INTERNATIONAL TELECOMMUNICATION UNION

Telecommunication Development Bureau (BDT)

2nd Experts Group Meeting on Opinion D Part 3 (ITU-D)
Geneva, 8 - 10 October 2001

Document IP Tel/9-E
8 – 10 October 2001
Original: English

IP TELEPHONY

TECHNICAL ISSUES

Dr. Charles M. Sarraf

Ericsson Lebanon Communications

WORKING DEFINITION OF "IP TELEPHONY"

IP Telephony is defined here as any telephony application that can be enabled across a packet-switched data network via the Internet Protocol (IP). An IP telephone is a telephone device that transports voice over a network using data packets instead of circuit switched connections over voice only networks.

IP Telephony refers to the transfer of Voice over the Internet Protocol (VoIP). Other Voice Over Packet (VOP) standards exist for Frame Relay and ATM networks.

In an IP Telephony connection, the voice signal is digitized, compressed and converted into IP packets, which are transmitted over the IP network, such as the Internet, Intranet and Local Area Networks (LANs), and shared with other IP traffic.

NETWORK ARCHITECTURES

IP Telephony Network

IP Telephones originally existed in the form of client software running on multimedia PCs for low-cost PC-to-PC communications over the Internet. Quality of Services (QoS) problems associated with the Internet and the PC platform itself resulted in poor voice quality due to excessive delay, variable delay, and network congestion resulting in lost packets, thus relegating VoIP primarily to hobby status.

A normal telephone call (Figure 1) is connected through an end-to-end circuit with a fixed bandwidth. With IP Telephony, a packet-based network is used where a number of calls plus data share the same network link.

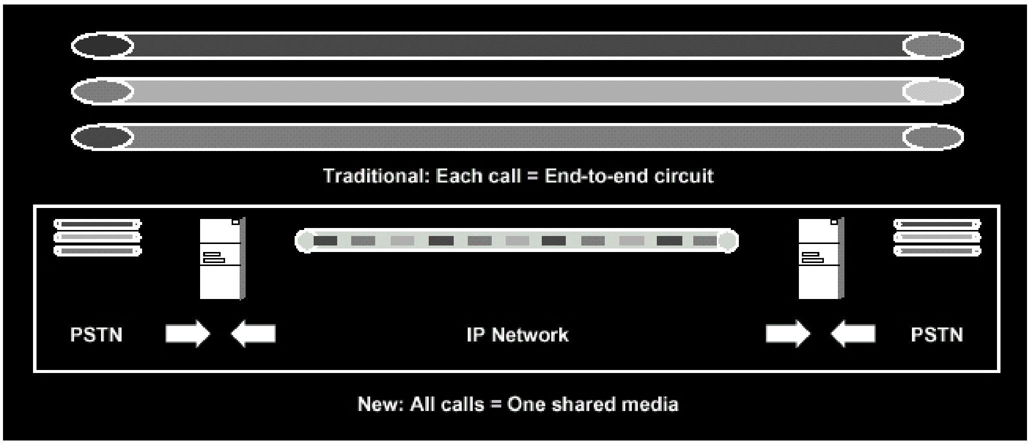
Traditional: Each call = End-to-end circuit
PSTN
IP Network
PSTN
New: All calls = One shared media

Figure 1

The data networking technology moves information at a much lower cost by making better use of the network capacity. Not only is a packet-based shared network more efficient than a fixed 64 kbits/s circuit switched connection, but it also compresses the voice signal.

Figure 2 illustrates a very simple protocol stack with the user application, such as IP Phone, is interfacing the IP-Telephony-signalling stack, which uses IP as an end-to-end protocol. The user application can be either a vendor-specific Voice Gateway implementation or PC client interface. The IP Telephony stack is the protocol handling communication between the two applications and the control entity.

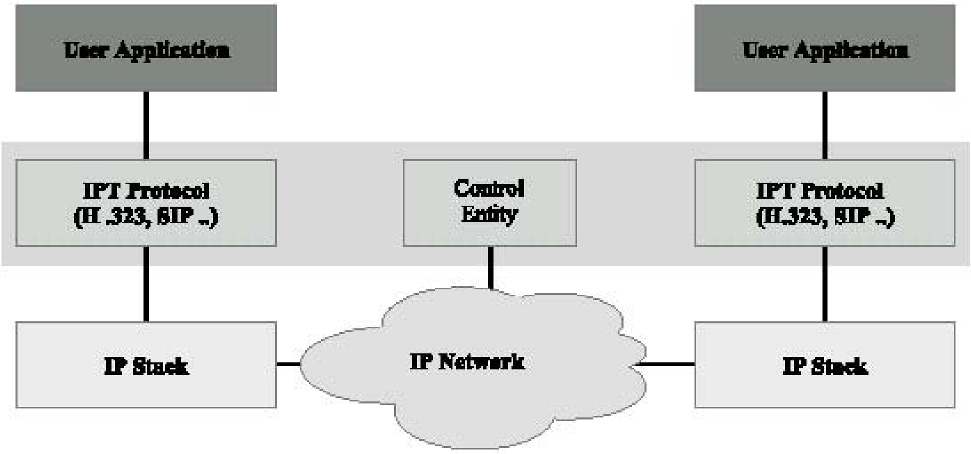

Figure 2

IP Telephony equipment, which can be categorized into client, access/gateway, and carrier infrastructure segments, should be configurable to capitalize on these different techniques but must also be sufficiently flexible to add new techniques as they become available. Gateways are the points where the IP network is interconnected with the switched network such as PSTN as demonstrated in Figure 3.

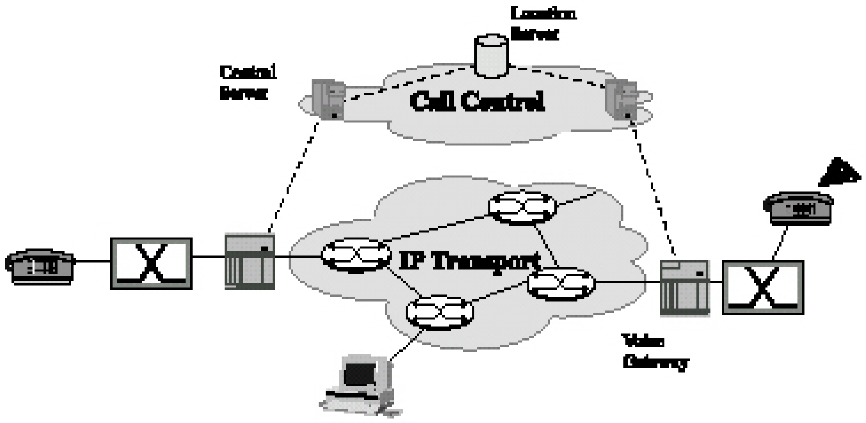

Figure 3.

Real-time voice traffic can be carried over IP networks in three different ways:

1. PC-to-PC Voice
2. PC-to-Phone
3. Phone-to-Phone Communication

PC-to-PC Voice

It can be provided for multimedia PCs (i.e. PCs with microphone and sound system) operating over an IP-based network without any interconnection with switched voice network such as PSTN or Mobile (Figure 4). PC applications and IP-enabled telephones can communicate using point-to-point or multi-point sessions.

PC-to-Phone

It can replace the analog or digital circuits that are serving as voice trunks (such as private links between company owned PBXs) or PSTN access trunks (links between a PBX and the carrier). A gateway serves as the inter-connecting point between the two networks.

Phone-to-Phone Communications

It appears like a normal telephone to the caller nut may actually consists of various forms of voice over packet network, all interconnected to the PSTN. Gateway functionality is required when interconnecting to circuit switched voice networks, such as PSTN, ISDN or Mobile Systems.

Network Capacities

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.

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

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 report. 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.

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.

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. Here are some of its functionalities:

- *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 necessary to indicate that a called terminal is busy, and to provide information for the Bandwidth Management function.

With H323, [Figure 4], the Gatekeeper provides the control functionality, but in an optional manner. The Gatekeeper must handle the registration and access control, but call setup and media negotiating do not have to go through the Gatekeeper.

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.

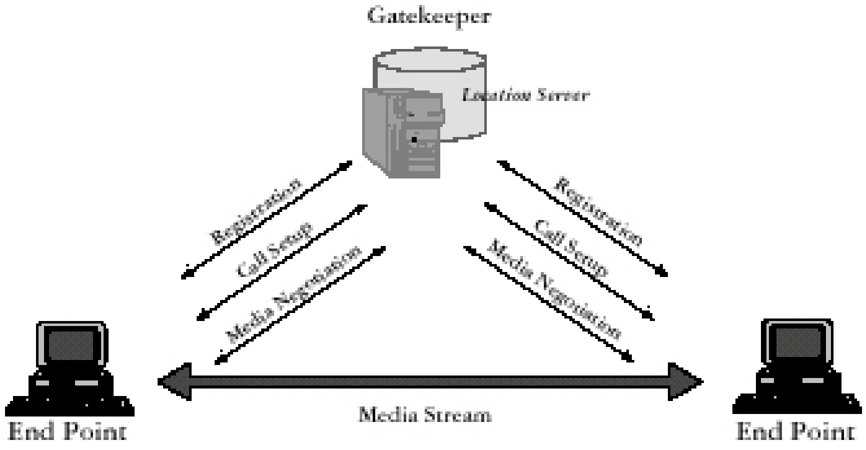

Figure 4

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.

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:

- *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.

Q.931 signalling

The Q.931 channel is a Transmission Control Protocol (TCP) based call protocol that is used for call setup and call release. The protocol is based on Integrated Service Digital Network (ISDN) Q.931, which is a well-proven protocol for this type of connection-oriented communication. It provides capabilities for handling a variety of supplementary services related to specific connections or users ans enables interworking with the SCN.

H.245 Media and Conference

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:

- *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.

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.

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 provides include [RFC2543].

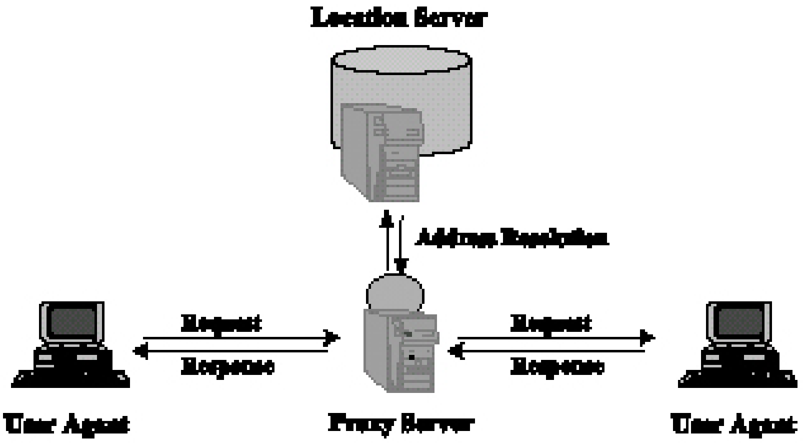

Figure 5. The SIP proxy mode of operation in which the proxy server acts as a call control point.

- 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.

The SIP System consists of two agents:

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.

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.

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.

The addressing used in SIP is based on a SIP Uniform Resource Locator (URL), which can take the form of either user@domain or user@IP address. The intention is to integrate SIP services into existing manner of services operations on the Internet.

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.

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).

Session Announcement Protocol (SAP)

SAP is often mentioned in connection with SIP, because both protocols are used to communicate the presence of conference sessions.

SAP is an announcement protocol for multicast conference sessions. A SAP client that announces a conference session periodically multicast address and port. The announcement is multicast with the same scope as the session it is announcing. This ensures that the recipients of the session also can be potential recipients of the session the announcement describes (bandwidth and other such constrains permitting). This type of operation makes SAP of little use as telephony call setup protocol; it is more useful for the announcement of a standard-type multicast session.

QUALITY OF SERVICE ISSUES

Voice Quality

Voice quality has evolved over the years to be consistently high and predictable, it is now an important differentiating factor for new VoP (Voice-over-Packet) networks and equipment. Consequently, measuring voice quality in a relatively inexpensive, reliable, and objective way becomes very important.

Voice quality means different things depending on your perspective. On the one hand, voice quality is a way to describe and evaluate speech fidelity, intelligibility, and the characteristics of the analog voice signal itself. On the other hand, voice quality can describe the performance of the underlying transport mechanisms.

Why is Voice Quality an Issue?

What users and their organizations expect from IP Telephony and VoIP is essentially PSTN quality and objective verification that they are receiving it.

Traditional circuit switched networks, such as PSTN, have long addressed the voice-quality problem by optimizing their circuit for the dynamic range of the human voice and the rhythms of human conversation. PSTNs have evolved to provide an optimal service for time-sensitive voice applications that require low delay, low jitter and constant and guaranteed bandwidth.

Unlike PSTNs, IP networks, which make a “best effort” to deliver packets, originally aimed to support non-real-time applications, such as file transfers and e-mail. These applications feature bursty traffic and sometimes-high-bandwidth demand but are not sensitive to delay or delay variation.

The absence of solutions ensuring Quality of Service (QoS) has been a deterrent to the widespread adoption of Voice over Internet Protocol (VoIP). Potential users often think that speech quality won’t be as good as what they are accustomed to the familiar public switched telephone network (PSTN).

Voice quality is subjective because it’s a measure of the intelligibility by the listener. However, perceptions drive decisions. VoIP service providers must extremely sensitive to the perceptions of their customers, because a decision to change service can be precipitated from such negative perceptions as

- When a user perceives unacceptable instantaneous quality, the user is likely to terminate the call prematurely.
- If a user perceive overall poor quality after completing a call, there is likely to be a harboring of residual dissatisfaction.

- If service providers achieve quality by over-provisioning their networks, the resulting high costs undermine the user's perception of value, despite excellent voice quality.

MOS Ratings

Real-Time Bandwidth

Many networks designs do not meet the real-time bandwidth requirements of speech. Data networks typically have been designed to make a "best effort" to deliver packets to destinations. Introducing voice signals into those networks requires methods that ensure this real-time transport, but voice quality can still suffer if these methods do not work properly.

Although real-time speech has a reasonably low-bandwidth requirement, it needs either a constant available bandwidth for linear CODECs or direct available bandwidth for low-bit-rate CODECs.

Delay

User-perceived delay can result from the time it takes for the system or network to digitize, packetize, transmit, route, and buffer a voice signal. This delay can interface with normal conversations and exacerbate network problems such as echo and talker overlap.

Echo is caused by the reflection of the transmitted signal, either acoustical or electrical in nature, back to the sender. It appears because of a reflection either in the network or in the terminal equipment. From telephony perspective, echo is the sound of the talker's voice returning to the talker's ear via the telephone's speaker. In other words, echo occurs when the talker's voice signal "leaks" from the transmitting path back into the receiving path. If the time between the original spoken phrase and the returning echo is 25 to 30 msec or if the echo's level is approximately –25 dB, no annoyance or disruption to voice conversations will probably occur. When the echo is loud enough and has enough delay – usually around 30 msec and more – so that the speaker perceives the echo, the quality of a voice call becomes problematic.

Talker overlap (or the problem of one talker stepping on the other talker's speech) becomes significant if the one-way delay becomes greater than 250 msec. The end-to-end delay budget is therefore the major constraint and driving requirement for reducing delay through a packet network.

Following are sources of delay in an end-to-end Voice over Packet call:

1. **Accumulating Delay (sometimes called algorithmic delay)**: This delay is caused by the need to collect a frame of voice samples to be processed by a single coder. It is related to the type of voice coder used.

2. **Processing Delay**: This delay is caused by the actual process of encoding and collecting the encoded samples into packet for transmission over the pocket network. The encoding delay is a function of both the processor execution time and the type of algorithm used.

3. **Network Delay**: This delay is caused by the physical medium and protocols used to transmit the voice data, and by the buffers used to remove packet jitter on the receive side. This delay is directly related to switching capacity of the packet network, which is a function of the capacity of the links and processing capabilities of nodes that route and transmit the packets.

   The jitter buffers add delay which is used to remove the packet delay variation that each packet is subjected to as it transits the packet network. This delay can be significant part of the overall delay since packet delay variations can be as high as 70-100 msec in some Frame Relay networks and IP networks.

4. **Jitter**: The delay problem is compounded by the need to remove jitter, a variable interpacket timing caused by the network a packet traverses. Removing jitter requires collecting packets and holding them long enough to allow the slowest packets to arrive in time to be played in the correct sequence. This cause additional delay.

   The two conflicting goals of minimizing delay and removing jitter have engendered various schemes to adapt the jitter buffer size to match the time-varying requirements of network jitter removal. This adaptation has the explicit goal of minimizing the size and delay of the jitter buffer, while at the same time preventing buffer underflow caused by jitter.

Packet Loss

Since IP – Internet Protocol provides unacknowledged connectionless services, it does not guarantee packets delivery, all voice packets are treated like data. Under peak loads and congestion, voice packets will be dropped equally with data packets. The data packets, however, are not time sensitive, and dropped packets can be appropriately corrected through the process of retransmission. Lost voice packets, however, cannot be dealt with in this manner.

Techniques to Improve Voice Quality

Since Voice Quality is affected by many factors clarity, end-to-end delay, and echo, many techniques are introduced to enhance the quality of service in a VoIP infrastructure, such as congestion management for packet loss improvement, admission control (RTP and RSVP) for jitter improvement, and echo cancellation for echo improvement.

Congestion Management

Congestion management features allow controlling congestion into a packet networks though congestion avoidance or congestion recovery techniques. The congestion management QoS feature offer four types of queuing protocols, each of which allows you to specify creation of a different number of queues, affording greater or lesser degrees of differentiation of traffic, and to specify the order in which that traffic is sent.

The four type of queuing, which constitute the congestion management QoS feature, are:

- **First-In, First-Out Queuing (FIFO):** FIFO entails no concept of priority or classes of traffic. With FIFO, transmission of packets out the interface occurs in the order the packets arrive.

- **Weighted Fair Queuing (WFQ):** WFQ offers dynamic, fair queuing that divides bandwidth across queues of traffic based on weights. WFQ ensures that all traffic is treated fairly, given its weight. WFQ ensures satisfactory response time to critical applications, such as interactive, transaction-based applications, that are intolerant of performance degradation.

- **Custom Queuing (CQ):** With CQ, bandwidth is allocated proportionally for each different class of traffic. CQ allows you to specify the number of bytes or packets to be drawn the queue, which is especially useful on slow interface.

- **Priority Queuing (PQ):** With PQ, packets belonging to one priority class of traffic are sent before all lower priority traffic to ensure timely delivery of those packets.

Lost-Packet Compensation

Some schemes used by Voice-over-Packet software to address the problem of lost packets are as follows:

- Interpolate for lost speech packets by replaying the last packet received during the interval when the lost packet was supposed to be played out; this scheme is a simple method that fills the time between noncontiguous speech frames. It works well when the incidence of lost frames is infrequent; it does not work well if there are a number of lost packets in a row or a burst of lost packets.

- Send redundant information at the expense of bandwidth utilization; this basic approach replicates and sends the n^{th} packet of voice information along with $(n+1)^{th}$ packet. This method has the advantage of being able to correct for the lost packet exactly; however, this approach uses more bandwidth and also creates greater delay.

- Use a hybrid approach with a much lower bandwidth voice coder to provide redundant information carried along in the (n+1)th packet; this reduces the problem of the extra bandwidth required but fails to solve the problem of delay.

Echo Cancellation

The echo canceller compares the voice data received from the packet network with voice data being transmitted to the packet network. The echo from the telephone network hybrid is removed by a digital filter on the transmit path into the packet. ITU standards G.165 and G.168 define performance requirements for echo cancellation.

Resource Reservation Protocol (RSVP)

RSVP has been developed by IETF (RFC 2205) in an attempt to solve the problems occurred from the network delay. 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 RSVP enabled, we can accomplish voice communication with tolerate delay on data network.

The RSVP requests will generally result in resources being reserved in each node along the data path. A common misunderstanding is that RSVP alone will give better quality of service. RSVP is a control protocol that sets up a reservation, but enforcement of the reservation needs to be done by another component of the architecture.

Real-time Transport Protocol (RTP)

The RSVP provides a method for establishing more reliable transmission through a routed network rather than normal best-effort routing. Although this can enhance the quality of service the network provides, it cannot completely eliminate queuing-introduced jitter.

The Real-time Transport Protocol compensates for the jitter by a “resynchronization” in the endpoints. RTP provides an end-to-end delivery service to support applications transmitting real-time data, such as interactive audio and video, over unicast and multicast networks services. The RTP does not specify any mechanisms to ensure timely delivery or other quality-of-services guarantees, but rather relies on the lower layers to provide that service. RTP is defined in IETF RFC 1889.

The RTP provides end-to-end delivery services, but it does not offer all of the functionality that is typically provided by a transport protocol, such as error recovery or flow/congestion control. In fact, the RTP typical runs on top of UDP to utilize its multiplexing other checksum services. Transport protocols other than UDP can carry RTP as well.

Delivery of the RTP packets is monitored by means of a control protocol that provides feedback to the source as well as the session participants. The RTP Control Protocol (RTCP) is based on the periodic transmission of control packets to all participants in the session, using the same distribution mechanism as the data packets in RTP.

The primary function of the RTCP is to provide feedback to an application regarding the quality of the data at the reception. The statistics provided by the RTCP, from both the sender and the receiver, include number of packets sent, number of packets lost, inter-arrival jitter, delay, etc.

This information can be used by the sender application to modify the transmission, for example, to another compression rate, in an attempt to improve the quality. This serves as an integral part of RTP's role as a transport protocol and is related to the flow and congestion control functions of other transport protocols.

The RTCP also includes identification of the users in the session, which may be used to control the participants in the session.

Real-Time Streaming Protocol (RTSP)

The RTSP 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: (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 sends 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.

SECURITY ISSUES

A network is vulnerable at any point where it contacts other networks or systems. These include internetworking devices like bridges, routers, and modems.

Generally IP networks are not secure, data are transmitted over the IP protocol are not encrypted and be read and monitored by a third party other than the destination host.

To provide a secure conversation on the LAN, many algorithms and protocols have been involved. Such algorithms are cryptography's algorithms. In addition to the Virtual Private Network (VPN) or any type of Virtual Local Area Network (VLAN), the IPSec protocol hepls us to achieve a secure network while being configured.

Cryptography's Algorithms

Data that can be read and understood without any special measures is called *plaintext* or *cleartext*. The method of disguising plaintext in such a way as to hide its substance is called '*encryption'*. Encrypting plaintext results in unreadable gibberish called *ciphertext*. Encryption can be used to ensure that information is hidden from anyone for whom it is not intended, even those who can see the encrypted data. The process of reverting ciphertext to its original plaintext is called '*decryption'*.

Virtual Private Network (VPN)

A virtual Private Network (VPN) is the extension of a private network that encompasses links across shared or public networks like the Internet. A VPN enables to send data between two computers across a shared or public internetwork in a manner that emulates the properties of a point-to-point private link. To emulate a point-to-point link, data is encapsulated or wrapped, with a header that provides routing or tunneling information allowing it to traverse the shared or public internetwork to reach its end point, similar to a tunnel. To emulate a private link, the data being sent is encrypted for confidentiality. Packets that are intercepted on the shared or public network are indecipherable without the encryption keys. The link in which the private data is encapsulated and encrypted is known as a Virtual Private Network connection.

Virtual Local Area Network (VLAN)

Virtual LANs (VLANs) can be viewed as a group of devices on different physical LAN segments, which can communicate with each other as if they were on the same physical LAN segment. Using VLANs has many benefits, increasing performance, improving manageability, simplifying software configuration, and increasing security options that is our issue.

VLANs have the ability to provide additional security not available in a shared media network environment. By nature, a switched network delivers frames only to the intended recipients, and broadcast frames only to other members of the VLAN. This allows the network administrator to segment users requiring access to sensitive information into separate VLANs from the rest of the general user community regardless of physical location.

IPSec

IPSec is a framework of open standards for ensuring secure private communications over IP networks. Based on standards developed by IETF, IPSec ensures confidentiality, integrity, and authenticity of data communications across a public IP network. IPSec provides a necessary component of a standards-based, flexible solution for deploying a network-wide security policy.

IPSec implements network layer (IP) encryption and authentication, providing an end-to-end security solution in the network architecture itself. Thus end systems and applications do not need any changes to have the advantages of strong security. Because the encrypted packets look like ordinary IP packets, they can be easily routed through any IP network, such as the Internet, without any changes to the intermediate networking equipment. The only devices that know about encryption are the end points. This feature greatly reduces both implementation and management costs.

INTEROPERABILITY ISSUES

Bandwidth

Bandwidth efficiency has been one of the early drivers of the evolution from circuit-switched to packet switched technology. That’s because packet-based IP networks use sophisticated compression algorithms that enable better utilization of existing digital data lines and maximize network capacity. For example a typical voice call over E1/T1 channel in the PSTN will take a full 64 Kbps bandwidth, while requirements for good quality VoIP transmission uses 6 Kbps to 13 Kbps, resulting in about 10 times the bandwidth capacity.

Traffic Volumes

While infrastructure savings are significant, the primary influencing factor behind growing IP applications is the new revenue opportunities that IP networks present. The increasing open IP network environment enables service providers to implement new services quickly and efficiently. As more carries shift to IP as the principal service delivery platform, development efforts will continue to drive new value added applications, such as e-commerce, call centers, distance learning, unified messaging, webcasting, vedio conferencing and interactive voice response (IVR), among others. According to research from Cahners In-Stat Group, revenues from IP services will increase dramatically from less than $3 billion in 2000 to nearly $70 billion by 2004.

Interconnectivity PSTN/IP

A workgroup called TIPHON, under ITU-T, was established in April 1997 to address the interoperability issues between voice applications using IP and SCN, Switched Circuit Network, applications.

The objective is to support a market for voice communication and related multimedia issues not only between users in IP-based networks and users in SCNs, but also between users in an SCN using IP-based networks for connection/truncking.

The TIPHON work process has been divided into four scenarios:

1. Voice communication from an Internet voice user to voice users in Packet-Switched Data Network (PSTN/ISDN/GSM). The objective is to give Internet users access to PSTN/ISDN/GSM voice services, including application supplementary and basic IN-based services, based on existing features in these networks.
2. Voice communication from users in PSTN/ISDN/GSM to voice users on the Internet, where identification of the called party is based on E.164 or IP numbering.
3. Voice communication between users of PSTN/ISDN/GSM using the Intenrt for the connection/truncking between the involved PSTN/ISDN/GSM
4. Voice communication from voice Internet users using SCN for the connection/truncking between the involved IP-based networks, whereby identification of the called party is based on E.164 or IP numbering.

The TIPHON objectives are set to support voice only, using H.323 in the IP network. TIPHON specifies mandatory messages and message interworking between SCN and the IP network.

The functional model is based on the H.323 model, which consists of three entities – Gatekeeper, Gateway, and Terminal, as defined in H.323. For a more flexible and scalable model, TIPHON has divided the Gateway into three separate functional entities:

- A signalling Gateway, which provides signalling mediation between IP domain and the SCN domain.

- A Media Gateway, which provides media mapping and/or transcoding functions, i.e., it terminates the SCN PCM signal and the packet media as defined in H.225 and performs address translation, echo cancellation, playing announcements, receiving/sending DTMF digits, etc.
- A Media Gateway Controller (MGC), which provides the H.323 signalling functions as defined in H.323, H.225 (RAS and Q.931), and H.245 and performs a mapping of the SCN signalling information to H.323 signalling. The MGC mainly provides media gateway control, such as resource monitoring, gateway control (connection control, resource management, and protocol translation), usage monitoring, and reporting.

The Gatekeeper is responsible for control and management of the elements in the network. This includes address resolution 9SCN to H.323 aliases and vice versa) and call routing: all the Gatekeeper aspects defined in H.323. In addition, the Gatekeeper definition in TIPHON includes charging, accounting, usage reporting, management, etc.

The back-end services provide functions such as authentication, billing, rating/tariffing, address resolution, etc.

Interdomain Settlement for Billing and Accounting has required that a protocol be specified for Interdomain pricing, authorization, and usage exchange. Such a protocol is being specified based on HTTP and eXchange Markup Language (XML); it will define authorization, pricing, and usage.

Every operator wants to offer a range of Quality of Service (QoS) levels and to charge subscribers accordingly. For this reason, it is imperative not only to maintain end-to-end quality within the operator’s network, but also to maintain the quality level across different peering operator domains.

To handle this end-to-end performance interoperability, TIPHON has defined four QoS classifications, which are used to describe the following three areas:

- End-to-end speech quality – this is characterized by comparatively subjective rating (Mean Opinion Score (MOS)). The recommended method id the “Absolute Category Rating” described in ITU-T P.800 and P.830.
- Call setup time – which is the delay experienced by the originator user between the time a number is dialed and a response tone is received.
- End-to-end delay – which is the time delay imposed on the voice signal from various elements in the IP network.

Obviously, to obtain the end-to-end QoS classification, these requirements must be applied to both the terminal devices and the network elements.

The Voice Gateway is being modeled from three functional elements:

- **SS7 Gateway (SG):** The SS7 Signalling Gateway performs an SS7 Signal Transfer Points (STP) function involving the SCN and a protocol conversion between an SS7 network and IP network transport protocols.

- **Media Gateway Controller (MGC):** The MGC handles registration, management, and control functionality of resources at the MG. The MGC can also perform protocol conversion between PSTN signalling protocols and IP Telephony.
- **Media Gateway (MG):** The MG terminates the switched-circuit connection from SCN, i.e. truncks or local loops, and interfaces into IP network. One of the main functions that the MG performs is media conversion from a PCM to a packetized format and vice versa. It also handles the media stream between two endpoints (e.g. a RTP/RTCP session).

APPLICATIONS AND SERVICES

As more carries shift to IP as the principal service delivery platform, in addition to voice and fax, development efforts will continue to drive new value added applications, such as e-commerce, call centers, distance learning, unified messaging, webcasting, vedio conferencing and interactive voice response (IVR).

CODING

Compression Algorithm

For traffic over switched Ethernet LANs where the QoS is excellent and there is plenty of bandwidth for voice and data functions, 64 Kbps G.711 PCM voice coding can be used. Various compression algorithms exist which have different performance characteristics: G.726 ADPCM which operates at 16, 24, 32, and 40 Kbps, G.723.1 which operates at 5.3 or 6.3 Kbps and G.729 which operates at 8 Kbps. Typically, voice algorithms that perform greater compression require much more processing power. It should be noted that high fidelity audio quality compression algorithms can also be used since IP Telephony is not subject to the 4 KHz bandwidth restrictions found in the PSTN. This would provide better sounding audio than PCM and allow music to be faithfully reproduced.

SUMMARY

IP Telephony and Voice-over-IP technologies allow the integration of voice and data communications, reducing costs and revealing new opportunities for both telecommunications service providers and corporate users.