Interworking Internet Telephony and Wireless Telecommunications Networks

Jonathan Lennox, Kazutaka Murakami, Mehmet Karaul, Thomas F. La Porta
Bell Laboratories, Lucent Technologies
{lennox,kmurakami,karaul,tlp}@bell-labs.com

Abstract— **Internet telephony and mobile telephony are both growing very rapidly. Directly interworking the two presents significant advantages over connecting them through an intermediate PSTN link. We propose three novel schemes for the most complex aspect of the interworking: call delivery from an Internet telephony (SIP) terminal to a mobile telephony (GSM) terminal. We then evaluate the proposals both qualitatively and quantitatively. We also describe our implementation of one of the proposals on the Bell Labs RIMA platform.**

Keywords— **Internet telephony, mobile telephony, SIP, GSM, mobility, interworking.**

I. Introduction

TWO of the fastest growing areas of telecommunications are wireless mobile telephony and Internet telephony. Second and third-generation digital systems such as the Global System for Mobile communications (GSM) [1], the Universal Mobile Telecommunications System (UMTS) [2], and wideband CDMA [3] are bringing new levels of performance and capabilities to mobile communications. Meanwhile, both the Internet Engineering Task Force's Session Initiation Protocol (SIP) [4] and the International Telecommunications Union's H.323 [5] enable voice and multimedia telephone calls to be transported over an Internet Protocol (IP) network. Subscribers to each of these networks need to be able to contact subscribers on the other. There is, therefore, a need to interconnect the two networks, allowing calls to be placed between them.

Some research has been performed investigating various aspects of interworking mobile communication systems with IP-based systems. The iGSM system [6] allows an H.323 terminal to appear to the GSM network as a standard GSM terminal, so that a GSM subscriber can have his or her calls temporarily delivered to an H.323 terminal rather than a mobile device. Several papers [7], [8], [9] describe a system for interworking GSM's in-call handover procedures with H.323. However, neither of these approaches solves the general interworking question: what is the best way for calls to be delivered and routed between the two networks?

As both mobile and Internet telephony are already designed to interconnect with the Public Switched Telephone Network (PSTN), the easiest way to interconnect them would be simply to use the PSTN as an intermediate link. This is, however, inefficient and suboptimal, as compared to connecting the networks by interworking the protocols directly, for a number of reasons.

First of all, routing calls via the PSTN can result in inefficient establishment of voice circuits. This is a common problem in circuit-switched wireless systems called "triangular routing," as illustrated in Figure 1. Because a caller's local switch does not have sufficient information to determine a mobile's correct current location, the signalling must travel to an intermediate switch which can locate the subscriber correctly.[1] This intermediate switch can be far away from the caller and the destination even if the two are located in a geographically close area. Since voice circuits are established at the same time as the call signalling message is routed, the voice traffic could be transported over a long, inefficient route.

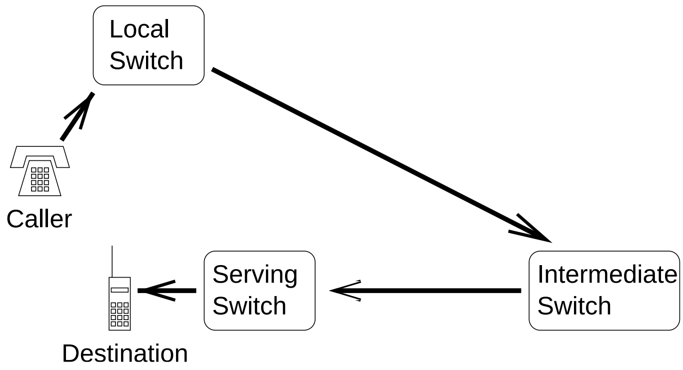

Fig. 1. Illustration of triangular routing in mobile networks

In Internet telephony, by contrast, the path of a call's media (its voice traffic, or other multimedia formats) is independent of the signalling path. Therefore, even if signalling takes a triangular route, the media travels directly between the devices which send and receive it. Since each device knows the other's Internet address, the packets making up this media stream are sent by the most efficient routes that the Internet routing protocols determine.

As we interwork Internet telephony with mobile telephony, we would like to maintain this advantage. We can accomplish this by supporting a direct IP connection between mobile base stations and IP terminals. With PSTN signalling, this is not possible, so IP telephony signalling must be used to establish this connection.

Another motivation for direct connection between mobile and Internet telephony is to eliminate unnecessary media transcod-

[1] There is an architectural difference here between the American mobile system based on ANSI 41 [10] and the European systems based on GSM MAP. In the American system, calls are always routed through a home mobile switching center, which is in a fixed location for each subscriber, so the voice traffic for all of the subscriber's calls travels through that switch. By contrast, GSM improves on this routing by sending calls through a gateway mobile switching center, which can be located close to the originating caller. However, as discussed in [11], there are some cases, such as international calls, where an originating PSTN switch does not have enough information to conclude that a call is destined for the GSM network, and thus routes it to the subscriber's home country. Because there is no way for circuit paths to be changed once they have been established, the call's voice traffic travels first to the user's home country and only then to his or her current location.

ing. The Real-Time Transport Protocol (RTP) [12], the media transport protocol common to both H.323 and SIP, can transport almost any publicly-defined media encoding [13]. Most notably, the GSM 06.10 encoding [14] is implemented by many clients. If a GSM mobile device talks to an RTP-capable Internet telephone with an intermediate PSTN leg, the media channel would have to be converted from GSM 06.10 over the air, to uncompressed (μ-law or a-law) audio over a PSTN trunk, and then again (likely) to some compressed format over the RTP media channel. The degradation of sound quality from multiple codecs in tandem is well known, and multiple conversions induce unnecessary computation. A direct media channel between a base station and an IP endpoint allows, by contrast, communication directly using the GSM 06.10 encoding without any intermediate transcodings.

Finally, on a broader scale, an integrated architecture supporting Internet and mobile telephony will evolve naturally with the expected telecommunications architectures of the future. Third-generation wireless protocols will support wireless Internet access from mobile devices. New architectures such as RIMA [15] for Mobile Switching Centers (MSCs) are using IP-based networks for communications between MSCs and base stations. In the fixed network, meanwhile, IP telephony is increasingly becoming the long-haul transport of choice even for calls that originate in the PSTN. The direct connection between Internet telephony and mobile networks takes advantage of all these changes in architecture and allows us to build on them for the future.

In this paper, we will consider the issue of how to interwork Internet telephony and mobile telecommunications, such that all the issues discussed above are resolved. For concreteness, we will illustrate our architecture using SIP for Internet telephony and GSM for mobile telephony.

The rest of the paper is structured as follows. Section II gives an architectural background on the mobility and call delivery mechanisms of GSM and SIP, to provide a basis for the following discussions. Section III proposes three different approaches to interworking GSM and SIP. Section IV provides mathematical and numerical analyses of the three proposals. In Section V, we discuss our implementation, and we finish with some conclusions in Section VI.

II. Background

In this section we review the mobility and call delivery mechanisms of GSM and of SIP.

GSM Mobility and Call Delivery

Some of the elements of a GSM network are illustrated in Figure 2. The MSC is a switching and control system in a wireless network. The MSC controlling the service area where a mobile is currently located is called its serving MSC. It routes calls to and from all the mobile devices within a certain serving area, and maintains call state for them. Associated with the serving MSC is a Visitor Location Register (VLR), a database which stores information about mobile devices in its serving area. (For the purposes of this paper we assume the predominant configuration in which the serving MSC and VLR are co-located.) Elsewhere in the fixed network we can find two other classes of entities. A Home Location Register (HLR) maintains profile information about a subscriber and keeps track of his or her current location. A gateway MSC directs calls from the PSTN into the mobile access network.

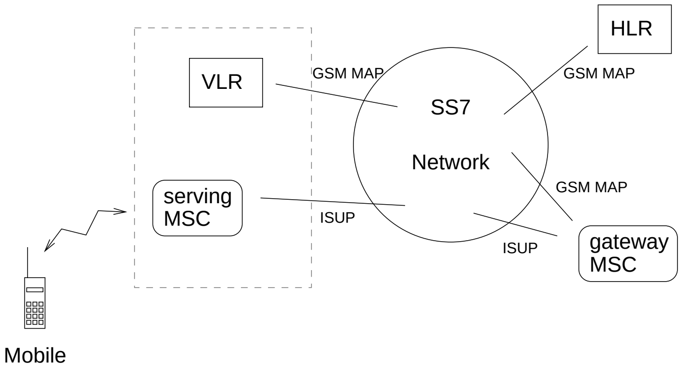

Fig. 2. Elements of a GSM Network

When a GSM mobile device first powers up or enters the serving area of a new serving MSC, it transmits a unique identification code, its International Mobile Subscriber Identity (IMSI) to the MSC. From the IMSI, the serving MSC determines the mobile's HLR and informs this HLR of the mobile's current location using the GSM Mobile Application Part (GSM MAP) protocol. The HLR stores this information and responds with profile data for the subscriber.

When a call is placed to a mobile subscriber, the public telephone network determines from the telephone number called (the Mobile Station ISDN number, or MSISDN) that the call is destined for a mobile telephone. The call is then directed to an appropriate gateway MSC. Call delivery from the gateway MSC is performed in two phases. In the first phase, the gateway MSC obtains a temporary routing number called a Mobile Station Routing Number (MSRN) in order to route the call to the serving MSC. For this purpose, the gateway MSC first locates the subscriber's HLR based on the MSISDN and requests routing information from it using GSM MAP. The HLR then contacts the VLR at the serving MSC. The VLR returns an MSRN that the HLR forwards to the gateway MSC. In the second phase, the gateway MSC routes the call to the serving MSC using the standard ISDN User Part (ISUP) protocol of the PSTN.

The MSRN is a temporarily assigned number which is allocated at the time the HLR contacts the VLR; it is valid only until the associated call is set up, and it is then recycled. This dynamic allocation of an MSRN is required because ISUP messages can only be directed to standard telephone numbers, and the quantity of these that can be allocated to a given serving MSC is limited. This has some costs, however, in the time needed to set up a call, as the serving MSC must be contacted twice during call setup.

When a subscriber moves from one location to another while a call is in progress, two possible scenarios result: intra-MSC or inter-MSC handovers. An intra-MSC handover occurs when a subscriber moves between the serving areas of two base stations controlled by the same serving MSC. In this case, the serving MSC simply redirects the destination of the media traffic. No signalling is necessary over the PSTN or GSM MAP. An inter-MSC handover, on the other hand, occurs when the subscriber moves from one serving MSC's area to another. The old serving

TABLE I
ANALOGOUS ENTITIES IN SIP AND GSM

GSM	**SIP**
HLR	Registrar
Gateway MSC	Home proxy server
Serving MSC	End system (for REGISTER)
MSISDN	User address (in INVITE)
IMSI	User address (in REGISTER)
MSRN	Device address

MSC contacts the new one in order to extend the call's media circuit over the PSTN. The old serving MSC then acts as an "anchor" for both signalling and voice traffic for the duration of the call.

All of the globally-significant numbers used by the GSM system — in particular, for the purposes of this paper, the MSRN, and the identifying number of the MSCs, in addition to the MSISDN — have the form of standard E.164 [16] international telephone numbers. Therefore they can be used to route requests in Signalling System no. 7 (SS7), the telephone system's signalling transport network.

SIP Mobility and Call Delivery

Architecturally, a pure SIP network (illustrated in Figure 3) is rather simpler than a GSM network, as it is significantly more homogeneous and much of the work takes place at the network layer, not the application layer. All devices communicate using IP, and all signalling occurs with SIP. Although many of the specific details are different, mobility in a SIP environment is conceptually similar to that of GSM. Table I lists some analogous entities in GSM and SIP networks.

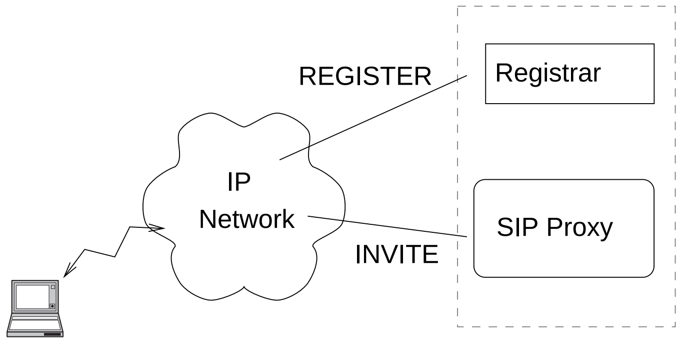

Fig. 3. Elements of a SIP network

There are two significant architectural differences between mobility in SIP and GSM. First of all, a SIP network does not have an intermediate device analogous to the serving MSC. Instead, end systems contact their registrars directly. Second, in SIP a two-phase process is not needed to contact the device during call establishment.

When a SIP subscriber becomes reachable at a new network address (either because she is using a new network device or because her device has obtained a new IP address through a mobility mechanism), the SIP device sends a SIP REGISTER to the user's registrar to inform it of the new contact location. This registration is then valid for only a limited period of time. Because end systems are assumed not to be totally reliable, registration information must be refreshed periodically (typically, once per hour) to ensure that a device has not disappeared before it could successfully de-register itself.

Unlike systems that use traditional telephone-network numbering plans, addresses in SIP are based on a "user@domain" format, similar to that of e-mail addresses. Any domain can, therefore, freely create an essentially unlimited number of addresses for itself. For the purposes of this discussion, it is useful to consider two types of addresses — "user addresses," analogous to an MSISDN number, to which external calls are placed, and "device addresses," roughly comparable to a non-transient MSRN. A device can create a temporary address for itself and have it persist for any period it wishes.

When a SIP call is placed to a subscriber's user address, a SIP INVITE message is directed to a proxy server in the domain serving this address. The proxy server consults the recipient's registrar and obtains his or her current device address. The proxy server then forwards the INVITE message directly to the device. Because the device address is not transient, the two-stage process used by GSM is not necessary. Once the call is established, media flows directly between the endpoints of the call, independently of the path the signalling has taken.

Though not explicitly defined as part of the basic SIP specification, in-call handover mobility is also possible within SIP. A mechanism for an environment based entirely on SIP, with mobile devices which have an Internet presence, is described in [17]. This mechanism does not use Mobile IP, as it suffers from a similar triangular routing issue as does circuit switching, and its handovers can be slow. Instead, it exploits SIP's in-call media renegotiation capabilities to alter the Internet address to which media is sent, once a device obtains a new visiting address through the standard mobile IP means. Therefore, Internet telephony calls can send their media streams to mobile devices' visiting addresses directly, rather than forcing them to be sent to the home addresses and then relayed by a home agent as in mobile IP.

III. ARCHITECTURE

In this section we describe our proposals for interworking SIP and GSM networks. In our design GSM mobile devices and their air interfaces and protocols are assumed to be unmodified. They use standard GSM access signalling protocols and GSM 06.10 media atop the standard underlying framing and radio protocols. Some GSM entities within the fixed part of the network, however, are upgraded to have Internet presences in addition to their standard GSM MAP and ISUP interfaces. Serving MSCs send and receive RTP packets and SIP signalling. In some of the proposals other GSM fixed entities, such as HLRs, have Internet presences as well. These entities still communicate with each other using GSM MAP and other SS7 signalling protocols, however.[2]

There are three primary issues to consider when addressing this interworking: how calls may be placed from SIP to GSM,

[2] It is possible that this SS7 signalling itself takes place over an IP network, using mechanisms such as the Stream Control Transmission Protocol (SCTP) [18], currently in development.

how they may be placed from GSM to SIP, and how in-call mobility (handovers) are handled. The second and third of these points are relatively straightforward, and we will address them first. The first one is more challenging and represents the main focus of this paper.

SIP/GSM Interworking: Calls from GSM to SIP

Calls originating from a GSM device and directed at a SIP subscriber are not, in principle, different from calls from the PSTN to a SIP subscriber. The primary issue when placing calls from a traditional telephone network to SIP is that traditional telephones can typically only dial telephone numbers, whereas SIP addresses are of a more general form, based roughly on e-mail addresses, which cannot be dialed on a keypad. Work is ongoing to resolve this problem, but the currently envisioned solution is to use a distributed database based atop the domain name system, known as "Enum," [19] which can take an E.164 international telephone address and return a SIP universal resource locator. For example, the E.164 number +1 732 332 6063 could be resolved to the SIP URI 'sip:lennox@bell-labs.com'.

Since globally significant GSM numbers take the form of E.164 numbers, several of the proposals below use Enum-style globally distributed databases in order to locate Internet servers corresponding to these addresses. However, for such databases it would not be desirable to use the actual global Enum domain, for security reasons.

SIP/GSM Interworking: In-Call Handover

As explained earlier, there are two categories of in-call handover: intra-MSC and inter-MSC. Intra-MSC handover does not need to be treated specially for SIP-GSM interworking. Because this happens between the serving MSC and the base stations, the network beyond the serving MSC is not affected. As an optimization, however, a serving MSC could use different IP addresses corresponding to different base stations under its control. In this case, a mechanism for SIP mobility as described before could be used to change the media endpoint address in mid-call.

Inter-MSC handover does affect SIP-GSM interworking, and remains for future study. We anticipate that a mechanism similar to that of [9], as described in the introduction, could be adapted to SIP for this purpose.

SIP/GSM Interworking: Mobile-Terminated Calls

The most complex point of SIP/GSM interworking is the means by which a SIP call can be placed to a GSM device. As discussed in the introduction, it is desirable to set up media streams directly between the calling party and the serving MSC. In order to accomplish this, SIP signalling must travel all the way to the serving MSC, as only the serving MSC will know the necessary IP address, port assignment conventions, and media characteristics.

We propose three methods as to how SIP devices can determine the current MSC at which a GSM device is registered. These have various trade-offs in terms of complexity, amount of signalling traffic, and call setup delay.

Proposal 1: modified registration

Our first proposal is to enhance a serving MSC's registration behavior. The basic idea is that a serving MSC registers not only with the subscriber's HLR, but also with a "Home SIP Registrar." This registrar maintains mobile location information for SIP calls.

The principal complexity with this technique lies in how the serving MSC locates the SIP registrar. Our proposal, illustrated in Figure 4, is to use a variant of the Enum database described above. Once the serving MSC has performed a GSM registration for a mobile device, it knows the mobile's MSISDN number. From this information, an Enum database is consulted to determine the address of the device's home SIP registrar, and the serving MSC performs a standard SIP registration on behalf of the device. A SIP call placed to the device then uses standard SIP procedures.

Because of authentication needs, this proposal uses either eight or ten GSM MAP messages (depending on whether authentication keys are still valid at the VLR) and six DNS messages per initial registration, and four SIP messages per initial or refreshed registration. Call setup requires a single SIP message and four DNS messages, though some DNS queries may be cached.

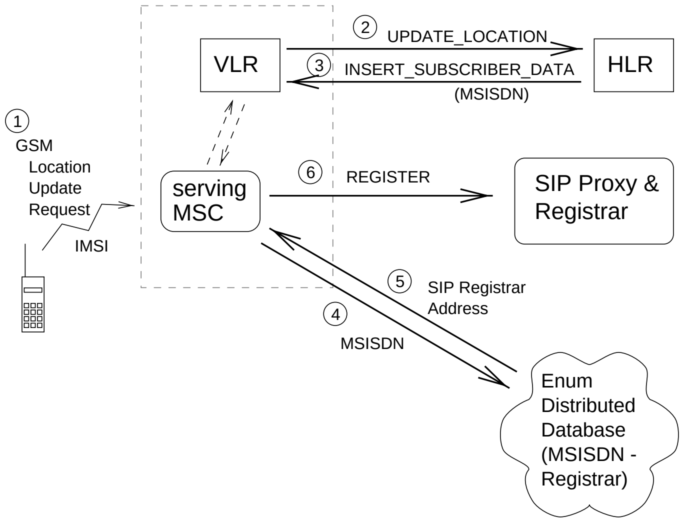

Fig. 4. Registration procedure for proposal 1

Compared to our other proposals, this proposal has two primary advantages. First, the only changes to the existing infrastructure are the modifications in the serving MSC and the addition of a variant Enum database to find registrars. Neither the SIP registrar and proxy server, nor the GSM HLR and gateway MSC, need to be altered. Second, because the complexity of the proposal occurs only in registration, call setup shares the single-lookup efficiency of SIP and is therefore relatively fast.

The disadvantages of this proposal, however, also arise due to the separation of the two registration databases. First, once a system requires the maintenance of two separate databases with rather incomparable data, the possibility arises that the information in the databases becomes inconsistent due to errors or partial system failure. This is especially true because of the differing semantics of SIP and GSM registrations — GSM registrations persist until explicitly removed, whereas SIP registrations have a timeout period and must be refreshed by the registering entity. Furthermore, when mobility rates are low, the dual registration procedure imposes significantly more signalling overhead than GSM registration alone, since SIP registrations must be refreshed frequently.

Proposal 2: modified call setup

By contrast, our second proposal does not modify the GSM registration procedure. Instead, it adds complexity to the call setup procedure. Essentially it adapts the GSM call setup to SIP. This is illustrated in Figure 5. When a SIP call is placed to a GSM user, the user's home SIP proxy server determines the MSISDN corresponding to the SIP user address, and queries the GSM HLR for an MSRN. The HLR obtains this through the normal GSM procedure of requesting it from the serving MSC's VLR. The SIP proxy server then performs an Enum lookup on this MSRN, and obtains a SIP address at the serving MSC to which the SIP INVITE message is then sent.

This approach uses either eight or ten MAP messages, as with standard GSM, for registration, and four MAP messages, six DNS messages, and one SIP message for a call setup.

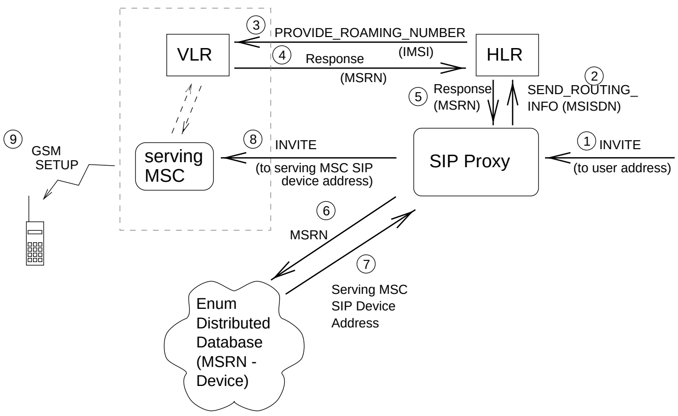

Fig. 5. Call setup procedure for proposal 2

Because this proposal does not modify the GSM registration database, it has several advantages over the previous proposal. Specifically, there is no possibility for data to become inconsistent, and the overhead of registration is as low as it is for standard GSM. However, both the signalling load and the call setup delay are high, as call setup now involves a *triple*-phase query: a GSM MAP query for the MSRN, an Enum lookup for the SIP device address, and finally the actual call initiation. Additionally, we have a new requirement that the SIP proxy server and the HLR need to be able to communicate with each other. This imposes additional complexity in both these devices, as it requires new protocols or interfaces.

Proposal 3: modified HLR

Our final proposal is to modify the GSM HLR. In this proposal, the serving MSC registers the mobile at the HLR through standard GSM means. The HLR then has the responsibility to determe the mobile's SIP device address at the serving MSC.

The overall registration procedure for this proposal is illustrated in Figure 6. When a serving MSC communicates with an HLR, the HLR is informed of the serving MSC's address, which, as mentioned earlier, is an E.164 number. The HLR performs a query to a specialized Enum database to obtain the name of the serving MSC's SIP domain, based on the serving MSC's address. While the previous two proposals treat the SIP device address as an opaque unit of information whose structure is known only to the serving MSC, this proposal takes advantage of its structure.

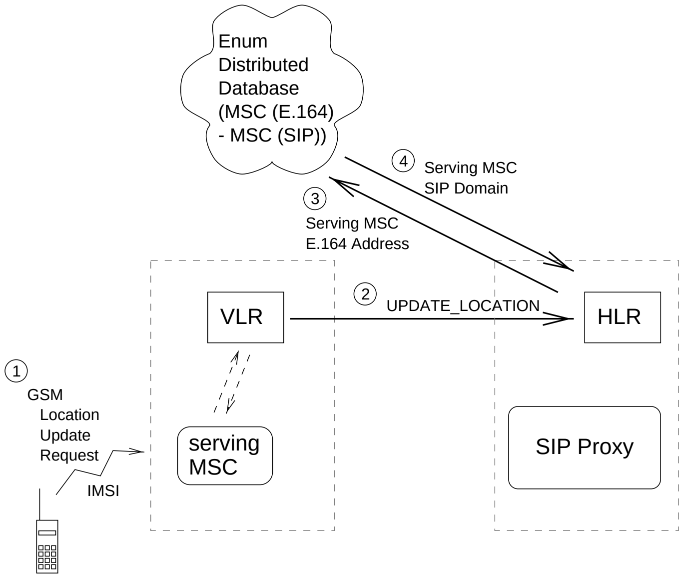

Fig. 6. Registration procedure for proposal 3

Figure 7 shows how a SIP call is placed. The SIP proxy server queries the HLR for a SIP address and the HLR returns an address of the form "MSISDN@hostname.of.serving.MSC" to which the SIP proxy then sends the call. This proposal uses either eight or ten MAP messages, and two DNS messages, for registration, and four DNS messages and one SIP message for call setup.

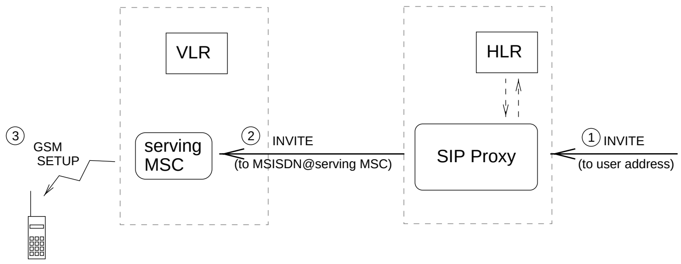

Fig. 7. Call setup procedure for proposal 3

This approach has the advantage that its overhead is relatively low for registration and quite low for call setup. The time requirements for call setup are similarly low. It does, however, require invasive modifications of HLRs. Additionally, the SIP proxy server and the HLR must be co-located, or else they must also have a protocol defined to interface them.

IV. Analysis

Two important criteria for evaluating the signalling performance of these three proposals for interworking SIP and GSM

TABLE II

MESSAGE WEIGHTS

Symbol	Parameter	Value
w_{sip}	Weight of a SIP message	1.0
w_{dns}	Weight of a DNS message	0.5
w_{map}	Weight of a MAP message	1.5

TABLE III

MOBILITY PARAMETERS

Symbol	Parameter	Value
$r_{\mathrm{in}}, r_{\mathrm{out}}$	Rate of call delivery / origination	variable
r_{bc}	Average boundary crossing rate	variable
$P_t(t)$	Boundary crossing rate prob. distribution ($P(t_0 \geq t)$)	$e^{-r_{\mathrm{bc}}t}$
s	Call / mobility ratio	$\frac{r_{\mathrm{out}}+r_{\mathrm{in}}}{r_{\mathrm{bc}}}$
P_{nr}	Prob. that a device is new to a serving MSC	50%
P_{ur}	Prob. that a device has a unique registrar at its serving MSC	20%
P_{us}	Prob. that a device has a unique serving MSC at its HLR/registrar	20%

TABLE IV

PROTOCOL PARAMETERS

Symbol	Parameter	Value
t_{sip}	SIP registration refresh interval	3 hr
t_{dns}	DNS cache time-to-live	24 hr
c_{auth}	Number of pieces of authentication data cached at VLR	5

TABLE V

WEIGHTED PACKET COUNTS FOR EACH PROPOSAL

Case	Formula
Modified Registration	
Registration	$r_{\mathrm{bc}}\big((8+2/c_{\mathrm{auth}})\,w_{\mathrm{map}}+ (2P_{\mathrm{nr}}+4P_{\mathrm{ur}})\,w_{\mathrm{dns}}+ 4\,(1+\sum_{i=1}^{\infty} P_t(it_{\mathrm{sip}}))\,w_{\mathrm{sip}}\big)$
Call setup	$r_{\mathrm{in}}\,(4P_{\mathrm{us}}w_{\mathrm{dns}}+1w_{\mathrm{sip}})$
Modified Call Setup	
Registration	$r_{\mathrm{bc}}\,(8+2/c_{\mathrm{auth}})\,w_{\mathrm{map}}$
Call setup	$r_{\mathrm{in}}\,(4w_{\mathrm{map}}+6P_{\mathrm{us}}w_{\mathrm{dns}}+1w_{\mathrm{sip}})$
Modified HLR	
Registration	$r_{\mathrm{bc}}\big((8+2/c_{\mathrm{auth}})\,w_{\mathrm{map}} +2P_{\mathrm{us}}w_{\mathrm{dns}}\big)$
Call setup	$r_{\mathrm{in}}\,(4P_{\mathrm{us}}w_{\mathrm{dns}}+1w_{\mathrm{sip}})$

are signalling load and call setup delay. A detailed study of call setup delay remains for future investigation. In this paper we focus on performance in terms of signalling load.

Each of the proposals involves the use of several different protocols, in varying ratios. In order to compare total signalling load imposed by each protocol, we assigned signalling messages of each protocol a weight. The default values of these weights are listed in Table II. We discuss the effect of these weights on the total signalling load in our sensitivity analysis later in this section.

Tables III and IV list the parameters for our model. We assume equal rates of call delivery r_{in} and r_{out}, as is commonly observed in European settings. We assign an exponential distribution to the probability $P_t(t)$ that a mobile remains in a particular MSC's serving area for longer than time t. DNS caching was accounted for by assigning the probabilities P_{nr}, P_{ur}, and P_{ns} to the likelihood that particular DNS queries have been performed recently, within the DNS time-to-live period.

Table V shows the equations for the weighted signalling loads for registration and call establishment in each proposal. These equations are based on the packet counts for each proposal in Section III.

Figure 8 graphs the total weighted signalling load (registration plus call setup costs) for each of the three proposals, as both the incoming call rate and the call / mobility ratio vary. The intersection line at which modified registration and modified call setup are equal is shown in bold.

From this graph, we can observe some general characteristics of the proposals' signalling load. First, the modified HLR proposal consistently has the lowest signalling load of the three, typically 20 – 30% less than the others. This corresponds to intuition, as it combines the "best" aspects of each of the other two proposals, unifying both an efficient registration and efficient call setup procedure.

Second, the relative signalling loads for the other two proposals depend on the values of the traffic parameters. Modified call setup is more efficient for a low incoming call rate or a low call / mobility ratio (i.e., fast mobility), while modified registration is more efficient when both parameters are high. A closer look at the equations in Table V reveals the reasons. Consider the relative efficiency of the two approaches for varying incoming call rates: modified call setup performs less well for high incoming call rates because its call setup procedure requires four additional GSM MAP messages and possibly two additional DNS messages compared to that of modified registration. Similarly,

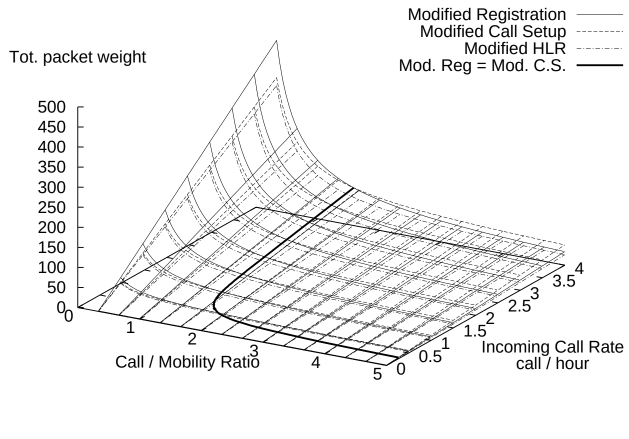

Fig. 8. Weighted signalling load of the three proposals

modified call setup outperforms modified registration for low call / mobility ratios because the latter has higher registration message overhead due to dual registration and SIP registration soft-state.

In order to increase the confidence in the above results, we performed sensitivity analyses to validate our choice of various parameters.

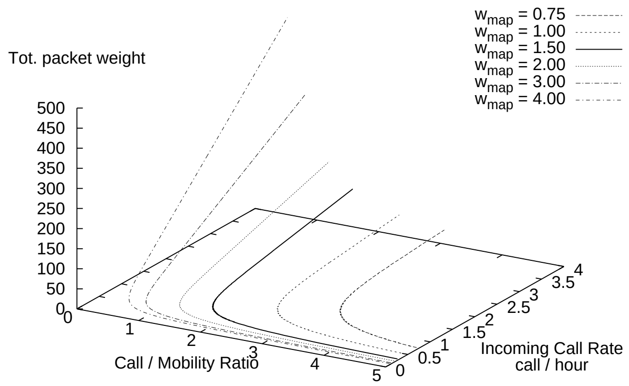

Fig. 9. Line of Intersection: Mod. C.S. = Mod. Reg. ($w_{\rm map}$ varying)

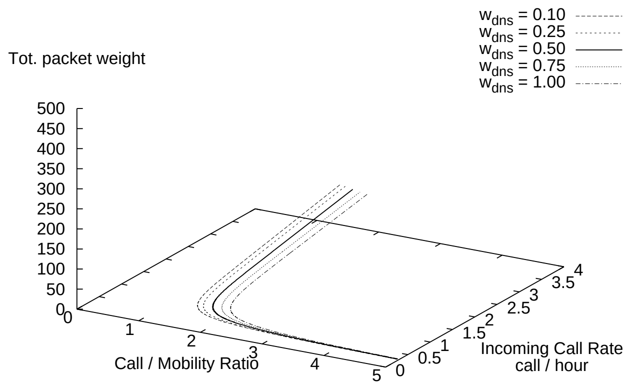

Fig. 10. Line of Intersection: Mod. C.S. = Mod. Reg. ($w_{\rm dns}$ varying)

Sensitivity analyses for the weights assigned to MAP and DNS messages are shown in Figures 9 and 10, respectively. These graphs illustrate how, as the protocol weighting changes, the position of the intersection line in Figure 8 changes.

Figure 9 shows that as the weight assigned to the MAP protocol increases, the area in which modified registration is more efficient — the right-hand side of the graph, where call rate and call/mobility ratio are both high — increases as well. This fits with the intuitive understanding of the approaches, as modified registration uses fewer MAP messages than modified call setup. Similarly, Figure 10 shows that as the weight assigned to the DNS protocol increases, the area in which modified registration is more efficient shrinks slightly. This also fits with intuition, as modified registration uses more DNS packets. However, the total packet load is generally less sensitive to the weight assigned to DNS messages, which explains why the lines in Figure 10 are relatively close to each other.

The signalling load of the modified HLR proposal is always less than the other two. Thus, it is not shown in our sensitivity graphs. In regards to the other two protocols, though the crossover point moves as the weights assigned to the protocols vary, these sensitivity analyses show that the general shape of the graph, and therefore the conclusions we draw from it, do not change.

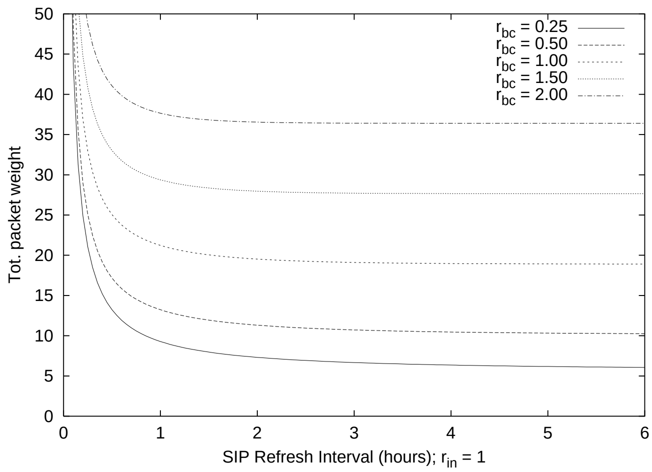

Fig. 11. Total weight of modified registration

Figure 11 shows the effect of various choices of values for the SIP registration timeout period. (This value only affects the modified registration proposal, as the other proposals do not use SIP registration.) The value for this parameter should be chosen so that the additional cost of SIP registration is relatively minor, that is, so that the graph has roughly flattened out. This optimal value therefore depends on the boundary crossing rate, but generally, a timeout of three hours is a good choice for most reasonable boundary crossing rates. This value can be larger than the standard value of one hour used by SIP, as serving MSCs can be assumed to be more reliable and available than regular SIP end systems.

V. Implementation

To prove the feasibility of our proposal, we implemented the modified call setup scheme atop the Enhanced Mobile Call Processing (EMCP) component of the Bell Labs Router for Integrated Mobile Access (RIMA) [15]. Figure 12 illustrates the overall architecture of this system. The modified call setup scheme was selected partly because it appears to be more applicable than modified registration scheme in the future mobile networks where a higher mobility rate is expected. It also requires substantially less modification to GSM equipment than the modified HLR scheme.

As opposed to traditional MSCs, RIMA is inherently IP based and uses packet networks for both transport and signalling. It is built on top of an IP router based network and is composed of a cluster of commodity processors and various gateways performing media conversion and transcoding. It supports standard circuit voice for wireless terminals like GSM phones and connects

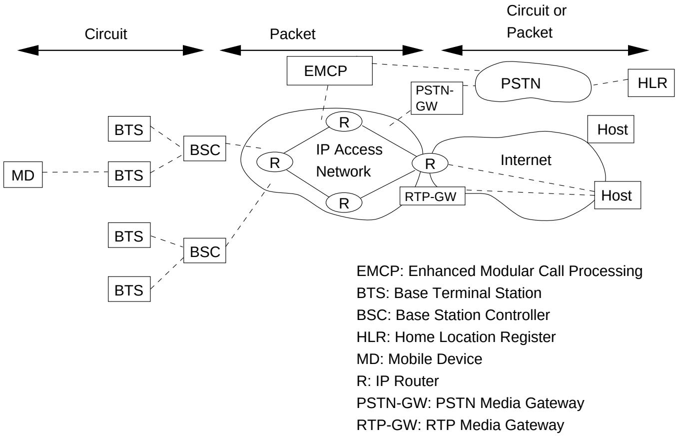

Fig. 12. RIMA-based Network

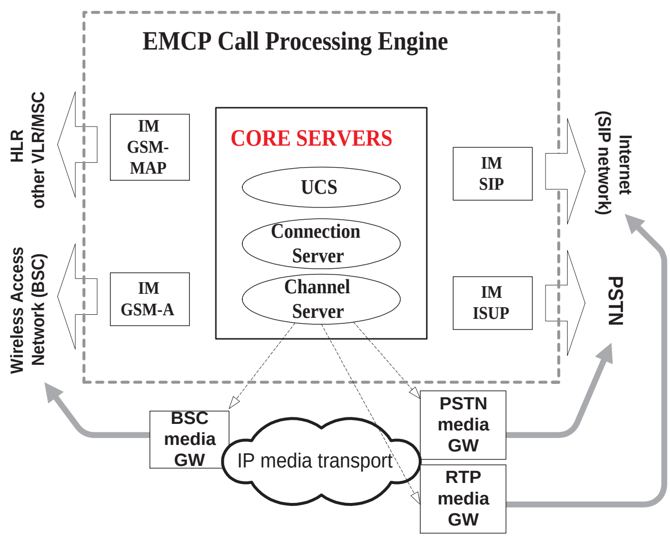

Fig. 13. Structure of EMCP Call Processing Engine

to existing circuit networks like the PSTN. It was designed with the idea in mind of connecting to packet voice networks like the Internet.

RIMA provides wireless access to mobile users through a packet based wireless access network. A RIMA network has four major components: a Base Station Controller (BSC), a PSTN media gateway (PSTN-GW), an RTP media gateway (RTP-GW), and the EMCP call processing engine, connected *via* an IP network.

Each BSC has an IP interface and translates voice and signalling information between circuit and packet format. It serves as a media gateway translating between circuit voice and RTP/IP packet voice. With respect to signalling, it terminates the standard GSM interface towards mobile devices to accommodate existing radio networks and tunnels these signals in IP packets on the RIMA wireless access packet network.

A PSTN-GW performs media conversion between RTP/IP packet voice in the RIMA access network and circuit voice over the PSTN. It is controlled by the call processing engine, and it may perform possible transcodings between different coding schemes such as compressed wireless (e.g. GSM speech) and PCM (e.g. (μ-law).

We added the RTP-GW to provide RIMA with media connections to the Internet. Though the RIMA access network uses RTP internally, it was useful to centralize advanced functionality such as buffering, jitter adaptation, and handling of the Real-Time Control Protocol (RTCP) into a single location. In this way, other RIMA entities do not need to support the entire suite of complex RTP behavior. The RTP-GW also performs transcoding between coding schemes as necessary, if for example a remote SIP endpoint does not indicate support for GSM encoding but wishes only to send and receive PCM. We implemented this gateway using the Bell Labs RTPlib [20] library, which we ported to the same single-board computers as the PSTN-GW.

RIMA's MSC and VLR functionality is realized by the EMCP call processing engine, whose structure is shown in Figure 13. It is deployed on a cluster of commodity processors such as workstations or single board computers. The engine is separated from the IP media transport network and can be viewed as a signalling gateway by IP telephony networks. It consists of a collection of functionally distributed servers. Call processing and mobility management tasks are accomplished by their collaboration.

The call processing engine is comprised of two server classes: core servers and interworking managers (IMs). Core servers perform call processing and mobility management tasks common to any wireless system. Interworking managers act as protocol gateways to internal core servers, isolating them from external signalling protocols thereby allowing the core servers to evolve independently of these protocols.

There are three core servers: a channel server, a connection server, and a user call server (UCS). The channel server manages switching device resources, such as transport channels and DSPs for vocoding, allocated during call setup and deallocated during call release. The connection server coordinates the allocation of channel resources to establish an end-to-end connection. The UCS maintains information on the registration status of mobile devices currently located within the service area of the RIMA system and records call activities involving a particular mobile device. The UCS also handles other mobility management tasks such as paging, handover, mobile user authentication, and ciphering.

Interworking managers allow core servers to accommodate different sets of standard interfaces. As originally developed, EMCP has interworking managers supporting the GSM A standard protocol between an MSC and a BSC (IM-GSM-A), GSM MAP to the HLRs (IM-GSM-MAP), and ISUP to the PSTN (IM-ISUP). To realize the architecture described in this paper, we added a new interworking manager, IM-SIP, which supports SIP towards the Internet. Implementing this IM was straightforward. Due to the modularity of the EMCP architecture, IM-SIP could use the same interfaces as IM-ISUP. Because we chose the modified call setup model, we did not have to alter EMCP's registration procedures.

For the Home SIP Proxy, we extended an experimental Bell Labs SIP proxy server and registrar to allow it to communicate with an HLR. This proxy server was programmed to recognize that certain blocks of addresses corresponded to GSM users. For these numbers it invokes a special procedure in which it asks the HLR for an MSRN. Because Enum has not yet been standardized, we instead used a table lookup to find SIP addresses corresponding to the MSRN returned.

VI. Conclusion

We proposed three novel schemes to directly interconnect GSM mobile and SIP Internet telephony systems. Compared with the conventional approach of routing a call through PSTN, direct interconnection prevents triangular routing and eliminates unnecessary transcodings along its path. We analyzed the signalling message load of three proposals under a wide range of call and mobility conditions. The modified HLR scheme always imposes less signalling burden, typically 20-30% less than the other schemes, although it requires significantly greater modification to GSM equipment. The efficiency of the other two proposals, modified registration and modified call setup, depends on the traffic parameters. When the incoming call rate and call / mobility ratio are both high, modified registration is more efficient. Modified call setup performs better otherwise. We further demonstrated our implementation of one of the proposed schemes, modified call setup, in the Bell Labs next generation wireless access system RIMA.

References

[1] European Telecommunications Standards Institute, "Digital cellular telecommunications system, network architecture," GSM 03.02 version 7.1.0 release 1998, European Telecommunications Standards Institute, Sophia Antipolis, France, Feb. 2000.

[2] European Telecommunications Standards Institute, "Universal mobile telecommunications system (UMTS), general UMTS architecture," 3G TS 23.101 version 3.0.1 release 1999, European Telecommunications Standards Institute, Sophia Antipolis, France, Jan. 2000.

[3] Bijan Jabbari, Ed., "Special issue on wideband CDMA," *IEEE Communications Magazine*, vol. 36, no. 9, Sept. 1998.

[4] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: session initiation protocol," Request for Comments 2543, Internet Engineering Task Force, Mar. 1999.

[5] International Telecommunication Union, "Packet based multimedia communication systems," Recommendation H.323, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[6] Herman C. H. Rao, Yi-Bing Lin, and Sheng-Lin Cho, "iGSM: VoIP service for mobile networks," *IEEE Communications Magazine*, vol. 38, no. 4, pp. 62–69, Apr. 2000.

[7] Wanjiun Liao, "Mobile internet telephony: Mobile extensions to H.323," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, New York, Mar. 1999.

[8] Wanjiun Liao, "Mobile internet telephony protocol: An application layer protocol for mobile internet telephony services," in *Conference Record of the International Conference on Communications (ICC)*, Vancouver, British Columbia, June 1999.

[9] Wanjiun Liao and Jen-Chi Liu, "VoIP mobility in IP/cellular network internetworking," *IEEE Communications Magazine*, vol. 38, no. 4, pp. 70–75, Apr. 2000.

[10] Telecommunications Industry Association and Electronics Industry Association, "Cellular radiotelecommunications intersystem operations," TIA/EIA ANSI-41-D, Telecommunications Industry Association, Arlington, Virginia, Dec. 1997.

[11] Yung-Jan Cho, Yi-Bing Lin, and Herman Chung-Hwa Rao, "Reducing the network cost of call delivery to GSM roamers," *IEEE Network*, vol. 11, no. 5, pp. 19–25, Sept. 1997.

[12] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," Request for Comments 1889, Internet Engineering Task Force, Jan. 1996.

[13] H. Schulzrinne, "RTP profile for audio and video conferences with minimal control," Request for Comments 1890, Internet Engineering Task Force, Jan. 1996.

[14] European Telecommunications Standards Institute, "Digital cellular telecommunications system, full rate speech," GSM 06.10 version 5.0.1, European Telecommunications Standards Institute, Sophia Antipolis, France, May 1997.

[15] Thomas F. La Porta, Kazutaka Murakami, and Ramachandran Ramjee, "RIMA: router for integrated mobile access," in *Proceedings of the 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC)*, London, United Kingdom, Sept. 2000, to appear.

[16] International Telecommunication Union, "The international public telecommunication numbering plan," Recommendation E.164, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, May 1997.

[17] Elin Wedlund and Henning Schulzrinne, "Mobility support using SIP," in *Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM'99)*, Seattle, Washington, Aug. 1999.

[18] R. R. Stewart et al., "Stream control transmission protocol," Internet Draft, Internet Engineering Task Force, June 2000, Work in progress.

[19] P. Faltstrom, "E.164 number and DNS," Internet Draft, Internet Engineering Task Force, May 2000, Work in progress.

[20] Henning Schulzrinne, Jonathan Lennox, Daniel Rubenstein, and Jonathan Rosenberg, "RTPlib: Bell Labs RTP library," Available from http://www.bell-labs.com/topic/swdist/.