocols exploit the locality of movement by confining movement-related changes and signaling to the AN.

Specific Terminology for Mobile ad hoc Networking

Cluster
: A group of nodes located within close physical proximity, typically all within range of one another, which can be grouped together for the purpose of limiting the production and propagation of routing information.

Cluster Head
: A cluster head is a node (often elected in the cluster formation process) that has complete knowledge about group membership and link state information in the cluster. Each cluster should have one and only one cluster head.

Cluster Member
: All nodes within a cluster *except* the cluster head are called members of that cluster.

Convergence
: The process of approaching a state of equilibrium in which all nodes in the network agree on a consistent collection of state about the topology of the network, and in which no further control messages are needed to establish the consistency of the network topology.

Convergence Time
: The time which is required for a network to reach convergence after an event (typically, the movement of an MN) which changes the network topology.

Laydown
: The relative physical location of the nodes within the *ad hoc* network.

Pathloss Matrix
: A matrix of coefficients describing the pathloss between any two nodes in an *ad hoc* network. When the links are asymmetric, the matrix is also asymmetric.

Scenario
: The tuple characterizing a class of *ad hoc* networks.

Security-Related Terminology

This section includes terminology commonly used around mobile and wireless networking. Only a mobility-related subset of the entire security terminology is presented.

Authorization-enabling Extension
: An authentication that makes a (registration) message acceptable to the ultimate recipient of the registration message. An authorization-enabling extension must contain an SPI.

Mobility Security Association

A collection of security contexts, between a pair of nodes, which may be applied to mobility-related protocol messages exchanged between them. In Mobile IP, each context indicates an authentication algorithm and mode, a secret (a shared key, or appropriate public/private key pair), and a style of replay protection in use. Mobility security associations may be stored separately from the node's IPsec Security Policy Database (SPD).

Registration Key

A key used as the basis of a Mobility Security Association between an MN and a foreign agent. A registration key is typically only used once or a very few times, and only for the purposes of verifying a small volume of authentication data.

Security Context

A security context between two routers defines the manner in which two routers choose to mutually authenticate each other, and indicates an authentication algorithm and mode.

Security Parameter Index (SPI)

An index identifying a security context between a pair of routers among the contexts possible in the mobility security association.

Stale Challenge

Any challenge that has been used by the MN in a Registration Request message and processed by the Foreign Agent by relaying or generating The Foreign Agent may not be able to keep records for all previously used challenges.

Unknown Challenge

Any challenge from a particular MN that the foreign agent has no record of having put either into one of its recent Agent Advertisements or into a registration reply message to that MN.

Unused Challenge

A challenge that has not been already accepted by the Foreign Agent challenge in a corresponding Registration Reply message; i.e., a challenge that is neither unknown nor previously used. The Mobile IPv6 specification includes more security terminology related to MIPv6 bindings.