Abstract

The CDMA community, under the umbrella of the 3rd Generation Partnership Project 2, has embarked on a standardization effort for wireless data based on Mobile IP. Important issues addressed include the link layer interface to a Mobile IP foreign agent; how link-layer mobility interacts with IP-layer mobility; how virtual private network services will be supported; and how to provide authentication, authorization, and accounting in a cellular Mobile IP environment. Members of 3GPP2 are also active in the Internet Engineering Task Force's Mobile IP, ROAMOPS, and AAA working groups. Based on our experiences in this effort, this article gives an overview of the issues we have encountered in standardizing a Mobile-IP-based network architecture in a cellular telephony environment, including current points of contention, and gives a summary of the current state of the standards.

An Internet Infrastructure for Cellular CDMA Networks Using Mobile IP

Peter J. McCann and Tom Hiller, Lucent Technologies

Access to the Internet via cellular radio networks is expected to become a critical part of future wireless operators' service offerings [1]. However, because mobility is an essential characteristic of cellular networks, operators are faced with a confusing array of choices for how to architect such a network. The Internet Engineering Task Force's (IETF's) Mobile IP protocol [2] provides a standard solution for wide-area mobility at the IP layer. However, Mobile IP by itself does not solve all problems involved in providing mobile Internet access to cellular users. A wide range of standards organizations are working on different approaches to the various issues, with the eventual goal of bringing these approaches into alignment as part of the International Mobile Telephony (IMT)-2000 effort for global standardization.

The 3rd Generation Partnership Project 2 (3GPP2) is a consortium of national standards bodies tasked with developing architectures and standards for third-generation cellular networks. These standards will apply to networks that make use of the cdma2000 air interface, the high-speed code-division multiple access (CDMA) standard that is an evolution of IS-95 [3], and that have traditionally used American National Standards Institute (ANSI)-41 [4] for intersystem signaling. The cdma2000 standard supports higher speeds and a packet data mode that does not require a continuous circuit-like reservation of radio resources. The packet data network portion of this architecture will make use of the IETF's Mobile IP protocol for network-layer mobility.

Participants of 3GPP2 include the Telecommunications Industry Association (TIA) of North America, which prior to the formation of 3GPP2 was working on similar issues in its TR45.6 study group and elsewhere; the Telecommunications Technology Association (TTA) of Korea; the China Wireless Telecommunication Standard (CWTS) of China; and the Association of Radio Industries and Businesses (ARIB)/TTC groups of Japan. Once technical specifications are completed in 3GPP2, they will be standardized by the participating standards organizations. Another effort, known as 3GPP, is currently underway to standardize a network architecture known as the Universal Mobile Telecommunications System (UMTS) for Global System for Mobile Communications (GSM) and wideband CDMA (W-CDMA), a CDMA standard developed for Europe. Here the core network will be based on the General Packet Radio Service (GPRS) specification developed over the past few years by the European Telecommunications Standards Institute (ETSI) rather than on Mobile IP. Lastly, recent agreements between carriers have prompted an effort to create a global third-generation (G3G) CDMA wireless standard. This standard will encompass multiple modes of operation: a direct-spread mode, which is based on the W-CDMA of UMTS, and a multicarrier mode based on cdma2000. It is not yet apparent whether the network-layer mobility aspects of this architecture will be harmonized.

A CDMA Mobile IP Packet Data Architecture

CDMA networks based on the IS-95 standard [3] are in widespread use today in providing voice services to mobile subscribers. Voice service is provided via a low-rate (13 or 8 kb/s) circuit connection over the air between a mobile node (MN) and a selection and distribution unit (SDU) via one or more base transceiver stations (BTSs). The SDU combines the signals from multiple BTSs and converts the low-rate compressed data to and from 64 kb/s traffic used by the public switched telephone network (PSTN). It is also responsible for selecting which BTSs currently in the range of the MN will transmit traffic in the forward direction, and for managing the power of transmissions to and from the MN. This provides some degree of mobility: as nodes move from BTS to BTS they can usually remain bound to the same SDU. This is known as *soft handoff*. The SDU functionality can be centralized (e.g., on a class 5 switch), but it may also be more distributed. Logically, the SDU is part of a larger base station (BS) complex consisting of the SDU, one or more BTSs, and other components. Figure 1 illustrates this basic architecture. Figure 1 also depicts the mobile switching center (MSC), which is responsible for interworking with the PSTN and managing one or more BSs.

The deployment of packet data services on these networks must coexist with the existing voice service. In particular, providing packet data should not add unnecessary cost to the network elements. In practice this implies that the network

1070-9916/00/$10.00 © 2000 IEEE

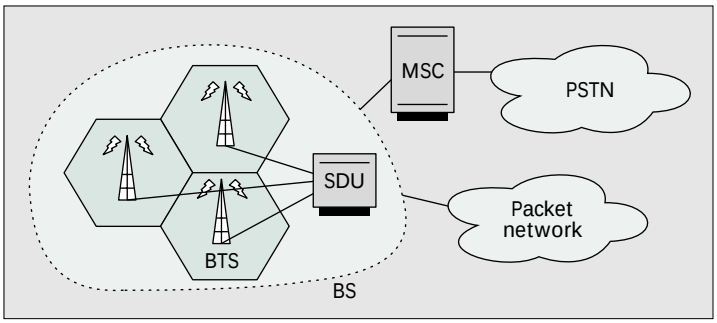

Figure 1. *The basic architecture of a CDMA cellular network.*

architecture should be kept unchanged as much as possible, and especially changes to the elements on the left side of Fig. 1 (which are more numerous) should be kept to a minimum. For packet data sessions, the radio resources will be released for use by other stations when they are not immediately needed. Also, the air interface and BTS-to-SDU links must be capable of supporting the higher data rates necessary for packet data. However, the basic message flows and traffic paths remain the same as for voice traffic. The SDU is the first network element that treats a packet data session as fundamentally different from a voice call. This is manifest in the fact that the SDU takes the data stream to a packet network rather than to the PSTN.

Because data traffic is less tolerant of the high loss rates typical of the wireless environment, the SDU runs a retransmission protocol called the Radio Link Protocol (RLP). This is a negative acknowledgment protocol where the recipient requests retransmission of missing data frames. Frames are buffered at the sender, and a small number (usually one or two) of retransmissions of each frame are attempted before giving up by dropping the current frame and moving on to the next. Because RLP performs buffering and retransmission, it may introduce quite a bit of latency, sometimes as much as 1 or 2 s under very adverse conditions. While RLP makes use of 20 ms frames as its basic unit of (re)transmission, it presents a simple unframed octet stream interface to both sender and receiver.

To interface to an IP network, some data link layer protocol must be used to frame the octet stream presented by RLP into IP packets. For now the protocol of choice for this function is the Point-to-Point Protocol (PPP) defined by the IETF [5]. This choice was motivated by the rich functionality provided by PPP as well as its widespread implementation on many platforms. However, PPP is a quite complex protocol, and some of its features overlap with those provided by Mobile IP. As we will see, an important consideration is how to properly partition the functionality between these two protocols.

The Mobile IP foreign agent (FA) sits at the IP layer, above PPP. Currently we are only concerned with IPv4, since the timeline for IPv6 deployment is uncertain. We will use the term *packet data serving node* (PDSN) to refer to the FA and the term *packet control function* (PCF) to refer to the element in the radio access network (either the SDU or some additional processing element) that connects to the PDSN. The PCF is responsible for relaying data to and from the PDSN and for insulating the PDSN from certain wireless-specific aspects of the architecture such as the current state of the air interface. The PDSN implements features of both a network access server (NAS) and a Mobile IP FA. These elements are depicted in Fig. 2 along with other elements involved in providing Mobile IP packet data services.

In the Mobile IP protocol, packets destined for the MN arrive at a stationary home agent (HA) on the MN's home network, where they are tunneled to an FA. The HA is informed of the MN's current whereabouts when the MN attaches to the FA and sends a *registration request* message. This message is received by the FA and relayed to the HA. The HA then readies itself to forward packets to the current location of the MN and returns a registration reply message confirming the tunnel establishment. Packets destined for the MN that arrive at the HA are encapsulated in an outer IP header containing the MN's *current care-of address*, which reflects the MN's current location. This encapsulation can work in one of two ways. In the first, called a *collocated care-of address*, encapsulated packets are sent directly to the MN, which then strips the outer header and processes the inner packet as normal. In the second, called an *FA-located care-of address*, the encapsulated packets are sent to the FA, which strips the outer header and delivers the inner packet directly to the MN across the last-hop link. In a wireless environment where radio resources are scarce, it is important to support the latter mode of operation so that the extra encapsulating header is not sent over the last-hop link. This mode also conserves address space because the care-of address of the FA can be shared by all connected MNs rather than assigning a unique care-of address to each MN.

To enable a wireless carrier to provide Internet service in a roaming environment, the PDSN must be connected to an authentication, authorization, and accounting (AAA) infrastructure. Standardization of such an infrastructure is an ongoing process in the IETF. This infrastructure will enable the network visited by a roaming user to query the home network for authentication credentials and will ensure payment for services rendered. This is in addition to the existing home/visiting location register (HLR/VLR) based authentication [4] for wireless voice services. These elements are depict-

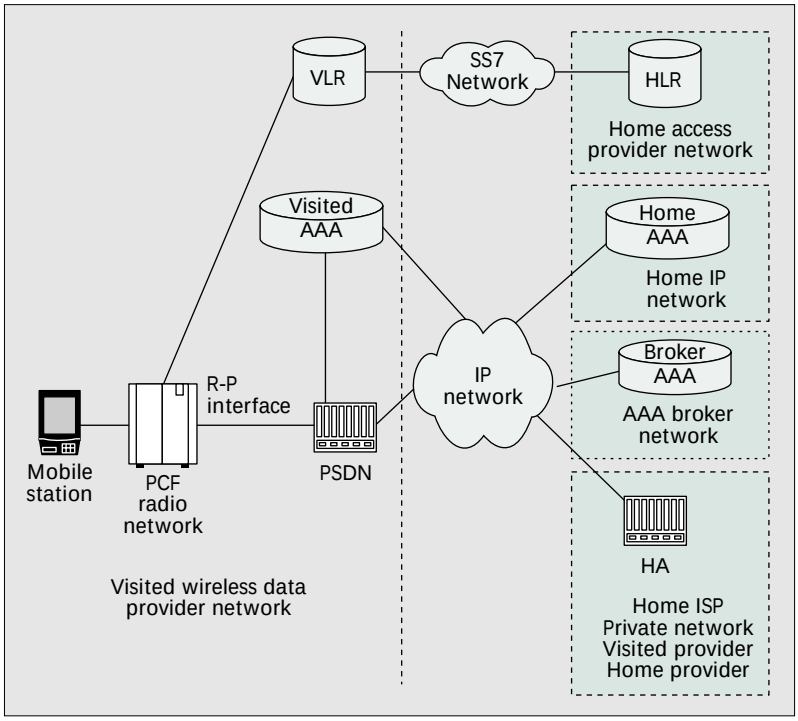

Figure 2. *The PCF, PDSN, and associated AAA infrastructure.*

ed in Fig. 2. We make use of recent extensions to the Mobile IP protocol that let us encode user credentials directly in a registration request, and to dynamically assign a home address to the MN. This essentially promotes authentication, network access control, and address configuration to the IP layer, removing it from the domain of PPP where it has traditionally been performed. By standardizing IP-layer mechanisms we can avoid dependence on any particular link-layer protocol, making the resulting system more robust against changes in the underlying technologies.

The remainder of this article is organized as follows. We first discuss the relationship of the PCF to the PDSN. Next, we compare various schemes for providing fine-grained mobility management in concert with Mobile IP. The article then discusses the requirements on an AAA infrastructure in support of Mobile IP. We present scenarios for providing remote access to private networks using Mobile IP. Finally, concluding remarks are presented.

The Foreign Agent Interface

The PCF is an abstraction for the radio access portion of the network, including at least the BTS and SDU and possibly other elements responsible for relaying data to the PDSN. The PCF and PDSN are connected by an IP network. The PDSN contains the Mobile IP FA, and for now we assume that PPP is also terminated there. From an architectural point of view, the PDSN is similar to existing NASs that provide dialup Internet access, in the sense that both terminate a link-layer protocol and are directly connected to the Internet. This is also the natural point to introduce AAA features, since this is the element that will receive credentials from mobile users.

The choice of terminating PPP in the PDSN, as opposed to earlier in the network at the PCF, was motivated partly by efficiency considerations and partly by uncertainty in the timeliness of widespread adoption of Mobile IP by MN host operating systems. First, if PPP were terminated in the PCF, a change in PCF would have necessitated reestablishment of PPP, requiring extra messages to be sent over the air. Centralizing the PPP termination point avoids disturbing the PPP state machine when a change of PCF occurs. Second, if an MN without Mobile IP client software attempts to use the system, the PCF-terminated PPP case would suffer address reconfiguration on every change of PCF, effectively tearing down any transport-layer connections that are in progress at the time. An alternative would be to transfer PPP state from one PCF to the next, but this would require additional interfaces and complexity, and would need to somehow route packets destined to the old address to the new PCF.

However, the choice to terminate PPP in the PDSN also comes at some cost. First, it impacts the scalability of the PDSN due to the large amount of state required by a PPP implementation. If PPP were terminated on the PCF, this processing could be more distributed. Second, implementation of quality of service (QoS) features is made more difficult by the centralized PPP architecture, because PPP requires that each frame be delivered in order. In contrast, multimedia services may require that each packet be treated differently according to its contents. High-priority traffic such as voice may need to be reordered with respect to lower-priority traffic on the network connecting the PCF and PDSN to avoid unnecessary queuing delays. For this reason, future versions of the CDMA architecture are expected to move the PPP termination point closer to the BTS and possibly to replace PPP altogether with a simpler link layer. However, it is a requirement of the architecture that it support a reasonable level of mobility and service transparency even for MNs that do not have Mobile IP clients and have standard implementations of PPP. This population of devices will be substantial for at least the next several years.

An alternative solution to the QoS problem is to run multiple simultaneous instances of RLP, each with different retransmission and buffering characteristics and relative priorities. High-priority traffic could use the appropriate instance of RLP. Separate instances of RLP may be required because of the potentially high latency introduced by RLP; with multiple RLPs some could be configured to perform zero retransmissions for carrying latency-sensitive multimedia traffic. All of these instances could be multiplexed under a single instance of PPP; this would be a variant of *multilink* PPP. In this case each link could use a separate connection between the PCF and PDSN. However, while some support for multiple simultaneous RLPs exists in the cdma2000 air interface standard, vendor support for this feature may not be widespread. Also, carrying multiple instances back to a computer connected via a single serial interface may present a problem.

With these considerations in mind, there are three main options for the protocol linking PCF and PDSN. Standardization of this interface will enable a service provider to purchase equipment containing the FA functionality of the PDSN from one vendor and wireless-specific functionality of the PCF from another vendor, spurring competition and hopefully lower prices. This interface is also sometimes called the R-P interface, as noted in Fig. 2, because it links the radio access network to the PDSN. The three options include:

- Existing wireless-specific interfaces
- The Layer-2 Tunneling Protocol (L2TP)
- A new Mobile IP-based L2TP

We examine each of these potential solutions in turn.

Existing Wireless-Specific Interfaces

Several interfaces exist already that could serve as the interface to a Mobile IP FA from a wireless network. These interfaces come from other standardization efforts or from vendor implementations that were never completely standardized. However, these interfaces suffer from being closed, vendor-specific, or incompletely specified, and they are usually specific to a given link-layer technology.

One such interface was developed for use between a class 5 switch and an interworking function (IWF), sometimes called the L interface. An IWF is intended to serve as the gateway to a data network. The L interface carried the octet stream from RLP over switched virtual circuit frame relay connections, and it also addressed transport of wireless-specific usage information from the switch to the IWF. This was standardized in ANSI Interim Standard (IS) 658, but the effort was not complete, and true interoperability was not achieved.

Another effort is ongoing to standardize a more advanced interface between the SDU and PCF within 3GPP2 itself. This effort is creating new messages to control IP tunnels between the SDU and PCF. While the tunnels themselves will be based on open IETF standards, the signaling messages are being created from scratch and will be CDMA-specific.

In light of these efforts it may seem superfluous to open a new IP-based interface between the PCF and PDSN. However, by choosing IP as the transport network between PCF and PDSN, we achieve independence from any given link-layer technology. An even more important consideration than adopting an IETF *protocol* may be the adoption of the IETF *process* for the standardization of this interface. Specifications should be more open, complete, simple, and available to a wider audience. In the remainder of this section we examine two possible directions for standardization of this interface within the IETF.

L2TP

The Layer-2 Tunneling Protocol was developed in the PPP working group of the IETF and recently attained RFC status [6]. This protocol was designed with the express purpose of separating an access point from a PPP termination point. L2TP defines two entities, an L2TP access concentrator (LAC) and an L2TP network server (LNS). A terminal connects to a LAC, upon which the LAC opens an L2TP connection to the appropriate LNS using a simple request-response signaling protocol.

In operation, the LAC receives data directly from a terminal and forms PPP link-layer frames. These frames are encapsulated in a special L2TP header and sent across the network to the LNS, where they are received and resequenced. In the opposite direction, the LNS receives packets destined for the terminal, creates PPP frames containing those packets, and sends them to the LAC, again encapsulated in a special L2TP header.

All processing of PPP messages and states is performed at the LNS; the LAC is simply responsible for receiving and transmitting PPP frames. However, in order to perform this task, the LAC needs some information from the LNS about the currently negotiated PPP options. Signaling messages are provided for this purpose.

L2TP provides a standardized solution for separating access points from PPP termination points, and seems to provide a good solution for the PCF-to-PDSN interface. However, as it stands now, L2TP does not support mobility (i.e., the handoff of a connection from one LAC to another), which is a requirement of our architecture. Proposed extensions to L2TP exist for handling this problem [7], but their adoption by the IETF is uncertain. Also, L2TP is designed to carry only PPP frames. As the network evolves to move the link-layer termination point closer to BTSs, L2TP would need to be replaced with some other protocol. As a consequence, 3GPP2 has decided to pursue another alternative.

Mobile IP-Based Layer-2 Tunneling

The 3GPP2 group is now considering layer-2 tunneling proposals based on Mobile IP with extensions to support a more general tunnel establishment protocol [8]. In this case, the PCF would establish a tunnel to the PDSN by sending a Mobile IP registration request to the PDSN and receiving a registration reply from the PDSN. Traffic between the two would then be encapsulated in a generic routing encapsulation (GRE) header [9]. However, unlike standard Mobile IP, the GRE tunnel would carry not IP packets, but rather link-layer data for PPP.

Mobility is achieved by adding the International Mobile Station Identifier (IMSI) to the registration request, which is a unique parameter associated with each MN. When the PDSN sees a registration request from a new PCF with the same IMSI, it knows to associate the new connection with the previous one from the old PCF. The state of the PPP running above the R-P tunnel will be unaffected by the handoff, making it transparent to the PPP implementation on the MN. This provides a layer of transparent mobility; the MN may move from one PCF to another while retaining the connection to the original PDSN. However, this layer is only intended for use within a single domain where the PCFs and PDSNs are on a single private network and have direct security associations. A change of PDSN will take place if the MN moves too far from its original point of attachment. Such a change will be visible to the MN, and would require a new Mobile IP registration with the HA. Sometimes, an MN will move back to a previously visited PCF while dormant and will not have an immediate indication that it has moved. In this case the PCF will signal the PDSN to send a new agent advertisement, which will trigger reregistration if needed.

To keep the interface simple, and in contrast to L2TP, no attempt will be made to transmit the currently negotiated PPP options from the PDSN to the PCF. Therefore, the data transmitted to the PDSN within each GRE packet will be in a raw format, consisting of exactly the bytes received by the PCF from the MN. Similarly, data transmitted to the PCF will be sent immediately to the MN after stripping out the GRE headers and without any further processing.

In addition to carrying user data, the PCF must transmit wireless-specific usage information to the PDSN so that accounting can be performed. The PDSN will merge these records with other data-specific usage information and transmit them to the accounting infrastructure. We will discuss the overall accounting architecture later.

Micro-Mobility

Several proposals exist for solving the so-called *micro-mobility* problem. Micro-mobility refers to situations where the MN changes its point of attachment to the network so frequently that basic Mobile IP tunnel establishment introduces significant network overhead in terms of increased signaling messages. Another oft-cited problem is the latency of establishing each new tunnel, which introduces delays or gaps during which user data is unavailable. This delay is inherent in the round-trip incurred by Mobile IP as the registration request is sent to the HA and the response sent back to the FA.

Micro-mobility proposals can be classified into two distinct categories. The first are the *hierarchical tunnel* approaches, characterized by their reliance on a tree-like structure of FAs. Encapsulated traffic from the HA is delivered to a root FA, which decapsulates it and reencapsulates it for one of its children FAs based on the MN's current point of attachment. As the MN moves from leaf to leaf in the tree, location updates are made to the least common ancestor node so that traffic is always tunneled properly. By keeping most location updates within the local access network, the cross-Internet signaling and latency are reduced substantially. These proposals sometimes require the MN to send new types of messages or to be otherwise aware that a hierarchical tunneling protocol is in use. Examples include the Regional Tunnel Management proposal [10].

The second distinct category consists of the *routing update* approaches, which attempt to avoid the overhead introduced by decapsulation and reencapsulation of traffic at each FA. In these proposals, IP routing is used to direct traffic toward the MN, and new signaling messages are introduced to update host-specific routes within the access network. These proposals usually require the use of a collocated care-of address, which will be used as the basis for routing packets, and often require the MN to send new types of route update messages or to be otherwise aware that a routing update protocol is in use. Examples include Cellular IP [11] and HAWAII [12].

Micro-mobility protocols are not yet needed in our network architecture. First, there are already several layers of mobility management underneath our FA; this situation is depicted in Fig. 3.

At one level, there is BTS-to-BTS handoff, keeping the same SDU/PCF. This is accomplished purely with wireless-specific link-layer protocols and is invisible to the PDSN. Such soft handoff is already part of voice service. The area covered by a single SDU may be quite large, as much as a large portion of a metropolitan area in some cases. At another level, our layer-2 tunneling scheme allows for transparent handoff from PCF to PCF, keeping the same PDSN. Such a handoff is

invisible to the PPP and FA protocols on the PDSN; this is dictated by our requirement to give reasonable levels of service to MNs that do not have Mobile IP software. Additionally, this handoff mechanism alleviates the need for cross-domain Mobile IP reregistration, authentication, and security negotiation. As MNs become more tolerant of link-layer disconnection and reconnection, our Mobile IP-based layer-2 tunneling protocol can evolve to become a layer-3 tunneling protocol, essentially becoming one of the hierarchical tunneling approaches described above, but with an additional measure of transparency.

Authentication, Authorization, and Accounting

Mobile IP as specified by RFC 2002 [2] does not provide scalable mechanisms for access control or accounting. While basic Mobile IP does specify extensions that can be used to authenticate an MN to an FA or an FA to an HA, these extensions are not mandatory and assume the existence of preconfigured shared secrets between these entities. In the CDMA cellular architecture, there will be many networks or domains in a public wireless data deployment. For example, each carrier that supports PDSNs represents a different domain. Home agents likewise will reside in many domains, such as wireless carriers, Internet service providers (ISPs), or private networks. Networks that support FAs (or provide the simpler Internet access service that does not use Mobile IP) will expect payment for wireless data services from the user or the user's home domain. To obtain assurances of payment, the cellular wireless data architecture must support scalable AAA services. The architecture realizes AAA services via AAA servers; taken as a whole, the collection of AAA servers in the network is sometimes referred to as the *AAA infrastructure*. The AAA infrastructure verifies user credentials and provides a service policy to the serving network for which the user is authorized. The *AAA infrastructure* also may provide reconciliation of charges between serving and home domains.

In the CDMA cellular network architecture, a user will have a home wireless carrier to which the user subscribes for wireless services. That home network will hold user profile and authentication information. When the user roams into the territory of a different wireless carrier (called the *serving network*), that carrier accesses the home wireless carrier for authentication information associated with the user as well as a service profile. The service profile indicates which radio resources the user is authorized to use, such as a maximum bandwidth or access priority. The intersystem signaling protocol between wireless carriers is ANSI-41 [4], which runs over the Signaling System No. 7 (SS7) network. ANSI-41 supports authentication as well as more general exchange of user parameters. In ANSI-41 the user profile is stored in an HLR owned by the home network and is temporarily retrieved into a VLR owned by the serving network. The home and serving networks must have a service roaming agreement in place.

In addition to a home wireless carrier, a user will have a home data (IP) network. The home IP network similarly holds user profile and authentication information. The home network could be another wireless carrier, an ISP, or a private network. This flexibility is especially desirable because the security information necessary to authenticate and authorize the user remains solely in the possession of the private network or home ISP and does not have to be exposed to or configured into wireless carrier networks. The user profile returned from the home network to the serving network will indicate policy associated with data-related features such as the need for virtual private network (VPN) services. The CDMA cellular network architecture will use IETF protocols for AAA. Examples include RADIUS, DIAMETER, and newer protocols being developed by the IETF's AAA Working Group. The ANSI-41 and IETF AAA servers are depicted in Fig. 2 along with their relationship to the PCF and PDSN.

The CDMA wireless data architecture therefore has two levels of authentication. One occurs in the wireless network, and the other in the data network. For the wireless network, the mobile identifies itself via an IMSI, and AAA functionality occurs via ANSI-41 and location registers. This authenticates the mobile device. For the data network, the mobile identifies itself via a network access identifier (NAI), which is of the form `user@homedomain`. Authentication of the user is based on a challenge from the PDSN. These mechanisms authenticate the user, not the terminal, which is important for supporting users who have only a transitory relationship with devices such as rental phones or walkup kiosks.

The data network AAA supports two mechanisms to authenticate a user, depending on the service the user requests. For simple Internet access service that does not use Mobile IP, the authentication protocol is the Challenge Handshake Authentication Protocol (CHAP) and is a part of PPP establishment. In CHAP, the PDSN challenges the user with a random value to which the MN must respond with a signature (e.g., using Message Digest 5, MD5, with a secret random value) and user NAI. The signature is verified by the home network. For Mobile IP, the FA sends a similar challenge with a random value to the MN [13]. Again, the mobile must respond to the challenge with a signature and NAI [14] that is verified by the home network,

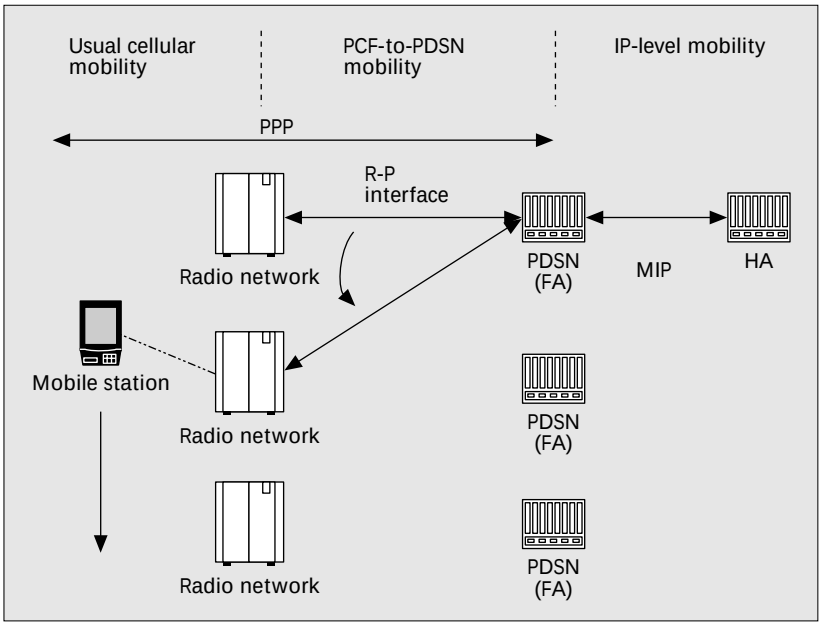

Figure 3. *Levels of mobility in a CDMA network.*

but this time the response is sent along with the Mobile IP registration request rather than during PPP establishment. Both of these mechanisms rely on shared secrets associated with the NAI which are stored in the home network, and both will be supported by the same AAA infrastructure.

Home and serving network AAA servers may have a direct bilateral relationship. The cellular architecture, however, will involve thousands of domains, because many of the domains are private networks owned by enterprises that seek wireless data service for their mobile work force. If the number of domains is small, the serving and home networks could have preexisting relationships (e.g., perhaps secured via IP Security, IPSec [15], associations). However, this would not be a scalable solution: it would require too many pairwise relationships. For scalability, the concept of a broker is introduced, as shown in Fig. 2. Brokers possess directories that allow AAA requests to be forwarded based on the NAI to home networks or to other brokers that may know the whereabouts of the home network. Brokers may also take on a financial role in settlement of accounts between domains and may process accounting records for the usage requests they authorize.

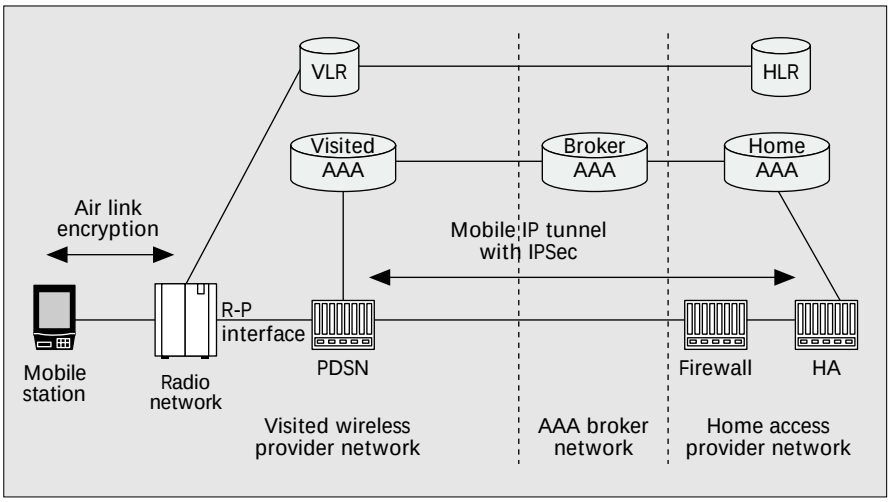

Figure 4. *Protection of user data by a concatenation of security mechanisms.*

Because the serving network will not provide service unless it is able to obtain authorization from the mobile user's home network, or from a broker that accepts financial responsibility, the AAA infrastructure needs to possess a reasonable level of reliability. This implies that servers must retransmit requests and switch over to backup servers when a failure is detected. It is important that this be done properly so that unnecessary outages are avoided. The new AAA protocols being developed by the IETF should address these problems.

Virtual Private Networking

Wireless carriers expect that one of the main uses of cellular wireless data will be remote access to private networks by a mobile work force. For example, a business traveler in a distant city may need to access Web and e-mail servers on his corporation's intranet which are protected by a firewall. From a wireless service provider's point of view, by far the simplest solution is to give the user mobile access to the public Internet and allow the MN to employ whatever techniques and security software is necessary to traverse the firewall and access the private network. This style of access is usually referred to as *voluntary tunneling* because it relies on the MN to voluntarily open some form of communication channel back to the private network and to be responsible for all aspects of security. For example, an MN could open an IPSec [15] tunnel back to the private network using prearranged security keys. All data to and from the private network would be encapsulated in the tunnel.

While voluntary tunneling provides a clean and secure end-to-end solution for access to private networks, greater efficiency and transparency can be achieved with some cooperation on the part of the wireless service provider. Voluntary tunneling leads to an extra layer of encapsulation over the last-hop wireless link, which consumes scarce and expensive radio resources. Also, complex encryption and decryption algorithms may not be suitable for implementation in small low-power wireless devices that nevertheless want access to private networks.

By relying on a sequence of concatenated protection mechanisms, it is possible to provide secure remote access to mobile users without requiring the extra tunnel overhead on the radio link or the implementation of computationally intense encryption algorithms on the MNs. Figure 4 depicts such a scenario.

Third-generation wireless physical layers will provide their own encryption of the RLP data as it is sent from MN to SDU. This transmission leg is shown on the left side of Fig. 4. Such encryption is designed especially for the wireless environment and will be less computationally intense than algorithms that are being developed for the Internet at large. Key distribution and identity verification will also be less computationally intense; these factors affect the latency with which a connection can be established.

Such wireless-specific encryption mechanisms only protect the data from the MN to the wireless network and PCF. The R-P interface from the PCF to the PDSN is expected to be over a secure private network; if this is not the case, IPSec should be added to the R-P tunnel. From the PDSN to the HA, protection is definitely needed because this leg will traverse the public Internet. This will most likely be based on IPSec, and will include mechanisms for distributing keys such as the Internet Key Exchange [16]. Verification of identity may leverage the AAA infrastructure described earlier to distribute some initial keying material; this will simplify the process and avoid reliance on a possibly slow public key distribution and revocation infrastructure. It is important for the HA to verify the identity of the PDSN because this node will have access to unprotected user data. It is important for the PDSN to verify the identity of the HA so that user traffic is not misdirected to an insecure location. Note that the HA may also wish to verify the identity of the PDSN even if it is in a wireless carrier network providing public Internet access; this could be used to prevent access to the HA from carriers that have no business relationship with the home wireless carrier. The same mechanisms outlined here can be used for this purpose.

In the private network access scenario it is expected that the HA will actually be owned and operated by the private network to which the user is gaining access. As such it will manage both security and mobility of the user, and will form security associations with PDSNs dynamically as the user moves from one to another. By relying on IETF standards such as Mobile IP and IPSec, such HAs should be easily available to private networks from a wide range of manufacturers.

The security provided by the concatenation of mechanisms

described here may not be adequate for all users because plain text traffic is exposed to the visited carrier network. If there are insecure links in this network, especially unencrypted radio links in the backhaul network, this could present problems. However, the additional performance benefits may outweigh the risks for many users, especially if the visited carrier is trusted and protected from outside attack. This level of security is at least as strong as that offered by the PSTN today, which is used by many corporations to carry unencrypted private traffic.

When an MN connects to a private network it may be assigned a private address. Because such addresses are not allocated by any global authority, they may not be globally routable or even unique. It is important to design the PDSN so that it can properly handle such traffic and use reverse tunneling [17] where appropriate. For instance, the PDSN must make use of the HA address and link-layer identification information to resolve potential collisions in the IP addresses assigned to different MNs. The ability to support private addresses facilitates acceptance of CDMA wireless data services by private network customers.

Conclusions

This article gives an overview of the issues we have encountered in standardizing a solution for wireless Internet access in a CDMA environment. Important considerations involve mobility management; authentication, authorization, and accounting of users; and access to virtual private networks. We present options for interfacing a Mobile IP foreign agent to the cellular network; the most likely choice for this interface is a mobile layer-2 tunneling scheme based on Mobile IP. The Mobile IP and AAA infrastructure will support integration with private networks and other ISPs by adopting IETF protocols, including IP Security where appropriate.

As the cost of wireless airtime continues to fall, Internet access is expected to become an important service for wireless carriers to offer consumers. It is vital that providers are ready to connect customers to an industry standard network architecture that provides the mobility and security features that users need. By relying on IETF protocols whenever possible, such an infrastructure can be delivered in a timely and competitive fashion.

As more and more services become IP-based, we envision the telephony-specific part of the cellular network becoming less and less important. Eventually the entire network infrastructure, including the part dedicated to voice service, could be replaced with an IP-centric solution. However, this eventuality is still some years in the future. For now it is important to coexist with the access network provided for voice and for voice service to interwork with the PSTN and existing second-generation cellular authentication mechanisms. The architecture presented here provides for this coexistence as well as evolution into the future.

Acknowledgments

Thanks to Lynne Sinclair for providing valuable feedback and discussion.

References

[1] C. J. Mathias, "Internet Access with No Strings Attached," *Bus. Commun. Rev.*, vol. 28, no. 6, June 1998, pp. 57–60.
[2] C. E. Perkins, "IP Mobility Support," IETF RFC 2002, Oct. 1996.
[3] Electronic Industries Alliance, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular Systems," EIA TIA/IS-95-A, 1995.
[4] M. D. Gallagher, *Mobile Telecommunications Networking with IS-41*, New York: McGraw-Hill, 1997.
[5] W. Simpson, "The Point-to-Point Protocol (PPP)," IETF RFC 1661, July 1996.
[6] W. Townsley *et al.*, "Layer Two Tunneling Protocol "L2TP"," IETF RFC 2661, Aug. 1999.
[7] M. Chuah *et al.*, "Mobile PPP," draft-ietf-pppext-mppp-01.txt, June 1999; work in progress.
[8] Y. Xu *et al.*, "Mobile IP Based Micro Mobility Management Protocol in the Third Generation Wireless Network," draft-ietf-mobileip-3gwireless-ext-04.txt, June 2000; work in progress.
[9] S. Hanks *et al.*, "Generic Routing Encapsulation (GRE)," IETF RFC 1701, Oct. 1994.
[10] E. Gustafsson, A. Jonsson, and C. E. Perkins, "Mobile IP Regional Registration," draft-ietf-mobileip-reg-tunnel-02.txt, Mar. 2000; work in progress.
[11] A. G.Valko, "Cellular IP: A New Approach to Internet Host Mobility," *Comp. Commun. Rev.*, vol. 29, no. 1, Jan. 1999, pp. 50–65.
[12] R. Ramjee *et al.*, "HAWAII: A Domain-based Approach for Supporting Mobility in Wide-Area Wireless Networks," *Int'l. Conf. Network Protocols*, 1999.
[13] C. E. Perkins and P. Calhoun, "Mobile IP Challenge/Response Extensions," draft-ietf-mobileip-challenge-12.txt, June 2000; work in progress.
[14] P. Calhoun and C. E. Perkins, "Mobile IP Network Access Identifier Extension for IPv4," IETF RFC 2794, Mar. 2000.
[15] S. Kent, and R. Atkinson, "Security Architecture for the Internet Protocol," IETF RFC 2401, Nov. 1998.
[16] D. Harkins and D. Carrel, "The Internet Key Exchange," IETF RFC 2409, Nov. 1998.
[17] G. Montenegro, "Reverse Tunneling for Mobile IP," IETF RFC 2344, May 1998.

Biographies

PETER J. MCCANN (mccap@research.bell-labs.com) received his B.S. degree in engineering and applied science from the California Institute of Technology in 1993, and M.S. and D.Sc. degrees in computer science from Washington University, St. Louis, Missouri, in 1995 and 1997, respectively. He is currently a member of technical staff at Bell Laboratories. His research interests include formal reasoning about concurrent mobile systems, and the design and architecture of networking protocols, including protocols for third-generation wireless networks. He is a member of the ACM and the IEEE Computer Society.

TOM HILLER (tom.hiller@lucent.com) works on third-generation wireless data standards and architecture at Lucent Technologies. For the last two years, he has been a principle contributor and editor of cdma2000 wireless data standards using Mobile IP and AAA. He is now focused on evolving the cdma2000 architecture to support wireless voice-over-IP services, as well as the creation of an all-IP cellular network. His prior work includes ATM, IP over ATM, and ISDN architectures and development within Lucent. He is a Distinguished Member of Technical Staff at Lucent Technologies, has a B.S.E.E./M.S.E.E. from the University of Illinois, and is a member of the ACM.