Introduction

SIP (Session Initiation Protocol) is a protocol developed to assist in providing advanced telephony services across the Internet. Internet telephony is evolving from its use as a "cheap" (but low quality) way to make international phone calls to a serious business telephony capability. SIP is one of a group of protocols required to ensure that this evolution can occur.

SIP is part of the IETF standards process and is modeled upon other Internet protocols such as SMTP (Simple Mail Transfer Protocol) and HTTP (Hypertext Transfer Protocol.) It is used to establish, change and tear down (end) calls between one or more users in an IP-based network. In order to provide telephony services there is a need for a number of different standards and protocols to come together - specifically to ensure transport (RTP), signaling inter-working with today's telephony network, to be able to guarantee voice quality (RSVP, YESSIR), to be able to provide directories (LDAP), to authenticate users (RADIUS, DIAMETER), and to scale to meet the anticipated growth curves. This introduction covers only SIP, but at the end of the paper there is a brief overview on associated standards.

SIP is described as a control protocol for creating, modifying and terminating sessions with one or more participants. These sessions include Internet multimedia conferences, Internet (or any IP Network) telephone calls and multimedia distribution. Members in a session can communicate via multicast or via a mesh of unicast relations, or via a combination of these. SIP supports session descriptions that allow participants to agree on a set of compatible media types. It also supports user mobility by proxying and redirecting requests to the user's current location. SIP is not tied to any particular conference control protocol.

In essence, SIP has to provide or enable the following functions:

Name Translation and User Location - Ensuring that the call reaches the called party wherever they are located. Carrying out any mapping of descriptive information to location information. Ensuring that details of the nature of the call (Session) are supported.

Feature Negotiation - This allows the group involved in a call (this may be a multi-party call) to agree on the features supported – recognizing that not all the parties can support the same level of features. For example video may or may not be supported; as any form of MIME type is supported by SIP, there is plenty of scope for negotiation.

Call Participant Management - During a call a participant can bring other users onto the call or cancel connections to other users. In addition, users could be transferred or placed on hold.

Call feature changes - A user should be able to change the call characteristics during the course of the call. For example, a call may have been set up as 'voice-only', but in the course of the call, the users may need to enable a video function. A third party joining a call may require different features to be enabled in order to participate in the call

Protocol Components
There are two components within SIP. The SIP User Agent and the SIP Network Server. The User Agent is effectively the end system component for the call and the SIP Server is the network device that handles the signaling associated with multiple calls.

The User agent itself has a client element, the User Agent Client (UAC) and a server element, the User Agent Server (UAS.) The client element initiates the calls and the server element answers the calls. This allows peer-to-peer calls to be made using a client-server protocol.

The SIP Server element also provides for more than one type of server. There are effectively three forms of server that can exist in the network - the SIP stateful proxy server, the SIP stateless proxy server and the SIP re-direct server. The main function of the SIP servers is to provide name resolution and user location, since the caller is unlikely to know the IP address or host name of the called party. What will be available is perhaps an email-like address or a telephone number associated with the called party. Using this information, the caller's user agent can identify with a specific server to "resolve" the address information – it is likely that this will involve many servers in the network.

A SIP proxy server receives requests, determines where to send these, and passes them onto the next server (using next hop routing principals). There can be many server hops in the network.

The difference between a stateful and stateless proxy server is that a stateful proxy server remembers the incoming requests it receives, along with the responses it sends back and the outgoing requests it sends on. A stateless proxy server forgets all information once it has sent on a request. This allows a stateful proxy server to fork requests to try multiple possible user locations in parallel and only send the best responses back. Stateless Proxy servers are most likely to be the fast, backbone of the SIP infrastructure. Stateful proxy servers are then most likely to be the local devices close to the User Agents, controlling domains of users and becoming the prime platform for the application services.

A re-direct server receives requests, but rather than passing these onto the next server it sends a response to the caller indicating the address for the called user. This provides the address for the caller to contact the called party at the next server directly.

Diagram 1 depicts a simple call set-up process.

PICTURE 1

SIP Protocol

SIP addresses users by an email-like address. Each user is identified through a hierarchical URL that is built around elements such as a user's phone number or host name (for example, SIP:user@company.com). Because of this similarity, SIP URLs are easy to associate with a user's e-mail address.

SIP provides its own reliability mechanism and is therefore independent of the packet layer and only requires an unreliable datagram service. SIP is typically used over UDP or TCP.

SIP provides the necessary protocol mechanisms so that end systems and proxy servers can provide services:

- User location
- User capabilities
- User availability
- Call set-up
- Call handling
- Call forwarding, including
    - The equivalent of 700-, 800- and 900- type calls
    - Call-forwarding no answer
    - Call-forwarding busy
    - Call-forwarding unconditional
    - Other address-translation services
- Callee and calling "number" delivery, where numbers can be any (preferably unique) naming scheme
- Personal mobility, i.e., the ability to reach a called party under a single, location-independent address even when the user changes terminals
- Terminal-type negotiation and selection: a caller can be given a choice how to reach the party, e.g., via Internet telephony, mobile phone, an answering service, etc.

- Terminal capability negotiation
- Caller and callee authentication
- Blind and supervised call transfer
- Invitations to multicast conferences

When a user wants to call another user, the caller initiates the call with an invite request. The request contains enough information for the called party to join the session. If the client knows the location of the other party it can send the request directly to their IP address. If not the client can send it to a locally configured SIP network server. If that server is a proxy server it will attempt to resolve the called user's location and send the request to them. There are many ways it can do this, such as searching the DNS or accessing databases. Alternatively, the server may be a redirect server that may return the called user location to the calling client for it to try directly. During the course of locating a user, one SIP network server can, of course, proxy or redirect the call to additional servers until it arrives at one that definitely knows the IP address where the called user can be found.

Once found, the request is sent to the user, and from there several options arise. In the simplest case, the user's telephony client receives the request—that is, the user's phone rings. If the user takes the call, the client responds to the invitation with the designated capabilities* of the client software and a connection is established. If the user declines the call, the session can be redirected to a voice mail server or to another user.

"Designated capabilities" refers to the functions that the user wants to invoke. The client software might support videoconferencing, for example, but the user may only want to use audio conferencing. Regardless, the user can always add functions—such as videoconferencing, white-boarding, or a third user—by issuing another invite request to other users on the link.

SIP has two additional significant features. The first is a stateful SIP proxy server's ability to split or "fork" an incoming call so that several extensions can be rung at once. The first extension to answer takes the call. This feature is handy if a user is working between two locations (a lab and an office, for example), or where someone is ringing both a boss and their secretary.

The second significant feature is SIP's unique ability to return different media types. Take the example of a user contacting a company. When the SIP server receives the client's connection request, it can return to the customer's phone client via a Web Interactive Voice Response page (IVR or could use the term Interactive Web Response or IWR), with the extensions of the available departments or users provided on the list. Clicking the appropriate link sends an invitation to that user to set up a call.

SIP Messages

A SIP request message consists of three elements:

Request Line
Header
Message Body

A SIP response message consists of three elements:

Status Line
Header
Message Body

The Request line and header field define the nature of the call in terms of services, addresses and protocol features. The message body is independent of the SIP protocol and can contain anything.

SIP defines the following methods (SIP uses the term 'method' to describe the specification areas):

Invite – invites a user to join a call.
Bye – terminates the call between two of the users on a call
Options – requests information on the capabilities of a server
Ack – confirms that a client has received a final response to an INVITE
Register – provides the map for address resolution, letting a server know the location of other users.
Cancel – ends a pending request, but does not end the call

The INFO method, for mid-session signalling, is also being added

Related Standards Activity.
New standards efforts are aimed toward using the IP network for non-voice interactions, while voice and fax are carried entirely on the PSTN. These efforts are going forward on a broad front; here are a few of the more exciting standards developments:

•**PSTN to Internet Interworking (PINT)**: The PINT standards effort defines "click-to-dial" services – interworking between a Web page and a PSTN gateway element, or PINT server. PINT is helping to standardize service delivery, including request to call, request to fax, request to hear content, and in the future, request to conference.

•**Services in the PSTN/IN Requesting Internet Services (SPIRITS)**: This working group is providing a framework for standardizing the mechanisms for controlling critical call altering and call completion processes from the Internet by exposing the Internet domain to the PSTN's call model. The goal is to create a framework for Internet call waiting-type services that will be applicable to wireline, wireless and broadband (cable and xDSL) environments. Proposed SPIRITS-enabled network events include voice mail arrival, incoming call notification, attempt to dial numbers, dropping dialed connection, completing Internet service provider (ISP) connections, attempt to forward calls and attempt to subscribe/unsubscribe to a PSTN service.

•**Signaling Translation (SIGTRAN)**: the IETF standard which deals with Signaling System 7 (SS7; specifically Transaction Capabilities Application Part or TCAP-over-IP);

•**Media Gateway Control Protocol (MEGACO)**: the IETF standard that media gateway controllers such as softswitches, service nodes and feature processors use to control packet/circuit gateways;

•**Telephone Number Mapping (ENUM)**: the IETF standard that defines address-mapping among phone numbers and Internet devices.

•**WAP Telephony Application (WTA)**: is providing good ideas for IN-style services creation. Some of these efforts are in their formative stages, but all are aimed at providing the interoperability of PSTN/IN and IP services that is the heart and soul of true convergence.

•**Telephony Routing over IP (TRIP)**: TRIP is a policy driven inter-administrative domain protocol for advertising the reachability of telephony destinations between location servers, and for advertising attributes of the routes to those destinations. TRIP's operation is independent of any signaling protocol, hence TRIP can serve as the telephony routing protocol for any signaling protocol.

•**SIP-TSI**: SIP-TSI is proposed as a method to enable voice and facsimile telephony applications running on an application server to communicate with a softswitch through a data network, according to a fully-specified Telephony Service Interface (TSI). Based upon IETF RFC 2543 on "Session Initiation Protocol"1 (SIP), this SIP-Telephony Service Interface (SIP-TSI) is capable of supporting a level of telephony application functionality commensurate with Time Division Multiplex device interfaces used in legacy PSTN voice and facsimile telephony applications.

SOFTWARE INTERFACES

A potentially rewarding area in standards work involves network applications programmer interfaces (APIs) between traditional network providers and applications service providers (ASPs)/enterprise networks. APIs have the potential to bridge the gap between traditional telecom services and complex telecom protocols. The telecom industry is considering numerous open services architecture APIs, including the following two:

•**The Parlay Group's open API**: The Parlay API supports the operation of a service API with any third parties (such as ASPs) that are known or unknown to the end user. By defining a policy-based infrastructure for 'untrusted' third parties, the Parlay API can securely automate the process of bridging services, such as inbound call center routing, messaging, notification and mobility/location determination from third parties into a network.

•**Java APIs for Interoperable Networks (JAIN)**: JAIN integrates JAVA-based middleware around SS7/IN protocol stacks. The value of these APIs to carriers is that services can be portable among vendors' platforms, putting them in the same programming environment with an exponentially growing number of Web services developers. JAIN has strong support among carriers, equipment manufacturers and applications providers.

The JAIN and Parlay Group organizations have recently come together and are proposing support for the Parlay APIs on top of the JAIN stacks. Since Parlay is an API, applications developers are not overly concerned about the underlying implementation technology.