VOICE OVER IP

Saengrawee Pratoomtong

Abstract

The data networking has been growing very rapidly for the past decade. The Internet has been use extensively all over the global. The popularity of Internet using has enabled a lot of Internet base application. Internet telephony is one of the applications that exploit the Internet to give an economical way to communicate. The new codec technology and the improvement in compression technique along with the rapid improvement in hardware and software system making the idea of the internet telephony possible. This survey introduces an overview of the Internet telephony to give a better understanding of this new service.

What is voice over IP?

“Voice over IP is a technique for sending real-time, full duplex voice communications over the Internet or internal IP network”[1]. The popular use of voice over IP is the Internet telephony.

How Voice can be sent over the Internet?

- Sample analog voice signal into digital samples. This is done with a Coder-Decoder chip (CODEC). The sample rate is 8000 samples/sec, which result in 64 Kbps data.
- In the voice data, there is silence, which enable the voice data to be compressed to minimize the amount of the data that need to be transmitted. This is done with a digital signal-processing chip, which conforms to international compression standards, such as G.729 or G.723.
- Create data packets form the compresses data bit stream and put them in a format that suitable to transmit over TCP/IP or UDP data network.
- At the other end, the packets are converts to a serial data stream, uncompressed and converted back to and analog voice signal.
- These entire processes, including call set-up and control are described in either the International Telecommunications Union (ITU-T) standard, H.323 or the Internet Engineering Task Force (IETF) standard, Session Initiate Protocol (SIP).

Three scenario of voice over IP communication.

- PC to PC. This required PC with a sound card, IP telephony software, such as Microsoft Net Meeting, Media Ring. Video is optional. This is the simplest case where the users need to know each other IP address and it require none of the gateway. The voice is directly transmitted to the IP network.
- PC to phone. This required a gateway that connect IP network to phone network. These gateways provide 2 ways interface between the Public Service Telephone
Network (PSTN) and the Internet. This allows voice conversations between users with standard phones. Gateways overcome the addressing problem, i.e. to address user on a multimedia PC, the user’s Internet Protocol (IP) address must be known but to address a remote user with a gateway product, only user’s phone number must be known. An example of the applications are Net2Phone, dialpad.com

- Phone to Phone. This required more gateways to connect IP network to phone network. The IP network could be intra-net of the Internet. Examples of the applications are Aplio H323 to phone interface module. For more product detail, visit http://www.applio.com, Cisco’s IP phone.

Protocols and Standards

Signaling protocols play the important role in the Internet telephony. The key roles are

- User location – This is to find a location of a specify user on the network so that the session establishment request can reach that user on the network. User can be reach by any mean at the same time such as by PC or phone. This function is very important for a user who’s PCs doesn’t have a permanent address. This is the case for almost all modem connections which address is to be issued to PCs dynamically using a Dynamic Host Configuration Protocol (DHCP)
- Session Establishment - This signaling protocol allows the called party to accept the call, reject it, or redirect it to another network location such as voice mail or a web page.
- Session negotiation – This allows the parties involved in the session to settle on a set of session parameter such as type of media stream (e.g. video, audio), speech and compression algorithm, multicast or unicast addresses and ports since the multimedia session being set up may comprise of various type of applications.
- Call Participant Management- This allows a new member to be added to a session and existing member to exit the session.
- Feature Invocation – Call feature such as hold, transfer, and mute, require communication between parties.

Several protocols needed to be combine to full fill these roles. There are 2 major standard for Voice over IP, H.323 (ITU-T standard) and SIP/SDP (IETF standard) to perform many of the above roles.

H.323

H.323 is an umbrella recommendation from ITU-T that sets standards for multimedia communications over Local Area Networks (LANs) that do not provide a guaranteed Quality of Service (QoS).

Major System Component for Voice over IP system of H.323

- Terminal
  The figure below describe the feature that the H.323 must have
H.323 call setup/release module performs signaling for establishment and termination of call connection based on Q.931 protocol. After call connection, H.245 protocol starts a capability negotiation between two endpoints such as voice compression algorithm to be used, and proceeds conferencing request for transfer of voice traffic. RAS (Registration, Admission, Status) is the function that registers its transport address such as IP and gets an admission from the gatekeeper when a user invokes a new call. RTP module transports a real-time media such as video, audio over a packet network. G.729/G.723/G.711 are the internationally accepted compression standards. T.120 is for data conferencing. It enables products from different vendors to interoperate without terminals assuming a prior knowledge of the other systems. It specifies the network interfaces and wire formats, along with data transmission facility.

- **Gateways**
  The gateways are the devices that communicate between the telephone signals and the IP endpoint. The IP endpoint usually speaks H.323 for media stream and more recently Session Initiation protocol (SIP). The gateways usually perform the following 6 functions: Search function, Connection Function, Digitizing function, Demodulation functions, Compression functions, Decompression and Remodulation functions.

- **Gatekeepers**
  Terminals are the LAN client endpoints that provide real-time two-way communications. When an endpoint is switched on, it performs a multicast discovery for a gatekeeper and registers with it. Thus the gatekeeper...
knows how many users are connected and where they are located. The collection of a gatekeeper and its registered endpoints is called as a **zone**. A gatekeeper is required to perform the following 4 functions: Address translation, Admissions control Bandwidth management, and Zone management

- **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 MP under the control of the MC handles the multiplexing of media streams. The following figure shows the interaction between all the H.323 components

---

**H.323 Protocol Stack**

The following figure 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. H.323 uses both TCP (reliable) and UDP (unreliable) communications. This is because the control signals and data require reliable transport since the signals must be received in the order in which they were sent and cannot be lost. On the other hand, audio and video streams lose their value with time. If a packet is delayed, it may not have relevance to the end user. So audio and video signals use the more efficient unreliable transport since it has a smaller delay. The reason why the TCP is called "Reliable" is
because it guarantees sequenced, error-free, flow-controlled transmission of packets. But since TCP has a flow control, the delay of the packet can occur since TCP will reduce the rate of transmit when congestion occurs. Unreliable transmission on the other hand doesn’t provide any flow control. Thus the delay is small compare to TCP.

<table>
<thead>
<tr>
<th>Data</th>
<th>Control and Signaling</th>
<th>Audio/Video</th>
<th>Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>T.120</td>
<td>H.225.0 Call Signaling</td>
<td>H.245 Conference Control</td>
<td>RTP/RTCP</td>
</tr>
<tr>
<td>TCP</td>
<td></td>
<td></td>
<td>UDP</td>
</tr>
</tbody>
</table>

The RAS channel is used for the communication between the endpoints and the gatekeeper. When the call is establish, the RAS channel multicast a Gatekeeper Request (GRQ) to find a gatekeeper and register with it. The gatekeeper will send either a Gatekeeper Reject (GRJ) when it does not want the endpoint to register with it or Gatekeeper Confirmation (GCF) message to indicate the willingness to be the gatekeeper for that endpoint. 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. After the gatekeeper is found, the endpoint will inform the gatekeeper of its transport and alias addresses and send Registration Request (RRQ) to a gatekeeper. After register, the endpoint or gatekeeper that has an alias address for an endpoint issue a Location request (LRQ) message to find the endpoint contact location. The gatekeeper with whom 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. 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.
H.225.0, the call signaling channel, is used when the network contains no gatekeeper to carry the control message. Call signaling messages are passed directly between the calling and called endpoints using the Call Signaling Transport Addresses. Thus the calling and called endpoints must know each other Call Signaling Transport Address in order to communicate directly.

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.

RTP and RTCP detail are described later in this survey.

For more specific information, see [4], [5]

**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. SIP is used to initiate a session between users. It provides user location services, call establishment and call participant management.

*Major System Component for Voice over IP system of SIP*

- **User Agents**
  A user agent is an end system acting on behalf of a user. A user agent contains 2 components, a *user agent client* (UAC) and a *user agent server* (UAS). The UAC task is to initiate the calls i.e. sending a SIP request while the UAS answering call i.e. sending a response.

- **Network Server**
  There are 3 types of the server. A *registration server* receives updates on the current locations of the users. A *proxy server* receives requests and forwards them to another server called next-hop server, which has more precise location information about the callee. The next-hop server could be another proxy server, a UAS, or a redirect server. A *redirect server* also receives requests and determines a next-hop server. But it returns the address of the next-hop server to the client instead of forwarding the request there as proxy server.

- **Internet Telephony Gateway**
  SIP allows a user on the Internet to call another user on the Internet. But if an Internet user wishes to call someone on that is not on the Internet but rather on the phone, the Internet Gateway Telephony must be used. This device is capable of converting signal and media between packet network and a telephone network (PSTN).
**SDP**

SDP is used to describe multimedia sessions for both telephony and distribution application like Internet radio. The protocol contain information about

- Media streams – It define a stream type i.e. Audio, video, data, control, application.
- Addresses- Indicate a destination address (unicast/multicast) for each stream.
- Ports- Indicate the UDP port number for sending and/or receiving for each stream.
- Payload types- the media formats which can be used during the session is also conveyed. For unicast sessions (traditional IP telephony), this is called a capability set.
- Start and stop times- Indicate the start, stop and repeat times of the session for broadcast-style session like a television program.
- Originator- Provide the session description names, the originator of the session and the way to contact that person for broadcast-style session.

*How SIP and SDP work correlated to set up a call?*

SIP defines several methods such as INVITE to invite a user to a call, BYE to terminate a connection between 2 users in a call, OPTION solicits information about capabilities but doesn’t set up a call, ACK is used to acknowledge the receive of invitation, CANCLE terminate search for user, REGISTER provide information about a user’s location to a SIP registration server.

A client sets up a call by issuing an INVITE request. This request contain a header file which provide information about the call such as callee’s (TO) and caller’s (FROM) address, subject. The invite message also contain the SDP body describe the session parameter acceptable to caller.

*An example of SIP transaction*
Figure 4 show how the SIP transaction flow. SIP user agent creates an INVITE request for sip: anne@honeywell.com. This request is forwarded to a local proxy (1). This proxy looks up Honeywell.com on the Domain Name Service (DNS) and obtains the IP address of the server handling SIP requests for this domain. It then forward request to that server (2). That server contains the information about user anne, but this user is currently logged in as a.user@university.edu. Therefore the user redirects this information to the requested proxy (3) to try this address. The local proxy looks up university.edu in DNS, and obtain the IP address of that SIP server. The request is then proxy there (4). The university server consults a local database (5), which indicates (6) that a.user@university.edu is known locally as a.pratoomt@cs.university.edu. The main university server proxies the request to the computer science server (7). This server know the IP address where the user is currently logged in, so it proxies the request there (8). The user accepts the call, and the response is returned through the proxy chain (9), (10), (11), (12) to the caller.

Other 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 [11]. The following figure (Fig 5) shows the signaling and transport protocols required for delivering voice over IP [10].
As described earlier, the VOIP packet consists of data, control and signaling, and video/audio. The audio/video packet relies on UDP transport protocol for the small delay and steady throughput rate while the data, control and signaling relies on TCP for the reliability of the packet. Either RTP or RSVP is used to transport the audio/video packet. RTCP is a control protocol, which work conjunction with RTP. H.323 controlling protocols are H.245, H.225.0/H.225.0 RAS that has been described in earlier section.

**RTP**

Real Time Transport Protocol (RTP) supports the transport of real-time media such as audio and video over packet network. RTP plays a key role in an Internet telephony system. It moves the actual voice among participants. Both SIP and H.323 using RTP as their transport protocol. RTP header contains information that helps the receiver to reconstruct the media and also contains information specifying how the codec bit streams are broken up into packets. RTP provides the following functions:

- Sequencing- Each RTP packet contains a sequence number. This can be used for loss detection and packet reordering.
- Payload Identification- In the Internet, network conditions such as packet loss or delay vary. So it’s necessary to be able to change the encoding of the media -the payload of RTP dynamically as the network vary. 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 the end of the frame, a frame marker bit stream has been provided.
- Source Identification- In a multicast session, many users are participating. An identifier called the synchronization source (SSRC) specifies the originator of the frame i.e. identify which participant sent the packet.
- Intermediate Synchronization- Packet within the same stream may suffer with the difference of the delays (jitter). Application use playout buffers [7-9] to compensate for delay jitter. RTP provides a time stamp, which are needed by the playout buffers.

**RTCP**

Real Time Control Protocol (RTCP) is a control protocol, which work conjunction with RTP. In RTP session, participants periodically send RTCP packets to obtain useful information. This information that RTCP provides to the participant are:

- QoS feedback- RTCP is used to report the quality of service. The information are the number of packet lost, Round Trip Time, jitter. This information is used by the source to adjust their rate.
- Session Control- RTCP allows the participant to indicate that they are leaving a session by using a BYE RTCP packet. Participant can also send a small notes accompany with that such as leaving out of the office.
• Identification- RTCP packet contain the information about the participant such as email address, name, phone number so that all the users can know the identities of the other users in the session.
• Intermediate Synchronization- Even though the audio and video are normally sent over different stream, in order to play them together at the receiver end, synchronization between them is required. RTCP provides the information that is required for synchronization the streams.

**RTSP**

Real Time Stream Protocol (RTSP) is used to control a stored media server. A stored media server is a device capable of playing prerecorded media from the disk to the network, and recording multimedia content to disk; RTSP offers control similar to a VCR remote control. For further information, see [3], [4].

**RSVP**

Resource Reservation Protocol (RSVP) can prioritize and guarantee latency to the specific IP traffic streams. So with RSVP, the voice communication with tolerable delay on a data network can be accomplish. RSVP makes receivers responsible for requesting a specify QoS such as the maximum delay that the receiver can tolerate. Thus the RSVP provides one of the solutions to improve the QoS.

**SAP**

This protocol is used for advertising the multicast session. For more information, see [6]

**MGCP**

This protocol allows the controller and signaling gateway to communicate i.e. the communication between call control element (call agent) and PSTN. For more information, see [12]

**GLP**

The Gateway Location Protocol is use by SIP in the Internet Telephony Gateway. To complete a call to a PSTN endpoint, an IP host must send a SIP invitation to the gateway. The GLP will provide the location of the gateway to complete the call, which minimized the PSTN leg to the call. For more information, see [3].
An example of how each protocol plays a role in an IP telephony call

John is sitting at his computer and his machine has a sounds “ring” and followed by speaker} Audio and video call from Joe. He accepts a call and talks for a while. He then decides that Alice needs to be in on the call too. He says ”Add Alice”, and the speech recognition software on his PC interprets the command. His client application consults a local Internet white pages directory and adds alice@ieee.org to the call. The call set up request reach Alice’s personal agent software. The agent has been instructed to ring her cell phone, home PC, and work PC all in parallel. To complete the call to the cell phone, Alice has instructed her agent to use the cheapest gateway that support credit billing. The agent find the appropriate gateway, and ring Alice at her cell phone, home and work PC all at once. Alice pick up in her car, joining the call voice only. During the conversation, Joe remembers that a video segment from a recent IEEE tutorial presentation would be helpful. He finds the media server with the content, and play the media stream into the conversation. Later, John decides to leave the call. He transfers Joe and Alice to the web page containing the addition information on the IEEE tutorial. Joe’s web browser jumps to the page and Alice’s phone display the text-only version, which they continue to discuss. Joe decides to add Bob to the call but Bob is not available. Bob agent return a web page containing his appointments, along with the hyperlink to a voice mail service.

The services contained in the call scenario above required many protocols in order to work. First the SIP, SDP are needed to allow Joe to call John, and to establish a multimedia session so that they can exchange audio and video. The actual audio and video are exchanged between session participants using RTP. The directory access protocols that use to access the white pages are the same as email service. These protocols are beyond the scope of this survey so they will not be discussed here. For more information, refer to [23]. During the call, Joe brings the media server and instructs it to play the video segment. This is accomplished by using RTSP. The intelligent agent concept described above interacts with the signaling protocol to provide advanced service. To realize this agent, a calling processing language is used. For more information about the calling processing language, see [24], [25]. The GLP helps the agent to select a gateway for terminating the call from the Internet on the telephone network.

**A new Reliable Signaling Transport Protocol (RSTP) VS TCP/UDP**

IP network based on best effort thus provide no guarantee of QoS. However call signaling usually demands high transport quality, including low delay and packet loss rate. UDP is used as the underlying protocol for the represented RSTP. The RSTP provide several functions such as Sequence Integrity Management, which is used to synchronize the sequence number in two peer, Connection management, Flow control, and Error control, Security Function. Hence the RSTP has many advantages over the existing mechanisms, such as low delay, high reliability and efficiency, and less implement complexity. But however the more efficient algorithm and simulation are needed to verify this claim.
TCP has a flow and error control but TCP required more OS resources used to manage them. The delay introduced by retransmission mechanism of TCP may also bring long delay. On the other hand UDP doesn’t provide any loss protection to the message transported. Even though ARQ enhancements allow for retransmission of lost or corrupted packets, it required at least additional 1.5 round trip time. Recently, version of UDP are being discussed which are supposed to provide guarantee against loss of the data, such as Reliable UDP (RUDP). However, the proposal is in working progress. Whether it will allow us to overcome all the problem of loss and delay and provide a protocol to meet the voice over IP signaling requirement is not yet to proved.

Normally, the voice packet is constructed as a UDP/IP packet, to avoid TCP/IP’s attempt to retransmit the corrupted packet. However TCP could be a better alternative for Fax transmission. This is because if lost packets occur during the negotiation of a page, the fax could be terminated. When TCP/IP is used and the host software hides the retransmission from the fax machine, there will be no impact.

Comparison of H.323 with SIP

H.323 was designed base on the Asynchronous Transfer Mode (ATM) and Integrated Service Digital Network (ISDN) signaling. So H.323 is complex and has overheads. SIP, on the other hand was design base on the Internet, so it is avoid the complexity and extensibility. SIP inherits most of the format of HTTP such as header field, encoding rules, error codes. But the H.323 has an advantage because of it share a larger chunk of the market. SIP is a newer standard so it’s still has a lesser share in a market. But one advantage of SIP is it’s backed up by IETF.

Current problem of the Voice over IP

- Standard – The biggest difficulties that Voice Over IP technology is facing today is the interoperability between Internet telephone products and internetworking with legacy PSTN-based systems and services. Currently, no two products are compatible. At this time the specific issue that have to be resolved are related to the codec format, the transport protocol, and directory services.
- Quality- The latency of the voice is big on the Internet. So the quality of Internet telephony is still not acceptable.
- Capacity- The Internet is an open network of many different ISP’s network; consequently, there is no way to get the guarantee of the network bandwidth. Packet sequence, and latency. Due to the increasing popularity of the Internet, the load on the Internet is increasing. This cause the network congestion which cause the packet loss which is the main parameter affecting the QoS. Currently the Internet telephony application repairs the lost packets with silence, which leads to the serious impact on the intelligibility of the speech.
- Social issue- in most country, governments or government-sanctioned entities retains monopolies for provisioning telephone service. Thus some of the
telephone service of the under development countries still lower than normal standard. This is a big obstacle for the growth of Internet telephone technology for these countries.

Some solutions to the current problem

- **Standard-** The ITU-T recommends H.323 to defined the core technology for Voice On the Net applications. And the IETF is also introducing SIP as a core standard as H.323. The Lucent Softswitch is also one of the interesting products that can help solve some of the incompatibilities of the system.
- **Quality-** There are some ways to improve the QoS such as using RSVP as a transport protocol, dedicated service lines with manage traffic loads, co-locating telephone access with backbone nodes, bigger routers, new network architecture.
- **Capacity-** Delay increase with each router hop. So one solution is to increasing router speed and reduces number of routers hop. Use a bigger router is one of the solutions since the big router can handle at least 10 times more traffic than a conventional router.
- **Social issue-** People are always embrace a new technology so nothing can prevent the growth of technology even though there’re some social issues.

Product and service supporting Voice Over IP

**Products**

- Gatekeepers- Ericsson H.323 gatekeeper, VocalTec gatekeeper, Nortel networks’ IP Connect, Elemedia H.323 gatekeeper GK2000S
- IP telephones- CISCO’s IP Phones, Selsius IP phone, Nokia Systems’IP Courier
- PC based Software Phones- VocalTec IPPhone v.501, Netscape’s Cooltalk, White Pine’s CU-SeeMe Pro, Microsoft Net Meeting

**Services**

- Pc to phone Services- VocalTec Surf&Call, Dialpad.com
- Pc to Pc services- Microsoft NetMeeting, VocalTec Iphone, TaoTalk.com
- Phone to Phone services- AT&T, America On-Line, IDT Corporation, USATEL VIA ONE Prepaid Calling Card
- Network services- Level 3’s IP Crossroad Service, QWest
- Services for the service providers- ITXC, IP Telephone for carriers by Delta Three and Ericsson, Cisco and VocalTec to jointly provide hybrid end-to-end services to carriers and service providers, Cisco AVVID
Motorola Vanguard Series
This is an award winning expandable network access and concentration platform that integrates LAN, analog/digital voice and future multimedia traffic.

Lucent Softswitch
The Lucent technologies Softswitch is a programmable, multi-protocol software system that allows communication between different signaling system such as SS7, SIP, H.323 and Q.931. It thus serves as a mediator between telephony and IP connectivity. It also solves inter operability problems between gateways from different vendors caused by signaling and protocol incompatibilities.

For more specific information on this section, see [14]

Conclusion
The Internet telephony caught the world’s attention since it provide and economical way of communication and it provide a varieties of function of communication that a ordinary telephone can’t provide such as it support Multicasting service, it integrate Web/email/presence and instant messaging application with the telephone, Multiple user session which allow more than 2 users to communicate, Ability to play a media stream into a conversation. There’s a lot of standard, protocol, and product to support this new technology. There’s some organization that try to develop and come up with a solution/standard to solve/improve this new service. Thus the voice over IP is a new technology that will develop and growing for more years to come.

Reference:
[2] Voice over IP, presentation slides. Raj Jain, Professor of computer and Information Science, OSU.
http://www.cs.columbia.edu/~hgs/papers/Schu9902_Internet.ps.gz
http://computer.org/internet/ic1999/w3073abs.htm
[12] "Media Gateway Control Protocol (MCGP)", August 1999, 111 pages,
ftp://www.ietf.org/internet-drafts/draft-huitema-megaco-mgcp-v0r1-05.txt
[13] DataBeam,” “A primer on the H.323 series standard”, 20 pages,
http://www.databeam.com/h323/h323primer.html
[22] “Quality Of Service for Voice over IP”, Barry Keepence, 3Com