

Overcoming Barriers to High-Quality Voice over IP Deployments



Executive Summary

Quality of Service (QoS) issues are critical to the successful deployment of Voice over Internet Protocol (VoIP) systems. Impairment factors, such as latency, jitter, packet loss, and echo, must be considered, along with the ways in which they interact with each other. The ability of a network to support high-quality VoIP calls and the configuration of various hardware and software parameters are important in ensuring caller satisfaction.



Table of Contents

Introduction.....	2
Understanding QoS	2
Measuring QoS.....	2
Factors Affecting QoS	2
Basic Delay: Latency	2
Variable Delay: Jitter.....	3
Packet Loss.....	4
Bandwidth.....	4
Echo	4
Optimizing QoS.....	5
Reduce Latency and Jitter	5
Compensate for Packet Loss	7
Ensure Enough Bandwidth Is Available	7
Start with a Good E-Model “R” Factor.....	7
Summary.....	7
References	8
Appendix A: Configuring Hardware and Software Optimally	8
Appendix B: Key to Board References	9

Introduction

Deployment of Voice over Internet Protocol (VoIP) is growing, partly because Quality of Service (QoS) issues have been successfully addressed. In a properly configured VoIP installation today, conversations in VoIP calls can sound as good as, and sometimes better than, circuit-switched calls.

In a VoIP system, QoS depends on how a call is set up. More than telephone equipment and server technology must be considered. The network itself must be “VoIP-ready,” and endpoint devices need to be tuned for QoS.

Creating a high-quality VoIP system depends on understanding basic QoS concepts, such as latency, jitter, packet loss, and echo. Specific strategies can then be put in place to identify potential deployment issues at a specific site and to resolve them should any ultimately arise.

Understanding QoS

Measuring QoS

Measures of quality in communications systems tend to be very subjective, and over the years several attempts have been made to predict a caller’s feelings about QoS using objective criteria.

The traditional measure of a user’s perception of quality is the Mean Opinion Score (MOS), as defined in Methods for Subjective Determination of Voice Quality (ITU-T P.800). In P-800, an expert panel of listeners rated pre-selected voice samples of voice encoding and compression algorithms under controlled conditions. A MOS score can range from 1 (bad) to 5 (excellent), and a MOS of 4 is considered toll quality. The Pulse Code Modulation (PCM) algorithm (ITU-T G.711) has a MOS score of 4.4.

ITU-T G.107 presents a mathematical model, known as the E-Model, which attempts to predict QoS scores using more objective impairment factors. TIA/EIA TSB116 provides a comparison of E-Model Rating Values (R) and MOS scores. Table 1 provides details based on data in TIA/EIA [TSB116, Figure 1]. An R-Value of 94 is equal to a MOS of 4.4.

Factors Affecting QoS

Impairment factors important to VoIP that are used as input parameters to the E-Model include the following:

- Delay
- Packet Loss
- Speech Compression
- Echo

These and other impairment factors must be controlled and mitigated in the network and at endpoints to ensure high QoS ratings and customer satisfaction.

Basic Delay: Latency

The transmission of voice data packets is not instantaneous, and latency is the term used to describe the time needed for all of the following to occur:

- A packet of voice data to move across the network to an endpoint
- Encoding and packetization at the transmitting end
- De-jittering and decoding at the receiving end

Total latency is also called end-to-end latency or mouth-to-ear latency.

R-Value	Characterization	MOS
90-100	Very satisfied	4.3+
80-90	Satisfied	4.0-4.3
70-80	Some Users Dissatisfied	3.6-4.0
60-70	Many Users Dissatisfied	3.1-3.6
50-60	Nearly All Users Dissatisfied	2.6-3.1
0-60	Not Recommended	1.0-2.6

Table 1. Comparison of R-Values and MOS Scores

Conversations generally involve “turn-taking” on 200 ms breaks. When network latency approaches this value, the flow of a conversation becomes distorted. The two parties may start to talk at the same time or interrupt each other.

At over 170 ms delay, even a perfect signal degrades rapidly in acceptability.

Note: For more discussion of delay, including the expected delay for IP coders, see IT-UT [G.114].

Variable Delay: Jitter

When discussing networks, latency is often given as an average value. However, VoIP is time-sensitive, and such an average can be misleading.

When a packet stream travels over an IP network, there is no guarantee that each packet in the network will travel over the same path, as in a circuit-switched network. Because they do not take the same path, intervals between packet arrival times vary since one packet may take more “hops” than the others, delaying its arrival considerably and causing it to have a much higher latency.

Parts of an IP network can also be momentarily overwhelmed with other processing duties in, for example, a router. Both of these circumstances cause latency to become irregular, and this irregular delay is called jitter.

To lessen the effects of jitter, packets are gathered in a jitter buffer at the receive end. Figure 1 shows that with the same average latency, increased jitter requires a larger jitter buffer, which consumes additional memory and yields greater latency.

Note: The jitter buffers in Figure 1 would handle 99% of all packet latencies.

The jitter buffer must be sized to capture an optimal proportion of the data packets while keeping the effective latency as low as possible. In Figure 1, packets on the left and right ends of the bell curve fall outside of the jitter buffer and are, in effect, lost.

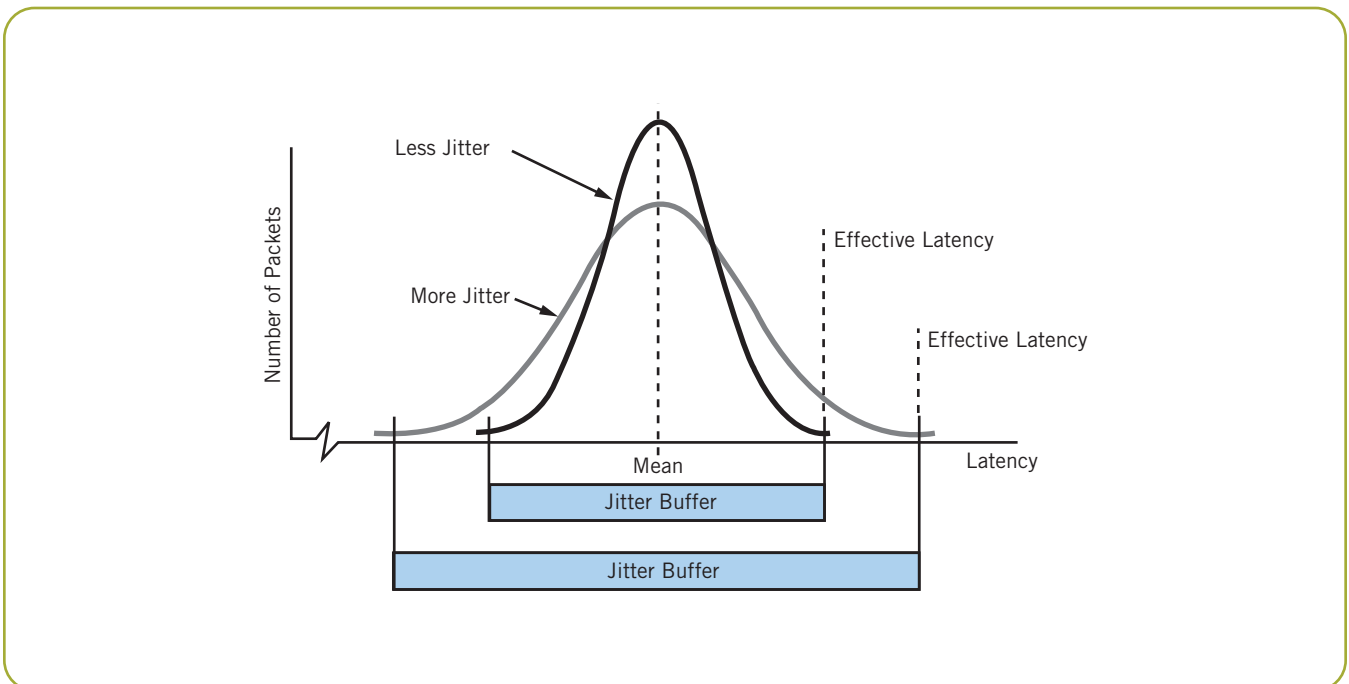


Figure 1. Jitter Increases Latency

Packet Loss

Because IP networks treat voice as if it were data, voice packets will be dropped just as data packets are dropped under severe traffic loads and congestion. Lost data packets can be re-transmitted, but, of course, this is not an acceptable solution for voice packets, which can contain up to 40 or even as many as 80 ms of speech information.

Packet loss, then, can significantly reduce QoS. Even a 1% loss can significantly degrade the user experience with the ITU-T G.711 voice coder (vocoder), which is considered the standard for toll quality. Other coders degrade even more severely because they compress the data more rigorously.

Lost packets are ignored in the calculation of jitter since they are, in effect, packets with infinite delay and would skew the calculations.

Bandwidth

Bandwidth is the raw data transmission capacity of a network. Inadequate bandwidth causes both delay and packet loss. Since IP network traffic is irregular, packets will often be delayed without some kind of prioritization.

Several techniques can be used to prioritize packets. These include Class of Service (CoS), Type of Service (ToS), Differentiated Services (DiffServ), and Internet Services (IntServ).

Class of Service (CoS)

The IEEE extension 802.1P describes Class of Service (CoS) values that can be used to assign a priority. Network devices that recognize three-bit CoS values deliver high-priority packets in a predictable manner. When congestion occurs, low-priority traffic is dropped in favor of high-priority traffic.

Type of Service (ToS)

RFC 1349 describes Type of Service (ToS), an in-band signaling of precedence for IP, which allows the Layer 3 IPv4 header to contain eight precedence values (0-7). These values are examined by routers and can also be used by Level 3 switches. DiffServ has superseded ToS.

Differentiated Services (DiffServ)

RFC 2474 redefines the ToS field in the Layer 3 IPv4 header as Differentiated Services (DiffServ). The redefinition is backwardly compatible. DiffServ defines a small set of Per-Hop Behaviors (PHBs) to define packet treatment. Examples are expedited forwarding, which simulate a virtual leased line; assured forwarding, which provides a high probability of delivery; and best effort. PHB is applied to each packet at each node, and the technique is highly scalable, performing classification at the edge.

Internet Services (IntServ RSVP)

RFC 2205 describes RSVP, an out-of-band QoS signaling protocol for reserving resources, such as bandwidth, for a “flow” or network path. Each flow is unidirectional, and two flows must be set up for each call. RSVP is not easily scalable, and DiffServ is expected to eclipse it.

Echo

Telephone handsets are designed to provide sidetone, which is necessary for a satisfactory user experience during a call. Microphone input is injected back into the earpiece, which allows callers to hear their own voice and feel that they are being transmitted correctly. Echo, however, is always a negative factor and should be minimized. In circuit-switched networks, latency is normally so low that echo is perceived in the same way as sidetone and is not a significant impairment.

On an analog phone, echo is generated at the 2- to 4-wire hybrid. On a digital or IP phone, echo is generated in the analog section of the phone (at the handset cord capacitive coupling, acoustic coupling, etc.). The [TSB116, 10] provides detailed diagrams of these types of connections.

In a VoIP configuration, the tail (that is, the time of a round trip from the gateway to the hybrid and back) on a gateway echo canceller only needs to handle delay on the circuit-switched leg of the call, as shown in Figure 2. An echo tail of 16 ms is usually adequate, with 32 ms required in France. Dialogic® DM/IP Boards, for example, provide a tail of 16 ms, but the Dialogic® IPT6720C IP Board provides a tail of 64 ms.

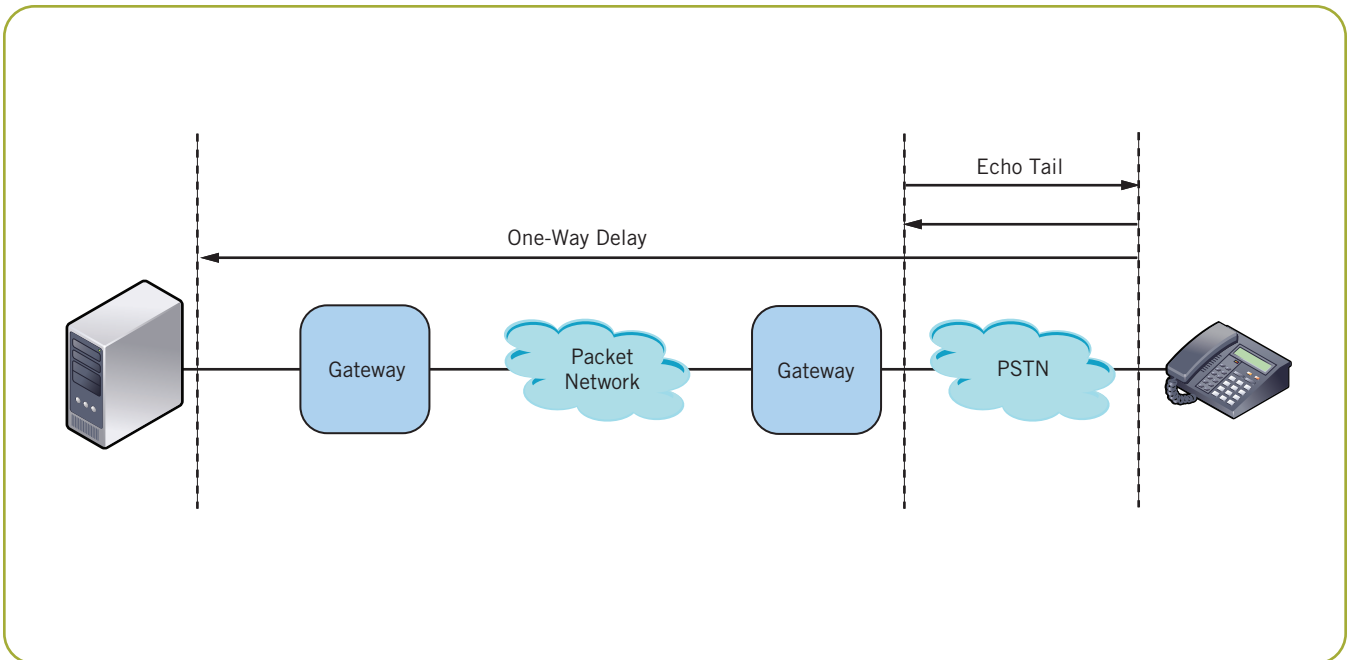


Figure 2. Echo and Delay in a VoIP Configuration

Echo Loudness must also be controlled. ITU-T G.168 recommends an Echo Loudness Rating (ELR) of ≥ 55 dB of echo path loss for echo cancellers in gateways. Echo cancellation is never perfect, and the more echo that is eliminated, the higher the computational load. The ITU recommendation is quite stringent and difficult to achieve in general.

Echo interacting with delay can compound a negative user experience. The [TSB116, 10] provides various Talker Echo Loudness Ratings (TELRs) mapped to delay, which shows how different loudness levels track to user satisfaction ratings.

Optimizing QoS

Some of the areas covered in this section correspond to the recommendations for IP telephony voice quality in TSB116.

Note: *A Summary of Voice Quality Recommendations for IP Telephony* appears in TIA/EIA [TSB116].

Reduce Latency and Jitter

Controlling delay is key to optimizing QoS, and should be kept well below 170 ms. Jitter does increase effective latency, so it is important to control both latency and jitter.

The size of a jitter buffer directly affects the latency perceived by the caller, so networks must be provisioned for low latency and low jitter both at the endpoint and from end-to-end.

Reducing Delay at an Endpoint

Several techniques can be used to reduce delay at an endpoint:

- Optimize jitter buffering — Dialogic's Packet Loss Recovery (PLR) algorithm provides an adaptive jitter buffer for Dialogic DM/IP Boards and Dialogic® HMP Software
- Optimize packet size — A packet size of 10 ms is currently preferred for reduced packetization latency. However, a larger number of smaller packets with relatively greater bandwidth overhead may be worse overall than a 20 ms packet size in some situations.
- Avoid asynchronous transcoding
- Use a stable packet size
- Use a low compression coder, such as G.711
- Ensure that network protocol stacks are efficient and correctly prioritized for VoIP traffic

Reducing End-to-End Delay by Prioritization

To reduce end-to-end delay, VoIP packets can be given a higher priority at Layer 2 and Layer 3 by using the following:

- Class of Service (CoS) — Implemented for Ethernet
- Type of Service (ToS) — Field in the IP header can be provisioned using the ipm_SetParm API on Dialogic® IP products
- DiffServ — Implemented at the router by static provisioning based on ToS bits
- RSVP for bandwidth reservation — Implemented at the router by static provisioning based on the transmitting port
- Policy-based network management
- Multi-Protocol Label Switching (MPLS) [RFC 3031] — Uses RFC 3212 CR-LDP for QoS

In order for prioritization to be effective, network stacks at endpoints must also prioritize the VoIP packets.

Dealing with Slow Links

On very slow links, a single large data packet could take up all available bandwidth for an unacceptable length of time. For example, a one-kilobyte data packet (equivalent to eight kilobits) traversing a 128 kilobits per second link

blocks the link for 62.5 ms, which delays six 10 ms packets of voice, greatly increasing jitter. Fragmenting large data packets and interleaving them with small VoIP packets, which are time-sensitive, can remedy this situation. Fragmentation and interleaving can be transparently provisioned between WAN routers with low-speed links.

Another technique to consider for reducing bandwidth demands on slow links is header compression, and IETF’s RFC 2508 describes an effective method for such compression. Normally, a header consists of nested headers for Real-time Transport Protocol (RTP), User Datagram Protocol (UDP), and Internet Protocol (IP), with a total size of 44 bytes. According to G.729A, the payload data at 20 ms in such a packet is only 20 bytes. However, by using a Compressed Real-time Transport Protocol (CRTP) header, the header is reduced to 2 to 4 bytes, which means the packet uses only about one-third of the bandwidth (24 bytes instead of 64 bytes). See Figure 3 for an illustration of this technique.

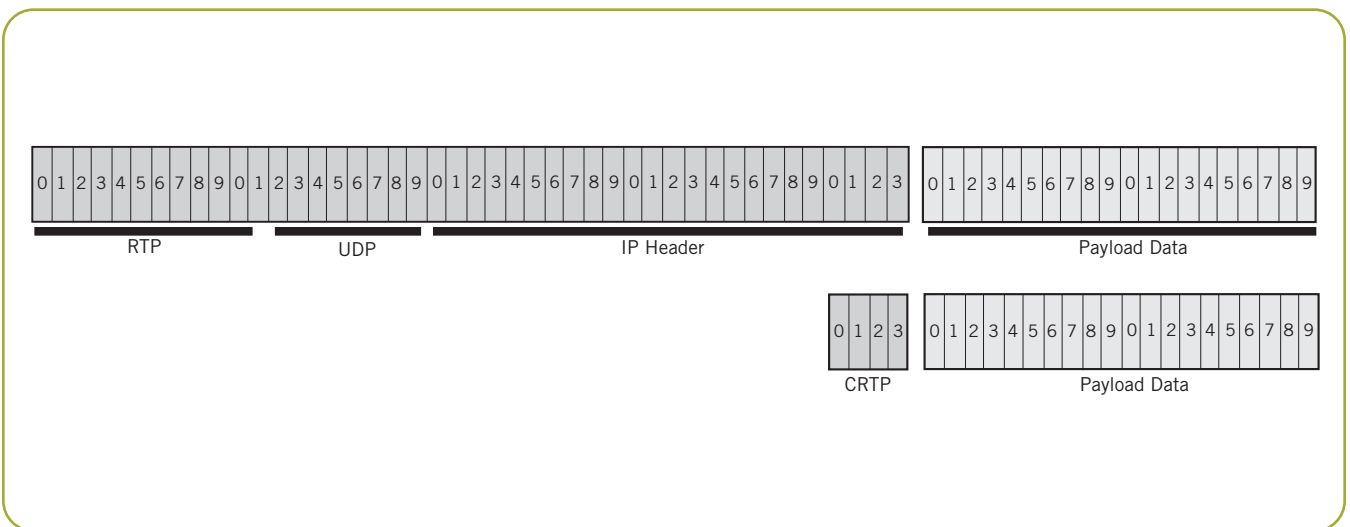


Figure 3. Header Compression

Compensate for Packet Loss

Packet Loss Concealment (PLC) or Packet Loss Recovery (PLR) algorithms at the endpoint can compensate for packet loss. Even a 5% rate of packet loss can be acceptable when the G.711 PLC algorithm is used. Many speech coders based on Codebook Excited Linear Prediction (CELP), such as [G.723.1], [G.728], and [G.729], have PLC built-in.

Note: See TIA/EIA [TSB116] for a discussion on Packet Loss. Figure I.1/G.711 provides an illustration of the dramatic effect of the frame erasure concealment algorithm for the G.711 vocoder [G.711].

Dialogic DM/IP boards and Dialogic HMP Software use PLC to repeat the previous packet when a packet is lost, and then inject silence for the remainder of the burst.

Dialogic's PLR algorithm is also valuable because it optimizes buffer size by adapting to current conditions. This algorithm is available with Dialogic DM/IP Boards and Dialogic HMP Software.

Payload redundancy (RFC 2198) can also be used to prevent packet loss, but it increases bandwidth requirements. It is supported with the `ipm_SetRemoteMedia` API in Dialogic IP products.

Ensure Enough Bandwidth Is Available

Speech compression also can be considered because it reduces bandwidth requirements end-to-end.

Speech compression algorithms are described in standards such as ITU-T G.723 and G.729. This kind of compression does reduce bandwidth requirements, but it also reduces perceived sound quality. In addition, packet loss has comparatively more serious consequences when high compression coders are used because more data is lost per packet. Even with small delays, compression coders are barely acceptable [TSB116]. For this reason, G.711 is considered the preferred voice coder, even though it has high bandwidth requirements.

G.711 also provides other advantages. As previously mentioned, G.711 Appendix I supplies a powerful packet loss concealment algorithm, and G.711 Appendix II provides two powerful tools for bandwidth conservation: a Voice Activity Detector (VAD) and a Comfort Noise Generator (CNG). The VAD senses when no voice is present, and sends sparse control packets rather than full packets of silence. The CNG plays background noise instead of no sound at all, which users find preferable to silence.

Bandwidth conservation techniques can provide about a 50% bandwidth savings simply by suppressing the normal silence in voice calls because the connection is full duplex with 64 Kbps in each direction. Since only one person talks at a time, the bandwidth consumed by the other person is empty and can be coded into silence packets.

Start with a Good E-Model "R" Factor

Use the best coder possible since the network degrades coder performance. G.711 is one option, although higher bandwidth coders such as [G.722] can potentially deliver superior sound quality if enough bandwidth is available. Low bandwidth coders, such as [G.723] or [G.729], are options if, for example, bandwidth is constrained. CRTP can be used to conserve bandwidth on slow links.

For more options, see Appendix A.

Summary

User perception of VoIP quality can radically improve when network and telephone equipment are set up correctly. Provisioning the network for QoS is paramount, but alone it is not enough. Impairment factors such as latency, jitter, packet loss, and echo, and the way these factors interact, are also critical. When careful attention is paid to controlling these factors and sufficient bandwidth is available, callers can be as satisfied with VoIP calls as they are with calls carried over a circuit-switched network.

References

[G.114] ITU-T G.114, *One-Way Transmission Time*, 2003.

[G.711] ITU Recommendation T-G.711, Appendix I, *Pulse Code Modulation (PCM) of Voice Frequencies: A high quality low-complexity algorithm for packet loss concealment with G.711*, 1999.

[G.722] ITU-T G.722, *7 kHz Audio-Coding Within 64 kbit/s*, 1993.

[G.723.1] ITU-T G.723, *Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*, 2006.

[G.728] ITU-T G.728, *Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction*, 1992.

[G.729] ITU-T G.729, *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)*, 2007.

[RFC 768] IETF RFC 768, *User Datagram Protocol*, August 1980.

[RFC 1889] RFC 1889, *RTP: A Transport Protocol for Real-Time Applications*, Audio-Video Transport Working Group, January 1996.

[RFC 2833] RFC 2833, *RTP Payload for DTMF Digits, Telephony Tones and Telephony Signals*, May 2000.

[TSB116] TIA/EIA TSB116, *Telecommunications IP Telephony Equipment Voice Quality Recommendations for IP Telephony*.

Appendix A: Configuring Hardware and Software Optimally

Before making specific changes to your hardware and software configuration, you can conduct a general VoIP readiness survey to understand your VoIP requirements and current traffic patterns. This survey can include steps such as identifying and remedying any network limitations that would inhibit effective VoIP deployment. You will then be ready to look at the specific configuration changes listed in the rest of this appendix.

1. Set Phone Defaults:

- Enable VLAN tag, which contains the CoS field
- Set CoS and the DiffServ Code Point (DSCP) for voice traffic
- Set coder to G.711
- Set packet to smallest (10 ms)
- Enable PLC
- Enable redundancy

2. Set NIC Defaults:

- Enable classification and tagging for VoIP with CoS and DSCP.

3. Set Parameters on Dialogic® Products:

- To optimize QoS, set certain specific parameters on Dialogic DM/IP Boards and Dialogic HMP software.
- If Dialogic® signaling stacks for SIP and H.323 are in use, do the following:
 - For setting coder and frame size for calls, use `gc_MakeCall` API
 - For setting ToS and [RFC 2833] redundancy level, use `gc_SetParm` API

- If other signaling stacks under split call-control are in use, do the following:
 - For setting coder and frame size for calls, use `ipm_SetMediaInfo` API
 - For setting ToS and RFC 2833 redundancy level, use `ipm_SetParms` API

4. Make Other Software Changes:

- Set the following parameters on a call-by-call basis:
 - Coder selection
 - PLC
 - Redundancy
 - Jitter buffer
- Record Real-time Control Protocol (RTCP) information with call details. This allows parameters to be tuned, and applications can query for RTCP information at any time with an `ipm` API.

Note: IETF [RFC 768] defines the User Datagram Protocol (UDP), which is considered an unreliable but fast protocol for IP. IETF [RFC 1889] uses UDP and specifies a media streaming protocol called the Real-time Transport Protocol (RTP). RTCP is a subset of the RTP specification.

- Use call quality alarms. This allows thresholds to be set, and administrators to be alerted when quality falls below those thresholds.

5. Consider Call Control Signaling and Media Streaming:

In general, call control signals should be given a higher priority than the media stream because a network segment congested with media streaming can cause an interruption in the transmission of call control signals. If you find this to be a problem, you may consider using a separate NIC for Session Initiation Protocol (SIP), H.323, or other control signals. Isolate them through the use of a separate Ethernet hub or use an Ethernet switch with significant headroom capacity.

Appendix B: Key to Board References

Two board categories are referenced in this paper. This appendix lists the current boards in each category.

Diallogic® DM/IP Boards

Diallogic® DM/IP241-1T1-PCI-100BT IP Board
Diallogic® DM/IP301-1ET1-PCI-100BT IP Board
Diallogic® DM/IP481-2T1-CPCI-100BT IP Board
Diallogic® DM/IP601-2E1-CPCI-100BT IP Board
Diallogic® DM/IP601-CPCI-100BT IP Board

Diallogic® IPT Boards

Diallogic® IPT1000C IP Board
Diallogic® IPT6720C IP Board
Diallogic® IPT4800C IP Board
Diallogic® IPT2400C IP Board
Diallogic® IPT1200C IP Board

To learn more, visit our site on the World Wide Web at <http://www.dialogic.com>

Dialogic Corporation

9800 Cavendish Blvd., 5th floor
Montreal, Quebec
CANADA H4M 2V9

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH PRODUCTS OF DIALOGIC CORPORATION OR ITS SUBSIDIARIES ("DIALOGIC"). NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN A SIGNED AGREEMENT BETWEEN YOU AND DIALOGIC, DIALOGIC ASSUMES NO LIABILITY WHATSOEVER, AND DIALOGIC DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF DIALOGIC PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT OF A THIRD PARTY.

Dialogic products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Dialogic may make changes to specifications, product descriptions, and plans at any time, without notice.

Dialogic is a registered trademark of Dialogic Corporation. Dialogic's trademarks may be used publicly only with permission from Dialogic. Such permission may only be granted by Dialogic's legal department at 9800 Cavendish Blvd., 5th Floor, Montreal, Quebec, Canada H4M 2V9. Any authorized use of Dialogic's trademarks will be subject to full respect of the trademark guidelines published by Dialogic from time to time and any use of Dialogic's trademarks requires proper acknowledgement.

The names of actual companies and products mentioned herein are the trademarks of their respective owners. Dialogic encourages all users of its products to procure all necessary intellectual property licenses required to implement their concepts or applications, which licenses may vary from country to country.