Efficient User Controlled Inter-Domain SIP Mobility Authentication, Registration, and Call Routing

Chaouki T. Abdallah
Joud S. Khoury
Henry N. Jerez

Follow this and additional works at: http://digitalrepository.unm.edu/ece_fsp

Recommended Citation
Efficient User Controlled Inter-Domain SIP Mobility
Authentication, Registration, and Call Routing

Joud S. Khoury*  Henry N. Jerez†  Chaouki T. Abdallah*
*School of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131-0001
†Corporation for National Research Initiatives, Reston, VA 20191
{jkhoury, chaouki}@ece.unm.edu
hjerez@cnri.reston.va.us

Abstract—Over the past decade, multimedia services have gained significant acceptance and played an important role in the convergence of IP networks. The proliferation of mobile devices and the nomadic user and computing lifestyles on current networks make mobility support a crucial ingredient of current IP-based multimedia systems. The Session Initiation Protocol (SIP) presents one approach towards supporting IP mobility. Additionally, SIP is increasingly gaining in popularity as the next generation multimedia signaling and session establishment protocol, and the SIP infrastructure is anticipated to be extensively deployed all over the Internet. We have lately proposed an approach to inter-domain SIP mobility which we call H-SIP. H-SIP is a user-controlled mobility scheme that improves personal and terminal mobility. H-SIP uses persistent identifiers and leverages the traditional SIP architecture to abstract any domain binding from users. This paper expands on our previous work and experimentally proves the efficiency of H-SIP in achieving inter-domain authentication and call routing through modeling and real-time measurements.

1. INTRODUCTION

The industry has recently witnessed a rapid increase in the popularity and deployment of packet switched multimedia services due to its low cost and broader data services. Additionally, mobile operators are currently promoting simultaneous Cellphone/Wi-Fi access [1] by introducing dual-mode phones that switch between the cellular network and the IP packet switched network. Efficient user mobility and identity persistence issues become critical in such networks and need to be carefully addressed in this context.

SIP [2] is a signaling and control protocol for handling multimedia sessions. The SIP architecture is proposed as an efficient candidate that can be reused to provide personal, terminal and session mobility [3], [4], [5], [6] with a readily available infrastructure. This avoids the redundancy introduced by simultaneous deployment with Mobile IP [7]. The successful reuse of SIP to simultaneously support multimedia communications and mobility leverages the issues emanating from SIP users roaming across multiple SIP domains. Briefly, SIP handles user location through the use of a proxy/location server that accepts user registration requests and updates the respective user location in a location repository. The protocol inherently implements location independence through the use of the uniform resource identifiers (URI) [8], which directly offers personal mobility due to its location independence. However, the SIP service model and identification scheme defines a user only within the domain boundaries of the service provider i.e. a user must associate with a specific proxy server that handles user authentication as well as initial traffic routing. The proxy maintains a unique account for the user, who in turn, is expected to coordinate with that same proxy irrespective of the user’s location. This requirement translates into unnecessary loads on the SIP server and on a particular domain. Additionally, it complicates the coordination of roaming users who must communicate with a central proxy server while roaming. Despite the possible presence of firewalls and other network restrictions on the foreign domain, roaming users are required to use the central home server instead of using the available local servers. Consequently, while URIs solve the location binding issue, they introduce the domain binding issue. Inefficient traffic routing is a direct consequence of such binding.

The Handle-SIP (H-SIP) [9] proposal addresses the inter-domain mobility and identification issues by introducing an abstraction framework on top of the traditional SIP architecture. User location and association are abstracted through the use of globally unique and persistent identifiers called handles that are part of the Handle System [10], [11], [12], [13]. This allows efficient load distribution and traffic routing as well as seamless and secure roaming. The major premise of H-SIP is the widespread deployment of the SIP infrastructure over the Internet. H-SIP solely enhances personal and terminal mobility [5] in traditional SIP architectures. It may be gradually deployed and fits seamlessly within the traditional SIP architecture allowing users to transparently roam across different SIP domains. This paper expands on our previous work with H-SIP [9] and experimentally proves the efficiency of H-SIP in achieving inter-domain authentication and call routing through modeling and real-time measurements.

The rest of the paper is organized as follows. Section II briefly overviews the principal design aspects of H-SIP and the mechanisms it employs to implement inter-domain authentication, registration and call routing. Section III details the implementation test-bed and section IV discusses the experimental model used for the evaluation of H-SIP. We then describe the performance results in section V. Finally,
Section VI outlines our future work before we conclude in section VII.

II. OVERVIEW OF H-SIP MOBILITY SCHEME

This section briefly overviews the H-SIP mobility scheme [9].

A. Sessions and Mobility

Recall that SIP defines a user as an entity that associates with a particular domain. Figure 1 depicts a simple scenario of a roaming user r_user who has a valid association with her home domain hdomain but is currently present in a foreign domain fdomain. SIP signaling traffic originating from (REGISTER) or terminating at (arrows 1,2,3: arbitrary SIP user trying to INVITE the roaming user) r_user must inefficiently pass through her home proxy server. Figure 1 identifies this traffic as traditional flow. All requests to/from the roaming user must go through the central home proxy server. Clearly, if the user is present in another country, her traffic would still have to go through her central home proxy (triangle routing) as depicted in Figure 1, despite the availability of a local proxy server in the foreign domain (Foreign Server). This results into significant delays that are not accepted for time sensitive applications. Even with SIP mobility management (SIPMM) [3], [4] support (personal, terminal and session mobility) enabled, the same scenario occurs. Inefficient per-call traffic routing, and service centralization, are obvious limitations that users roaming in these traditional and Mobile SIP environments have to suffer from.

Obviously, the use of URIs to identify users and the inherent dependence of the URI on a particular domain, complicates message routing. One solution is to abstract the actual identifier eliminating per-call coordination to minimize the signaling traffic in highly mobile environments.

B. Abstraction layer

H-SIP uses handles as globally unique identifiers to locate and identify SIP architectural elements. Users will identify each other, as well as the SIP servers they associate with using handles instead of URIs and domain names respectively. This abstraction allows the system to route calls independent of user location and domain association. A handle is a persistent name that can be associated with a set of attributes. Some of these attributes can describe location, permissions, administrators and state. The fact that handles are defined independently of any of the attributes or public keys of the underlying objects, makes them persistent identifiers [14].

Briefly, the Handle System [10], [11], [12], [13] is intended to be a means of universal basic access to registered digital objects [15]. It provides a distributed, secure and global name service for administration and resolution of handles over the Internet. The Handle System is extremely scalable and suitable to operate in highly mobile environments.

A possible realization of the handles for user r_user and proxy server fproxy is depicted in Figure 2. The handle has several fields including administration and security information (public or secret keys). Administrators have the privilege to modify the fields inside the handle provided the administrator succeeds to authenticate to the Handle System using the private/secret key.

![Fig. 1. A Reference inter-domain roaming scenario](image1)

![Fig. 2. Sample handle structures. A. user handle B. proxy handle (dashed arrow shows relation between the two handles)](image2)

C. Authentication and Registration

Recall that with the H-SIP [9] abstraction layer, each user owns and administers her own handle, specifying the set of administrators (also handles) that have rights over her handle. Among those administrators, the user should include the handle of any SIP proxy server that she wishes to register with, which could be any foreign server(s) that she trusts (dashed arrow in Fig. 2 shows that r_user trusts fproxy). Authentication proceeds by the user providing credentials to the local proxy (Figure 1), which in turn validates the credentials against the Handle System. Note that the user has picked the password and made it part of her handle (Fig. 2.A, SIP.PWD), thus eliminating the need to maintain internal accounts on the proxy server. The user’s credentials are valid for all realms provided the correct administrative privileges are

---

1 The largest individual Handle System implementation to date is deployed at the Los Alamos National Laboratory, which is intended to support more than a half billion identifiers while providing internal resolution services to one of the largest archival collections in the United States.
are set in the Handle System. This property is essential, as it allows a particular authenticated SIP message to traverse multiple domains instead of requiring re-authentication for each domain on the path of the message.

D. Call Routing

After authenticating and registering with the foreign server, the user's handle-to-URL mapping (Fig. 2A, SIP.URL) remains fresh allowing correspondent users to reach her simply by addressing her handle. From the perspective of a user, all other users seem to belong to one local domain and abstraction is complete. In the scenario of Figure 1, call routing proceeds when the caller c.user attempts to INVITE the roaming user r.user (callee) using the latter's handle 2118/r.user as shown in Figure 1 (arrows a,b,c,d). The proxy server resolves the handle into the SIP.URL field which is 2118.r.user@x.y.z.w, and forwards the call accordingly using the natural SIP call flow. Consequently, H-SIP eliminates the need to go through a home proxy for session setup or redirection. This renders the call route more efficient eliminating unnecessary overhead and significant round-trip times.

E. User-Controlled Mobility

The H-SIP mobility scheme is user-controlled in the sense that the user is responsible for the administration of her SIP identifier now that the latter is domain independent. Consequently, routing the user's calls through a local server simply requires the user to add the server's persistent identifier (handle) to her persistent identifier's admin list. This indicates that the user trusts the local server to route her calls. The other aspect of user control is the distributed service model that potentially eliminates the need for service level agreements SLAs between domains. Instead, the local domain proxy will directly challenge the user and grant her access provided sufficient credentials exist within the user's persistent identifier. These credentials can include trust information as well as financial information that the actual server can validate before allowing the user to route traffic through.

III. IMPLEMENTATION AND TEST-BED

We have implemented the H-SIP functionality as an extension to two open source SIP servers, the JAIN-SIP Proxy [16] and the SIP Express Router (SER) [17]. The JAIN-SIP proxy server is an open source JAVA based SIP proxy built on top of the JAIN-SIP-1.1 API. SER is an open source, configurable SIP server that is widely deployed in the research community. Our implementation is a JAVA based H-SIP API that can be easily called from both proxy servers to expose the H-SIP interface operations. The operations mainly enable inter-domain authentication, registration and call routing using handles. All the results depicted heretofore are based on the JAIN-SIP proxy. Note that the H-SIP mobility scheme requires no modifications to the User Agents (UAs) and is solely deployed through extensions to the SIP proxy servers.

Our test-bed is a realization of the roaming framework depicted in Figure 1. The roaming user r.user is located in a foreign domain fdomain while her home domain is hdomain. We study the registration process that follows from the model. We also study the call establishment process in which a correspondent user c.user attempts to establish a call with the roaming user.

We are running three modified SIP servers on three separate domains:

1) ece.unm.edu located at the University of New Mexico, Albuquerque, New Mexico. [Server IP: 129.24.24.106]
2) cnri.reston.va.us located in Reston, Virginia. [Server IP: 132.151.9.104]
3) istec.org located in Panama. [Server IP: 168.77.202.59]

A roaming user is allowed to move across the domains while establishing connectivity within each domain using the respective local server. The three servers are running Fedora Core 4 kernels and are identical in terms of work load and processing speed (AMD Athlon 1.1 GHz processors).

To use the framework, users are expected to be able to manage their own handles. The Handle System provides a free administration tool [13] for this purpose. This tool is currently implemented in JAVA.

Currently, users wishing to associate with a proxy server are required to specify the latter's IP address or domain name. Since our approach exploits handles instead of domain names, we have implemented a specialized gateway that translates between DNS and handle protocols. The gateway is responsible for protocol translation, specifically, handle to DNS. We refer to this gateway as the Handle-DNS proxy (HDP). HDP is a modified DNS server that communicates using the BIND protocol and implements extra functionality allowing it to associate canonical names and aliases inside its particular naming zone with handles. HDP will therefore resolve canonical names inside its naming zone using the Handle System and will, in addition, allow any common DNS server to resolve DNS entries in the format: <handle>:[DNS proxy domain] to the actual value of the INET_HOST attribute of that particular handle. Details about this gateway which can be extended to become a plug-in for current DNS servers are presented in [18]. The ease of deployment of our framework and success of the conducted experiments encourages us to pursue this work.

IV. EXPERIMENTAL MODEL

In this section, we will compare the performance of roaming and non-roaming environments in terms of registration and call establishment delays. The non-roaming case represents traditional SIP signaling between a user and her home SIP server irrespective of roaming, while the former case represents our proposed H-SIP approach. Both cases were tested under the same conditions regarding UA/SIP processor speed and work loads. As User Agents, we used the Java SIP
A. Registration

Figure 3 shows a comparison of the data flow involved in the registration process for roaming (H-SIP) and non-roaming (traditional SIP) environments. The roaming user \( r_{user} \) is located in fdomain, ece.unm.edu. For the non-roaming case, we tested with hdomain being either cnri.reston.va.us (Reston, Virginia) or issec.org (Panama). Registration here assumes the basic digest authentication [20] with the proxy server. The model considers how the UA refreshes its registration with the proxy/registrar continuously (we used a registration TTL of 1 minute). If we estimate the server’s average processing time of the REGISTER request including digest authentication by \( \alpha \), then from Figure 3.A, the average registration time \( t_A \) as seen by \( r_{user} \) is given by,

\[
t_A = \frac{1}{n} \sum_{1 \leq i \leq n} t_i = t_1
\]

(1a)

since,

\[
t_1 = t_2 = \ldots = t_n \approx \alpha + 2y
\]

(1b)

where \( t_i \) is time consumed by the \( i^{th} \) REGISTER, and \( y \) is the average round-trip communication delay between \( r_{user} \) and the registering SIP server. Obviously, \( t_A \) includes a relatively significant communication delay as a result of the presumably large geographical separation between the roaming user and her home SIP server.

Now, if we consider H-SIP, then from Figure 3.B, the average registration time \( t_B \) as seen by \( r_{user} \) is given by,

\[
t_B = \frac{1}{n} \sum_{1 \leq i \leq n} t_i'
\]

(2a)

and,

\[
t_1' \approx \alpha + x + u
\]

(2b)

\[
t_2' = \ldots = t_n' \approx \alpha
\]

(2c)

where \( t_i' \) is time to perform the \( i^{th} \) REGISTER, \( x \) is the average handle resolution delay and \( u \) is the average handle update delay. First, we note that the round-trip delay \( y \) is negligible in this case due to the existence of \( r_{user} \) and the SIP server on the same local network. Besides, note here that the server will issue one handle resolution and one update for the first REGISTER request only. Additionally, the first handle resolution is always cached internally on the server. Unless \( r_{user} \) moves to another domain or un-registers, no handle resolution/update is required. This means that subsequent REGISTER requests \( (t_2' \text{ to } t_n') \) will read the cached value. So, for \( i > 1 \), \( x = u = 0 \).

Here, \( t_A \) and \( t_B \) are measured by \( r_{user} \) as the average time between sending the REGISTER request and receiving the 200 OK response. In Figure 3, \( r_{user} \) performs a DNS lookup on a Handle-DNS lookup in scenarios A and B respectively. These lookups are excluded from the performance metrics i.e. \( t_1 \) and \( t_2 \) do not include the DNS lookups for the SIP servers. However, we closely examine and compare the two lookup times and show the performance results in section V-C.

B. Call establishment

Figure 4 focuses on the call establishment process according to the test-bed, where \( c_{user} \) will try to establish a call with the roaming user \( r_{user} \). Again we compare the INVITE message flow of H-SIP to the traditional SIP. For this section, cdomain is ece.unm.edu [New Mexico], the hdomain is issec.org [Panama] and the fdomain is cnri.reston.va.us [Virginia]. For brevity, the 100 TRYING messages are excluded from Figure 4. We focus on the INVITE message flow and the servers require no authentication. We denote by \( r_t(m,n) \) the round-trip communication delay between nodes \( m \) and \( n \). For example, in Figure 4, \( r_t(cdomain,hdomain) \) is the round-trip delay between
the cdomain SIP server and that of hdomain. If we estimate the server’s average processing time of the INVITE request by \( \beta \), and the UA’s average processing time of the INVITE request by \( \gamma \), then again from Figure 4.A,

\[
\lambda \approx r_t(c_{\text{user,cdomain}}) + \beta + \phi \quad (3a)
\]

\[
\phi \approx r_t(\text{edomain,hdomain}) + \beta + \mu \quad (3b)
\]

\[
\mu \approx r_t(\text{fdomain,r_user}) + \gamma \quad (3c)
\]

and therefore,

\[
\lambda \approx r_t(c_{\text{user,cdomain}}) + r_t(\text{edomain,hdomain}) + r_t(\text{fdomain,r_user}) + 2\beta + \gamma \quad (3d)
\]

where \( \lambda \) is the average call establishment time (the INVITE/RINGING round-trip) as seen by \( c_{\text{user}} \), \( \phi \) is the average INVITE/RINGING round-trip time between the cdomain and hdomain SIP servers and \( \mu \) is the average INVITE/RINGING round-trip time between the hdomain SIP server and \( r_{\text{user}} \).

Now, if we consider H-SIP, then from Figure 4.B,

\[
\lambda' \approx r_t(c_{\text{user,cdomain}}) + \beta + x + \phi' \quad (4a)
\]

\[
\phi' \approx r_t(\text{fdomain,fdomain}) + \beta + \mu' \quad (4b)
\]

\[
\mu' \approx r_t(\text{fdomain,r_user}) + \gamma \quad (4c)
\]

and therefore,

\[
\lambda' \approx r_t(c_{\text{user,cdomain}}) + r_t(\text{fdomain,fdomain}) + r_t(\text{fdomain,r_user}) + 2\beta + x + \gamma \quad (4d)
\]

where \( \lambda' \) is again the average call establishment time as seen by \( c_{\text{user}} \), \( \phi' \) is the average INVITE/RINGING round-trip time between the cdomain and fdomain SIP servers, \( \mu' \) is the average INVITE/RINGING round-trip communication time between the fdomain SIP server and \( r_{\text{user}} \), and \( x \) is the average handle resolution delay. Note here that each INVITE request will require the server to issue one fresh handle resolution.

Both \( \lambda \) and \( \lambda' \) are measured by \( c_{\text{user}} \) as the time between sending the INVITE request and receiving the RINGING response.

V. PERFORMANCE MEASUREMENTS

The measurements of the registration times, call establishment times and communication delays in this section were all averaged from 10,000 samples dispersed over a 10 day period i.e. \( n = 1000 \) samples a day\(^3\).

A. Registration

Examining equations 1a and 2a, we deduce that \( t_B \leq t_A \) for sufficiently large \( n \). In the general case of a roaming user, the round-trip delay \( 2y \) is expensive. Our real-time measurements are listed in Table I. Clearly our measurements show that equations 1a and 2a hold for an \( \alpha \approx 50\text{ms} \). We see here that \( t_A \) is approximately \( 5t_B \) or \( 7t_B \) for the virginia and panama cases respectively. Obviously, the H-SIP approach outperforms the traditional approach as long as \( y \) is significant, which is often the case for roaming subscribers. The value of \( \alpha \) directly depends on the implementation of the SIP server which is the JAIN-SIP Proxy server [16] in this case. Besides, \( x \) is random and it directly depends on the location of the Local Handle Server (LHS) [11] storing the particular handle.

B. Call Establishment

In comparing the call establishment time given by equations 3d and 4d, we are interested in computing the performance enhancement or degradation \( \Delta \lambda = \lambda' - \lambda \). We consider the following variables to verify our model:

\[
\tau_1 = r_t(\text{hdomain,r_user}) - r_t(\text{fdomain,r_user}) \geq 0 \quad (5a)
\]

\[
\tau_2 = r_t(\text{hdomain,hdomain}) - r_t(\text{fdomain,fdomain}) \geq 0 \quad (5b)
\]

\[
\sigma = \tau_1 + \tau_2 > 0 \quad (5c)
\]

where \( \sigma \) is the difference in the cumulative round-trip delay between the traditional SIP case and the H-SIP case respectively. Equation 5a is true in general due to the presumably large geographical separation between the roaming user and her home server versus using a local server. We also argue that equation 5c holds in general based on our model i.e. the cumulative round-trip delays for \( c_{\text{user}} \) to reach \( r_{\text{user}} \) is smaller in the H-SIP scenario. However, \( \tau_2 \) in equation 5b can be positive or negative since it directly depends on the cdomain-hdomain separation versus that of cdomain-fdomain. For our particular setup, \( \tau_2 > 0 \).

It follows from equations 3 and 4 that,

\[
\Delta \mu = \mu - \mu' \approx \tau_1 \quad (6a)
\]

\[
\Delta \phi = \phi - \phi' \approx \tau_2 + \Delta \mu \quad (6b)
\]

\[
\Delta \lambda \approx \Delta \phi - x \quad (6c)
\]

Notice here that,

\[
\Delta \lambda \approx \tau_1 + \tau_2 - x = \sigma - x \quad (6d)
\]

Consequently, H-SIP outperforms the traditional SIP approach, in general, as long as \( \sigma > x \). Our conducted real-time measurements are listed in Table II. In Table II, we verify the validity of our model by separately comparing the real-time call setup measurements to the round-trip delays, thus asserting equations 6.

C. DNS vs Handle-DNS resolution

Obviously, the only performance degradation introduced by our approach is the handle resolution overhead. The handle resolution measurements that this paper reflected are all

---

\(^1\)The experiments were conducted during the month of December in 2006.
TABLE II
(A) COMPARISON OF AVERAGE CALL SETUP DELAYS (AS SHOWN IN FIG. 4) (B) AVERAGE ROUND-TRIP COMMUNICATION DELAYS, TRANSPORT=UDP

<table>
<thead>
<tr>
<th>Call Setup Delays [ms]</th>
<th>Round-trip Delays [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \lambda$</td>
<td>$\Delta \phi$</td>
</tr>
<tr>
<td>88</td>
<td>168</td>
</tr>
</tbody>
</table>

Fig. 5. Handle implementation performance measurement as of August 2005. Acquired through the courtesy of Mr. Sam Sun and CN-NIC.

Fig. 6. Handle load performance measurement as of August 2005. Acquired through the courtesy of Mr. Sam Sun and CN-NIC.

The system has also been extensively tested by CNRI along with load assessment. Results, as shown in Figure 6, are very encouraging as they show that the Handle System can potentially replace the DNS system.

VI. FUTURE WORK

We intend to leverage the scalable resolution, security, and administration services of the Handle System and use it to replace the DNS system within the SIP protocol. According to RFC 3263, the main reason SIP needs to use DNS is to enable the originator domain proxy to locate the SIP proxy in the destination domain (IP, port and transport protocol). The other need for DNS in SIP is for the terminating proxy to identify a backup for the originating proxy in the case the latter fails. Replacing the DNS with SIP requires carefully examining all the DNS resolutions performed by UA clients and proxy servers according to RFC 3263 and formalizing the fields to be resolved within the handles. Finally, part of our current research is to focus on implementing a structured peer-to-peer form of the Handle System to expedite lookups and resolutions and eliminate single points of failure.

VII. CONCLUSION

In this paper, we have reviewed the H-SIP mobility scheme and experimentally proven its efficiency in achieving inter-domain authentication and call routing through modeling and real-time measurements. On top of addressing the domain resolution itself, the H-SIP approach minimizes the signaling traffic needed by a roaming user to join the SIP infrastructure and be ready to initiate calls. The user is efficiently utilizing the services of a local server with no need for per-call coordination with her home server. If the home server is located in another continent per se, the round trip times for registration/redirect messages from the roaming user to home proxy/correspondent user respectively become significant. H-SIP optimizes the association, authentication and call routing times for the roaming user by always trying to address the mobile user through the closest available server in her vicinity.

VIII. ACKNOWLEDGEMENTS

We would like to thank the Corporation for National Research Initiatives CNRI for their support as well as the Ibero American Science and Technology Consortium (ISTEC) and the University of New Mexico that provided the VoIP infrastructure for the test bed of this work.

REFERENCES