

# Voice over packet

## An assessment of voice performance on packet networks

**This white paper explores the major factors that influence the quality of voice services delivered over packet networks and recommends ways to minimize their impact.**

## Contents

<b>Overview</b> .....	<b>3</b>
<b>Packet voice quality issues</b> .....	<b>3</b>
<b>Summary</b> .....	<b>11</b>
<b>Endnotes</b> .....	<b>11</b>

## Executive summary

This paper discusses the main factors that affect voice performance in packet networks and offers recommendations on how to control them. The four key quality factors are:

1. Performance of the speech codec
2. Delay
3. Packet loss
4. Echo control

A thorough understanding of these quality factors provides a solid grounding for assisting customers in selecting the voice quality targets to meet their business needs and for planning and provisioning a network to achieve these targets.

This is the first of three white papers discussing the voice performance of the packet-based Succession portfolio from Nortel Networks. The second paper will cover the Nortel Networks process for selecting voice quality requirements and developing specifications to achieve those targets. The final paper in the series will address the network voice validation/verification process used by Nortel Networks. This process helps assure that a given solution can fulfill your requirements.

## Overview

Convergence of the telephone network and the Internet is driving the move to packet-based transmission for telecommunications networks. Integration of voice and data onto a single network offers significantly improved efficiency for both private and public network operators. Data is carried most efficiently on packet networks. Because data has overtaken voice as the major type of telecom traffic, and data traffic volume continues to grow faster than voice traffic volume, it is not surprising that the integrated network uses packet-based transmission.

Packet-based transmission of digital voice is a logical step, but it has some important implications for voice quality. First, the operating assumptions of packet networks are based on the requirements for data. Packet sizes, packet header overhead, the sizes of queuing and other buffers, etc., were chosen for optimal efficiency of data transfer.

Second, access links (which are dedicated to voice in switched-circuit networks) may be shared between voice and data in the packet environment. In some cases, this will severely constrain the data rate of a voice channel running over a packet-access link. Interworking of modern packet networks with the traditional circuit-switched PSTN, PBXs, and special networks, such as wireless, also poses quality management issues.

Finally, there are some specific challenges associated with providing many of the standard voice features, such as handsfree and conferencing, that voice telephony users have come to rely on.

Frame relay, ATM, and IP networks are the three common types of packet networks. From an end-to-end voice quality point of view, the differences among these transport technologies are generally not important, although each may present different challenges when meeting the requirements for packetized voice. Consequently, we will refer to “packet networks” throughout this discussion, rather than to specific transmission platforms.

## Packet voice quality issues

High end-to-end voice quality in packet transmission depends principally on these factors:

- The speech codec used
- End-to-end delay across the network and variation in the delay (jitter)
- Packet loss across the channel
- Echo control

Selecting the right speech codec is essential. Codec performance includes the baseline quality (that is, without impairments) and the performance with impairments present, such as background noise and lost/late packets. To prevent excessive degradation from transcoding, it is necessary to control

whether and where transcodings occur and what combinations of codecs are used.

It is also essential to control the end-to-end delay. When end-to-end delay exceeds 150–200 milliseconds one-way (300–400 milliseconds round-trip), the connection is noticeably impaired. Anyone who has talked on a call with a single satellite hop has experienced this effect. (A geostationary satellite connection adds about 300 milliseconds of one-way delay.)

In IP transmission, packets sometimes get “lost.” These packets may have been late or may have been discarded in the network because of congestion. The missing information degrades the output signal, and a packet loss concealment (PLC) algorithm may be needed to smooth over the gaps.

The delay through the packet network exacerbates any echo that may be present. Echo control with the right characteristics at the appropriate places in the connection protects against echo at both ends. Echo control becomes essential when packet network equipment is interconnected with circuit-switched equipment operating with two-to-four-wire conversions or “hybrids.”

There are other parameters that can affect performance. For example, an end device that doesn’t support the designated loss/level plan may not provide a usable voice connection.<sup>1</sup>

Though transmission levels and terminal performance are vital to strong end-to-end voice performance, these issues have been thoroughly addressed by industry standards, so there is no need to repeat the discussions here.

Voice quality issues in packet transmission are similar to those in digital wireless transmission. Wireless networks use low bit-rate codecs and are susceptible to channel impairments and increased end-to-end delay. Echo control within the wireless network and at the interface between the wireless and wireline networks is essential. Nortel Networks is a major supplier of wireless infrastructure and has extensive experience in delivering digital voice over wireless. This experience gives us a head start in delivering high-quality voice on packet networks.

The following sections describe the effects of the four performance factors, where and how they interact, the range of acceptable operation, and how they can be managed in the packet network environment.

### Speech codecs

A codec (**coder-decoder**) converts the analog voice signal to a digitized bitstream at one end of a call and returns it to its analog state at the other (codecs are also used to convert from one digital form to another, a process known as “transcoding”). In telephone networks, one of two techniques is generally used: waveform coding or CELP (code excited linear predictive) coding.

Waveform codecs directly or indirectly code the amplitude of the signal at each point, while CELP codecs are based on a model of the acoustics of the vocal tract during speech production.

ITU Rec. G.711<sup>7</sup> defines the PCM (pulse code modulation) coding that is used in much of the circuit-switched (TDM) digital network. G.711 is a waveform codec, and operates at 64 kbps in almost all telephony applications. G.726<sup>8</sup> defines ADPCM (Adaptive Differential Pulse Code Modulation), also a waveform codec. G.726 reduces the data rate, but also degrades the quality of the reproduced signal. The processing delay for both of these codecs is less than 1 millisecond, which is negligible. The main delay associated with the use of these codecs in packet networks is the packetization delay. This delay is equivalent to the duration of signal contained in each packet, typically between 10 and 40 milliseconds.

CELP codecs work on chunks of speech called “frames.” They use a model of speech production to remove redundancy from the signal, allowing transmission at a lower data rate—typically between 4 and 16 kbps for telephony applications. CELP codecs generally create more delay than do waveform codecs. A CELP speech frame cannot be generated until the encoder collects all the speech for the duration of that frame. This means there is a delay of one whole frame before the codec can begin processing. Some codecs also look ahead into the following frame to improve compression without sacrificing good reproduction. This adds more delay, because the encoder must wait for the look-ahead speech to be collected before processing the current frame.

If each processed frame from a CELP codec is put into a separate packet for transmission, no additional packetization delay is imposed. If additional

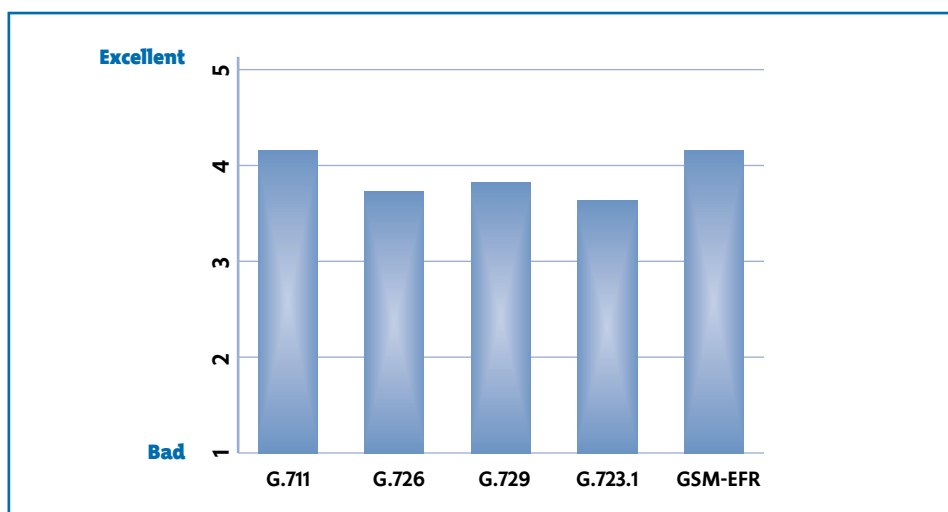


Figure 1. Results of subjective survey assessing codec quality with clean-input speech

**Table 1. Characteristics of speech codecs commonly used in packet networks**

Codec	Type	Bit rate	Frame size *	Total delay **	Other information
G.711	PCM	64 kbps	Depends on packet size		Codec of choice for high-quality, PSTN-equivalent voice service
G.726	ADPCM	32 kbps	Depends on packet size		Often used for multiplexing on 64 kbps channels and is specified for many low-powered wireless systems. Generally considered to be at lower limit of “toll” quality range.
G.729	CS-ACELP	8 kbps	10 ms	25 ms	ITU’s 8-kbps coding standard. Good delay characteristics (due to short frame) and acceptable voice quality. Has overtaken G.723.1 as codec of choice for applications requiring compression.
G.729A	CS-ACELP	8 kbps	10 ms	25 ms	Reduced-complexity version of G.729. The voice quality is equivalent to G.729. Identical decoders allow the two codecs to interwork seamlessly.
G.723.1	MP-MLQ	6.3/5.3 kbps	30 ms	67.5 ms	Default standard for off-the-shelf toll-bypass clients. Baseline voice quality generally inadequate for commercial telephony applications. Total delay compromises delay budget. Nortel Networks includes this codec in our products only for interworking with other vendors’ products.
GSM-EFR	ACELP	12.2 kbps	20 ms	40 ms	Wireless codec

\* A “frame” is a chunk of speech processed as a unit by a compression codec. GSM-EFR and many other wireless codecs use 20-ms frames. G.711 and G.726 are waveform codecs whose output is not chunked into frames. The packet size for these codecs is arbitrary. Where 20-ms packetization is used, the total delay for G.711 and G.726 is 20 ms.

\*\* The total delay for an ordinary implementation of the codec on a DSP, based on twice the frame size plus the “look-ahead”; assumes one frame per packet. Specific implementations will vary.

frames are put into the same packet, then those frames must be processed before the packet can be sent, and the delay increases by the amount of the frame length for each additional frame.

Packet transmission offers the flexibility to use different codecs as needed. In choosing a codec for a particular call or application, there are several considerations: the compression rate needed, the desired voice quality, the delay that the codec adds to the connection, how well the codec allows missing packets to be smoothed over, and whether a packet loss concealment algorithm must be added externally or is already built into the codec.

The encoding delay of the codec is an integral component of the end-to-end delay. Because compression codecs add

significant delay, the delay budget defining the distribution of allowable delay to the various network elements may require adjustment to accommodate a long encoding delay.

Table 1 shows some of the characteristics of the codecs most commonly chosen for point-to-point voice-over-IP applications.

Figure 1 shows subjective ratings of the voice quality of these codecs. We asked a group of users to rate the quality of various samples on a 5-point scale, ranging from 1 (bad) to 5 (excellent). The test method followed ITU Rec. P. 800.<sup>3</sup> Each mean is the average of 240 judgments. Differences greater than about 0.1 are real (that is, similar differences would occur if the study were repeated).

The input speech for these tests was clean (no background noise) and the channel was perfect (no missing packets or corrupted data). Therefore, these results reflect the very best that each codec can do. Because the ratings were gathered in a listening test, the effects of delay are not included in these ratings.

When packet loss is introduced, these codecs will show different amounts of degradation, depending on the effectiveness of the associated packet loss concealment (PLC) algorithm. The effects of packet loss are discussed below.

## Delay

The end-to-end delay (sometimes called “latency”) is the time between the generation of a sound at one end of a call and its reception at the other end. The delay includes the time taken to encode the sound as a digital signal, the signal’s journey through the network, and the regeneration of the signal as a sound at the receiving end. Delay causes two different impairments. First, as delay increases, echo becomes more noticeable. Second, when the delay becomes long enough, it disrupts conversation dynamics, making communication difficult. The effects of delay on conversation have been thoroughly studied at the Nortel Networks Subjective Assessment Laboratory.

Based on this work, Nortel Networks researchers have determined the delay thresholds at which conversation impairment occurs. It is especially important to note that while some impairments such as echo or noise can be removed from the signal, no measures can be taken to mitigate delay.

In the conventional PSTN, the largest part of the end-to-end delay is the propagation time of the transport medium. Processing causes some delay, but this is generally no more than a few milliseconds. In contrast, packetized voice encounters significant processing delay and additional delays created by queuing and jitter buffers. To minimize these queuing and propagation delays, network processing must be streamlined and packets carrying

**Table 2. Types and causes of one-way delay**

Delay sources	Ranges	Other info
<b>Propagation delay</b>	1-100 ms terrestrial; approximately 300 ms for geostationary satellite	Shortest propagation delay is local, longest is halfway around globe
<b>Processing delay</b>		
Codec	20-100 ms	Includes encoding and packetization delay for single IP hop, one frame per packet
Other DSP	0-30 ms	Can be PLC, noise suppression, silence suppression, echo cancellation
<b>Packet management</b>		
Delay for jitter buffer	1-20 ms	Depends on network utilization and whether congestion control is used
Multiple frames per packet	10-60 ms	Add time of additional frames beyond one
Interleaving	5-90 ms	Depends on size of frames and packets
<p><i>Note: The end-to-end delay is equal to the sum of the delay components. Contributing factors include the specific elements used to transmit the call and the call’s path through the network.</i></p>		

interactive voice communication must be given the most direct route through the network. Descriptions of the components contributing to end-to-end delay follow.

### Processing delay

Processing delay includes the time needed for encoding and decoding the speech, collecting the voice data into packets, and other DSP features such as echo cancellation, noise reduction, and packet loss concealment. These functions may be indispensable for achieving acceptable voice quality, but their contribution to delay must be taken into account.

### Propagation delay

Propagation delay is associated with sending a signal over a substantial distance. For instance, a fiber-optic trunk imposes a propagation delay of

about 5 microseconds per kilometer. As technological innovation reduces the number of repeaters required for fiber networks, fiber transmission speeds are increasing, but physical constraints forestall huge reductions in end-to-end propagation delay. However, the topology of the network can be controlled to keep propagation delay to a minimum by ensuring that packets take the most direct routes. For instance, backhaul to a central router or switching point may add unnecessary delay.

### Buffering delay

Data waits in a buffer for transmission or processing. Buffers are used for queuing at routers and to control packet arrival time at the decoder. Travel time through the network can vary for individual packets. Because the voice playback speed must be constant, a jitter buffer is used to remove variation

**Table 3. ITU recommended end-to-end (one-way) delay times<sup>4</sup>**

Total one-way delay	Recommendation for use
0-150 ms	Acceptable
50-400 ms	Acceptable for some applications, care required to assure user satisfaction
+400 ms	Unacceptable for general network planning

(jitter) in the flow of packets to the decoder. The delay imposed by the jitter buffer depends on the variation in delay across the network. Where congestion control is used, a very short jitter buffer is sufficient. Uncontrolled jitter causes packet loss (see section below on packet loss).

Poorly designed edge devices and congested network nodes can cause additional delay, and delay is further compounded by multiple conversions from packet to TDM and back. Network component failures resulting in a change to a less direct path can also increase delay.

Delay is conventionally characterized as one-way, on the assumption that the paths in each direction are symmetrical (of course, in real networks this may not be the case). The total round-trip delay for the examples given is twice the one-way delay.

*Table 2* summarizes the delay that can be expected during the stages of transmission through a packet network.

Delay can significantly impair conversation. Delay destroys simultaneity, disrupting natural turn-taking. Longer delays cause simultaneous starts and awkward silences. It becomes difficult to interrupt the other party, and attempts to do so may appear especially impolite because of the difference in perceived timing at the two ends of the call. Delay can even affect one party's perception of the honesty, intelligence, or attentiveness of the other. For business calls, such misperceptions can cause serious problems, especially when

sensitive negotiation is involved and when callers who don't know each other well must rely on immediate impressions.

The ITU (International Telecommunications Union) recommends the delay limits shown in *Table 3*. Maintaining less than 150 milliseconds of one-way delay on connections prevents delay impairments.

**Echo impairment and control**

Echo in the network results from coupling between the transmit path and the receive path, which causes the outgoing speech to be sent back to the talker. The severity of an echo depends on two factors: the amplitude of the echoed signal and the time it takes to return to the talker. Amplitude is a function of the strength of the coupling between the transmit and receive channels. It is characterized as the "echo path loss," which is the difference in level (in dB) between the original input speech and the echoed signal. For a given echo path loss (i.e., constant level), the longer the time between the original speech and the returning echo, the louder the echo will seem.

Echo that is inaudible in the circuit-switched network may become noticeable with packet transmission because of the increased delay. Interconnections between packet networks and circuit-switched networks are especially

susceptible to echo impairment. Reflections from two- to four-wire hybrids used on analog lines in circuit-switched networks create strong echo; loss planning in the PSTN and rules for private networks connecting to the PSTN are largely intended to keep hybrid and other echoes below the threshold of audibility. The delay associated with packet transmission violates the engineering assumptions of the circuit-switched network. Therefore, echo control at the interface between the networks is essential to protect users at both ends from hearing echo.

While fully digital networks have no echo paths in the network, they can still be subject to echo from coupling in the end devices. Acoustic coupling, where the microphone picks up the output of the receiver, is one potential source. Electrical pickup between analog circuits (crosstalk) is another. Such echo is usually lower level than hybrid echo, but may be audible with long delay.

Following are descriptions of several echo-reduction techniques. They can be used alone or in combination, depending on the application and the level of echo expected.

**Echo cancellers**

An echo canceller is a device that looks for echo (a delayed signal on the return path that is strongly correlated with a signal seen on the incoming path) and

uses an adaptive filter to model the echo and then subtract it from the return signal. An echo canceller can improve the echo path loss of a connection by up to 26-30 dB with the adaptive filter. Any residual echo is removed using a non-linear processor, which removes all signals below a certain threshold.

### Echo suppressors

An echo suppressor or voice switch detects a signal on the incoming or outgoing path and switches attenuation into the other path to reduce the level of any returning signal. This suppression technique can be used in speakerphones, headsets, and wireless handsets, where acoustic coupling is common. Voice switching is a simpler function than echo cancellation, but is less transparent to the conversation dynamics and can add its own impairments to the speech signal.

### Loss/level planning

Echo control on calls with shorter delays can be managed effectively by introducing loss in the path. Loss planning is a key strategy in the management of network echo. In the circuit-switched network, the specified loss depends on the expected propagation delay. The loss is chosen to ensure that the received signal is audible, while at the same time echo is differentially attenuated (because the echo returns through the network, it “sees” the loss twice). Certain features such as volume controls or automatic gain control can defeat the loss plan by

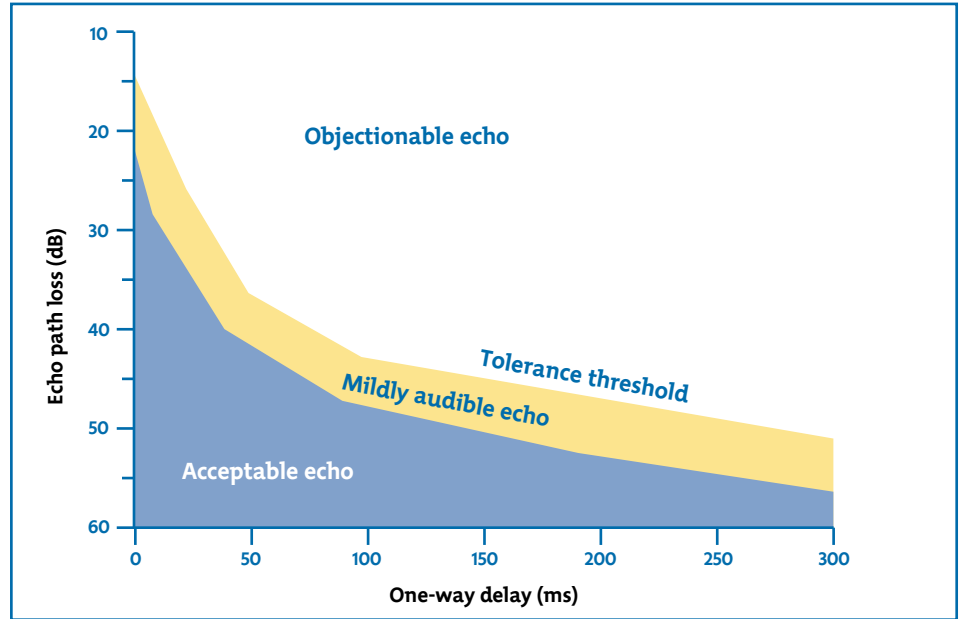


Figure 2. Talker echo tolerance as a function of delay

adding gain, which may make echo audible. Loss planning cannot prevent audible echo where delays exceed 20-25 ms one way. However, it is still necessary to control signal levels to ensure that echo control devices function properly. For instance, the level of any echo must be low enough that an echo canceller will not mistake it for direct speech.

Echo control is required at the interface between a packet network and a circuit-switched network where hybrids may be present. Echo cancellers deployed at the interface need to have tail coverage (the maximum round-trip delay of the echo path) of at least 48 milliseconds. This is because the loss plan protects circuit-switched trunks with up to 20-25 ms one-way delay (i.e., approximately 45 ms round trip) against echo without the need for cancellers. On longer trunks, the PSTN employs echo cancellers, which will protect the user at the packet end of the call.

Figure 2 shows how the level of the echo signal (vertical) and the delay in the echo path (horizontal) combine to determine the subjective degradation. The echo level is shown as echo path loss (the difference in dB between the original signal and the echo signal); note that as the numbers get larger, the echo gets quieter. These data were gathered in experiments where subjects were asked to rate the quality of connections with controlled amounts of delay and echo. The blue area across the bottom of the chart shows the region where the echo signal is not audible. The yellow area above it corresponds to mildly audible echo, which is tolerable to the user. Above that, the echo is objectionable. The relation shows that the level of the echo must diminish as the delay increases to achieve acceptable performance.

## Packet loss

In the traditional circuit-switched telephone network, a call is assigned a physical connection between end-points, and the circuit remains dedicated to that channel for the duration of the call. In contrast, packet networks break voice, fax, and data into small samples or packets of information. Each packet has a header that identifies where the packet is going and provides information for reassembly when the packet arrives at the destination. Packets travel independently and they are interspersed with packets from other network traffic along the way. Travel time through the network varies for individual packets.

Unless the network is precisely matched to the peak traffic load, packets sometimes fail to arrive at the destination. These lost packets create gaps in voice communications, which can result in clicks, muting, or unintelligible speech. In transmitting data, the remedy for packet loss is to resend the missing packets, but this solution doesn't work for time-sensitive voice conversations.

Generally, there are two ways to lose packets. They can be lost at network nodes because of an over-flow in the buffer or because a congested router deliberately discards them to reduce congestion. These packets are truly gone, and will never arrive at the destination. Network outages from disabled devices or fiber cuts can also result in lost packets. These events may result in large packet losses; these will be spread

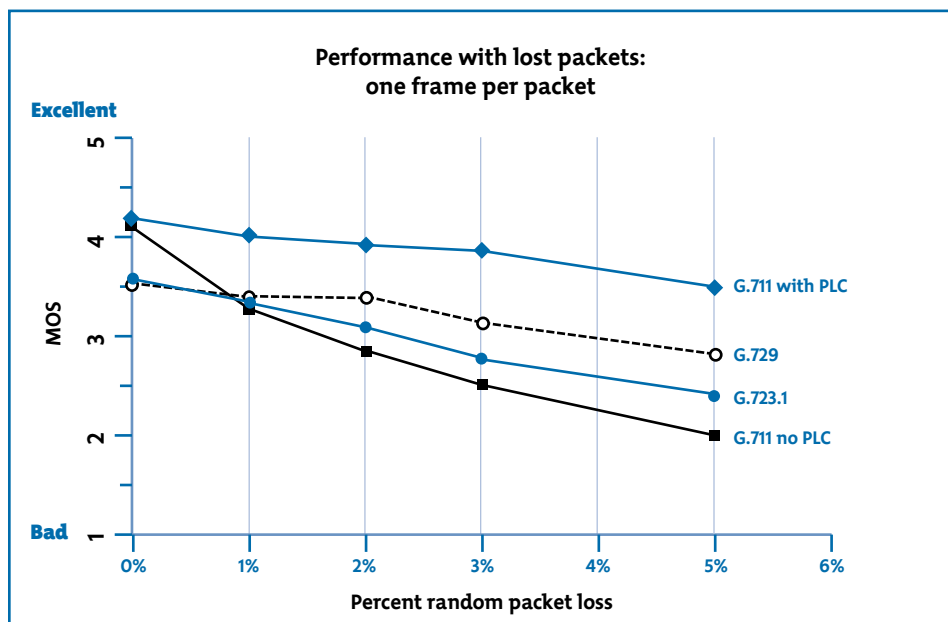


Figure 3. Packet loss effects for three common speech codecs (G.711 data based on 10-ms packets)

across the many different virtual channels that the network is handling at that time.

Second, packets can be delayed if they take a longer route or spend time in a device queue, causing variability in arrival time at the receiving end. The jitter buffer is used to smooth out the variability by holding packets for input to the decoder. The delay introduced by the jitter buffer is tuned to the expected network delay variation. That delay determines the longest time that a packet can take to arrive and still be in time to be decoded. Packets arriving after the prescribed delay lose their turn and are as good as lost, since the voice playout cannot wait for the late packets to show up.

In a network running without call admission control, and without a quality of service (QoS) protocol enabled, packet loss is uncontrollable in the face of congestion. The consequences of congestion depend on the

type of network, the proportion of voice and data traffic, the number of hops, and the duration of the event. The number of late packets can be minimized by increasing the size of the jitter buffer. However, a longer jitter buffer increases the end-to-end delay.

The Nortel Networks Subjective Assessment Laboratory in Ottawa has studied the perceived voice degradation associated with various rates of packet loss. The tests examined different codecs and different packet loss concealment options. Figure 3 shows user ratings of voice samples with packet loss ranging from 0 to 10 percent. Four curves are shown: G.711 without PLC, G.729<sup>9</sup>, G.723.1<sup>10</sup>, and G.711 with the T1.521 Annex B PLC algorithm<sup>5</sup>. The quality for G.711 without concealment drops dramatically as random packet loss increases. Where a packet loss concealment algorithm is used, the user satisfaction ratings remain

relatively high. G.729 and G.723.1 fall between G.711 with no PLC and G.711 with PLC. For voice quality to remain acceptable, the packet loss rate should remain low (1 to 2 percent or less).

### **Avoiding lost packets and missing data**

The best way to prevent late and lost packets is to engineer the network to preclude or minimize delays and other contributing factors. This means that congestion control (call admission control) must be in place to prevent the router queues from filling, which causes variation in delay and possibly overflow.

Strategies for minimizing lost packets are described below. Some require networkwide implementation; others can be used on a single channel to improve the quality on an individual call.

### **QoS protocols**

Implementing quality of service (QoS) protocols in the network devices expedites the transmission of voice packets at the various gateways and routers, minimizing jitter and its resultant lost packets. This is most effective if the network is carrying a substantial proportion of data traffic. If the network is carrying a high proportion of voice traffic, there may still be queuing delays in the routers, with associated increase in jitter and lost packets.

### **Call admission control**

In networks with a high proportion of voice traffic, call admission control can prevent congestion by limiting the number of calls that can be active through various nodes in the network. This is analogous to “fast busy” in the circuit-switched network. Where there is no call admission control and the number of calls increases above the recommended utilization, the quality of all calls in the network declines as the delay, jitter, and packet loss increase.

### **Adaptive jitter buffer**

When a voice packet arrives at the destination, it is held in the jitter buffer until the decoder is ready for it. Late packets (described above) are discarded. An increase in the packet loss rate at the decoder may mean that more packets are arriving late. An adaptive algorithm can be used to adjust the jitter buffer delay as the packet loss rate rises and falls. This adjustment helps minimize the number of late packets when the system is congested and avoids adding unnecessary delay when congestion eases. The buffer is adjusted during silent periods, so the temporal shift in the signal is transparent to users.

### **Interleaving**

Where multiple frames are sent in the same packet, interleaving of speech data across different packets helps minimize the effect of missing packets on output voice. An interleaved packet would contain multiple nonconsecutive frames. For example, if two frames are

sent in each packet, the first packet might contain frames 1 and 3; the next, frames 2 and 4; the third, frames 5 and 7; and so on. Because the frames in each packet are non-consecutive, the PLC algorithm can more effectively repair the gaps left by a missing packet. Interleaving increases the processing delay on the call, because frames are sent out of sequence. In cases where packet loss is due to buffer overflow and/or the discarding of packets at congested routers, interleaving is superior to simply lengthening the jitter buffer.

### **Sending duplicate data**

Sending redundant data also corrects for voice packet loss. For this remedy, information from a packet is copied into the next packet in the sequence and is used if the original packet is lost or delayed. With some codecs, such as G.729, even incomplete data can be useful in repairing the gap. Because decoding of the duplicate data must await the arrival of another packet if the original is lost, this method of loss suppression imposes extra delay.

### **Concealing missing data**

Packet loss concealment procedures can camouflage gaps in the output signal. The simplest techniques require little extra processing power, and the most sophisticated can restore speech to a level approximating the quality of the original. Concealment techniques are most effective for about 40-60 ms of missing speech; gaps longer than about 80 milliseconds are generally muted.

One of the basic PLC processes simply smooths the edges of gaps to eliminate audible clicks. Slightly more advanced algorithms replay the previous packet in place of a lost one, but this can cause harmonic artifacts such as tones or beeps. Good concealment methods use variation in the synthesized replacement speech to make the output more like natural speech. Better PLC algorithms preserve the spectral characteristics of the talker's voice, and maintain a smooth transition between the estimated signal and the surrounding original. The most sophisticated PLCs use CELP or a similar technique to determine the content of a missing packet by examining the previous one.

The G.729 and G.723.1 codecs have built-in PLC and their quality drops slowly with increasing amounts of packet loss. G.711 and G.726 have no component PLC algorithm, but an external one can be added when these codecs are used in a packet environment. A T1 standard has been adopted defining the PLC to be used with G.711<sup>5</sup>. There is no comparable standard for G.726, but Nortel Networks has developed a proprietary method for concealing missing packets of G.726. Note that when good packets are restored, G.711 recovers immediately, whereas G.726 (ADPCM) and CELP based codecs require a short time to readapt.

## Summary

Packet networks that carry voice must be carefully engineered and managed to ensure prompt packet delivery and minimal packet loss. Even in well-engineered networks, losses are inevitable during periods of congestion unless there are congestion control mechanisms in place. Strategies such as interleaving and redundancy may be used to help disperse the effects of packet loss. Packet loss concealment techniques are useful at the receiving end to smooth over missing data.

Delay over interactive voice channels can cause serious impairment. Delays of 10 to 150 milliseconds are transparent to the user, provided the appropriate echo control is in place. Longer delays can degrade the character of the interaction between the users, even where the appropriate echo control is present. Delay cannot be mitigated once it has been introduced into the path. The ITU provides guidelines for acceptable end-to-end delays and for acceptable echo performance as a function of delay<sup>4</sup>.

Echo control must be in place where packet networks are connected to circuit-switched networks. Echo cancellers at the interface must have sufficient tail coverage and echo path loss enhancement to protect against hybrid echo from the circuit-switched end. Echo control may also be needed where the end devices do not provide

sufficient separation between acoustic elements (acoustic coupling) or analog electrical paths (crosstalk). Speaker-phones and headsets often require echo control either in the device itself or in its interface to the network.

Modeling of network performance is helpful in understanding the interactions and tradeoffs among these and other performance parameters. Nortel Networks has developed tools to estimate the voice quality delivered by a particular network design, and to define the required packet layer behavior to achieve a given service quality for voice. The second paper in this series describes the modeling tools and process used by Nortel Networks to determine the performance requirements for packet network operating parameters.

## Endnotes

- <sup>1</sup> Requirements for the analog and digital performance of the end devices (e.g., telephone sets, computer soft-clients, and voice gateways between switched-circuit and packet environments) are similar to the requirements of ordinary analog and digital telephone sets, and are generally well understood. They include the transmission characteristics (sending and receiving levels, frequency responses, etc.), circuit noise limits, minimum terminal coupling loss, and so on. These requirements have been described in detail in TIA/EIA-810-A. The preparation of TIA 810 was sponsored by Nortel Networks and others in TIA Committee TR41<sup>2</sup>. If the end device does not meet the requirements defined in TIA 810, the end-to-end quality will suffer.
- <sup>2</sup> TIA/EIA-810-A (in press) Transmission Requirements for Narrowband Voice over IP and Voice over PCM Digital Wireline Telephones, Standard for Telecommunications: Telephone Terminal Equipment. Interim document number SP 4352-URV.
- <sup>3</sup> P.800: ITU-T (1993). Methods for subjective determination of transmission quality, Recommendation P.800 (08/96). Geneva: International Telecommunications Union.
- <sup>4</sup> G.114: ITU-T (1996). One-way transmission time. Recommendation G.114 (02/96). Geneva: International Telecommunications Union
- <sup>5</sup> ANSI T1.521 (2000). American national standard for packet loss concealment for use with ITU-T Recommendation G.711, and ANSI T1.521A (2001): Annex B to ANSI/T1.521. American National Standards Institute.
- <sup>6</sup> G.107: ITU-T (1998). The E-Model, a computational model for use in transmission planning. Recommendation G.107 (12/98). Geneva: International Telecommunications Union
- <sup>7</sup> G.711: ITU-T (1988). Pulse code modulation (PCM) of voice frequencies, Recommendation G.711. Geneva: International Telecommunications Union.
- <sup>8</sup> G.726: ITU-T (1990). 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), Recommendation G.726. (Includes Annex A, Extensions of Recommendation G.726 for Use with Uniform—Quantized Input and Output—General Aspects of Digital Transmission Systems.) Geneva: International Telecommunications Union.
- <sup>9</sup> G.729: ITU-T (1996). Coding of Speech at 8-kbit/s Using Conjugate-Structure Algebraic Code-Excited Linear-Prediction (CS-ACELP), Recommendation G.729. (Includes Annex A, Reduced Complexity 8-kbps CS-ACELP Speech Codec, and Annex B, Silence Compression Scheme for G.729 Optimized for Terminals Conforming to Recommendation V.70.) Geneva: International Telecommunications Union.
- <sup>10</sup> G.723.1: ITU-T (1996). Dual-rate speech coder for multimedia communications; Recommendation G.723.1. (Includes Annex A, Silence Suppression, and Annex C, Channel Coding Scheme for use in wireless applications.) Geneva: International Telecommunications Union



**In the United States:**

Nortel Networks  
35 Davis Drive  
Research Triangle Park,  
North Carolina 27709  
USA

**In Canada:**

Nortel Networks  
8200 Dixie Road,  
Suite 100  
Brampton, Ontario L6T 5P6  
Canada

**In Europe:**

Nortel Networks  
Maidenhead Office Park  
Westacott Way  
Maidenhead Berkshire SL6 3QH  
UK

**In Asia:**

Nortel Networks Singapore Pte Ltd  
151 Lorong Chuan #02-01  
New Tech Park,  
Singapore 556741

**In Australia:**

Nortel Networks Australia Pty Limited  
380 St. Kilda Road  
5th/6th Floor  
Melbourne, Victoria  
Australia 3004

For more information, contact your Nortel Networks representative, or  
call 1-800-4 NORTEL or 1-800-466-7835 from anywhere in North America.

<http://www.nortelnetworks.com>

\*Nortel, Nortel Networks, the Nortel Networks corporate logo, and DMS are trademarks of Nortel Networks. All other trademarks are the property of their owners.

Copyright © 2001 Nortel Networks. All rights reserved. Information in this document is subject to change without notice. Nortel Networks assumes no responsibility for any errors that may appear in this document.

**74007.25/09-01**