Integration of 802.11 and Third-Generation Wireless Data Networks

M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, L. Salgarelli
Bell Labs, Lucent Technologies, NJ USA

Abstract— **The third-generation (3G) wide area wireless networks and 802.11 local area wireless networks possess complementary characteristics. 3G networks promise to offer always-on, ubiquitous connectivity with relatively low data rates. 802.11 offers much higher data rates, comparable to wired networks, but can cover only smaller areas, suitable for hot-spot applications in hotels and airports. The performance and flexibility of wireless data services would be dramatically improved if users could seamlessly roam across the two networks. In this paper, we address the problem of integration of these two classes of networks to offer such seamless connectivity. Specifically, we describe two possible integration approaches - namely tight integration and loose integration and advocate the latter as the preferred approach. Our realization of the loose integration approach consists of two components: a new network element called IOTA gateway deployed in 802.11 networks, and a new client software. The IOTA gateway is composed of several software modules, and with co-operation from the client software offers integrated 802.11/3G wireless data services that support seamless inter-technology mobility, Quality of Service (QoS) guarantees and multi-provider roaming agreements. We describe the design and implementation of the IOTA gateway and the client software in detail and present experimental performance results that validate our architectural approach.**

I. Introduction

Recent trends indicate that local area wireless networks based on IEEE 802.11 standards and third-generation wide area wireless networks such as CDMA2000 and UMTS will co-exist to offer Internet access to end users. The two technologies offer characteristics that complement each other perfectly. The 802.11 standards allow the realization of economical Wireless LANs that support data rates anywhere from 1 Mbps to 54 Mbps based on the distance to the base station (often called Access Points) [1]. However, 802.11 Access Points can cover areas of only a few thousand square meters, making them suitable for enterprise networks and public hot-spots such as hotels and airports. On the contrary, wireless networks built using the 3G standards [6], [3] require significant capital investments, support limited peak rates that range from 64 Kbps to nearly 2 Mbps as a maximum, but offer a much wider area of coverage that enables ubiquitous connectivity. The deployment of architectures that allow users to seamlessly switch between these two types of network would present several advantages to both service providers and users. By offering integrated 802.11/3G services, 3G operators and Wireless Internet Service Providers (WISP) could capitalize on their investments, attract a wider user base and ultimately facilitate the ubiquitous introduction of high-speed wireless data. Users would benefit from the enhanced performance and lower overall cost of such a combined service.

The design of a network architecture that efficiently integrates 3G and 802.11 is a challenging task, particularly when the objective is to make the interoperation between the two technologies as seamless and as efficient as possible, both from the end-user's and from the operator's perspectives. Wireless LANs, originally targeted at enterprise and home networks, lack many of the capabilities which are essential in public environments. These capabilities include unified and universally accepted authentication, accounting and billing mechanisms; the integration of mobility mechanisms with QoS and application-level services; the support for heterogeneous network architectures through the implementation of roaming agreements. Conversely, although these characteristics are present by design in 3G networks, their implementation depends on specific wireless access architectures such as CDMA2000 [3] or UMTS [6], and their extension to other wireless technologies such as 802.11 presents several compatibility issues. Depending on the level of inter-dependence that one is willing to introduce between 802.11 and 3G, the design of integrated multi-technology wireless systems can lead to network architectures that have fundamentally different properties.

A. Research Contributions

In this paper, we describe two possible approaches to the design of an integrated 3G/802.11 network architecture, defined as tightly-coupled and loosely-coupled interworking. We analyze the two designs qualitatively, and advocate the latter as the preferred approach. We then introduce our realization of the loose approach, which targets specifically the integration of 802.11 and 3G CDMA2000. Our solution, which has been operational in our testbed since January 2002, consists of two key components: a new network element called IOTA access gateway, deployed in the 802.11 network, and a new service access software on the client devices. We describe the design and implementation of the IOTA gateway and its various components such as mobility agents, integrated web cache, RADIUS server, accounting support etc. in detail. We also present detailed design of the client software. We describe experimental performance results that validate our integration approach. Finally, we briefly describe how our approach can be easily extended for integration of 802.11 and UMTS networks.

B. Outline

The rest of the paper is organized as follows. In Section II we describe an example service scenario for integrated access to 3G and 802.11 networks, and highlight the issues that motivated our work. Section III describes the *tightly-coupled* and *loosely-coupled* integration architectures and motivates the choice of the latter as the preferred approach. In Sections IV and V we introduce in detail the key components of our realization of loosely-coupled architecture, namely the IOTA 802.11 gateway and its corresponding client software. Section VI reports experimental results that characterize the performance of our system and validate our architectural choices. Section VII describes the related research work reported in literature. Finally, Section VIII describes the conclusions and future work.

II. Service Scenario in Integrated 802.11 and 3G networks: problem statement

In 802.11 networks, Access Points (AP) bridge the wireless and wired parts of the network. However, the current 802.11 protocol suite only defines the physical and media access control layers but not the layers above. There are three implications to this. First, authentication procedures vary from provider to provider, depending on the particular architecture and set of authentication protocols that they decide to deploy. Second, existing standards do not define the characteristics of the services offered to users, for example with respect to QoS guarantees. Finally, there is currently no agreed-upon mobility-management mechanism that would allow users to seamlessly roam across different 802.11 networks managed by different providers.

In 3G networks, Base Stations (BS) together with Radio Network Controllers (RNC) bridge the wireless and wired network. There are two dominating 3G standard suites – CDMA2000 and UMTS. In the case of CDMA2000, the Packet Control Function (PCF) and Packet Data Service Nodes (PDSN) channel data packets to the Internet through the provider's core network. In the case of UMTS, the Serving and Gateway GPRS Service Nodes (SGSN and GGSN) provide logically similar functionalities. Unlike 802.11, 3G standards cover also the layers above the media access, so protocols that deal with authentication procedures, QoS guarantees, and mobility management are standardized. Users are guaranteed that they can seamlessly roam across 3G networks owned by different providers, assuming that they share a roaming agreement.

Now, let us consider an example of an ideal service scenario. A user, John Doe, has a laptop/handheld that has both a 3G and an 802.11 interface. The 802.11 service that many airports offer is appealing to him, because of the high bandwidth he could enjoy. However, given that 802.11 can offer only spot coverage, John would need to sign-up with many 802.11 providers in order to receive service in the places he most visits. Furthermore, he would need to manually setup and tear-down his wireless connection as he travels from one place to the other. John is therefore attracted by the ubiquitous coverage of 3G, and thus he decides to sign up with a 3G carrier, which, in turn, has roaming agreements with many 802.11 service providers. When John travels to a place, such as an airport concourse, where there is such an 802.11 service provider, his machine should be able to transparently switch to the 802.11 access. When John leaves the coverage of the 802.11 provider, his machine should seamlessly switch to the 3G access.

There are several issues that must be resolved to enable such scenario. First, as a subscriber of the 3G carrier, John's machine is configured with a security association (a user identity and a secret key) with the carrier. However, prior to his trying to access the 802.11 network, the 802.11 provider does not know anything about John. Therefore, there must exist a secure mechanism through which the 802.11 provider can authenticate John by interacting with the Authentication, Authorization and Accounting (AAA) server of the 3G carrier. Second, when the switching occurs, John may have several ongoing network sessions (e.g., network radio, voice chat, etc), and these sessions should be transparently maintained. Third, as a related point, the switching should happen automatically and transparently without John's intervention. Fourth, the 802.11 provider should be able to honor the service level, such as QoS guarantees, that the carrier has agreed to provide to John, while enforcing the policies that John's contract with the 3G carrier foresees. As a minimum, this means that the 802.11 provider has to obtain John's user profile from the carrier infrastructure (most likely the AAA server) and be able to map the local service characteristics to the desired service described in the profile. Finally, the accounting and billing infrastructures of the 3G carrier and the 802.11 provider must be interfaced to enable periodic revenue sharing and settlement and to allow the 3G carrier to generate a common bill to the customer. Typically, the last two issues are addressed by establishing roaming agreements between the providers and therefore, efficient mechanisms are required to setup the same.

There are two fundamentally different architectural approaches to solve the above issues. We outline the two approaches and contrast their characteristics in the next section.

III. Two Architectures for 802.11 and 3G Integration

Depending on the degree of inter-dependence that one is willing to introduce between the 3G network and the 802.11 network, there are two different ways of integrating the two wireless technologies. We define them as *tightly-coupled interworking* and *loosely-coupled interworking*.

A. Tightly-coupled Interworking

The rationale behind the tightly-coupled approach is to make the 802.11 network appear to the 3G core network as another 3G access network. The 802.11 network would emulate functions which are natively available in 3G radio access networks. In this architecture, utilized by WISP No.1 in Figure 1, the "802.11 gateway" network element appears to the upstream 3G core as either a PCF, in the case of a CDMA2000 core network, or as an SGSN, in the case of

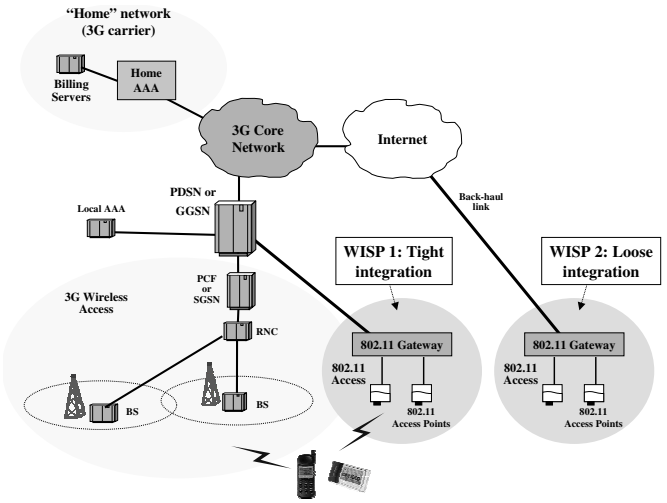

Fig. 1. 3G and 802.11 integration: tightly-coupled vs. loosely-coupled architectures.

UMTS. The 802.11 gateway hides the details of the 802.11 network to the 3G core, and implements all the 3G protocols (mobility management, authentication, etc.) required in a 3G radio access network. Mobile Nodes in this approach are required to implement the corresponding 3G protocol stack on top of their standard 802.11 network cards, and switch from one physical layer to the next as needed. All the traffic generated by clients in the 802.11 network is injected using 3G protocols in the 3G core. The different networks would share the same authentication, signaling, transport and billing infrastructures, independently from the protocols used at the physical layer on the radio interface.

However, this approach presents several disadvantages. Since the 3G core network directly exposes its interfaces to the 802.11 network, the same operator must own both the 802.11 and the 3G parts of the network. In fact, in this case, independently operated 802.11 islands could not be integrated with 3G networks. Today's 3G networks are being deployed using carefully engineered network-planning tools, and the capacity and configuration of each network element is calculated using mechanisms which are very much specific to the technology utilized over the air interface. By injecting the 802.11 traffic directly into the 3G core, the setup of the entire network, as well as the configuration and the design of network elements such as PDSNs and GGSNs have to be modified to sustain the increased load.

The configuration of the client devices also presents several issues with this approach. First, as described earlier, the 802.11 network cards would need to implement the 3G protocol stack. It would also mandate the use of 3G-specific authentication mechanisms based on Universal Subscriber Identity Module [6] or Removable User Identity Module (R-UIM) cards [2] for authentication on Wireless LANs, forcing 802.11 providers to interconnect to the 3G carriers' SS7 network to perform authentication procedures. This would also imply the use of 802.11 network interface cards with built-in USIM or R-UIM slots or external cards plugged separately into the subscriber devices.

For the reasons described above, the complexity and the high cost of the reconfiguration of the 3G core networks and of the 802.11 gateways would force operators that chose the tightly-coupled approach to become uncompetitive to 802.11-only WISPs.

B. Loosely-coupled Interworking

Like the previous architecture, the loosely-coupled approach calls for the introduction of a new element in the 802.11 network, the 802.11 gateway. However, in this design (WISP No.2 in Figure 1), the gateway connects to the Internet and does not have any direct link to 3G network elements such as PDSNs, GGSNs or 3G core network switches. The user population that accesses services of the 802.11 gateway may include users that have locally signed on, as well as mobile users visiting from other networks. We call this approach *loosely-coupled interworking* because it completely separates the data paths in 802.11 and 3G networks. The high speed 802.11 data traffic is never injected into the 3G core network but the end user still achieves seamless access.

In this approach, different mechanisms and protocols can handle authentication, billing and mobility management in the 3G and 802.11 portions of the network. However, for seamless operation to be possible, they have to interoperate. In the case of interoperation with CDMA2000, this requires that the 802.11 gateway supports Mobile-IP functionalities to handle mobility across networks, as well as AAA services to interwork with the 3G's home network AAA servers. This would enable the 3G provider to collect the 802.11 accounting records and generate a unified billing statement indicating usage and various price schemes for both (3G and 802.11) networks. At the same time, the use of compatible AAA services on the two networks would allow the 802.11 gateway to dynamically obtain per-user service policies from their Home AAA servers, and to enforce and adapt such policies to the 802.11 network.

Since the UMTS standards do not yet include support for IETF protocols such as AAA and Mobile-IP, more adaptation is required to integrate with UMTS networks. Mobile-IP services would need to be retrofitted to the GGSNs to enable seamless mobility between 802.11 and UMTS. Common subscriber databases would need to interface to Home Location Registers (HLR) for authentication and billing on the UMTS side of the network, and to AAA servers for the same operations to be performed while clients roam to 802.11 networks.

There are several advantages to the loosely-coupled integration approach. First, it allows the independent deployment and traffic engineering of 802.11 and 3G networks. 3G carriers can benefit from other providers' 802.11 deployments without extensive capital investments. At the same time, they can continue to deploy 3G networks using well-established engineering techniques and tools. Furthermore, while roaming agreements with many partners can result in widespread coverage, including key hot-spot areas, subscribers benefit from

having just one service provider for all network access. They no longer need to establish separate accounts with providers in different regions, or covering different access technologies. Finally, unlike the tightly-coupled approach, this architecture allows a WISP to provide its own public 802.11 hot-spot, interoperate through roaming agreements with public 802.11 and 3G service providers, or manage a privately installed enterprise Wireless LAN.

It should be clear that the loosely-coupled approach offers several architectural advantages over the tightly-coupled approach, with virtually no drawbacks. Therefore, we advocate the loosely-coupled approach as the preferred architecture for the integration of 802.11 with 3G networks, and we will use it as a reference throughout the rest of the paper.

IV. The IOTA 802.11 gateway

Using the framework provided by the loosely-coupled architecture described above, we designed and implemented a gateway system called IOTA[1]. Each gateway system serves multiple 802.11 access points in a hot-spot, and controls the traffic from these APs before it can reach the back-haul link.

A mobile node that roams into a hot-spot obtains 802.11 access under the control of the gateway. After successful authentication and Mobile-IP registration, the gateway allows the mobile node to access the network. Furthermore, the gateway provides QoS services and collects accounting data. Therefore, the gateway integrates a number of sub-systems, as shown in Figure 2: RADIUS server, Mobile-IP agent, dynamic firewall, QoS module, and accounting module. All the IOTA sub-systems rely on an on-disk database to store persistent information about each client's session. Thus, the state of the gateway can be preserved and restored even in the event of a system reboot.

All the IOTA building blocks are implemented as software modules, and run on top of the Linux Operating System. The design of the gateway software allows it to be scalable, so that it could be implemented on hardware of varying power, depending on the size of the 802.11 network. Furthermore, the design allows for a very inexpensive solution by not requiring custom-built hardware. Each of the IOTA gateways used in our laboratory is implemented on off-the-shelf rack-mountable PC servers.

The rest of this section first describes the details of the building blocks of IOTA, which is currently targeted at the integration of 802.11 hot-spots with CDMA2000 networks. Then, it introduces a service mode called Simple-IP that supports roaming but not seamless handoff. Section IV-H explains how we plan to extend the system to integrate with UMTS as well.

A. RADIUS server

The IOTA gateway contains a complete RADIUS AAA server. The server enables roaming agreements between the 3G providers and 802.11 WISP, and also provides authentication services to the 802.11 cloud.

[1]IOTA stands for Integration Of Two Access technologies(3G and 802.11).

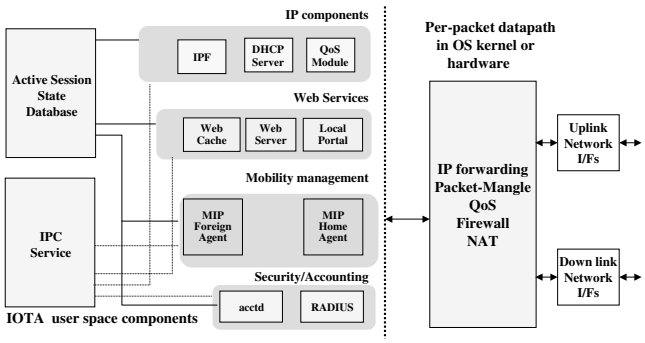

Fig. 2. Software architecture of the IOTA gateway.

Our server can be used to authenticate clients in two different ways. For Wireless LANs that implement the 802.1X [5] port-access control protocol, and that use the Extensible Authentication Protocol (EAP [11]) to transfer authentication information between the client and the network, the IOTA AAA server functions as an EAP relay. In this mode, it passes authentication information between the 802.11 APs and the client's Home AAA server. Our server supports IETF standardized EAP methods such as TLS [9], MD5 [11], One Time Password (OTP) [11], as well as legacy authentication methods such as PAP and CHAP [22]. In addition, it also implements novel authentication mechanisms such as the Shared Key Exchange [16], which has been highly optimized for the support of roaming clients in wireless networks.

For Wireless LANs that do not implement 802.1X, the IOTA AAA server interacts with the Mobile-IP Foreign Agent module (refer to Section IV-B), to authenticate the client with its Home AAA server based on the Mobile-IP mechanisms specified in [20].

In both cases, the presence of the AAA server on the gateway allows for an easy implementation of per-user policies. In fact, being on the path of the authentication exchange, the AAA server can obtain user profiles from their Home AAA server, and pass them on to the other IOTA modules for implementation and enforcement on the local network. At the same time, the IOTA AAA server serves as the Foreign AAA and can relay the RADIUS packets to a remote Home AAA via broker networks, allowing the efficient implementation of roaming agreements without any direct interaction between the 3G provider and the WISP.

B. Mobile-IP agent

The IOTA system implements a very scalable and efficient Mobile-IP agent functionality [14], which supports the roles of both Home and Foreign Agents (HA, FA). The Foreign Agent is used to manage the mobility of clients that move across different sub-networks, as well as across different wireless technologies. In fact, CDMA2000 uses Mobile-IP Foreign Agents in the PDSNs, and calls for the use of Mobile-IP to support seamless internetwork handoffs. By extending this functionality into the 802.11 network, the integration of the two mobility management mechanisms becomes automatic.

The Home Agent is used to support an emerging standard called "dynamic Home Agent allocation". In this case, during the initial authentication phase, the AAA infrastructure can allocate a Home Address and a corresponding Home Agent dynamically, every time a client session commences. This allows the HA to be allocated closer to the FA, reducing the length of the network path between them, and thus reducing the IP tunneling overhead. With this optimization, the mobile station's IP address is no longer well known across sessions, but it remains the same for a single Mobile-IP session.

C. Dynamic firewall

The IOTA gateway supports a dynamic stateful firewall service, implemented using the Linux IP Filter architecture. All the IOTA modules use the IOTA Packet Filter library (IPF), which is an abstraction layer on top of the IP Filter architecture, to install complex sets of packet filtering rules that depend on per-user policies. Such policies are dynamically obtained from the subscriber's Home AAA, hence the term "dynamic firewall service".

The Mobile-IP agents and the AAA server, upon successful authentication, install through IPF sets of rules that implement two major functionalities: firewalling and packet-mangling. The firewalling rules serve the dual purpose of protecting the clients from malicious attacks coming from the Internet (such as PING floods, TCP syn floods, etc.), and of protecting the IOTA gateway itself against traffic coming from malicious clients. IPF installs firewall rules that match layer-2 information, such as the MAC address of the clients. Therefore, attacks such as IP address spoofing become difficult to perpetrate.

The packet-mangling rules deal with the automatic redirection of user's traffic to local services, such as a local DNS server or the IOTA web-cache (refer to Section IV-F). Once again, these rules are all implemented on a per-user basis, depending on the user's profile downloaded from their Home AAA server.

D. QoS module

The IOTA system provides Quality of Service in the form of multiple service classes, each with a guaranteed minimum bandwidth. For example, an IOTA system can be configured with three classes (*Gold, Silver, Bronze*) and each class can be guaranteed a minimum bandwidth such as 750 Kbps for Gold, 250 Kbps for Silver and 125 Kbps for Bronze. If extra bandwidth is available, users can exceed their minimum rate, with higher class users getting the priority to grab excess resources. Users are assigned to their corresponding class based on information contained in their user profile, which is obtained by the IOTA gateway during the authentication phase, as explained in Section IV-A. The rate guarantee is valid only within the wireless access network, because the IOTA system does not control the entire network path between the communicating peers. To achieve true end-to-end QoS, some kind of QoS infrastructure (such as the IETF's differentiated-services, integrated-services or MPLS) must be present over the entire network path.

Our scheme is novel in that we provide QoS in 802.11 networks without air-link QoS mechanisms. While numerous research activities attempted to solve the fairness issues and to ensure different QoS levels in 802.11-type multiple access networks, virtually all of the existing proposals approach the problem at the MAC layer level, mostly by manipulating the back-off mechanism. We took a different approach. Our key idea is to control the amount of traffic which competes for resources, instead of prioritizing traffic when congestion occurs. The IOTA system, located between the 802.11 APs and back-haul link (Figure 1), controls all the traffic to and from the hot-spot, and manages the bandwidth for each user. We first estimate the capacity of the wireless link[2], and then shape the downstream traffic (i.e., packets from the Internet to mobile hosts) at the IOTA system to prevent excessive traffic from reaching to the wireless link. The upstream traffic (i.e., packets from mobile hosts to the Internet) is controlled similarly but in an indirect way, by relying on the higher-layer congestion control mechanisms (e.g., TCP). If a host pumps more traffic than its fair share into the network, IOTA drops or delays its packets so that the host can detect congestion and slow down the traffic generation. IOTA can accelerate the congestion detection at the client, by sending explicit ICMP source-quench messages.

The IOTA system manages bandwidth in two spots where congestion can occur, namely (1) the 802.11 APs, and (2) the back-haul link to the Internet that can be over-subscribed. The IOTA gateway uses SNMP queries to 802.11 APs to detect new user arrivals and user movements, and maintains the up-to-date user population map across APs. This map and the user profile obtained from the Home AAA are used to to determine each user's fair share of bandwidth. Depending on the pattern of user population, the 802.11 link or the back-haul link becomes the bottleneck, which results in the traffic shaping of some (or all) of the users' traffic. The IOTA system also provides admission control. Specifically, in case the wireless link bandwidth or the back-haul bandwidth is already entirely allocated to existing users, the gateway can be configured to either reject new users by blocking all their traffic, or to degrade them to the best-effort class, which does not get any rate guarantee.

We implemented the IOTA rate adaptation mechanism using a simple token bucket scheme with low performance overhead. In our scheme, we assign two token buckets for each user, one for upstream traffic, the other for downstream traffic. Since it works at the IP layer, this mechanism will co-exist with future QoS mechanisms that the IEEE 802.11e standards may mandate.

E. Accounting module

The potential to share usage revenue is one of the key business motivations for a 3G carrier and a 802.11 service provider to sign a roaming agreement with each other. To

[2]For example, the actual link capacity (in terms of total throughput) of a 802.11b network is around 4 to 6 Mbps depending on the vendors.

support this, after a user is authenticated and authorized to use a foreign 802.11 network, the IOTA gateway must collect accounting data of the user session and forward them to the home accounting server for billing purposes.

Since the IOTA gateway supports three different operation modes, there are three entities that may authenticate users and request services from the accounting sub-system. If Mobile-IP is used, the entity is the Foreign Agent. If, as explained later in Section IV-G, the Simple-IP mode is used, the entity is the web authenticator. If 802.1X is used, the local AAA server is involved in the exchange of EAP messages and is also one such entity. These entities, which will be referred to as the *applications* in this section, request accounting services by triggering accounting `start` and `stop` operations.

We provide the accounting mechanism but do not mandate the specific pricing policies such as time-based, usage-based, or flat-price scheme. Therefore, all potentially relevant accounting data of a user session are collected. They include start and stop times, duration, packet and octet counts. The accounting subsystem obtains these data from different sources. It obtains the time and duration data from the system clock when the `start` and `stop` triggers happen. It obtains the packet and octet counts from the kernel through a special call to the IPF module. The accounting subsystem also obtains auxiliary information such as user identity, IP address, MAC address, etc. from the active-session database.

These data are then transmitted to an accounting server using accounting start, stop, and interim-update messages. Currently, our system uses RADIUS [15][21] to send these messages, but in the future we may support other protocols such as the DIAMETER [17] or the protocols required by UMTS [4].

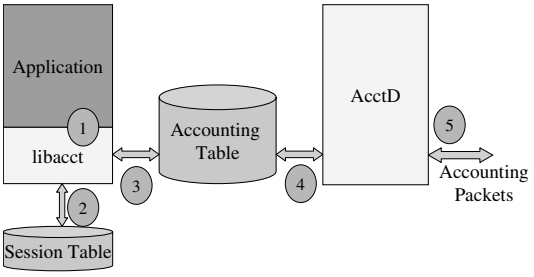

Fig. 3. IOTA accounting architecture.

Figure 3 illustrates the architecture of the IOTA accounting subsystem. The application links with a library called `libacct`. Five steps are involved for the generation of accounting messages: (1) The application triggers an accounting operation (`start` or `stop`); (2) Upon a trigger, `libacct` collects all necessary accounting information; (3) `libacct` then persistently stores the information into a table kept in the local database and returns control to the application immediately – this design makes accounting operations non-blocking yet reliable to the application; (4) A daemon, called `acctd`, periodically polls the accounting table; (5) It then formats the information into RADIUS `acct-start` and `acct-stop` messages. It also generates periodic RADIUS `acct-interim-update` messages for active sessions. The transmission of these messages to an accounting server are done in the background and may involve retries and failovers.

Regarding the reliable transmission of messages, `acctd` is configurable to use either one of the two re-transmission strategies: *loyalistic* – in which `acctd` retries a number of time on the same server before failing over to another server; or *opportunistic* – in which `acctd` fail-overs to other servers in a round-robin fashion.

F. Integrated web cache

Often, WISPs will choose to over-subscribe the back-haul link that connects their 802.11 network to the rest of the Internet. For example, while a single 802.11 access point may have a throughput of 11 Mbps, the back-haul link may be a 1.5-Mbps cable-modem link. Intuitively, a web cache placed on the hot-spot allows re-use of frequently visited web content and should save the bandwidth of the back-haul link. However, when clients access the network using Mobile-IP, in order for the web-cache to be effective, it needs to be integrated with the Foreign Agent.

Figure 4A illustrates what would happen if the web-cache was not an integrated part of the gateway. With the presence of a layer-4 switch, users' web requests to a web server get directed to the cache. In the case of a cache-miss, the cache would forward the requests to the web-server and would obtain a response. In the case of a cache-hit, the cache would already have the response in its own local disk. Either case, the cache would forward the response back to the users. However, in case of Mobile-IP service, the requests coming from the users would appear to have come from their home addresses. Therefore, the cache would forward the response back to their home networks, where the home agent would tunnel the response back to the gateway. As a result, while the cache was intended to reduce the traffic on the back-haul link, in this setup it would not eliminate any traffic even for cache-hits. In fact, the presence of the cache would double the traffic volume on the back-haul for cache-misses.

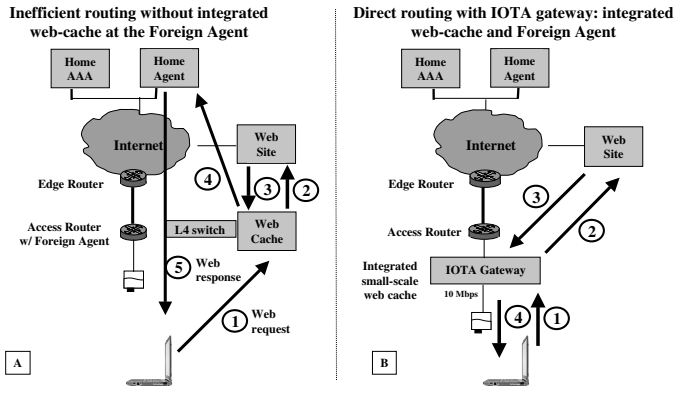

Fig. 4. The IOTA gateway integrates a web-cache, minimizing Mobile-IP overheads.

Figure 4B illustrates the scenario where the web-cache is an integral part of the IOTA gateway. When the user is registered

with the Foreign Agent, the agent uses the IPF module to add a packet-mangling rule to the per-user set of firewall policies. The rule redirects all web requests (TCP port 80) from the user to the local web cache, and all return traffic back to the user, avoiding the round-trip to the home network. Only with this integrated approach can the cache eliminate network traffic on the back-haul link for cache-hits and become effective.

G. Simple-IP operation

Although the ideal integration of 802.11 with 3G should support seamless inter-technology handoffs, short term deployments may offer an intermediate type of service, often referred to as *Simple-IP*. The Simple-IP service offers integrated authentication and billing. However, it does not support seamless mobility, and requires manual user intervention to switch network access. In this service, a session is authenticated via a web browser, while local network information such as client's IP address and default IP router is acquired using DHCP [13]. This allows the end users to access the service without any specialized software and still receive some of the benefits that we discussed thus far.

In addition to the Mobile-IP service, the IOTA gateway provides simultaneous support for the Simple-IP service. Specifically, we implemented a DHCP server and a web-based authentication system. Once the client starts up, it gets its IP address through DHCP. At the first attempt of accessing the Web, the IPF packet mangling routines redirect the client's web browser to the local authentication page served over a Secure Socket Layer (SSL) connection. Our Simple-IP authentication system, by means of the IOTA AAA server, authenticates the user to their Home AAA either with their username and password combination, or with a One Time Password (OTP) mechanism that delivers single-use passwords through the cellular Short Message Service (SMS). Upon successful authentication, the web-server uses the IPF APIs to configure the gateway's firewall according to the downloaded user policy. The IOTA gateway also supports private addressing schemes, using the NAT implementation included in the Linux IP Filter architecture.

H. Integration with UMTS

The current UMTS standards do not include support for the IETF AAA and Mobile-IP protocols. Therefore, the integration of the IOTA gateway with UMTS is somewhat more complicated than the case with CDMA2000. Although it is expected that the definition of usage for AAA and Mobile-IP within UMTS will soon become standardized [7], until then seamless inter-technology handoffs between 802.11 and UMTS networks can be handled with a Mobile-IP overlay onto the UMTS network. This introduces Mobile-IP at the GGSN, combining the Foreign Agent functionality with support for normal GGSN functionality as outlined in [8]. In this case, mobility within the UMTS network would be handled with the normal SGSN-GGSN procedures, whereas inter-technology handoffs with 802.11 networks would be handled with Mobile-IP procedures. The same client software would work for both UMTS and CDMA2000, with Mobile-IP registrations being invoked when moving under a new foreign agent (i.e. GGSN in the UMTS network). User authentication can be done through Mobile-IP procedures using a smart card (or SIM) to generate the required authenticator fields for the Mobile-IP messages. This IP-layer authentication procedure would be handled by a AAA server, either combined with or completely separate from the normal HLR functionality. Finally, an added software module would convert the generated RADIUS accounting messages into the CDR format that is required to reuse existing UMTS billing systems [4].

V. IOTA CLIENT SOFTWARE

The support of seamless mobility across 802.11 and 3G networks requires Mobile-IP client software that can work across multiple interfaces. Such a client must also intelligently select and activate the ideal interface depending on the network conditions.

Since no such client is readily available today, we implemented multi-interface Mobile-IP client software for Linux and Windows XP. Our current implementation supports Ethernet, 802.11b (also known as Wi-Fi), and CDMA2000 (Qualcomm handset and Sierra Wireless 1x-RTT card) interfaces, and is easily extensible to other types of interfaces.

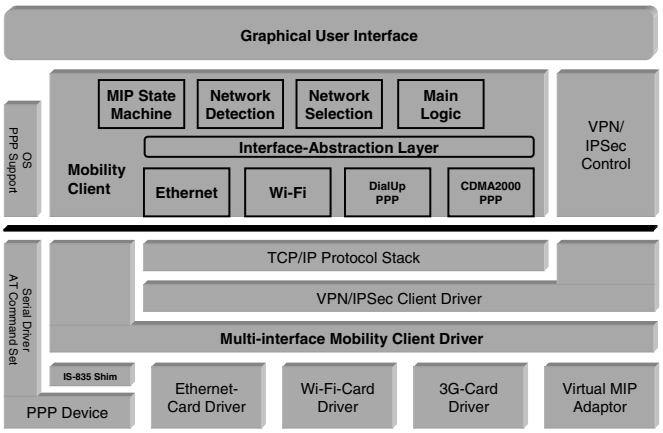

Fig. 5. IOTA Client Software Architecture.

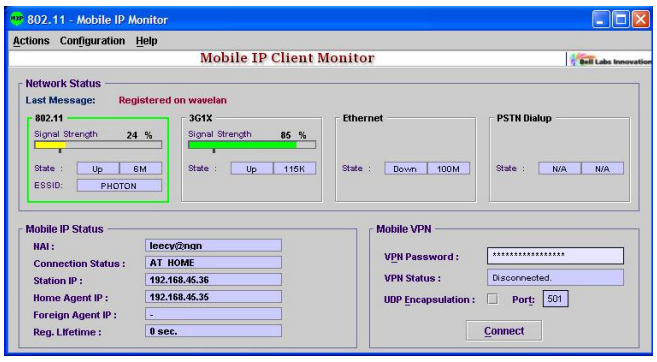

Fig. 6. IOTA Client Graphical User Interface.

The client software architecture for the Windows XP platform is shown in Figure 5[3]. We implemented the mobility client in three parts: a client GUI (Graphical User Interface) and a mobility client task in user space, and a device driver that stays below the network protocol stack in the OS kernel. The user space task includes a complete Mobile-IP stack and performs most of the mobility management. The driver offers the abstraction of a single virtual interface to the OS protocol stack. As a result, the virtual interface hides all the details about mobility from the applications, which therefore are unaware of any intra or inter-technology handoff. The mobility client task uses a driver API to monitor and select the actual network devices. The GUI allows the user to configure, monitor, and control the state of the client. By running IPSec over Mobile-IP, our solution also supports VPN (Virtual Private Network) operation that many enterprises require. Currently our client incorporates the Lucent IPSec client, but it is also expected to interoperate with other IPSec implementations. Figure 6 shows the graphic user interface of our client implementation.

A. Interface Selection Algorithm

At any given time, the client selects one of its configured physical interfaces as its *current* interface and registers with the mobility agent on that interface. To avoid data loss, it maintains association with the *current* interface while probing for an alternate *better* interface.

We designed a novel interface-selection algorithm that uses the current signal strength and the priority of the interfaces to select the active interface. The algorithm avoids unnecessary oscillations between two interfaces that may happen when their radio signal strengths are close. Four variables are considered in this algorithm: normalized signal strength, priority, low threshold, and high threshold. In the following, we denote these values as s_i, p_i, L_i, and H_i for an interface i, where $s_i, L_i, H_i \in [0, 100]$, and $p_i \in \{1, 2, 3\}$. The client periodically computes the weight w_i for each interface i, and switches to the interface that has the highest weight.
If i is the current interface,

$$w_i = \begin{cases} 1000 * p_i + 2s_i & \text{if } s_i \geq L_i \\ 2s_i & \text{if } s_i < L_i \end{cases}$$

If i is not the current interface,

$$w_i = \begin{cases} 1000 * p_i + s_i & \text{if } s_i \geq H_i \\ s_i & \text{if } s_i < H_i. \end{cases}$$

Our goal is to introduce a hysteresis effect and let the client stay with the current interface as much as possible so as to prevent oscillation. At startup, the client latches on to an interface with the highest priority and best signal strength within that priority. After that, it stays with the current interface i, until one of the following occurs: (1) the current signal strength s_i drops below its low threshold L_i, or (2) another interface j with a higher priority receives a signal strength s_j above its high threshold H_j, or (3) another interface j with the same priority receives a signal strength s_j above its high threshold H_j and s_j is more than twice the signal strength s_i of the current interface.

Although this mechanism seems to perform very well in our testbed, we are currently in the process of evaluating several variations of this algorithm, and we plan to formally characterize its performance and follow up on the study of the interface-selection problem in another paper.

[3]Here we omit the slightly different architecture for Linux.

VI. Experimental results

In this section, we report some experimental results that characterize the performance of our IOTA system. The IOTA gateways used in these experiments are implemented on servers with 800 MHz, dual-Pentium CPUs, 256 MB memory, and 9GB SCSI-II disks.

A. Performance of Mobile-IP agents

The performance of mobility management in IOTA can be characterized as the sum of two components: (1) the time needed to discover the presence of a Mobile-IP Foreign Agent on a new interface, and (2) the time needed to receive a Mobile-IP registration reply, after sending a registration request to that agent.

In Mobile-IP, agent discovery is performed through agent advertisements, which are sent by Foreign and Home agents periodically, as well as any time they receive an ICMP agent solicitation from clients. According to [14], the advertisements should be sent out at a random time (between 0 and a maximum allowed for router advertisements) after the router receives an agent solicitation. The maximum is tunable in our system, and we have set it to 500ms. On average, we observed that in our testbed clients receive advertisements 200ms after the solicitation.

After agent discovery, the time it takes for a client to register with the IOTA gateway's Foreign Agent varies depending on three possible states that the client could be in. (1) In case the IOTA system has no state information about the client, this is a *first-registration delay,* **f**, and it includes the overhead of AAA authentication, setting up packet filters, and creating tunnels between the Home and the Foreign agents. (2) The *re-registration delay,* **r**, is the time taken to re-register the client with the same gateway in an on-going registered session. This overhead includes AAA authentication, but it requires no time for tunnel or filter set up. Finally, (3) the *switching-registration delay,* **s**, is the time taken for registration when switching to an interface after the client had registered with the mobility agent on that interface at least once, i.e. when the receiving agent already had state information about the client. This includes the AAA authentication overhead, and tunnel set up at the home agent, but does not include the time taken for filter creation. It should be noted that, under the assumption of overlapping coverage of the 802.11 and 3G network, the above registration delays happen in the background and *do not* introduce any switching latency or service disruption visible

TABLE I

IOTA MOBILE-IP REGISTRATION DELAYS (ALL IN MILLISECONDS)

	FirstReg **f**	ReReg **r**	SwitchReg **s**
Ethernet	370	40	50
802.11b	410	40	60
CDMA2000	390	260	260

at application level[4].

Table I shows the preliminary results from our prototype systems. We observed that the time taken for re-registrations and switching-registrations are very small, under 60ms in both 802.11 and Ethernet, and tolerable in CDMA2000. The first-registrations times cost the most, since that involves setting up Mobile-IP tunnels as well as packet filters. We expect the first-registration procedures to complete much quicker after we optimize the filter and tunnel set up.

Adding the agent discovery delay (200ms) to the registration delays (410ms) leads to worst-case total switching times ranging from 570ms to 610ms. Such sub-second latencies shuld be more than tolerable, and would allow for seamless handoffs for moving speeds in the range of a few tens of kilometers per hour.

Finally, we also measured the re-registration time under varying forwarding load. We varied the TCP traffic through the IOTA gateway (using Ethernet) from 10Mbps to 100Mbps, using a home-grown traffic generator. The IOTA system was able to sustain close to 100Mbps forwarding load and still provide re-registration times of the order of 40-50ms.

B. Performance of QoS mechanisms

In this section, we demonstrate the performace characteristics of the IOTA rate adaptation mechanism which enables QoS guarantees. In the following three scenarios, we used three MS-Windows laptops wirelessly connected to a single 802.11 AP. On each laptop, we ran an FTP application to download a large file from an external server. We configured the back-haul connection of the IOTA gateway to be a 10 Mbps Ethernet.

The first scenario (Figure 7) illustrates restricting per-user traffic to 3.5 Mbps. At first, a single user gets 3.5 Mbps. As a second and a third users arrive, they all get equal share of the available bandwidth which is around 4.5 Mbps (which is lower than the capacity of an 802.11b cell; this is due to contention among users and uplink control traffic).

In the second scenario, we enabled the class-based configuration with Gold, Silver and Bronze classes with maximum rates of 1.5 Mbps, 1 Mbps, and 0.5 Mbps, respectively. Figure 8 shows that the QoS level of each class is maintained quite well. The slightly higher actual throughput than the specified maximum rate is attributed to the selection of token bucket parameters.

The third scenario (Figure 9) shows how class-based queuing works with a background load of $\approx$ 3Mbps. A single Gold user (max rate 1.5Mbps) is able to access all of the 1.5Mbps initially. However, as Silver (max rate 1Mbps) and Bronze (500Kbps) users arrive, the available bandwidth is shared proportionately to their maximum rate. The jittery periods are due to the rate adjustments and their length depends primarily on the IOTA rate adaptation algorithm.

[4]I.e., the overlapping coverage guarantees that there is no packet-loss during the handoffs.

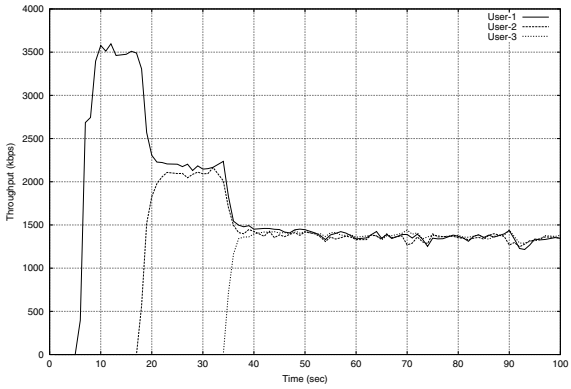

Fig. 7. QoS experiment: scenario 1.

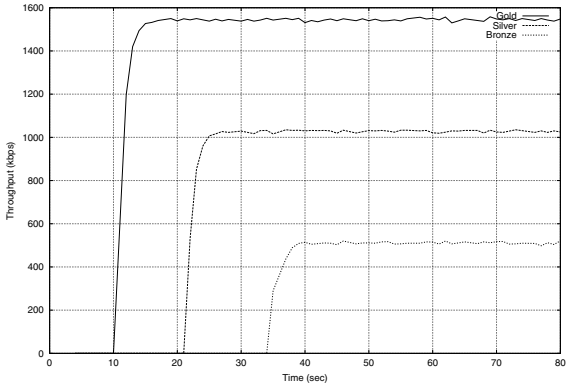

Fig. 8. QoS experiment: scenario 2.

VII. RELATED WORK

Ala-Laurila et al. proposed a solution that combines GSM/GPRS subscriber-management and billing mechanisms with 802.11 access technology[10]. They assumed that user terminals (laptops or PDAs) are equipped with GSM SIM readers and use authentication procedures similar to those in GSM/GPRS networks. They used a special protocol called NAAP that runs on top of UDP/IP to transport authentication messages. However, they did not study the use and implication of dual-interface (GSM/GPRS and 802.11) terminals. Therefore their system supports roaming but not seamless hand-off. In their model, if the two networks use two different access technologies, the user has to manually configure the terminal to use a different network interface. Finally, their system does not provide QoS guarantees in 802.11 access network and does not optimize web delivery over mobile-IP sessions.

Park studied how ISP subscribers visiting a foreign GPRS/UMTS network can authenticate themselves and use

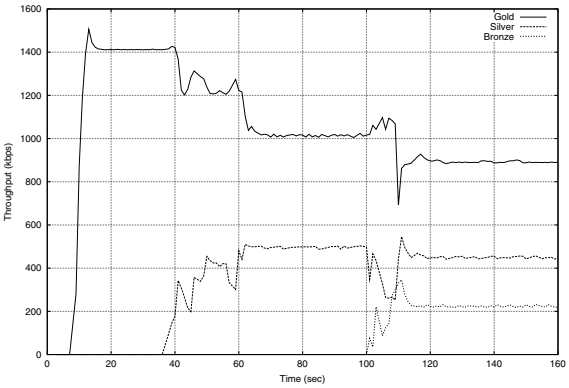

Fig. 9. QoS experiment: scenario 3.

the GPRS/UMTS network [19]. In other words, he focused on the case where the access network is a GPRS/UMTS network. The integration of 802.11 access services with cellular services was not a part of his study. In contrast, in our study the access network can be a 3G network or a 802.11 network.

Weinstein et al. proposed a scenario where 802.11 access networks complement rather than compete with cellular access networks [18]. They noticed the importance of dual-mode radios and co-ordinated AAA, but they did not address the issue of seamless inter-technology hand-off. It is not clear whether their design ideas have been implemented in a working prototype.

Brustoloni et al. proposed an architecture called *microISP* for hot-spot operators offering services in airports, hotels, etc. [12]. In their architecture, an operator leases a high-speed back-haul link to a conventional ISP, and provides high-speed Internet access to transient users using 802.11 access network. Integration of such access with 3G services is not a focus of their study. Also, in their work, there is no notion of roaming agreement, and the users are expected to settle payment individually for each of their session. In contrast, we believe that roaming agreement is an essential ingredient for a large-scale deployment of 802.11-access service.

VIII. Conclusions and future work

In this paper, we described the issues in the integration of third-generation wireless networks with local-area wireless technologies such as 802.11. We described an ideal integrated service scenario that formed the basis of our work. We introduced two architectural choices for the integration, termed as tightly-coupled and loosely-coupled interworking, and using qualitative analysis showed that the latter is the preferred approach. We described in detail our realization of the loosely-coupled architecture in the form of the IOTA 802.11 gateway and corresponding service access client software. We presented sample experimental results characterizing the performance of the IOTA system to demonstrate the validity of our design and implementation choices.

Our on-going work focuses on the modifications of the architecture to support UMTS as outlined in Section IV-H, as well as the gathering of more experimental results to complete the performance characterization of the IOTA gateway. Additional features such as the use of state-transfer protocols between access gateways for efficient inter-network handoffs and R-UIM support at the client will also be introduced in the near future.

References

[1] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. ANSI/IEEE Std 802.11: 1999 (E) Part 11, ISO/IEC 8802-11, 1999.

[2] Removable User Identity Module Standard for CDMA 2000 Spread Spectrum Systems. C.S0023-0, 3GPP2, June 2000.

[3] Wireless IP Network Standard. P.S0001-A-1, Third Generation Partnership Program 2 (3GPP2), 2000.

[4] Charging and billing: 3G call and event data for the Packet Switched (PS) domain. TS 32.015, ETSI, 2001.

[5] Local and Metropolitan Area Networks: Standard for Port Based Network Access Control. Technical report, IEEE P802.1x, 2001.

[6] General Packet Radio Service (GPRS) Service Description (Stage 2). TS 122 060, ETSI, 2002.

[7] Interworking between the Public Land Mobile Network (PLMN) supporting Packet Based Services and Packet Data Networks (PDN). TS 29.061 Version 4.5.0, Release 4, ETSI, June 2002.

[8] Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service description. TS 23.060 Version 3.12.0, Stage 2, Release 1999, ETSI, June 2002.

[9] B. Aboba and D. Simon. PPP EAP TLS Authentication Protocol. RFC 2716, IETF, October 1999.

[10] J. Ala-Laurila, J. Mikkonen, and J. Rinnemaa. Wireless lan access network architecture for mobile operators. *IEEE Communications Magazine*, pages 82–89, November 2001.

[11] L. Blunk and J. Volbrecht. PPP Extensible Authentication Protocol (EAP). RFC 2284, IETF, March 1998.

[12] J. Brustoloni and J. Garay. Microisps: Providing convenient and low-cost high-bandwith internet access. *Computer Networks*, 33(1–6):789–802, 2000.

[13] R. Droms. Dynamic Host Configuration Protocol. RFC 2131, IETF, March 1997.

[14] C. Perkins (Editor). IP Mobility Support for IPv4. RFC 3220, IETF, January 2002.

[15] C. Rigney et. al. Remote Authentication Dial In User Service (RADIUS). RFC 2865, IETF, June 2000.

[16] L. Salgarelli et. al. EAP SKE authentication and key exchange protocol. Work in progress - Internet Draft, IETF, April 2002. draft-salgarelli-pppext-eap-ske-01.txt.

[17] P. Calhoun et. al. Diameter Base Protocol. Work in progress - Internet Draft, IETF, July 2001. draft-ietf-aaa-diameter-07.txt.

[18] S. Weinstein et al. Wireless LAN and Cellular Mobile – Competition and Cooperation. Technical Talk, IEEE New Jersey Coast Section. Available from http://www.ewh.ieee.org/r1/njcoast/events/weinstein.ppt, May 2002.

[19] J.H. Park. Wireless internet access for mobile subscribers based on the grps/umts network. *IEEE Communiations Magazine*, pages 38–49, April 2002.

[20] Charles Perkins and Pat Calhoun. Mobile IPv4 Challenge/Response Extensions. RFC 3012, IETF, November 2000.

[21] C. Rigney. RADIUS Accounting. RFC 2866, IETF, June 2000.

[22] W. Simpson. PPP Challenge Handshake Authentication Protocol (CHAP). RFC 1994, IETF, August 1996.