



Echo Analysis for Voice over IP

Version History

Version Number	Date	Notes
1	June 22, 2001	This document was created.
2	November 30, 2002	Added information on loudness.

Echo Analysis for Voice over IP defines echo and describes where it occurs in a voice network. It examines the basic aspects of echo analysis and describes the effects of various network elements on echo. This document also explains how echo is measured and how echo cancelers work to estimate and eliminate echo. It looks at customer and service provider expectations about echo, and explains how to configure gateways to minimize echo. It points out that the normal characteristics of packet-based networks might unmask preexisting problems in the time-division multiplexing (TDM)-based voice infrastructure. Finally, it outlines a process for locating and eliminating loud echos and long echos, and concludes with a real-life case study involving PBX-based echo in an international voice network.

This document contains the following sections:

- [Echo Analysis Overview, page 1](#)
- [Echo Analysis Basics, page 2](#)
- [Effects of Network Elements on Echo, page 6](#)
- [Echo Canceler, page 11](#)
- [Customer Expectations About Echo, page 18](#)
- [Service Provider Expectations About Echo, page 18](#)
- [Configuring Gateways to Minimize Echo, page 18](#)
- [Process for Locating and Eliminating Echos, page 20](#)
- [Echo Analysis Case Study, page 22](#)
- [Related Documents, page 27](#)

Echo Analysis Overview

In a voice telephone call, an echo occurs when you hear your own voice repeated. An echo is the audible leak-through of your own voice into your own receive (return) path.

This document discusses basic concepts applicable to echo analysis, explains echo cancellation, and provides a process for locating and eliminating echos.

Echo Analysis Basics

Every voice conversation has at least two participants. From the perspective of each participant, there are two voice paths in every call:

- Transmit path—The transmit path is also called the send or Tx path. In a conversation, the transmit path is created when a person speaks. The sound is transmitted from the mouth of the speaker to the ear of the listener.
- Receive path—The receive path is also called the return or Rx path. In a conversation, the receive path is created when a person hears the conversation. The sound is received by the ear of the listener from the mouth of the speaker.

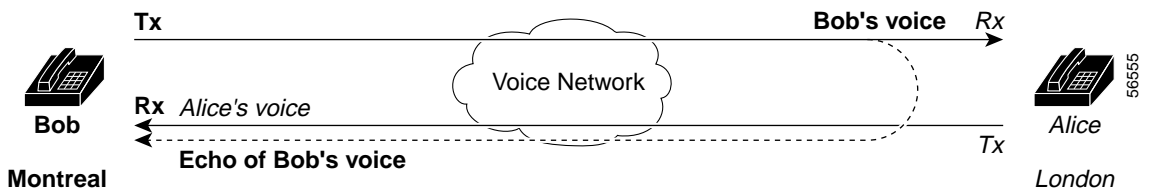
Figure 1 shows a diagram of a simple voice call between Bob and Alice. From Bob's perspective, the transmit path carries his voice to Alice's ear, and the receive path carries Alice's voice to his ear. Naturally, from Alice's side these paths have the opposite naming convention: the transmit path carries her voice to Bob's ear, and the receive path carries Bob's voice to her ear.

Figure 1 Simple Telephone Call



As mentioned, an echo is the audible leak-through of your own voice into your own receive (return) path. Figure 2 shows the same simple telephone call where Bob hears an echo.

Figure 2 Simple Telephone Call with an Echo



Bob hears a delayed and somewhat attenuated version of his own voice in the earpiece of his handset. Initially, the source and mechanism of the leak are undefined.

One significant factor in echo analysis is the round-trip delay of the voice network. The round-trip delay of the network is the length of time required for an utterance to go from Bob's mouth, across the network on the transmit path to the source of the leak, and then back across the network again on the receive path to Bob's ear.

Two basic characteristics of echo are as follows:

- The louder the echo (echo amplitude), the more annoying it is.
- The longer the round-trip delay (the “later” the echo), the more annoying it is.

Table 1 shows how delay can affect the quality of a voice conversation.

Table 1 *Effect of Delay on Voice Quality*

One-Way Delay Range (ms)	Effect on Voice Quality
0–25	This is the expected range for national calls. There are no difficulties during conversation.
25–150	This is the expected range for international calls using a terrestrial transport link and IP telephony, which includes only one instance of IP voice. This range is acceptable for most users, assuming the use of echo control devices.
150–400	This is the expected range for a satellite link. Delays in this range can interrupt the normal flow of a conversation. A high-performance echo canceler must be used and careful network planning is necessary.
Greater than 400	This is excessive delay and must be avoided by network planning.

Locating an Echo

In Figure 2, Bob experiences the echo problem, which means that a signal is leaking from his transmit path into his receive path. The fact that Bob hears an echo illustrates one of the basic characteristics of echo: Perceived echo most likely indicates a problem at the other end of the call. The problem that is producing the echo that Bob hears, the leakage source, is somewhere on Alice's side of the network (London). If Alice was experiencing echo, the problem would be on Bob's side (Montreal).

The perceived echo leak is most likely in the terminating side of the network for the following reasons:

- Leak-through happens only in analog circuits. Voice traffic in the digital portions of the network does not leak from one path into another.

Analog signals can leak from one path to another—electrically from one wire to another (called crosstalk), or acoustically through the air from a loudspeaker to a microphone. Also, analog signals can be reflected in the hybrid transformer in the tail circuit. (See the section “[Effect of Hybrid Transformers on Echo](#).”) When these analog signals have been converted to digital bits, they do not leak.

All digital bits are represented by analog signals at the physical layer and these analog signals are subject to leakage. The analog signals that represent bits can tolerate a good deal of distortion before they become too distorted to be properly decoded. If such distortion occurred in the physical layer of the PSTN, the problem would not be echo. If you had connectivity at all, you would hear digital noise instead of a voice echo.

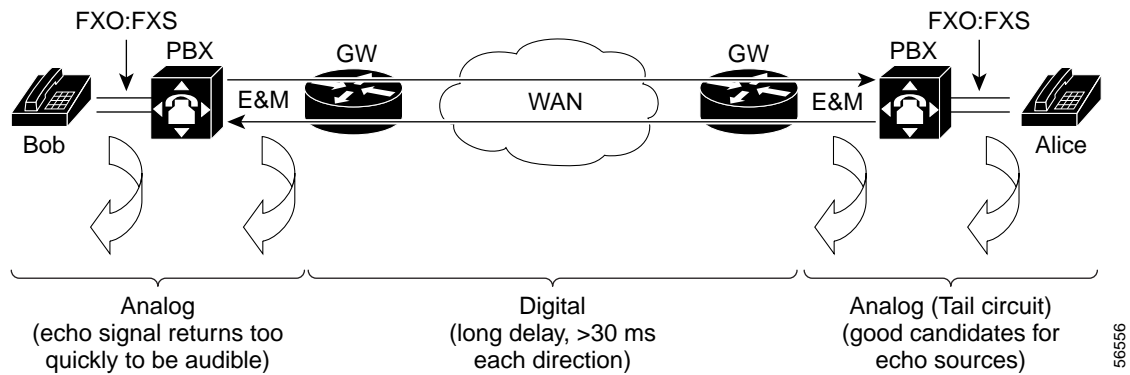
- Echos arriving after very short delays, about 25 milliseconds (ms), are generally imperceptible because they are masked by the physical and electrical sidetone signal.

This point is a corollary to the previous assertion that echos become increasingly annoying with increasing mouth-to-ear delay. A certain minimum delay is needed for an echo to become perceptible. In almost every telephone device, some of the Tx signal is fed back into the earpiece so that you can hear yourself speaking. This feedback is known as sidetone. The delay between the actual mouth signal and the sidetone signal is negligible, and sidetone is not perceived as an echo.

Also, your skull resonates during speech (an acoustic sidetone source) and the human auditory system has a certain integration period that determines the minimum time difference between events that will be perceived as separate events rather than a single one. Together, these phenomena create a minimum mouth-to-ear delay of about 25 ms before an echo signal can be perceived.

Given these two reasons—that echos must be delayed by at least 25 ms to be audible, and that leaks occur only in the analog portion of the network—you can deduce much about the location of the echo source. Figure 3 shows possible sources of echo in a simple Voice over IP (VoIP) network.

Figure 3 Potential Echo Paths in a Network with Both Analog and Digital Segments



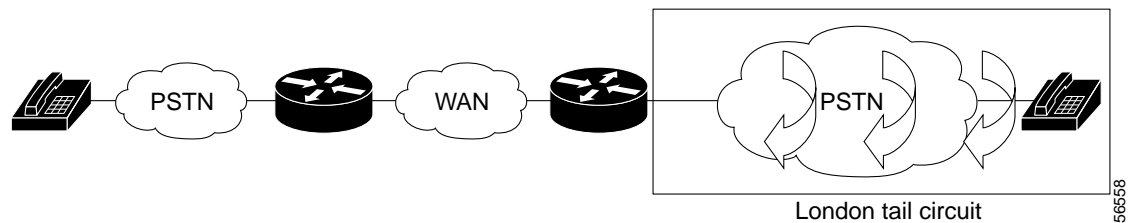
In this typical VoIP network, the digital packet portion of the network is located between two analog transmission segments. Bob in Montreal is connected by a 2-wire analog foreign exchange station (FXS) circuit to a local PBX, which is connected to a local VoIP gateway by a 4-wire analog receive and transmit (E&M) circuit. The Montreal gateway communicates with the London gateway through an IP network. As we will discuss later, this packet transmission segment has an end-to-end latency greater than 30 ms. At the London end of the call, the gateway is connected in the same way to Alice's telephone (by E&M to the PBX and by FXS to the terminal).

Tail Circuits

The analog circuit in London is known as the *tail circuit*. It forms the tail or termination of the call from the perspective of the user experiencing the echo—Bob in this case.

A packet voice gateway is a gateway between a digital packet network and a public switched telephone network (PSTN). It can include both digital (TDM) and analog links. The tail circuit is everything connected to the PSTN side of a packet voice gateway—all the switches, multiplexers, cabling, and PBXs between the voice gateway and the telephone. Figure 4 shows that the PSTN can contain many components and links, all of which are potential echo sources.

Figure 4 Echo Sources in the PSTN

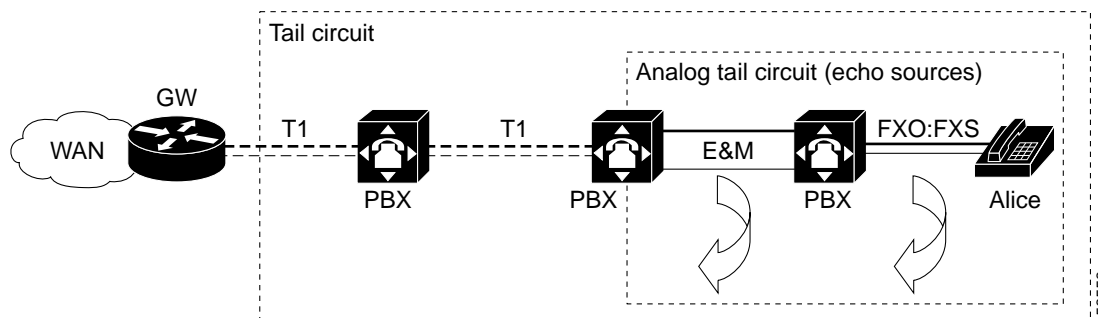


Packet voice gateways have two types of PSTN interfaces: digital (ISDN, BRI, T1/E1) or analog (E&M, FXO, FXS). Because bits do not leak, you can further refine your search for echo sources to the analog elements of the tail circuit. You can extend the echo-free digital zone out from the gateway to the point of digital-to-analog conversion in the PSTN, as shown in Figure 5.

56556

56558

Figure 5 Tail Circuit with Both Analog and Digital Links



Echo Sources

Suppose you want to locate potential sources of echo in the network in [Figure 3](#). You know that bits do not leak, so you can disqualify the digital segment of the system. Therefore, the leak causing Bob's echo must be located in either the tail circuit in Montreal or the tail circuit in London. Any leak in the Montreal tail circuit would not have a long enough delay to be perceptible—echos there would be masked by Bob's sidetone. So the source of the echo must be the London tail circuit.

Remember that an echo problem has three ingredients, as follows:

- An analog leakage path between analog Tx and Rx paths.
- Sufficient delay in echo return for echo to be perceived as annoying.
- Sufficient echo amplitude to be perceived as annoying.

The digital packet link in [Figure 3](#) requires a relatively long time for the analog signals entering this link to exit from the other side: the end-to-end delay of the link. This digital packet link delay comprises Tx, IP network, and Rx delays as follows:

- Tx delay includes primary processing delay (5 ms) and packetization delay (20 ms).
- IP network delays includes transport propagation delay, packet processing delay within the node, and serialization delay (depending on the interface choice).
- Rx delay includes jitter buffer delay (20 to 30 ms) and processing delay (5 ms).

Delay Calculation Example

The one-way transport delay for Bob's call from Montreal to Alice in London can be calculated by using the following two formulas.

- For PSTN fiber optic circuits:
National delay = 3 ms + (0.005 ms/km * distance in km)
- For PSTN international circuits:
Optic delay = transmitter delay + receiver delay + (0.005 ms/km * distance in km)

The distance from the Montreal local phone to the Montreal international gateway is 5 km, and the transport delay (national) is calculated (to the closest millisecond) as follows:

$$3 \text{ ms} + (0.005 \text{ ms/km} * 5 \text{ km}) = 3 \text{ ms}$$

The distance from the Montreal international gateway to New York City (Oceanic Cable) is 1000 km, and the transport delay (national) is calculated as follows:

$$3 \text{ ms} + 0.005 \text{ ms/km} * 1000 \text{ km} = 8 \text{ ms}$$

The distance from New York (Oceanic Cable) to London (Oceanic Cable) is 3000 km, and the transport delay (optic) is calculated as follows:

$$13 \text{ ms} + 10 \text{ ms} + 0.005 \text{ ms/km} * 3000 \text{ km} = 38 \text{ ms}$$

The distance from the London (oceanic cable) to the London international gateway is less than 1 km, so the transport delay (national) is negligible.

The distance from the London international gateway to the London local phone is only a few kilometers, so the transport delay (national) is about 3 ms.

The preceding calculations show a total one way transport delay of 52 ms. To this transport delay of 52 ms, you need to add the delays for codec packetization (20 ms) and the jitter buffer (20 ms). This calculation yields a total one-way delay of 92 ms. Therefore, the echo that Bob hears is delayed well into the range of audibility.

Effects of Network Elements on Echo

The following network elements in a VoIP network can have an effect on echo: loudness, hybrid transformers, telephones, routers, and quality of service (QoS). These elements are discussed in the following sections:

- [Effect of Loudness on Echo](#)
- [Effect of Hybrid Transformers on Echo](#)
- [Effect of Telephones on Echo](#)
- [Effect of Routers on Echo](#)
- [Effect of QoS on Echo](#)

Effect of Loudness on Echo

As mentioned previously in this document, the louder an echo is, the more annoying it is. To better understand the part that loudness contributes to echo, it is appropriate to examine a few characteristics of sound and its transmission.

Voice in the Air

Voice is carried over an air path in the form of acoustic pressure waves that are measured in units of Newtons per square meter or Pascals (Pa). Pressure is also expressed as units of sound pressure level (SPL), also known as the decibel (dB), commonly used to designate the loudness of sound and noise. The reference formula for SPL in dB is $20 \log(P/0.00002)$ where P is in Pascals. To provide a better evaluation of the human impression of the strength of a sound (the loudness), SPL is normally expressed in dBA, where A designates a frequency-weighted curve modeled on the frequency response of the human ear. To help you get a feeling for sound pressure levels (in dBA), see the approximate SPLs of some common sound sources shown in [Table 2](#).

Table 2 *Sound Pressure Level of Common Sounds*

Source	SPL (dBA)
Faintest audible sound	0
Whisper	20

Table 2 Sound Pressure Level of Common Sounds (continued)

Source	SPL (dBA)
Quiet residence	30
Soft stereo in residence	40
Speech range	50–70
Cafeteria	80
Pneumatic jackhammer	90
Loud crowd noise	100
Accelerating motorcycle	100
Rock concert	120
Jet engine (75 feet away)	140

Loudness Rating

The primary role of the phone set is to convert the air waves into electrical waves and vice versa. To allow an end-to-end evaluation of the voice loudness during a phone call conversation, engineers developed the concept of Loudness Ratings (LR), where LRs are defined as a set of weighted, unitless losses in dB for the acoustic and electrical conversion characteristics of a particular phone set. The engineers also created the concept of a fixed loss plan, or as it is commonly called, a Network Transmission Loss Plan (NTLP), that assigns fixed loudness losses to devices along the voice path. In the fixed loss plan tables, the LR values are actually the Equivalent Loudness Rating (ELR) values that include the loss of the analog loop, if applicable.

The following are four of the most common loudness characteristics used in VoIP circuit planning:

- Send Loudness Rating (SLR)—the loudness between the Mouth Reference Point (MRP) and the electrical interface.
- Receive Loudness Rating (RLR)—the loudness between the electrical interface and the Ear Reference Point (ERP).
- Overall Loudness Rating (OLR)—the total loudness loss between the MRP and ERP in a connection. OLR is calculated as follows:

$$\text{OLR} = \text{SLR}_{\text{talker}} + [\text{sum}]_{\text{attenuations}} + \text{RLR}_{\text{listener}}$$
- Talker Echo Loudness Rating (TELRL)—the loudness loss between the talker's mouth and the ear via the echo path. TELRL is calculated as follows:

$$\text{TELRL}(A) = \text{SLR}(A) + \text{loss in top path} + \text{ERL}(B) \text{ or } \text{TCLw}(B) + \text{loss in bottom path} + \text{RLR}(A),$$
 where ERL is the echo return loss of the hybrid or echo canceler, and TCLw is the weighted terminal coupling loss of the digital phone set.

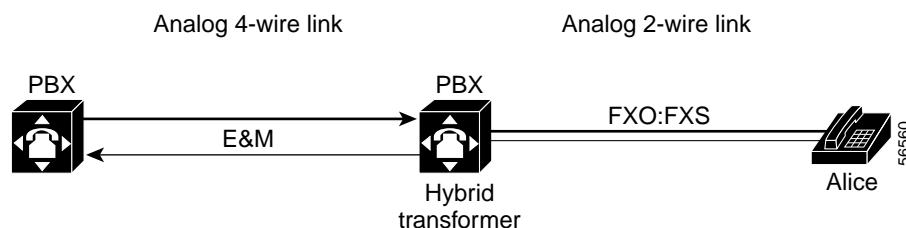
Effect of Hybrid Transformers on Echo

Echo sources are points of signal leakage between analog transmit and receive paths. Hybrid transformers are often prime culprits for this signal leakage.

The typical analog telephone terminal is a 2-wire device: A single pair of conductors is used to carry both the Tx and Rx signals. For analog trunk connections, known as 4-wire transmission, two pairs of conductors carry separate Tx and Rx signals. Digital trunks (T1/E1) can be considered virtual 4-wire links because they also carry separate Tx and Rx signals.

A hybrid is a transformer that is used to interface 4-wire links to 2-wire links. Figure 6 shows a hybrid transformer in an analog tail circuit.

Figure 6 *Detail of Analog Tail Circuit with a Hybrid Transformer*



Because a hybrid transformer is a nonideal physical device, a certain fraction of the 4-wire incoming (Rx) signal will be reflected into the 4-wire outgoing (Tx) signal. A typical fraction for a properly terminated hybrid in a PBX is about -25 decibels (dB), meaning that the reflected signal (the echo) will be a version of the Rx signal attenuated by about 25 dB. For a PSTN POTS termination, the expected value is between 12 and 15 dB. Echo strength is expressed in dB as a measurement called echo return loss (ERL). The ERL for the -25 dB reflected signal fraction equals $+25$ dB. The relation between the original source and the ERL is as follows:

$$\text{Original source amplitude} = \text{Echo amplitude} + \text{ERL}$$

Therefore, an ERL of 0 dB indicates that the echo is the same amplitude as the original source. A large ERL indicates a negligible echo.

Remember that an echo must have both sufficient amplitude and sufficient delay to be perceived. For local calls with one-way delay from 0 to 25 ms, an echo strength of -25 dB relative to the speech level of the talker, is generally quiet enough to not be annoying. For a one-way delay in the range of 25 to 150 ms, the ERL should exceed 55 dB to eliminate the perception of echo from the end-user perspective, as recommended in ITU-T Recommendation G.168 on echo cancellation. In these cases, echo cancellation is required.

The ERL is not a property of the hybrid alone, however. It depends on the load presented by the terminating device, which could be a telephone or another PBX. The hybrid has a certain output impedance that must be balanced by the input impedance of the terminating device. If the impedances are not matched, the returning echo fraction will be larger (the ERL will be smaller) and the echo will be louder.

You can expect a certain amount of impedance mismatch (a few tens of ohms)—a normal hybrid connection will yield ERLs in the range of 15 to 20 dB. However, one device could be provisioned for an output impedance of 900 ohms and the terminating device provisioned with an input impedance of 600 ohms, for example, which would yield a large echo, and would be expressed by a small ERL.

The main point to remember about hybrids is this: Ensure that output and input impedances are matched between the hybrid and the terminating device.

Effect of Telephones on Echo

As stated, the analog tail circuit is the portion of the PSTN circuit between the point of digital-to-analog (D/A) conversion and the telephone terminal. When digital telephones are used, this point of D/A conversion occurs inside the terminal itself, and the term weighted Terminal Coupling Loss (TLC_w) is used instead of ERL to represent the isolation between Rx and Tx on the analog portion of the digital phone. As a general rule, extending the digital transmission segments closer to the actual telephone will decrease the potential for echo.

The analog telephone terminal itself presents a load to the PBX. This load should be matched to the output impedance of the source device (FXS port). Some (inexpensive) telephones are not matched to the output impedance of the FXS port and are sources of echo. Headsets are particularly notorious for poor echo performance.

Acoustic echo is a major concern for hands-free speakerphone terminals. The air (and the terminal plastics) provides mechanical or acoustical coupling between the loudspeaker and the microphone. Speakerphone manufacturers combat this coupling with good acoustic design of terminals, directional microphones, and acoustic echo cancelers and suppressors in the terminal. However, this is a very difficult problem and speakerphones are inherently good echo sources. If you are hunting an echo problem and the terminating tail circuit involves a speakerphone, eliminate the speakerphone.

Effect of Routers on Echo

The belief that adding routers to a voice network creates echos is a common misconception. Digital segments of the network do not cause leaks; so technically, routers cannot be the source of echos. Adding routers (and IP/PSTN gateway routers) to the network, though, adds delay to the network—delay that can make a previously imperceptible echo perceptible. The gateway itself does not add echo unless you are using an analog interface to the PSTN and the output impedance is incorrectly provisioned with respect to the PBX. It is more likely that the echo was already in the analog tail circuit but was imperceptible because the round-trip delay was less than 50 ms.

For example, suppose you were visiting London and you wanted to call a friend who lives on the other side of town. This call is echo free. But when you call the same friend (whose telephone is on the same tail circuit) from the United States over a link with a much longer round-trip delay, the echo is obvious and annoying. The only change has been the insertion of delay.

VoIP technologies impose a fundamental transmission delay due to packetization and the buffering of received packets before playout at the receiving endpoint. This delay is generally much smaller than the delay associated with, for example, satellite links—but it is usually sufficient to make a previously unnoticeable echo objectionable.

There are other technologies that add delay, such as wireless, satellite, and long distance bypass over Frame Relay, ATM, or IP. These technologies require that echo cancellation capability be a part of their design.

End-to-End Voice Call Delays

Analog transmission is very fast, limited only by the propagation speed of electrons in a wire (which is lower than the speed of light, but still very fast) or photons in a fiber-optic link. TDM transmission is similarly very quick. A transcontinental PSTN call in the United States has a typical round-trip delay of about 10 to 20 ms. A local PSTN call has a shorter typical round-trip delay of only a few milliseconds. Such short delays mean that even relatively loud echos in the PSTN remain imperceptible as echo because they are masked by sidetone.

Imagine a call between Bob and Alice over a VoIP transmission link such as the one shown in [Figure 3](#). Consider the path Bob's voice takes from Montreal to London. Bob speaks into his mouthpiece and the analog signal arrives at the Montreal PBX within 1 ms. At the PBX, his analog voice signal is converted to a digital pulse code modulation (PCM) stream and arrives at the Montreal IP gateway after only 1 ms more of delay. So Bob's voice takes 2 ms to go from his mouth to the voice gateway. The gateway sends out packets every 20 ms, which means that each packet contains 20 ms of voice payload. Therefore, the voice gateway must wait to collect 20 ms of Bob's voice before it can fill the first packet. The first packet leaves the Montreal gateway 22 ms after Bob starts talking. Assuming that the WAN is very quick and uncongested, this first packet arrives at the London voice gateway after only 5 ms of transit. So the London gateway gets the packet 27 ms after Bob starts speaking.

This packet is not played out from the London gateway to Alice immediately upon receipt, however. The Montreal gateway delivers new packets faithfully at 20 ms intervals, but the vagaries of packet transmission mean that packets arrive in London at nonconstant intervals: Packet 2 might be 1 ms late, packet 3 might be 4 ms late, and so on. If the London gateway played out packet 1 immediately, it would be caught short 20 ms later when packet 2 was due but had not yet arrived—and Bob's voice would be interrupted.

The London gateway puts incoming packets into a buffer. The deeper the playout buffer, the longer packets will wait before being played. The minimum buffer depth you can safely use is one packet, or 20 ms in our case. So packet 1 arrives at time 27 ms and is played out to the London PSTN tail 20 ms later at time 47 ms. It takes 2 more milliseconds to go from the London gateway across the PSTN to Alice's earpiece, for a total of 49 ms for Bob's words to go from Bob's mouth to Alice's ear. This is the end-to-end delay of the voice transmission system: 45 ms in the WAN and 4 ms in the PSTN.

You could increase the packet transmission rate to reduce the end-to-end delay, but increasing the packet transmission rate would increase the bandwidth necessary for the call because it would increase the ratio of header size (which is a constant) to payload size (which you would reduce).

As a general rule, the end-to-end latency for a packet transmission link has a fundamental minimum of about two to three packet sizes (in milliseconds). Even if the packet transit time was instantaneous, one packet's worth of time still is required to fill the first packet. Even an unrealistically ideal, "fast-as-light" gateway and network face this fundamental, minimum delay.

Any echo source in the London tail circuit will send an echo all the way back across the WAN, facing another 47 ms of delay. The echo will return to Bob's earpiece after a round trip—almost 100 ms of delay—which is quite enough to make the echo audible.

Therefore, the use of a packet transmission link imposes an extra delay of at least two to three packet sizes that was not present before. Echos occur in the analog tail circuit, not the packet network, and they existed before any routers were added. Adding the delay makes the existing, inaudible echo an audible echo. The delay of the packet network cannot be reduced below a fundamental limit. Cisco voice gateways already operate very close to this minimum delay (50 to 80 ms end-to-end is typical). Because of these long delays, all VoIP gateways employ echo cancelers to reduce the amplitude of returning echos. However, the best solution to echo problems is always to remove the source of the echo.

In summary, remember the following effects:

- Network delay increases user annoyance for an echo of equal strength.
- Adding routers does not cause echo; it exacerbates existing echo problems.

Effect of QoS on Echo

QoS might improve your end-to-end network delay for a given level of congestion—the shorter the delay, the less annoying a given echo becomes. However, you will never be able to reduce the delay below the “danger zone” (25 ms) for echo perception with any form of QoS because the minimum delay inherent in VoIP networks is long enough for echos to be perceptible. QoS can help in other ways (for example, packet loss and jitter), but it cannot, by itself, eliminate echo.

Echo Canceled

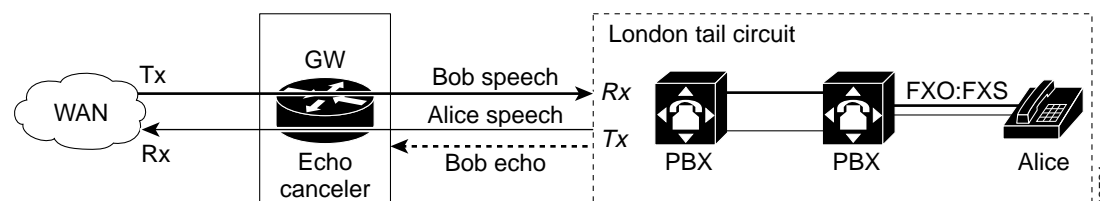
An echo canceler is a component of a voice gateway that reduces the level of echos that have leaked from the Rx path (from the gateway out into the tail circuit) into the Tx path (from the tail circuit into the gateway). Rx and Tx here are from the perspective of the voice gateway—London, in the case of the examples thus far described.

Echo cancelers have the following properties:

- Echo cancelers face into the PSTN tail circuit.
- An echo canceler eliminates echos in the tail circuit on its side of the network.

Figure 7 shows the location of the echo canceler in London.

Figure 7 *Echo Canceler in London Eliminates Bob's Echos in London Tail Circuit*



From the perspective of the echo canceler in the London voice gateway, the Rx signal is Bob's voice coming across the packet network from Montreal. The Tx signal is a mixture of Alice's voice and the echo of Bob's voice, which comes from the London tail circuit, and will be sent to Montreal.

The echo canceler in the London gateway looks out into the London tail circuit and is responsible for eliminating Bob's echo signal from the London Tx signal and allowing Alice's voice to go through unimpeded. If Alice were hearing an echo in London, the source of the problem would be in Montreal and the echo canceler in Montreal would eliminate it.



Note

Delay and jitter in the WAN do not affect the operation of the echo canceler because the tail circuit is static, and the tail circuit is where the echo canceler operates.

Basics of Echo Canceler Operation

The role of the echo canceler is to remove the echo portion of the signal coming out of the tail circuit and headed into the WAN. The echo canceler removes the echo by learning the electrical characteristics of the tail circuit and forming its own model of the tail circuit in its memory. Using this model, the echo canceler creates an “estimated echo” signal based on the current and past Rx signal (Bob's voice). Bob's voice is run through this functional model to come up with an estimate of what Bob's echo signal would

sound like. This estimated “Bob echo” is then subtracted from the actual Tx signal coming out of the tail circuit. The time that is required (after the user starts talking) to adjust the level of attenuation needed to the original signal is called the convergence time. Mathematically, the subtraction of the estimated “Bob echo” is performed as follows:

$$\begin{aligned}
 & \text{Tx signal sent from the gateway back to Bob} \\
 &= \text{Tx signal} - \text{“estimated Bob's echo”} \\
 &= (\text{Alice's voice} + \text{Bob's echo}) - \text{“estimated Bob's echo”} \\
 &= \text{Alice's voice} + (\text{Bob's echo} - \text{estimated Bob's echo}) \\
 &= \text{Alice's voice (if the estimation is accurate)}
 \end{aligned}$$

The quality of the estimation is continuously improved by monitoring the estimation error.

Echo Canceler Components

A typical echo canceler includes circuits that perform two functions: A convolution processor (CP) and a nonlinear processor (NLP). Each function includes multiple stages of signal processing, as described in the following sections.

Convolution processor

The convolution processor first stage captures and stores the outgoing signal toward the far-end hybrid. The CP then switches to monitoring mode and waits for the echo signal to return. When the echo signal returns, the CP must estimate the signal level of the incoming echo signal and then subtract the attenuated original voice signal from the echo signal. Of course, these processes are running full time when the near end talks, so it is much more dynamic than the description.

The time required to adjust the level of attenuation needed to the original signal is called the convergence time. Because this process requires that the voice signal be stored in memory, the EC has limited coverage of tail circuit delay, normally 64, 96, and up to 128 ms. After convergence, the convolution processor provides about 18 dB of echo return loss enhancement (ERLE). Because a typical analog phone circuit provides at least 12 dB of echo return loss (ERL), that is the echo path loss between the echo canceler and the far end hybrid, the expected permanent ERL of the converged echo canceler is about 30 dB or greater.

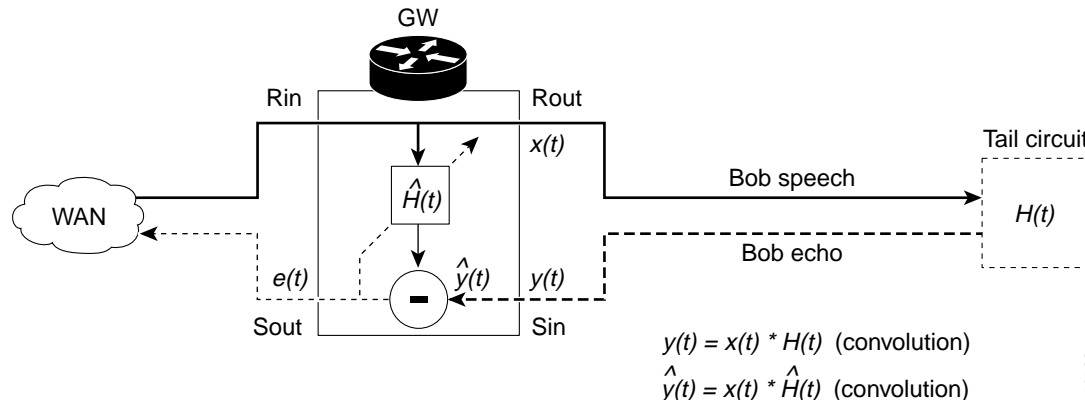
Nonlinear processor

In single-talk mode, the nonlinear processor replaces the residual echo at the output of the echo canceler with comfort noise based on the actual background noise of the voice path. The background noise normally changes over the course of a phone conversation, so the NLP must adapt over time. The NLP provides an additional loss of at least 25 dB when activated. The NLP should be active only in single-talk mode, that is, when one person is talking and the other is silent. In a double-talk situation, the NLP must be deactivated because it would create a one-way voice effect by adding 25 to 30 dB of loss in only one direction.

To completely eliminate the perception of echo, the TELR should be greater than 65 dB in all situations. To reflect this reality, ITU-T Recommendation G.168 regarding echo canceler requires an ERL equal to or greater than 55 dB. SLR, RLR and CLR along the echo path should allow another 10 dB to meet the expected TELR. As you can see, CP, NLP and LRs must be optimized to make sure that echo is canceled effectively.

Figure 8 shows a simplified version of echo canceler single-talk operation.

Figure 8 Echo Canceled Single-Talk Operation: Training



The key to echo canceler operation is that the tail circuit can be functionally represented by a mathematical formula. For the moment, assume that Alice is not talking. The tail circuit is a black box with an input (Bob's speech) and an output (Bob's echo). There exists a formula that describes the relationship between these two signals—a recipe for transforming the input signal into the output signal. If you knew what the formula was, you could simulate the black box in software. Then you could record the input signal and use the formula to predict what the output signal should sound like.

This prediction is precisely what an echo canceler does. Bob's voice signal, $x(t)$, enters the real tail circuit and emerges as the echo signal $y(t)$. The input-output relationship (impulse response) of the real tail circuit is $H(t)$. $H(t)$ is a mathematical representation of the transformation applied to $x(t)$ to obtain $y(t)$.

The echo canceler stores an estimate of this impulse response, denoted $\hat{H}(t)$. The echo canceler has access to the signal $x(t)$, Bob's voice, and runs this signal through $\hat{H}(t)$ to obtain a “virtual” echo signal $\hat{y}(t)$. This virtual echo is subtracted from the real echo, and the resulting signal $e(t)$ (error signal) is ideally zero. The echo is canceled.

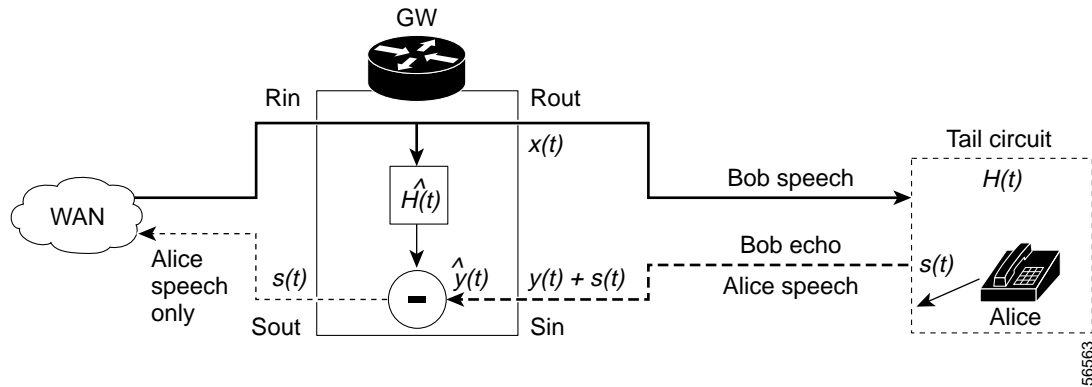
How does the echo canceler obtain the formula for $H(t)$? The simple answer is through trial-and-error. The precise answer is the use of a gradient descent algorithm to drive the coefficients of an adaptive finite impulse response (FIR) filter.

The echo canceler starts out with an all-zeros formula for $\hat{H}(t)$. Naturally, this is a very poor guess and the error signal $e(t)$ is large. A control method exists that allows the formula for $\hat{H}(t)$ to *wiggle* or adapt in a controlled fashion. If a wiggle causes the error to decrease, the formula keeps wiggling in that direction. If the wiggle causes the error to increase, the formula stops wiggling in that direction and starts wiggling in the opposite direction. Gradually the error decreases, the wiggles get smaller, and $\hat{H}(t)$ becomes a better and better estimate of the true $H(t)$. This period of wiggling is known as the *adaptation* or *convergence* period— $\hat{H}(t)$ wiggles until its formula converges on the true formula $H(t)$.

Alice is not talking in the previous example. If Alice is talking, the signal coming back from the tail circuit is a mixture of Alice's voice and Bob's echo. This condition is known as double talk. (See [Figure 9](#).) Double talk obscures the clean relationship of $H(t)$ that the formula is trying to estimate; therefore, convergence occurs only when Alice is silent. This does not mean that echo canceling stops. The whole point of converging is to provide a method of estimating Bob's echo signal. When Alice talks, the formula continues to generate echo estimates and subtract these estimates from the incoming signal. In this way, only the portion of the signal from Bob's echo is stripped out. Bob hears Alice's voice with no echo from his own speech.

[Figure 9](#) shows the operation of the echo canceler during double talk.

Figure 9 Echo Canceler Operation: Double Talk



Measuring Echo

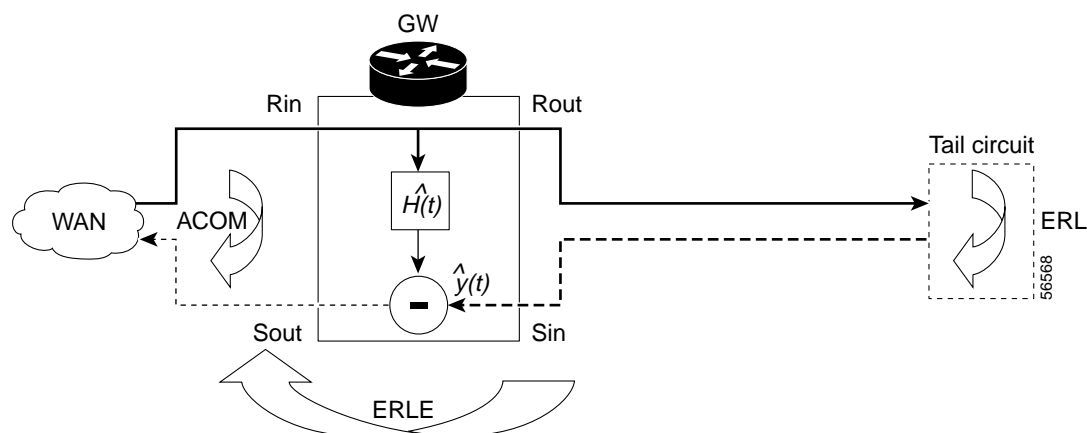
Following are descriptions of the primary measurements of relative signal levels used by echo cancelers. They are all expressed in dB.

- Echo return loss (ERL)—the reduction in the echo level produced by the tail circuit without the use of an echo canceler. So if an Rx speech signal enters the tail circuit from the network at a level of X dB, the echo coming back from the tail circuit into the Sin terminal of the echo canceler is $(X - \text{ERL})$.
- Echo return loss enhancement (ERLE)—the additional reduction in echo level accomplished by the echo canceler. An echo canceler is not a perfect device; the best it can do is attenuate the level of the returning echo. ERLE is a measure of this echo attenuation performed by the echo canceler. It is the difference between the echo level arriving from the tail circuit at the echo canceler and the level of the signal leaving the echo canceler.
- Acombined (ACOM)—the total echo return loss seen across the Rin and Sout terminals of the echo canceler. ACOM is the sum of $\text{ERL} + \text{ERLE}$, or the total echo return loss seen by the network.

56563

Figure 10 depicts the location of the three primary echo return loss measurements.

Figure 10 Location of ERL, ERLE, and ACOM



As shown in Figure 11, the echo return loss measurements are derived as follows:

ERL = Echo return loss through tail = $R_{out} - S_{in}$ (dB)

ERLE = Echo return loss enhancement through echo canceler = $S_{in} - S_{out}$ (dB)

ACOM = Combined echo return loss through system = $R_{in} - S_{out}$ (dB)

TELR Calculation Example

Following are examples of how TELR is calculated for three scenarios of a telephone conversation.

- Single talk operation (training)

TELR = talker phone SLR + WAN attenuation + tail circuit attenuation + tail circuit ERL + tail circuit attenuation + WAN attenuation + talker phone RLR

TELR = 8 dB + 0 + 0 + 12 dB + 0 + 9 dB – 6 dB = 23 dB

- Single talk operation (after convergence)

TELR = talker phone SLR + WAN attenuation + ACOM (ERLE + ERL) + WAN attenuation + talker phone RLR

TELR = 8 dB + 0 + 55 dB (CP:1 8 dB, NLP: 25 dB, local loop ERL: 12 dB) + 9 – 6 = 66 dB

- Double Talk Operation

TELR = talker phone SLR + WAN attenuation + ACOM (ERLE + ERL) + WAN attenuation + talker phone RLR

TELR = 8 dB + 0 + 30 dB (CP:18 dB, local loop ERL: 12 dB) + 9 – 6 = 41 dB

Echo Canceled Coverage

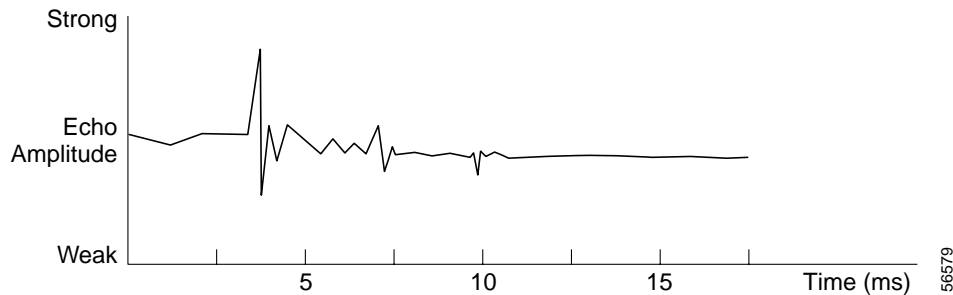
Echo canceler coverage (also known as tail coverage or tail length) specifies the length of time that the echo canceler stores its approximation of an echo, $\hat{H}(t)$, in memory. You can think of coverage as the echo canceler's cache. It is the maximum echo delay that an echo canceler will be able to eliminate.

Previously we noted that the echo canceler faces into a static tail circuit. A tail circuit has an input and an output. If a word enters a tail circuit (signal $x(t)$ in Figure 9), the echo (output signal $y(t)$ in Figure 9) is a series of delayed and attenuated versions of that word, depending on the number of echo sources and the delays associated with them. After a certain period of time, no more signal will come out. This time period is known as the *ringing time* of the tail circuit.

Think of the original echo source as a pebble tossed in still water and the echos as the series of attenuated ripples that the pebble produces. The ringing time is the time required for all of the ripples to disperse. Therefore, to fully eliminate all echos, the coverage of the echo canceler must be as long as the ringing time of the tail circuit.

Figure 11 is an example tail circuit impulse response. The peaks correspond to individual echos in the tail circuit. We see that this system has three echos: a strong one at about 3 ms and two weaker ones at about 7 ms and 9 ms. After about 12 ms, there is no significant energy in the impulse response. The amplitudes of the peaks correspond to the strength of the echo—the higher the peaks, the stronger the echo, and the smaller the ERL.

Figure 11 Example Tail Circuit Impulse Response $H(t)$



To cancel all three echos, you should provision an echo canceler facing into such a tail circuit for at least 12 ms of tail coverage. An echo canceler with 5 ms of coverage would perform fairly well with this circuit because the primary echo falls within the 5 ms window. The second two echos, though, would remain uncanceled because the echo canceler would discard its approximation of those echos from its memory.

It is important to stress again that the echo canceler faces into a static tail circuit—it eliminates echos in its own tail circuit that would be experienced by callers on the other end of the network. Echo cancelers *are not aware of the rest of the network*; therefore, tail coverage has nothing to do with the WAN, the round trip delay, or whether the network delay is changing.

Many people assume incorrectly that the long delays associated with VoIP require that the echo cancelers have equally long tail coverage. However, *only the tail determines the needed coverage*. Remember that analog transmission is very quick—almost all simple tails ring for only a few milliseconds. Longer ringing times exist when the tail is very complex (for example, a large number of PSTN hops, multiple D/A conversions), or when it contains long-distance trunks. If the tail of your VoIP system contains another VoIP link, then your tail will be far too long to cover. In that case, the embedded VoIP link requires its own echo canceler on its own tail.

We recommend that you avoid such embedded VoIP links. We suggest that you provision all your echo cancelers to their maximum tail coverage all the time.

Uncancelable Echo

An echo is uncancelable under the following conditions:

- The echo is too loud. The echo canceler algorithm cannot detect the signal as an echo, so it provides no attenuation.
- The echo is delayed beyond the time window of the echo canceler. Tail circuits that involve multiple PSTN hops, some long distance trunks, and alternating series of digital and analog links can have ringing times that exceed the tail coverage window.
- The echo is acoustically distorted. The echo canceler cannot recognize a distorted signal as echo.

Verifying Echo Canceler Operation

The quickest way to determine if you have a working echo canceler in the circuit is to make a call and immediately begin to say something like “TAH TAH FISH” repeatedly. The person on the other end of the line should be silent. If you are calling a voice-mail system, wait for the announcer to stop talking before starting the experiment.

If the terminating tail circuit has cancelable echos and if the echo canceler is enabled, you will hear echo for the first few utterances and then it will die away. After a few seconds of speech, the echo should be gone or at least very quiet compared to the echo level at the beginning of the call. This is the signature of a working echo canceler. Recall that an echo canceler starts out with no knowledge of the tail circuit that it is looking into. It needs to observe a certain amount of speech and echo flowing through the tail circuit to form the virtual tail circuit model. This training period is known as the *convergence time* of the echo canceler. You should expect convergence within the first few seconds of active speech.

If you try this experiment and do not obtain echo reduction with time, there are two possible explanations: The echo canceler is disabled or broken, or the echo is uncancelable (either too loud or delayed beyond the tail coverage of the canceler). Try making calls to other destinations and looking for the standard “echo die away” behavior.

The surest way to determine if your echo canceler is working is to first run the previously described test when the echo canceler is off, and then again when the echo canceler is on. If you do not find the standard “echo die away” behavior, then perform following steps to determine if your echo canceler is working:

1. Telnet to the destination voice gateway and check the provisioning of the voice ports (POTS). (Remember, the echo canceler you are interested in is the echo canceler in the destination voice gateway.)
2. Disable the echo canceler by issuing the **no echo-cancel enable** voice-port configuration command, then shut down and reopen the voice port by issuing the **shutdown** and **no shutdown** commands.
3. Make a call to a destination telephone and listen for echo by saying something like “TAH TAH FISH.” If you do not hear any echo, try different destination phones until you do. When you have found an echo that persists throughout the call, save the destination number.
4. Re-enable the echo canceler by using the **echo-cancel enable** voice-port configuration command, set coverage to maximum by using the **echo-cancel coverage** voice-port configuration command, and shut down and reopen the voice port. You should hear the echo die away within the first few seconds of speech. If the echo persists, the problem is in your echo canceler.

If the echo diminishes but is still noticeable, try to locate the source of the echo path and eliminate the echo. Clearly, the echo canceler is working but it is unable to give sufficient ERLE. Occasionally, tiny bursts of echo might emerge during the conversation, especially if the talker makes a quick, loud, bursty sound. This is normal echo canceler behavior. If these types of echos are loud enough to be unacceptable, you need to identify and eliminate the source of the echo in the tail circuit.

Customer Expectations About Echo

Because of the fundamental delays associated with VoIP technologies, existing echos will be more annoying than with TDM, and even the normal operation of an echo canceler will be more apparent. Customers of VoIP networks need to be educated to expect the standard echo canceler operation previously described so that they do not confuse these types of echos with abnormal echos. Abnormal echos persist throughout a call and do not fade.

Service Provider Expectations About Echo

Echo problems are relatively rare in the PSTN, which has short delays; they are much more common over cellular and satellite long distance calls. Interestingly, they are also much more readily tolerated in cellular and long distance calls because customers have been educated to have lower expectations for such calls.

As long as VoIP calls continue to be terminated in analog tails, echo will be a problem. One major obstacle to widespread VoIP implementation is that many tail circuits have preexisting delays that will become noticeable only when service providers introduce digital segments to the networks.

These problems will gradually be solved as digital networks extend out toward homes and telephone endpoints. Until then, how much echo can be expected? One call in 50? 100? 1000? Even if customers are trained to complain only when an echo problem is persistent and repeatable, a service provider cannot hunt down and destroy every echo complaint. No one has sufficient resources to do this task, and hunting down an echo is a necessarily intrusive process.

The challenge is to determine when an echo complaint is both solvable and worth solving. You know that the echo source is in the destination tail circuit. For an echo problem to be solved, the tail circuit needs to be accessible.

In an enterprise application where the PBXs are in the basement, for example, it is relatively easy to solve echo problems by examining levels and impedances in the customer PBX. Look for consistency and commonality in the echo problems. If every call going through a particular PBX or transmission link exhibits echo, then you can concentrate on that particular link. That is a problem worth solving. If you receive an isolated echo complaint for a particular destination phone number in the PSTN that does not share any links with other echo complaints, then you might find yourself hunting down a single telephone echo complaint, which is usually not worth the resources.

The goal of service providers in eliminating echos, therefore, is to identify clusters of echo complaints, look for common links, and then fix the echos. A PSTN has many *dirty tails*, and it is unrealistic to expect that every echo can be eliminated. The best that you can do is to make sure that your own network and tails are clean, which requires care in installation and provisioning, especially when connecting gateways to analog equipment.

Configuring Gateways to Minimize Echo

Echos live in the analog tail circuit, not in the gateway. The gateway has an echo canceler that can attenuate manageable echos, but gateways cannot affect the root causes of the echo problems. An echo on a gateway can be fixed in only three ways:

- Ensure that the echo canceler is enabled with maximum coverage.
- Match output impedances and levels with the analog telecom equipment attached to the gateway analog voice ports.

- Provision the network in compliance with a fixed loss plan for IP voice gateways.

You must adjust the audio levels of analog voice ports to comply with a fixed-loss plan, whenever possible. You can adjust the audio level of either the outputs or the inputs of a voice port on a gateway.

Lowering the Sin input audio level (also called *increasing the input attenuation* or *adding a loss pad*) correspondingly decreases the level of any echos (increases the ERL of the tail). However, lowering the Sin input audio level also decreases the audio level of the Tx speech signal for every call (Alice's voice in this example).

Similarly, lowering the Rout output audio level correspondingly decreases the level of any echos, but also decreases the audio level of the Rx speech signal for every call (Bob's voice in this example).

In the end, you could be helping the echo canceler for calls to tails with poor ERL, but hurting voice quality by reducing levels for *all* calls through that particular voice port. You should adjust audio levels to alleviate echos only as a temporary workaround while you attempt to eliminate the echo source in the tail circuit.

Fixed Loss Plan for IP Voice Gateways

Fixed loss plans were introduced a long time ago to help maintain received speech loudness at an optimal listening level for the conversation for any type of voice terminal and network equipment. A section of Part 68 of Federal Communications Commission (FCC) rules and regulations enforces a loss plan for any components connected to the PSTN. TIA/EIA/TSB122-A specifies the fixed loss plan for voice gateways and satisfies the Part 68 document. TIA/EIA/TSB122-A was based on TIA/EIA/464-C, which specifies the fixed loss plan for PBXs.

TIA/EIA/TSB122-A provides loudness loss tables based on the type of endpoints, as described in [Table 3](#).

Table 3 TIA/EIA/TSB122-A Loudness Losses for Endpoints

Port Definition	Voice Gateway Designation	PBX Designation
Analog On Premise Station	ONS	ONS
Analog Off Premise Station	OPS	OPS
Digital Station	DGS	DGS
Wide Area Network (IP Connection)	WAN	—
Digital Access Line (to Digital CO)	DAL	DAL
Foreign Exchange Office (to Analog CO)	FXO	AAL(A)
Foreign Exchange Digital (to Analog CO)	FXD	AAL(D)
Analog Tie Trunk	ATT	ATT

The most common loudness losses used in IP telephony planning are described in [Table 4](#).

Table 4 Loudness Losses Commonly Used in IP Telephony Planning

Port to Port	IP Gateway Loss	PBX Loss	FCC Loss Limits
ONS to ONS	6 dB	6 dB	—
ONS to ATT	3 dB	3 dB	-3 dB to 3 dB

Table 4 Loudness Losses Commonly Used in IP Telephony Planning (continued)

Port to Port	IP Gateway Loss	PBX Loss	FCC Loss Limits
ATT to ONS	3 dB	3 dB	—
ONS to DAL	3 dB	3 dB	-3 dB to 3 dB
DAL to ONS	9 dB	9 dB	—
OPS to DAL	0 dB	0 dB	—
DAL to OPS	6 dB	6 dB	—
ONS to WAN	0 dB	—	—
WAN to ONS	9 dB	—	—
ATT to WAN	-3 dB	—	—
WAN to ATT	3 dB	—	—
DAL to WAN	0 dB	—	—
WAN to DAL	0 dB	—	—
DGS to WAN	0 dB	—	—
WAN to DGS	0 dB	—	—


Note

Losses involving analog ports should be implemented before encoding or after decoding. Digital padding should be avoided wherever possible, because it could increase quantization distortion.

Process for Locating and Eliminating Echos

Before we look at the process for eliminating echos in the tail circuit, we will summarize the process for dealing with echos in general:

1. Identify which tail circuit is causing the echo. Remember, the echo is caused by the tail circuit on the opposite side of the network from the caller hearing the echo.
2. Check for speakerphones or headsets. If the destination telephone is a speakerphone or headset, this is probably the source of the echo. Try replacing the speakerphone or headset with a better quality handset and see if the echo dies away normally.
3. Telnet to the destination voice gateway and check that the echo canceler is enabled and that the coverage is set to maximum.
4. Test for normal echo canceler behavior as described in the earlier section, “Verifying Echo Canceler Operation.”

If the echo persists, and you have verified that the echo canceler is working properly, you can conclude that the echo canceler cannot fix the echo for one of two reasons:

- The echo is too loud (called a *loud echo*).
 - The echo is too delayed (called a *long echo*).
 - The echo is acoustically distorted.
5. Identify which type of echo you are experiencing, either long or loud.
 6. Locate and mitigate the echo source.

After you have verified that the echo canceler is working properly, you still need to determine the cause of the echo: Is the problem insufficient ERL in the tail, or is the echo delayed beyond the coverage of the echo canceler? Most persistent echos are loud echos. Delayed echos are common, however, when the tail circuit involves a long-distance PSTN link, a series of alternating digital and analog links, or any other link with high latency.

Identifying a Loud Echo

You can use the voice gateway itself to measure the ERL of the tail circuit by using the gateway echo canceler statistics reporting function. For a Cisco VoIP gateway, output from the **show call active voice EXEC** command contains valuable statistics.

To generate these statistics, first establish a voice call over the gateway. Then type the **show call active voice EXEC** command without pressing the Return key. Make a loud continuous sound into the mouthpiece or hold down a button on your touch-tone keypad to generate a sound, and then press Return to display the call statistics.



Note

You can also use commercial test devices (including handheld telecom level meters) to measure ERL for a particular destination circuit.

Remember, you need to look at the *destination* voice gateway. [Figure 10](#) shows that the ERL is the difference in the reported Tx and Rx levels. Ideally, you want your gateway to have an ERL of at least 15 dB. If your ERL is less than 10 dB, you probably have insufficient ERL in the tail circuit. Repeat the test outlined in the previous paragraph using louder and softer noises and verify that the ERL is consistent and that when you vary your volume, the levels vary accordingly. If these tests are consistent, you can be confident that the tail circuit is not providing enough echo loss for the echo canceler to be able to eliminate the echo.

Identifying a Long Echo

You can identify a long echo problem with a technique similar to the one described previously for loud echos. The signature of a loud echo problem is that the echo is somewhat attenuated but still noticeable. The echo is the same regardless of whether the echo canceler is enabled. If you determine that the ERL is reasonable (greater than 10 dB) but the echo is still persistent, then the problem might be a long echo.

If the problem is a long echo, there is not much that you can do to solve it. If the tail includes a long distance hop, make sure that the PBX terminating the long distance hop has its own echo canceler turned on. If possible, extend the digital portion of your network as close as possible to the endpoint.

Locating and Eliminating Echos in the Tail Circuit

Because of the variety of possible network scenarios, it is difficult to give specific instructions for finding and eliminating an echo in a tail circuit. There are some general steps that you can perform, however, to track down the source of an echo and eliminate it.

Draw a diagram of the tail circuit, including all the digital and analog links between the destination voice gateway and the destination telephone. This diagram will likely form a tree; from the voice gateway out, each device will have one or more potential destination branches. You need to identify the break point off the main branch for which calls give consistent echo.

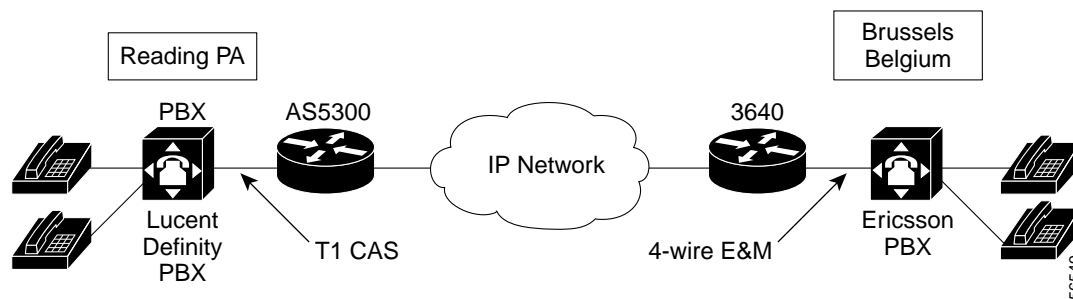
For example, the gateway might be connected to a PBX with three output cards. If many of the calls through one of these ports exhibit echo, then you have narrowed the source of the problem tail to the circuits attached to that voice port. Look for clusters of echo associated with common links. If you trace your tail out to the uncontrolled PSTN, then remember that a certain percentage of PSTN tails always will not provide sufficient ERL and you will be unable to correct them. When you find a link that is giving insufficient ERL, examine the levels and provisioning of the devices at both ends of the link.

Echo Analysis Case Study

The following case study describes how Cisco worked with an enterprise customer to eliminate echo in a VoIP network. The customer is a large manufacturing company with headquarters in Reading, Pennsylvania, and several plants in the United States and overseas. One plant, located in Brussels, Belgium, previously used the PSTN for intersite calling, which resulted in high toll charges. Because the customer already had a data network in place, the logical choice was to implement a combined voice and data network. Because traffic at the headquarters was required to cross the Ethernet backbone to the PBX, the customer decided to use IP for voice traffic. It was calculated that the customer would save \$3000 a month by installing three voice trunks across the data infrastructure.

Figure 12 shows the network topology between the headquarters and the remote site in Brussels.

Figure 12 Case Study Customer Topology



The Brussels site has 4-wire E&M trunks connected from an Ericsson PBX to the Cisco 3640 router. In Reading, a Cisco AS5300 access server is connected to a Lucent Definity GR3 PBX. All the proper QoS considerations and dial plan configurations were discussed and properly planned, and will not be discussed in this document.

Echo Problem Description

When the voice and data network was first implemented, users experienced substantial echos, so the customer reverted to using the PSTN for calls between the headquarters and the Brussels site. The customer initially believed that the Cisco routers were causing the echo, but we explained that our routers function like a 4-wire circuit and that it was not possible for leakage between the two voice paths to create echo.

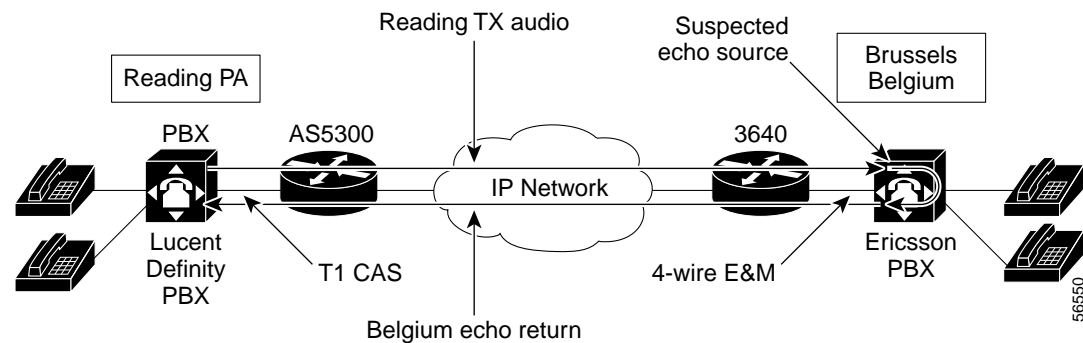
After testing calls between headquarters and Brussels, we noticed large amounts of echo and determined that the echo was being heard only on the headquarters end of the calls; therefore, the source of the echo was in the Brussels tail circuit—between the Cisco 3640 router and the telephone in Brussels.

Initially, we thought this might be a case of loud echo, which means an echo caused by insufficient ERL in the tail circuit. We ruled out the possibility of a long echo—an echo delay longer than the echo canceler's coverage. The Cisco 3640 had echo cancelers active on the Brussels tail circuit, and the

Brussels tail was connected only to the PBX. Long echo was not a possibility because the PBX would not cause a delay long enough to cause long echo. If calls from headquarters were dropping off the Brussels PBX or being routed to a third destination, long echo could then have been a possibility.

Eventually we discovered that a hybrid in the tail circuit was converting signals from 4-wire to 2-wire. Hybrids can be a common echo source. [Figure 13](#) shows how the hybrid was deployed in the customer's network.

Figure 13 Echo in Customer Topology



We explained to the customer that the echo problem probably existed before VoIP was implemented but that it had not been perceivable because the PSTN delay was below the noticeable threshold. Packet-based networks create some small delays (as a result of packet encoding, queueing delays, and jitter buffers) that might unmask preexisting echo problems. This condition is normal and is characteristic of a packet-based network.

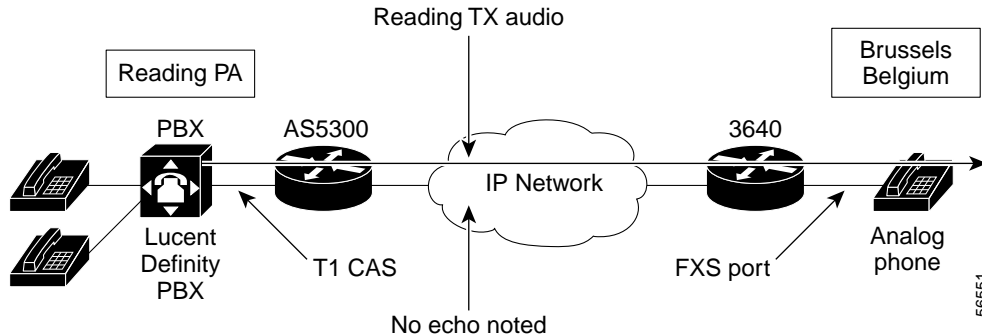
We set out to resolve the echo issue by proving that the problem was the PBX in Brussels and by proposing a solution to eliminate the echo problem. We looked at the following issues:

- Source of the echo
- Audio levels of the PBX
- ERL of the PBX
- Impedance settings

To thoroughly check the network, we ordered a commercial test set for the Brussels site. Before the test set was delivered, we ran a simpler preliminary test. We had an FXS module shipped to the customer site in Brussels from the local Cisco Technical Assistance Center (TAC). We instructed the customer's on-site personnel to install and configure the FXS module in the existing Cisco 3640 to allow calls from the FXS port on the Brussels 3640 to the PBX in Reading. When we established calls between the Brussels 3640 and the PBX in Reading, there was no perceivable echo and the quality was very clear.

This test indicated that if the 4-wire to 2-wire conversion occurred on the router (as opposed to the Ericsson PBX), no echo was present. Therefore, the Ericsson PBX was most likely causing the echo. The simplest solution to such an echo problem would be to connect only FXS ports from the Cisco 3640 into the PBX. This configuration would allow the router to perform the 4-wire to 2-wire conversion, and the FXS ports would appear as central office (CO) trunks to the Ericsson PBX. Although this configuration would not provide as much flexibility as the 4-wire E&M trunks, it would not take away any functionality from the customers because they used an auto-attendant. [Figure 14](#) shows this FXS test configuration.

Figure 14 FXS Test Configuration



Eliminating the Echo

After our test set arrived, we arranged to have a Cisco representative in Reading and an Ericsson representative on site in Brussels. The process to eliminate the echo was as follows:

1. Verify proper impedance levels on the Ericsson PBX in Brussels.
2. Verify proper audio levels.
3. Measure the ERL of the Ericsson PBX.

Verifying Proper Impedance Levels

The Ericsson representative verified that the impedance of the 4-wire E&M circuits was set for 600 ohms, which matched the configuration on the Cisco 3640.

Verifying Proper Audio Levels

We verified proper audio level settings from the Reading site to the Brussels site. The test set had the ability to connect to the Lucent PBX like any 2-wire analog phone; it also had a dial pad that allowed our test set to initiate a call to Brussels. After we established a call to Brussels, we injected a 1004-Hz tone at 0 dB into the Lucent PBX. We then measured the audio levels at various points along the voice path. These levels were verified in accordance with Cisco audio guidelines.

We entered a **show call active voice EXEC** command on the Reading router to verify the audio levels. The level on the Reading router measured -3 dB, which was the correct level according to the Cisco guidelines.

If the levels had needed to be adjusted, we would have entered the **input gain** voice-port configuration command. For example:

```
voice-port
  input gain 3
```

This command increases the level into the VoIP network by 3 dB. For these input gain changes to take effect, you need to hang up and reestablish the call.

After we verified the proper audio settings on the Reading router, we entered a **show call active voice EXEC** command on the Cisco 3640 in Brussels. This router displayed a -7 dB audio setting heading toward the Ericsson PBX. Even though the -7 dB level itself was acceptable, the optimal level is -12 dB at the phone on the PBX because different PBXs have different loss levels. Figure 15 and Table 5 depict the level adjustment configuration and the levels that were seen.

Figure 15 Audio Level and Echo Test Setup

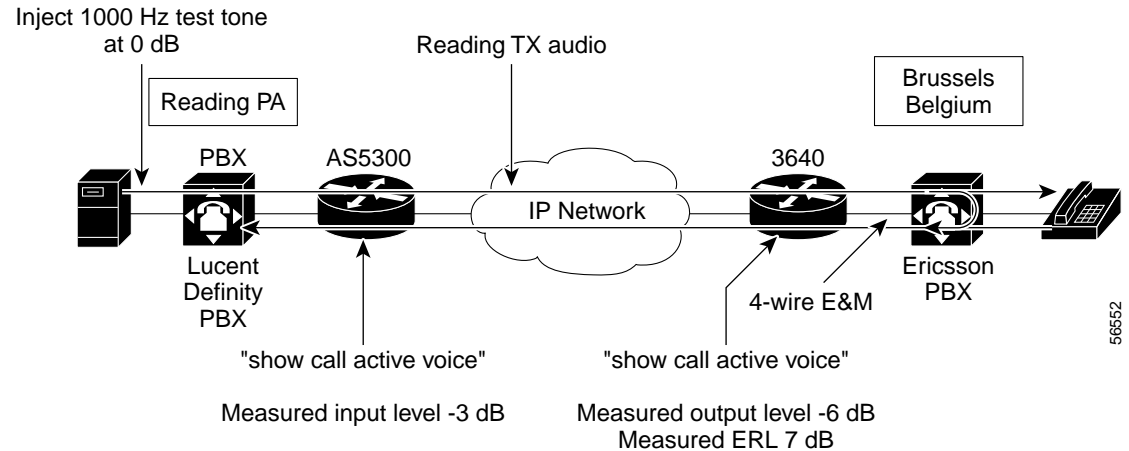


Table 5 show call active voice Command Output

Reading AS5300	Brussels 3640
Reading# show call active voice	Brussels# show call active voice
CoderTypeRate=g729r8	CoderTypeRate=g729r8
NoiseLevel=0	NoiseLevel=0
ACOMLevel=0	ACOMLevel=0
OutSignalLevel=-79	!This is the output level, R(out)
!This is the input level	OutSignalLevel=-7
InSignalLevel=-3	!This is the input level, S(in)
	InSignalLevel=-14
	InfoActivity=2
	ERLLevel=7
	!ERL = R(out) - S(in)
	!ERL = (-7) - (-14) = 7 dB
	!ERL should be > 15 dB

Measuring ERL

Because the audio levels were acceptable to the customer, we did not adjust them. However, we did raise and lower the audio levels during the ERL test. We sourced a tone from Reading and measured the echo on the Cisco 3640 router in Brussels. (Note that you do not need an official test set for echo testing. You can use DTMF tones or your own voice to get an approximate indication of level mismatches.)

We applied the same 1004-Hz tone at 0 dB to the Reading PBX and then we again entered the **show call active voice EXEC** command to display the ERL level. The ERL represents the level of the echo coming out of the PBX in relation to the signal into the PBX. Notice that in Table 5 the ERL level is -14 dB, which means that, in relation to the signal going into the PBX, the echo is coming back at a level only 7 dB less than what was going in.

The International Telecommunication Union Telecommunication Standardization Sector (ITU-T) recommendation G.131 states that the ERL of a PBX should be greater than 15 dB. The ERL was substantially higher than an echo canceler can effectively nullify; therefore, the echo problem was with the Brussels PBX. To further verify the problem location, we adjusted the audio level into the PBX up and down. When we adjusted the audio level, the ERL remained constant.

We ran the same test with the FXS port plugged into the Ericsson PBX, as shown in Figure 16. Note the measured ERL of 19 dB. Table 6 shows output from the **show caller active voice** EXEC command. That call exhibited no echo.

Figure 16 ERL Test Using the FXS Port in Brussels

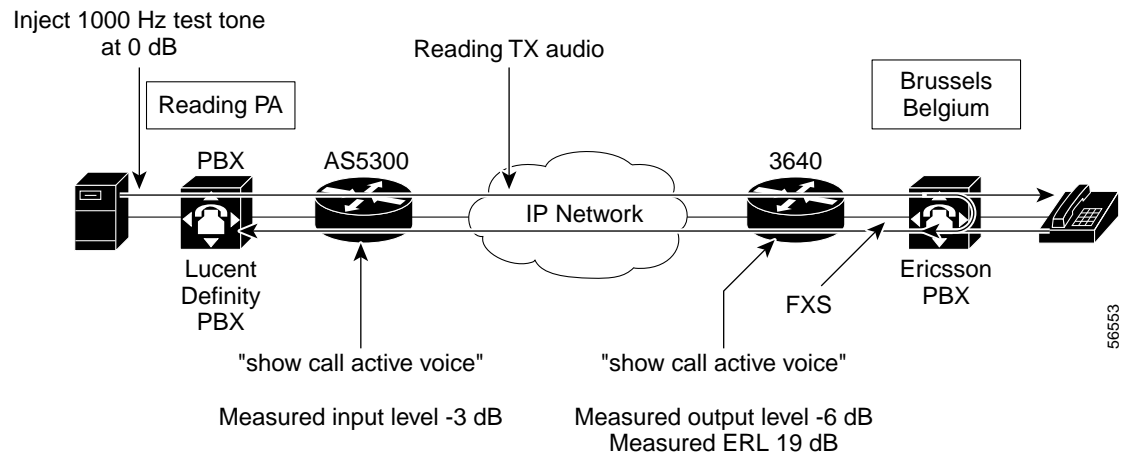


Table 6 show call active voice Command Output for FXS Test

Reading AS5300	Brussels 3640
<pre> Reading# show call active voice CoderTypeRate=g729r8 NoiseLevel=0 ACOMLevel=0 OutSignalLevel=-79 !This is the input level InSignalLevel=-3 </pre>	<pre> Brussels# show call active voice CoderTypeRate=g729r8 NoiseLevel=0 ACOMLevel=0 !This is the output level, R(out) OutSignalLevel=-7 !This is the input level, S(in) InSignalLevel=-27 InfoActivity=2 ERLLevel=20 !ERL = R(out) - S(in) !ERL = (-7) - (-27) = 20 dB !ERL is > 15 dB </pre>

Case Study Summary

The customer was satisfied with our testing results and decided to use our suggested workaround of using FXS ports, which appeared as CO trunks to the Brussels PBX, out of the Brussels Cisco 3640 router. This solution reduced some of the inward dialing flexibility of the network, but because all inbound calls were handled by an auto-attendant, no functionality was lost.

This case study illustrates the importance of educating customers about proper expectations of packet-based networks. Specifically, you should stress that the normal characteristics of packet-based networks may unmask preexisting problems in the TDM-based voice infrastructures.

This particular kind of echo problem—where the echo is PBX-based—is the easiest to solve. It is much more difficult to solve a case where the tail circuit is the PSTN and calls to only some locations are being affected. Not only are such cases difficult to troubleshoot, but they also present the challenge of your convincing the customer that the problem is in the PSTN, not the VoIP network. In reality, this type of echo problem is not related just to VoIP. It is essentially media-independent, and can exist wherever added delays in the network might exist.

As a general recommendation, an end-to-end implementation of a fixed-loss plan should be undertaken first, and the appropriate values applied to the gateway interfaces. Then, the field measurements can be applied to the port configurations. This process is preferred instead of adjusting the levels based on the field measurements without the fixed-loss plan in place.

Related Documents

- *Cisco IOS Voice, Video and Fax Configuration Guide*, Release 12.2
- Sheno, K. *Digital Signal Processing in Telecommunications*. Prentice Hall PTR; 1995
- *Voice over IP Fundamentals*, Cisco Press, 2000

