



A Survey of SIP In Distributed Systems

by

Fei Yao

Li Zhang

Wen Hu

Supervisors: Morten Goodwin Olsen
Andreas Häber

Project report for IKT 404 in Spring 2007

Agder University College
Faculty of Engineering and Science

Grimstad,

Status: Draft

Keywords: SIP Distributed systems VoIP IMS

**Abstract:**

SIP is a signalling protocol used for establishing, modifying and terminating sessions with one or more participants on the Internet. The emergence of SIP promises simple and efficient handling of multimedia sessions among multiple users. This article gives a survey of SIP in distributed systems. First, we provide an overview of SIP by presenting its characteristics and functionality, and also describe the relationship between SIP and distributed systems. Secondly, we present current issues about SIP, such as QoS, security and NAT/Firewall. Finally, usages of SIP for now are discussed and it includes VoIP and IMS.

Version Control

<i>Version</i> ¹	<i>Status</i> ²	<i>Date</i> ³	<i>Change</i> ⁴	<i>Author</i> ⁵
1.0	draft	18-04-2007	3. Background	Li Zhang, Wen Hu, Fei Yao
1.1	draft	24-04-2007	4. SIP used for now	Li Zhang, Wen Hu, Fei Yao
1.2	draft	30-04-2007	Comments from Morten	Li Zhang, Wen Hu, Fei Yao
1.3	draft	05-05-2007	5. Discussion	Li Zhang, Wen Hu, Fei Yao
1.4	draft	08-05-2007	6. Conclusion	Li Zhang, Wen Hu, Fei Yao
1.5	draft	11-05-2007	Comments from Morten	Li Zhang, Wen Hu, Fei Yao
1.6	draft	13-05-2007	Comments from Andreas Häber	Li Zhang, Wen Hu, Fei Yao
1.7	final	14-05-2007	Figures and references	Li Zhang, Wen Hu, Fei Yao

1 **Version** 1.5

2 **Status** FINAL

3 **Date** 2007-05-11

4 **Change** Improvement based on draft v 1.4

5 **Author** Fei Yao, Li Zhang, Wen Hu

Table of Contents

1 Introduction.....	1
1.1 Report outline.....	1
2 Problem description	1
3 Background (Review of literature).....	2
3.1 Introduction of SIP.....	2
3.1.1 SIP overview	2
3.1.2 SIP elements.....	2
3.1.3 SIP Messages.....	3
3.1.3.1 Start-line.....	4
3.2.3.1.1 Request.....	4
3.2.3.1.2 Response	5
3.2.3.1.3 Message header fields	5
3.2.4 Examples of SIP UA Behavior.....	6
3.2 SIP used in distributed systems.....	8
4 Current issues about SIP	10
4.1 SIP Security.....	10
4.1.1 SIP security mechanisms.....	10
4.1.1.1 Network and transport layer security	11
4.1.1.2 SIPS URI scheme.....	12
4.1.1.3 HTTP Authentication	12
4.1.1.4 S/MIME	12
4.1.2 Implementing Security Mechanisms.....	14
4.2 NAT/firewall traversal.....	14
4.2.1 ALG.....	14
4.2.2 STUN	15
5. Usages of SIP.....	16
5.1 SIP and VoIP	16
5.2 SIP used in IMS	17
5.2.1 Architecture of IMS.....	17
5.2.2 Protocols and interfaces in IMS	18
5.2.3 SIP signaling in IMS	19
5.2.3.1 Registration	19
5.2.3.2 Call session setup.....	22
6. Conclusion	24
Appendix.....	25
A1 Glossary & Abbreviations	25
A2 Reference	26

Figure List

Figure 1 UAC and UAS	3
Figure 3 Proxy mode.....	7
Figure 4 Redirect mode.....	7
Figure 5 SIP session flow.....	8
Figure 6 C/S architecture of distributed system	9
Figure 7 Example of a tunneled “message/sip” body.....	13
Figure 8 Typical NAT Configurations [10]	14
Figure 9 STUN operations [20].....	15
Figure 10 General architecture of SIP-based VoIP [12]	16
Figure 11 3GPP IMS architecture [16].....	18
Figure 12 Protocols used in IMS [18]	19
Figure 13 Complete registration flow in the IMS [16].....	21
Figure 14 IMS call session setup	22

1 Introduction

The next generation of communication systems will provide high quality multimedia services in a flexible way. In 2000, SIP (Session Initiation Protocol) was selected by the Third Generation Partnership Project (3GPP) as the call control protocol for the 3G IP-based mobile networks. SIP [11] is a signaling protocol which has been developed to establish sessions in an IP network without knowing details about the session.

SIP is one of the key innovations driving the current evolution of communication system. Its main utility is in Internet telephony. However, SIP is not limited to it, it is already employed for voice sessions, transmission, push-to-talk, or other Internet based communication mechanism, with examples such as distributed games, shared application, shared text editors and white boards. What is more, the most important thing about SIP is that it can combine these applications to provide a large-scale and seamless service.

In the report, we will give a survey of SIP in distributed systems. We will focus on relationship between SIP and distributed systems and the newest research trends in its related fields, and point out the future research direction and the promising foreground application.

1.1 Report outline

This report is organized as follows. Section 1 gives an introduction of the report which is the current chapter. Section 2 will describe the problem of the project. In section 3 we present the relationship between SIP and distributed system, and the basic theory about SIP. Current issues about SIP will be shown in section 4. Section 5 discuss how SIP work in the future communication field compactly. We conclude by some remarks and future work.

2 Problem description

With the ascent of the Internet, the communication providers and users wanted to introduce some new service which are based on IP and are more open. The IETF (Internet Engineering Task Force) came up with SIP. SIP is a signaling protocol which has been developed to establish sessions in an IP network without knowing details about the session. As a mature all-IP technology, how SIP provides a better service and uses in a wide range for now and future communication field is our concern.

3 Background (Review of literature)

3.1 Introduction of SIP

3.1.1 SIP overview

The Session Initiation Protocol (SIP) [4] is a signaling protocol for initiating, managing and terminating multimedia sessions across packet networks. SIP is an end-to-end, client-server session signaling protocol. Therefore, SIP allows two or more participants to negotiate how they are going to communicate. The participants are able to communicate by multicast, unicast or both methods. SIP is not only used for Internet telephony, but also for voice sessions, or other Internet based communication mechanism, with examples such as distributed games, shared application, shared text editors and white boards having been demonstrated in practice [5]. Once the session has been established by using SIP, the two participants will select another method to communicate and be unrelated to SIP. For example, the audio and video packets are usually transported by means of RTP (Real-time transport protocol), and SDP (Session Description Protocol) is used to describe the media content of the session. SIP can also used to update sessions' information, for example, to add another stream to a conversation. Finally, SIP is used to terminate the session.

3.1.2 SIP elements

In SIP, a UA is the endpoint entity and it has the function of generate and answer requests. Soft phones, SIP telephones and some computer application (Windows Messenger) could be used as UAs. In SIP, the UA is divided into user agent client (UAC) and user agent server (UAS). The role of UAC is to originate a request, while UAS is to respond to the request. The figure 1 is the relationship between UAC and UAS. It is possible that two UAs can communicate with each other without any SIP infrastructure. However, the approach is not practical for a public service [6]. Therefore, some special UAs which works as network servers are exist, as well, namely proxy server, redirect server, etc.

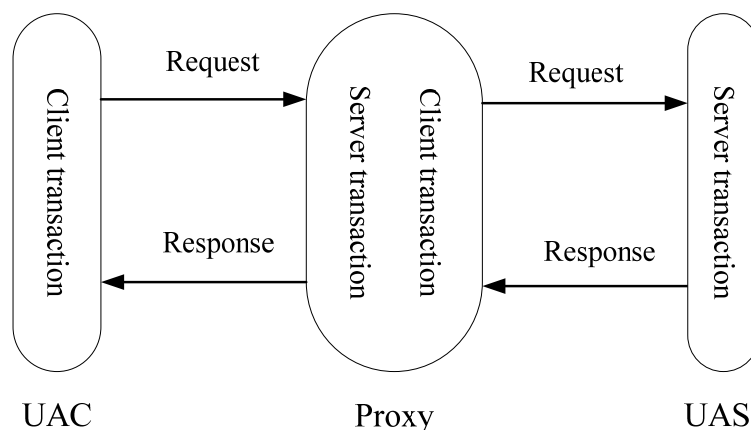


Figure 1 UAC and UAS

The role of proxy server is to send a request to a server which determines the next hop signaling. There are two kinds of proxy server: stateful and stateless proxy server. If a proxy is stateful it will keep track of received messages in the past, while the stateless proxy server will forget the information as soon as transaction over. Therefore, stateful proxy server is able to transact multi-user's requests and respond, while the efficiency of stateless proxy server is higher than stateful server. Proxy server is backbone of SIP.

Redirect servers are used to redirect callers to other servers.

The main function of Registrar is to register UAS. In SIP, UAS has to register in a server in order to let UAC find it by the server.

Additionally, there is also another special UA we should mention here, that is, Back-to-Back User Agent (B2BUA). "The Back-To-Back User Agent (B2BUA) is a Session Initiation Protocol (SIP) based logical entity that can receive and process INVITE messages as a SIP User Agent Server (UAS). It also acts as a SIP User Agent Client (UAC) that determines how the request should be answered and how to initiate outbound calls. Unlike a SIP proxy server, the B2BUA maintains complete call state and participates in all call requests". [21]

3.1.3 SIP Messages

SIP messages for set up or terminate the multimedia communications usually with five different aspects which are user location, user availability, user capabilities, session setup and session management. User location is used to show which system will be used for communication while the user capabilities show what media and media parameters will be used. Session management has some operations to the sessions such as transfer the sessions, modify session parameters, etc.

SIP messages are defined into two different types that are Request and Response. All the messages consist of a start-line and message header fields and have an optional message body.

3.1.3.1 Start-line

3.2.3.1.1 Request

In Request messages the start-line is called Request-Line which contains three parts: method, Request-URI and SIP protocol version. The format is show as below:

Request-Line: method name SP Request-Line SP SIP protocol version [4]

(SP stands for space between two parts.)

● method

The method has SIP basic request methods and some SIP extended request methods.

There are six SIP basic request methods;

- (1) INVITE: used for invite user or server attend a session.
- (2) ACK: Client uses ACK to tell the server that the response to INVITE request is accepted.
- (3) CANCEL: used to cancel a request.
- (4) OPTIONS: this method is used to query servers about their capabilities.
- (5) BYE: Clients show that they want to release the sessions to servers.
- (6) REGISTER: register an address with contact information.

There are many SIP extended request methods [4], following we describe some of the methods.

- (1) INFO method [22]: used to transfer extend messages of application layer.
- (2) Extension supported by the SIMPLE protocol [23]:
SIMPLE is SIP instant message (IM) and presence (P) extension. Compare with SIP, SIMPLE add another SIP request method: MESSAGE. And there are other two extend method: SUBSRIBER, NOTIFY.
-MESSAGE [23]: used to carry the content of instant message in request body.
-SUBSCRIBE [4]: used to destine the notification for status change of the port.
-NOTIFY: used to inform the status change situation.
- (3) REFER method [24]: used to specify transfer requests.

● Request-URI

A Request-URI "is a SIP or SIPs URI or a general URI". [4] In a Request-URI contains the address that the request is being addressed. A SIP or SIPs URI has the format as showed below:

sip (or sips): password@host:port; uri-parameters ? headers [4]

The host and the port show where the request is to be sent.

Uri-parameter has the form as: parameter-name=" parameter-value

3.2.3.1.2 Response

In Response messages the start-line is called Status-Line which contains three parts: SIP protocol version, status-code and reason-phrase CRLF. The format is show as below:

Status-Line: SIP protocol version SP Status-code SP Reason-Phrase CRLF

- **Status-code**

The status-code is a 3 digit integer result code and the first digit shows different types of response. In RFC 3261 it defines those types as following shows:

- 1XX: Provisional-request received continuing to process the request
- 2XX: Success-the action was successfully received, understood, and accepted.
- 3XX: Redirection- further action needs to be taken in order to complete the request.
- 4XX: Client Error- request contains bad syntax or cannot be fulfilled at this server.
- 5XX: Server Error-the server failed to fulfill an apparently valid request.
- 6XX: Global Failure-request cannot be fulfilled at any server.” [4]

3.2.3.1.3 Message header fields

In the message header fields there are many different types of header, we use an example as showed in figure 2 to describe several headers that are in common use.

Route header and Record-Route header:

Both Route header and Record-Route header belong to route set field. In those header contains SIP or SIPs URI which describe the path to transfer a request.

Via header:

Via header records the port used for transaction and tell the address where the response will be sent. The bottom via header is inserted by the User Agent who initialize the request, and other via above are insert by proxies on the router path.

From header, To header and Call-ID header:

From, To and Call-ID headers together built up the dialog information that can identify an unique dialog.

-From header field usually contains a SIP URI and a display name to show the identity of the user that initialized the request.

-To header field used to indicate the addresses that the request will be received, these addresses are showed by a SIP or SIPs URI or any other URI schemes.

-Call-ID field is shows in the form:

cryptographically random identifiers @ host name (or IP address)

CSeq header:

CSeq stands for Command Sequence Number, it used for identify and order transactions. It consists of a sequence number which can be a arbitrary 32 digit integer and the request method.

Contact header:

Contact header field including a SIP or SIPs URI that indicate the address where the request will be accepted. If there is a displayed name in the contact header, then all the URI parameter has to be enclosed in "<>".

Max-Forwards header:

This header field is used to limit the number of hops during the process that the request be transferred. If a request message contains a Max-Forwards and the value is zero, then the SIP hops cannot be re-sent, a 483 (Too Many Hops) response will be return.

```
Request-Line: INVITE sip:bob@open-ims.test SIP/2.0
Message Header
  Route: <sip:orig@scscf.open-ims.test:6060;lr>
  Record-Route: <sip:mo@pcscf.open-ims.test:4060;lr>
  Via: SIP/2.0/UDP 192.168.1.7:4060;branch=z9hG4bK17b.09719f96.0
  Via: SIP/2.0/UDP 192.168.1.12:5060;rport=5060;branch=z9hG4bK2099595392
  From: "Alice" <sip:alice@open-ims.test>;tag=1990381184
  To: <sip:bob@open-ims.test>
  Call-ID: 1621887217@192.168.1.12
  CSeq: 20 INVITE
  Contact: <sip:alice@192.168.1.12:5060>
  Max-Forwards: 16
  User-Agent: UCT IMS Client
  Subject: IMS Call
  Expires: 120
  Privacy: none
  Require: sec-agree
  Proxy-Require: sec-agree
  Supported: 100rel
  Allow: INVITE, ACK, UPDATE, INFO, CANCEL, BYE, OPTIONS, REFER,
SUBSCRIBE, NOTIFY, MESSAGE
  Content-Type: application/sdp
  Content-Length: 322
```

Figure 2 example of an INVITE message header fields

3.2.4 Examples of SIP UA Behavior

Here we have two examples of SIP UA behavior, the following figure 3 and figure 4 show them respectively. It is worth noting that there is an important step before any of these two modes can be executed: the user must first register to a SIP server about its current location, in this case, IP 131.161.1.112.

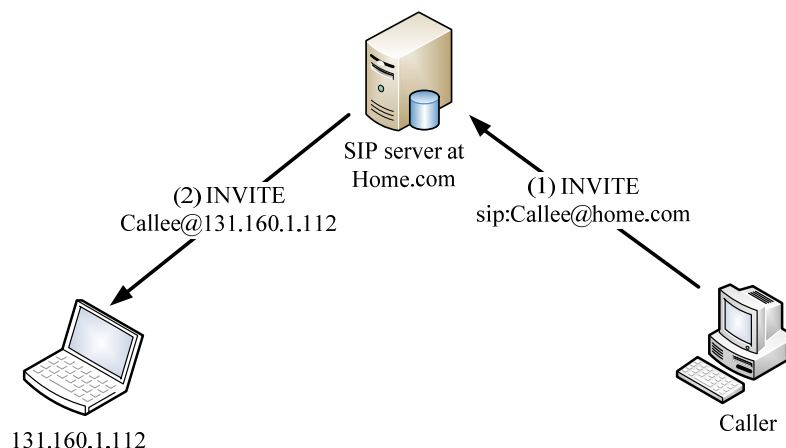


Figure 3 Proxy mode

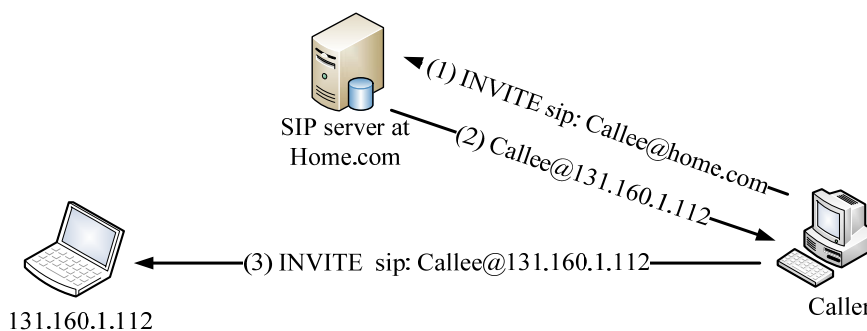


Figure 4 Redirect mode

The difference here is that in redirect mode, Proxy will tell the caller about the callee's address while in Proxy mode, Proxy will act as a mid-ware by calling the callee directly.

As an example, a caller, A, would like to speak to B and will use SIP to establish a voice session among them. Then SIP message carrying a session description will be sent from A's proxy server to B's proxy server. SIP distributes the session description and invites B to participate in the session. B's server will search for B by means of Location server and the Location server will indicate that user B logs off, so the calling fails. In order to know whether user B logs on or not, user A subscribes to B's Presence service. User A will re-send a request to user B when A is aware of B's appearance. The conversation starts when the connection is established. Finally, user B confirms the message sent by A who wants to terminate the conversation. The following figure 5 is the SIP session flow.

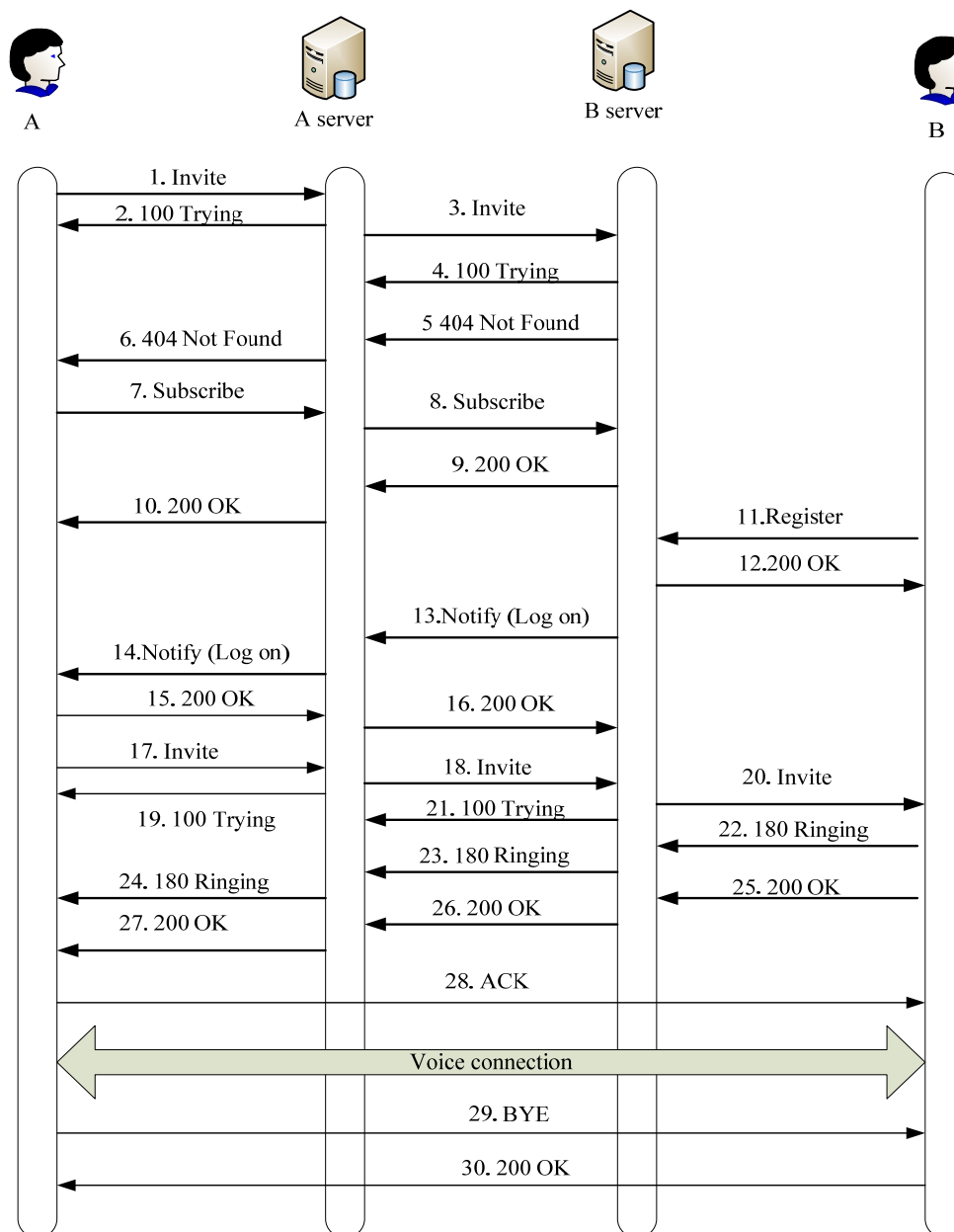


Figure 5 SIP session flow

3.2 SIP used in distributed systems

With the development of network, distributed systems become more and more popular. Nowadays, distributed systems, particularly the Web and other Internet-based applications and services, have an unprecedented interest and importance.

Distributed system is the process of aggregating the power of several computing entities to collaboratively run a single computational task in a transparent and coherent way, so that they appear as a single, centralized system. [1]

As the definition described above, we can see there are two aspects. The first one deal

with hardware: the machines are autonomous. The second one deal with software: the users think they are dealing with a single system [2]. Both are essential and have some goal; to connect users and resources in a transparent, open, and scalable way [1].

One of the most important structures of distributed systems is also called C/S (client/server) architecture. C/S architecture is one of the most successful architecture in developing software application. It has been the basic idea in software design and development. In this kind of structure, it can be understood as: The Client sends a request to the Server, and then, the Server will use some methods to process this request, and send back the result to the Client. Note that, when the Server processes the request from the Client, itself can also be used as a Client in another C/S architecture. The figure 6 shows the main idea about C/S architecture in distributed systems.

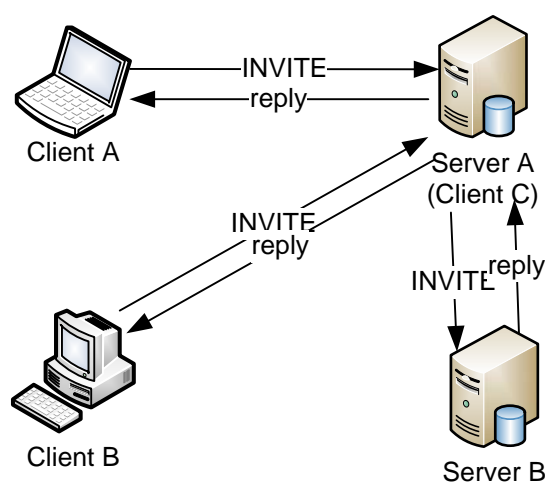


Figure 6 C/S architecture of distributed system

From the figure 6, we can find that SIP is used in the C/S architecture. SIP is a signaling protocol for initialing, managing and terminating multimedia sessions across packet networks. It is an end-to-end, client-server session signaling protocol. [3]

In recent years, people always discuss about the specification in SIP and H.323. The H.323 is used mostly before because of it is originated from ISDN, which is more stable and mature. It is widely accepted in enterprise solution. But with the research and the development in SIP, especially as SIP is chosen as the protocol in IMS (IP Multimedia Subsystem) by 3GPP, people tend to use SIP as the key protocol in NGN (next generation network).

4 Current issues about SIP

4.1 SIP Security

As we know, attacks from the Internet are usually hard to trace and there is potential susceptibility to danger for communication on the Internet, so security threats in the Internet environment is common. SIP transports information through IP networks, therefore, SIP security is considered as a very important problem and the solutions for this problem are still under improving nowadays. SIP security including two different aspects, one is the confidentiality and integrity of the messages, another is the IP network security.

For SIP-based networks, there are two kinds of possible threats: external threats and internal threats [7]. External threats usually occur when the voice and signaling packets traverse some unsafe networks, network equipments or proxies. Internal threats usually launched by a SIP caller or SIP callee which make the situation more difficult and complex. Usually, enterprises consider the users within firewalls are trustworthy, so once those safe users have security problems, it will be very dangerous to the SIP-based network because the difficulty of identifying and tracing the attack.

In [7], the author summarizes threats of general network and application-level security issues and come up with five different types of attacks:

- Denial of service (DOS) attacks: Attacks can stop the service to SIP proxy servers or gateways in the network, usually attackers send lots of unauthorized packets to make the servers stop working.
- Eavesdropping: Eavesdropping in network usually are voice packets or Real-Time Transport Protocol (RTP) media stream been unauthorized intercepted and decoded, in this way the attacker steal the signaling messages.
- Packet spoofing: Attackers camouflage to legal users and sending data with threats.
- Replay: Attackers retransmit a legal spurious message to the callee so that the callee has to reprocess this spurious message.
- Message Integrity: Attackers insert offensive data in to messages to destroy the integrality of sent messages.

Following we will describe how to ensure the security of communications when using SIP protocol under kinds of threats from the IP network.

4.1.1 SIP security mechanisms

SIP chooses to reuse existing security models defined in HTTP and SMTP space,

instead of define new security mechanisms.

SIP cannot use fully encryption to protect messages because there are some message fields need to be visible so that SIP requests can routed correctly. Those message fields usually are Request-URI, Via, From and To, Route fields. So SIP protocol provides multimedia end-to-end, hop-by-hop security mechanisms. End-to-end security mechanisms related to the SIP proxies of participants in a call session, such as SIP authentication and encryption of SIP message bodies. Since SIP has no specific secure functions for hop-by-hop basis, so it mainly relies on network and transport layer security.

Following details some common security mechanisms used in SIP.

4.1.1.1 Network and transport layer security

Network and transport layer security “encrypts signaling traffic, guaranteeing message confidentiality and integrity” [4].

At the network and transport layer, IP Security (IPSec) and Transport Layer Security (TLS) used to provide security.

- IPSec

IPSec is a set of protocols that used to ensure the security of IP communications. It operates at the network layer and there are two models of IPSec operation: transport mode and tunnel mode [8].

In transport mode which is used for host-to-host communications only the message of the IP packet is encrypted. In tunnel mode the whole IP packet is encrypted and this packet will be encapsulated in order to routing. Different from the transport model, the tunnel mode can use not only for host-to-host communications but also for host-to-network or network-to-network communications.

IPSec can provide a suite of security services, such as to reject replaced packets, data authentication, encryption, access control and so on by choosing required protocol and algorithms for use. Two protocols, Authentication Header (AH) and Encapsulating Security Payload (ESP) are used in IPSec for traffic security.

- TLS

The TLS operates at transport-layer and provides endpoint authentication by using cryptography. It usually used for host-to-host communications that do not require pre-existing trust relationship.

The TLS protocol exchange records, the “TLS record protocol is a layered protocol” [9], each record can be transmitted, optionally compressed, encrypted with a message authentication code (MAC).

There are four record protocols in common use that are: handshake protocol, alert protocol, change cipher spec protocol and the application data protocol [9]. The TLS record protocols support additional record types, if there are any new record protocols

then the type values should be set higher than the ContentType values for those four record types.

4.1.1.2 SIPS URI scheme

The SIPS URI scheme use the same syntax as the SIP URI, however, the scheme is “sips” instead of “sip”. A SIPS URI illustrates the resource be contacted. It can be used as a Request-URI in the From or To header fields to show hops where the request is forwarded before it reaches the destination. That means the TLS is to be used between UAC and the domain that owns the Request-URI.

4.1.1.3 HTTP Authentication

In reference [4] it introduced that SIP provides a challenge capability that is based on authentication in HTTP. The reuse of the HTTP Digest authentication in SIP can protect against the replay attacks and provide message authentication.

The Digest authentication in SIP follows the fundamental rules described below:

The URI= SIP-URI or SIPS-URI;

The digest-uri-value=Request-URI;

That is in the authentication message should contain the URI and the digest-uri-value, and the URI has to be SIP URI or SIPS-URI. The URI in the request line and the URI in the Authorization header field may point to different users;

The entity-body=MD5 (“”) =”d41d8cd98f00b204e9800998ecf8427e”.

4.1.1.4 S/MIME

SIP messages carry MIME bodies. S/MIME secures the MIME bodies by encrypting MIME bodies in SIP without modification of message header.

S/MIME provides end-to-end authentication, integrity as well as confidentiality. When S/MIME used for SIP header security, it can compress the SIP messages in the MIME bodies and those compressed SIP messages are used to check the integrity. When a UAS receives the request with a tunneled “message/sip” S/MIME body, then the response there should also be a tunneled “message/sip” body. If the “message/sip” body includes To, From, Call-ID or Cseq information then the signed MIME body can allow limited authentication. “Any message bodies that require integrity must be attached to the ‘inner’ message” and the message senders should copy all header fields appears in the request message in the signed body [4].

The following figure 7 is an example of the use of a tunneled “message/sip” body [4]:

```

INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com; branch=z9hG4bKnashds8
To: Bob<sip:bob@biloxi.com>
From: Alice<sip:alice@atlanta.com>;tag=1928301774
Call-ID: a84b4c76e66710
CSeq: 314159 INVITE
Max-Forward: 70
Contact: <sip:alice@pc33.atlanta.com>
Content-Type: multipart/signed;
  Protocol="application/pkcs7-signature";
  Micalg=sha1; boundary=boundary42
Content-Length: 568
-bonndary 42
Content-Type: message/sip
INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com; branch= z9hG4bKnashds8
To: Bob<sip:bob@biloxi.com>
From: Alice<sip:alice@atlanta.com>;tag=1928301774
Call-ID: a84b4c76e66710
CSeq: 314159 INVITE
Max-Forward: 70
Contact: <sip:alice@pc33.atlanta.com>
Content-Type: application/SDP
Content-Length: 147
V=0
o=UserA 2890844526 2890844526 IN IP4 here.com
s=Session SDP
c=IN IP4 pc33.atlanta.com
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000
--boundary42
Content-Type: application/pkcs7-signature; name=smime.p7s
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename=smime.p7s;handling=required
  ghyHhHUujhJhjH77n8HHGTrfvbnj756tbB9HG4VQpfyF467GhIGfHfYT6
  4VQpfyF467GhIGfHfYT6jH77n8HHGghyHhHUujhJh756tbB9HGTrfvbnj
  n8HHGTrfvhJhjH776tbB9HG4VQbnj7567GhIGfHfYT6ghyHhHUujpfyF4
  7GhIGfHfYT64VQbnj756
--boundary42-

```

Figure 7 Example of a tunneled “message/sip” body

4.1.2 Implementing Security Mechanisms

As described in section 26.3 of [4], for implement the security mechanisms the UAs should be able to implement TLS as well as behavior as a TLS server. All SIP elements should not only support TLS but also support SIPS URI scheme.

Proxy servers redirect servers, registrars should implement Digest authorization and they should be configured with at least one Digest realm. They maybe also implement IPSec or other security protocols. [8]

If a UA can validate certificates for TLS or IPSec, then it should also have the ability to check S/MIME certificates.

4.2 NAT/firewall traversal

NAT (Network Address Translation) [10] is “a method by which IP addresses are mapped from one realm to another in an attempt to provide transparent routing to hosts.” NAT is a common practice used in networks; however, it doesn’t work well with SIP. For example, if SIP server “A” is behind a NAT gateway, SIP server “B” won’t be able to contact it and it will give rise to failed calls or missing audio. Furthermore, most NATs and firewall will prevent incoming TCP connections and UDP traffic [11]. As a result, how to make SIP traversal NAT and firewall is becoming an important issue. Figure 8 shows the typical NAT configuration.

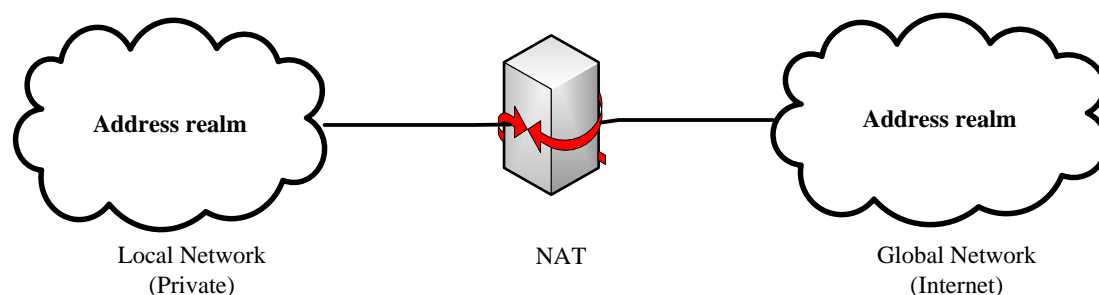


Figure 7 Typical NAT Configurations [10]

4.2.1 ALG

ALG (Application layer gateways) [10] embedded in the NAT is one of solutions. ALG is designed to be as a gateway which able to identify H.323 and SIP protocol. It doesn’t just simply examine packet header, but performs the application layer functions required for a particular protocol to traverse a NAT. When the NAT relays an IP packet form one network to another, it will first examine whether the packet contains SIP message or not. If the packet contains SIP message, it’s time for the ALG work.

The ALG will analyze data in packet payload, namely, in application layer, which contains SIP or H.323 message. The message includes port numbers of receiving audio and video data from endpoints. According to analyze which port needs use, the ALG will automatically open the port, while the other unused ports keep closed. In this way, the SIP message is able to traversal NAT.

By means of this mechanism, however, the operation of network will be affected due to the overload of firewalls. Additionally, if there are multi-NAT/firewalls in the network, it is needed that each NAT/firewalls should be updated to support the ALG.

4.2.2 STUN

Besides the ALG solution, STUN (Simple traversal of UDP over NAT) is another way to traversal NAT. The STUN protocol [11] is *“widely used by SIP endpoints to discover what their IP address and port look like on the other side of a NAT (the endpoint can then put its external addresses in all the right places, sacrificing talking to other endpoints behind the same NAT).”* STUN requires support from clients.

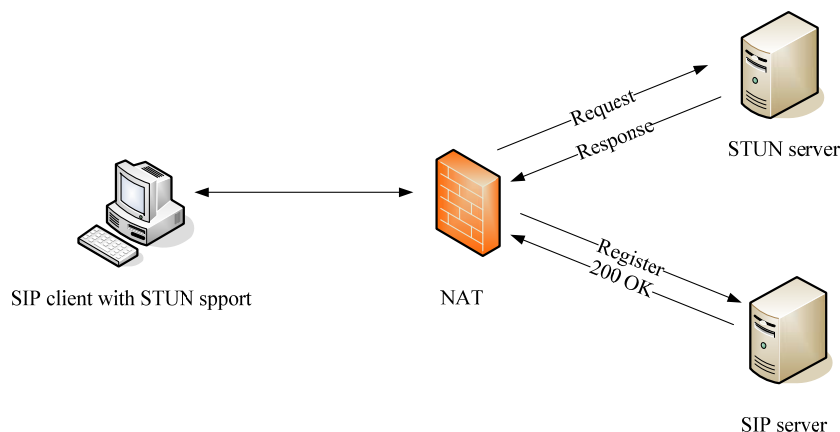


Figure 8 STUN operations [20]

The operation of STUN is shown in figure 9 [20]. STUN clients in a private network sent a request and the STUN server responds with the IP address and port in the public network by NAT which replace the original IP address and port in Contact header of SIP Registrar. Then the SIP user agent uses the modified IP address and port to register the SIP server. The registrar server is able to find the callee. In this way, SIP could successfully traversal NAT without modifications of NAT and network. It is worth noting that the connection will be closed automatically if NAT discovers it is inactive for a certain time. Therefore, a packet should be sent periodically every certain second in order to keep alive.

A drawback of STUN, however, is that STUN doesn't work with symmetric NAT. If a

client is behind a symmetric NAT, this address will not allow the establishment of the call because this type of NAT changes mapping according to the address source and of the address destination [20].

Except the stated two solutions, there are a lot of methods which developed to get SIP working in connection with firewalls and NATs [25]. Each method has this own advantage as well as disadvantage. We have to select the suitable way in practice.

5. Usages of SIP

5.1 SIP and VoIP

Today's IP networks have become an important alternative to the traditional public switched telephone networks (PSTN) to make calls, due to its lower costs, greater consumer control and location flexibility. Nowadays H.323 [13], which standardized by the ITU-T, is adopted as a protocol between IP telephone gateways by most countries. IP network is only regarded as a transmission medium in the whole IP telephone system and the gateway of IP telephone is worked as an interface between IP network and switched telephone network. However, the next generation network (NGN) [26] will provide end-to-end communication by purely utilizing IP technology. As a mature all-IP technology, SIP is a good alternative for VoIP.

The figure 10 is the general architecture of SIP-based VoIP systems. The architecture of SIP follows the client-server model and is similar to that of HTTP [14]. So SIP is easy to extend, what is more, SIP-T (SIP for telephones) [15] defined in IETF draft is a mechanism that uses SIP to facilitate the interconnection of PSTN with carrier class VoIP network.

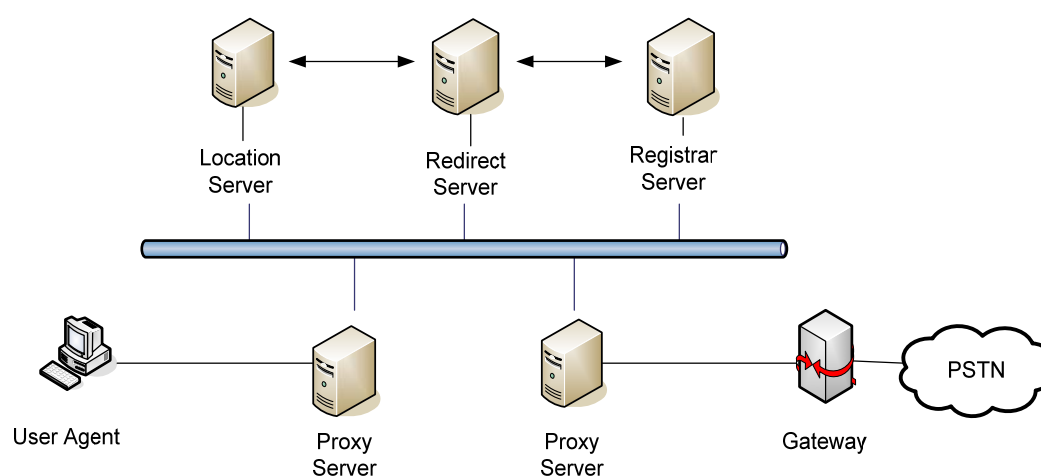


Figure 9 General architecture of SIP-based VoIP [12]

Many current IP services are lack of QoS guarantees from underlying network providers. It is also a typical problem for VoIP. With better infrastructure support, increasing network speeds and emergence of a widely acceptable and efficient protocol – SIP [12], VoIP is beginning to show its true potential.

5.2 SIP used in IMS

IMS adopted SIP protocol to control the signaling procedure as the core protocol in the 3rd generation of mobile communication. Besides, many companies also want to explore new and promising business opportunities, e.g., in enterprise solutions in the VoIP/SIP/IMS area, the interoperability of SIP and IMS shows its importance. People have discusses of this topic with regards to SIP solution recent years. IMS could provide the service with QoS, so that the quality of a VoIP conversation can have dramatically improvements.

As the key element in NGN, IMS (IP Multimedia Subsystem) plays an important role in offering key features such as QoS, security, group management, and instant voice message. IMS makes it easier for operators to provide new services, in comparison with GSM where this is very limited. The IMS is an open, standardized architecture that aims to merge multimedia services across the cellular world and IP networks, using the same standard protocols for both mobile and fixed IP services.

5.2.1 Architecture of IMS

IMS was first released in the 3rd Generation Partnership Project (3GPP) Release 5. It is further improved in Release 6 and Release 7. The 3GPP specify functions instead of nodes, so 'IMS architecture is a collection of functions linked by standardized interfaces.' [16] Figure 11 shows an overview of the IMS architecture. From the figure we can see that the main elements in IMS are CSCFs, HSS, MGCF, MGW. Besides, there are many different interfaces between components. In this, SIP is used as the main signaling protocol.

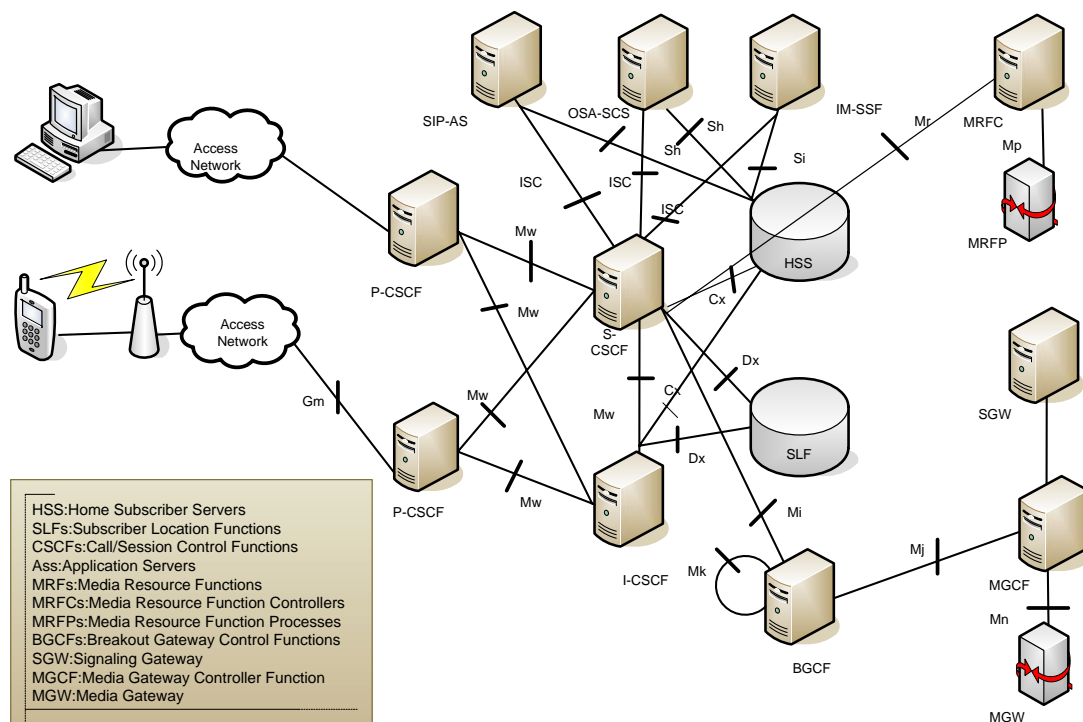


Figure 10 3GPP IMS architecture [16]

The HSS (Home Subscriber Server) is used as the user database of the IMS. It stores the authentication information of users, subscription-related information (user profiles) and users' service profiles. [17]

The Call/Session Control Function is a SIP server and it processes the call session. It is very important to IMS. CSCF includes: P-CSCF (Proxy-CSCF), I-CSCF (interrogating-CSCF), S-CSCF (Serving-CSCF). The P-CSCF locates in visited network or in the home network. It is a SIP proxy server that is the first point of contact for the IMS terminal. The I-CSCF is a SIP proxy located at the edge of an administrative domain'. [16] It is the entry point from visit network to home network, and is also the main connection of IMS and other PLMN. The DNS (Domain Name System) records of the domain stores the address of the I-CSCF. 'There may be multiple I-CSCFs within an operator's network.' [18] The S-CSCF is the key element of the signaling plane in IMS. It locates in the home network and it supports the operators to take charge of the session and the registration services.

5.2.2 Protocols and interfaces in IMS

From the figure 12, we can see that IMS has chosen SIP as session control protocol. SIP protocol is used in User Agent and Server systems. It supports user's mobility, uses Client-Server solution in HTTP protocol and it can combine many other IETF protocols. Besides, SIP supports forking and this makes SIP easy to carry out the service. [19] Expect these features, SIP is easy to program compared to mobile network protocols and widely used in the internet. It also enables convergence of fixed and mobile

internet, and can be used to implement all types of peer-to-peer applications.

The other protocols used in IMS are AAA (Authentication, Authorization and Accounting), RTP (Real-time Transport Protocol) and RTCP (Real-time Transport Control Protocol). All protocols in IMS are adopted from Internet standards (IETF).

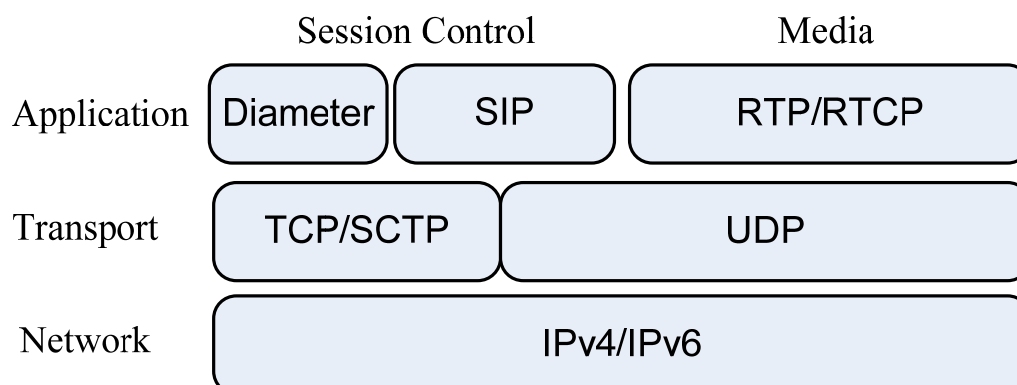


Figure 11 Protocols used in IMS [18]

5.2.3 SIP signaling in IMS

SIP in IMS is used in registration, call session setup, Instant Messaging, presence and etc. Here we discuss the SIP signaling in IMS with two examples, one is SIP signaling flow for registration and another is SIP signaling flow for call session setup. Following we will illustrate those two flows.

5.2.3.1 Registration

IMS can support SIP registration without authentication as well as with authentication, however, in practice, the IMS user always requests authentication to use the IMS services, and this is “IMS-level” registration.

As showed in the figure 13, first the SIP terminal creates a SIP REGISTER request including four parameters which are the registration URI, the Public User Identity, the Private User Identity and the Contact address.

The registration URI: this is used to identify the home network domain, and it included in the *Request-URI* of the REGISTER request.

The Public User Identity: it is a SIP URI that represents the user ID under registration. And it is included in the *To* header field value of the REGISTER request.

The Private User Identity: this is only used for authentication, we can find it in the in the *username* parameter of the *Authorization* header field value of SIP REGISTER request.

The Contact address: this is a SIP URI which includes the IP address of the IMS terminal or a host name where the user is reachable. It is contained in the SIP *contact*

header.

The SIP terminal sends the request to the P-CSCF. The P-CSCF inserts a P-Visited-Network-ID that shows where the P-CSCF is located, and the P-CSCF also inserts a Path header with its own SIP URI to request the home network to forward all SIP requests. The P-CSCF discovers an I-CSCF in the home network and forwards the SIP REGISTER request to it.

The I-CSCF queries the HSS with a Diameter User-Authorization-Request (UAR) message and the HSS answers with the Diameter User-Authorization-Answer (UAA) message and telling the user information to I-CSCF. Eventually, the I-CSCF forwards the REGISTER request to the S-CSCF.

The S-CSCF creates a Diameter Multimedia-Auth-Request (MAR) message and sends it to the HSS. The HSS then stores the S-CSCF URI in the user data for further query and answers with a Diameter Multimedia-Auth-Answer (MAA) message. In this way, the S-CSCF has downloaded the authentication data from the HSS. After this, the S-CSCF creates a SIP 401 (Unauthorized) response and forwards it to the SIP terminal via I-CSCF and P-CSCF

In response, the SIP terminal sends a new REGISTER request to the P-CSCF. Because the authentication was successful before, so when the request received by the S-CSCF, it sends a Diameter Server-Assignment-Request (SAR) message to the HSS to inform it that the user is now registered and at the same time to download the user profile. Then, the S-CSCF sends a 200 (OK) response to the SIP terminal showing that the REGISTER request is successful.

By this stage, the SIP terminal sends the SUBSCRIBER request for the event: reg to the P-CSCF and P-CSCF proxies the request to the S-CSCF. The S-CSCF acts as a SIP notifier and sends a 200 (OK) response, and the P-CSCF forwards the response to the terminal. Besides, the S-CSCF also sends a NOTIFY request and the P-CSCF forwards it to the terminal. At the end, the terminal answers with a 200 (OK) response and the P-CSCF forwards the response to the S-CSCF.

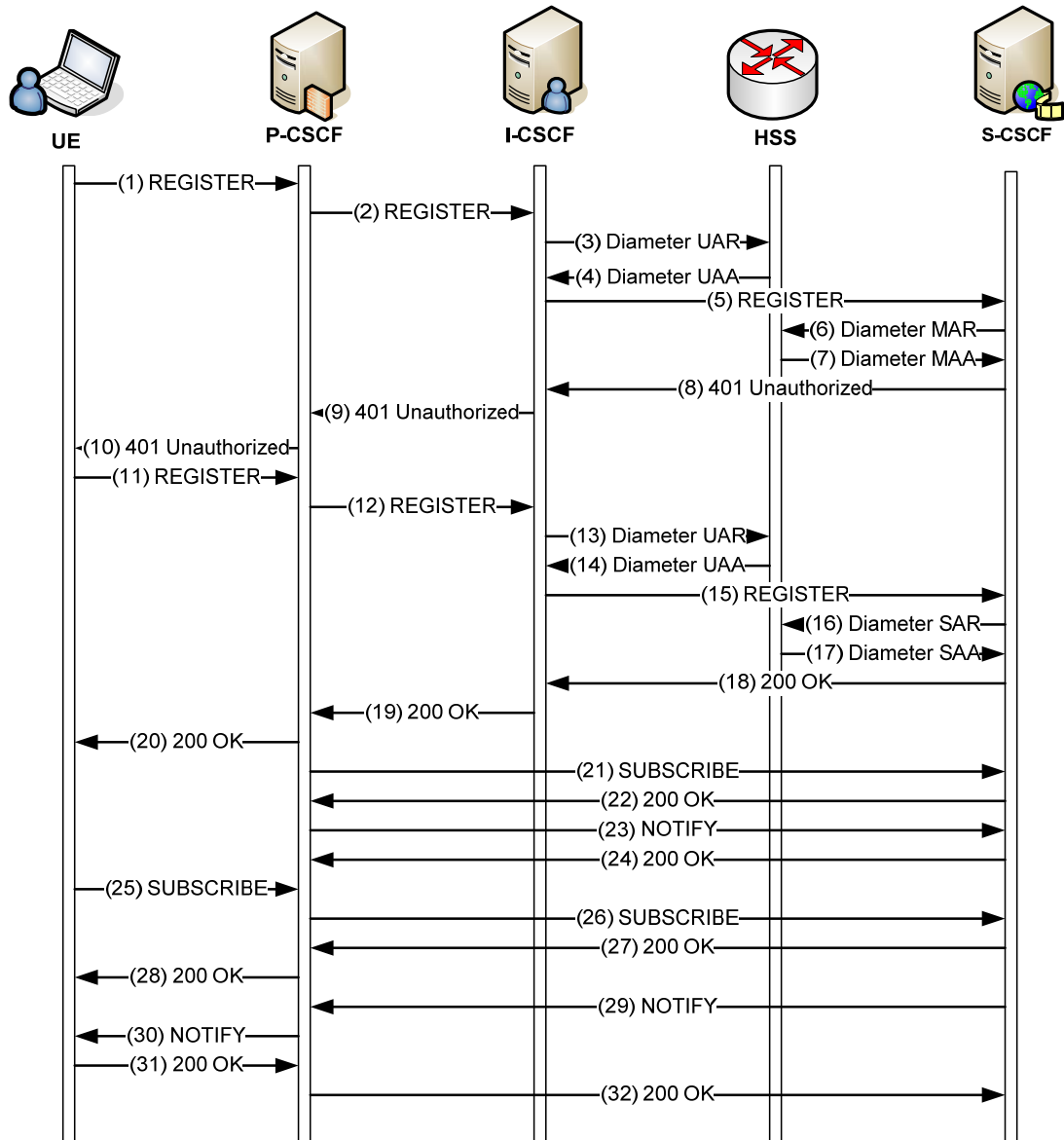


Figure 12 Complete registration flow in the IMS [16]

5.2.3.2 Call session setup

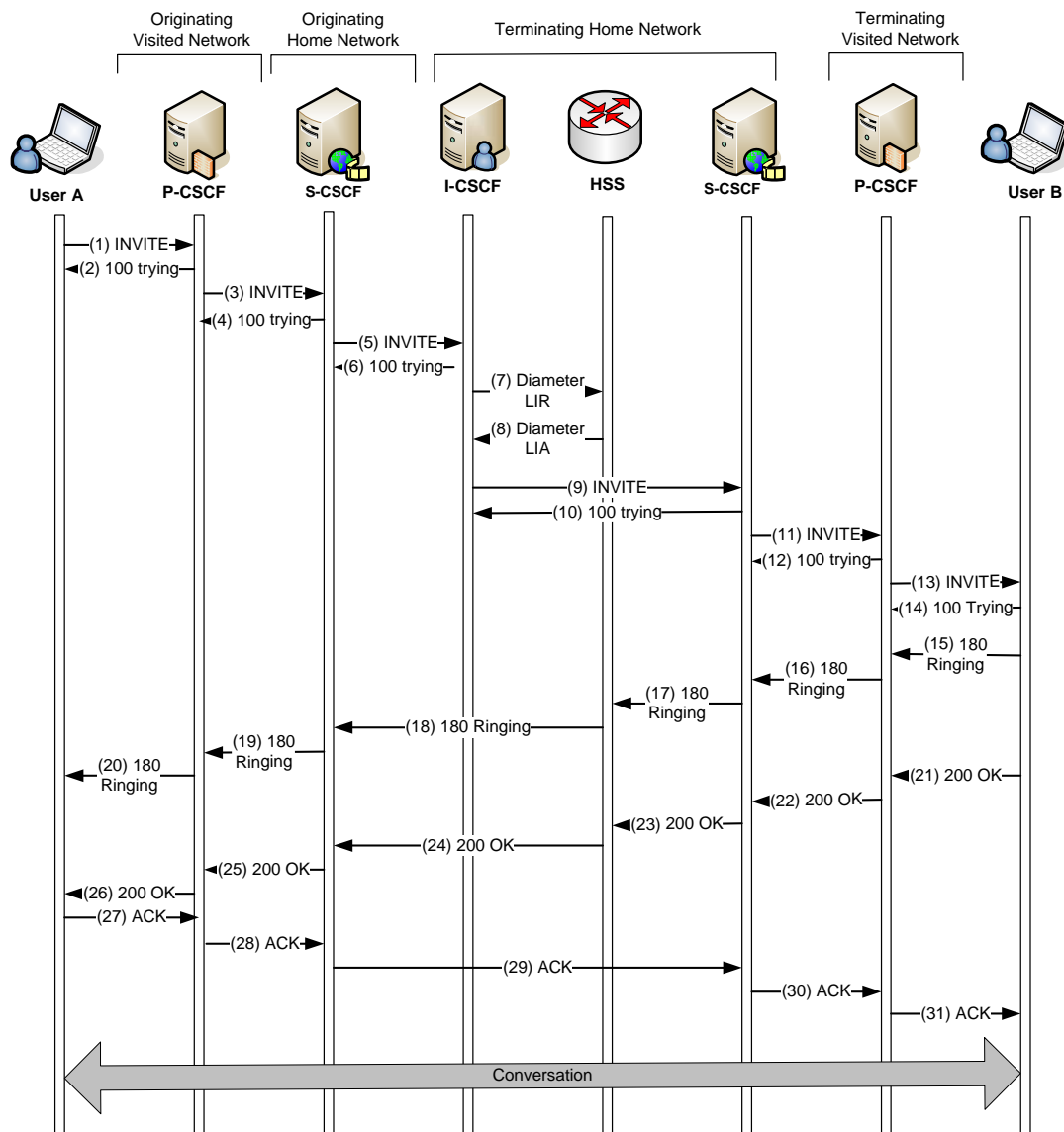


Figure 13 IMS call session setup

As showed in the figure 14, after the User A initiate an INVITE request, then:

The Originating P-CSCF Processes the INVITE response

The P-CSCF first checks if in the Router header contains the value that the Service-Router header field the IMS terminal received. Since the Router header containing the value, so P-CSCF knows that the Router header is correctly populated. Then, P-CSCF checks whether a P-Preferred-Identity header is in the INVITE request. if it exists, then P-CSCF deletes it in the INVITE and inserts a P-Asserted-Identity header field and set its value to a SIP registered Public User Identity of the user. The P-CSCF records the router and inserts a Record-Router header field with its own SIP URI when a SIP proxy wants to remain in the path for subsequent operation.

The Originating S-CSCF Processes the INVITE Request

The S-CSCF tries to route the SIP request based on the destination in the Request-URI. Since the request is forwarded within the home work, therefore S-CSCF still keeps the P-Access-Network-Info header field in the request.

The Terminating I-CSCF Processes the INVITE request

The S-CSCF has found the I-CSCF at the SIP server in the home network. The I-CSCF is used to identify the callee in the Request-URI of the INVITE request and has to forward the SIP request to the S-CSCF allocated to the callee.

To discover the address of the S-CSCF which is allocated to the callee, the I-CSCF queries the HSS where the address of the S-CSCF is stored with a Diameter Location-Information-Request (LIR) message. After the HSS received the message, it gets the stored S-CSCF address and inserts it in a Diameter Location-Information-Answer (LIA) message, sending it to the I-CSCF. Till then, I-CSCF can route the INVITE request to the S-CSCF that is allocated to the callee

The Terminating S-CSCF Processes the INVITE Request

The S-CSCF in the terminating network used to take care of the callee receives the INVITE request. First, it verifies the callee by the Request-URI in the request, and it adds its own SIP URI to the Record-Router header field value to remains in the path. Then it continues with the processing of the INVITE request. It creates a new Request-URI with the contents of the Content header field value that was registered by the callee. And it sets the Router header to that of the Path header which contains the P-CSCF in. The S-CSCF is retargets, so it inserts a P-Called-Party-ID header field which is set to the original Request-URI. Then S-CSCF forwards the INVITE request including the P-Called-Party-ID header field and at the end the request will reach to the P-CSCF.

The Terminal P-CSCF Processes the INVITE request

The P-CSCF keeps the P-Asserted-Identity header field in the INVITE request if no security is required. Also, the P-CSCF extracts the Public User Identity from the P-Called-Party-ID header of the SIP INVITE request and identifies the Public User Identity of the callee.

The Callee's Terminal Processes the INVITE Request

The UserB checks the P-Asserted-Identity header field and got the result that it presents, so the IMS terminal extracts the identity of the caller. Then it inspects the value of the P-Called-Party-ID to determine where the INVITE request is addressed to. At the same time the IMS terminal inserts a Contact header whose value is a SIP URI that including the IP address and the port number.

6. Conclusion

To summarize our whole task, we mainly illustrated the current issues about SIP including QoS, security and NAT/Firewall traversal, and we also discussed how will SIP work for future communication field, such as VoIP and IMS.

SIP has becoming more popular nowadays. It is already deployed, and the base of users is growing rapidly. For example, the IMS is the new IP multimedia systems and it uses SIP, so the access modes have nothing to do with the communication and CSCFs can act as SIP servers to transfer services directly. SIP also offers the promise of supporting a wide range of services beyond basic telephony, including instant messaging, presence management and voice-enabled web –based e-commerce.

Nowadays communication trends to wireless messages transmission and multimedia messages, under these goals the H.323 and SIP will be used in long time. However, because of SIP is more related to IP, so in the future SIP will play more and more important roles in the IMS field, and for VoIP field, it seems that SIP will replace H.323 and become the dominant standard.

As we have seen, with the improvement of interoperability of SIP, it will become main protocol for media communicate networks. The next generation of mobile phones will use SIP as the primary signaling technology.

Appendix

A1 Glossary & Abbreviations

3GPP	Third Generation Partnership Project
ACK	Acknowledge
ALG	Application layer gateways
C/S	Client/Server
DOS	Denial of service
HTTP	Hyper Text Transfer Protocol
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IPsec	Internet Protocol Security
IT	Internet Technology
MAC	message authentication code
NAT	Network Address Translation
NGN	next generation network
PSTN	Public switched telephone networks
QoS	Quality of Service
RTP	Real-time transport protocol
S/MIME	Secure Multipurpose Internet Mail Extensions
SDP	Session Description Protocol
SIMPLE	SIP instant message (IM) and presence (P) extension
SIP	Session Initiation Protocol
SIP-T	Session Initiation Protocol for Telephones
SMTP	Simple Mail Transfer Protocol
STUN	Simple traversal of UDP over NAT
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
URI	Uniform Resource Identifier
UA	User agent
UAC	User agent client
UAS	User agent server
VoIP	Voice over Internet Protocol

A2 Reference

- [1] “*Distributed computing*”, 14th May,2007
URI: en.wikipedia.org/wiki/Distributed_system
- [2] Andrew S. Tanenbaum Maarten van Stean, “*Distributed Systems – Principles and Paradigms*”, Prentice-Hall, Inc, 2002 pp 1—10.
- [3] “*Session Initiation Protocol*”, 14th May,2007
URI: http://en.wikipedia.org/wiki/Session_Initiation_Protocol
- [4] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, and J. Peterson, “*SIP: Session Initiation Protocol*”, RFC 3261, IETF, June 2002
URI:<http://tools.ietf.org/html/rfc3261>
- [5] Henning Schulzrinne and Elin Wedlun, . “*Application-Layer Mobility Using SIP*” , ACM Mobile Computing, Vol. 4, pp.47-57, 2005
- [6] Antonio Vilei, Gabriella Convertino and Fabrizio Crudo, “*A New UPnP Architecture for Distributed Video Voice over IP*”, ACM, 2006
- [7] “*Overview of SIP Security*”, Cisco IOS SIP Security Application Guide
URI:http://www.cisco.com/univercd/cc/td/doc/product/software/ios123/123cgcr/vvfax_c/callc_c/sip_c/sipsec_c/sipsecov.htm#wp1010792
- [8] “*IPSec*”, 14th May,2007
URI: <http://en.wikipedia.org/wiki/IPSec>
- [9] T.Dierks, C.Allen, “*The TLS Protocol Version 1.0*”, RFC 2246, IETF,January 1999,
URI:<http://tools.ietf.org/html/rfc2246>
- [10] Jae Cheon Han, Wook Hyun and Sun Ok Park, “*An Application Level Gateway for Traversal of SIP Transaction through NATs*”, Advanced Communication Technology, Vol 3. 2006
- [11] Robert Sparks, “*SIP: basics and beyond*”, ACM press, pp 22-33. 2007
- [12] Aameek Singh and Arup Acharya. “*Using Session Initiation Protocol to build Context-Aware VoIP Support for Multiplayer Networked Games*”, ACM, pp98-105, 2004
- [13] ITU-T Recommendation H.323, “*Packet-base Multimedia Communication Systems*”, Sep. 1999.
- [14] H. Schulzrinne and J. Rosenberg, “*Signaling for Internet Telephony*”, Proceedings of the Sixth International Conference on Network Protocols. 1998
- [15] A. Vemuri and J. Peterson, “*SIP for Telephones (SIP-T): Context and Architectures*”, IETF Internet Draft, Feb.2001
- [16] Gonzalo Camrillo, Miguel A. Garcia-Martin, “*The 3G IP Multimedia Subsystem (IMS)*”, ISBN 0-470-01818-6, May 2006

- [17] 'IP Multimedia Subsystem', 14th May, 2007
URL: http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem#Architecture
- [18] Lian Wu; Anders Aasgaard, "Migration of VOIP/SIP Enterprise Solutions towards IMS", Master Thesis Report, Agder University College, June 2006
URL: http://ikt.hia.no/aml/papers/Report_2006_Lian_and_Anders-final.pdf
- [19] Ning Xue, "Next generation of VoIP protocol", Telecommunication Technology, August 2005
- [20] Peter Koski, Jorma Ylinen and Pekka Loula. "The SIP-Based System Used in Connection with a Firewall". Telecommunications, Feb. 2006
- [21] "Back-to-Back User Agent", 14th May, 2007
URI: <http://www.vovida.org/applications/downloads/b2bua/>
- [22] S. Donovan, *The SIP INFO Method*, IETF RFC 2976, October, 2000;
URI: <http://www.rfc-editor.org/rfc/rfc2976.txt>
- [23] B. Campbell, J. Rosenberg, H. Schulzrinne, C. Huitema, and D. Gurle, *Session Initiation Protocol (SIP) Extension for Instant Messaging*, IETF RFC 3428, December, 2002;
URI: <http://tools.ietf.org/html/rfc3428.html>
- [24] R. Sparks, *The Session Initiation Protocol (SIP) Refer Method*, IETF RFC 3515, April, 2003;
URI: <http://tools.ietf.org/html/rfc3515.html>
- [25] "NAT Traversal for Multimedia over IP Services" (visited Feb. 2006)
URL: <http://www.newport-networks.com/whitepapers/nattraversal1.html>
- [26] L. Chae-Sub and D. Knight, "Realization of the next-generation network," *Communications Magazine, IEEE*, vol. 43, pp. 34-41, 2005.