Voice over IP : Protocols and Standards

Rakesh Arora, *arora@cis.ohio-state.edu*

Abstract

This paper first discusses the key issues that inhibit Voice over IP (VOIP) to be popular with the users. Then I discuss the protocols and standards that exist today and are required to make the VOIP products from different vendors to interoperate. The main focus is on H.323 and SIP (Session Initiation Protocol), which are the signaling protocols. We also discuss some hardware standards for internet telephony.

See Also: Voice over IP - Products, Services and Issues | Voice over IP (Lecture by Dr Jain) | Voice over ATM | H.323 and Associated Protocols | VOIP References | Books on Voice over IP and IP Telephony
Other Reports on Recent Advances in Networking
Back to Raj Jain's Home Page

Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/

TABLE OF CONTENTS

INTRODUCTION

Voice over IP (VOIP) uses the Internet Protocol (IP) to transmit voice as packets over an IP network. So VOIP can be achieved on any data network that uses IP, like Internet, Intranets and Local Area Networks (LAN). Here the voice signal is digitized, compressed and converted to IP packets and then transmitted over the IP network. Signaling protocols are used to set up and tear down calls, carry information required to locate users and negotiate capabilities.One of the main motivations for Internet telephony is the very low cost involved. Some other motivations are:

- Demand for multimedia communication
- Demand for integration of voice and data networks

1.1 Main Issues

For VOIP to become popular, some key issues need to be resolved. Some of these issues stem from the fact that IP was designed for transporting data while some issues have arisen because the vendors are not conforming to the standards. The key issues are discussed below [Munch98]:

- Quality of voice
  As IP was designed for carrying data, so it does not provide real time guarantees but only provides best effort service. For voice communications over IP to become acceptable to the users, the delay needs to be less than a threshold value and the IETF (Internet Engineering Task Force) is working on this aspect. To ensure good quality of voice, we can use either Echo Cancellation, Packet Prioritization (giving higher priority to voice packets) or Forward Error Correction [Micom] .

- Interoperability
  In a public network environment, products from different vendors need to operate with each other if voice over IP is to become common among users. To achieve interoperability, standards are being devised and the most common standard for VOIP is the H.323 standard, which is described in the next section.

- Security
  This problem exists because in the Internet, anyone can capture the packets meant for someone else. Some security can be provided by using encryption and tunneling. The common tunneling protocol used is Layer 2 Tunneling protocol and the common encryption mechanism used is Secure Sockets Layer (SSL).

- Integration with Public Switched Telephone Network(PSTN)
  While Internet telephony is being introduced, it will need to work in conjunction with PSTN for a few years. We need to make the PSTN and IP telephony network appear as a single network to the users of this service.
- Scalability
  As researchers are working to provide the same quality over IP as normal telephone calls but at a much lower cost, so there is a great potential for high growth rates in VOIP systems. VOIP systems needs to be flexible enough to grow to large user market and allow a mix of private and public services.

Back to Table of Contents

2. H.323 STANDARD

This is the ITU-T's (International Telecommunications Union) standard that vendors should comply while providing Voice over IP service. This recommendation provides the technical requirements for voice communication over LANs while assuming that no Quality of Service (QoS) is being provided by LANs. It was originally developed for multimedia conferencing on LANs, but was later extended to cover Voice over IP. The first version was released in 1996 while the second version of H.323 came into effect in January 1998. The standard encompasses both point to point communications and multipoint conferences. The products and applications of different vendors can interoperate if they abide by the H.323 specification.

2.1 Components of H.323

H.323 defines four logical components viz., Terminals, Gateways, Gatekeepers and Multipoint Control Units (MCUs). Terminals, gateways and MCUs are known as endpoints. These are discussed below [DataBeam]:

2.1.1 Terminals

These are the LAN client endpoints that provide real time, two way communications. All H.323 terminals have to support H.245, Q.931, Registration Admission Status (RAS) and Real Time Transport Protocol (RTP). H.245 is used for allowing the usage of the channels, Q.931 is required for call signaling and setting up the call, RTP is the real time transport protocol that carries voice packets while RAS is used for interacting with the gatekeeper.These protocols have been discussed later in the paper. H.323 terminals may also include T.120 data conferencing protocols, video codecs and support for MCU. A H.323 terminal can communicate with either another H.323 terminal, a H.323 gateway or a MCU.

2.1.2 Gateways

An H.323 gateway is an endpoint on the network which provides for real-time, two-way communications between H.323 terminals on the IP network and other ITU terminals on a switched based network, or to another H.323 gateway. They perform the function of a "translator" i.e. they perform the translation between different transmission formats, e.g from H.225 to H.221. They are also capable of translating between audio and video codecs. The gateway is the interface between the PSTN and the Internet. They take voice from circuit switched PSTN and place it on the public Internet and vice versa. Gateways are optional in that terminals in a single LAN can communicate with each other directly. When the terminals on a network need to communicate with an endpoint in some other network, then they communicate via gateways using the H.245 and Q.931 protocols.

2.1.3 Gatekeepers

It is the most vital component of the H.323 system and dispatches the duties of a "manager". It acts as the central point for all calls within its zone (A zone is the aggregation of the gatekeeper and the endpoints registered with it) and provides services to the registered endpoints. Some of the functionalities that gatekeepers provide are listed below [DataBeam][H.323]:

Address Translation: Translation of an alias address to the transport address. This is done using the translation table which is updated using the Registration messages.
Admissions Control : Gatekeepers can either grant or deny access based on call authorization, source and destination addresses or some other criteria.
Call signaling : The Gatekeeper may choose to complete the call signaling with the endpoints and may process the call signaling itself. Alternatively, the Gatekeeper may direct the endpoints to connect the Call

Signaling Channel directly to each other.
Call Authorization: The Gatekeeper may reject calls from a terminal due to authorization failure through the use of H.225 signaling. The reasons for rejection could be restricted access during some time periods or restricted access to/from particular terminals or Gateways.
Bandwidth Management: Control of the number of H.323 terminals permitted simultaneously access to the network. Through the use of H.225 signaling, the Gatekeeper may reject calls from a terminal due to bandwidth limitations.
Call Management: The gatekeeper may maintain a list of ongoing H.323 calls. This information may be neccesary to indicate that a called terminal is busy, and to provide information for the Bandwidth Management function.

2.1.4 Multipoint Control Units (MCU)

The MCU is an endpoint on the network that provides the capability for three or more terminals and gateways to participate in a multipoint conference. The MCU consists of a mandatory Multipoint Controller (MC) and optional Multipoint Processors (MP). The MC determines the common capabilities of the terminals by using H.245 but it does not perform the multiplexing of audio, video and data. The multiplexing of media streams is handled by the MP under the control of the MC. The following figure [Fig1] shows the interaction between all the H.323 components

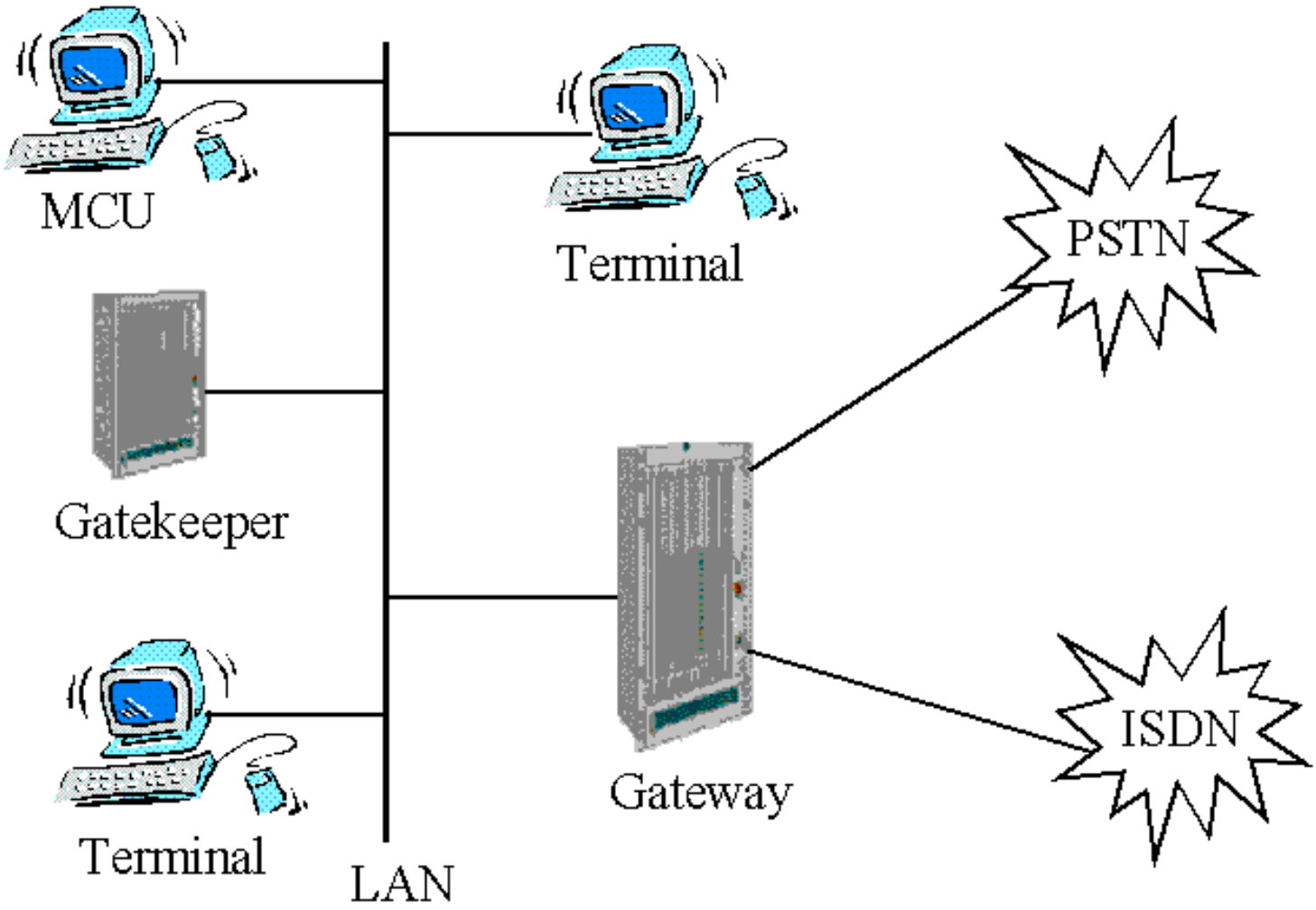

Fig 1. Components of H.323

Back to Table of Contents

2.2 H.323 Protocol Stack

The following figure [Fig 2] shows the H.323 protocol stack. The audio, video and registration packets use the unreliable User Datagram Protocol (UDP) while the data and control application packets use the reliable Transmission Control Protocol (TCP) as the transport protocol. Except for the T.120 protocol, the other protocols are described in the paper. The T.120 protocol is used for defining the data conferencing part.[Toga99]

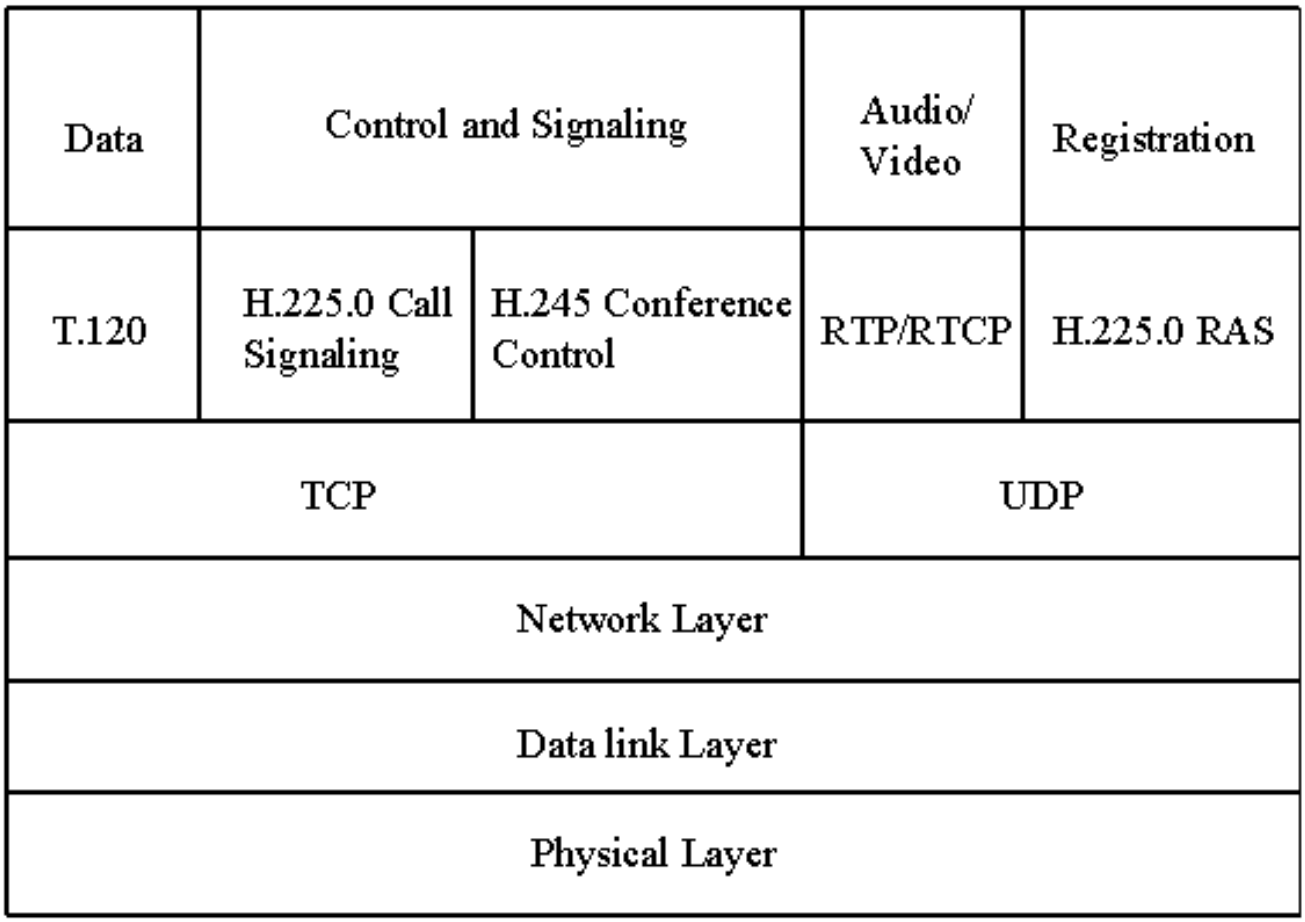

Fig 2. The protocol stack of H.323

Back to Table of Contents

2.3 Definitions

2.3.1 Zone

The collection of a gatekeeper and the endpoints registered with it is called a zone.

2.3.2 Network Address

For each H.323 entity, a network address is assigned and this address uniquely identifies the H.323 entity on the network. An endpoint may use different network addresses for different channels within the same call.

2.3.3 Alias Address

The alias address provides an alternate method of addressing the endpoint. It could be an email address, a telephone number or something similar. An endpoint may have one or more alias addresses associated with it and is unique within a

zone.

2.3.4 TSAP Identifier

For each network address, each H.323 entity may have several TSAP (Transport layer Service Access Point) identifiers. These TSAP identifiers allow multiplexing of several channels sharing the same network address. Endpoints have one well known TSAP identifier defined : the Call Signaling Channel TSAP identifier. Gateways also have one well known TSAP identifier defined : the RAS channel TSAP identifier and one well known multicast address defined : Discovery Multicast Address [H.323]

Back to Table of Contents

2.4 Control and Signaling in H.323

H.323 provides three control protocols viz., H.225.0/Q.931 Call Signaling, H.225.0 RAS and H.245 Media Control. H.225/ Q.931 is used in conjunction with H.323 and provides the signaling for call control. For establishing a call from a source to a receiver host, the H.225 RAS (Registration, Admission and Signaling) channel is used. After the call has been established, H.245 is used to negotiate the media streams.

2.4.1 H.225.0 : RAS

The RAS channel is used for the communication between the endpoints and the gatekeeper. Since the RAS messages are sent over UDP (an unreliable channel), so it recommends timeouts and retry counts for messages. The procedures defined by the RAS channel are [H.323]:

Gatekeeper discovery
This is the process that an endpoint uses to determine the gatekeeper with which it should register. The endpoint normally multicasts a Gatekeeper Request (GRQ) message asking for its gatekeeper. One or more gatekeepers may respond with the Gatekeeper Confirmation (GCF) message thereby indicating the willingness to be the gatekeeper for that endpoint. The response includes the transport address of the gatekeeper's RAS channel. Gatekeepers who do not want the endpoint to register with it can send a Gatekeeper Reject (GRJ) message. If more than one gatekeeper responds with GCF, then the endpoint may choose the gatekeeper and register with it. If no gatekeeper responds within a timeout interval, the endpoint may retransmit the GRQ.

Endpoint Registration
This is the process by which an endpoint joins a zone and informs the gatekeeper of its transport and alias addresses. All endpoints usually register with the gatekeeper that was identified through the discovery process. An endpoint shall send a Registration Request (RRQ) to a gatekeeper. This is sent to the gatekeeper's RAS channel Transport Address. The endpoint has the network address of the gatekeeper from the gatekeeper discovery process and uses the well known RAS channel TSAP Identifier. The gatekeeper shall respond with either a Registration Confirmation (RCF) or a Registration Reject (RRJ). The gatekeeper shall ensure that each alias address translates uniquely to a single transport address. An endpoint may cancel its registration by sending an Unregister Request (URQ) message to the gatekeeper. The gatekeeper shall respond with an Unregister Confirmation (UCF) message. A gatekeeper may cancel the registration of an endpoint by sending an Unregister Request (URQ) message to the endpoint. The endpoint shall respond with an Unregister Confirmation (UCF) message

Endpoint Location
An endpoint or gatekeeper which has an alias address for an endpoint and would like to determine its contact information may issue a Location request (LRQ) message. The gatekeeper with which the requested endpoint is registered shall respond with the Location Confirmation (LCF) message containing the contact information of the endpoint or the endpoint's gatekeeper. All gatekeepers with which the requested endpoint is not registered shall return Location Reject (LRJ) if they received the LRQ on the RAS channel

Admissions, Bandwidth Change, Status and Disengage
The RAS channel is also used for the transmission of Admissions, Bandwidth Change, Status and Disengage messages.

These messages are exchanged between an endpoint and a gatekeeper and are used to provide admissions control and bandwidth management functions. The Admissions Request (ARQ) message specifies the requested Call bandwidth. The gatekeeper may reduce the requested call bandwidth in the Admissions Confirm (ACF) message. An endpoint or the gatekeeper may attempt to modify the call bandwidth during a call using the Bandwidth Change Request (BRQ) message.

2.4.2 H.225.0 Call Signaling

The call signaling channel is used to carry H.225 control messages. In networks that do not contain a gatekeeper, call signaling messages are passed directly between the calling and called endpoints using the Call Signaling Transport Addresses. It is assumed that the calling endpoint knows the Call Signaling Transport Address of the called endpoint and thus can communicate directly. In networks that do contain the gatekeeper, the initial admission message exchange takes place between the calling endpoint and the gatekeeper using the gatekeeper's RAS channel transport address. The call signaling is done over TCP (reliable channel).

Call Signaling channel Routing
Call Signaling messages may be passed in two ways. The first way is Gatekeeper Routed Call Signaling where the call signaling messages are routed through the gatekeeper between the endpoints. The other alternative is Direct Endpoint Call Signaling where the call signaling messages are passed directly between the endpoints. Admissions messages are exchanged with the gatekeeper over the RAS channel, followed by an exchange of call signaling messages on a Call Signaling Channel which inturn is followed by the establishment of the H.245 Control Channel.

Control Channel Routing
When gatekeeper routed call signaling is used, there are two methods to route the H.245 Contol Channel. The first alternative establishes the H.245 Control Channel directly between the endpoints while in the second case, the establishment of the H.245 Control Channel is done through the gatekeeper.

2.4.3 H.245 Media and Conference Control

H.245 is the media control protocol that H.323 systems utilize after the call establishment phase has been completed. H.245 is used to negotiate and establish all of the media channels carried by RTP/RTCP. The functionality offered by H.245 are [Toga99]:

- Determining master and slave: H.245 appoints a Multipoint Controller (MC) which is held responsible for central control in cases where a call is extended to a conference
- Capability Exchange: H.245 is used to negotiate the capabilities when a call has been established. The capability exchange can occur at any time during a call, thereby allowing renegotiations at any time.
- Media Channel Control: After conference endpoints have exchanged capabilities, they may open and close logical channels of media. Within H.245 media channels are abstracted as logical channels (which are just identifiers)
- Conference Control: In conferences, H.245 provides the endpoints with mutual awareness and establishes the media flow model between all the endpoints.

Back to Table of Contents

2.5 Call Setup in H.323

The procedure to set up a call involves [Maddux99]:

- Discovering a gatekeeper which would take the management of that endpoint
- Registration of the endpoint with its gatekeeper
- Endpoint enters the call setup phase
- The capability exchange takes place between the endpoint and the gatekeeper
- The call is established

- When the endpoint is done, it can terminate the call. The termination can also be initiated by the gatekeeper

Back to Table of Contents

---

3. SESSION INITIATION PROTOCOL (SIP)

This is the IETF's standard for establishing VOIP connections. It is an application layer control protocol for creating, modifying and terminating sessions with one or more participants. The architecture of SIP is similar to that of HTTP (client-server protocol). Requests are generated by the client and sent to the server. The server processes the requests and then sends a response to the client. A request and the responses for that request make a *transaction*. SIP has INVITE and ACK messages which define the process of opening a reliable channel over which call control messages may be passed. SIP makes minimal assumptions about the underlying transport protocol. This protocol itself provides reliability and does not depend on TCP for reliability. SIP depends on the Session Description Protocol (SDP) for carrying out the negotiation for codec identification. 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. The services that SIP provide include [RFC2543]:

- User Location: determination of the end system to be used for communication
- Call Setup: ringing and establishing call parameters at both called and calling party
- User Availability: determination of the willingness of the called party to engage in communications
- User Capabilities: determination of the media and media parameters to be used
- Call handling: the transfer and termination of calls

3.1 Components of SIP

The SIP System consists ot two components [Jones99]:

3.1.1 User Agents:

A user agent is an end system acting on behalf of a user. There are two parts to it: a client and a server. The client portion is called the User Agent Client (UAC) while the server portion is called User Agent Server (UAS). The UAC is used to initiate a SIP request while the UAS is used to receive requests and return responses on behalf of the user.

3.1.2 Network Servers:

There are 3 types of servers within a network. A registration server receives updates concerning the current locations of users. A proxy server on receiving requests, forwards them to the next-hop server, which has more information about the location of the called party. A redirect server on receiving requests, determines the next-hop server and returns the address of the next-hop server to the client instead of forwarding the request.

3.2 SIP Messages

SIP defines a lot of messages. These messages are used for communicating between the client and the SIP server. These messages are:

INVITE: for inviting a user to a call
BYE: for terminating a connection between the two end points
ACK: for reliable exchange of invitation messages
OPTIONS: for getting information about the capabilities of a call
REGISTER: gives information about the location of a user to the SIP registration server.
CANCEL: for terminating the search for a user

Back to Table of Contents

3.3 Overview of SIP operation

Callers and callees are identified by SIP addresses. When making a SIP call, a caller first needs to locate the appropriate server and send it a request. The caller can either directly reach the callee or indirectly through the redirect servers. The Call ID field in the SIP message header uniquely identifies the calls. Below I briefly discuss how the protocol performs its operations [RFC2543].

3.3.1 SIP Addressing

The SIP hosts are identified by a SIP URL which is of the form sip:username@host. A SIP address can either designate an individual or a whole group.

3.3.2 Locating a SIP server

The client can either send the request to a SIP proxy server or it can send it directly to the IP address and port corresponding to the Uniform Request Identifier (URI).

3.3.3 SIP Transaction

Once the host part of the Request URI has been resolved to a SIP server, the client can send requests to that server. A request together with the responses triggered by that request make up a SIP transaction. The requests can be sent through reliable TCP or through unreliable UDP.

3.3.4 SIP Invitation

A successful SIP invitation consists of two requests: a INVITE followed by ACK. The INVITE request asks the callee to join a particular conference or establish a two party conversation. After the callee has agreed to participate in the call, the caller confirms that it has received that response by sending an ACK request. The INVITE request contains a session description that provides the called party with enough information to join the session. If the callee wishes to accept the call, it responds to the invitation by returning a similar session description.

3.3.5 Locating a User

A callee may keep changing its position with time. These locations can be dynamically registered with the SIP server. When the SIP server is queried about the location of a callee, it returns a list of possible locations. A Location Server in the SIP system actually generates the list and passes it to the SIP server.

3.3.6 Changing an Existing Session

Sometimes we may need to change the parameters of an existing session. This is done by re-issuing the INVITE message using the same Call ID but a new body to convey the new information.

Back to Table of Contents

3.4 Sample SIP Operation

Here a basic example of a SIP operation is given where a client is inviting a participant for a call. A SIP client creates an INVITE message for arora.32@osu.edu., which is normally sent to a proxy server. This proxy server tries to obtain the IP address of the SIP server that handles requests for the requested domain. The proxy server consults a Location Server to determine this next hop server. The Location server is a non SIP server that stores information about the next hop servers

for different users. On getting the IP address of the next hop server, the proxy server forwards the INVITE to the next hop server. After the User Agent Server (UAS) has been reached, it sends a response back to the proxy server. The proxy server in-turn sends back a response to the client. The client then confirms that it has received the reponse by sending an ACK. The exchange of messages is shown in the figure below (Fig 3). In this case, we had assumed that the client's INVITE request was forwarded to the proxy server. However, if it had been forwarded to a redirect server, then the redirect server returns the IP address of the next hop server to the client.The client then directly communicates with the UAS [Schulzrinne99b].

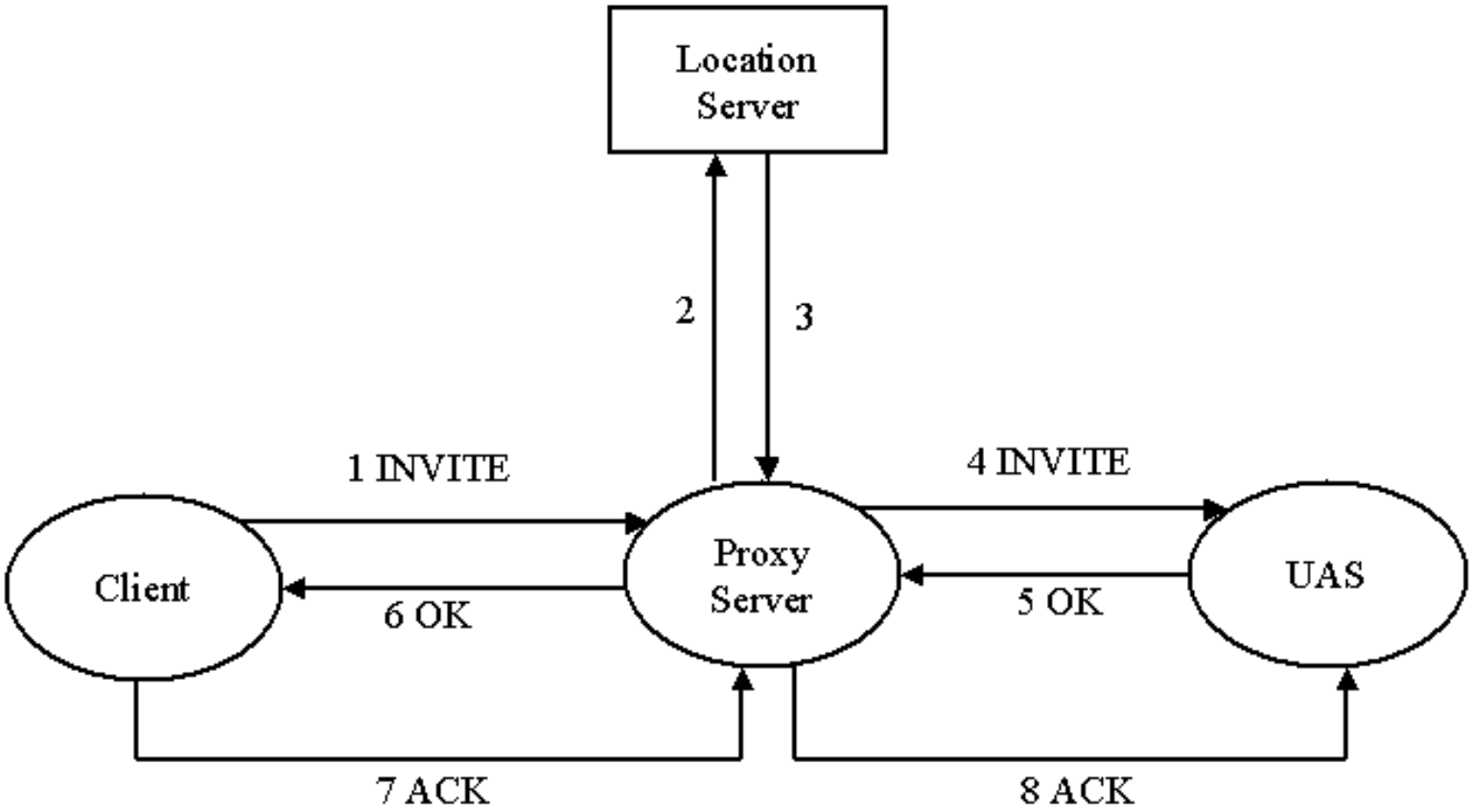

Fig 3. Example of a SIP operation

Back to Table of Contents

---

4. COMPARISON OF H.323 WITH SIP

The proponents of SIP claim that since H.323 was designed with ATM and ISDN signaling in mind, so H.323 is not well suited for controlling the voice over IP systems. They say that H.323 is inherently complex, has overheads and thus inefficient for VOIP. They also claim that H.323 lacks the extensibility required of the signaling protocol for VOIP. As SIP has been designed by keeping the Internet in mind, so it avoids both the complexity and extensibility pitfalls. SIP reuses most of the header fields, encoding rules, error codes and authentication mechanisms of HTTP. H.323 defines hundreds of elements while SIP has only 37 headers, each with a small number of values and parameters. H.323 uses a binary representation for its messages, which is based on ASN.1 while SIP encodes its messages as text, similar to HTTP. H.323 is not very scalable as it was designed for use on a single LAN and so has some problems in scaling though newer versions have suggested techniques to get around the problem. H.323 is still limited when performing loop detection in complex multi-domain searches. It can be done statefully by storing messages but this technique is not very scalable. On the other hand, SIP uses a loop detection method by checking the history of the message in the header fields, which can be done in a stateless manner. The advantage of SIP is that it is backed up by IETF, one of the most important standard

bodies while the advantage of H.323 is that it has a much larger chunk of the market presently [Schulzrinne98]. The table given below (Table1) lists the differences in a tabular form.

H.323	SIP
Complex protocol	Comparatively simpler
Binary representation for its messages	Textual representation
Requires full backward compatibility	Doesnt require full backward compatibility
Not very modular	Very modular
Not very scalable	Highly scalable
Complex signaling	Simple signaling
Large share of market	Backed by IETF
Hundreds of elements	Only 37 headers
Loop detection is difficult	Loop detection is comparatively easy

Table 1. Comparing H.323 with SIP

Back to Table of Contents

---

5. SUPPORTING PROTOCOLS

SIP works in conjunction with RSVP (Resource Reservation Protocol), RTP/RTCP (Real-time Transport Protocol), RTSP (Real-time Streaming Protocol), SAP(Session Announcement Protocol) and SDP (Session Description Protocol). RTP/RTCP is used for transporting real time data, RSVP for reserving resources, RTSP for controlled delivery of streams, SAP for advertising multimedia sessions and SDP for describing multimedia sessions. H.323. too works in conjunction with RTP and RTCP (Real-time Control Protocol). The present day voice gateways usually compose of two parts: the signaling gateway and the media gateway. The signaling gateway communicates with the media gateway using MGCP (Media Gateway Access Protocol). MGCP can interoperate with both SIP and H.323 [Huitema99]. The following figure (Fig 4) shows the signaling and transport protocols required for delivering voice over IP [Schulzrinne99b].

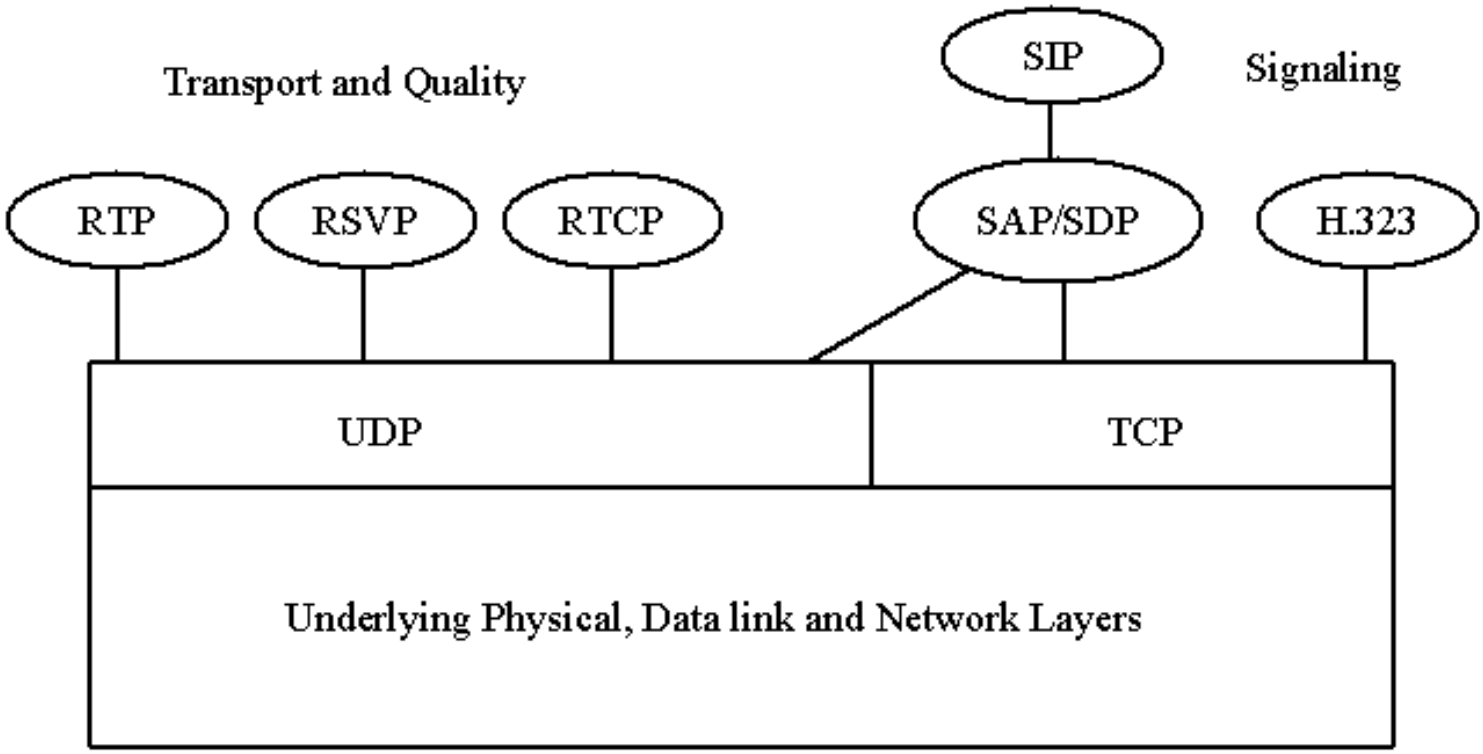

Fig 4. Signaling protocols SIP and H.323 with some of its supporting protocols

Back to Table of Contents

5.1 Media Gateway Control Protocol(MGCP)

It is a protocol that defines communication between call control elements (Call Agents) and telephony gateways. Call Agents are also known as Media Gateway Controllers. It is a control protocol, allowing a central coordinator to monitor events in IP phones and gateways and instructs them to send media to specific addresses. It resulted from the merger of the Simple Gateway Control Protocol and Internet Protocol Device Control. The call control intelligence is located outside the gateways and are handled by external call control elements, the Call Agent. MGCP assumes that these call control elements or Call Agents will synchronize with each other to send coherent commands to the gateways under their control. It is a master/slave protocol, where the gateways are expected to execute commands sent by the Call Agents. It has introduced the concepts of connections and endpoints for establishing voice paths between two participants, and the concepts of events and signals for establishing and tearing down calls. Since the main emphasis of MGCP is simplicity and reliability and it allows programming difficulties to be concentrated in Call Agents, so it will enable service providers to develop reliable and cheap local access systems. [IDMGCP][Huitema99]

5.1.1 Endpoints and Connections

Endpoints are the sources or sinks of data. An example could be an interface on a gateway that terminates a trunk connected to a PSTN switch. Connections may be either point to point or multipoint. A connection is either an association between two endpoints (point to point) or it is an association between multiple endpoints (multipoint). Once the association is established, data transfer can take place. Connections can be established over a number of bearer networks viz., TCP/IP, ATM etc.

5.1.2 Events and Signals

A call agent may ask to be notified about certain *events* occuring in an endpoint, such as off-hook, on-hook, dialed digits, and may request that a certain *signal* be applied to an endpoint such as dial-tone, busy tone or ringing. Events and signals are grouped in packages that are supported by a particular type of endpoint e.g., one package may support a certain group of events and signals for analog access lines.

5.1.3 Creating Connections

Connections are created on the call agent at each endpoint that will be involved in the call. When the two endpoints are located on gateways that are managed by the same call agent, the creation is done via the following three steps [IDMGCP]:

- The Call Agent asks the first gateway to create a connection on the first endpoint. The response sent by the gateway includes a session description that contains pertinent information required by third parties to be able to send packets to the new connection that has been created.
- The Call Agent then sends the session description of the first gateway to the second gateway and asks it to create a connection on the second endpoint. The second gateway responds by sending its own session description.
- The Call Agent uses a modify connection command to provide this second session description to the first endpoint. Now communication can occur in both directions.

When the two endpoints are located on gateways that are managed by the different call agents, these two call agents shall exchange information through a call agent to call agent signaling protocol, in order to synchronize the creation of the connection on the two endpoints.

5.1.4 Commands

The MGCP implements the media gateway control interface as a set of transactions. The transactions are composed of a command and a mandatory response. There are 8 types of command [Huitema99]:

CreateConnection: The CreateConnection command is used to attach an endpoint to a specific IP address and port. To create a connection, a CreateConnection request is required for the remote endpoint also. If the request is successfully acknowledged by the gateway, then a ConnectionId is returned that uniquely identifies the connection.
ModifyConnection: This command is used by the call agent to modify the parameters of an active connection. The ConnectionId is passed to identify the connection.
DeleteConnection: This command is used by either the call agent or the gateway to delete an existing connection. The response includes a list of parameters about the status of the connection.
NotificationRequest: If a call agent wants to be informed about the occurrence of specified events in an endpoint, then it can send this request to the gateway. Events could be: off hook transition, flash-hook, continuity tone detection, etc. A notification may be requested for a continuity tone detection event in a gateway.
Notify: The response to the NotificationRequest is sent via the Notify command by the gateway. The notify command includes a list of events that the gateway observed.
AuditEndpoint: This command is used by the call agent to get details about the status of an endpoint/several endpoints and the response from the gateway contains the requested information
AuditConnection: To obtain information for a specific connection of an endpoint, the call agent uses this command. The connection is identified by the ConnectionId and the response from the gateway contains the requested information.
RestartInProgress: This command is used by the gateway to indicate that an endpoint or a bunch of endpoints have been taken in/out of service. It also includes a parameter that indicates the type of restart (graceful restart/ forced restart/delayed restart)

Back to Table of Contents

5.2 RTP and RTCP (Real-time Transport Protocol and Real-time Control Protocol)

RTP supports the transfer of real-time media (audio and video) over packet switched networks. It is used by both SIP and H.323. The transport protocol must allow the receiver to detect any losses in packets and also provide timing information so that the receiver can correctly compensate for delay jitter. The RTP header contains information that assist the receiver to reconstruct the media and also contains information specifying how the codec bitstreams are broken up into packets. RTP does not reserve resources in the network but instead it provides information so that the receiver can recover in the presence of loss and jitter.[Chunlei97][RFC1889]

The functions provided by RTP include:

- Sequencing: The sequence number in the RTP packet is used for detecting lost packets
- Payload Identification: In the Internet, it is often required to change the encoding of the media dynamically to adjust to changing bandwidth availability. To provide this functionality, a payload identifier is included in each RTP packet to describe the encoding of the media
- Frame Indication: Video and audio are sent in logical units called frames. To indicate the beginning and end of the frame, a frame marker bit has been provided
- Source Identification: In a multicast session, we have many participants. So an identifier is required to determine the originator of the frame. For this Synchronization Source (SSRC) identifier has been provided.
- Intramedia Synchronization: To compensate for the different delay jitter for packets within the same stream, RTP provides timestamps which are needed by the play-out buffers.

RTCP is a control protocol and works in conjunction with RTP. In a RTP session, participants periodically send RTCP packets to obtain useful information about QoS etc. The additional services that RTCP provides to the participants are:

- QoS feedback: RTCP is used to report the quality of service. The information provided includes number of lost packets, Round Trip Time, jitter and this information is used by the sources to adjust their data rate.
- Session Control: By the use of the BYE packet, RTCP allows participants to indicate that they are leaving a session
- Identification: Information such as email address, name and phone number are included in the RTCP packets so that all the users can know the identities of the other users for that session.
- Intermedia Synchronization: Even though video and audio are normally sent over different streams, we need to synchronize them at the receiver so that they play together. RTCP provides the information that is required for synchronizing the streams.

Back to Table of Contents

5.3 Real-Time Streaming Protocol (RTSP)

RTSP, the Real Time Streaming Protocol, is a client-server protocol that provides control over the delivery of real-time media streams. It provides "VCR-style" remote control functionality for audio and video streams, like pause, fast forward, reverse, and absolute positioning. It provides the means for choosing delivery channels (such as UDP, multicast UDP and TCP), and delivery mechanisms based upon RTP. RTSP establishes and controls streams of continuous audio and video media between the media servers and the clients. A media server provides playback or recording services for the media streams while a client requests continuous media data from the media server. RTSP acts as the "network remote control" between the server and the client. It supports the following operations:[Chunlei97][RFC2326]

- Retrieval of media from media server: The client can request a presentation description, and ask the server to setup a session to send the requested data. The server can either multicast the presentation or send it to the client using unicast.
- Invitation of a media server to a conference: The media server can be invited to the conference to play back media or to record a presentation.
- Addition of media to an existing presentation: The server or the client can notify each other about any additional

media that has become available.

Features of RTSP include:

- RTSP is an application level protocol with syntax and operations similar to HTTP, but works for audio and video. It uses URLs like those in HTTP.
- An RTSP server needs to maintain states, using SETUP, TEARDOWN and other methods.
- Unlike HTTP, in RTSP both servers and clients can issue requests.
- RTSP is implemented on multiple operating system platforms and it allows interoperability between clients and servers from different manufacturers.

Back to Table of Contents

5.4 Resource Reservation Protocol (RSVP)

The network delay and Quality of Service are the most hindering factors in the voice-data convergence. The most promising solution to this problem has been developed by IETF viz., RSVP. RSVP can prioritize and guarantee latency to specific IP traffic streams. RSVP enables a packet-switched network to emulate a more deterministic circuit switched voice network. With the advent of RSVP, VOIP has become a reality today. With RSVP enabled, we can accomplish voice communication with tolerable delay on a data network. RSVP requests will generally result in resources being reserved in each node along the data path. RSVP requests resources in only one direction, therefore it treats a sender as logically distinct from a receiver, although the same application process may act as both a sender and a receiver at the same time. RSVP is not itself a routing protocol, it is designed to operate with current and future unicast and multicast routing protocols. In order to efficiently accommodate large groups, dynamic group membership, and heterogeneous receiver requirements, RSVP makes receivers responsible for requesting a specific QoS. A QoS request from a receiver host application is passed to the local RSVP process. The RSVP protocol then carries the request to all the nodes along the reverse data path to the data source. RSVP has the following attributes [RFC2205]:

- It is receiver oriented
- It supports both unicast and multicast
- It maintains soft state in routers and hosts, providing graceful support for dynamic membership changes
- It provides transparent operation through routers that do not support it

Back to Table of Contents

5.5 Session Description Protocol (SDP)

SDP is intended for describing multimedia sessions for the purpose of session announcement, session invitation etc. The purpose of SDP is to convey information about media streams in multimedia sessions to allow the recipients of a session description to participate in the session. SDP includes the following information: [RFC2327]

- Session name and purpose
- Address and port number
- Start and stop times
- Information to receive those media
- Information about the bandwidth to be used by the conference
- Contact information for the person responsible for the session

The above information is conveyed in a simple textual format. When a call is set up using SIP, the INVITE message contains an SDP body describing the session parameters acceptable to the calling party. The response from the callee includes a SDP body describing the capabilities of the callee. In general, SDP must convey enough information to be able to join a session and to announce the resources to be used to non-participants that may need to know. The media information that SDP sends are: type of media (audio or video), transport protocol (RTP, UDP etc) and media format

(MPEG video, H.263 video etc).

Back to Table of Contents

5.6 Session Announcement Protocol (SAP)

This protocol is used for advertising the multicast conferences and other multicast sessions. A SAP announcer periodically multicasts an announcement packet to a well known multicast address and port (port number 9875). A SAP listener learns of the multicast scopes using the Multicast Scope Zone Announcement Protocol and listens on the well known SAP address and port for those scopes. There is no rendezvous mechanism – the SAP announcer is not aware of the presence or absence of any SAP listeners. A SAP announcement is multicast with the same scope as the session it is announcing, ensuring that the recipients of the announcement can also be potential recipients of the session being advertised. If a session uses addresses in multiple administrative scope ranges, it is necessary for the announcer to send identical copies of the announcement to each administrative scope range. Multiple announcers may announce a single session, as an aid to robustness in the face of packet loss and failure of one or more announcers. The time period between repetitions of an announcement is chosen such that the total bandwidth used by all announcements on a single SAP group remains below a preconfigured limit. Each announcer is expected to listen to other announcements in order to determine the total number of sessions being announced on a particular group. SAP is intended to announce the existence of a long-lived wide area multicast sessions and involves a large startup delay before a complete set of announcements is heard by a listener. In order to reduce the delays inherent in SAP, it is recommended that proxy caches be deployed. A SAP proxy is expected to listen to all SAP groups in its scope and maintain an up-to-date list of all announced sessions along with the time each announcement was last received. SAP also contains mechanisms for ensuring integrity of session announcements, for authenticating the origin of an announcement and for encrypting such announcements. [IDSAP]

Back to Table of Contents

---

6. HARDWARE STANDARDS

Some hardware standards for computer telephony have come up over the past few years. They attempt to provide interoperability among the telephony products form different vendors. Two of these standards (SCBus and S.100) are discussed below:

6.1 SCBUS

The SCBus is a high speed digital TDM (Time Division Multiplexing) bus developed for computer telephony. It is a standalone component of SCSA (Signal Computing System Architecture) that makes it easier to build more scalable systems using devices from multiple vendors. It provides tight integration of hardware resources from different vendors. The features provided by SCBus include [SCSA]:

- It is based on a single distributed switching model
- It provides clock management for real-time communications
- it allows developers to build large distributed systems
- It supports 16 synchronous serial data lines for real-time communication between devices in a single node

The SCBus standard has been endorsed by American National Standards Institute (ANSI) and telephony products from several vendors are based on it [Jain98].

6.2 S.100

S.100 is a standard API (Application Programming Interface) for computer telephony. It provides an effective way to develop computer telephony applications in an open environment. S.100 is based on a client-server model and the client applications use a collection of services to allocate, configure, and operate hardware resources. The implementation details of call processing hardware and switch fabrics are hidden by S.100 so as to allow portable applications to be written.
The services provided by S.100 can be mainly categorized under the following heads [ECTF] [Jain98]:

- Session/Event Management: Session/Event Management is the collection of services that allow a client to authenticate itself to a S.100 server and allows it to manage message communication between the client and the server. It provides a logical channel between the client application and the server, and an associated event queue through which the application can receive events from the server.
- Group Management: The function provided by it allows a group to be treated as a single entity by the application. It configures the group, keeps track of the resources owned by the group and the session that owns the group.
- Resource Allocation and Management: In order for a group to actually perform an operation, all of the hardware components required by the operation must be allocated to the application and properly configured. The resource management service takes care of this issue.
- Run Time Control: It is a mechanism provided by the S.100 server which allows a group resource currently performing an operation to modify that operation as the result of a condition detected by another resource in the group.

The S.100 standard has been endorsed by Enterprise Computer Telephony Forum (ECTF)

Back to Table of Contents

---

7. SUMMARY

In this paper, we discussed the signaling protocols H.323 (ITU-T standard) and SIP (IETF standard). We compared both the protocols and noted that although H.323 has more share of the market at present, but SIP is a much better protocol given its simplicity and scalability. We also discussed MGCP, which is a gateway protocol whereby the Call Agent controls the signaling gateway. For both H.323 and SIP, we need some real-time protocols that does the actual transport. RTP and RTCP are used for the real-time transport and controling. RTSP is used to provide controlled delivery of media streams. We also saw some protocols that are required in conjunction with SIP so as to advertise the session (SAP) and give a description of the session (SDP). RSVP is used to reserve resources in the network and thereby provide some Quality of Service. Finally, we discussed two hardware standards, viz SCBus and S.100.

A table summarizing the key protocols and standards can be found in Appendix A.

Back to Table of Contents

APPENDIX A : FUNCTIONS OF THE KEY PROTOCOLS AND STANDARDS

H.323	Key ITU-T protocol that provides interoperability
H.225	Provides call signaling and registration
H.245	Negotiates the usage of the media channels
SIP	IETF standard for providing voice over IP

MGCP	Gateway protocol that defines communication between the call agent and the signaling gateway
RTP	Provides real-time transport over packet switched networks
RTCP	Control protocol that provides feedback to the application
RSVP	Responsible for providing QoS by reserving resources
RTSP	Provides control over delivery of real-time media streams
SDP	Describes the multimedia session
SAP	Advertises the multicast conferences/ sessions
SCBus	Hardware standard endorsed by ANSI
S.100	Hardware standard endorsed by ECTF

Back to Table of Contents

References

[H.323] ITU, "Packet Based Multimedia Communications Systems", Feb 1998, 125 pages
This describes the H.323 standard in detail

[Toga99] J Toga, J Ott, "ITU-T Standardization activities for interactive multimedia communications on packet-based networks: H.323 and related recommendations", IEEE Computer Networks, Feb 1999, pp. 205-223
This paper describes the H.323 standard and also discusses the security mechanism that has been put in place in version 2.

[DataBeam] DataBeam, "A primer on the H.323 series standard", 20 pages,
http://www.databeam.com/h323/h323primer.html
This paper is a good starting point to learn about the H.323 standard

[Schulzrinne99a] H Schulzrinne, J Rosenberg, "The IETF Internet Telephony Architecture and Protocols", IEEE Network, May/June 1999 pp. 18-23
http://207.127.135.8/ni/private/1999/may/schulzrinne.html (IEEE Membership required)
This paper discusses protocols that provide a partial solution for internetworking Internet telephony and PSTN.

[Schulzrinne99b] H Schulzrinne, J Rosenberg, "Internet Telephony: architecture and protocols - an IETF perspective", IEEE Computer Network, Feb 1999 pp. 237-255, http://www.cs.columbia.edu/~hgs/papers/Schu9902_Internet.ps.gz
This paper discusses the IETF protocol components that are required for VOIP.

[IDMGCP] "Media Gateway Control Protocol (MCGP)", August 1999, 111 pages,
ftp://www.ietf.org/internet-drafts/draft-huitema-megaco-mgcp-v0r1-05.txt
This Internet Draft gives a detailed explanation of the MGCP protocol

[Schulzrinne98] H Schulzrinne, J Rosenberg, "A Comparison of SIP and H.323 for Internet Telephony", July 1998, 4 pages, Proc NOSSDAV'98
http://www.cs.columbia.edu/~hgs/papers/Schu9807_Comparison.pdf
This paper compares the ITU's standard with the IETF standard.

[RFC2326] "Real Time Streaming Protocol (RTSP)", April 1998, 80 pages, http://www.ietf.org/rfc/rfc2326.txt
This RFC explains the RTSP protocol in length.

[RFC2543] "SIP : Session Initiation Protocol", March 1999, 132 pages, ftp://ftp.isi.edu/in-notes/rfc2543.txt
This RFC explains how the sessions are initiated.

[RFC2327] "SDP : Session Description Protocol"ftp://ftp.isi.edu/in-notes/rfc2327.txt

This RFC explains the description of the session that occurs during the handshaking of the client and the SIP server.

[IDSAP] "Session Announcement Protocol" http://www.ietf.org/internet-drafts/draft-ietf-mmusic-sap-v2-03.txt
This Internet Draft explains how the different sessions are advertised in a unicast/multicast manner.

[Rizzetto99] D Rizzetto, C Catania, "A Voice over IP Service Architecture for Integrated Communications", IEEE Internet Computing May/June 1999 pg 53-62
http://computer.org/internet/ic1999/w3053abs.htm
This paper proposes an architecture that integrates the circuit-switched communications with the Internet.

[Jain98] Raj Jain, "Voice over IP: Issues and Challenges", Nortel, Canada, August 14, 1998 and Southwestern Bell, Atlanta, October 21, 1998, 42 slides, http://www.cis.ohio-state.edu/~jain/talks/voip.htm
This slideshow gives a good introduction to VOIP and its standards

[Huitema99] C Huitema, J Cameron et al, "An Architecture for Residential Internet Telephony Service" IEEE Network May/June 1999 pg 50-56
http://computer.org/internet/ic1999/w3073abs.htm
This paper proposes an architecture based on the decomposition of the gateway functionality. The MGCP protocol is discussed.

[Maddux99] Michel Maddux, "Compaq CustomSystems Gatekeeper Implementation", Compaq white paper, April 1990, 19 pages http://www.digital.com/info/LIW0JF/LIW0JFHM.HTM
This paper give a good overview of H.323 standard and MGCP

[Rosenberg99] J Rosenberg,J Lenox, H Schulzrinne, "Programming Internet Telephony Services", IEEE Network May/June 1999 pp. 42-49
This paper discusses programming issues and brings up two solutions - one for trusted users and another for untrusted users.

[Polyzois99] C Polyzois, K Purdy, et al, "From POTs to PANs :A Commentary on the Evolution of Internet Telephony" IEEE May/June 1999 pg 58-64
http://computer.org/internet/ic1999/w3083abs.htm
This paper discusses the evolution of Internet telephony and the motivation behind it.

[SCSA] SCSA ,"SCBus Hardware Model", 4 pages, http://www.scsa.org/info/scbus.htm
This paper gives a brief overview of SCBus

[ECTF] ECTF, "S.100 Revision 1.0 Media Services C Language. Application Programming Interfaces", 420 pages, http://www.ectf.org/ectf/pubdocs/s100r1-0.pdf
This document describes the S.100 standard in detail.

[Micom] Micom, "Voice/Fax over IP: Internet, Intranet and Extranet", 47 pages, http://www.saintrochtree.com/cgi-bin/go2.pl?http://www.micom.com/WhitePapers/whtpaper.pdf
This white paper gives us an introduction to VOIP

[Chunlei97] Chunlei Liu, "Multimedia Over IP: RSVP, RTP, RTCP, RTSP", Jan 1998, 23 pages, http://www.cis.ohio-state.edu/~jain/cis788-97/ip_multimedia/index.htm
This survey paper explains the RSVP, RTSP and RTP protocols.

[Jones99] R Jones, J Cruz, "Carrier Class Voice over IP" , August 1999, 9 pages, http://www.digital.com/info/LIW0PF/
This white paper gives us a brief introduction of Internet telephony

[Munch98] Bjarne Munch, "IP Telephony – Today/Tomorrow/Ever?", Ericsson July 1998, 13 pages
http://www.ericsson.com/ipservices/literature/pdf/whitepaper.pdf
This white paper discusses the key issues that need to be resolved for VOIP to become popular

[Techguide] "Voice over IP", 24 pages, http://www.techguide.com/comm/sec_html/voiceip.shtml
It is a good introduction to P

[RFC2205] "Resource Reservation Protocol", Sept 1997, 97 pages
http://www.ietf.org/rfc/rfc2205.txt
This RFC explains the RSVP protocol in detail

[RFC1889] "RTP: A Transport Protocol for Real-time applications", Jan 1996, 65 pages
http://www.ietf.org/rfc/rfc1889.txt
This RFC describes RTP, the real-time end-to-end transport protocol

Books

[Black99] Ulyess Black, "Voice over IP", 1999, 328 pages, Prentice Hall
This book explains the protocols for VOIP
[Goralski99] W Goralski, M Kolon, "IP Telephony", 1999, 468 pages, McGraw Hill
This book gives a good overview of Voice over IP.

Back to Table of Contents

List of Acronyms

SIP	Session Initiation Protocol
ITU	International Telecommunications Union
SAP	Session Announcement Protocol
MGCP	Media Gateway Control Protocol
SDP	Session Description Protocol
RSVP	Resource Reservation Protocol
RTP	Real Time Transport Protocol
RTCP	RTP Control Protocol
MCU	Multipoint Control Unit
UAS	User Agent Server
UAC	User Agent Client
RAS	Registration, Admission and Status
TSAP	Transport layer Service Access Point

Back to Table of Contents

Last Modified: November 23,1999.
Note: This paper is available on-line at http://www.cis.ohio-state.edu/~jain/cis788-99/voip_protocols/index.html