WHITE PAPER

Addressing Echo in VoIP Systems

Understanding and Monitoring Echo Cancellation for Optimal VoIP Performance

Echo is a troubling problem for telecom professionals. Most of us have suffered through a telephone call where we had to try to talk with a lot of echo on the wire. It's very distracting.

Voice over IP does not create echo. However, due to the temporal aspect of echo, VoIP systems can and do increase the amount of echo heard during a telephone conversation. Echo cancellation is one of the more complex parameters that need tuning to preserve VoIP call quality. And IP networking specialists are increasingly finding that they must understand and monitor echo cancellation to manage their VoIP system.

This white paper gives an overview of echo issues, provides a brief introduction to echo cancellation, and describes the echo metrics provided by the NetQoS® VoIP Monitor VoIP management solution.

Echo in Voice over IP Systems

Echo is a troubling problem that is well known to telecom professionals as a frequent source of end-user complaints. Most of us have suffered through a telephone call where we had to try to talk with a lot of echo on the wire. It's very distracting; most people find it hard to think straight, hear their own words repeated, and try to talk over them.

Voice over IP systems do not create echo, but due to the temporal aspect of echo, VoIP networks can and do increase the amount of echo heard during a telephone conversation. This paper gives an overview of echo issues and describes some of the solutions.

What Is Echo?

Echo is your voice coming back to you, as if you were repeating yourself. During a normal, two-person phone conversation, your voice is transmitted from your mouth to the ear of the person at the other end, and their voice is returned from their mouth to your ear. However, in any conversation, a certain amount of your own voice is also part of what you hear, whether you are talking face-to-face with someone who is sitting in your office, or talking to someone on the phone. This experience of hearing your own voice is not echo. Commonly called "sidetone," it's a normal aspect of talking and listening.

Your own voice becomes echo when it comes to your ear with a significant delay from the time you spoke. Sidetone is scarcely noticeable when the delay between your speaking and hearing is less than 25 milliseconds. Within that time window, the human brain does not perceive the sound as echo.

Echo, then, is very much a function of latency. Once you can hear your own voice more than 25 ms later, the possibility of perceiving it as echo arises. Twenty-five to 150 ms is a typical delay range for international telephone calls, which is why echo cancellation is necessary for such calls. Voice over IP calls don't actually create additional echo, but they also have a delay budget in the range of 150 ms to preserve audio quality, so VoIP systems commonly employ echo cancellation as well.

Among the various settings and parameters that need tuning to preserve optimal call quality in a VoIP system, echo cancellation is one of the more complex, least understood factors that IP networking specialists must address when they encounter VoIP for the first time. To understand echo cancellation and the metrics associated with it, we need to look at some other aspects of echo that affect the implementation of echo cancellation on voice gateways.

Echo Is Never Digital

Echo is always caused by the analog components in the telephony system. The digital stream of packets traveling in one direction of a VoIP call cannot "bleed into" the digital stream of packets in the other direction, nor are the packets played back at the receiving end of the call. The same is true for the digital parts of the Public Switched Telephone Network (PSTN): while the underlying electrical signals carrying the bits over the traditional switched telephone network are, indeed, analog, the corruption of those signals results in digital noise or other problems, but not in echo.

Strictly speaking, echo is never caused by voice over IP. In fact, what happens is that the longer delays introduced by all voice over IP systems reveal echo that was imperceptible with the shorter delays of the PSTN. By delaying existing echo signals longer, the VoIP network causes them to fall outside that 25 ms window and become audible to us.

What Causes Echo?

To reiterate what we've said so far, echo is the reflection or return of the speaker's voice to the speaker. It has an analog source, and it usually occurs at the far end of a conversation. Cisco Systems explains that the main two types of echo have different sources:

- Hybrid echo—Caused by an impedance mismatch in a hybrid circuit, such as a two-wire to four-wire interface, which allows the Tx signal to appear on the Rx path.
- Acoustic echo—Caused by poor insulation between the earpiece and the microphone in telephone handsets and hands-free devices.

At several places along a phone circuit, your voice can get into the return channel and come back to you. The first interface where echo may occur is at the transition between a 4-wire and a 2-wire interface. Analog telephone handsets are 2-wire devices. At some point in the path, perhaps in a local PBX, there is a hybrid interface that converts the network 4-wire interface to the 2-wire interface. Impedance mismatches here will reflect some of the energy back into the network, creating a potential source of **hybrid echo**.

Another common source of echo is the basic hardware: the mouthpiece of the phone at the far end may be too close to the earpiece, or it may be poorly insulated, so that your voice is heard and forwarded on the same return channel as the one on which the person at the far end is speaking. Therefore, the analog phone itself is a possible source of **acoustic echo**. Even more suspect these days is the speaker phone function of the phone at the far end of the call. Speaker phones broadcast the voice and simultaneously listen to the voices of the speakers in the room. It is all too easy for speaker phones—especially cheap ones—to send back some of the far end voice as part of what they are "hearing."

Delay is a necessary condition for echo, so it is rare for components that are close to the speaker—that is, on the speaker's side of the call—to cause echo. Even if part of the transmitted signal is being reflected back to the speaker by means of the return channel, the propagation delays are so brief that it will never be heard as echo. But several required components in every VoIP system exacerbate delay. The extra latency starts with the codec, which translates the analog signal into digital packets and places these packets on the wire. Latency is often increased by network components, such as routers, by geographical distance, and by jitter buffers in the IP phones. Any network congestion only makes it worse.

Because echo is rarely caused by a local component and has an analog source, the main suspects when an echo problem crops up are usually part of the "tail circuit" connecting the remote speaker to the PSTN. See Figure 1, below, for an illustration. The *voice gateway* device shown in the diagram allows analog phone calls from the PSTN to enter the IP network, and vice-versa.

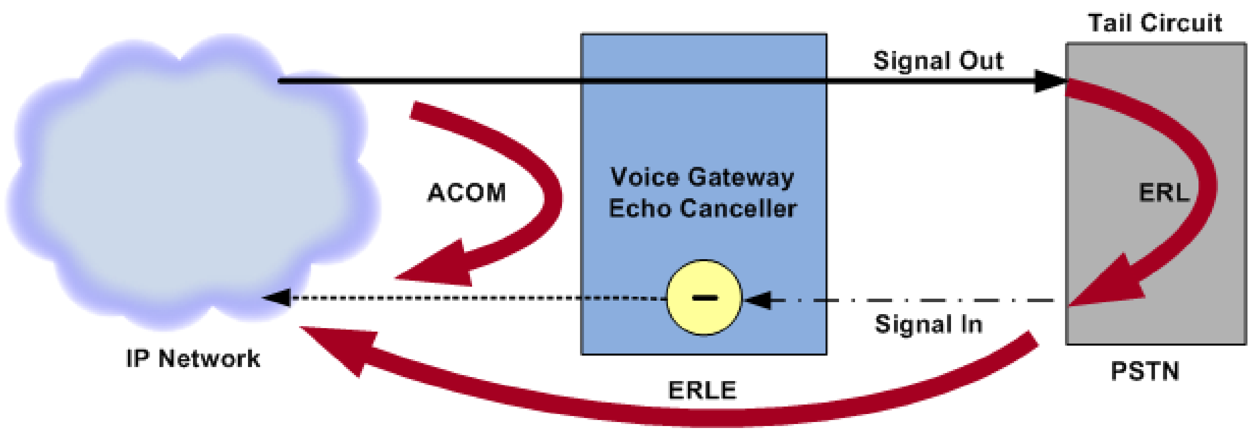

Figure 1. Echo Measurements & Echo Cancellation

Echo Metrics and Cancellation

If echo usually occurs at the far end of a call, echo cancellation is ideally done at the far end of the call as well. However, it can be done at any analog boundary in the network, with varying degrees of success. In one sense, you cannot eliminate an echo originating at the far end of the circuit if you don't control the equipment at the far end. However, you can minimize it using *echo cancellation* (or ECAN) devices.

One place where telecom or network engineers try to minimize echo in a VoIP system is at the voice gateway connecting an IP network to the PSTN. Even though the echo is originating at the far end of the call, echo cancellation here can work, within limits.

An echo cancellation device operates by comparing the signal going into the tail circuit with the signal coming back out. Basically, the ECAN remembers the signal pattern of the signal entering the tail circuit and examines the data exiting the circuit to see if it contains this pattern. If I say, "Is that good?" and you reply, "Fine" the ECAN is remembering "Is that good?" and looking for it mixed in with "Fine."

Two dimensions of the ECAN's work are important: echo strength, or volume, and echo delay.

Echo Volume

Not surprisingly, the louder the echo, the more distracting it is. Echo cancellation in effect consists of attenuating the amplitude of the echo part of the signal so that the echo is not easily heard. As some point, the echo becomes so quiet that it disappears, as compared to the foreground volume of the call.

Volume is typically measured in decibels (dB). You've probably seen the charts that show that a whisper is in the range of 20 dB while a jet engine or a rock band you don't like is around 120 dB.

Echo strength, which is equivalent to volume, is measured as echo return loss (ERL), which we discuss below. The ERL sets some boundaries on the ECAN's function because the echo must be weak enough to be distinguished from regular speech. In practice, this means that echo must be **6 dB quieter** than the speech it appears alongside for the ECAN to be able to suppress it. If the echo is louder than this, it typically falls into the volume range of the actual replying speaker and cannot be safely removed without endangering the conversation contents. So, if an echo were actually present in a telephone conversation and so loud that the ECAN could not distinguish it from the conversation, you would be more or less doomed to a very bad call with high levels of echo.

Echo Delay Times

Delay is the other key dimension of echo cancellation. As the ECAN examines the signal returning from the tail circuit, looking for a pattern that matches the signal sent into the tail circuit, it applies a *convergence time* algorithm to find the echo portion. Basically, this means that ECANs obey a time window that limits their operation. They look for the signal going into the tail circuit to be repeated in the signal coming back *within a specific time*, like 12 ms. (Typical values are from 8 to 64 ms.)

ECAN Limitations

Because the ECAN is actually modeling the response of the tail circuit mathematically, it starts each conversation with no knowledge of which part of the signal is legitimate speech and which part is echo and has to build its model. This usually takes a few seconds. After that, the ECAN's ability to discern and remove echo reaches its operating state for that call, and most echoes fade.

If a signal included multiple reverberations at different delays, like ripples in a pond, the ECAN would see perhaps the first two, but miss the one that arrived after 12 ms had passed. Typically, the energy level of each successive reverberation is reduced, so the resulting echo would potentially be quite soft anyway.

One other condition that can exceed the ability of an ECAN is distortion. If the echo itself is so distorted that it no longer matches the pattern seen as the signal was sent into the tail circuit, the ECAN will not be able to recognize it as echo.

ECAN Metrics

Taking into consideration the fact that echo cancellation is really all about echo *suppression* and not echo *prevention*, you've probably already deduced that echo is extremely common in any telephony environment. And as we mentioned above, it plagues VoIP systems due to their multiple sources of delay. It's therefore important to closely monitor the echo levels in a given VoIP system. Unless they're using a mobile phone and expecting slightly inferior service, users will complain stridently if echo becomes noticeable during their phone calls.

When monitoring echo levels in a phone system, or when troubleshooting a reported issue with echo, telecom professionals apply a well-known set of metrics that express the effectiveness of echo cancellation. These metrics also apply to VoIP.

Echo Loss Metrics

Echo return loss, or ERL (see Figure 1), is a measurement applied to echo that measures the **loss of volume** between the original signal and the echo. In other words, an ERL of zero is the worst case; it means that the echo is fully as loud as the original signal. As the delay grows longer, an ERL of up to 55 dB (and at least 6 dB) is necessary to soften the echo enough to avoid distraction. Applying this principle to VoIP, with long delays (up to 150 ms for acceptable call quality), the echo part of the signal needs to be at least 15 dB quieter than the original voice in order to avoid the perception of echo. So high ERL values are good. Note that ERL and related values mentioned in this document are all part of various ITU standards.

The amount of echo suppression is measured in the ERLE, or echo return loss enhancement, which expresses how much quieter the ECAN was able to make the echo, in dB. In other words, ERLE is a measure of what the actual echo canceller is able to accomplish. While ERL measures the "native" echo coming from the tail circuit or far end of the call, ERLE is the amount of additional echo attenuation the ECAN provides. Taken together, they exactly equal the *ACOM* value.

Another standard metric applied to echo cancellation, the ACOM value is the view of echo from the IP side of the echo canceller. As defined in ITU G.168, ACOM is the "combined" echo return loss through the system—the attenuation of echo from all possible means.

ACOM resembles ERL: it is a measure of the degree to which an echo signal has been attenuated. The difference is that ACOM is measured on "our side" of the ECAN device (as shown in Figure 1, above).

Therefore, because it includes all sources of echo loss in each direction of the circuit, ACOM is the best gauge of echo strength. Like ERL, ACOM should be high.

Signal In and Signal Out Metrics

The Signal In and Signal Out metrics are useful for testing a circuit to see whether it is introducing echo. Cisco Systems' echo testing procedures use these metrics to measure the effects of altering the signal strengths when tuning echo levels.

Signal In is the audio signal traveling in the direction of the IP network, measured as it enters the ECAN from the tail circuit (shown as "Signal In" in Figure 1). It contains echo that needs to be canceled. Signal Out is the audio signal coming out of the ECAN and going into the tail circuit (shown as "Signal Out" in Figure 1)—from the IP network to the PSTN. The Signal Out stream contains an estimation of the amount of echo in the audio stream. Both metrics are measured by the gateway on the PSTN side of the gateway's ECAN.

ERL is used along with Signal In and Signal Out to tune the echo canceller. Remember that ERL must be at least 6 dB to distinguish the echo portion of the signal from the voice itself. To enhance this difference and enable echo cancellation, the voice gateway ports perform "input gain" and "output attenuation" on the signals. Input gain is performed at "Signal In" in the diagram, before the gateway's ECAN sees the echo, and output attenuation is performed at Signal Out, after the gateway's ECAN has seen and cached the original signal. Thus, these metrics provide reference points for adjusting the overall strength of the signals to enable echo cancellation. They are used in echo troubleshooting to adjust the signal gain or attenuation performed by the gateway so that:

```
Signal Out – Signal In > 6 dB.
```

Addressing Echo Problems in a VoIP System

The **NetQoS VoIP Monitor** product ensures the availability and performance of your voice over IP (VoIP) system by passively monitoring and reporting on VoIP call setup and call quality metrics. NetQoS VoIP Monitor has several features that make it uniquely capable of measuring echo levels in voice over IP telephone calls, sending alerts when echo crosses a threshold, and helping track down the source of an echo problem.

The VoIP Monitor system continually monitors both call setup performance and call quality. ACOM is reported for all call legs that include a voice gateway. One of the default call quality threshold settings instructs the system to raise alerts if ACOM values are too low. Like ERL, ACOM should be high; a VoIP Monitor alert is sent by default when ACOM drops to 15 dB. We stated earlier that ECANs use a limit of 6 dB to distinguish echo and avoid suppressing the actual contents of a conversation. If the ERL value is too low, the echo signal that returns to the gateway might be too loud, falling within 6 dB of the conversation signal. With the default threshold settings, the VoIP Monitor Management Console flags call quality as excessively bad when ACOM measurements, or the sum of ERL + ERLE, fall to 6 dB. ACOM measurements are available per voice gateway, so you can easily spot a gateway where excessive echo is causing call quality to deteriorate.

In addition to threshold monitoring, NetQoS VoIP Monitor also reports ERL and ACOM values in real time for "watched" calls. The VoIP Monitor Call Watch feature collects additional diagnostic data from selected VoIP calls. During a Call Watch, the VoIP Collector actively gathers detailed quality metrics for all calls made to and from a selected IP phone by polling the phone and any associated gateway, if a call to the PSTN is watched. The collected data is presented in a series of charts, which are displayed and updated in real time, as the watched calls are in progress.

ACOM

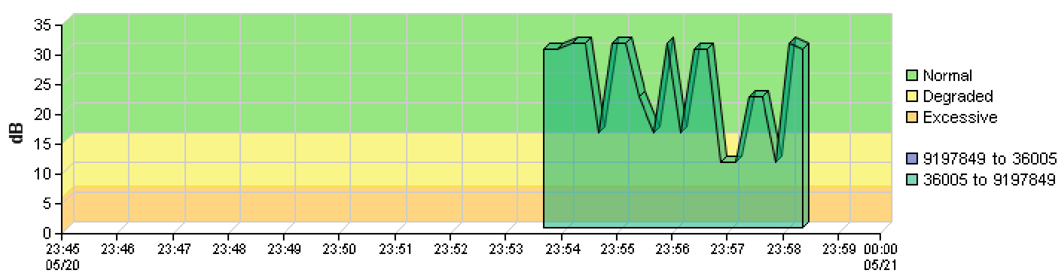

Figure 2. Real-time ACOM measurements are compared to the Degraded and Excessive performance thresholds.

The Call Watch Report also includes Signal In and Signal Out metrics. In addition to contributing to the information-gathering phase of troubleshooting a call-quality problem, these values can help engineers tune their echo cancellers. In order to tune the ECAN, test signals are introduced into a voice gateway circuit from an IP phone. Gain and attenuation are applied through the command line at the incoming and outgoing gateway interfaces until the desired values of ERL and ERLE are obtained.

Cisco Systems' typical echo testing procedures expect you to gather these metrics by frequently running a command from the command-line interface during an active phone call. Now you can simply start up a VoIP Monitor Call Watch for a phone connected to the gateway under test, make a phone call to the PSTN to or from that phone, and take a look at the metrics as they are reported in real time. The signal levels are presented in a graph format that is continually updated during the tuning procedure.

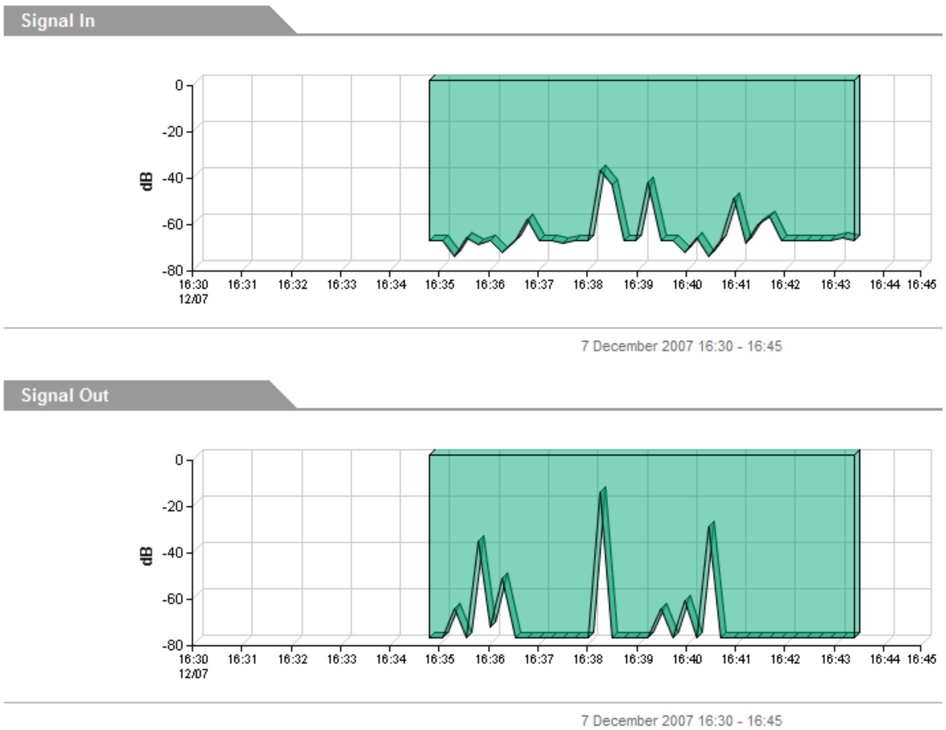

Figure 3. Signal In and Signal Out metrics are useful for tuning an echo canceller.

Summary

Echo is a difficult problem and not one that can be readily solved without access to the end-to-end circuit. Echo cancellation devices at the boundary of the VoIP system attempt to reduce echo that is of no concern in the PSTN so that it cannot be heard in the IP telephony environment. NetQoS VoIP Monitor provides a quick, overall rating of the success of this effort in the form of ACOM values for all gateway calls made on the monitored network, as well as some additional details, such as ERL, Signal In, and Signal Out, that are essential for troubleshooting echo problems.

Some Helpful References

http://www.netqos.com/solutions/voip_monitor/index.html
http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a0080149a1f.shtml
http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a008019ab88.shtml#topic6
http://www.netqos.com/offer/voip%20ebook/index.html
http://www.performance-edge-journal.com/subscribe.html
http://www.cisco.com/web/CA/events/pdfs/IP_Telephony_Advanced_Topics_And_Troubleshooting.pdf
http://www.itu.int/rec/T-REC-G.168-200408-S/en

About NetQoS VoIP Monitor

NetQoS VoIP Monitor is a network-based call setup and call quality monitoring product that tracks the call quality users experience, provides alerts on call performance problems, and isolates performance issues to speed troubleshooting and MTTR. NetQoS VoIP Monitor is integrated with the NetQoS Performance Center so you can monitor VoIP quality of experience while managing network quality of service, from a single, Web-based console. With NetQoS VoIP Monitor, you can assess the performance of your Cisco® Unified Communications Manager (CallManager) IP PBX by tracking, evaluating, and reporting on key metrics without deploying server agents or probes. For every call, NetQoS VoIP Monitor reports on user call quality and the underlying network performance metrics associated with it. NetQoS VoIP Monitor also breaks out performance data from the IP and Public Switched Telephone Network (PSTN) legs of calls that pass through voice gateways traveling to endpoints in the PSTN. This data lets you know if it is your network that is responsible for less-than-optimal call quality.

About NetQoS

NetQoS is the fastest growing network performance management products and services provider. NetQoS has enabled hundreds of the world's largest organizations to take a Performance First approach to network management—the new vanguard in ensuring optimal application delivery across the WAN. By focusing on the performance of key applications running over the network and identifying where there is opportunity for improvement, IT organizations can make more informed infrastructure investments and resolve problems that impact the business. Today, NetQoS is the only vendor that can provide global visibility for the world's largest enterprises into all key metrics necessary to take a Performance First management approach. More information is available at **www.netqos.com**.

NetQoS Global Headquarters
5001 Plaza On The Lake
Austin, TX 78746
Phone: 512.407.9443
Toll-Free: 877.835.9575
Fax: 512.407.8629

NetQoS EMEA
1650 Arlington Business Park
Theale Reading, RG7 4SA
Phone: + 44 (0) 118 929 8032
Fax: + 44 (0) 118 929 8033

NetQoS APAC
NetQoS Singapore Representative Office
Level 21, Centennial Tower
3 Temasek Ave., Singapore 039190
Phone: + 65 6549 7476
Fax: + 65 6549 7001

Website: www.netqos.com

E-mail: sales@netqos.com

WP rev1 20070816