Realizing Mobile Wireless Internet Telephony and Streaming Multimedia Testbed

A. Dutta, P. Agrawal, S. Das, M. Elaoud, D. Famolari, S. Madhani, A. McAuley, B. Kim, P. Li, M. Tauil
Telcordia Technologies Inc., 445 South Street, Morristown, NJ 07960

S. Baba, Y. Ohba, T. Kodama, N. Nakajima
Toshiba America Research Inc., Morristown, NJ 07960

Jyh-Cheng Chen
Department of Computer Science, National Tsing Hua University, Taiwan

Henning Schulzrinne
Computer Science Department, Columbia University, New York, NY 10027

Abstract

Streaming real-time multimedia content over the Internet is gaining momentum in the communications, entertainment, music and interactive game industries as well as in the military. In general, streaming applications include IP telephony, multimedia broadcasts and various interactive applications such as multi-party conferences, collaborations and multiplayer games. Successfully realizing such applications in a highly mobile environment, however, presents many research challenges. In order to investigate such challenges and demonstrate viable solutions, we have developed an experimental indoor and outdoor testbed laboratory. By implementing standard IETF protocols into this testbed, we have demonstrated the basic functionalities required of the mobile wireless Internet to successfully support mobile multimedia access. These requirements include signaling, registration, dynamic configuration, mobility binding, location management, Authentication Authorization and Accounting (AAA), and Quality of Service (QoS) over a variety of radio access network (RAN) technologies (e.g. 802.11b, CDMA/GPRS). In this paper, we describe this testbed and discuss important design issues and tradeoffs. We detail the incorporation and inter-relation of a wide catalog of IETF protocols - such as SIP, SAP, SDP, RTP/RTCP/RTSP, MGCP, variants of Mobile-IP, DRCP, HMMP, PANA, and DSNP - to achieve our goals. We believe that the results and experiences obtained from this experimental testbed will advance the understanding of the pertinent deployment issues for a Mobile Wireless Internet.

I. INTRODUCTION

Streaming multimedia is gaining momentum as a killer application for the next generation of telecommunication services. Efficiently providing flexible and programmable multimedia services at low cost has been a main motivation behind the transition from traditional circuit switched networks to packet based networks. Many of these transitional efforts have focused on wired networks. As personal communication and ubiquitous access becomes more prevalent, however, it is necessary to develop robust solutions that can support mobile, wireless streaming applications, such as IP-telephony and video conferencing. Supporting multimedia over wireless and mobile links requires the consideration of several factors, such as signaling, registration, configuration, quality of service, bandwidth management, mobility management, and authentication, among others.

We have designed a comprehensive testbed, based on a suite of standard IETF protocols, where we have implemented the necessary functional components to support mobile multimedia in a next generation wireless network. Where possible, we have implemented alternative approaches and have presented performance analysis to gain insight into the pertinent design questions surrounding mobile multimedia. The full suite of implemented or emulated protocols is shown figure 1.

In [1] we described a basic framework for a wireless Internet testbed. In the present work, we elaborate on this basic framework and discuss the comprehensive design of an extended multimedia testbed. Some of the important features of this testbed include mechanisms to support seamless roaming across different carrier domains, billing and network man-
agreement features, quality of service support, as well as IPv6 compatibility. The testbed also demonstrates integration of PSTN-based networks with cellular-based networks by means of interaction with signaling protocols such as MGCP [2], H.323 [3], and gateways based on SIP [4]. In addition, the testbed supports localized streaming services for mobile users by leveraging wireless multicast in the local domain.

Figure 2 shows a Mobile Wireless Internet Telephony architecture in a heterogeneous network environment upon which the multimedia testbed is built. It shows a scenario where a mobile starts from a home network and moves away to a visited domain while communicating with the corresponding host. In the process the mobile is subjected to cell, subnet and domain handoff. MAAAQ stands for mobility agent, AAA and quality of service. SLA is the service level agreement between two domain control agents. IDCA is the Inter Domain Control Agent that acts like a broker agent between domains.

This paper is organized as follows. Section II discusses the functional requirement of the architecture and describes how these requirements are realized. In section III we describe the testbed architecture and define some of the hardware and software components associated with the testbed. Section IV describes details of several functional features implemented, associated sequence of operation and performance analysis for each feature. Section V concludes the paper with some discussion.

II. TESTBED FUNCTIONAL REQUIREMENT

Many of the design features in the proposed Wireless Internet Telephony Testbed are motivated by a number of recently proposed architecture for 3G/4G type networks such as 3GPP [5], 3GPP2 [6], MWIF [7]. MWIF has recently merged with Open Mobile Alliance because of its relevance with Mobile Internet. References [8] provides a testbed involving wireless and wired LAN systems but does not discuss many of the signaling issues related to Internet Telephony. Reference [9] provides an analysis between Internet Telephony and Wireless Telecommunication Networks but has not focused on subnet or domain mobility. Design criterion of our multimedia testbed have taken into consideration many of the issues and requirement described in [10]. Some of the most important design criterion that are different from the available architectures in the literature are: application layer techniques that make it more attractive for the application service providers; policy based mobility management that provides a flexibility of choosing a specific technique based on the application type; added capability of IP and PSTN interaction using softswitch with mobility features.

Our testbed emulates a wireless Internet and is comprised of Linux, Windows and Solaris platforms. The following standard IETF protocols are implemented as software components: SIP (Session Initiation Protocol), and SDP (Session Description Protocol) [11] for signaling and media description respectively, RTSP [12] (Real-Time Streaming Protocol) for multimedia streaming; Mobile IP [13] or its variants [14], HAWAII [15], TeleMIP [16] for address binding; DHCP [17] or its variant Dynamic Rapid Configuration Protocol [18] for client configuration; Diffserv based protocols for QoS such as DSVPN (Dynamic Service Level Specification Negotiation Protocol) [19]. Wireless connectivity is provided by a variety of heterogeneous access networks including Bluetooth, IEEE 802.11b, CDMA and GPRS. Wireless connectivity spans over both internal laboratory testbed and external testbed using commercially available carrier networks.

Following describes a list of functional requirements that have been realized in the testbed.

A. Signaling for the multimedia clients

This architecture is built upon the vision of next generation wireless networks where clients are IP end-points. Additionally, we implement a softswitch in our testbed to account for possible transition scenarios and migration paths where IP and non-IP end-points must communicate. Because of the distributed nature of these networks, the Session Initiation Protocol is a natural choice to provide signaling between clients. SIP is used in the initiation, and tear-down of multimedia calls and SIP servers and user agents are key components of our signaling architecture. SIP functionality can also be integrated into a call agent for demonstrating IP-PSTN call features. While SIP was designed primarily to handle signaling for multimedia calls, there has been numerous proposals to extend SIP to handle mobility as well as discussed later in the paper.

B. Handoff

Hand-off allows an established call/session to continue without interruptions when a mobile station (MS) moves from one cell to another. Successful hand-off requires registration,
configuration, dynamic address binding, and location management functions. The hand-off process must perform the necessary AAA to ensure the integrity, authenticity, privacy, and confidentiality of a user’s location. Furthermore, hand-offs should maintain QoS requirements whenever possible, striving to meet the delay requirements of real-time applications while minimizing lost data.

In end-to-end wireless IP paradigm, three logical levels of hand-off procedure can be defined:

i. Cell Hand-off (Micro): It allows an MS to move from one cell to another in a subnet within an administrative domain. One subnet may consist of multiple cells. IP address of the mobile host remains same in this case.

ii. Subnet hand-off (Macro): It allows an MS to move from a cell within a subnet to an adjacent cell within another subnet that belongs to the same administrative domain.

iii. Domain hand-off (Domain): It allows an MS to move from one subnet within an administrative domain to another subnet in a different administrative domain.

Each of the above levels of handoff has been prototyped in the testbed and performance measurements have been recorded.

C. Dynamic Configuration

Dynamic configuration involves providing an end-terminal with the appropriate information to participate in the network. This may mean supplying such parameters as an IP address, as well as the location of support servers such as DNS and SIP proxies. There are several standard ways of registration in IPv4 networks, including DHCP [17], PPP [20] and Mobile IP [13], while IPv6 supports stateless auto-configuration. DHCP is the primary method of configuration in wired networks, however, it does not perform well in wireless networks due to excessive latencies. The Dynamic Rapid Configuration Protocol [18] is a lightweight version of DHCP that performs configuration faster and more efficiently, making better use of scarce wireless bandwidth. It does so by shrinking the message size, minimizing the number of messages in transaction and limiting the use of broadcast. DRCP performs subnet discovery by detecting link-layer channel changes or by server discovery, similar to router discovery in Mobile IP.

D. Location Management and Registration

Registration is a process by which a network is made aware of the existence and location of a mobile station and its associated user. This process comprises several steps such as sending a registration request from the MS to the network, registering with the SIP server with the new IP address, performing an AAA process by the network for proper authentication, and sending appropriate responses to the MS as well as location management entities to ensure that the network is aware of MS’s current location.

It is important to establish security associations between MSs and the networks that serve them. There have been a few AAA protocols created to handle these security associations as a client moves between subnets within a domain. Generally, the Home AAA server or an intermediate broker agent (SIP Central Point of Contact [21]) must be contacted when the user moves into a new domain for the first time to establish its credentials. It is important to complete the registration process in a timely manner during the handoff process, and thus it is critical to minimize the time spent creating new security associations.

E. Mobility Binding

Binding allows MSs to maintain their TCP and UDP streams without interruption when they move. Binding is typically handled by Mobile IP [13] or one of its variants [14], [15], [16]. Native version of Mobile IP however suffers from several drawbacks such as triangular routing and encapsulation. MIP with Route Optimization and MIPv6 however address these shortcomings. We have implemented several versions of Mobile IP [22], [23], [24] in a co-located mode while interacting with DHCP or DRCP. We have also implemented an application layer mobility management technique based on SIP [25] to handle personal, and terminal mobility for streaming application that are typically RTP/UDP based. Such SIP based mobility techniques provide an alternative approach to Mobile IP for mid-session mobility, however, they have difficulty supporting TCP applications. Such binding support for TCP application though is provided by the the Host Mobility Management Protocol (HMMP) [26], a framework built on SIP-extensions capable of binding TCP sessions and [27] a SIP based mobility proxy.

F. Authentication, Authorization and Accounting

The testbed implements AAA through Diameter [21] running on Network Access Servers (NAS) and AAA servers. In addition, we are developing and prototyping a new protocol called Protocol for carrying Authentication for Network Access (PANA) [28]. PANA is implemented as a user level protocol to enable a flexible access control independent of underlying link-layer technology and configuration protocol.

G. Security

Supporting mobile multimedia requires a multi-layer security design including user-based access control, packet encryption and end-to-end security for both signaling and media traffic. Such a multi-layer security framework has been implemented in the testbed and has been shown to interoperate with both MIP and SIP-based approaches to mobility. In the SIP case, clients use PGP based authentication while registering with the SIP servers. Additionally, PANA provides user-based authentication between the mobile client and the first hop access router and works in conjunction with an AAA server to provide authentication. Packet encryption is provided over the air by using IP-Sec [29] mechanism between the mobile client and the edge router and Secured RTP [30] for end-to-end encryption. Lastly, layer two port-based security is provided by 802.1x.
III. TESTBED ARCHITECTURE COMPONENTS

Figure 3 shows the proposed testbed along with its hardware, software and protocol components. We elaborate on these components in the sections below.

A. Hardware Components

A.1 Mobile Station (MS)

Mobile stations are multimedia laptops and PDAs capable of running multimedia applications such as (wb, vic, rat). These devices connect to the available access networks through either built-in support or PCMCIA interface cards. These terminals are primarily Linux-based, however, the testbed also supports Windows clients.

A.2 Base Station (BS)

Base stations provide last-hop connectivity to mobile users. Our testbed includes several types of base stations including Bluetooth, 802.11b and CDMA/GPRS based access points. While the CDMA/GPRS access points are controlled by commercial service providers, the other access points are fully within our control. They are SNMP enabled, which provides increased manageability and helps in providing location-based services. Additionally, we have built outdoor components to interface with our indoor testbed. Figure 4 shows a roof top yaggi-array antenna that provides coverage up to two miles.

A.3 Radio Access Network

The Radio Access Network represents the wireless and back-haul infrastructure that provides mobile stations with access to a greater regional, wireline IP backbone. The IMT-2000 standards envision multiple RANs providing access to a variety of link-layer technologies. Our testbed attempts to emulate this condition by integrating heterogeneous access technologies.

A.4 Base Station Controller (BSC)

Base Station Controllers are multi-port switches (e.g. VLAN switches) that can control base stations connected to each of its ports. Such VLAN switches may institute the IEEE 802.1p scheme for providing class of service for multimedia traffic to end clients. BSCs are also capable of filtering multicast packets, and can prioritize traffic destined for particular mobile hosts, thus providing QoS mechanisms in terms of bandwidth.

A.5 ERC (Edge Router and Controller)

An ERC is a routing and control system that connects a wireless access network to a regional wireline IP network. A single ERC may support several RANs. An ERC comprises two functional entities, an edge router (ER) and an Edge Control Agent (ECA). The ER functions as an IP router, while the ECA is an intelligent agent that interacts with a Domain Control Agent (DCA) (described below) to control the RANs and support necessary network-wide control tasks. Control Agent functionality can be distributed within a domain.

In the testbed, ERCs are Linux PCs acting as routers with multiple ethernet interfaces. There are multiple base stations connected to the interface of each router via Cisco’s VLAN switches. In this architecture an ERC provides more functionality than just a router. The ERC also runs many of the server and client software such as the PANA server, the Diameter client, QoS Local Node (QLN), IPSec, DRCP, and Home Agent. In Figure 3 the four ERCs illustrate the agents that define the controller part of ERC.

B. Protocol/software Components

B.1 Domain Control Agent

The domain control agent (DCA) provide session management and allows users to interact with network control systems and entities. Additionally, the DCA supports 1) mobility management, 2) authentication, authorization and accounting and 3) QoS management. As an implementation option,
each administrative domain may distribute the domain control agent features across the domain. Parts of domain control agent are described later on.

B.2 SIP server/SIP User Agent

A SIP user agent runs on all MSs. The client version of SIP is implemented with tcl/tk and C code, and provides a user interface for managing multimedia calls. A SIP server provides proxy/re-direct and registration functionality. A multimedia call between two wireless clients can be established using either direct mode or proxy mode. Additionally, SIP User Agents and servers also provide personal and terminal mobility support. The testbed implements a multi-layered mobility management approach, such as [31], where RTP/UDP based applications are handled via SIP signaling.

B.3 DRCP server/client

DRCP servers are responsible for leasing IP addresses to clients. The testbed contains a DRCP server within each subnet which will provide moving clients with IP addresses, if necessary. A DRCP daemon runs on each client and interacts with the 802.11 driver, Mobile IP and SIP user agent during the handoff process. The daemon will request a new IP address upon boot up and after it learns that the terminal has entered a new subnet. Additionally, DRCP provides a fallback mechanism and will default to DHCP in the absence of DRCP servers.

B.4 Mobile IP Server/Client

In our testbed, we use Mobile IP primarily to support mobility for non-real-time traffic, such as TCP. We have incorporated a variety of available Mobile IP implementations, including MosquitoNet [22], SUN [23], and Dynamics HUT [24]. Each of these implementations supports co-located care-of addresses, which are provided by the DRCP servers in our testbed. After moving into a new subnet, the mobile terminal obtains a care-of address from the DRCP server and sends a registration message to the Mobile IP home agent.

B.5 Application Servers

The testbed is comprised of a number of application servers responsible for executing specific software functionality. In particular, we employ a Real Stream server, an Apache HTTP server, an IMT-2000 emulator and a Mobile Application Server (MAS). As a practical matter, these servers can all reside on the same hardware.

In the absence of an in-house CDMA 2000 based infrastructure, IMT 2000 QoS functionality has been emulated. Specifically, we use the IMT-2000 emulator performs the following tasks; assigns priority to particular kinds of applications (signaling, data transfer, ftp, http), provides variable speed transmission (e.g., 14.4 kb/s, 384 kb/s, 2Mb/s), and enforces variable terminal states (active, dormant). This emulation service allows our testbed to deliver differentiated service quality to meet client requests and thus better approximates the conditions of an IMT-2000 based system.

The MAS is a Java middleware platform for mobile services. It provides the basic building blocks required to provision mobile wireless applications, such as user profile management, location handling, device detection, content adaptation and e-wallet functions. The MAS is modular, flexible, and provides an open API for 3rd party service creation. Users can thus utilize a variety of access technologies (CDPD, cellular, 802.11, SMS, etc.) and content delivery mechanisms (HTML, WML, VoiceXML, etc.) while sparing application developers the details of each of the underlying approaches. The MAS can also interface with traditional PSTN elements such as PBX-es, A/IN elements such as SCPs, as well as NGN elements such as Call Agents.

IV. FUNCTIONAL FEATURES AND PERFORMANCE

Following subsections describe some of the features that can be demonstrated in the multimedia testbed.

A. Multimedia Session Establishment

Our testbed supports multimedia session establishment through SIP. The correspondent host (CH) initiates a multimedia call with the mobile client and they negotiate their capabilities by means of the Session Description Protocol (SDP). This multimedia call can be set up in direct mode or proxy mode. The proxy mode utilizes a SIP proxy server that provides additional security and authentication features. Once both parties agree on a set of media preferences, the call is established. The call may involve any combination of voice, video and data. SIP only assists in the call setup phase, the actual RTP/UPD media flows delivered between two end-points using standard routing mechanisms.

We have used a white board (wb) application to demonstrate data, (RAT) to demonstrate voice, and (VIC) to demonstrate video communication. In addition, we use our IMT-2000 emulation software to assign priority among the multimedia and signalling flows.

B. Seamless Terminal Mobility

Terminal Mobility can be categorized broadly into two kinds: pre-session mobility and mid-session mobility. Pre-session mobility can be handled by pre-registration mechanisms. Mobile IP and DHCP/DRCP protocols provide traditional ways to handle pre-session mobility at the network-layer. In our testbed, however, we use an application-layer technique enabled by SIP that provides pre-session mobility by means of registration and re-direction using a unique URI.

Mid-session mobility, as discussed previously, provides continuous binding and can be sub-divided into three different categories such as, micro, macro, and domain categorized by the type of movements of clients. We employ both network-layer (Mobile IP) and application-layer (SIP) techniques to support mid-session mobility.

Micro-mobility is handled by the link-layer. Thus, neither
mobile IP nor SIP must be involved to maintain session continuity of the multimedia stream.

Macro mobility involves the mobile client moving to a different subnet. The client discovers it is in a new subnet (via advertisements), which triggers a DRCP DISCOVER message. The mobile client now gets configured with a new co-located IP address by the DRCP server. When using Mobile IP, the client sends a registration message to the home agent with the new IP address. Data now gets tunneled through the home agent and decapsulated at the mobile client. Using SIP, however, the mobile issues a re-invite to the corresponding host, which simply redirects the traffic to the new address.

Figure 5 shows SIP signaling messages and RTP flows associated with a typical hand-off sequence due to terminal mobility.

Figure 5. SIP Re-Invite Handoff Sequence

Domain mobility has been demonstrated by configuring a separate DNS (Domain Name Service) and AAA domain. As part of domain mobility Mobile Station has to interact with visited AAA server, home AAA server, home SIP server and PANA server during its movement. A complete sequence of secured domain mobility and associated performance has been shown later in the security section. Experiments show that several multimedia application (audio, video, and data) do not have any discontinuity problems even if the cars move at a speed of 45 mph. Dynamic DNS is implemented to provide the dynamic DNS update as the IP address of the mobile changes.

A comparison of the latency for both MIP and SIP based mobility management with various packet sizes is shown in Figure 6. It shows that SIP based terminal mobility offers better performance comparison compared to MIP for both data (D) and signaling and data combined (SD).

C. Fast and Optimized handoff methods

Fast handoff of multimedia streams reduces transient data loss and latency. Fast handoff mechanisms can be applied at different stages of a mobility management framework. It is typically dependent upon speeding up the following mobility stages: detection of link-layer signals, registration with the new subnet/domain and re-direction of the stream after detection of the new subnet. The standard version of DHCP takes about 5-15 sec with ARP (Address Resolution Protocol) checking [32] to complete a handoff. Suppressing the ARP checking saves roughly 5 seconds. DRCP, however, it takes only roughly 100 msec. This reduced latency results because of the factors mentioned in the earlier section.

While DRCP takes care of faster configuration, there are several approaches to take care of re-direction of transient data traffic as in [33] [34] [35] [36] [37] [38]. We are also implementing IDMP-based [37] and SIP-based [38] fast-handoff approaches as network layer and application layer alternatives respectively. Figure 7 shows DRCP setup time as multiple clients obtain IP addresses from the server and figure 8 shows the packet-loss-ratio gain due to the SIP-based fast-handoff technique.

D. Mobility-aware Quality of Service

To provide quality of service to mobile clients roaming between subnets and domains, we use the Dynamic SLS negotiation protocol [DSNP], which is based on Diffserv approach [39] but uses two levels of hierarchical servers within a domain (subnet level and domain level). The higher (Domain) level server keeps track of the QoS requirements and distributes this information to other routers in the domain where the traffic is shaped accordingly. Details of this mechanism can be found in [19]. Both incoming and outgoing traffic can be shaped at the edge router (ERC) before being delivered. For real-time (RTP/UDP) traffic, RTCP feedback [40] can be used to provide QoS requirements to the domain level.
server. The RTCP based feedback approach, in conjunction with Diffserv, provides an application layer solution for the mobile environment. Figure 9 shows throughput results. Several handoff sequences are shown when the mobile moved between different access points across two different domains and subnets. Rate decreases during handoff due to packet loss. Rate reaches peak after handoff or SLS change due to shaper initialization. SIP and DSNP has been integrated in the testbed thus making sure that the signaling and RTP packets are assigned proper priority after the mobile changes its point of attachment.

In addition, a proxy based multicast solution has been realized in the testbed that provides flexible streaming services [44]. Here join and leave operations are handled at the application layer using RTCP feedback. In this case, each ERC acts like a localized server capable of doing application layer scope-based multicasting and provide localized advertisements.

**E. Localized IP Multicast**

IP multicast has been deployed in the testbed to achieve better bandwidth efficiency, facilitate mobile content distribution and provide flexible streaming services such as traffic incident alerts and localized advertisements. There are many issues related to multicast mobility (e.g., maintaining group membership while changing cells, subnets and domains). Underlying multicast support has been realized using DVMRP and PIM along with a set of Cisco and Linux based routers running mrouted.

SIP signaling can take advantage of IP multicast to invite a group of people to a conference, invite a stream server to a conference, move from a two party conference to a multiparty conference and help provide virtual soft-handoff for unicast streams. QoS for real-time multicast traffic is being handled in the testbed by an extension of DSNP using RTCP based feedback approach [41]. A variety of streaming services have been experimented with multiple servers across subnets. Some of the proactive mechanisms proposed in [41] and [42] to reduce the join and leave latency have been implemented in the testbed as well. As an alternative arrangement UMTP (UDP Multicast Tunneling Protocol) [43] has also been implemented to handle non-multicast enabled networks using application layer tunneling.

In addition, a proxy based multicast solution has been realized in the testbed that provides flexible streaming services [44]. Here join and leave operations are handled at the application layer using RTCP feedback. In this case, each ERC acts like a localized server capable of doing application layer scope-based multicasting and provide localized advertisements.

**F. Softswitch/SIP-PSTN integration**

In addition to providing mobile multimedia support between IP end-points, this testbed also provides a way of integrating with PSTN components by using a call agent (Media Gateway Controller) [2] and a SIP server. The Call Agent is based on the Media Gateway Control Protocol, where the Media Gateway Controller is resident on a server, and controls the non-IP devices connected to the gateway. This gateway is usually a media gateway, whose interface is connected to the IP cloud. The other interface connects to a standard analog phone, or a PBX. The media gateway converts the analog signal to an IP stream, and has an MGCP slave agent which is controlled by the Media Gateway Controller. In this testbed, the gateway is actually a Cisco router (2600) with an FXO (Foreign Exchange Office) board that connects to a standard analog phone. The call agent maintains a database of the PSTN end-points, and has access to an intelligent database that provides AIN (Advanced Intelligent Networking) functionality. In order to provide scalability or domain connectivity there can be multiple instances of call-agents. The protocol interaction between the call-agents can be based on SIP-T, an open standard protocol currently under discussion within IETF. SIP to Analog phone integration has been realized by using a pair of Mediatrix gateway [45], analog with wireless phones and a windows-based SIP server that keeps a mapping between the end points. In this scenario each mediatrix box has a SIP user agent installed, which generates and terminates the SIP signal.

A standard Call Agent (softswitch) and SIP server configuration has been shown in figure 10. This prototype shows the possible interaction between PSTN, analog phones and SIP based IP phones (software and hardware) using call agents and SIP-PSTN gateways.

Wireless call agent and terminal mobility have been real-
ized by making the residential gateway mobile and using Mobile IP. Application layer mobility is possible by adopting a SIP based mobility management where SIP re-Invite can be sent to the SIP-PSTN gateway.

G. Mobility over Heterogeneous Access

In addition to 802.11b, our current testbed has been extended to cover heterogeneous access networks such as CDMA 1XRTT and GPRS based networks provided by commercial carriers such as Verizon and Voice Stream. A separate DMZ (De-Militarized Zone) has been setup so that specific signals (MIP bindings and SIP registration) and RTP data stream both RAT (Robust Audio Tool) and VIC (Video Conferencing) on specific ports can pass back and forth between the commercial network and the 802.11 based enterprise networks as the mobile moves between the two. The triggering decision for Mobile IP binding, or SIP Re-Invite, is based on the signal-to-noise ratio, network congestion or other cost factor observed by the interface. It is important to mention that DHCP is used to provide the IP address to the mobile in the LAN environment where as PPP is being used when the mobile gets connected to the Verizon or Sprint network. PPP connection has an added overhead delay for setting up the connection. Thus switch over from LAN to heterogeneous network takes more time than switching back. From experimental evaluation it was found that it took about 5-10 sec while switching from 802.11 to CDMA network but took under 1 sec while switching from CDMA to 802.11 network. Using a make-before-break approach, however, means that both interfaces are active at the same time and handoff latencies do not translate into lost data.

Figure 11 shows the handoff performance on the mobile that uses SIP based mobility management as the mobile moves back and forth between cellular (CDMA1XRTT) and 802.11b network. The handoff values may differ if a Mobile IP based approach is used.

H. Mobility Support for IPv6

Our testbed has been enhanced to support SIP based wireless telephony features over IPv6. Linux kernel version 2.4.9 with patch from USAGI projects [46] is being used in the routers and linux hosts. Both SIP based mobility and MIPv6 have been experimented with. MIPL Mobile IPv6 for Linux [47] is adopted to support mobility in the testbed. Several experiments were carried out for real-time communication including analyzing the effect of Duplicate Address Detection (DAD) [48] in the disruption of SIP based multimedia calls. SIP user agent has been modified to support audio and video calls using RAT and VIC over IPv6 based network. A complete handoff analysis for SIP and Mobile IP based mobility has been shown in [49].

During the experiment with Mobile IPv6 a delay of 1900 ms was observed with DAD (Duplicate Address Detection) and 1.5 ms without DAD during movement between subnets [49]. However when the mobile node comes back to its home network within a specific period of time, delay associated with DAD becomes negligible. Home agent usually caches the binding association and proxies on behalf of the mobile. Thus it protects its address from other nodes. SIP based mobility in an IPv6 environment offers comparable handoff delay but is limited by the processing speed on the end hosts.

I. Multi-layer Security Framework

A flexible multilayered security framework has been implemented using SIP, AAA, DRCP, PANA, Diameter, IPSec protocols and set of IP firewall rules. The following describes some of the security models in detail.

I.1 SIP-AAA Model

We implemented a SIP-AAA interaction model to explore how SIP signaling can interact with AAA infrastructure in a mobile environment. In this model, when the SIP server receives a SIP Register message, it consults with the home AAA server for authentication and authorization by using Diameter protocol. The database of the SIP user account has been replicated in the home AAA server for authentication purposes.

SIP registration is usually done in the SIP server after the client obtains a new address. Our SIP-AAA interaction provides a mechanism so that a communicating user’s activities are monitored securely for accounting and auditing purposes. In this model, SIP registration is authenticated only after consulting with AAA. Besides authentication via the home SIP
server and home AAA server, a user is also authenticated via interaction between a local AAA server and a home AAA server using Diameter.

I.2 PANA-AAA Model

PANA offers user authentication for network access within an AAA (Diameter) framework. The ERC in which the PANA server resides maintains an association between the user identity such as an NAI (Network Access Identifier) and lower-layer identity such as an IP address for each user. The ERC also has firewall functionality such that only packets sent from/to the authorized users can pass through. Since the association between the user identity and lower-layer identity dynamically changes as a result of hand-off, the ERC updates the access control list of the firewall if and only if there is a change in the association and the resulting PANA registration, or a local authentication, is successful. This means any signaling message such as SIP Register or Re-invite messages will not pass through the firewall until the access control list is updated.

This model provides application layer authentication based on user NAI. When the MS moves into a new domain, it performs a PANA (Protocol for carrying Authentication for Network Access) [28] registration with the PANA server in the domain. Since the PANA server has no pre-established security association with the MH at the time of PANA registration, the PANA server consults with the home AAA server directly or indirectly through a local AAA server by using Diameter protocol to authenticate the user on the MS. Once the PANA registration is successful, a Local Security Association (LSA) is established between the MS and PANA server. Intra-domain hand-off can then be performed locally and quickly at the PANA server without contacting the home AAA server. Local authentication is also performed periodically in order to detect when the user silently disappears from the domain due to, e.g., battery exhaustion or bad radio conditions.

It is possible to combine PANA with various kinds of access control. In the testbed, PANA is used to dynamically control a router firewall so that full network access is authorized only for hosts associated with authenticated PANA clients. If a PANA re-authentication fails, any previously opened holes in the firewall will be immediately closed.

I.3 Packet Encryption

Packet-based encryption protects information both over the last hop and end-to-end. Although 802.11b provides WEP (Wired Equivalent Privacy) based encryption scheme, we have used IPSEC to secure the packets on the last hop wireless networks in order to make it link-layer independent.

We use PANA for distributing IKE credentials to an authorized host. When the host is authorized as a result of PANA authentication, the IKE credentials are carried in a PANA message and are transferred from the PANA authentication agent to the host. The credentials are then used for establishing an IPsec tunnel between a host and an access router, which provides a secure unicast communication channel in the access network. The dynamic distribution of the IKE credentials enables hosts to roam among different administrative domains since there is no need to pre-configure the credentials. When the mobile attaches to a new subnet, a new IPsec tunnel is established between the mobile host and the new edge router. This IPsec tunnel secures the signaling and data only on the last hop in the new domain.

While an IPSEC based encryption mechanism helps secure the last hop wireless channel, it is essential to also provide end-to-end security. Since the Real-time traffic is RTP/UDP based, secured RTP (SRTP) [30] is used to provide encryption to different types of multimedia traffic such as audio, video and data. Setting up a secured RTP session involves exchanging the RTP key between the mobile host and correspondent host for each multimedia instance. The RTP key exchange takes place over SIP by exchanging INVITE messages. Since the RTP key is part of the SDP parameters, it is protected using S/MIME (Secured MIME) [50]. This ensures that both signaling and data can be secured end-to-end.

J. Secured Multimedia call Sequence

Many different protocols must interwork across several layers to demonstrate secured and seamless mobility. Figure 12 provides an operational flow as a mobile moves from domain to domain.

After being configured with an IP address by DRCP, the MS provides the PANA agent with the appropriate NAI to open the firewall controlled by the ERC. Then the CH initiates a SIP call to the MS using a SIP proxy server. The MS moves towards another domain and experiences a domain handoff (domains are segregated as AAA domains not DNS domains in this case). During inter-domain mobility the entire suite of testbed protocols interact with each other in a sequential manner to provide registration, configuration, user authentication, profile verification, signaling, personal mobility, mobility binding and security features. While PANA provides a session based authentication, the Freeswan version of [51] IPsec has been implemented between the client and the first hop router to provide packet based security. Packets get detunneled beyond the ERC1 and tunneling takes places again between ERC2 and MS after the handover. In case of Mobile IP binding, mobile node can interface with IPsec and firewalls that get triggered by PANA agent. In case of SIP mobility SIP-Revite signal gets delayed a bit because of the IPsec tunneling setup.

K. Secured mobility measurement

We took measurements of the timing associated with each atomic operation for the various mobility scenarios. Several tools such as tcpdump, rtptools [52], ethtereal [53] were used on the mobile host and correspondent hosts, and the output was analyzed for specific messages associated with different ports. This allowed us to determine the timing for each of the operations associated with signaling, moving between cells,
triggering to obtain an IP address, sending the PANA messages, interacting with the Diameter server and SIP server, interacting with both AAA servers during the domain handoff, and Mobile IP registration or SIP Re-Invite with IP-Sec tunnels.

As an example for initial SIP based call set up a typical INVITE message was about 455 UDP bytes, ringing was about 223 bytes, OK takes 381 bytes, ACK was 216 bytes, REGISTER and its OK messages were about 370 bytes and 412 bytes. Subsequent de-registration and re-registration messages were of 372 and 425 bytes respectively followed by OK messages which were of size 510 and 410 bytes. A typical Re-Invite after subnet change and respective OK messages were 450 and 380 UDP bytes respectively. This handoff delay constitutes several components at different layers including 802.11b channel change, subnet discovery, IP address acquisition, local authentication by means of PANA, and delay due to SIP Re-invite. A complete Re-invite, OK and ACK sequence took about 500 msec including the processing time at the end hosts, but CH could start forwarding the data to the MS as soon as it receives the Re-Invite message thus helping to reduce the time for media redirection by about 350 msec, since this will eliminate the timing associated with ACK and OK. Address acquisition by means of DRCP is done within 100 msec, but it does not include the extra time needed to check the channel change or periodic DRCP server advertisement. SIP re-registration does not affect the media re-direction to the new address, since it is independent of the Re-Invite process and is used mostly for location management. A typical complete registration process so that the client’s new IP address gets updated in the SIP server is about 150 msec. From the experimental results it was observed that it may take up to 1 sec from the time it lost connectivity with the old access point until the host gets configured with the new channel number associated with the new access point. As the MS binds to the new access point and listens to the server advertisement, DRCP Discover process sets in by the client. According to [54] a typical beacon interval from an access point is about 100 msec that contributes to the layer 2 delay. It is normal to assume that rest of the time is used to process the beacon and set up the channel number in the application before a layer 2 association is established. Table 1 shows the timing associated with different functional components during subnet and domain handoff measured in number of seconds. Both Average values (denoted as Ave) and standard deviation denoted as (SDV) are shown in seconds. As is evident domain handoff (denoted as D) takes more time than subnet handoff (denoted as S) because of associated AAA interaction.

It is noteworthy that these performance parameters strongly depend on media used, number of hops, authentication mechanism used, processing speed of the correspondent and mobile hosts and background cross traffic. Studies [55] show that voice conversations can tolerate up to 2 percentage packet loss and 200 ms one-way delay while still delivering acceptable quality. Tolerance value is higher if it is a streaming audio or video instead of two way conversation. Reference [56] states a hard call setup delay limit of 2 seconds for PSTN/Internet telephony interworking and 1.6 seconds for SIP based call setup. It is interesting to note that values obtained during our experiment for call setup and packet loss during handoff fall within the performance bounds.

In the experiments overall timing for RTP packet interruption due to DRCP, PANA and SIP amounts to about 1 sec. As it turns out, the bulk of the time is consumed to process SIP signaling (Re-Invite, ACK, OK) messages that amount to 600 msec. The extra time is due to the security association involved during domain hand-off. With proper optimization
this can be reduced to 200 msec, including the processing
time on the end hosts. These values may be different if the
correspondent host is also moving since the calls will then be
proxied. In optimized mode, the total interruption due to sub-
et and domain handoff would be limited to 400 msec, with
domain handoff taking slightly more time because of AAA
interaction. Secured mobile communication has been demon-
strated by integrating both Mobile IP and SIP based mobility
techniques along with other suite of protocols such as IP-Sec,
PANA and AAA. IP-Sec tunneling between the client and
ERC1 and ERC2 adds an overhead of about 53 bytes for UDP
packets that comprise the IP headers, SPI, Sequence number
and authentication header.

Figures 13 and 14 show a typical handoff performance an-
alyzing the timing and sequence of protocols associated with
different types of traffic (e.g. signaling, media, control) when
Mobile IP and SIP are used as mobility protocols respectively.
As can be seen from figure 12, a successful Mobile IP regis-
tration is complete only after the IP-Sec tunnel is set up and
most of the handoff time is due to time taken by IP-Sec tun-
el setup. Y axis shows different set of protocols that have
been integrated and their sequence of execution on the mo-
bile client as the mobile moves along and performs a domain
handoff. X axis shows the timing associated with the mobile’s
movement. As can be evident there is a gap in RTP packet se-
quence received on the mobile until Mobile IP registration is
complete. Figures 15 and 16 illustrate the sequence of RTP
packets being received at the mobile station for both SIP and
MIP based mobility respectively. It is evident from these fig-
ures that SIP based mobility offers less packet loss compared
to MIP based mobility.

### Table I
**Handoff Values in Seconds**

<table>
<thead>
<tr>
<th>Handoff</th>
<th>RTP1</th>
<th>RTP2</th>
<th>DRCP</th>
<th>PANA</th>
<th>SIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave(S)</td>
<td>0.322</td>
<td>1.81</td>
<td>0.079</td>
<td>0.002</td>
<td>0.227</td>
</tr>
<tr>
<td>SDV(S)</td>
<td>0.07</td>
<td>0.492</td>
<td>0.033</td>
<td>0.0005</td>
<td>0.255</td>
</tr>
<tr>
<td>Ave(D)</td>
<td>0.241</td>
<td>1.89</td>
<td>0.08</td>
<td>0.045</td>
<td>0.289</td>
</tr>
<tr>
<td>SDV(D)</td>
<td>0.061</td>
<td>0.306</td>
<td>0.014</td>
<td>0.002</td>
<td>0.254</td>
</tr>
</tbody>
</table>

**Fig. 15.** RTP loss with SIP

**Fig. 16.** RTP loss with Mobile IP

### V. Conclusions and Discussions

This paper provides an architectural and implementation
perspective of a mobile wireless Internet telephony and
streaming multimedia testbed. This testbed has realized sev-
eral functional components needed to provide seamless mul-
timedia operation over the wireless Internet. In addition to
describing the functional components, and sequence of opera-
tion, we have provided performance measurement for each of
the associated atomic operations. Lessons learnt during build-
ing this comprehensive multimedia testbed and the results of
prototyping will be valuable while deploying it in the mo-
bile wireless Internet. During the process of implementation,
we ended up changing some of the initial design parameters
in favor of optimized methods and implementation friendly.
During mobility experiments, we did encounter problems due
to layer 2 interference if the adjacent base stations were us-
ing neighboring channels. Heterogeneous mobility involv-
ing CDMA based carrier networks often had fluctuation in
its signal strength based on the time of the day. Operating
systems choice affects the ease of implementation to a great
extent. Linux operating systems were favored for any server
implementation where as some of the clients were windows
based. Implementing and demonstrating multilayer secured
framework across carrier domain was the most challenging.
However providing desired QoS for the mobile and applica-
tion layer mobility management seemed to be very effective
during demonstration. We also discovered that it is more opti-
mized to distribute the functionality across the domain instead
of running several application on the same hardware platform.
During the process of testbed evolution we needed to upgrade
the kernels several times. Thus any solution based on appli-
cation layer approaches turned out to be easier to migrate to
the new version. Thus this architecture may be useful for the
Application Service Providers (ASPs) who do not have access
to underlying network components.

### VI. Acknowledgement

Authors would like to acknowledge other contributing
ITSUMO members, Ted Lu, Aileen Cheng, Parmesh Ra-
manathan of University of Wisconsin, Xiaotao Wu and
Jonathan Lennox of Columbia University towards their help
during the course of realizing this multimedia testbed.