HEISON CHAK

echo in VoIP systems

Heison Chak is a system and network administrator at SOMA Networks. He focuses on network management and performance analysis of data and voice networks. Heison has been an active member of the Asterisk community since 2003.

heison@chak.ca

HELLO, HELLO, HELLO, HELLO, . . .

Many of us have had experience with echo, whether it was speaking loudly in an empty stadium, shouting in the great outdoors in front of a mountain, hearing your own voice while talking on a telephone, or even singing Karaoke with the echo feature enabled on a mixer. It all boils down to the persistence of a sound after its source has stopped.

Although there are two main types of echo, acoustic and electronic, both can be generalized as reflection points being introduced that cause the audible echo.

In the case of an acoustic echo, sound waves get bounced against objects. Since these objects may have very different physical characteristics and can be varying distances from the source of the sound, the resulting sound arrives at the speaker's ears at different times and amplitudes. It is important to understand that echo may only be heard by the speaker. Consider a speaker standing at one end of a tunnel or an empty stadium, shouting out to his listener. Only the speaker will hear his own voice being reflected by the surroundings. Similarly, if a climber is shouting out to her friend on top of a mountain, the friend may be able to hear the message clearly with no echo. Yet the climber may hear her own voice being reflected by the mountain and surrounding reflection objects even after she has stopped shouting.

Electronic echo reflection points found in telephony systems are generally related to analog-to-digital conversation, 2–4-wire hybrid, and impedance mismatch. In a telephony system, when a caller experiences echo, it's likely that the reflection point is at the remote end of the conversation. In fact, the remote party of the conversation may find the communication sounding perfectly fine, simply because he or she is beyond or behind the reflection point, much like the mountaineer yelling from the bottom of a mountain to her friend at the top. This is commonly known as the far-end problem.

Echo in VoIP Systems

When sound is applied to the handset transmitter of a VoIP device, it is digitized, processed, encoded, and transmitted via the network interface as IP packets. The near-end telephone intentionally sends some of the electrical signal from the transmitter to the receiver. These signal components, known as side-tone, are used to simulate the natural expectation of a user to hear his or her own voice while speaking. Since side-tone originates locally, it carries only a small amount of delay, too short to be perceived by a user.

In a pure VoIP system (i.e., an IP-IP call), there is no chance of echo being introduced until the packets arrive at the far-end telephone. Electrical crosstalk in the far-end telephone can occur and couple some of the received signal in its transmit path, leading to leakage of the speaker's voice in the return path. The handset on the far-end telephone can generate an acoustic output that is coupled back to the transmitter as well. The combined electrical and acoustic signals get digitized, processed, encoded, and transmitted to the near-end phone, where they appear at the receiver. Since these packets carry much longer delay than those generated locally on the near-end devices, they become a more noticeable side-tone with a much longer delay—in other words, echo.

The perceived quality of the connection can often be impacted by the amount of delay these signals carry as well as by their amplitude. Typically, when the signal delay exceeds 15–20 ms, the user perceives the delay, and the experience becomes more unpleasant if the delay has a high amplitude. Fortunately, analog telephone systems have low enough latency that signals are generally not perceived as echo.

VoIP packets are known to have low resilience to delay and latency; the effect is even more undesirable when it comes to echo. Besides typical transmission delay of IP networks, VoIP packetization intervals (framing duration of received audio for transmission) and jitter buffers can contribute further delay.

VoIP and POTS

In traditional analog telephone systems, whenever a 2–4-wire hybrid is used, some receive signals will leak back into the transit path, owing to the imperfection of coupling. Similar coupling issues can be seen if the user's telephone impedance mismatches that of the phone service provider. Because of the short delay characteristics of local calls on the PSTN (Public Switched Telephone Network), signals created by such 2–4-wire hybrid reflection generally are not perceived as echo. For long distance and international calls, telephone service providers feature echo cancellers to take care of the added transmission delay. An echo canceller is a software algorithm that tries to remove the portion of the signal caused by the transmitter picking up receive output of the telephone; it does so by attenuating delay samples in uncompressed form (e.g., 12-bit PCM). Generally, echo cancellers are disabled for local calling, as they are simply redundant.

When VoIP systems are communicating with POTS phones, echo can originate from a number of places:

- VoIP to local numbers: The PSTN provider has its echo canceller disabled, as it assumes that local calls have short latency; it has not factored in the possible added delay inherited from an IP network. (A VoIP user may experience echo, mainly electrical.)
- Poor echo canceller on a gateway: An inexpensive media gateway usually will not feature a hardware echo canceller; echo cancellation performed in software may not converge fast enough to cope with the ever-changing characteristics of a call, especially on a busy system. (A VoIP user may experience echo, mainly electrical.)
- Underpowered echo canceller: With dedicated hardware, if an echo canceller has a relatively short echo cancellation capacity, it will not be able to remove echo effectively. (A VoIP user may experience echo, mainly electrical.) A typical T1/PRI echo canceller is capable of removing undesired signal up to 128 ms across all channels.
- Impedance mismatch on a gateway: The FXO (Foreign Exchange Office) interface on a media gateway may not couple with the PSTN, so

the voice of the POTS user gets reflected into his or her return path. (The POTS user may experience echo, mainly electrical.)

- Poorly designed hybrid on a gateway: The quality of each 2–4-wire hybrid that comprises a telephone call tends to correlate with the amount of echo a user perceives. Good circuit design can ensure better coupling, minimizing reflection points. This can be compensated for by an echo canceller, if applied correctly. (POTS and VoIP users may experience echo, mainly electrical.)
- Hands-free telephones: These tend to have higher output amplitude on the speaker and higher sensitivity, because a user is usually sitting further away from the phone than with a handset telephone. As a result, the acoustic coupling of the speaker to the microphone is high and can result in echo. With the added latency of VoIP, echo from hands-free telephones can become quite strong. (POTS and VoIP users may experience echo, mainly acoustic.)
- PC to PC: During a pure IP-IP call, when packets arrive, electrical and acoustic echo may be introduced. If a PC-to-PC call involves a speaker and a microphone that have the sensitivity characteristics just described, echo becomes inevitable. The same applies to PC-to-PSTN calls. (VoIP and POTS users may experience echo, mainly acoustic.)

Both acoustic and electrical echo can be compensated for by proper application of echo cancellation. If an echo canceller is applied too far from the source, the undesired signal may outrun the capability of the canceller (e.g., a canceller with 128 ms of capacity will not remove a delayed copy of someone's voice of 400 ms). In the case of a call between a VoIP device (either a PC or an IP handset) and the PSTN, if the call is identified by the PSTN as a local call, the echo canceller may be disabled, making the call vulnerable to the sorts of echo problems just described.

VoIP applications are often challenged by echo problems, impacting call quality. It is important to remember that echo is usually a far-end problem. If echo cannot be avoided, placing an echo canceller closest to the source can ensure more effective compensation, giving faster convergence and less DSP-intense operations.

Although echo is often viewed as a far-end problem, this may not always be the case. Consider a telephone call where one end is on a mobile phone. The conversation starts out fine but deteriorates as soon as a generic headset is used on the mobile phone (in which case only the mobile user experiences echo). As soon as the generic headset is swapped with one that is specifically designed for the phone, the echo problem diminishes. In this case, no changes are made to the system except that a near-end device has been replaced. It is possible that there is impedance mismatch between the generic headset and the mobile phone, which suggests a near-end problem.

It is difficult to convince the user of a poorly designed telephone that his phone is coupling the signal it receives and sending the delayed signal back to the other party on a well-designed phone. In effect, the user of the well-designed telephone will hear a delayed copy of her own voice, while the conversation sounds just fine to the user of the poorly designed telephone. The situation can be misleading. The user of the poorly designed telephone believes that his telephone is working well, whereas the user of the well-designed telephone may wonder whether her telephone set is the source of echo, when just the opposite is true.

REFERENCE

http://microtronix.ca/echo_problems.htm.