Real-Time Payments for Mobile IP

Hitesh Tewari and Donal O'Mahony, Trinity College Dublin

ABSTRACT

The Mobile IP protocol has evolved from providing mobility support for portable computers to support for wireless handheld devices with high mobility patterns. A new category of micromobility protocols has been proposed to deal with the increased signaling loads that will be generated with large populations of such devices on a network. We argue that the authentication schemes presently employed in these networks do not scale well for large numbers of nodes, and that the lack of accounting procedures prevents the mass deployment of these networks. We envisage that future access networks will be operated by independent service providers, who will charge users for access to services in the fixed network but may not have long-term contractual relationships with them. These access networks may also employ a variety of micromobility protocols for fast handover support. We present a scheme based on hash chains, which allows for fast authentication of datagrams for secure updating of router entries within the access network, and real-time accounting of network usage by mobile nodes. Such a system will alleviate problems of fraud in mobile networks and eliminate the need for interoperator billing agreements.

INTRODUCTION

The Mobile IP protocol [1] was designed with a view of supporting mobile hosts with relatively slow mobility patterns. In Mobile-IP-based networks, users have an enduring relationship with a network operator, and interoperator roaming agreements are required to allow mobile users to access resources in third-party networks. Payment for usage of network resources may take place after the services are used, and call detail records (CDRs) must be exchanged by the operators for billing purposes. In the future, when the number of mobile users and network operators becomes large, these existing approaches to authentication and billing in Mobile IP networks will not be sufficient.

The mobility of small handheld communicators such as PDAs, on the other hand, is much greater than that of traditional laptop computers. Research has been carried out to support fast handovers in situations where a mobile node moves rapidly between base stations under the control of a single administrative domain. Such techniques are classified as micromobility protocols. Signaling messages in these networks have the potential to create or modify routing table entries within the routers in the network, and therefore must be authenticated prior to any changes to the tables. Failure to implement proper security procedures can allow malicious nodes to masquerade as legitimate hosts and carry out denial-of-service attacks.

The current approach to providing secure and trusted communications in Mobile IP networks is based on authentication, authorization, and accounting (AAA) protocols [2], which are currently under development in the Internet Engineering Task Force (IETF). In this article, we outline a lightweight scheme to address the security issues in Mobile-IP-based next-generation wireless networks. Our proposal allows a user to arrive in a new access network and avail themselves of network services by paying the service provider in real time. The routing nodes within the access network are able to efficiently verify the authentication and accounting information carried in the datagrams during a call. In the remainder of this article, we give a brief overview of Mobile IP and some of the well-known micromobility architectures that have been proposed. We then discuss the current approaches to security in Mobile-IP-based networks, followed by an ideal set of mobility requirements for next-generation networks. We provide details of our proposed protocol and finally present some concluding remarks.

MOBILITY SUPPORT IN IP NETWORKS

The IETF Mobile IP protocol is a well-known approach for mobility support in IP networks. It allows a mobile node (MN) to maintain its existing transport layer connections when it moves from one subnetwork to another. Mobile IP is a network-layer protocol that is completely transparent to the higher layers, and does not require any changes to the existing Internet hosts or routers. In normal IP routing, packets are routed

0163-6804/03/$17.00 © 2003 IEEE

from a source to a destination on a hop-by-hop basis, where a router makes a routing decision based on the network part of the address. Thus, an IP address is used for routing and also specifies a point of attachment for a node on the Internet. For a host to continue to receive packets while it is roaming in a foreign network, it needs to keep its IP address. Mobile IP solves this problem by adopting a two-tier addressing approach. In a Mobile IP network, a roaming MN has two network addresses. The first is known as the home address, and remains unchanged regardless of where the node is attached on the Internet. The second is known as the care-of address (CoA) and changes at each new point of attachment in the network. An MN using its home address appears to be able to receive packets in a foreign network through a node in the home network known as the home agent (HA).

When a host roams into a foreign network it obtains a new CoA, which it registers with its HA. Registration messages must be authenticated by the HA, prior to any updates to the CoA for an MN. Unauthenticated signaling messages can allow a malicious host to trick an MN's HA into adding a false CoA for the node, and result in datagrams being redirected to an unknown network by the HA. In Mobile IPv4, the router in the visited network that cooperates with the HA to deliver datagrams to an MN is known as the foreign agent (FA). When the HA receives datagrams for a roaming MN, it tunnels them to the node using the CoA as the destination IP address. The FA detunnels the packets to reveal the MN's original home address and forwards them to the node. Since the packets arrive at the MN with their home address as the destination, they are processed correctly by the upper layers.

Mobile IP solves the *macromobility* problem in wide area networks by updating the CoA at the HA whenever a mobile host changes its point of attachment on the Internet. Rapid proliferation in the number of wireless devices and networks has led to considerable research into *micromobility* protocols in both academia and industry. The main aim of these protocols is to minimize the number of signaling messages that have to be carried over the core network to support host mobility. Micromobility architectures support mobility within a local network or domain. With Mobile IP, any change in a MN's point of attachment results in a signaling message being sent to the home network, even though most of the path between the MN and the HA remains unchanged. On the other hand, with micromobility protocols, local handoffs result in signaling messages being generated that are limited to the administrative domain in which the MN is currently roaming. The signaling messages are intercepted by a gateway node or router, which usually also acts as the CoA for the MN. As long as an MN remains within the administrative domain of a single operator or service provider, there is no need to update its route entry in the home network. Only when an MN moves between two separate networks does one need to employ the Mobile IP protocol.

CELLULAR IP

Cellular IP [3] is a micromobility protocol based on cellular telephony concepts such as passive connectivity, paging, and support for fast intradomain handover for mobile nodes. A Cellular IP (CIP) network consists of a number of interconnected routers and radio endpoints. The leaf nodes are usually the base stations, and the point of attachment to the Internet is a host known as the gateway. The gateway node embodies both the HA and FA functionality, and is responsible for filtering out all signaling messages that are specific to the CIP network. Data packets transmitted by an MN are routed toward the gateway on a hop-by-hop basis. Each intermediate CIP node creates a mapping of the address of the MN and the neighbor that forwarded the packet in its route cache. As long as the MN has data to send, the CIP nodes along the path to the gateway keep an up-to-date mapping for the MNs point of attachment on the network. In cases where an MN does not have any data to send but wishes to maintain a valid route cache entry, it periodically sends a *route update* message toward the gateway node. Page and route update messages can be used to create entries within the Cellular IP caches, which can result in changes to the routing of packets within the network. A CIP node must therefore authenticate all signaling messages prior to any cache modifications. Cellular IP employs a fast session-key management scheme that allows authentication of control packets by CIP nodes by making use of a *shared network key*. This aids fast handoffs within the access network but also has some drawbacks. All CIP nodes have knowledge of the network key, and repeated use of the same can lead to a decrease in the effective security of the key. Data packets can refresh a cache entry but are not required to be authenticated by a CIP node. A malicious node can inject data packets into the network and keep cache entries alive to disrupt traffic flows. There are no mechanisms in place to account for network usage within a CIP network. Once a node stops transmitting signaling or data packets, the cache entries expire and are deleted by a CIP node. All records of the MN having ever been present on the network are subsequently lost.

HIERARCHICAL MOBILE IP

The Hierarchical Mobile IP (HMIP) protocol [4] makes use of a hierarchy of FAs, which allow an MN to register locally within a visited domain. This reduces the number of signaling messages that have to be sent to the home network and also improves handover performance, since it minimizes the latency associated with updating the CoA with an HA in a possibly distant network. When an MN first arrives in a new domain, it performs a registration with its home network and registers the address of the gateway FA (GFA) as its CoA. When the MN moves between FAs under the same GFA, localized or regional updates take place while the CoA remains the same. Each intermediate FA maintains a visitor list of foreign MNs that are currently roaming within the network. During

The HMIP protocol makes use of a hierarchy of foreign agents, which allow a mobile node to register locally within a visited domain. This reduces the number of signaling messages that have to be sent to the home network and also improves handover performance.

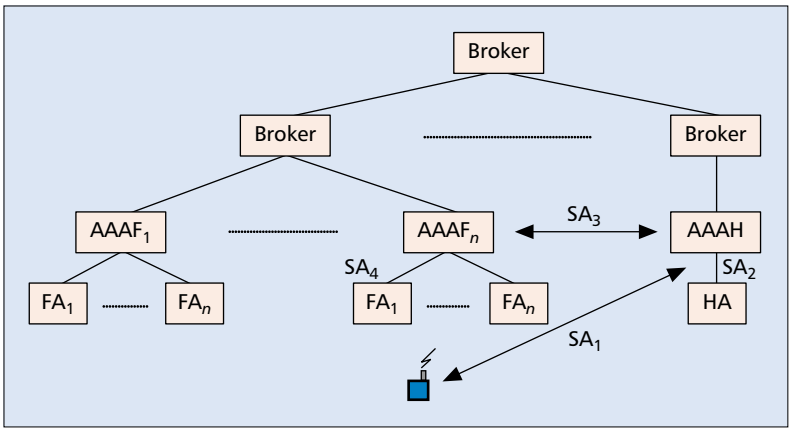

Figure 1. *The AAA trust model for Mobile IP.*

handover, regional registration messages travel only as far as the *crossover router* or *switching FA*, while the remainder of the path to the gateway remains the same. All intermediate nodes must authenticate any registration messages they receive. A key may be distributed to the MN and to the domain in which it is currently roaming. The key is used to prove the authenticity of registration messages generated by the MN. An MN has a preconfigured secret with its HA, while the FAs within a domain may share security associations.

HAWAII

HAWAII [5] is a micromobility proposal from Lucent Technologies that provides efficient intradomain mobility. The HAWAII scheme divides the network infrastructure into a number of hierarchically arranged domains. A domain is connected to the Internet via a *domain root router*, and may consist of several routers and base stations. Each MN in a HAWAII network is assigned a collocated CoA and a home domain. An MN retains its CoA as long as it remains within a domain, and therefore the HA does not need to be contacted. The MN sends a registration message to its nearest base station, which adds a forwarding entry for the node. The base station initiates a HAWAII signaling message toward the domain root router. Intermediate routers also add a forwarding entry to their caches for the MN, which is used during the reverse path for delivering datagrams to the MN. Location management information is created or updated by explicit signaling messages by MNs in a HAWAII network. A HAWAII node must be able to verify the authenticity of signaling messages originating from an MN, and must be able to generate authenticated replies that the MN can also verify. To achieve this goal the protocol designers suggest that three separate security associations be in place: the first between base stations and the MN, the second between the base stations and the HA, and the third between the HA and the MN.

There are a number of similarities in the design and operation of the above mentioned micromobility protocols [6]. We also observe that in each of the above protocols there are certain key nodes or routers that maintain a route cache, which holds soft-state mappings of the current location of mobiles within the access network. Regardless of how the cache is structured, a router must authenticate all signaling messages it receives from a MN prior to any cache modifications. Second, an MN usually has a security association with its home network and must establish a temporary security association in the access network. This may involve the use of a trusted key-distribution server or a dialog with the home network. To aid fast handoffs and maintain quality of service (QoS), the requirement to contact a remote entity must be kept to a minimum. Finally, there is a gateway or domain router, which connects the access network to a core network such as the Internet. This node can be used to collate accounting information of network usage by roaming MNs.

AAA and Security

The AAA working group within the IETF has been working on the definition of a general AAA infrastructure for network access. Recently the work has been adapted to support the Mobile IP protocol, to allow for authentication and collection of accounting information of network usage by MNs [7]. Figure 1 shows the entities involved in authenticating and registering an MN in a foreign network using the AAA infrastructure. A foreign domain contains one or more AAA servers (AAAF) and multiple FAs. The FAs interact with an MN to authenticate its credentials. An FA has a security association with its local AAA server, which in turn may have further security associations with other AAA servers. If the AAAF cannot verify the credentials of an MN, it can contact the MN's home AAA server (AAAH) with whom it must share a security association.

A security association at a very minimum consists of a shared secret between two entities. In the AAA trust model for Mobile IP, an MN shares a security association SA_1 with the AAA server in its home domain. The AAAH in turn shares a security association SA_2 with the HA. It is also necessary for the AAAH and AAAF to share a security association SA_3 so that the AAA server in the foreign domain can verify the credentials of a roaming MN. Finally, the FA must share a security association SA_4 with the AAAF in order to allocate local resources to an MN. For scalability reasons the concept of *brokers* (AAAB) is employed, which means that a foreign domain does not need to keep security associations with every possible home domain. The use of brokers in the system requires that the two administrative domains have security associations with the broker. The broker then becomes privy to security exchanges between the two domains and also has to be trusted.

Once an MN has been authenticated, three session keys are generated by the AAAH. Each session key that is generated by the AAAH is generally distributed to two entities. The method by which the key is encoded depends on the security association between the entities. The mobile-home key $K_{MN,HA}$ is shared between the

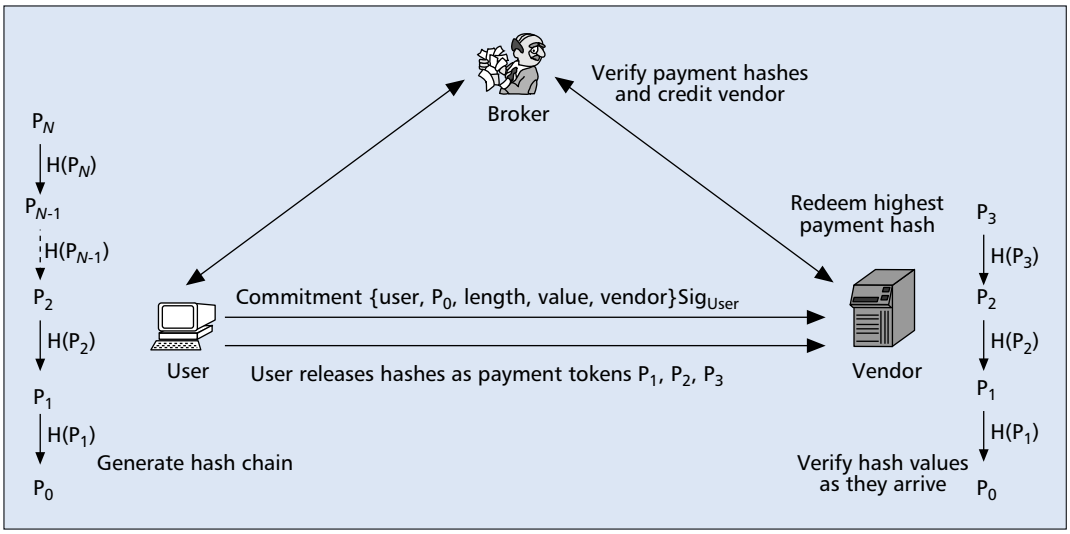

Figure 2. *Micropayments using hash chains.*

MN and the HA. It is encrypted using security association SA_2 for the HA and SA_1 for the MN. For MNs currently roaming in a foreign network, this key has to be transported via the AAAF and the serving FA in the foreign network. The mobile-foreign key $K_{MN,FA}$ is shared between the MN and the FA. It is encrypted using SA_3 for the FA and SA_1 for the MN. The AAAF forwards the key to the correct FA using security association SA_4. Finally, the foreign-home key $K_{FA,HA}$ is shared between the FA and the HA. It is encrypted using SA_3 for the FA and SA_2 for the HA. Once the session keys have been distributed, there is no need to invoke the AAA protocols until the keys expire. During intradomain handover the new FA will contact the AAAF and obtain the session keys $K_{MN,FA}$ and $K_{FA,HA}$, which were previously assigned to the old FA.

From the above discussion it becomes increasingly clear that the overheads incurred in setting up the AAA security associations and the transferring of cryptographic material can be quite substantial. In micromobility environments, where an MN may change its point of attachment within the access network frequently, this may lead to degradation in the QoS since routers in the new path will need to be informed of the existing security associations prior to making any routing decisions. In addition, interdomain handover requires that the MN obtain a new set of sessions keys from the AAAH server and have them distributed to all entities concerned. We now present a lightweight cryptographic technique that allows us to address some of these problems.

Micropayment Technology

Numerous methods for electronically paying for goods or services across a network have been proposed over a number of years. These systems are usually electronic equivalents of physical payment methods such as cash, checks or credit cards [8]. Each of these *macropayment* instruments have a minimum transaction overhead, since they require an online connection to a bank or trusted third party (TTP) to verify the authenticity of the payment tokens. In addition, most systems make use of computationally expensive cryptographic operations such as public key cryptography. These two factors combined make the use of macropayment systems prohibitive for payments of a few cents.

In contrast, micropayments are a family of payment systems designed to allow repeated small valued payments (e.g., 1/10 of a cent in a single transaction). Micropayment research has concentrated on repeated payments at a single vendor, and many of the schemes make use of one-way functions [9] to generate a chain of hash values. A one-way function is one where it is easy to compute $y = f(x)$ but computationally expensive to reverse the transaction. Hash functions such as MD5 or SHA are computationally less expensive than symmetric key algorithms such as AES and asymmetric key algorithms such as RSA.

Micropayments Using Hash Chains

A user generates a *hash chain* of length N by applying a hash function N times to a random value P_N the *root* of the hash chain, to obtain a final hash P_0, the *anchor* of the hash chain. The user commits to the chain by digitally signing the anchor with his/her private key. For each payment, the user releases the *pre-image* of the last hash value. For example, for the first payment the user releases the hash value P_1. The receiver can apply the same hash function to the value P_1 to obtain the anchor P_0. Since the hash function is one-way, only the user could have generated the hash value.

Figure 2 shows the overall payment process where a user generates a hash chain of length N. The user *commits* to the anchor of the chain P_0, the length of the chain, the value of each hash, and the vendor at which he/she wishes to spend the chain. Prior to payment the user forwards the commitment to the vendor, who can verify its authenticity offline. For each micropayment the user releases the next payment hash in the chain. The vendor can redeem the hashes at the broker with whom the user has an account at a

Micropayments are a family of payment systems designed to allow repeated small valued payments. Micropayment research has concentrated on repeated payments at a single vendor, and many of the schemes make use of one-way functions to generate a chain of hash values.

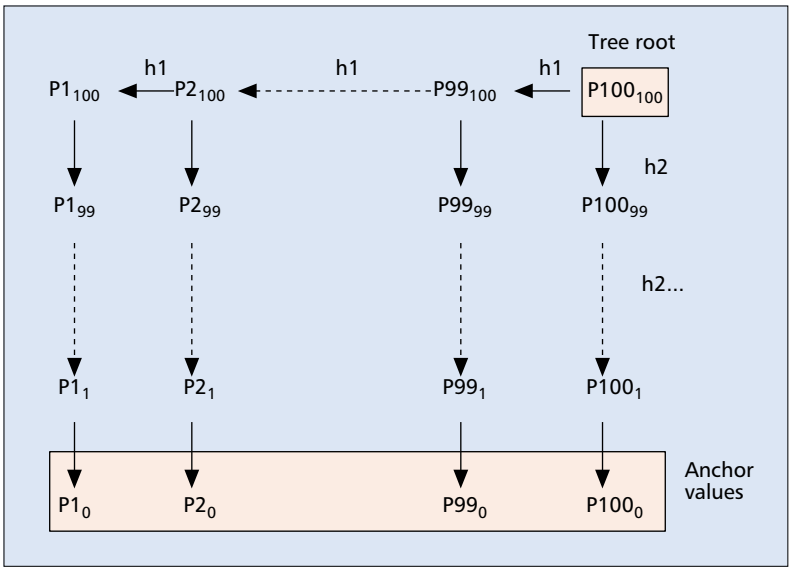

Figure 3. *UOBT generation.*

later date by presenting the highest payment hash along with the signed commitment.

Unbalanced One-Way Binary Tree (UOBT)

The UOBT scheme is an efficient hash chain scheme where the root of each chain is derived from another hash chain [10]. The scheme is ideal when a large number of hashes are needed and the device has limited storage capabilities. Only the *tree root* value has to be stored on the device to be able to reconstruct the entire UOBT. Figure 3 shows an example UOBT where $P100_{100}$ is the tree root.

We repeatedly apply the hash function *h1* (e.g., SHA) to this value to obtain the *backbone hash chain* $\{P100_{100} \ldots P1_{100}\}$. Each of these hash values is used as the *secret root* value for deriving the individual subchains by applying the hash function *h2* (e.g., MD5). For example, the value $P2_{100}$ is the root of the P2 chain, which consists of the values $P2_{100}$, $P2_{99}$, and so on until the anchor of that chain $P2_0$ is reached. The result is a UOBT with a backbone chain length of 100, with each subchain also consisting of 100 hash values and an overall hash chain tree consisting of 10,000 hashes. The signed commitment consists of a hash of the concatenation of the anchors of each of the subchains $\{P1_0, P2_0 \ldots P99_0, P100_0\}$ signed with the private key of the user or broker.

On small devices with limited storage such as a PDA, it may not be possible to store all the hash values of a long hash chain. If a UOBT is used with a backbone chain length equal to the length of each subchain, it can be shown that the average computational overhead to compute the next hash is $n^{1/2} - 1$, where n is the number of values in the UOBT, and $n^{1/2}$ is the square root of n. Thus, a 100×100 UOBT requires 99 hashes on average to compute a hash value. There is an initial overhead in transporting the set of anchors to the broker for the commitment to be signed. However, once the initial exchange has taken place, the number of cryptographic operations performed during the communications session are greatly reduced.

Protocol Overview

We envisage that in the future, there will be a large number of independent network operators and service providers, who will provide users with wireless access to other fixed or mobile nodes on the Internet. These access networks will have dedicated high-speed connections via a gateway (GW) node into the core IP network. Small or medium-sized networks may consist of a number of radio cells and may support micromobility protocols for fast handover within the access network. Each user in the system will be uniquely identifiable by his/her network access identifier [11] which is of the form user@realm. A user who arrives in a new access network can immediately start accessing services in the core network once he/she has registered with the service provider and bought provider-specific payment tokens from a trusted broker (BK). In order to be able to receive incoming calls, a user must register his/her CoA with his/her preferred location management server (LS). The LS keeps a mapping of the user's network access identifier (NAI) to his/her current CoA. Figure 4 shows the entities that make up the system.

A service provider (SP) generates revenue by charging for usage of network resources from roaming mobiles and may also provide other value-added services. We employ micropayment technology, which allows the routers within the access network to authenticate datagrams prior to making a routing decision and the SP to be paid in real time for service provision. This eliminates the need for any long-lived contracts between the MN and the operators. Existing credit-based mobile billing systems trust users to pay their bills based on strong identity verification and credit history checks. They also place total trust in the operator to bill the user for the correct amount, but provide no mechanisms to prove the authenticity of the CDRs generated by the operator. Unlimited credit with post-fact punishment is too open to abuse in mobile networks. With a large population of mobile users and independent network operators, it is desirable to remove the need to trust them and thereby minimize fraud in the system.

A mobile user registers with the SP to obtain a unique network address for the access network in which he/she currently finds him/herself. This address serves as the new CoA for the user while he/she remains in the access network. The user subsequently purchases provider-specific payment tokens from his/her BK via the GW node, as payment for using resources in the access network. The GW authenticates the BK commitment and broadcasts the payment details to the routing nodes within the access network. In order to be able to receive incoming calls, the MN sends an authenticated update of its new CoA to its chosen LS. It also attaches the number of payment tokens the location server requires to keep the CoA information alive within its database.

In addition to the existing fixed servers in the network, we now have a new category of mobile users and therefore handle data calls differently

from voice calls. To make a data call, the MN obtains the IP address of the destination node by querying a domain name server (DNS), and forwards datagrams toward the destination via the GW node in the access network. For voice calls, the MN first pays its LS to obtain the current CoA for the correspondent node, and then forwards datagrams to the CN via the gateway. In each case, the MN attaches the required amount of payment tokens along with each datagram, to enable them to be transported through the access network. The hash values also perform an authentication function, and are used by the routing nodes within the access network to update their soft-state route mappings for roaming MNs. The payment parameters are removed by the GW node prior to releasing the datagrams into the core network. The service provider periodically deposits the highest payment hash in each chain used by an MN with its BK to redeem payment for network usage.

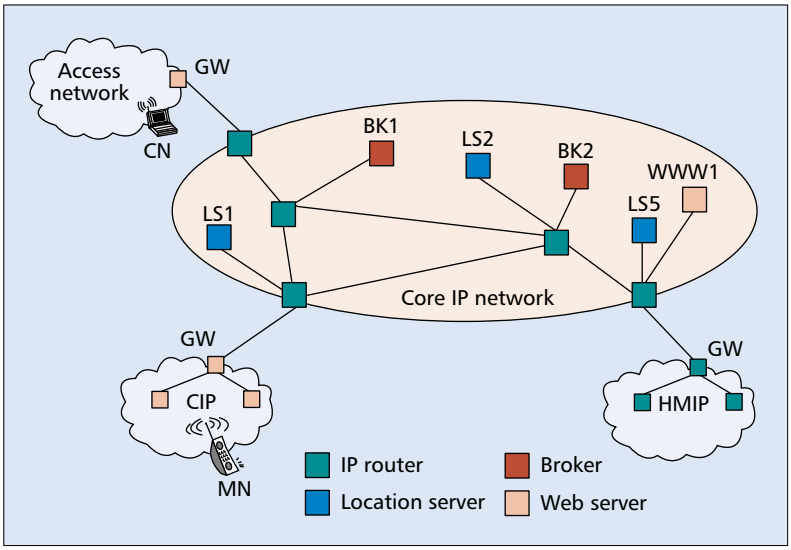

Figure 4. *The network model.*

Design Goals

The transition from macro- to micromobility-based architectures has been outlined along with the associated security issues. A vision of the architecture of future Mobile-IP-based access networks has also been presented. We now discuss the design goals of our authentication and accounting solution for Mobile-IP-based networks.

Real-Time Payment — A mobile user should be able to pay in real time for network usage in an access network without the need to contact a home network. Furthermore, if the location management functionality can be provided as a paid service by a VASP, subscription with a home operator can be completely removed. By removing the need for subscription-based billing and location management, the requirement for a home network can be eliminated.

Authentication and Router Updates — Users should not have to maintain a separate security relationship with each access network they may use from time to time. In addition, signaling messages transmitted by an MN in order to add or update the soft-state routing entries within the routers in the access network must be quickly and efficiently authenticated. This should be done without the need for extensive key exchanges between the various entities or the need to contact a third party.

Lightweight Cryptographic Procedures — Authentication and accounting related cryptographic data that accompanies datagrams in the access network should be kept to a minimum. This will lead to fast processing of datagrams prior to routing and minimize storage requirements on intermediate nodes. Nodes outside the access network should not have to understand the mobility or payment messages. Packets that traverse the core network should appear as normal IP datagrams to intermediate routing nodes.

Offline Payment Verification — There should be no need to contact a third party to verify the validity of payment tokens by an accepting entity. The payee should be able to present the payment tokens at a later date to a BK and be guaranteed payment. There should be multiple BKs in the system, which allows payment tokens to be redeemable at the recipient's BK who in turn trusts the issuing BK.

Identified Payment Tokens — Payment tokens should redeemable only by the specified entity. This will prevent double-spending of tokens by cheating nodes or eavesdroppers in the network.

Personal Mobility — A user should be reachable using a globally unique identifier and be able to obtain an access-network-specific address on demand. The payment tokens and associated security keys for a user can be stored on a smart card. This allows them to be used in any mobile device owned or rented by the user. However, the smart card should be used purely as a secure portable device, and there should be no dependency on the cryptographic hardware to maintain the security of the system.

Location Privacy — The location of a mobile user should be known to as few entities as possible in the system. One should be required to pay an LS to obtain the current CoA for a user. This will minimize the number of unsolicited messages sent to a node.

In brief, we wish to remove unnecessary trust from the system, reduce the online communications overhead of contacting a home network, and allow real-time payment for usage of network resources anywhere by anyone who holds valid payment tokens.

Registration in the Access Network

On entering or waking up in a new access network, an MN's first task is to obtain a network address. It first obtains the address of the GW node, which is periodically advertised by the base stations in the network, and sends a registration request to it. The MN identifies itself by attaching a digitally signed copy of its NAI and

The mobility and payment related fields are kept within the mobility header and are required only by the nodes within the access network. The mobility header is removed by the gateway node prior to any datagrams leaving the access network.

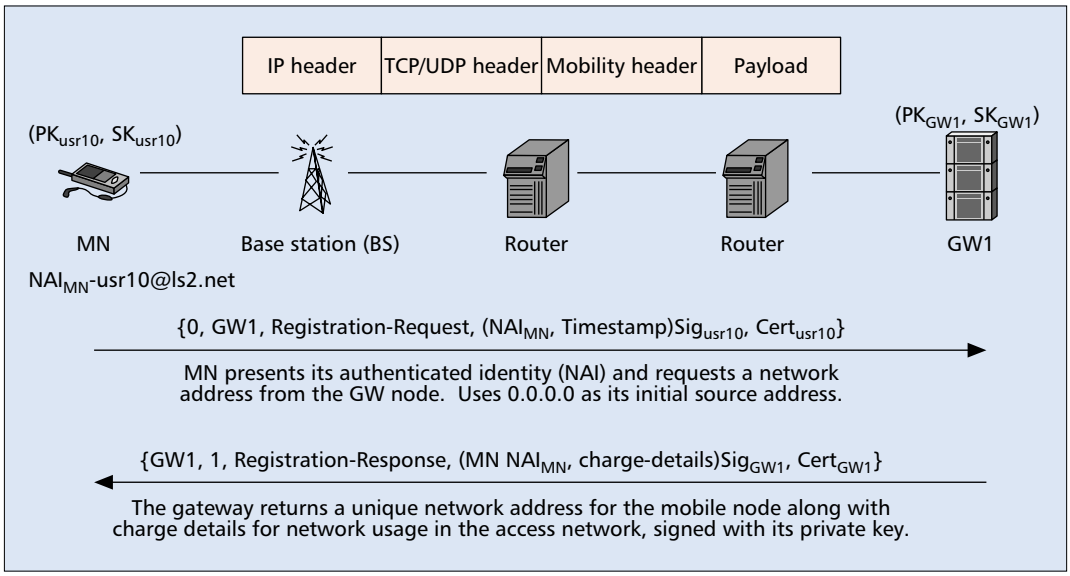

Figure 5. *Registration procedure and datagram format.*

a timestamp, and uses a null IP address as its initial source address. An MN is issued an NAI when it establishes an association with an LS. It may additionally get its public key certified by the LS or any other widely known TTP. The GW returns a unique network-specific address, along with its charge details for network usage signed with its private key in response. The charge details at a very minimum may specify the charge for carrying data and voice traffic to and from the core network. Figure 5 shows the datagram format, which contains an additional header we refer to as the *mobility header*, and the message exchanges between the mobile and GW nodes. The mobility and payment-related fields are kept within the mobility header and are required only by the nodes within the access network. The mobility header is removed by the GW node prior to any datagrams leaving the access network.

Payment Chain Purchase from a Broker

Prior to making any calls, an MN must purchase access-network-specific payment tokens from its BK, whom the service provider must also trust. The MN generates a UOBT of the desired length from a secret tree root to obtain, for example, the set of anchors $\{P1_0 \ldots P100_0\}$. The secret roots of the UOBT from which the individual subchains are generated do not leave the user's device during the chain purchase protocol. The payment chains of the UOBT can only be spent at the specified SP and have no monetary value until committed to by a BK. To obtain this commitment, the MN attaches a macropayment for the broker in the Purchase-Request message. The request is sent to the GW node in the access network, which forwards it to the identified broker (BK2) and monitors the reply.

The message consists of the payment details such as the NAI of the user, the identity of the SP where the chains are to be spent, and the UOBT identifier along with the set of anchors, the length of each chain, and the value of a payment token in the chain. The payment details are signed with the private key of the user, the authenticity of which the BK can verify. The macropayment is encrypted with the public key of the BK, which ensures that no unauthorized entity is able to access the payment details during transit. The MN and BK already possess each other's public key certificates, since they have a pre-established accounting relationship. Figure 6 shows the payment chain purchase protocol in detail.

The BK commits to the hash chains or promises to honor their value by digitally signing the payment chain commitment (Comm), consisting of the payment details sent by the user and an expiry date associated with the chains. The commitment shows that each payment hash from the chains represents a prepaid value redeemable at the BK. The commitment is returned to the user via the GW node in the access network. The GW verifies the BK commitment, and broadcasts the relevant payment details (Auth) to all internal routing nodes in the access network. The payment details are stored in an authentication cache (Auth Cache) in each router. Hash tokens are released subsequently to the SP by the MN, and serve as payment throughout the duration of a call. Since the hash tokens are generated using a one-way function, they can serve the dual purpose of authentication. They are used by the routers in the access network to authenticate signaling messages and datagrams, prior to updating their route cache entry for an MN. If the MN already possesses valid payment tokens for the access network, it may present these to the GW node, who can authenticate the same and broadcast the relevant details to the routing nodes within the access network.

Location Management

In order to receive incoming calls, a user must periodically update his/her CoA details at his/her designated LS. Other users in the network wishing to contact the user can identify the serving

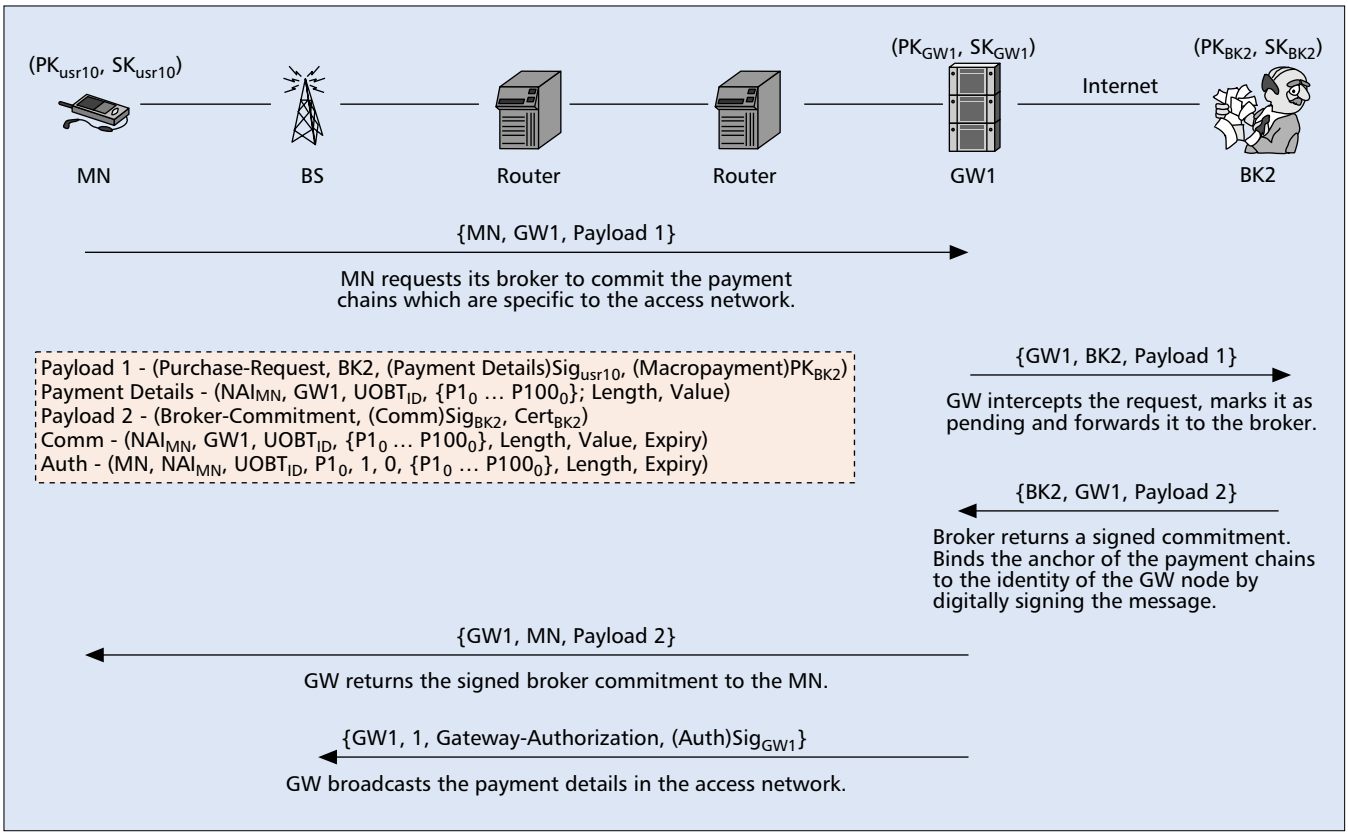

Figure 6. *Payment chain purchase.*

LS from the user's NAI, and obtain the current CoA for the user. Figure 7 shows the sequence of steps involved in the process. The MN sends a message to its LS via the GW node in the access network. The mobility header contains a payload identifier (Data), the UOBT identifier ($UOBT_{ID}$), a hash token (PX_Y) as payment for network usage in the access network, the subchain number X, and the hash number Y. Intermediate routing nodes verify the authenticity of the hash token by hashing back to the last hash value stored in their Auth Cache. A valid hash value allows them to add to or update both their authentication and route caches. The Auth Cache always contains the last valid hash token received by the node. This aids in fast verification of the hash tokens since a node is not required to hash back to the anchor of the chain each time.

The GW node removes the mobility header and forwards the payload to the specified destination address. The datagram is treated as a regular IP datagram by routers in the core network. The contents of the payload are decrypted by LS2 to reveal a Location-Update message, which contains the new CoA for the user. The user also attaches the required hash tokens as payment to the LS for hosting this information. The payment tokens are from a separate UOBT that the user purchased previously in order to spend at the LS. Alternatively, a user may have an accounting relationship with an LS. He/she may be required to digitally sign the Location-Update message so that the LS can verify its authenticity and debit his/her account.

Call Delivery: Data

From time to time a node may wish to contact other fixed nodes in the core network, such as a Web server or a VASP. The address of such a server is well known and remains unchanged over a period of time. The MN may query a DNS to obtain the IP address of the server, and may need to pay the SP for the same. Once the MN has the network address of the server, it attaches the required payment tokens along with each datagram to be transported through the access network towards the destination. Intermediate routing nodes on the path to the gateway can authenticate the hash tokens, by hashing back to the last stored hash value in their authentication cache. The GW as before removes the mobility header and forwards the datagrams.

Note that to pay the SP multiple hash tokens, the MN does not have to attach multiple hash values. Instead it can just attach the correct hash value further up the chain. Sending $P1_5$ after $P1_1$ is the equivalent of sending four payment hashes, $P1_2$–$P1_5$, since they can be obtained from $P1_5$ by repeatedly applying the correct hash function. Also, since the datagrams are handled in the normal manner by intermediate routers in the core network as well as the destination node, existing applications can be used without modification. The charge details the MN negotiated with the SP for accessing data services in the core network allow for a negotiated amount of data to be delivered to the MN in response to a request. However, the GW node can at any stage demand that the MN release further payment

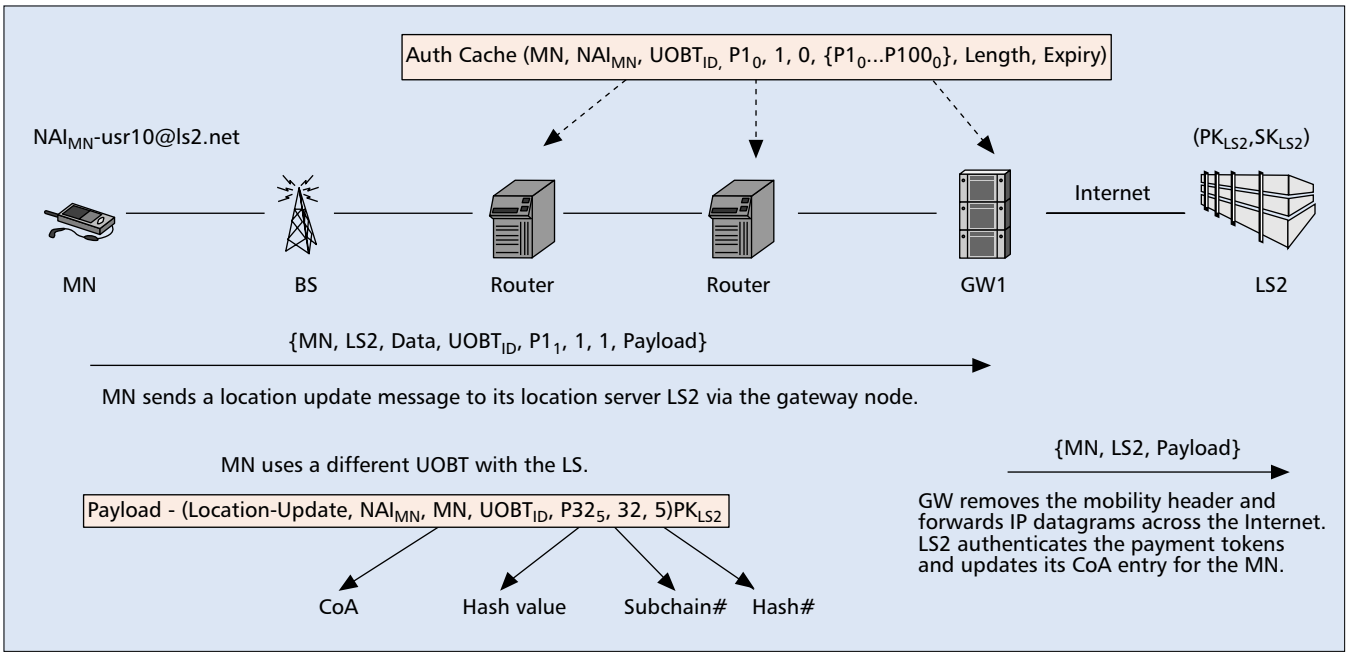

Figure 7. *Updating the care-of address.*

tokens if the traffic levels exceed the agreed contract details. Datagrams on the reverse path are delivered without modification to the MN. After a call finishes, the MN can use the unspent hashes in the remaining chains on other calls. During a call, the routers in the access network ensure that the hashes sent with a datagram have not already been used.

Handover and Authentication

During handover, an MN changes its point of attachment from one base station to another in the access network. New nodes in the path to the GW must hash back to the anchor of the current chain, while existing nodes are only required to hash back to the last value stored in their Auth Cache. Figure 8 shows the message exchanges when the MN moves between base stations. Prior to moving, the MN was using the first subchain (P1), and the last hash value it released was the 53rd ($P1_{53}$) hash in that chain. When the MN changes base stations, it transmits a datagram with the hash value ($P1_{57}$) from the P1 chain. Any new routers in the path toward the GW node verify the hash value by hashing back to the current subchain anchor ($P1_0$) and consult their Auth Cache. However, existing routers in the new path to the gateway only hash back to the last stored hash value ($P1_{53}$) in their Auth Cache. It has been shown that an ordinary desktop machine can perform over 100,000 hash operations/s. Therefore, there should only be a minimum amount of delay in setting up the new path in the access network. If an MN has no data to send but wishes to keep its route entries alive in the access network, it can forward the required number of hash tokens with a network-specific keep alive message.

Call Delivery: Voice

Whenever a user wishes to make a voice call to another user in the network, the MN must first obtain the current CoA for the destination node. The MN does this by requesting its LS to obtain the CoA from the correspondent node's (CN's) LS, LS1, which can be identified from the NAI of the CN (cn@ls1.net). The MN transmits a datagram destined for its LS, LS2, via the GW node, the payload of which contains a Location-Request message. As far as the GW is concerned, the MN has transmitted another data packet and it needs to validate the payment tokens. The GW removes the mobility header and forwards the packet toward the destination. The MN's LS decrypts the contents of the payload to reveal a Location-Request message from the MN, along with the payment tokens required by it to complete the operation.

The LS uses the NAI_{CN} field in the message to obtain the address of the LS that hosts the CoA information for the CN. It sends its own location request message to LS1 asking it for the current CoA for the CN along with a micro or macropayment for LS1. LS1 responds with a Location-Response message, which is forwarded to the MN. The details are signed with the private key of LS1 and encrypted with the recipient's public key so that intermediate nodes are not able to obtain any details of the whereabouts of an MN. Once the MN has the network address of the CN, it can forward datagrams directly to the CN via the GW node, as long as it pays the SP for network usage in the access network. The MN indicates in the mobility header to the GW node that it is sending voice data in the payload field to the CN, and attaches the required number of payment tokens with each datagram. Intermediate routing nodes are able to verify the authenticity of the datagrams prior to making a routing decision. If the MN exhausts the current chain, it can immediately start using the next available subchain of the UOBT.

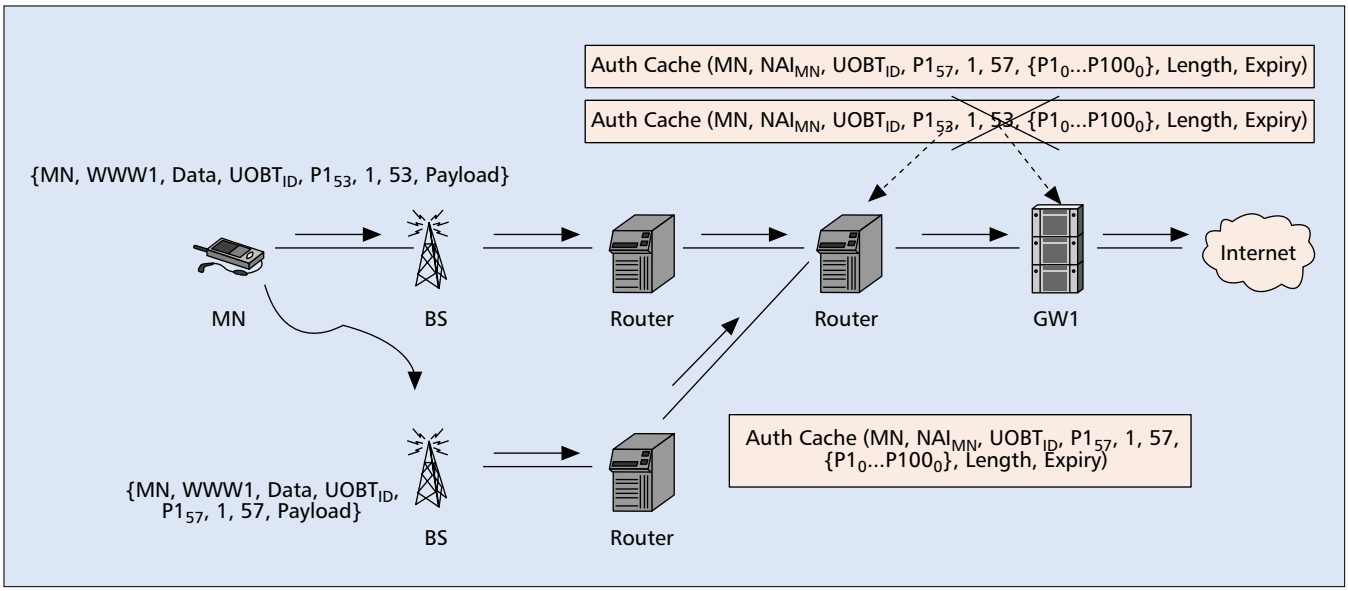

Figure 8. *Handover in the access network.*

Broker Clearing

Periodically an SP contacts its BK and deposits payment tokens it has collected for usage of network resources by mobiles. It encrypts the payment details with the public key of the BK who issued the payment chains. The SP's BK, BK1, forwards the message to the issuing BK, BK2, who verifies the payment tokens and checks for double-spending. BK1 marks the transaction as pending and forwards the message to its correct destination. The actual payment details are encrypted with the public key of BK2 and not accessible to BK1. A reply is sent back to BK1 authorizing it to credit the SP's account for the specified amount. The reply from BK2 also contains a payment receipt, which is encrypted with the public key of the SP and forwarded to the SP by BK1. The BKs in the system have accounting relationships and periodically transfer funds between each other to settle user accounts. The value of unspent hashes can be reclaimed by a user once the chains have expired and the spent hashes have been deposited by the SP.

Discussion

Our solution provides a means of allowing real-time payment and authenticated router updates in the access network. A micropayment scheme using hash functions and offline broker contact allows the solution to be efficient and scalable. To aid performance, digital signatures are only used at call setup and to generate the broker commitment. Subsequent authentication and accounting procedures are achieved using only hash tokens. The proposed protocol is generic enough to easily be integrated in any of the micromobility schemes we have studied. We make use of an additional mobility header that is access-network-specific. The gateway node in the access network removes this header prior to any datagrams being forwarded onto the core network. On the reverse path the datagrams are delivered without modification to the MN. This ensures that existing Internet applications can work without any modifications.

With the exception of purchasing payment chains and location management procedures, the overhead incurred by including the mobility header in each datagram is 16 bytes for the hash value (assuming we make use of MD5 to generate the subchains) and another 8 bytes to represent the payload type, UOBT identifier, subchain number, and hash number. Thus, a total of 24 bytes is added as part of the mobility header within the access network. In addition, each router in the access network must also maintain an authentication cache it must consult prior to making any modifications to the route cache. Registering the care-of address at a location management server is an optional feature, which may or may not be used by all users in the system. A user may only wish to access data services or make outgoing calls from his/her MN, in which case he/she is not obliged to register his/her current CoA with a location management server. Only if a user wishes to be reachable for incoming calls does he/she need to update his/her care-of address periodically at a location server, and pay for the privilege.

Hash chains are of a finite length, and there is a possibility that a node may run out of hash values during a session. In the case of a UOBT, if there are unused subchains, the user can switch to the next subchain immediately. If there are no further subchains available, it can result in the dropping of a connection. This is particularly true for real-time communications such as voice telephony or videoconferencing where the source may transmit a large number of datagrams during a session. However, this applies to all payment protocols that have a fixed amount in the user's purse.

When a mobile node moves between access networks, any ongoing calls may be dropped. This is due to the fact that the MN will need to register in the new network and obtain another IP address. In addition, it will require new

We believe that in the future there will be a large number of micro and pico-cellular mobile networks based on IP, which will provide the next generation of telecommunications services to a very large user population.

provider-specific payment tokens to pay for network usage. If, on the other hand, an MN already possesses valid payment tokens from a prior visit, it may use them instead. If the user was involved in a data call such as accessing a Web server, there may be a small delay before the connection can be resumed. However, for voice calls a new connection will have to be set up with the destination node.

CONCLUSION

With the widespread use of PDAs and the availability of low-cost wireless networking hardware, there has been considerable growth in the number of wireless access networks. Currently such networks are operated by individual organizations, and are usually closed to users who belong to other network operators or organizations. One of the reasons for this is that such closed networks do not have any AAA provisioning policies in place, and thus cannot deal with nodes with which they do not have a pre-established security relationship.

With a large number of network operators and service providers, it is necessary to guarantee payment and remove any complex trust relationships involved in billing. We identify a number of problems with the current AAA mechanisms for Mobile-IP-based access networks. The desirable properties of an authentication and accounting system are then drawn up, and a solution that securely achieves those goals is proposed. To allow for an efficient solution, with minimal computational costs during packet authentication, a scheme based on hash chain trees is used. The use of hash chain trees is an advantage over ordinary hash chains; it allows efficient generation and storage of hash values within a mobile device, which may have limited storage capacity. We have also kept the number of cryptographic keys required to a minimum, while allowing for fast verification of the authentication data carried within the datagrams.

We believe that in the future there will be a large number of micro- and picocellular mobile networks based on IP, which will provide the next generation of telecommunications services to a very large user population. These networks will require secure and scalable AAA provisioning. Our solution provides an efficient means of authentication of signaling messages and accounting of resources in such networks.

REFERENCES

[1] C. Perkins, Ed., "IP Mobility Support," IETF RFC 2002, Oct. 1996.
[2] S. Glass *et al.*, "Mobile IP Authentication, Authorization and Accounting Requirements," IETF RFC 2977, Oct. 2000.
[3] A. Campbell *et al.*, "Design, Implementation, and Evaluation of Cellular IP," *IEEE Pers. Commun.*, vol. 7, no. 4, Aug. 2000, pp. 42–49.
[4] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regional Registration," Internet draft, draft-ietf-mobileip-reg-tunnel-06, work in progress, Mar. 2002.
[5] R. Ramjee *et al.*, "IP-Based Access Network Infrastructure for Next-Generation Wireless Data Networks," *IEEE Pers. Commun.*, vol. 7, no. 4, Aug. 2000, pp. 34–41.
[6] A. Campbell *et al.*, "Comparison of IP Micromobility Protocols," *IEEE Wireless Commun.*, vol. 9, no. 1, Feb. 2002, pp. 72–82.
[7] P. Calhoun, T. Johansson, and C. Perkins, "Diameter Mobile IPv4 Application," Internet draft, draft-ietf-aaa-diameter-mobileip-09, work in progress, Mar. 2002.
[8] D. O'Mahony, M. Peirce, and H. Tewari, *Electronic Payment Systems for E-Commerce*, 2nd ed., Artech House, 2001.
[9] L. Lamport, "Password Authentication with Insecure Communication," *Commun. ACM*, vol. 24, no. 11, Nov. 1981, pp. 770–72.
[10] S. Yen, L. Ho, and C. Huang, "Internet Micropayment Based on Unbalanced One-way Binary Tree," *Proc. CrypTEC '99*, Hong Kong, July 1999, pp. 155–62.
[11] B. Aboba, "The Network Access Identifier," IETF RFC 2486, Jan. 1999.

BIOGRAPHIES

HITESH TEWARI (htewari@cs.tcd.ie) is a lecturer in computer science at Trinity College Dublin, Ireland, and works as part of the Networks and Telecommunications Research Group, which is involved in the development of an experimental 4G network. His current research interests cover network security, electronic payment systems, mobile communications, and ad hoc networks. He has co-authored a book, *Electronic Payment Systems for E-Commerce*, 2nd ed., which is a best seller in its category.

DONAL O'MAHONY (omahony@cs.tcd.ie) received B.A., B.A.I., and Ph.D. degrees from Trinity College Dublin, Ireland. After a brief career in industry at SORD Computer Systems in Tokyo and IBM in Dublin, he joined Trinity College as a lecturer in computer science in 1984. At Trinity, he coordinates a research group working in the areas of networks and telecommunications. Within this group, projects are ongoing in a wide range of areas including electronic commerce, network security, and mobile communications technology. He is co-author of the leading text on electronic payment systems with Michael Peirce and Hitesh Tewari. He was awarded a Fulbright scholarship in 1998 and was made a fellow of Trinity College in 2001.