# Comparison of H.323 and SIP for IP Telephony Signaling

Ismail Dalgic[a], Hanlin Fang[b]

[a]3Com Corporation
Technology Development Center
5400 Bayfront Plaza, M/S 3219, Santa Clara, CA 95052

[b]Cisco Systems
170 West Tasman Dr., San Jose, CA 95134

**ABSTRACT**

Two standards currently compete for the dominance of IP telephony signaling: the H.323 protocol suite by ITU-T, and the Session Initiation Protocol (SIP) by IETF. Both of these signaling protocols provide mechanisms for call establishment and teardown, call control and supplementary services, and capability exchange. We investigate and compare these two protocols in terms of Functionality, Quality of Service (QoS), Scalability, Flexibility, Interoperability, and Ease of Implementation. For fairness of comparison, we consider similar scenarios for both protocols. In particular, we focus on scenarios that involve a gatekeeper for H.323, and a Proxy/Redirect server for SIP. The reason is that medium-to-large IP Telephony systems are not manageable without a gatekeeper or proxy server. We consider all three versions of H.323. In terms of functionality and services that can be supported, H.323 version 2 and SIP are very similar. However, supplementary services in H.323 are more rigorously defined, and therefore fewer interoperability issues are expected among its implementations. Furthermore, H.323 has taken more steps to ensure compatibility among its different versions, and to interoperate with PSTN. The two protocols are comparable in their QoS support (similar call setup delays, no support for resource reservation or class of service (CoS) setting), but H.323 version 3 will allow signaling of the requested CoS. SIP's primary advantages are (i) flexibility to add new features, and (ii) relative ease of implementation and debugging. Finally, we note that H.323 and SIP are improving themselves by learning from each other, and the differences between them are diminishing with each new version.

**Keywords:** IP Telephony, Voice over IP, H.323, SIP, signaling, call control

## 1. INTRODUCTION

Telephony service today is provided for the most part over circuit-switched networks, which are referred to as Public Switched Telephone Networks (PSTN). This service is known as Plain Old Telephone Service (POTS). A new trend that is beginning to emerge in recent years is to provide telephony service over IP networks, known as *IP telephony*, or *Voice over IP*.

An important driving force behind IP Telephony is cost savings, especially for corporations with large data networks. The high cost of long-distance and international voice calls – thanks to layers of local and international carriers – is the crux of the issue. A significant portion of this cost originates from regulatory taxes imposed on long-distance voice calls. Such surcharges are not applicable to long-distance circuits carrying data traffic; thus, for a given bandwidth, making a data call is much less expensive than making a voice call.

In addition to the cost savings for long-distance voice calls, carrying voice traffic on the data network within a business building or campus also can achieve substantial cost savings, since the operation of today's proprietary PBX setups is relatively cost-inefficient.

There are other very significant motivating factors for carrying voice traffic over data networks as well. A very important benefit of IP Telephony is the integration of voice and data applications, which can result in more effective business processes. Examples of such applications are integrated voice mail and e-mail, teleconferencing, computer-supported collaborative work and automated and intelligent call distribution. Another benefit is the enabling of many new services both for businesses and for customers. The flexibility offered by IP Telephony by moving the intelligence from the network to the end stations, as well as the open nature of IP networks, are the factors that enable new services.

Furthermore, many of the existing services that require a fee today, such as caller-id, call-forwarding, and multi-line presence become trivial to implement; therefore, such services are likely to be offered for free for competitive reasons.

In order for IP Telephony to gain mainstream acceptance and ultimately replace traditional Plain Old Telephone Service (POTS), two conditions have to be met. First, the *quality* of the voice communication must be at least at the same level as POTS. The two primary aspects of voice quality are the end-to-end delay, and the voice clarity (which depends on many factors, including the voice digitization and compression scheme used, and the amount of lost or late-arrived packets). Therefore, the IP network must be designed such that it can meet the delay and packet loss requirements of the telephony application.

The second condition for the acceptance of IP Telephony is the ease of operation and functionality offered to the end-user at least at the same level as in PSTN. This requires the IP Telephony architecture to provide a *signaling infrastructure* that offers at least the same capabilities and features as the Signaling System 7 (SS7) architecture in PSTN[1]. More specifically, the signaling infrastructure must:

- provide the *functionality* required to set up, manage, and tear down calls and connections;
- be *scalable* to support a very large number of registered endpoints (in the order of billions worldwide), and a very large number of simultaneous calls (in the order of millions worldwide);
- support *network management* features for policy control, accounting, billing, etc;
- provide a mechanism to communicate and set up the *Quality of Service* requested by the end points;
- be *extensible* to help with adding new features easily;
- support *interoperability* among different vendors' implementations, among different versions of the signaling protocol, and with different signaling protocols.

Two standards compete for IP Telephony signaling. The older and currently more widely accepted standard is the ITU-T recommendation H.323[2], which defines a multimedia communications system over packet-switched networks, including IP networks. The other standard, Session Initiation Protocol (SIP)[3], comes from the IETF MMUSIC working group. In this paper, we compare these two standards for IP Telephony signaling from the above listed points of view. Other papers exist which compare H.323 and SIP; among them the most notably is the one by Schulzrinne and Rosenberg[4]. Our paper differs from the previous comparisons in the following ways: (i) we consider the new features introduced in H.323 version 3; (ii) we present a comprehensive evaluation, considering many different aspects; and (iii) we focus on the features and characteristics that are relevant to IP telephony.

The remainder of the paper is organized as follows. In Section 2, we provide a brief overview of the two signaling protocols. In Section 3, we describe our comparison in terms of the areas listed above. We present our concluding remarks in Section 4.

## 2. OVERVIEW OF H.323 AND SIP

### 2.1. H.323 Overview

| Name | The description of protocols |
|---|---|
| H.323 | Specification of the system |
| H.225.0 | Call control (RAS), call setup (Q.931-like protocol), and packetization and synchronization of media stream |
| H.235 | Security protocol for authentication, integrity, privacy, etc. |
| H.245 | Capability exchange communication and mode switching |
| H.450 | Supplementary services including call holding, transfer, forwarding, etc |
| H.246 | Interoperability with circuit-switched services |
| H.332 | For large size conferencing |
| H.26x | Video codecs including H.261 and H.263 |
| G.7xx | Audio codecs including G.711, G.723, G.729, G.728, etc |

Table 1: ITU-T recommendations that are part of the H.323 specification.

H.323 is an ITU-T Study Group 16 recommendation that specifies a system and protocols for multimedia communications over packet-switched networks. In particular, H.323 consists of a set of protocols that are responsible for encoding, decoding, and packetizing audio and video signals, for call signaling and control, as well as for capability exchange. Most of the existing H.323 implementations are based on the second version of the standard, which was decided in February 1998. The third version is expected to be decided in September 1999. In this paper, we discuss both versions of the standard.

H.323 is an umbrella specification, and various aspects of H.323 are specified in different ITU-T recommendations. Table 1 shows the recommendations that are part of the H.323 specification.

### 2.1.1. H.323 Endpoint Types

H.323 defines four major components for a network-based communication system: Terminals, Gateways, Gatekeepers and Multipoint Control Units (MCUs). (See Figure 1.)

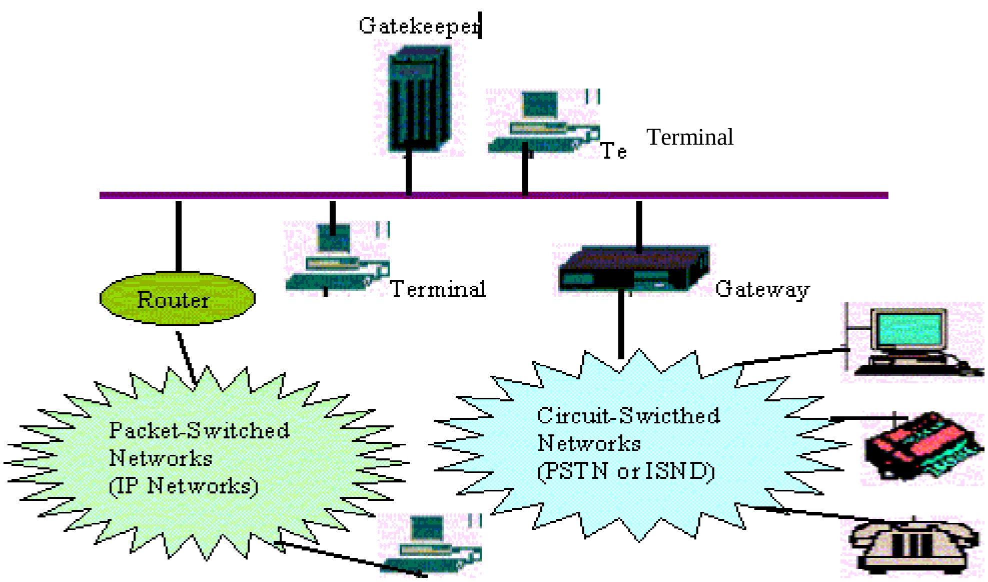


Figure 1: H.323 endpoint types.

Terminals are client endpoints on IP-based networks that provide real-time, two-way communications with other H.323 entities. H.323 terminals are required to support the following three functional parts:

**Signaling and Control:** H.323 must support H.245, a complex standard for channel usage and capabilities, in addition to a Q.931-like protocol defined in H.225 for call signaling and establishment, as well as Registration/Administration/Status (RAS) protocol defined in H.225 for communication with gatekeepers. All of these protocols use ASN.1 encoding for their messages.
**Real-time communication:** H.323 terminals must support RTP/RTCP, a protocol for sequencing audio and video packets.
**Codecs:** Codecs are pieces of software that compress audio/video before transmission and decompress them back after receiving compressed packets. For interoperability purposes, every H.323 terminal is required to support the G.711 audio codec. Other audio and video codecs are optional.

Gateways provide the connection path between the packet-switched network and the Switched Circuit Network (SCN, which can be either public or private). The gateway is not required when there is no connection to other networks. In general, a gateway deflects the characteristics of a LAN endpoint to a SCN endpoint, and vice-versa. Gateways perform call setup and control on both the packet-switched network and on the SCN, and they translate between transmission formats and between communication procedures. Some gateways can also translate between different codec standards for audio and/or video (referred to as *transcoding*), with the purpose of reducing the bandwidth of the audio/video flow if the SCN bandwidth is limited.

Gatekeepers are optional on an H.323 system, but they have certain mandatory functions if they are present. Gatekeepers perform four required functions: Address Translation (from alias addresses or phone numbers to transport addresses[*]), Admission Control, Bandwidth Control and Zone Management. Gatekeepers can also support four optional functions: Call Control Signaling, Call Authorization, Bandwidth Management and Call Management. When a gatekeeper is present on an H.323 system, all other types of endpoints are required to register with the gatekeeper and receive its permission prior to making a call.

Multipoint Control Units (MCU) support conferencing between three or more endpoints. The MCU typically consists of a Multipoint Controller (MC) and zero or more Multipoint Processors (MP). MC provides the control functions such as negotiation between terminals and determination of common capabilities for processing audio and video. MP performs the necessary processing on the media streams for a conference. Such processing typically involves audio mixing and audio/video switching.

### 2.1.2. Channels Defined in H.323

H.323 uses the concept of channels to structure the information exchange between communication entities. A channel is a transport-layer connection, which can be either unidirectional or bi-directional. In particular, H.323 defines the following types of channels:

**RAS Channel:** This channel provides a mechanism for communication between an endpoint and its gatekeeper. The RAS (Registration, Admission, and Status) protocol is specified in H.225.0. Through the RAS channel, an endpoint registers with the gatekeeper, and requests permission to place a call to another endpoint. If permission is granted, the gatekeeper returns the transport address for the call signaling channel of the called endpoint.
**Call Signaling Channel:** This channel carries information for call control and supplementary service control. The Q.931-like protocol used over this channel is specified in H.225.0 and H.450.x. When the call is established, the transport address for H.245 Control Channel is indicated on this channel.
**H.245 Control Channel:** This channel carries the H.245 protocol messages for media control with capability exchange support. After the call participants exchange their capabilities, logical channels for media are opened through the H.245 control channel.
**Logical Channel for Media:** These channels carry the audio, video, and other media information. Each media type is carried in a separate pair of uni-directional channels, one for each direction, using RTP and RTCP.

H.323 specifies that the RAS channel and the logical channels for media are carried over an unreliable transport protocol, such as UDP. The H.245 control channel is specified to be carried over a reliable transport protocol, such as TCP. H.323 versions 1 and 2 specify that the call signaling channel is carried over a reliable transport protocol. In version 3, this channel can optionally be carried over an unreliable transport protocol.

## 2.2. SIP Overview

IETF has also specified a multimedia communications protocol suite. In the IETF architecture, the media flows are carried using RTP, just like in H.323. Therefore, the main difference between H.323 and IETF specifications is how the call signaling and control is achieved.

The primary protocol that handles call signaling and control in the IETF specification is SIP. SIP is an application layer control protocol that can establish, modify and terminate multimedia sessions or calls. There are two major architectural elements to SIP: the user agent (UA), and the network server. The UA resides at the SIP end stations, and contains two components: a user agent client (UAC) which is responsible for issuing SIP requests, and a user agent server (UAS), which responds to such requests. There are three different network server types, a redirect server, a proxy server, and a registrar. A basic SIP call does not need servers, but some of the more powerful features depend upon them. To the

[*] When the packet-switched network is an IP network, a transport address is defined as the combination of an IP address and a TCP or UDP port number.

first degree of approximation, the SIP User Agent is equivalent to a H.323 terminal (or the packet-network side of a gateway), and the SIP network servers are equivalent to a H.323 gatekeeper.

The most generic SIP operation involves a SIP UAC issuing a request, a SIP proxy server acting as end-user location discovery agent and a SIP UAS accepting the call. A successful SIP invitation consists of two requests: INVITE followed by ACK. The INVITE message contains session description that informs the called party what type of media the caller can accept and where it wishes the media data to be sent. SIP addresses are referred to as SIP Uniform Resource Locators (SIP-URLs), which are of the form sip:user@host.domain. SIP message format is based on the Hyper Text Transport Protocol (HTTP) message format, which uses a human-readable, text-based encoding.

Redirect servers process an INVITE message by sending back the SIP-URL where the callee is reachable. Proxy servers perform application layer routing of the SIP requests and responses. A proxy server can either be stateful or stateless. A stateful proxy holds information about the call during the entire time the call is up, while a stateless proxy processes a message and then forgets everything about the call until the next message arrives. Furthermore, proxies can either be forking or non-forking. A forking proxy can, for example, ring several phones at once until somebody takes the call. Registrar servers are used to record the SIP address (called a SIP URL) and the associated IP address. The most common use of a registrar server is to register after start-up, so that when an INVITE request arrives for the SIP URL used in the REGISTER message, the proxy or redirect server forwards the request correctly. Note that usually a SIP network server implements a combination of different types of servers.

SIP is used to establish, modify, and terminate multimedia sessions. However, it only handles the communication between the caller and the callee, the endpoint addressing, and user location. There needs to be a description about a multimedia session within a SIP request and response message, as well as an announcement for a session. IETF Session Description Protocol (SDP)[5] is used together with SIP to accomplish all the call signaling functions in IP telephony. Roughly speaking, SIP is the equivalent of RAS and the Q.931-like protocol in H.323. SDP is the equivalent of H.245.

## 3. COMPARISON OF H.323 AND SIP FOR IP TELEPHONY SIGNALING

We compare H.323 and SIP in terms of Functionality, Quality of Service (QoS), Scalability, Flexibility, Interoperability, Security, and Ease of Implementation. For fairness of comparison, we consider similar scenarios for both protocols. In particular, we focus on scenarios that involve a gatekeeper for H.323, and a Proxy/Registrar server for SIP. The reason is that medium-to-large IP Telephony systems are not manageable without a gatekeeper or a proxy server.

### 3.1. Functionality

In addition to the basic telephone call service, both SIP and H.323 support some call control services, advanced features, and capability exchange. Roughly, the services they provide are similar but with different approaches. We will discuss the detailed signaling procedure for some services, then summarize their characteristics.

#### 3.1.1. Basic Call Setup and Tear Down

H.323 v2 call setup is based on reliable transport protocol. Therefore, the call setup needs a two-phase connection: TCP connection and call connection. The coming H.323 v3 supports both TCP and UDP, which simplifies the call setup procedure. (See Figures 2, and 3 for call setup in H.323 v2 and v3, respectively.) SIP call setup procedure is similar to H.323 v3. (See Figure 4 for Call setup in SIP.)

The tear down procedure is a reverse of the call setup. Either caller or callee can terminate a call by RELEASE COMPLETE (in H.323) or BYE (in SIP) message.

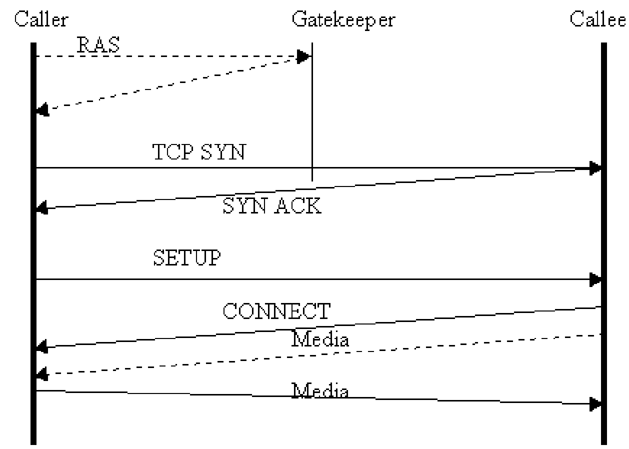

Figure 2: Call Setup in H.323 v2

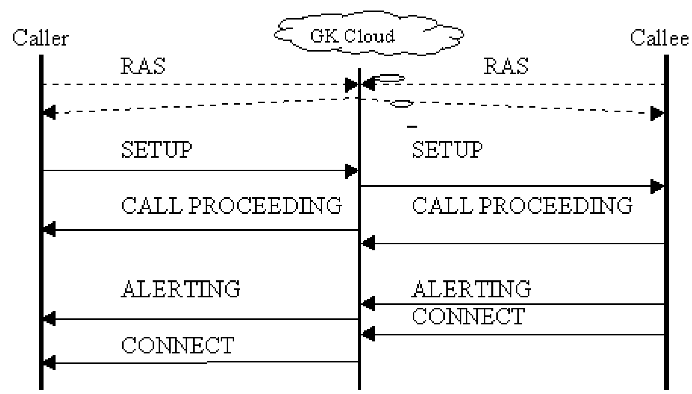

Figure 3: Call Setup, H.323 v3 using UDP (both Endpoints registered, Gatekeper routed call setup)

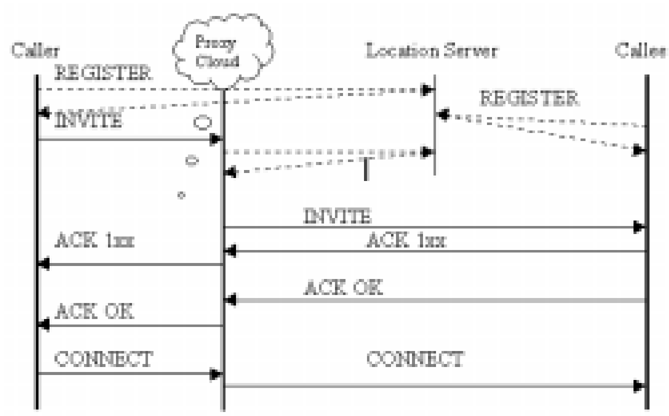

Figure 4: Call Setup with SIP (both endpoints registered, proxy routed call setup)

### 3.1.2. Call control services

SIP and H.323 both support call hold, call transfer, call forwarding, call waiting, conferencing, and some other supplementary services. In the following, we examine some example supplementary services, namely, call hold, call transfer, call forwarding, and call waiting.

**Call Hold**

Call Hold is defined as one call party disconnecting the voice communication without terminating the call, with the ability to reestablish the voice communication at a later time. When the call is on hold, optionally some music can be played, so that the party on hold knows that the call is still active.

H.323 defines two scenarios in call hold service: Near-end Call Holding and Remote-end Call Holding. Both can work with or without a gatekeeper. Gatekeepers only pass SS-HOLD (Supplementary Service-HOLD) operation transparently. (Thus, we illustrate the signaling message flow without gatekeepers for simplicity.)
*Near-end Call Hold:* Hold is invoked at the holding endpoint as a local procedure. (See Figure 5)
*Remote-end Call Hold:* The holding endpoint sends a hold request to the remote endpoint requiring the held endpoint to provide Music on Hold (MOH) to the held user. (See Figure 6)

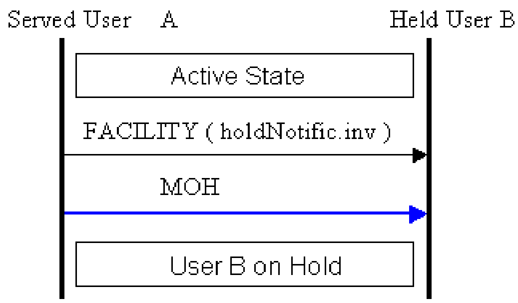


Figure 5: Signaling flow for Near-end call hold without a gatekeeper in H.323

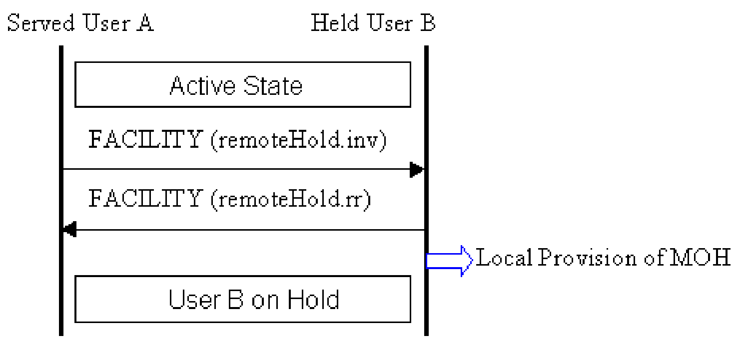


Figure 6: Signaling Flow for Remote-End Call Hold without a gatekeeper in H.323

SIP uses a simpler approach to achieve the same call hold functionality as H.323. For a Near-end Call Hold, no protocol assistance is needed. The client just continually receives media stream from a server but does not generate any response. (See Figure 7.) To achieve Remote-end Call Hold, the holding side needs to send an INVITE message to other side, indicating a NULL set of receiving capability for any kind of media (See Figure 8.) MOH can be implemented by asking an RTSP server to play to the IP address or phone number provided in the RTSP SETUP request.

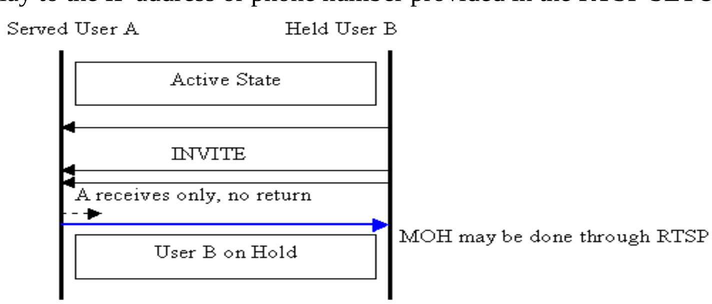


Figure 7: Near-end Call Hold in SIP

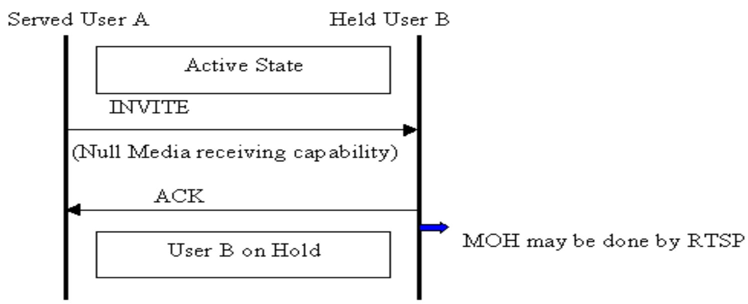


Figure 8: Remote-end Call Hold in SIP

**Call Transfer**

Call Transfer enables a user to transfer an established call to a third party. Both H.323 and SIP support three types of Call Transfer: Blind Transfer, Alternative Transfer, and Operator-Assisted Transfer. The signaling flow diagrams of Blind Transfer and Operator-Assisted Transfer for both H.323 v3 and SIP are provided in Figures 9, 10, and 11.

Blind Transfer works as follows:
Originator  A connects with B
A asks B to connect with C
A simply disconnects with B without any acknowledgement of connection between B and C

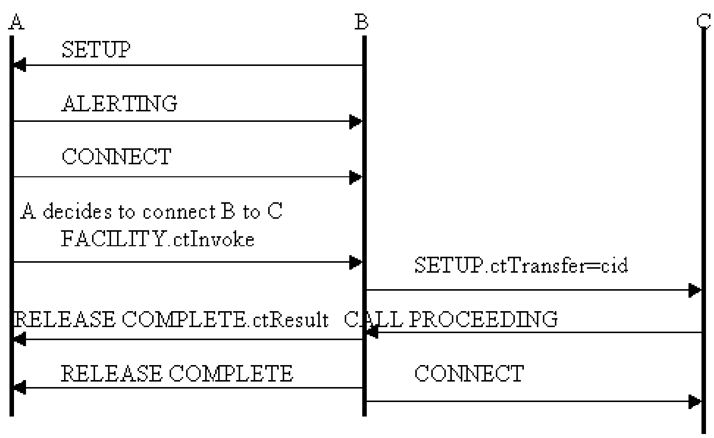


Figure 9: Signaling Flow for Blind Call Transfer in H.323

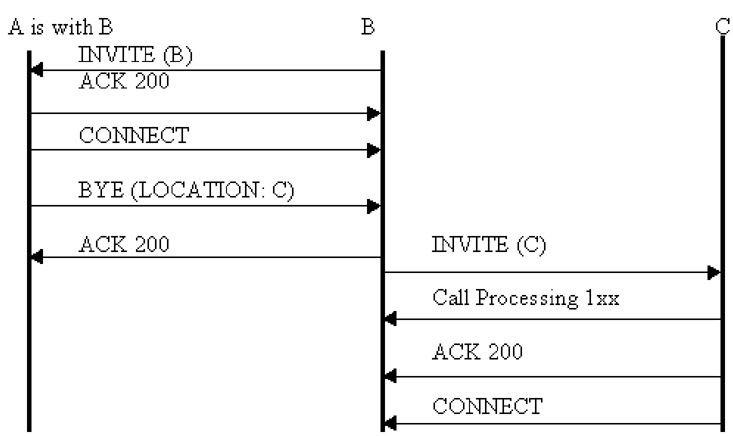


Figure 10: Signaling Flow for Blind Call Transfer in SIP

Operator-Assisted Transfer works as follows:
Originator B sets up a connection with the operator A
A puts B on HOLD, then sets up another connection with C
B and C set up the connection between them
A releases the connection with B
A releases the connection with C

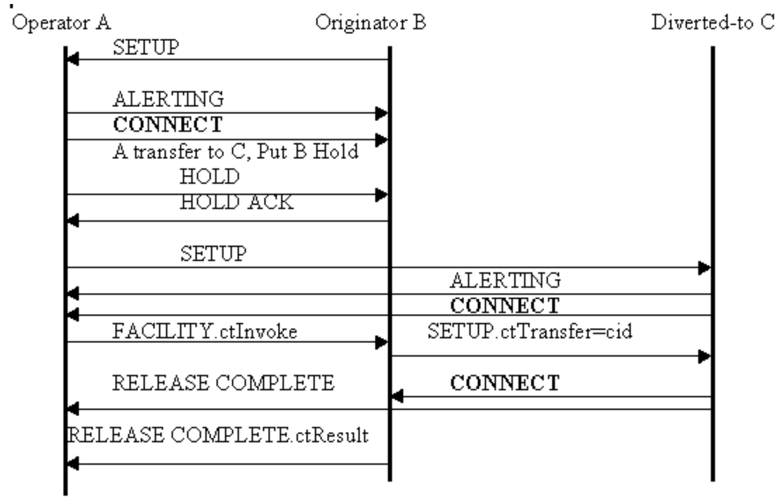


Figure 11: Operator-Assisted Call Transfer in H.323

The procedure of Operator-Assisted Call Transfer in SIP is very similar to that in H.323, except that the equivalent SIP messages are sent out.

**Call Forwarding**

Call Forwarding permits the called party to forward particular pre-selected calls to other addresses. H.323 defines the following operation models for call forwarding: Call Forwarding immediate/delayed with rerouting, Call Forwarding partial rerouting in gatekeeper, CFU/CFB invoked by the gatekeeper, and CFNR invoked by the gatekeeper.

Call Forwarding services provided by SIP are usually instantiated with the LOCATION header fields, which contain the forwarding destination.  SIP supports Call Forwarding Busy, Call Forwarding no Response, and Selective Call Forwarding.

A more general model, Call forwarding partial rerouting in gatekeeper/proxy is chosen as an example. (See Figures 12 and 13) The messages used for call forwarding are different in H.323 and in SIP. However, the call flows are very similar.

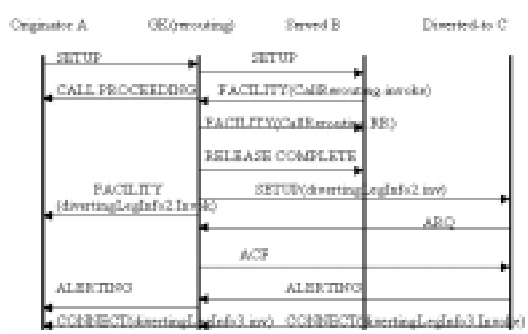


Figure 12: Signaling Flow for Call Forwarding partial rerouting in GK (H.323)

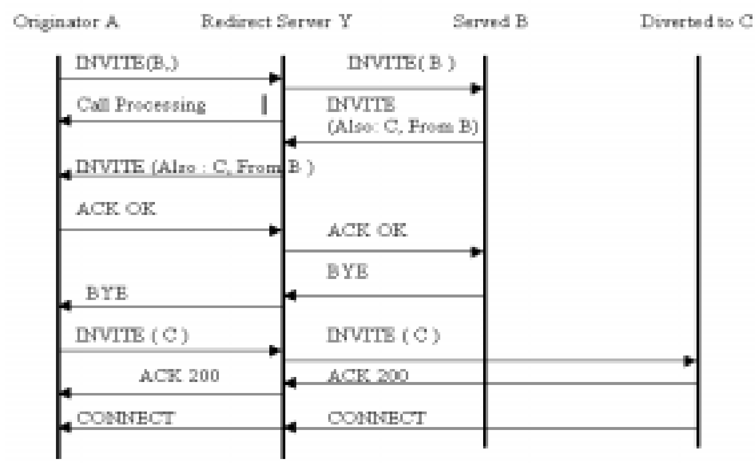


Figure 13: Signaling Flow for Call Forwarding with redirect server (SIP)

**Call Waiting**

Call Waiting allows the called party to receive a notification that a new party is trying to reach it while it is busy talking to another party.

In Figure 14 we illustrate the call waiting signaling flow for H.323. We do not show the gatekeeper for simplicity, since gatekeepers only pass on SS-CW (Supplementary Service-Call Waiting) operations transparently. We consider that party C calls the party B while B is in another call with A.

Actions at the served endpoint B: B returns an ALERTING message to C. B also optionally starts a timer, and locally provides a call indication to the user. If the served user B likes to accept the waiting call, B stops the timer, and sends a CONNECT message to the calling point.

Action at the calling endpoint C: On receipt of an ALERTING message, the calling endpoint may indicate call waiting to the calling user. Then the calling user may wait until the waiting call gets accepted, release the call, or choose other supplementary services.

SIP can provide call waiting service using the Call-Disposition header field, which allows the UAC to indicate how the server is to handle the call. The following is an example of Call Waiting Service provided by SIP.
The called party B is temporarily unreachable (e.g. it is in another call).
The caller indicates that it wants to have its call queued rather than rejected immediately via a "Call-Disposition: Queue" header field.
If the call is queued, the server returns "181 Queued"
When the callee becomes available, it will return the appropriate status response.
A pending call can be terminated by a SIP BYE request.

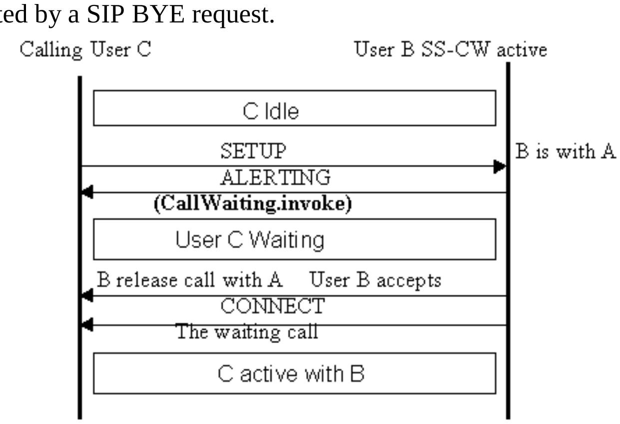


Figure 14: Signaling Flow for Call Waiting (H.323)

**Other Supplementary Services**

Other supplementary call control services supported by both H.323 and SIP are Call Park and Call Pickup, Call Completion on Busy Subscriber (SS-CCBS) or Camp-on, and Call Identification (Call Screening). H.323 and SIP signaling flow diagrams for those services are very similar. Table 2 lists all the supplementary services supported in H.323 along with the H.450 specification where the service is defined. The third column of the table indicates whether or not SIP supports the same service.

| Service Name | H.323 | SIP |
|---|---|---|
| Call Transfer Supplementary Service for H.323 | H.450.2 | Yes |
| Call Diversion Supplementary Service for H.323 | H.450.3 | Yes |
| Call Hold Supplementary Service for H.323 | H.450.4 | Yes |
| Call Park and Call Pickup Supplementary Services for H.323 | H.450.5 | Yes |
| Call Waiting Supplementary Service for H.323 | H.450.6 | Yes |
| Massage Waiting Indication Supplementary service for H.323 | H.450.7 | No |
| Conference out of Consultation Supplementary Service for H.323 | H.450.8 | No |
| Call Completion on Busy Subscriber for H.323 | H.450.9 | Yes |

Table 2: Supplementary services in H.323 and SIP.

### 3.1.3. Third-Party Control in SIP

Third-Party control is defined as the ability for a party to set up a call between two other parties without necessarily participating in the call. This feature is currently available only in SIP, although work is in progress to add the same functionality to H.323. The flexible and powerful SIP headers, which are very similar to those used in the Hyper-Text Transfer Protocol (HTTP) make the implementation simple. Third-party control is useful for many scenarios, including:

- A secretary dials for a manager
- Auto-dialer hands call to a telemarketer
- Attended call transfer
- Operator service

### 3.1.4. Capability Exchange

The capability exchange procedures are intended to ensure that multimedia signals that are transmitted can be received and processed appropriately by the receiving terminal.

H.323 uses the H.245 protocol for exchanging capabilities. The complete set of what a terminal can receive and decode is made known to the other terminal by transmission of its capability set. The terminal's total capabilities are described by a set of CapabilityDescriptor structures, each of which is a single SimultaneousCapabilities structure and a capabilityDescriptorNumber[6]. With this structure, very precise information about each terminal's capabilities can be expressed in a relatively compact structure.

SIP uses SDP for capability exchange. The caller can use an OPTION request to find out the capability of the callee. Currently, SIP does not have the full negotiation flexibility of H.245, due to the limited expressiveness of SDP. For example, SIP does not support asymmetric capabilities (receive or transmit only) and simultaneous capabilities of audio & video encoding.

## 3.2. Quality of Service (QoS)

Quality of Service is a term that encompasses many different aspects. The relevant QoS parameters for multimedia flows are the bandwidth, maximum delay, delay jitter, and packet loss rate. It is important for the call signaling and control protocols to provide support for communicating the required QoS parameters with the goal of meeting the required QoS levels. In addition to the above QoS parameters, call setup delay is another parameter that affects the perceived QoS, and it is highly dependent on the particular signaling protocol used. Call setup delay is also dependent on the transport protocol that is used to carry the signaling messages, in particular when some signaling messages are lost, and need to be retransmitted.

Therefore, here we first consider the QoS support by the signaling protocols for the multimedia flows. Then, we examine the call setup delay. Since the call setup delay is affected by error detection and error correction mechanisms, we then describe such mechanisms in H.323 and SIP.

### 3.2.1. QoS Support for Multimedia Flows

Gatekeepers in H.323 provide a rich set of control and management functions, including address translation, admission control, bandwidth control, and zone management. Some optional functions inside gatekeeper are call control signaling, call authorization, bandwidth management and call management. By contrast, SIP does not supply the management or control functions by itself but relies on other protocols.

Admission Control determines whether the network has sufficient resources to support the QoS required for a call, and accepts or rejects the call accordingly. In order to do admission control, the protocol must handle bandwidth management, call management, and bandwidth control. These are supported by H.323 but not by SIP.

In their current stages, neither H.323 nor SIP supports resource reservation by itself. Both of them recommend using external means for resource reservation, such as the Resource Reservation Protocol (RSVP) by the IETF IntServ working group[7]. RSVP addresses the needs of applications that require QoS, promising per-flow service. The reliance of IntServ on Per-flow State and per-flow processing is an impediment to its deployment in the Internet at large networks[8].

Recently, a new Differentiated Services architecture has drawn more attention. Unlike the IntServ model, in which applications explicitly request QoS reservations, Differentiated Services provides a simpler model: Network providers define a set of service levels that is general, simple and application independent. The applications and the users match their needs to specific service levels based on their performance and policy constrains[9].

H.323 v3 can offer some Differentiated Services based on QoS parameter negotiation (Bit rate, delay, and jitter)[10]. Upon initiation of a call, a terminal may request one of three service classes defined: "Guaranteed Service," "Controlled Service," and "Unspecified Service." Neither SIP nor the older versions of H.323 support a similar functionality.

### 3.2.2. Call Setup Delay

We define call setup delay as the number of round trips needed for establishing audio communication between the call participants. Call setup delay is very large in H.323 v1; it has been reduced significantly with the fast call setup procedure in H.323 v2. SIP and H.323 v3 both have a significantly more efficient call setup, resulting in relatively small delays.

H.323 uses H.225/Q931 signaling procedures to establish a connection between caller and callee. Depending on whether a gatekeeper is being used or not, a H.323 v1 call can take about 6 to 7 round-trip times, including setting up the Q.931 and H.245 TCP connections.

The fast call setup method is an option specified in H.323 v2 that reduces the number of delays up to three roundtrips involved in establishing a call and initial media streams by including H.245 logical channel information in the SETUP and CONNECT messages. With the fast call setup method, only G.711-based voice communication can be established between the two call parties, since the capabilities are not exchanged. If the parties wish to establish other types of media channels, they can optionally perform the H.245 capability exchange procedures after the G.711 channel is established.

In H.323 v3, either UDP or TCP can be used to carry call setup messages. Using UDP has the advantage that there is no roundtrip delay associated with establishing a transport-layer connection. When UDP is used, the call setup delay can be 1.5 or 2.5 round trips, depending on whether or not a gatekeeper is involved.

SIP call setup is very similar to the one in H.323 v3. However, if the UDP call setup fails, H.323 v3 has some advantages over SIP. H.323 v3 sets up a UDP connection and a TCP connection almost simultaneously, and provides an efficient mechanism to close the TCP connection if the UDP set up is successful. If the UDP setup fails, TCP can take over immediately. SIP operates UDP and TCP sequentially. This increases the call setup delay if the UDP connection is not available.

### 3.2.3. Error detection and correction

**Packet Loss**

H.323 v1 and v2 are based on the reliable transport protocol. They achieve the reliability based on transport protocol. The use of TCP would simplify the state machine for the call control protocol, since it has its own flow control, window control, and retransmission mechanisms to ensure the reliability.

H.323 v3 specifies its own retransmission policies for both sender and receiver for support TCP and UDP[11]. The sender starts two timers, T1 and T4, after it sends call setup PDU. If T1 expires before the caller received a response from the callee or GK, the sender retransmits the SETUP packet and starts a timer T3. If T3 expires, another retransmission is performed, and T3 is restarted. After a total of N1 transmission, the caller stops retransmission and reverts to the use of TCP call signaling instead of UDP.

With the first transmission of response, the receiver starts a timer T1. If T1 expires, the callee retransmits the packet and restarts a timer T3. If T3 expires, the callee sends another transmission of response and restarts T3. After the response message has been sent for a total N1 times, the callee stops re-transmitting and starts a timer T5. After T5 expires, the callee discards the conference/call identification information and associated state, and regards the setup of this session as failed.

The reliability of SIP is messages is achieved by having the client retransmit requests every 0.5 seconds until either a progress report (1xx) or final status (>200) response has been received. The server simply retransmits the original final response until an ACK is received. The client retransmits an ACK for every final message.

**Loop Detection**

Another common type of error is call forwarding loops, which may occur especially when multiple gatekeepers or SIP network servers are involved in setting up a call. There is no provision to prevent loops in H.323 v1 and v2. H.323 v3 defines a PathValue field to indicate the maximum number of gatekeepers that signaling message should traverse before being discarded. Using the PathValue field can reduce the rate of loop occurrence, but not as efficiently as the loop detection algorithm used in SIP. When a loop occurs without knowing the names of gatekeepers, the signaling messages will not be stopped until reaching the PathValue. Thus how to define a proper value for PathValue becomes a critical issue. Furthermore, when the network configuration changes, the PathValue may possibly need to be changed.

SIP provides loop detection algorithms similar to the one used in BGP (Border Gateway Protocol) to prevent searching loop[3]. It works through the *via* header field. Before a proxy server redirects any request, it checks the *via* field. If its own name is already in there, then a loop must have occurred. If its name is not on the list of via field, then the proxy server posts its name on the list, then transmits the request to another proxy server or endpoint.

**Fault Tolerance**

Errors happen in networks for a variety of reasons. The signaling protocol should have the capability to bypass network faults and provide a normal service whenever possible. A client should not notice the error during the services. This capability is called Fault Tolerance.

H.323 v3 provides better fault tolerance than SIP by redundant gatekeepers and endpoints. During registration, a gatekeeper may indicate alternate gatekeepers to the registering endpoint, which may be used in the event of a primary gatekeeper failure. Likewise, an endpoint may indicate a backup, redundant or alternate Transport Address. This allows an endpoint to have a secondary network interface or a secondary H.323 endpoint as a backup.

### 3.3. Scalability

In a fully operational system, it should be possible for every Internet host to act as an IP telephony client (albeit not all of them participating in a call simultaneously). It is estimated that the worldwide number of Internet users will reach 500 million by the year 2000, with the number increasing exponentially. Therefore, scalability of the current signaling protocols is extremely important. Many different aspects affect a system's scalability. We compare the scalability of H.323 and SIP in terms of their Complexity, Endpoint Location, Server Processing, Inter-Server Communication, Global Addressing, and Multipoint Communication.

#### 3.3.1. Complexity

H.323 is a rather complex protocol. It includes H.225 for call signaling, H.245 for call control, H.332 for large conferences, H.450.x (x=1,2,…9) for supplementary services, H.235 for security and encryption, and H.246 for interoperability with circuit-switched services. Many services require interactions between those sub-protocols, which increases the complexity but decreases the scalability.

On the other hand, SIP and SDP are less complicated. A basic SIP Internet Telephony implementation can be done using four headers (To, From, Call-ID, and Cseq) and three request types (INVITE, ACK, and BYE)[3]. This simplifies programming and maintenance, and better scalability is a consequence.

#### 3.3.2. Server Processing

No connection states are required in UDP. Therefore, large backbone servers based on UDP can operate in a stateless manner. This significantly reduces the memory requirements and improves the scalability.

In SIP, a transaction through servers and gateways can be either stateful or stateless. The stateless model simplifies the memory management, and the stateful model provides the sufficient information to forward the response correctly. H.323 v1 and v2 server processing is stateful, in which TCP was chosen as the transport protocol. Gatekeepers must hold the call states, as well as the TCP connections for the duration of a call. H.323 v3 supports the stateless processing model just as in SIP.

#### 3.3.3. Endpoint Location

The current mechanism for logical addressing and addressing resolution in H.323 standard is to utilize aliases (E.164 or H323ID) and a mapping mechanism supported by gatekeepers. When a client likes to make a connection, the gatekeeper may either return the endpoint's address to the client (in direct call model) or route the SETUP message to the called endpoint (in GK routed model). H.323 v3 defines a mechanism for inter-gatekeeper communication, which aids in locating an endpoint registered in a different zone or administrative domain.

SIP chooses an e-mail-like address, referred to as a *SIP URL*[3]. When a client wishes to send a request, it either sends it to a locally configured SIP proxy server or a SIP redirect server, independent of the Request-URL, or sends it to the IP address and port corresponding to the Request-URL. In the former case, SIP redirect or proxy server obtains information about a callee's possible location(s) from SIP location servers. SIP does not specify a means to locate endpoints registered in other administrative domains, and suggests the use of external mechanisms such as DNS.

### 3.4. Flexibility

A well-defined protocol should have the capability to extend the current functionality for further development, and should allow implementers to customize sub-components depending on individual interest. In this section, H.323 and SIP will be evaluated in terms of those aspects.

#### 3.4.1. Extensibility of Functionality

IP telephony technology is not yet mature. It is likely that new signaling capabilities and functionality need to be added. Also, different vendors may want to support additional features.

H.323 chooses vendor-defined NonStandardParam field in ASN.1 as its extension mechanism. NonStandardParam consists of vendor codes and an opaque code for that particular vendor. This approach has some limitations, as only the NonStandardParam field can be extended.

SIP offers a more flexible mechanism by providing a hierarchical namespace of feature names and hierarchically organized numerical error codes. The client inputs feature information in a SIP Require header. If there are some required features that the server does not support, the server sends back a hierarchical error code to the client. New features can either be registered with Internet Assigned Numbers Authority (IANA) or hierarchically derived from the feature owner's Internet domain name depending on whether the new feature can be derived directly.

### 3.4.2. Ease of customization

To customize the services, H.323 requires more interactions between its sub-protocols. For example, if the conference size changes from small to big, then a different protocol, H.332, has to be used. Also, since H.323 requires the full compatibility between each version, the customization will definitely increase the size of code. SIP uses its relatively simple header fields to handle those interactions. The text-based encoding makes the customization in SIP much easier than that in H.323.

## 3.5. Interoperability

An IP Telephony signaling protocol needs to cooperate with different versions, implementations, and other signaling protocols crossing worldwide networks. This capability is defined as interoperability.

### 3.5.1. Interoperability among Versions

The fully backward compatibility in H.323 enables all implementations based on different H.323 versions to be seamlessly integrated. This is important for customers. If a customer has bought client side products implemented by H.323 v2, he or she does not have to change anything when server side is upgraded to H.323 v3. The new server supports every function the customer had before with the older server.

In SIP, a newer version may discard some old features that are not expected to be implemented any more. This approach saves code size and reduces protocol complexity, but loses some compatibility between different versions. Some products implemented on the older version may not be supported in the new version.

### 3.5.2. Interoperability among Implementations

Different vendors on the market may implement the same protocol with different approaches. It is possible that endpoints in a large-scale system use different vendors' products, and simply complying with the standard does not guarantee interoperability among the different products.

H.323 provides an implementers' guide, which clarifies the standard and helps towards interoperability among different implementations. Furthermore, International Multimedia Teleconferencing Consortium (IMTC) iNOW! Activity Group defined an interoperability profile, which combines, clarifies and complements existing standards to provide a complete IP telephony interoperability protocol[12]. IMTC also organizes interoperability events, where different vendors can test their implementations against each other.

SIP, being in an earlier stage of development, thus far has not provided an implementation agreement. The interoperability between different implementations is still uncertain, although the first interoperability tests began recently.

### 3.5.3. Interoperability with Other Signaling Protocols

To support traditional telephony services, the VoIP signaling protocols have to support ISDN Signaling System 7 (SS7). SS7 performs out-of-band signaling in support of the call-establishment, billing, routing, and information-exchange functions of the public switched telephone networks (PSTN). There are two signaling specifications available in SS7 for different interfaces: Q.931 used for User-to-Network Interface (UNI) and ISUP used for Network-to-Network Interface (NNI).

H.323 embraces the more traditional circuit-switched approach based on the ISDN/Q.931 protocols. Q.931-like signaling messages are used in H.323 procedures, which makes it easier to interoperate with ISDN/Q.931. However, the call setup messages of H.323 are only a subset of those in SS7/ISUP. Because there is no established standard for relaying of SS7/ISUP messages over an H.323 network, H.323 can only translate a portion of SS7 messages in the conversion.

The H.32x family of recommendations offers specific standards to interoperate with other circuit-switched networks; for example, H.320 for ISDN and B-ISDN, H.324 for GSTN. Within those standards, the interoperability by gateways is well defined.

For SIP, no translation function for SS7 signaling messages is provided. Although there is an Internet draft on the subject, "A Functional Description of SIP-PSTN Gateway,"[13] the detailed information about signaling information transportation is not yet worked out.

### 3.6. Ease of Implementation

H.323 signaling messages are binary encoded using ASN.1 PER (Packet Encoding Rules)[14]. Within a system, the information represented using an abstract syntax must be mapped into some form for presentation to human users. Similarly, this abstract syntax must be mapped into some local format for storage. Such mappings require a special parser, which makes implementation and debugging more complicated.

SIP messages are text-based, using ISO 10646 in UTF-8 encoding[15]. Text-based encoding allows easy implementation in languages such as JAVA, Tcl and Perl, and easy debugging.

## 4. CONCLUSIONS

In Table 3, we present a summary of the comparison presented in this paper. Our major conclusions are as follows. In terms of functionality and services that can be supported, H.323 version 2 and SIP are very similar. However, supplementary services in H.323 are more rigorously defined. Therefore, fewer interoperability issues are expected among its implementations. Furthermore, H.323 has better compatibility among its different versions and better interoperability with PSTN. The two protocols are comparable in their QoS support (similar call setup delays, no support for resource reservation or class of service (QoS) setting), but H.323 version 3 will allow signaling of the requested QoS.

SIP's primary advantages are its flexibility to add new features and its relative ease of implementation and debugging. Finally, we note that H.323 and SIP are improving themselves by learning from the other side, and the differences between them are diminishing with each new version.

## REFERENCES


1. Bell Atlantic, "Signaling System 7 (SS7) Tutorial", http://www.webproforum.com/bellatlantic2/topic01.html
2. ITU-T Recommendation H.323, "Packet-Based Multimedia Communications Systems", February 1998.
3. M. Handley, H. Schulzrinne, E. Scholler, J. Rosenberg, " SIP: Session Initiation Protocol", RFC2543, IETF, March 1999.
4. H. Schulzrinne, J. Rosenberg, "A Comparison of SIP and H.323 for Internet Telephony", Network and Operating System Support for Digital Audio and Video (NOSSDAV), Cambridge, England, July 1998.
5. V. Jacobson and M. Handley," SDP: Session Description Protocol", RFC2327, IETF, April 1998, work in process.
6. ITU-T Recommendation H.245, "Control protocol for multimedia communication," February 1998.
7. Braden, R., Zhang, L., Berson, S., herzog, S. And Jamin, S.," Resource Reservation Protocol (RSVP) - version 1 Functional Specification," RFC 2205, proposed standard, September 1997.
8. Y. Bernet, R.Yavatkar, P. Ford, F. Baker, L, Zhang, " A Framework for End-to-End QoS Combining RSVP/Intserv and Differentiated Services ," Internet Draft, March, 1998.
9. S. Blake, et. al, "An Architecture for Differentiated Services," RFC 2475, December 1998.

10. "H.323 Differentiated Services and Their Protocol Architectures," ITU Telecommunication Standardization Sector, APC-1492, February, 1999
11. ITU-T Standardization Sector, "H.323 Annex E: call signaling over UDP", APC 1441, September, 1998
12. International Multimedia Teleconferencing Consortium, “iNOW!™ Standards-Based IP Telephony Interoperability Profile,” February 23, 1999
13. Steve Donovan, Mattew Cannon, "A Functional Description of SIP-PSTN Gateway", Internet Draft, IETF, November 1998.
14. ITU-T Recommendation X.691, “Information Technology - ASN.1 encoding rules – Specification of Packed Encoding Rules (PER),” Dec. 1997.
15. ISO/IEC 10646-1:1993, “Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane”, 1993.

| | **H.323 v1** | **H.323 v2** | **H.323 v3** | **SIP** |
|---|---|---|---|---|
| ***FUNCTIONALITY*** | | | | |
| **CALL CONTROL SERVICES:** | | | | |
| **Call Holding** | No | Yes | Yes | Yes |
| **Call Transfer** | No | Yes | Yes | Yes |
| **Call Forwarding** | No | Yes | Yes | Yes |
| **Call Waiting** | No | Yes | Yes | Yes |
| **ADVANCED FEATURES:** | | | | |
| **Third Party Control** | No | No | No | Yes |
| **Conference** | Yes | Yes | Yes | Yes |
| **Click-for-Dial** | Yes | Yes | Yes | Yes |
| **Capability Exchange** | Yes&Better | Yes &Better | Yes &Better | Yes |
| ***QUALITY OF SERVICE*** | | | | |
| **Call Setup Delay** | 6~7 RT | 3~4 RT | 2~3 RT | 2~3 RT |
| **RELIABILITY:** | | | | |
| **Packet Loss Recovery** | Through TCP | Through TCP | Better | Better |
| **Fault Detection** | Yes | Yes | Yes | Yes |
| **Fault Tolerance** | N/A | N/A | Better | Good |
| ***MANAGEABILITY*** | | | | |
| **Admission Control** | Yes | Yes | Yes | No |
| **Policy Control** | Yes | Yes | Yes | No |
| **Resource Reservation** | No | No | No | No |
| ***SCALABILITY*** | | | | |
| **Complexity** | More | More | More | Less |
| **Server Processing** | Stateful | Stateful | Stateful or stateless | Stateful or Stateless |
| **Inter-Server Communication** | No | No | Yes | Yes |
| ***FLEXIBILITY*** | | | | |
| **Transport Protocol Neutrality** | TCP | TCP | TCP/UDP | TCP/UDP |
| **Extensibility of Functionality** | Vendor Specified | | | Yes, IANA |
| **Ease of Customization** | Harder | Harder | Harder | Easier |
| ***INTEROPERABILITY*** | | | | |
| **VersionCompatibility** | N/A | Yes | Yes | Unknown |
| **SCN Signaling Interoperability** | Better | Better | Better | Worse |
| ***EASE OF IMPLEMENTATION*** | | | | |
| **Protocol Encoding** | Binary | Binary | Binary | Text |

Table 3: Comparison summary