White Paper | March 2009

radisys®

Improving Audio Quality in VoIP Conferencing Services

CONTENTS

Executive Summary

Audio conferencing services based on circuit-switched networks and audio bridging equipment have provided hosted conferencing users with a benchmark for pricing and quality in voice communications. While next generation networks based on Voice over Internet Protocol (VoIP) technology introduce economic benefits with new feature capabilities for conferencing service providers (CSPs), they also present new technical challenges in maintaining acceptable voice quality. Delivering good voice quality is an important requirement in any VoIP conferencing system, as poor voice quality will increase the costs associated with customer churn, while impacting the bottom line by reducing revenue growth prospects.

Voice Quality Enhancement (VQE) encompasses an integrated set of features designed to overcome common audio quality problems in VoIP conferencing services, including noise, packet loss and echo. A comprehensive VQE solution also measures VoIP quality metrics, which are used in ongoing voice quality measurement associated with service level agreements.

Many features inherent in a VQE solution require sophisticated digital signal processing algorithms. The rapid, scalable execution of these algorithms dictates a product specifically designed for real-time IP packet processing. Fortunately, in a next-generation VoIP audio conferencing architecture, a network element already exists with carrier-class real-time IP packet processing power. And that network element is the IP media server (see sidebar).

This white paper describes an innovative approach towards VoIP audio quality improvements in VoIP conferencing applications where VQE capabilities are integrated into the IP media server itself. By performing both voice quality enhancement and VoIP media processing in the same network element, CSPs and their application developers can realize economic and technical benefits in their delivery of multi-party VoIP conferencing services with superior audio quality.

What is an IP Media Server?

The world is filled with numerous enhanced telecommunication services and capabilities such as unified communications, multimedia conferencing, IP contact centers and ringback tones. And while the application and signaling logic of all these applications has many differences, the underlying requirements for processing the actual voice and video media streams share many similarities, such as playing an audio or video clip, collecting digits from a phone or mobile terminal, bridging multiple signals into a conference mix or transcoding media streams using different encoding standards.

An IP media server is a common, shared IP media processing resource used in a broad range of IP audio, video, fax and speech applications inside a next-generation VoIP and IMS network. Because it is a critical, centrally deployed resource in the core of the network already involved in sophisticated real-time IP packet processing, an IP media server is also the ideal technology and network location to address common sources of poor audio quality such as noise, packet loss or echo.

VoIP Audio Quality—Challenges and Solutions

The three most common sources of VoIP audio quality problems in a VoIP or IMS network are noise, packet loss and echo. This section discusses each of these VoIP audio quality challenges and describes the conceptual solutions to overcome quality problems.

1. Audio Noise

Gone are the days when people were constrained to quiet office and residential environments. Today, with mobile phones and the Internet, people are calling from their cars, airports and from just about anywhere, and these environments are flooding the mouthpiece with all kinds of unwanted sounds that ultimately get onto the call. Audio noise comes from background sounds like dogs barking, televisions blaring, planes flying overhead and people conversing nearby. Making matters worse, callers are using laptops and mobile phones, which are typically saddled with marginal equipment such as low cost earphones and microphones.

This section describes a combination of mechanisms that reduce and help manage the disturbing effect of audio noise: noise gating, noisy line detection and noise reduction.

Noise Gating

Noise gating is a simple yet effective mechanism to reduce background noise, as shown in Figure 1. When no speech is detected on a line, its signal is attenuated (e.g., decreased amplification), which prevents unnecessary noise from being inserted into a VoIP recording or conference mix. Noise attenuation is configurable, so the conferencing application can avoid making the signal unnaturally quiet when the noise gate is applied to an audio signal.

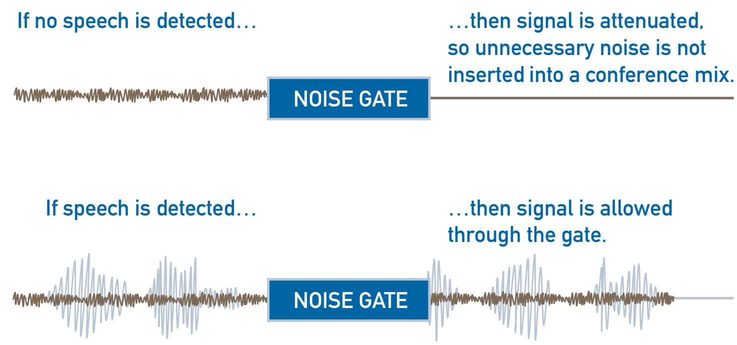

Figure 1. *Noise Gating*

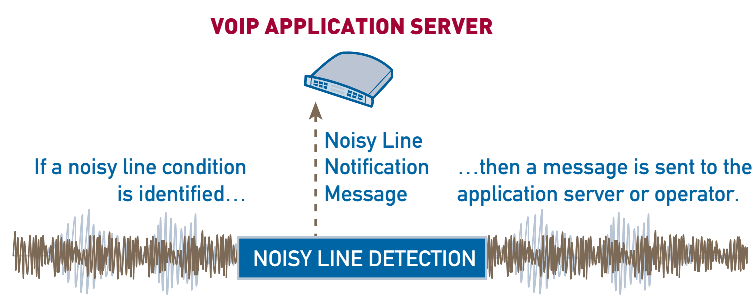

Figure 2. *Noisy Line Detection*

Key Benefits of Noise Gating:

- Reduces background noise using a simple yet effective mechanism
- Supports configurable attenuation

Noisy Line Detection

There are times on a conference call when some lines are very noisy and disrupt the productivity of the entire call. The responsible caller may be unable to mute, could walk away or fail to realize that he or she has an offending line. Noisy line detection measures the noise on audio ports and sends a noisy line notification message to the VoIP application server if a predefined threshold is exceeded, as shown in Figure 2. A second message is sent if the noise subsequently falls below the threshold.

Noisy line detection actively looks for and distinguishes four noise conditions, which are described in Table 1.

The noisy line detection feature does not automatically apply a treatment to a noisy line. Instead, noisy line detection reports noisy line conditions to the application server, which can then notify the conference moderator (e.g. the conferencing system could "whisper" to the moderator that Participant X has a noisy line). The application or moderator then can choose to mute the noisy line or leave the noisy line unchanged if, for example, the offending user is an important contributor to the conference.

Key Benefits of Noisy Line Detection:

- Notifies the application server of noisy line conditions, initiating possible corrective action
- Enables quick remedial action by the application server or the operator (e.g., mute line)

Noise Reduction

While a noise gating function described earlier provides a relatively simple solution to eliminating noise when no speech is detected, noise reduction goes a step further by using digital processing techniques to remove the noise and leave the important speech signal intact, as shown in Figure 3. This provides benefits in many VoIP applications, such as removing noise from VoIP audio recordings or noisy caller lines in a conference mix.

The amount of attenuation applied against the noise is a function of the noise reduction algorithm and can vary for the duration of the call. Generally, an attenuation of 10 to 20 dB is sufficient to dampen the noise without making the speaker sound unnaturally quiet.

Noise reduction algorithms require lots of processing power, and many VoIP lines today have adequate quality without noise reduction. So a solution that indiscriminately applies noise reduction to all VoIP lines, including lines with acceptable quality, would result in a wasteful use of processing resources, which lowers overall system capacity and/or raises the price per port. A more sophisticated approach is to dynamically identify and rank the noisy lines, and then selectively apply noise reduction to the worst offenders.

Noise Conditions	Examples
Background Noise	Environment sounds such as those from airports, fans, machinery, traffic or crowds of people talking
Impulsive Noise	Loud, short duration sounds, like lightning strikes, other electrical noise and gunfire
Continuous Signal Noise	Tonal signals, such as "music-on-hold," someone humming humming and DTMF tones caused by pressing phone keys
Low Signal-To-Noise Ratio (SNR)	Someone speaking too softly to be clearly heard over any noise in the signal

Table 1: *Noise Types Distinguished by Noisy Line Detection*

Figure 3. *Noise Reduction*

Key Benefits of Noise Reduction:

- Filters out noise without impacting the speaker's signal
- Reduces noise continuously, whether speech is detected or not

2. Dropped Packets

The Internet is an amazing network of interconnected computers, but it's not perfect. The network employs the IP protocol, which does not guarantee packet delivery. Hence, when IP networks get busy or congested, packets can get lost or delayed. While lost packets are not critical for many data applications, packet loss in real-time VoIP services can cause significant audio quality problems. Without special technology to compensate for dropped packets, the result is an abnormal audio signal that might sound 'choppy.'

Packet Loss Concealment

Packet loss concealment is a technique for replacing audio from lost or unacceptably delayed packets with a prediction based on previously received audio. Whereas any voice repair technology would have difficulty recovering from extreme packet loss in

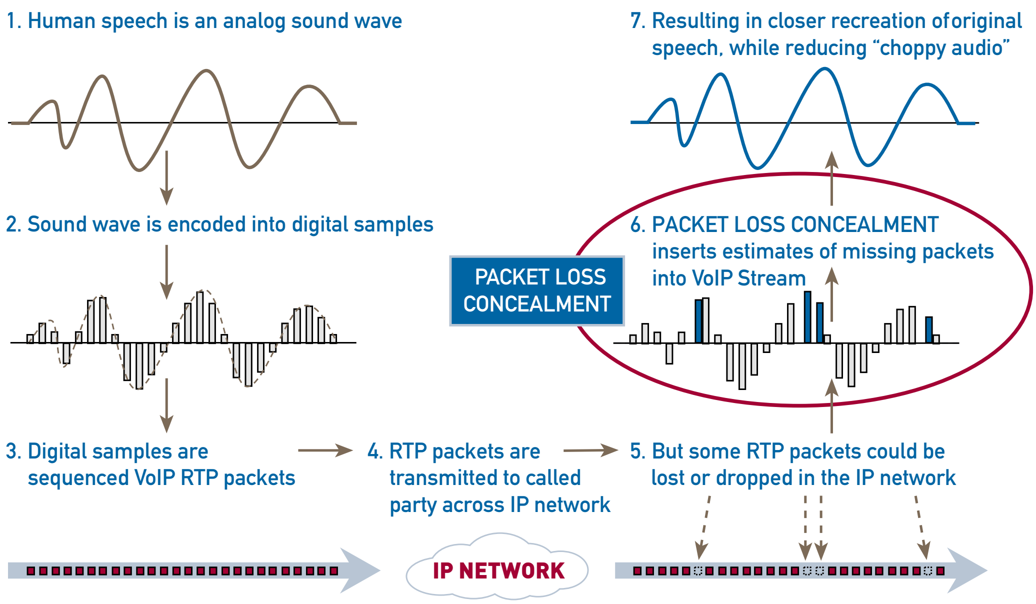

Figure 4. *Packet Loss Concealment*

abnormal conditions, packet loss concealment is designed to perform intelligent restoration of lost or delayed packets for a large majority of congested network scenarios.

Figure 4 depicts the process of packet loss concealment. Spoken words produce an analog sound wave (1), which is encoded into digital samples (2) and sequenced into Real-time Transport Protocol (or RTP) audio packets (3). These packets are transmitted across an IP network (4), however some packets may be lost or dropped (5). Packet loss concealment inserts estimates of the missing packets (6) resulting in a speech rendering that closely resembles the original (7).

Key Benefits of Packet Loss Concealment:

- Softens any breaks in the voice signal
- Reduces the occurrences of choppy audio

3. Acoustic Echo

Anyone who has yelled into a large cave or in a canyon has heard the echo of their own voice. The concept is the same in telephony: in a point-to-point audio conversation where a receiver's terminal equipment has improper echo isolation, the sender might hear a muted, but irritating echo of their own voice after a delay.

Figure 5. *Acoustic Echo Cancellation (AEC)*

An acoustic echo is created when sound emanating from the receiver's speaker (e.g., handset or speakerphone) is transmitted back by the receiver's microphone. This is depicted in Figure 5, where the Sender (on the left) transmits a signal to the Receiver, and an acoustic echo is created when some speech energy 'bounces back.' In a VoIP conferencing application, all participants will hear an echo except for the guilty party with the device causing the echo. Since nobody can quickly answer the basic question, "Who's causing the echo?," troubleshooting echo issues in a VoIP conference call can be difficult and frustrating. There are many sources of acoustic echo, including:

- Poor room acoustics
- Marginal microphones for soft terminals
- Low quality cellular handsets
- Deficient echo control in the terminal device itself

Acoustic Echo Cancellation

Acoustic echo cancellation (AEC) technology is designed to detect and remove the sender's transmit (Tx) audio signal that bounces back through the receive (Rx) path. By removing the echo from the signal, overall speech intelligibility and voice quality is improved.

AEC in a VoIP network is particularly challenging. In a traditional voice network, once a voice circuit is established through the PSTN, the round-trip echo delay is constant. However, in a VoIP network, packet delay is a variable, hence the echo delay is also a variable for the duration of the call, which makes the echo cancellation algorithms in any VoIP quality improvement product more complex and processor-intensive than an equivalent echo cancellation solution in a circuit-switched network.

Key Benefits of Acoustic Echo Cancellation:

- Removes a sender's audio echo from the receive path
- Addresses variable packet delay inherent in IP networks

Voice Quality Metrics

Technology to remove audio quality impairments in a VoIP network is an important part of any solution. But along with the functions to improve VoIP quality, service providers also need a standard, objective way to measure voice quality in order to accurately monitor performance levels and uphold service level agreements (SLAs) with customers.

Voice quality metrics can be divided into three groups: packet, audio and acoustic echo cancellation (AEC). All statistics are captured for each call leg of a conference call to help with granular troubleshooting of audio quality problems and performance measurement. Packet statistics measure performance with respect to packet throughput, loss and delay, while audio statistics measure speech and noise power levels. AEC statistics measure echo delay and echo cancellation performance. Examples of packet, audio and AEC statistics are shown in Tables 2, 3 and 4, respectively.

Packet Statistics	Calculation	Service Level Agreement
Duration	Overall lifetime of a port	Activation time
Packets and Octets Sent	Real-Time Transport Protocol (RTP) packets and payload (Octets exclude headers and padding)	Throughput
Packets and Octets Received	RTP packets and payload	Throughput
Packet Loss	The ratio of lost packets to total packets transmitted	Performance (loss)
Average Jitter	The difference between the RTP timestamp (time sent) and the actual arrival time of the packet	Performance (delay)

Table 2: *Packet Statistics*

Audio Statistics	Calculation	Service Level Agreement
Peak Amplitude	The peak audio amplitude over the lifetime of the connection	Speech volume consistency within a conference mix
RMS Speech Power Min/Max/Ave	Root mean square (RMS) speech power over the lifetime of the connection and the most recent talk-burst	Individual speech volume level
RMS Noise Power Min/Max/Ave	Noise power over the lifetime of the connection and the most recent talk-burst	Individual noise level
DC Offset	The running average of the DC offset, a potential source of noise in a VoIP system	Systemic noise level
Signal to Noise Ratio Min/Max/Ave	The difference between the speech power and noise power	Relative noise level

Table 3: *Audio Statistics*

AEC Statistics	Calculation	Service Level Agreement
Echo Flag	Turned on if echo detected; indicates which conference endpoints are generating echo	Echo detection
Bulk Echo Delay	Round trip delay time between original signal and the beginning of the echo received on a return path	Performance (echo)
Echo Return Loss (ERL)	Difference in dB between the original signal amplitude and its echo (the smaller the ERL, the larger the echo)	Performance (echo)
Echo Return Loss Enhancement (ERLE)	Difference in dB of the echo level before and after echo cancellation	Performance (echo)

Table 4: *Acoustic Echo Cancellation (AEC) Statistics*

Key Benefits of Voice Quality Metrics:

- Provides objective measurements for administering service level agreements (SLAs)
- Facilitates the troubleshooting of audio quality issues in the network

Voice Quality Enhancement for VoIP Conferencing Applications

Voice Quality Enhancement (VQE) encompasses an integrated set of features designed to improve VoIP quality and generate statistics needed for ongoing performance monitoring. This requires sophisticated digital signal processing algorithms that perform rapid real-time IP packet processing, a key component in next-generation VoIP audio conferencing architecture. As such, VQE can be deployed in an existing IP media server, which provides the requisite carrier-class real-time IP packet processing power.

IP media servers, also known as the Multimedia Resource Function (MRF) in an IMS architecture, are specifically designed to deliver real-time IP media processing as a common, shared resource for a broad range of VoIP and IMS applications in a next-generation network.

They also deliver real-time processing of codec algorithms, transcoding of codecs and sophisticated audio mixing for conferencing applications. Since media server and VQE tasks are interrelated and require the rapid execution of IP packet processing algorithms, it makes sense to integrate the functions of both into a single network element.

Radisys is the leading supplier of IP media server products and technologies for service providers, telecom equipment manufacturers and application developers. Radisys Convedia media servers are specifically designed to process thousands of real-time IP audio and video packet streams simultaneously in high-availability configurations, making them the ideal network platform for integrated and cost-efficient voice quality improvements for multi-party VoIP telecommunication services.

Integrated Voice Quality Enhancements (VQE) on Radisys Convedia Media Servers

Voice Quality Enhancements (VQE) is a new feature set that has been integrated with the existing IP media processing features of the Radisys Convedia media server. Through this integration, Radisys has introduced an innovative, cost-effective solution for all real-time IP media processing requirements in a single network element.

Specifically, VQE introduces all of the features outlined in this document, including:

- Noise Gating
- Noisy Line Detection
- Noise Reduction
- Packet Loss Concealment
- Acoustic Echo Cancellation
- Voice Quality Metrics

VQE features can be applied to individual lines or groups of lines associated together through audio mixing (i.e. a conference call). In addition, VQE configuration and default parameters can be set at a global level through the Radisys Convedia media server management system. Alternatively, VQE global configuration parameters can be overwritten by the controlling VoIP application server to deliver granular line-by-line quality control through a standards-based control interface.

Figure 6 represents a simplified VoIP audio conferencing infrastructure. The conferencing application server orchestrates the signaling and call control to terminate all the VoIP call legs onto the IP media server. It then commands the IP media server, using an open standards-based control protocol, to perform IP media processing, including:

- Feature-rich VoIP audio mixing functions
- Interactive voice response (IVR) dialogs, like playing prompts and collecting digits for user PIN or feature control
- Audio recordings, such as participant entry announcements or conference recordings

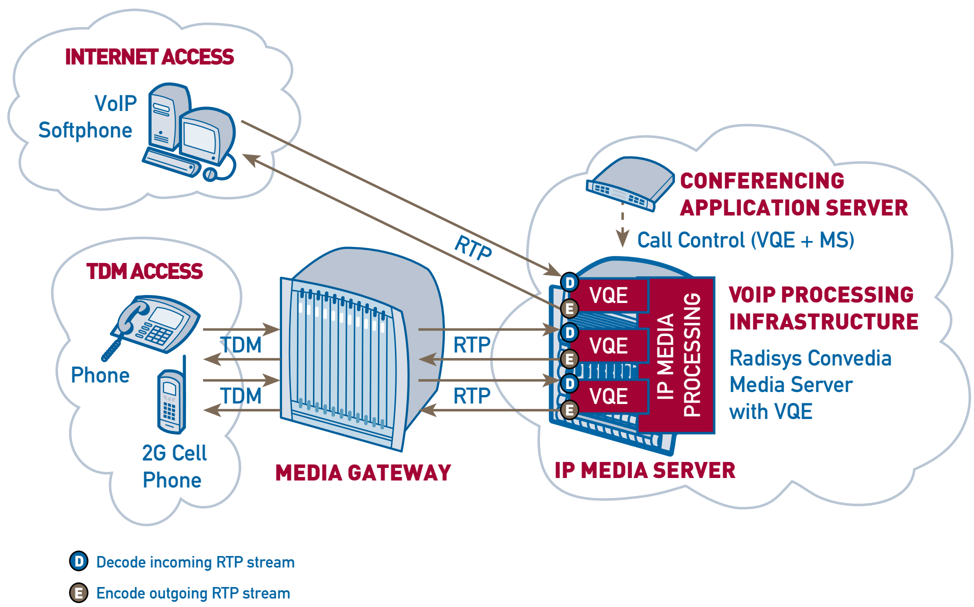

Figure 6. *VQE Functions Integrated in the IP Media Server*

Transcoding is the process of decoding and encoding RTP media streams to and from an internal digital signal format conducive to rapid real-time digital processing inside the media server, as shown in Figure 6. Each transcoding step in a voice path adds a small delay, typically 10-20 ms. IP media servers already perform carrier-scale transcoding in order to execute the audio mixing algorithms required for a VoIP audio conferencing service. Adding VQE into the IP media does not introduce any additional transcoding steps, so end-to-end call path delay is only slightly increased due to the additional internal processing of the VQE algorithms.

Contrast this converged solution with an alternative approach where VQE is performed somewhere else in the VoIP conferencing infrastructure. If a conferencing service provider were to use this segmented approach, two or more transcoding steps would be required—one for each network element performing VQE, as well as the transcoding required by the media server—resulting in an increased end-to-end delay that actually exacerbates voice quality problems.This alternative approach also adds unnecessary complexity in VoIP call leg setup and teardown, because the conferencing application server would need to coordinate call setup and teardown across multiple control interfaces: one or more interfaces for the VQE network element(s) and one for the IP media server.

VoIP quality improvements, in particular noise reduction and acoustic echo cancellation, are processing-intensive functions that can drive up the amount of network equipment required and the cost per port for the overall audio quality solution. Radisys has responded with a unique configuration capability that allows the network operator to specify the percentage of media server ports that should receive noise reduction and echo cancellation treatment. First, Radisys media servers dynamically monitor and measure noise and echo conditions across all lines, and then they rank the lines based on detected quality problems. Based on a configurable percentage, the packet streams on the worst lines are identified for full echo cancellation and noise reduction processing. This unique configuration flexibility allows the operator to strike a balance between improving VoIP audio quality on the worst lines only and minimizing the overall VQE solution cost.

In a VoIP conferencing application, implementing voice quality technology in the IP media server has unique advantages over other VQE approaches. The IP media server is the only media processing element with real-time knowledge of the configuration and state of every VoIP conference, such as the individual settings for each call leg (e.g. gain, mute, 1-way only) and which VoIP call legs are associated with each conference mix. The IP media server can take advantage of this knowledge to apply VQE intelligently, thereby optimizing the use of IP packet processing resources for VQE functions. For example, a 1-way listen-only connection for a supervisor in an IP contact center conference mix does not require echo cancellation because a 1-way media stream cannot generate an echo. Another example is a participant temporarily muted in a conference call, which would also permit echo cancellation, packet loss concealment or noise reduction to be temporarily turned off.

A summary of the benefits of the integrated VQE approach are listed in Table 5.

Benefits of Integrated VQE Solution	Improvement
Optimizes IP Packet Processing Resource Utilization	Applies VQE functions intelligently due to its unique knowledge of endpoint relationships in a conference mix
Lowers Overall Call Delay	Reduces transcoding operations performed on call legs
Decreases Network Complexity	Minimizes the number of network elements to manage
Simplifies Application Development	Allows the application server to control both voice quality and IP conference mixing features using a single, common standards-based interface
Lowers Overall Solution Cost	Doesn't require a separate VQE network element, which reduces CAPEX and ongoing OPEX

Table 5: *Benefits of Integrating VQE Functions into the IP Media Server*

Conclusions

The economic benefits of VoIP audio conferencing services are compelling, but consumers won't migrate to VoIP unless they experience good audio quality. There are many sources of poor VoIP quality in an audio conference mix, including noise, lost packets and acoustic echo. It's imperative that next-generation VoIP audio conferencing deployments address these challenges and deliver good voice quality to accelerate service adoption while increasing customer satisfaction. Otherwise, poor voice quality will impact the bottom line by reducing revenue growth prospects and increasing costs associated with customer churn.

Through the integration of the new VQE capabilities with existing IP media processing features, Radisys Convedia media servers now offer conferencing service providers and VoIP application developers an innovative, cost-effective solution to all real-time IP media processing requirements in a single network element. Radisys integrated VQE technology cost-effectively overcomes voice quality challenges associated with an IP packet network, which can increase end-user satisfaction in next-generation VoIP conferencing services and help accelerate the growth of VoIP and IMS service revenues.

Corporate Headquarters
5435 NE Dawson Creek Drive
Hillsboro, OR 97124 USA
503-615-1100 | Fax 503-615-1121
Toll-Free: 800-950-0044
www.radisys.com | info@radisys.com