SESSION INITIATION PROTOCOL

Introduction

Session Initiation Protocol (SIP) is one of the most common protocols used in VoIP technology. It is an application layer protocol that works in conjunction with other application layer protocols to control multimedia communication sessions over the Internet.

SIP is a signaling protocol used to create, modify, and terminate a multimedia session over the Internet Protocol. A session is nothing but a simple call between two endpoints. An endpoint can be a smartphone, a laptop, or any device that can receive and send multimedia content over the Internet.
SIP takes the help of SDP (Session Description Protocol) which describes a session and RTP (Real Time Transport Protocol) used for delivering voice and video over IPnetwork.

SIP can be used for two-party (unicast) or multiparty (multicast) sessions.

Other SIP applications include file transfer, instant messaging, video conferencing, online games, and steaming multimedia distribution.

Basically SIP is an application layer protocol. It is a simple network signaling protocol for creating and terminating sessions with one or more participants. The SIP protocol is designed to be independent of the underlying transport protocol, so SIP applications can run on TCP, UDP, or other lower-layer networking protocols.

Typically, the SIP protocol is used for internet telephony and multimedia distribution between two or more endpoints. For example, one person can initiate a telephone call to another person using SIP, or someone may create a conference call with many participants.

The SIP protocol was designed to be very simple, with a limited set of commands. It is also text-based, so anyone can read a SIP message passed between the endpoints in a SIP session.

SIP - Network Elements

There are some entities that help SIP in creating its network. In SIP, every network element is identified by a **SIP URI** (Uniform Resource Identifier) which is like an address. Following are the network elements −

- User Agent
- Proxy Server
- Registrar Server
- Redirect Server
- Location Server

User Agent

It is the endpoint and one of the most important network elements of a SIP network. An endpoint can initiate, modify, or terminate a session. User agents are the most intelligent device or network element of a SIP network. It could be a softphone, a mobile, or a laptop.

User agents are logically divided into two parts

- **User Agent Client (UAC)** − The entity that sends a request and receives a response.
- **User Agent Server (UAS)** − The entity that receives a request and sends aresponse.

SIP is based on client-server architecture where the caller‘s phone acts as a client which initiates a call and the callee‘s phone acts as a server which responds the call.

Proxy Server

It is the network element that takes a request from a user agent and forwards it to another user.

- Basically the role of a proxy server is much like a router.
- It has some intelligence to understand a SIP request and send it ahead with the help of URI.
- A proxy server sits in between two user agents.
- There can be a maximum of 70 proxy servers in between a source and a destination.

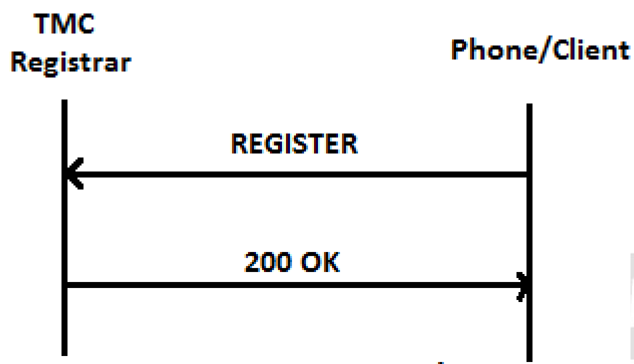
There are two types of proxy servers

- **Stateless Proxy Server** − It simply forwards the message received. This type of server does not store any information of a call or a transaction.
- **Stateful Proxy Server** − This type of proxy server keeps track of every request and response received and can use it in future if required. It can retransmit the request, if there is no response from the other side in time.

Registrar Server

The registrar server accepts registration requests from user agents. It helps users to authenticate themselves within the network. It stores the URI and the location of users in a database to help other SIP servers within the same domain.

Take a look at the following example that shows the process of a SIP Registration.



SIP Registration Link

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

Here the caller wants to register with the TMC domain. So it sends a REGISTER request to the TMC's Registrar server and the server returns a 200 OK response as it authorized the client.

Redirect Server

The redirect server receives requests and looks up the intended recipient of the request in the location database created by the registrar.

The redirect server uses the database for getting location information and responds with 3xx (Redirect response) to the user.

Location Server

The location server provides information about a caller's possible locations to the redirect and proxy servers.

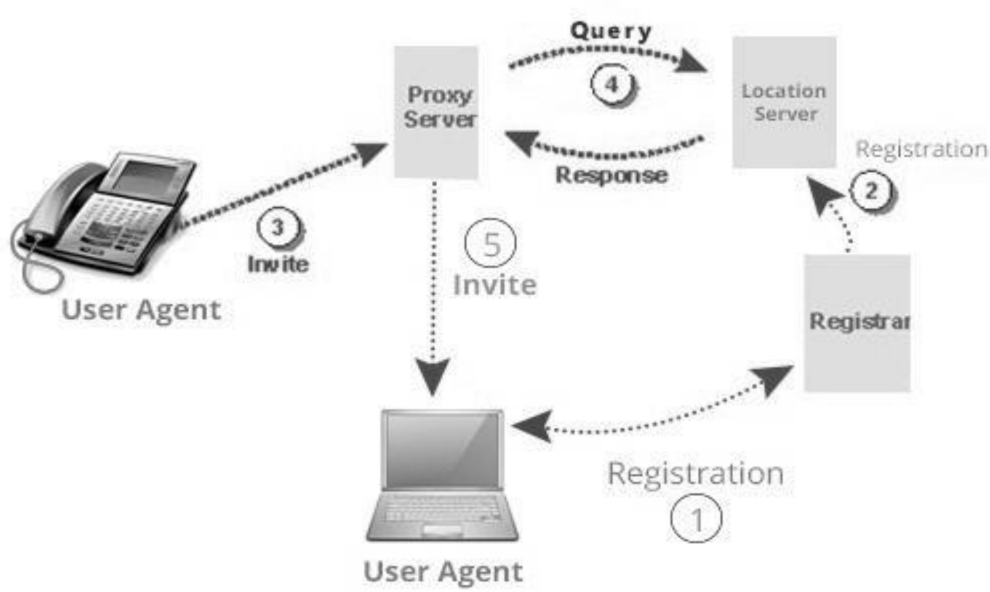


Fig.2.16 Call flow

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

SIP – System Architecture

SIP is structured as a layered protocol, which means its behavior is described in terms of a set of fairly independent processing stages with only a loose coupling between each stage.

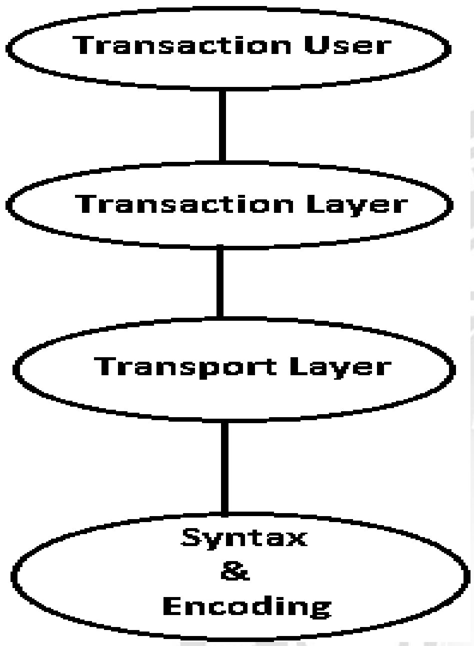


Fig.2.17 SIP – System Architecture

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

- The lowest layer of SIP is its syntax and encoding. Its encoding is specified using an augmented Backus-Naur Form grammar (BNF).
- At the second level is the transport layer. It defines how a Client sends requests and receives responses and how a Server receives requests and sends responses over the network. All SIP elements contain a transport layer.
- Next comes the transaction layer. A transaction is a request sent by a Client transaction (using the transport layer) to a Server transaction, along with all responses to that request sent from the server transaction back to the client. Any task that a user agent client (UAC) accomplishes takes place using a series of transactions. Stateless proxies do not contain a transaction layer.
- The layer above the transaction layer is called the transaction user. Each of the SIP entities, except the Stateless proxies, is a transaction user.

SIP - Basic Call Flow

The following image shows the basic call flow of a SIP session.

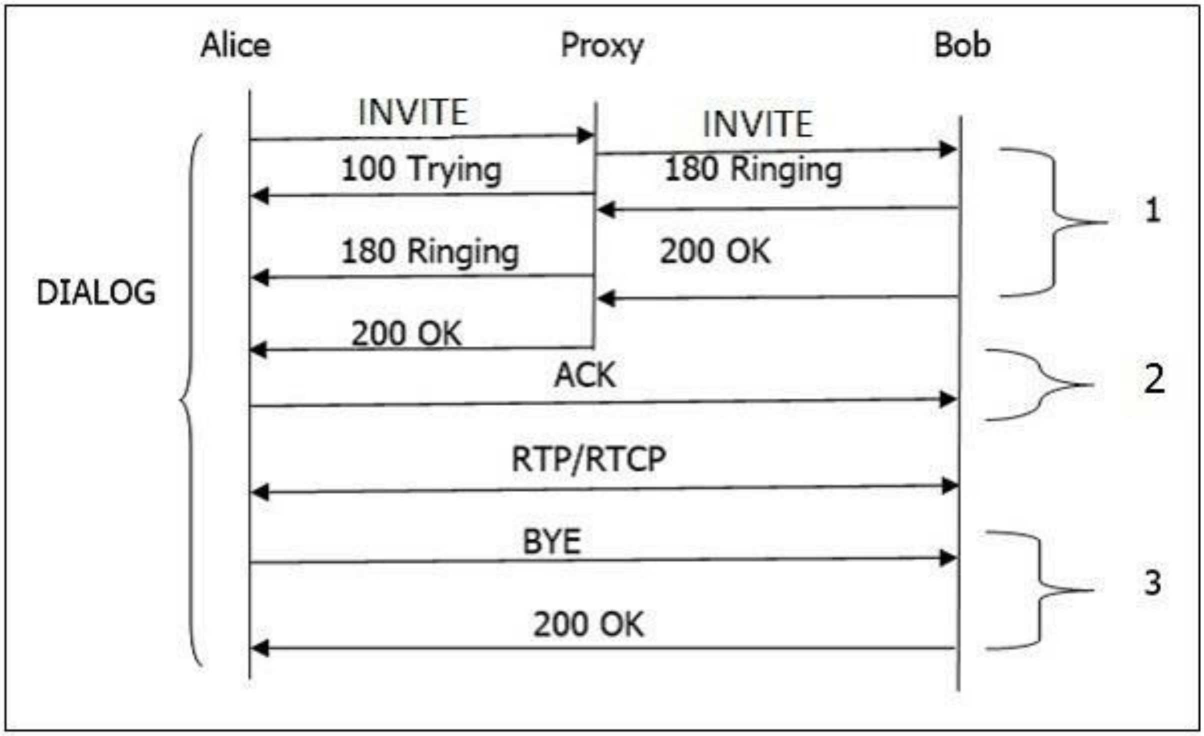


Fig.2.18: SIP session

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

Given below is a step-by-step explanation of the above call flow

- An INVITE request that is sent to a proxy server is responsible for initiating a session.
- The proxy server sends a **100 Trying** response immediately to the caller (Alice) to stop the re-transmissions of the INVITE request.
- The proxy server searches the address of Bob in the location server. After getting the address, it forwards the INVITE request further.
- Thereafter, **180 Ringing** (Provisional responses) generated by Bob is returned back to Alice.
- A **200 OK** response is generated soon after Bob picks the phone up.
- Bob receives an **ACK** from the Alice, once it gets **200 OK**.
- At the same time, the session gets established and RTP packets (conversations) start flowing from both ends.
- After the conversation, any participant (Alice or Bob) can send a **BYE** request to terminate the session.
- **BYE** reaches directly from Alice to Bob bypassing the proxy server.
- Finally, Bob sends a **200 OK** response to confirm the BYE and the session is terminated.

- In the above basic call flow, three **transactions** are (marked as 1, 2, 3) available. The complete call (from INVITE to 200 OK) is known as a **Dialog**.**SIP Trapezoid**

How does a proxy help to connect one user with another? Let us find out with the helpof the following diagram.

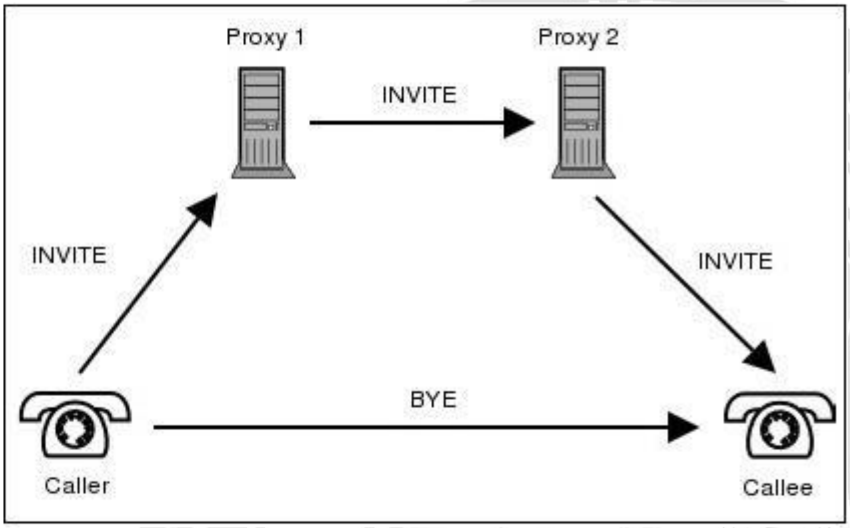


Fig.2.19SIP trapezoid

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

The topology shown in the diagram is known as a SIP trapezoid. The process takes place as follows −

- When a caller initiates a call, an INVITE message is sent to the proxy server. Upon receiving the INVITE, the proxy server attempts to resolve the address of the callee with the help of the DNS server.
- After getting the next route, caller‘s proxy server (Proxy 1, also known as outbound proxy server) forwards the INVITE request to the callee‘s proxy server which acts as an inbound proxy server (Proxy 2) for the callee.
- The inbound proxy server contacts the location server to get information about the callee‘s address where the user registered.
- After getting information from the location server, it forwards the call to its destination.
- Once the user agents get to know their address, they can bypass the call, i.e., conversations pass directly.

SIP - Messaging

SIP messages are of two types − **requests** and **responses**.

- The opening line of a request contains a method that defines the request, and a Request-URI that defines where the request is to be sent.
- Similarly, the opening line of a response contains a response code.

Request Methods

SIP requests are the codes used to establish a communication. To complement them, there are **SIP responses** that generally indicate whether a request succeeded or failed.

These SIP requests which are known as METHODS make SIP message workable.

- METHODS can be regarded as SIP requests, since they request a specific action to be taken by another user agent or server.
- METHODS are distinguished into two types −
    - Core Methods
    - Extension Methods

Core Methods

There are six core methods as discussed below.

INVITE

INVITE is used to initiate a session with a user agent. In other words, an INVITE method is used to establish a media session between the user agents.

- INVITE can contain the media information of the caller in the message body.
- A session is considered established if an INVITE has received a success response(2xx) or an ACK has been sent.

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

- A successful INVITE request establishes a dialog between the two user agents which continues until a BYE is sent to terminate the session.
- An INVITE sent within an established dialog is known as a re-INVITE.
- Re-INVITE is used to change the session characteristics or refresh the state of a dialog.

BYE

BYE is the method used to terminate an established session. This is a SIP request that can be sent by either the caller or the callee to end a session.

It cannot be sent by a proxy server.

BYE request normally routes end to end, bypassing the proxy server. BYE cannot be sent to a pending an INVITE or an un established session.

REGISTER

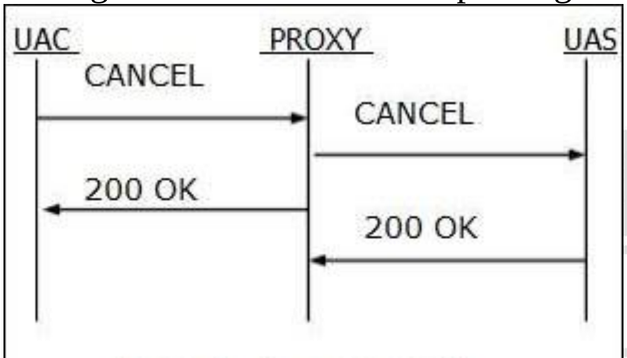
REGISTER request performs the registration of a user agent. This request is sentby a user agent to a registrar server.

- The REGISTER request may be forwarded or proxied until it reaches an authoritative registrar of the specified domain.
- It carries the AOR (Address of Record) in the To header of the user that is being registered.
- REGISTER request contains the time period (3600sec).
- One user agent can send a REGISTER request on behalf of another user agent. This is known as third-party registration. Here, the From tag contains the URI of the party submitting the registration on behalf of the party identified in the To header.

CANCEL

CANCEL is used to terminate a session which is not established. User agents use this request to cancel a pending call attempt initiated earlier.

- It can be sent either by a user agent or a proxy server.
- CANCEL is a hop by hop request, i.e., it goes through the elements between the user agent and receives the response generated by the next stateful element.

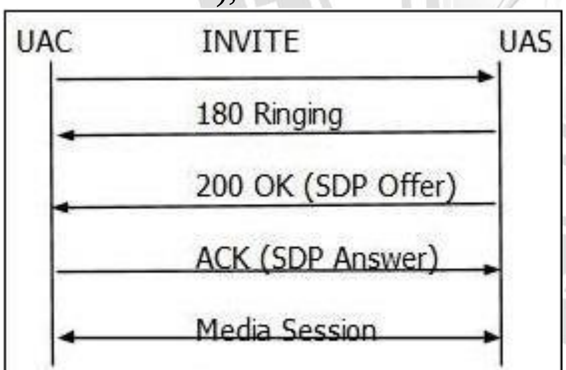


(hop by hop request)

[**Source: Text book-** Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

ACK

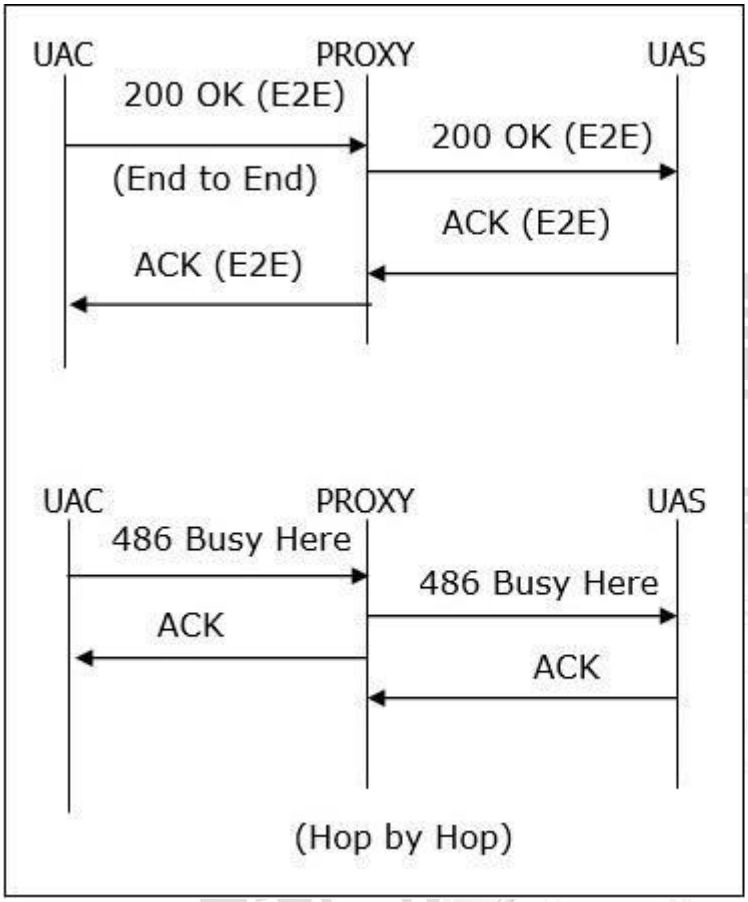
ACK is used to acknowledge the final responses to an INVITE method. An ACK always goes in the direction of INVITE.ACK may contain SDP body (media characteristics), if it is not available in INVITE.



(SDP exchange in ACK)

[**Source: Text book-** Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

- ACK may not be used to modify the media description that has already been sent in the initial INVITE.
- A stateful proxy receiving an ACK must determine whether or not the ACK should be forwarded downstream to another proxy or user agent.
- For 2xx responses, ACK is end to end, but for all other final responses, it works on hop by hop basis when stateful proxies are involved.



[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

SIP - Headers

A header is a component of a SIP message that conveys information about the message. It is structured as a sequence of header fields.

SIP header fields in most cases follow the same rules as HTTP header fields. Header fields are defined as Header: field, where Header is used to represent the header field name, and field is the set of tokens that contains the information. Each field consists of a fieldname followed by a colon (":") and the field-value (i.e., field-name: field-value).

SIP Headers - Compact Form

The following image shows the structure of a typical SIP header.

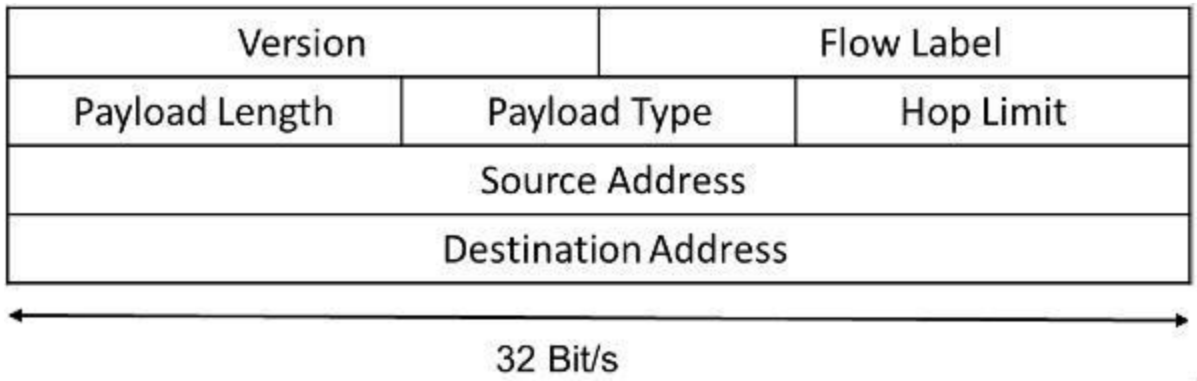


Fig.2.20 SIP Header

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

Headers are categorized as follows depending on their usage in SIP −

SIP - Mobility

Personal mobility is the ability to have a constant identifier across a number of devices. SIP supports basic personal mobility using the REGISTER method, which allows a mobile device to change its IP address and point of connection to the Internet and still be able to receive incoming calls.

SIP can also support service mobility – the ability of a user to keep the same services when mobile

SIP Mobility During Handover(Pre-call)

A device binds its Contact URI with the address of record by a simple sip registration. According to the device IP address, registration authorizes this information automatically update in sip network.

During handover, the User agent routes between different operators, where it has to register again with a Contact as an AOR with another service provider.

For example, let‘s take the example of the following call flow. UA which has temporarily received a new SIP URI with a new service provider. The UA then performs a double registration −

The first registration is with the new service operator, which binds the Contact URI of the device with the new service provider‘s AOR URI.

The second REGISTER request is routed back to the original service provider and provides the new service provider‘s AOR as the Contact URI.

As shown later in the call flow, when a request comes in to the original service provider‘s network, the INVITE is redirected to the new service provider who then routes the call to the user.

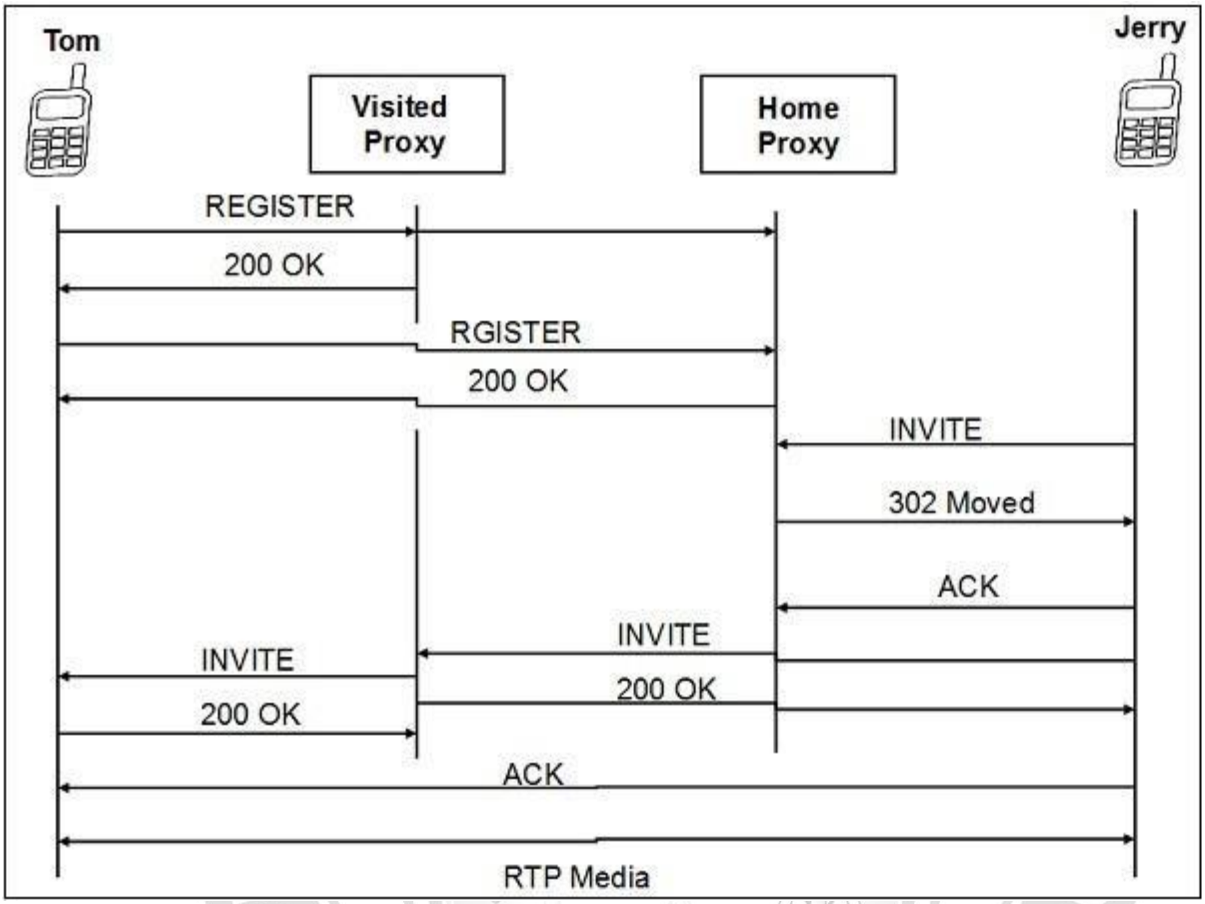


Fig.2.21 SIP Mobility During Handover

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

The first INVITE that is represents in the above figure would be sent to sip:registrar2.in; the second INVITE would be sent to sip: sip:Tom@registrar2.in, which would be forwarded to sip:Tom@172.22.1.102. It reaches Tom and allows the session to be established. Periodically both registrations would need to be refreshed.

Mobility During a Call(re-Invite)

User Agent may change its IP address during the session as it swaps from one network to another. Basic SIP supports this scenario, as a re-INVITE in a dialog can be used to update the Contact URI and change the media information in the SDP.

Take a look at the call flow mentioned in the figure below.

Here, Tom detects a new network,

Uses DHCP to acquire a new IP address, and

Performs a re-INVITE to allow the signaling and media flow to the new IP address.

If the UA can receive media from both networks, the interruption is negligible. If this is not the case, a few media packets may be lost, resulting in a slight interruption to the call.

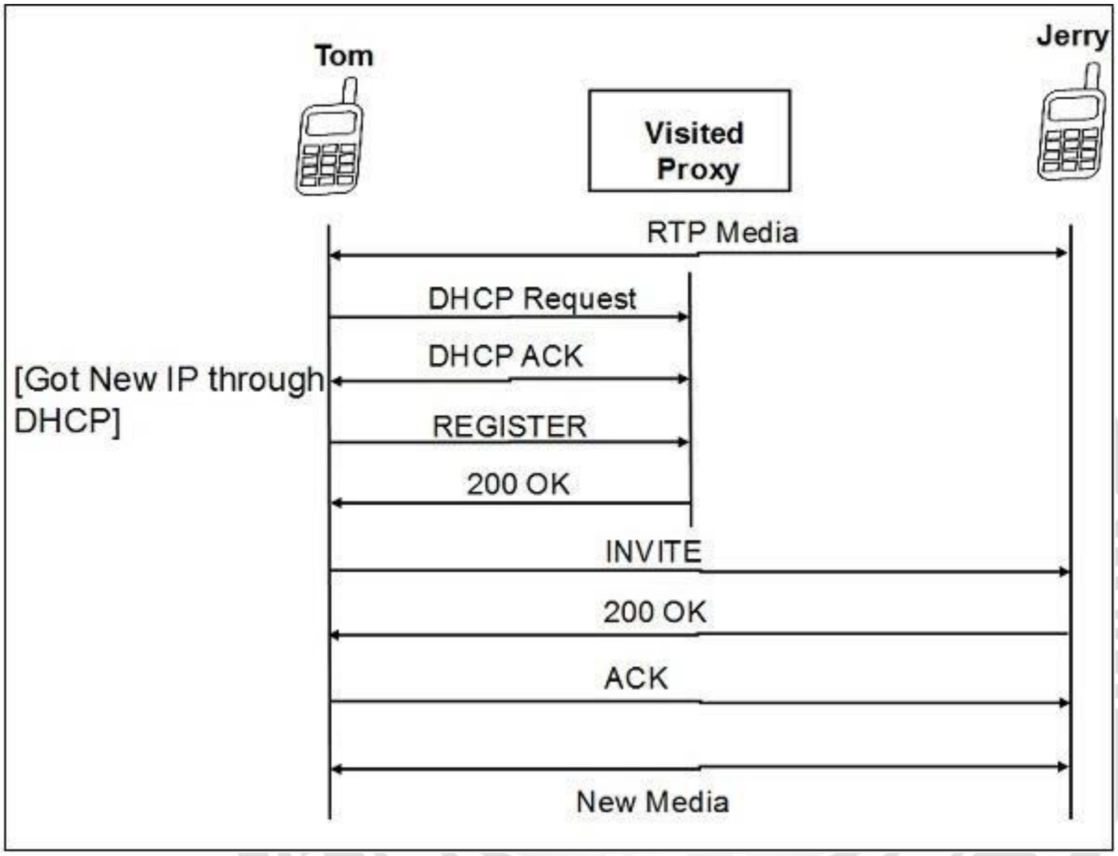


Fig.2.22 SIP Mobility During a Call(re-Invite)

[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

The re-INVITE would appear as follows −

The re-INVITE contains Bowditch‘s new IP address in the Via and Contact header fields and SDP media information.

Mobility in Mid call (With replace Header)

In mid call mobility, the actual route set (set of SIP proxies that the SIP messages must traverse) must change. We cannot use a re-INVITE in mid call mobility

For example, if a proxy is necessary for NAT traversal, then Contact URI must be changed — a new dialog must be created. Hence, it has to send a new INVITE witha Replaces header, which identifies the existing session.

Note − Suppose A & B both are in a call and if A gets another INVITE (let‘s say from C) with a replace header (should match existing dialog), then A must accept the INVITE and terminate the session with B and transfer all resource to newly formed dialog.

The call flow is shown in the following Figure. It is similar to the previous call flow using re-INVITE except that a BYE is automatically generated to terminate the existing dialog when the INVITE with the Replaces is accepted.

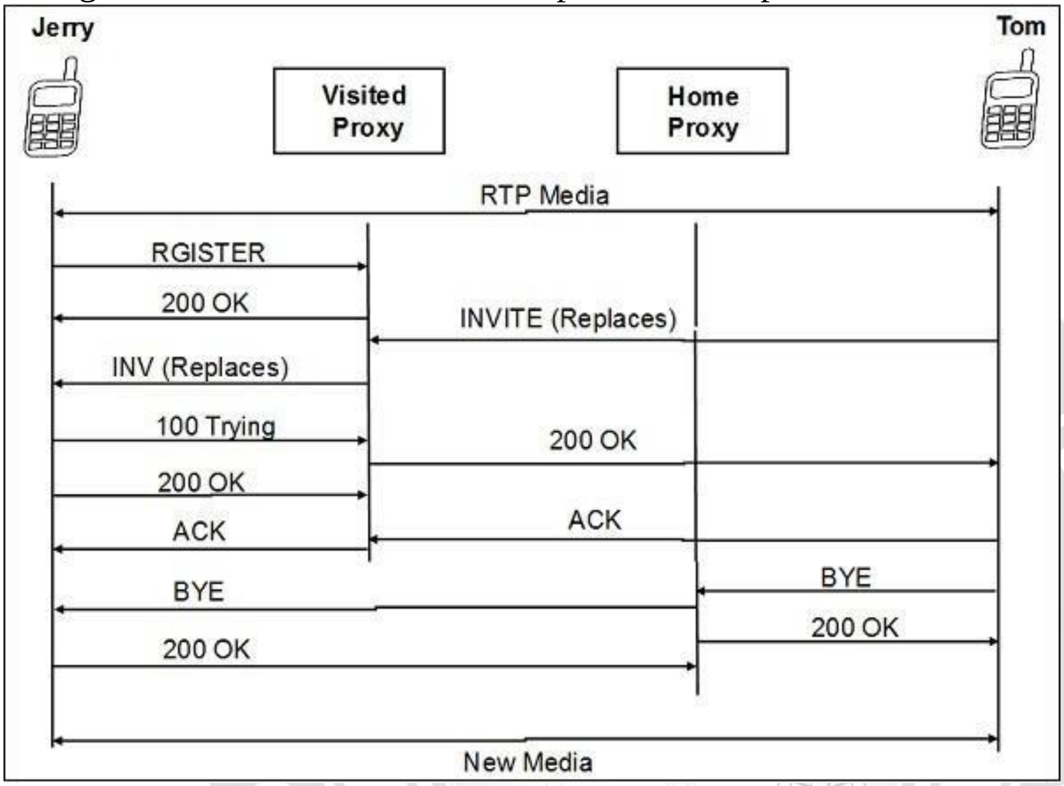


Fig 2.23. SIP Mobility in Midcall (With replace Header)
[Source: Text book- Mobile Communications, Second Edition, Pearson Education by Jochen Schiller]

Given below are the points to note in this scenario

- The existing dialog between Tom and Jerry includes the old visited proxy server.
- The new dialog using the new wireless network requires the inclusion of the new visited proxy server.
- As a result, an INVITE with Replaces is sent by Tom, which creates a new dialog that includes the new visited proxy server but not the old visited proxy server.
- When Jerry accepts the INVITE, a BYE is automatically sent to terminate the old dialog that routes through the old visited proxy server that is now no longer involved in the session.
- The resulting media session is established using Tom‘s new IP address from the SDP in the INVITE.

Service Mobility

Services in SIP can be provided in either proxies or in UAs. Providing service mobility along with personal mobility can be challenging unless the user‘s devices are identically configured with the same services.

SIP can easily support service mobility over the Internet. When connected to Internet, a UA configured to use a set of proxies in India can still use those proxies when roaming in Europe. It does not have any impact on the quality of the media session as the media always flows directly between the two UAs and does not traverse the SIP proxy servers.

Endpoint resident services are available only when the endpoint is connected to the Internet. A terminating service such as a call forwarding service implemented in an endpoint will fail if the endpoint has temporarily lost its Internet connection. Hence some services are implemented in the network using SIP proxy servers.