

VoIP Overview

VoIP Protocol Overview

The voice over IP (VoIP) protocol suite is generically broken into two categories, control plane protocols and data plane protocols. The control plane portion of the VoIP protocol is the traffic required to connect and maintain the actual user traffic. It is also responsible for maintaining overall network operation (router to router communications). The data plane (voice) portion of the VoIP protocol is the actual traffic that needs to get from one end to another.

Common LAN traffic demonstrates a good example of differentiating between control plane and data plane traffic. A user may “surf the web” (HTTP) or send e-mail (SMTP) across a network. This constitutes data plane (user) traffic. On the other hand, the routers in that network are also communicating over the same LAN using OSPF (Open Shortest Path First) or RIP (Router Information Protocol). This traffic is never visible to the user, but it is required to route the user traffic. This constitutes control plane traffic.

Within the VoIP suite of protocols, voice packets are commonly referred to as the data plane. Likewise, signaling packets are commonly referred to as the control plane. This document will examine the VoIP protocol suite in this manner, data plane protocols and control plane protocols.

VoIP Protocol Stack

As its name implies, VoIP utilizes IP as its basic transport method. VoIP utilizes both the TCP and UDP protocols over IP. The following diagram shows the protocol stack for a VoIP network.

It is important to note that VoIP works with any protocol stack that supports IP. End users of VoIP can add enterprise VoIP systems to their existing infrastructure relatively quickly and easily.

| | | |
|-----------------|------------|-----------------|
| VoIP | | |
| TCP | UDP | |
| IP | | |
| HDLC | ATM | ETHERNET |
| PHYSICAL | | |

Data Plane Protocols (The Voice)

RTP and cRTP

Both Real-Time Protocol (RTP) and Compressed Real-Time Protocol (cRTP) are currently available using any of the control plane protocols defined in this document. Since the voice traffic within a VoIP network can often take a different path than the signaling traffic, it makes sense that they are independent protocols.

RTP

RTP is the protocol that supports user voice. Each RTP packet contains a small sample of the voice conversation. The size of the packet and the size of the voice sample inside the packet will depend on the CODEC used.

RTP Protocol Stack

The following diagram shows the RTP protocol stack.

| |
|---------------------|
| Voice Sample |
| CODEC |
| RTP |
| UDP |
| IP |

RTP information is encapsulated in a UDP packet. If an RTP packet is lost or dropped by the network, it will not be retransmitted (as is standard for the UDP protocol). This is because a user would not want a long pause or delay in the conversation due to the network or the phones requesting lost packets. The network should be designed, though, so that few packets are lost in transmission.

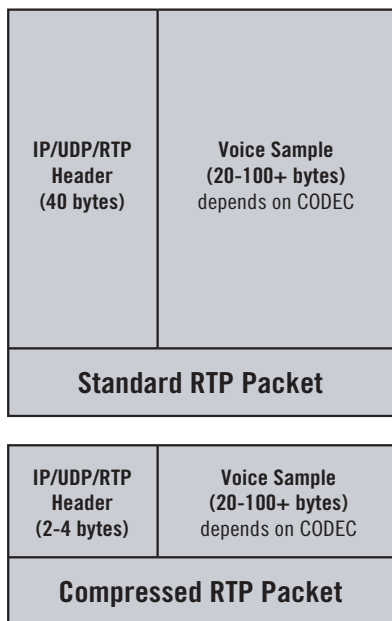
RTP Header

The RTP frame contains several pieces of information to identify and manage each individual call from endpoint to endpoint. This information includes a timestamp, a sequence number, and conversation synchronization source information. A complete list of RTP header information is provided in Appendix A.

Compressed RTP

A variant of RTP is compressed RTP (cRTP). Compressed RTP eliminates much of the overall packet header. By eliminating this overhead, a more efficient packet is placed onto the network. With a system running cRTP, a user can place approximately twice as many calls as compared to a system running standard RTP.

Compressed RTP is used on point-to-point WAN links. Point-to-point, in this case, is not implying a PPP layer 2 framing format. The link layer may be any standard WAN link layer protocol (frame relay, HDLC, PPP, or Cisco HDLC). The following diagram of a cRTP frame demonstrates why it can only be used on a point-to-point link.



Since the IP header is compressed with the UDP and RTP headers down to a maximum of 4 bytes, there is no room for an IP address. Therefore, the packet cannot be routed. It can only be placed on a point-to-point link that requires no addressing.

The issue with cRTP, similar to any form of compression, is that it takes more processing cycles of the router to process the packet. The router must recreate each header as the packet arrives so that the packet can be routed on the LAN to the IP phone. When cRTP is used, router utilization can be a factor in the overall call quality.

RTCP

Real-Time Control Protocol (RTCP) is a data plane protocol that is not always used. This protocol allows the endpoints to communicate directly regarding the quality of the call. RTCP affords the endpoints the ability to adjust the call in real time to increase the quality of the call.

RTCP also aids significantly in the troubleshooting of a voice stream. Traditional VoIP analyzers sit at specific locations on a circuit and base their derived results from only the packets that they capture. With RTCP enabled, the analyzer can see the end-to-end quality as well as the quality at the point at which the analyzer is inserted. This capability allows the user to sectionalize problems much more quickly.

RTCP XR

RTP Control Protocol Extended Reports (RTCP XR) is a newer extension of the RTCP concept. It defines a set of metrics that can be inexpensively added to call managers, call gateways, and IP phones for call quality analysis. RTCP XR messages are exchanged periodically between IP phones and gateways.

With RTCP XR messages enabled, an analyzer sitting midstream on a voice call can see and decode the messages. These messages can also be retrieved via SNMP requests and can be fed into a larger network performance management system.

RTCP XR provides information on the following call quality metrics.

Packet Loss/Discard – The endpoints of a phone call examine each RTP packet and identify lost packets using the sequence numbers. The endpoints also identify those packets that arrive too late and are discarded by the endpoint. These RTP packets are referred to as discarded packets.

Delay – RTCP XR reports on the round trip delay detected using RTCP and adds reporting information on the full envelope delay. The envelope delay includes the CODEC and jitter buffer.

SNR and Echo – RTCP XR reports on the signal-to-noise ratio (SNR) at each endpoint. If the endpoint is equipped with an echo canceller, RTCP XR reports on the un-canceled echo level.

Overall Call Quality – Using simple embedded algorithms, RTCP XR can report MOS ratings or R factor values for the call.

Configuration Information – RTCP XR can report on the overall configuration of an endpoint, including jitter buffer size.

CODECS

There is a wide range of voice CODECs (coder/decoder or compression/decompression) protocols available for VoIP phone implementation. The most common voice CODECs include G.711, G.723, G.726, G.728, and G.729. A brief description of each CODEC follows.

G.711 – Converts voice into a 64 kbps voice stream. This is the same CODEC used in traditional TDM T1 voice. It is considered the highest quality.

G.723.1 – There are two different types of G.723.1 compression. One type uses a CELP compression algorithm and has a bit rate 5.3 kbps. The other type uses an MP-MLQ algorithm and provides better quality sound. This type has a bit rate of 6.3 kbps.

G.726 – This CODEC allows for several different bit rates, including 40 kbps, 32 kbps, 24 kbps, and 16 kbps. It works well with packet to PBX interconnections. It is most commonly deployed at 32 kbps.

G.728 – This CODEC provides good voice quality and is specifically designed for low latency applications. It compresses voice into a 16 kbps stream.

G.729 – This is one of the better voice quality CODECs. It converts voice into an 8 kbps stream. There are two versions of this CODEC, G.729 and G.729a. G.729a has a more simplified algorithm over G.729, allowing the end phones to have less processing power for the same level of quality.

Voice Quality Metrics

Overall Quality Factors

There are several factors that affect the quality of a VoIP call in an operational environment. The following section describes these factors.

CODEC

The choice of CODEC is the first factor in determining the quality of a call. Generally, the higher the bit rate used for the CODEC, the better the voice quality. Higher bit rate CODECs, however, take up more space on the network and also allow for fewer total calls on the network.

Network

The biggest factor in call quality is the design, implementation, and use of the network that the voice calls are riding on. A typical VoIP call will start on a LAN at a CPE, go through a WAN connection to a provider cloud, and then go back out the other end. The CPE LAN and WAN links are most vulnerable to over utilization and errors. Most VoIP quality issues are typically caused at these links.

There are several ways a network can affect a VoIP call, including packet jitter, packet loss, and delay.

Packet jitter – This is caused by changes in the inter-arrival gap between packets at the endpoint. The packets should arrive evenly spaced to allow for a seamless conversion into analog voice. If the packet gap changes, the user could experience degradation in quality. If the packet gap gets sufficiently large, the phone's packet jitter buffer will not be able to wait for the late packet, and the phone will drop the late packet. There are three different types of packet jitter – RFC jitter, instantaneous jitter, and absolute jitter.

RFC jitter, or “smoothed jitter”, is defined by an ITU standard and essentially assigns a standardized value to the packet jitter of a call. The advantages of this metric are that it is defined by a standard organization and the equipment measuring this type of jitter should generate the same results. The disadvantages of RFC jitter are that it is a fluctuating average, and it eliminates spikes in the jitter that can cause packets to be dropped by the phone's jitter buffer. For these reasons, RFC jitter is not a very useful statistic.

Instantaneous jitter is the actual inter-packet jitter measurement, measuring the arrival time of each packet. There is no smoothing algorithm to eliminate spikes. Instantaneous jitter is the most realistic jitter measurement. The jitter buffer uses the instantaneous jitter measurement to determine which packets it will keep and which packets it will drop.

Absolute jitter is very different from RFC jitter or instantaneous jitter. Both RFC and instantaneous jitter rely on the current packet gap to determine their values. Absolute jitter represents the changes in inter-packet arrival times as compared to the previous packet gap.

Packet loss – This is the actual loss of voice packets through a network. Packet loss is often caused by congestion at one or more points along the path of the voice call or by a poor quality link (one that experiences physical layer errors).

Delay – Delay, sometimes referred to as envelope delay, is the time it takes for the voice to travel from the handset of one phone to the ear piece of the other phone. Envelope delay is the sum of the delay caused by the CODEC of choice, jitter buffer in the phone, and the path time it takes for the packets to get through the network. A large delay can make conversation difficult.

Echo

Echo is a common problem for VoIP networks. It is important to note that, unlike packet jitter, packet loss, and delay, echo is not caused by the IP network. Echo is an analog impairment. It is extremely difficult to passively monitor for echo. The best way to detect echo is by placing a call onto the network with a known “good” device or capturing the voice packets of a call and playing them back for analysis.

There are two types of echo on analog voice networks – hybrid echo and acoustic echo. Hybrid echo is generated by impedance mismatches at various analog or digital points on the network. The most common location for the generation of hybrid echo is at a 2-wire to a 4-wire conversion point. Acoustic echo is generated at the phone. It occurs when the voice leaving the speaker is picked up by the microphone.

Measuring Quality

There are many methods for measuring voice quality on a VoIP network. The following section describes these methods.

Intrusive: Non-real-time, two-ended methods

These methods involve sending known voice samples across a network from one endpoint to a receiving endpoint. The receiving endpoint does a comparison analysis of the degraded sample with the original. Because of the complexity of the signal comparisons, intrusive testing algorithms are computationally intensive and are not viable for real-time quality measurements. The following section describes the most common algorithms for this comparison.

PSQM

Perceptual Speech Quality Measurement (PSQM) is designed to avoid the subjective nature of Mean Opinion Score (MOS) rating and the effort it takes to get people into a room and listen to voice calls. PSQM measurements are performed by generating a known signal into the phone and then measuring what comes out the other end (post CODEC). The two signals are compared, and a PSQM value is derived.

However, PSQM was only designed to test the compression/decompression of CODEC functions. The algorithms that are used do not support overall end-to-end call quality through the network. Basically, PSQM does not have the ability to account for the effects of packet loss and packet jitter on voice quality.

PESQ

Perceptual Evaluation of Speech Quality (PESQ) was developed to expand the PSQM measurement from CODEC analysis to include distortion, filtering, and other channel impairments. PESQ is still not capable of handling all of the issues that could occur in a network, including excessive delay and packet loss.

PAMS

Perceptual Analysis and Measurement System (PAMS) is similar to PSQM and PESQ, but it is designed to access the signal at an analog interface. Like PSQM and PESQ, a known signal is injected into the system, and the output is measured. It is not designed to test the overall end-to-end quality of a call.

Passive: Real-time, single-ended methods

These methods passively calculate voice quality without a reference voice sample. They are most commonly used in the turn-up and testing of actual networks.

E-Model (R-Factor)

The E-model produces a single value called an R-factor. This value is derived from a variety of factors, including delay and other network impairments. Originally the E-Model was intended for use in network planning and design. The goal of the E-Model is to measure MOS without using all of the people that are typically required to provide an accurate MOS rating. R-factors range from 0 (extremely poor) to 100 (high quality). Any R-factor below 50 is unacceptable. TDM-based phone calls have a maximum R-factor of 94.

There are three main variations of R-factor – R_{CQE} , R_{LQE} , and R_{NPE} .

- R_{CQE} : This is the call quality estimate. It is the estimated quality of the call in both directions.
- R_{LQE} : This is the listening quality estimate. This metric removes delay impairments.
- R_{NPE} : This is the network performance estimate. This metric removes CODEC degradation impairments, allowing the user to determine how the network is handling the raw packets.

Refer to Appendix E for a more detailed description of the calculation of R-factor, including the different variables in the equation.

MOS

MOS (Mean Opinion Score) assigns a value to the overall quality of the delivered voice through a network. This measurement scheme takes into account both the CODEC and the network. MOS ratings have a range from 1 (bad) to 5 (excellent). A true MOS rating is determined by people listening to the same call and rating it from 1 to 5.

Test devices can measure MOS ratings through complicated algorithms based on the data from large groups of listeners rating calls. The test devices can then provide overall and per call MOS ratings to give network operators an accurate view of how their network is performing. This is currently the most common VoIP call quality measurement.

The MOS score of a call can be increased using the packet loss control (PLC) algorithm. This algorithm can be applied at the phone or at the media gateway to mask packet loss to the end user. For low levels of packet loss, the algorithm detects the lost packets and plays back a small sample of speech, typically from the last received packet. This effectively tricks the end user's ear, masking the packet loss; thus affecting the overall MOS score.

Control Plane Protocols (The Signaling)

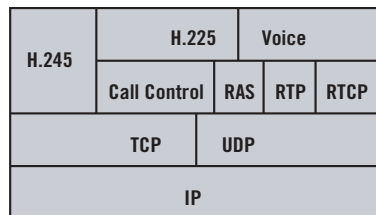
The control plane is used for the various signaling protocols, allowing users of VoIP to connect their phone calls. There are several different types of VoIP signaling available today, including H.323, SIP, SCCP, MGCP, MEGACO, and SIGTRAN. The most prevalent types of signaling protocols today, H.323 and SIP, are discussed in this document.

H.323 Standard Overview

H.323 was the first widely adopted and deployed VoIP protocol suite. The H.323 standard was developed by the International Telecommunications Union- Technology Standardization Sector (ITU-T) for transmitting audio and video over the Internet. Over the past 10 years, this standard has gone through several revisions and additions to encompass more features, scalability, and stability. The current version of H.323 is Version 5.

H.323 Protocol

The overall protocol stack for H.323 (see below) is made of up several parts. Each part is responsible for specific tasks, such as call setup and phone registration.



H.245

H.245 is the media control portion of the H.323 protocol suite. H.245 establishes a logical channel for each call (endpoint to endpoint). During H.245 negotiation, capabilities and preferences are exchanged by each endpoint. The choice of CODEC for the call is part of this exchange.

H.225

H.225 represents the basic signaling messages that are also used when dealing with ISDN or GR-303. For H.225, these messages include setup, alerting, connect, call proceeding, release complete, and facility messages. They are based on the Q.931 signaling scheme and are defined as follows.

Setup – This message attempts to connect a call. The calling party sends this message to the called party.

Alerting – This message is sent from the called party back to the calling party to let the caller know that the far end is being alerted (ringing).

Connect – This message informs the calling party that the called party has accepted the call. The conversation can begin at this point.

Call Proceeding – This message informs the endpoints that the call is up and running. Call Proceeding messages are exchanged at specific intervals during the call.

Release Complete – This message is sent by the party (called or calling) who disconnects the call first.

Facility – These messages represent various control messages. They are often seen when a gateway is required to connect a call.

RAS

RAS (registration, admission, and status) protocol deals with element (phone) management. The RAS logical channel is established between the IP phones and the gatekeeper that manages those phones. Without appropriate RAS communications, an IP phone will not be able to place or receive calls.

A more complete list of RAS and H.245 messages is provided in Appendix B along with a sample of a basic signaling ladder.

SIP Overview

Session Initiation Protocol (SIP) is designed to manage and establish multimedia sessions, such as video conferencing, voice calls, and data sharing. SIP is still in its early stages of deployment and is a growing and evolving protocol standard. This is the standard that many element manufacturers are using to develop products.

There are several key features of SIP that make it so attractive:

1. URL addressing scheme – This allows for number portability that is physical location independent. Addressing can be a phone number, an IP address, or an e-mail address. The messages are very similar to those used by the Internet (HTTP).
2. Multimedia – SIP can have multiple media sessions during one call. This means that users can share a game, instant message (IM), and talk at the same time.
3. It is a “light” protocol and is easily scaleable.

The two components that make up a SIP system include user agents and network servers.

User Agents – User agents represent the phone (user agent client) and the server (user agent server). The user agent client (UAC) initiates media calls. The user agent server (UAS) responds to those requests for setup on behalf of the UAC. The UAS is also responsible for finding the destination UAC or intermediate UAS.

Network Servers – These elements include redirection, proxy, and registrar servers. Redirection servers do not process calls and only respond with information containing the appropriate address of the next server. Proxy servers contain features of both a client and a server. The proxy server can receive requests and response messages. It can also adjust the header information prior to forwarding the request to the next proxy server or back to the user client. The registration server registers new clients in the database and updates other databases.

SIP Protocol

As with HTTP, SIP messages can be broken into two major categories, including messages from clients to servers and messages from servers back to clients.

Message Headers

Each message has a message header. The message header identifies the message type, calling party, and called party. There are four basic message types.

General Headers – This message header applies to request and response messages.

Entity Headers – This message header provides information about the message body type and the length.

Request Headers – This message header enables clients to include additional request information.

Response Headers – This message header enables the server to include additional response information.

More information regarding message headers is provided in Appendix C.

Request Messages/Methods

Request messages, or methods, are somewhat similar to the Q.931 messages used in ISDN. Request messages are initiated by a client to a server. SIP, a “light” protocol, has only a few request messages that it uses to connect calls. The following section defines the SIP request messages.

Invite – An invite message, as the name implies, is a request from a client to speak to another client. It contains the media type and other capabilities of the client.

Acknowledgement – This message is a response to an invite message. It represents the final message in the invite process and the beginning of the media exchange (voice).

Bye – This message is sent by either client to end a call. The server is the first to receive the bye message followed by the opposite client.

Options – This message allows the client to collect information on other clients and the servers.

Cancel – This message cancels any message exchanges that are in progress but not yet completed.

Registration – This message registers a client with a server and allows the client to use the services on the network.

Response Messages

In keeping with the “light” design of SIP and its Internet friendliness, SIP designers borrowed most of the response messages from HTTP.

There are two categories of response messages, provisional and final. Provisional messages are sent during a request/response process as details are worked out. Final messages, as the name implies, are the final response messages to a series of request/response messages.

There are five classes of response messages, including success, client error, server error, global failure, and informational. Each message class has several message types. Specific response messages are listed in Appendix C.

Other Signaling Protocols

SCTP, TALI, MGCP, and SCCP are other protocols that perform signaling functions on a VoIP network. It is important to note that multiple signaling protocols can exist in some portions of the network.

SCTP

Stream Control Transmission Protocol (SCTP) is a protocol format used for transmitting traditional TDM signaling protocols over IP. The main signaling protocol carried over SCTP is SS7 (this is also referred to as SIGTRAN).

Since SCTP carries traditional SS7 traffic, the protocol must meet the same guidelines defined for SS7. These guidelines include:

1. It must be compatible with UDP.
2. It must support acknowledged and error-free transfer of data.
3. It must support the segmentation of SS7 messages.
4. It must allow for network level fault tolerance.

Since SCTP is essentially SS7 over IP, an SS7 guide should be consulted to better understand this protocol and its message types. The SCTP protocol stack is as follows.

| | |
|------------------------------|-------------|
| SS7 Application Layer | |
| TCAP | ISUP |
| SCCP | |
| MTP3 | |
| SCTP | |
| IP | |

TALI

Transport Adaptation Layer Interface (TALI) is a standard very similar to SCTP. The TALI protocol is recommended for ISUP and TCAP messages transported over TCP/IP protocols. The TALI protocol stack is as follows.

| | |
|------------------------------|-------------|
| SS7 Application Layer | |
| TCAP | ISUP |
| SCCP | |
| MTP3 | |
| TALI | |
| TCP | |
| IP | |

As with SCTP, the 56 kbps DS0 with MTP1 and MTP2 layers are replaced by a new physical layer (typically Ethernet), MAC, IP, and TCP layers.

MGCP

Media Gateway Control Protocol (MGCP) is a combination of Cisco's Simple Gateway Control Protocol (SGCP) and IPDC (Level 3 protocol). The main feature of MGCP is the capability of breaking a telephony gateway into two basic parts, a call control and a media element.

In an MGCP system, there are a set of IP phones, a call controller, signaling gateways, and media gateways. The gatekeeper is a combination of a signaling gateway (SG) and a media gateway (MG). Signaling is converted from traditional PSTN signaling to a packet domain signaling protocol. Voice is converted from the PSTN G.711 CODEC to the CODEC of choice in the packet domain.

The CC manages the call routing. Multiple MGs are supported by one CC in an MGCP cloud. This allows for central management and control of the edge elements.

One of the main focuses of MGCP is the control and management of calls from a PSTN phone to another PSTN phone through the packet domain. Since the use of PSTN will not go away any time soon, the majority of calls over the next few years will be routed over an IP network to and from PSTN network elements.

SCCP

Skinny Client Control Protocol (SCCP) is Cisco's proprietary protocol used between a Cisco call manager and Cisco VoIP phones. This protocol is based heavily on the H.323 standard.

With SCCP architecture, the vast majority of H.323 processing power resides in the call manager. The end phones run the Cisco Skinny Client. The client requires very little processing power. This keeps the cost of the phones down as well as offers a more scaleable architecture than the standard H.323 architecture.

Testing VoIP Deployments

As carriers deploy VoIP services to their business customers, turning up and troubleshooting the service will become critical. Unlike traditional TDM services, a simple BER test will not fully qualify the service. The ability to place and receive calls for turn-up and monitor the calls for troubleshooting will be required.

Turn-up

To effectively turn-up a VoIP service, both signaling and call quality will need checking prior to customer hand off.

Signaling – A technician needs to place and receive calls through the network to make sure that the link is properly provisioned with the correct signaling protocol (H.323, SIP, etc). Calls should be placed within the VoIP cloud as well as from the VoIP cloud to the PSTN. These calls should include local and long distance calls to multiple exchanges.

By confirming that all possible types of calls can be placed, the technician can confidently connect the end-user's CPE equipment knowing that any signaling issues will not be within the carrier's cloud.

Call Quality – While checking the various call options, the technician can monitor the quality of the RTP stream. The two main values that the technician will want to examine are the R-factor (derived from the E-model) and an interpreted MOS rating. The technician can then compare these values with the SLA defined by the carrier.

Troubleshooting

There are two types of troubleshooting that technicians will be called upon to perform, catastrophic failure and intermittent issues.

Catastrophic failure – When a circuit is non-operational, troubleshooting becomes very similar to turn-up. The technician will be able to terminate the circuit back into a test device and begin the process of checking connectivity to the local elements through pings, trace routes, and phone call placement. The problem can be sectionalized to the CPE or carrier and then fixed by the appropriate party.

Intermittent issues – Intermittent issues represent the most difficult problems to solve. Since the error condition does not occur during every phone call, or even every day, the customer will most likely not allow the circuit to be taken down and tested. The technician will need to monitor the circuit and determine which calls have issues, when the issues have occurred, and what the environment is during those times.

It is absolutely **critical that the technician monitor every call on the circuit**. It is common for complaints to come in from a specific user representing one problem. The problem, however, may be from an element (CTMS or router, for example) or from a path in the network that affects many users on the network. An analyzer monitoring calls should be able to look at all of the calls at once so that larger problems can be identified. Depending on the CODEC and the voice sample time (ms) in each packet, a gigabit Ethernet link can process up to 27,000 calls simultaneously.

| Voice Sample Time (ms) | Frame Rate (frame/s) | G.711 | | G.729 | |
|------------------------|----------------------|--------------------|-------------------------|--------------------|-------------------------|
| | | Voice Size (bytes) | Maximum Number of Calls | Voice Size (bytes) | Maximum Number of Calls |
| 20 ms | 50 | 180 | 4,845 | 20 | 12,755 |
| 30 ms | 33.3 | 240 | 5,902 | 30 | 17,379 |
| 40 ms | 25 | 360 | 5,708 | 40 | 21,187 |
| 60 ms | 16.7 | 540 | 6,056 | 60 | 27,120 |

Environment, in this case, refers to the traffic riding on the network at the time of a failure. Other applications can use router processing time or even bandwidth, causing calls to drop or become difficult to listen to. The only way to determine if CPE traffic is the culprit of the VoIP issues is to monitor the whole circuit and measure call quality in real time.

When all of the factors (packet loss, jitter, and delay) for a VoIP call are added up, a call quality rating (either MOS-derived or R-factor) is calculated. Customers will base their quality complaints on a MOS rating using their ear. On the other hand, technicians will base their values on a derived R-factor rating using test devices. If the technician's derived R-factor rating and the customer's rating of the call differ, the call must be captured and listened to. This is the only way to successfully work with customers. Therefore, tools used for troubleshooting and turn-up should have the ability to play back voice and to emulate the actual VoIP phone being used.

Appendix A – RTP Header Information

The RTP Header has the following frame format.

| | Bit 1 | Bit 2 | Bit 3 | Bit 4 | Bit 5 | Bit 6 | Bit 7 | Bit 8 |
|---------|-------------------|--------------|---------|--------------|------------|-------|-------|-------|
| Byte 1 | Version | | Padding | Extension(X) | CSRC Count | | | |
| Byte 2 | Marker | Payload Type | | | | | | |
| Byte 3 | Sequence Number | | | | | | | |
| Byte 4 | | | | | | | | |
| Byte 5 | | | | | | | | |
| Byte 6 | Timestamp | | | | | | | |
| Byte 7 | | | | | | | | |
| Byte 8 | | | | | | | | |
| Byte 9 | | | | | | | | |
| Byte 10 | SSRC | | | | | | | |
| Byte 11 | | | | | | | | |
| Byte 12 | | | | | | | | |
| Byte 13 | | | | | | | | |
| to | CSRC (0-60 bytes) | | | | | | | |
| Byte 72 | | | | | | | | |

Version – This is the RTP version number. It is currently set to 2.

Padding – This gives the number of bytes at the end of the payload that are considered padding (not voice) and should be ignored. Padding is often used when encryption is enabled to keep the packets at a fixed length.

Extension (X) – If set, the header is extended.

CSRC Count – This provides the number of CSRC headers that follow the fixed header.

Marker – The interpretation of the marker is defined by a profile. It is intended to allow for the marking of significant events, such as frame boundaries, in the packet stream.

Payload Type – This field identifies the format of the RTP payload and determines its interpretation by the application. The following list contains the possible payload types, as defined by RFC3351.

Sequence Number – This number increments by one for each RTP data packet sent. In addition, it may be used by the receiver to detect packet loss and to restore packet sequence. The initial value of the sequence number is chosen randomly.

Timestamp – This reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations. The resolution of the clock must be sufficient for the desired synchronization accuracy and for measuring packet arrival jitter. The clock frequency is dependent on the format of the data carried as payload. It is specified statically in the profile, or payload format specification, that defines the format. It may also be specified dynamically for payload formats defined through non-RTP means. If RTP packets are generated periodically, the nominal sampling instant, as determined from the sampling clock, is used. A reading of the system clock is not used. The initial value of the timestamp is random, as is the sequence number.

SSRC – This identifies the synchronization source. The value is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC. Although the probability of multiple sources choosing the same identifier is low, all RTP implementations must be prepared to detect and resolve collisions. If a source changes its source transport address, it must also choose a new SSRC to avoid being interpreted as a looped source.

CSRC – This is a list identifying the contributing sources for the payload contained in the packet. The number of identifiers is given by the CC field. CSRC identifiers are inserted by mixers, using the SSRC identifiers of contributing sources.

| Payload Type # | Name | Media (Audio/Video) |
|----------------|------------|---------------------|
| 0 | PCMU | Audio |
| 1 | Reserved | Audio |
| 2 | Reserved | Audio |
| 3 | GSM | Audio |
| 4 | G723 | Audio |
| 5 | DVI4 | Audio |
| 6 | DVI4 | Audio |
| 7 | LPC | Audio |
| 8 | PCMA | Audio |
| 9 | G722 | Audio |
| 10 | L16 | Audio |
| 11 | L16 | Audio |
| 12 | QCELP | Audio |
| 13 | CN | Audio |
| 14 | MPA | Audio |
| 15 | G728 | Audio |
| 16 | DVI4 | Audio |
| 17 | DVI4 | Audio |
| 18 | G7209 | Audio |
| 19 | Reserved | Audio |
| 20 | Unassigned | Audio |
| 21 | Unassigned | Audio |
| 22 | Unassigned | Audio |
| 23 | Unassigned | Audio |
| Dynamic | G726-40 | Audio |
| Dynamic | G726-32 | Audio |
| Dynamic | G726-24 | Audio |
| Dynamic | G726-16 | Audio |
| Dynamic | G729D | Audio |
| Dynamic | G729E | Audio |
| Dynamic | GSM-EFR | Audio |
| Dynamic | LS | Audio |
| Dynamic | RED | Audio |
| Dynamic | VDVI | Audio |
| 24 | Unassigned | Video |
| 25 | CelB | Video |
| 26 | JPEG | Video |
| 27 | Unassigned | Video |
| 28 | Nv | Video |
| 29 | Unassigned | Video |
| 30 | Unassigned | Video |
| 31 | H261 | Video |
| 32 | MPV | Video |
| 33 | MP2T | Audio/Video |
| 34 | H263 | Video |
| 35-71 | Unassigned | Not defined |
| 72-76 | Reserved | N/A |
| 77-95 | Unassigned | Not defined |
| 96-127 | Dynamic | Not defined |
| Dynamic | H263-1998 | Video |

Appendix B – H.323 RAS Messages, H.245 Messages, and a Sample H.323 Signaling Ladder

RAS Messages

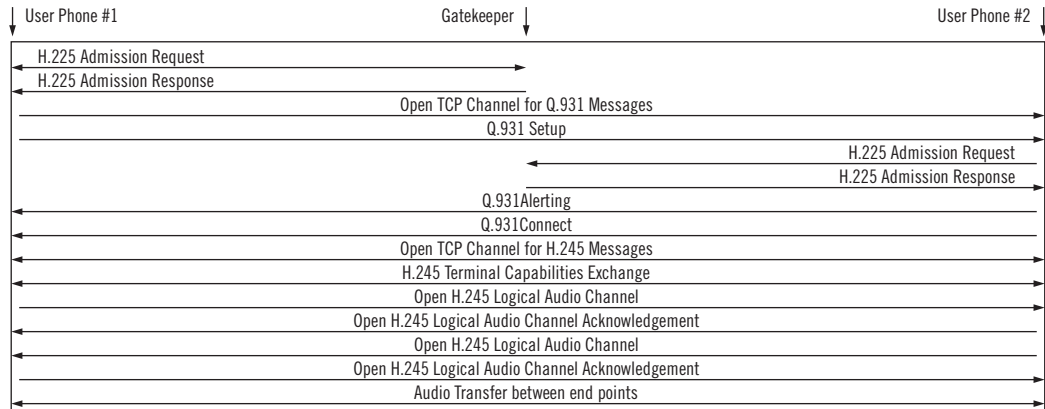
| <i>Message Type</i> | <i>Message Name</i> | <i>Description</i> |
|---------------------|------------------------|---|
| RRQ | Registration Request | Sent from phone to gatekeeper to begin registration process. |
| RCF | Registration Confirm | Sent from gatekeeper to phone to confirm registration. |
| RRJ | Registration Rejection | Sent from gatekeeper to phone rejecting the registration request. |
| URQ | Unregister Request | Sent from phone to gatekeeper to unregister the phone. |
| UCF | Unregister Confirm | Sent from gatekeeper to phone to confirm unregistration. |
| URJ | Unregister Reject | Sent from gatekeeper to phone to indicate that the phone was never registered. |
| BRQ | Bandwidth Request | Sent by phone to gatekeeper to request an increase or decrease in the required bandwidth. |
| BCF | Bandwidth Confirm | Sent by gatekeeper to phone to confirm bandwidth increase or decrease. |
| BRJ | Bandwidth Reject | Sent by gatekeeper to phone rejecting the bandwidth change request. |
| LRQ | Location Request | Sent by phone to gatekeeper to request an endpoint address. |
| LCF | Location Confirm | Sent by gatekeeper to phone confirming an endpoint address. |
| LRJ | Location Reject | Sent by gatekeeper to the phone rejecting the endpoint address request. |
| ARQ | Admission Request | Sent by phone to gatekeeper to request call initiation. |
| ACF | Admission Confirm | Sent by gatekeeper to phone to confirm the call initiation request. |
| ARJ | Admission Reject | Sent by gatekeeper to phone rejecting the call initiation request. |
| IRQ | Information Request | Sent from gatekeeper to phone requesting a status update. |
| IRR | Information Response | Sent from phone to gatekeeper in response to a status request. This message can also be sent periodically in the absence of an IRQ message. |
| SE | Status Enquiry | Can be sent from a gatekeeper to an endpoint or from one endpoint to another. Endpoints usually send these messages during calls for status. The gatekeeper can use these messages to keep track of active calls. |

H.245 Messages

| <i>Message Type</i> | <i>Description</i> |
|---------------------------|--|
| Capability Exchange | Consists of various messages that exchange the capabilities of the phones prior to a call being set up, including CODEC choice and media capabilities (video, data, or voice). |
| Master/Slave | For any call, one endpoint will be the master and the other will be the slave. This allows for the resolution of disputes when endpoints request features. |
| Round Trip Delay | Allows the phones to determine the delay across the cloud and back to the requesting endpoint. It also acts like a ping message. If the phone is not active, it will not respond to the round trip delay request. |
| Logical Channel Signaling | These messages open up the local channel for the RTP packets (voice) to travel across a network. If a gatekeeper is involved, the gatekeeper provides actual IP addresses of each endpoint so that the RTP stream can travel the shortest route. |

Sample H.323 Signaling Ladder

This ladder assumes a single gatekeeper involved in the call. The ladder will be different if no gatekeeper is used or if multiple gatekeepers are used.



Appendix C – SIP Message Headers and Response Messages

SIP Message Headers

The following list contains the specific message headers in each SIP Message Header type.

| SIP Message Header Type | Specific Header |
|-------------------------|---------------------|
| General Headers | Accept |
| | Accept Encoding |
| | Accept Language |
| | Call ID |
| | Contact |
| | CSeq |
| | Date |
| | Via |
| Entity Headers | Content Encoding |
| | Content Length |
| | Content Type |
| Request Headers | Authorization |
| | Contact |
| | Hide |
| | Max Forwards |
| | Organization |
| | Priority |
| | Proxy Authorization |
| | Proxy Require |
| Response Headers | Route |
| | Require |
| | Response Key |
| | Subject |
| | User Agent |
| | Unsupported |
| | Warning |
| | WWW Authenticate |
| | Server |

Response Messages

Response messages, messages from the server to the client, are each assigned a unique status code that defines the message type. The status codes are grouped together by response classes.

Informational Responses

| Status Code | Message |
|-------------|----------------------|
| 100 | Trying |
| 180 | Ringing |
| 181 | Call Being Forwarded |
| 182 | Queued |
| 183 | Session Progress |
| 200 | OK |

Success Responses

| Status Code | Message |
|-------------|---------------------|
| 300 | Multiple Choices |
| 301 | Moved Permanently |
| 302 | Moved Temporarily |
| 303 | See Other |
| 305 | Use Proxy |
| 380 | Alternative Service |

Client Error Responses

| Status Code | Message |
|-------------|--|
| 400 | Bad Request |
| 401 | Unauthorized |
| 402 | Payment Required |
| 403 | Forbidden |
| 404 | Not Found |
| 405 | Method Not Allowed |
| 406 | Not Acceptable |
| 407 | Proxy Authentication Required |
| 408 | Request Timeout |
| 409 | Conflict |
| 410 | Gone |
| 411 | Length Required |
| 413 | Request Entity Too Large |
| 414 | Request URL Too Large |
| 415 | Unsupported Media Type |
| 420 | Bad Extension |
| 480 | Temporarily Not Available |
| 481 | Cell Leg or Transaction Does Not Exist |
| 482 | Loop Detected |
| 483 | Too Many Hops |
| 484 | Address Incomplete |
| 485 | Ambiguous |
| 486 | Busy Here |

Server Error Responses

| Status Code | Message |
|-------------|-----------------------|
| 500 | Internal Server Error |
| 501 | Not Implemented |
| 502 | Bad Gateway |
| 503 | Service Unavailable |
| 504 | Gateway Time-out |
| 505 | Version Not Supported |

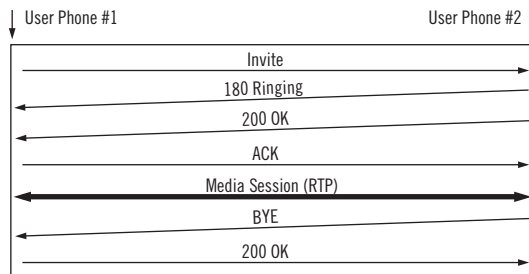
Global Failure Responses

| Status Code | Message |
|-------------|---------------------|
| 300 | Multiple Choices |
| 301 | Moved Permanently |
| 302 | Moved Temporarily |
| 303 | See Other |
| 305 | Use Proxy |
| 380 | Alternative Service |

Client Error Responses

| Status Code | Message |
|-------------|-------------------------|
| 600 | Busy Everywhere |
| 603 | Decline |
| 604 | Does Not Exist Anywhere |
| 606 | Not Acceptable |

Appendix D – Sample SIP Signaling Ladder



Appendix E – R-Factor Determination

R-factor is determined using the following equation:

$$R = R_o - I_s - I_d - I_{e-eff} - I_{recency} + A$$

The components of the R-factor equation are described in the following table.

| Component | Description |
|----------------------------|---|
| <i>R_o</i> | This is the signal-to-noise ratio. |
| <i>I_s</i> | This is the combination of all impairments that occur simultaneously with the voice signal. |
| <i>I_d</i> | This is the impairments caused by delay. |
| <i>I_{e-eff}</i> | This is the impairments caused by a low bit rate CODEC. It also includes impairments due to packet loss and rejection. |
| <i>I_{recency}</i> | This is the impairments resulting from significant packet loss. Significant packet loss is detected if there are eight or more packets lost in a row. |
| <i>A</i> | This is the advantage factor, which allows for the compensation of impairment factors. |

Appendix F – Bibliography

1. Davidson, J. and Peters, J., “Voice over IP Fundamentals”, Cisco Press, Indianapolis, IN, 2003.
2. Khasnabish, B., “Implementing Voice over IP”, Wiley and Sons, Hoboken, NJ, 2003.
3. RFCs: The protocols referenced in this document are available on the IETF’s RFC Web page at <http://www.ietf.org/rfc.html>.
4. “What is ‘Voice Quality?’”, Telchemy, Inc, 2003. This article is available on Telchemy’s Web site at http://www.telchemy.com/references/voice_quality.html.
5. Wright, David, “Voice over Packet Networks”, Wiley and Sons, Hoboken, NJ, 2001.

All statements, technical information and recommendations related to the products herein are based upon information believed to be reliable or accurate. However, the accuracy or completeness thereof is not guaranteed, and no responsibility is assumed for any inaccuracies. The user assumes all risks and liability whatsoever in connection with the use of a product or its application. JDSU reserves the right to change at any time without notice the design, specifications, function, fit or form of its products described herein, including withdrawal at any time of a product offered for sale herein. JDSU makes no representations that the products herein are free from any intellectual property claims of others. Please contact JDSU for more information. JDSU and the JDSU logo are trademarks of JDS Uniphase Corporation. Other trademarks are the property of their respective holders. ©2005 JDS Uniphase Corporation. All rights reserved. 30137172 501 1205 VOIPTERM.WP.ACC.TM.AE

Test & Measurement Regional Sales

| | | | | |
|--|---|--|--|---|
| <p>NORTH AMERICA TEL: 1 866 228 3762 FAX: +1 301 353 9216</p> | <p>LATIN AMERICA TEL: +55 11 5503 3800 FAX: +55 11 5505 1598</p> | <p>ASIA PACIFIC TEL: +852 2892 0990 FAX: +852 2892 0770</p> | <p>EMEA TEL: +49 7121 86 2222 FAX: +49 7121 86 1222</p> | <p>WEBSITE: www.jdsu.com</p> |
|--|---|--|--|---|